



Cite this: *Mol. Syst. Des. Eng.*, 2026, **11**, 195

Exploring the temperature stability of CRISPR-Cas12b using molecular dynamics simulations

Yinshao Jia,^a Katelynn Horvath,^b Santosh R. Rananaware,^a Piyush K. Jain^{acd} and Janani Sampath^a  

The thermal stability of CRISPR-Cas nucleases is a critical factor for their successful application in ‘one-pot’ diagnostic assays that utilize high-temperature isothermal amplification. To understand the atomistic mechanism of stabilization in a previously engineered variant of the thermostable BrCas12b protein, we performed all-atom molecular dynamics (MD) simulations on the wild-type and mutant forms of apo BrCas12b. High-temperature simulations reveal a small structural change along with greater flexibility in the PAM-interacting domain of the mutant BrCas12b, with marginal structural and flexibility changes in the other mutated domains. Comparative essential dynamics analysis between the wild-type and mutant BrCas12b at both ambient and elevated temperatures provides insights into the stabilizing effects of the mutations. Our findings offer comprehensive insights into the important protein motions induced by these mutations. These results provide insights into thermal stability mechanisms in BrCas12b that may inform the future design of CRISPR-based tools.

Received 30th July 2025,
Accepted 24th October 2025

DOI: 10.1039/d5me00140d
rsc.li/molecular-engineering

Design, System, Application

In this work, we employ all-atom molecular dynamics (MD) simulations to unravel the mechanism behind the enhanced thermostability of an engineered BrCas12b variant. This is crucial for developing improved one-pot CRISPR-based diagnostic assays for infectious diseases, which rely on thermophilic BrCas12b for sensitivity. We assessed the system’s structural dynamics by tracking global unfolding at high temperature *via* RMSD and RMSF at the domain level. Subsequent analyses, including principal component analysis (PCA) and residue/domain-level cross-correlation maps, revealed subtle structural changes and increased flexibility within the PAM-interacting (PI) domain. These analyses also illuminated how mutations alter the characteristic “open–close” motions and introduce novel stabilizing interactions. These detailed insights into the dynamic effects of specific mutations are vital for the rational design and optimization of Cas12 proteins, enabling the development of more sensitive and stable diagnostic platforms. This work not only enhances current diagnostic capabilities but also opens exciting avenues for future therapeutic applications involving engineered Cas12 proteins.

Introduction

CRISPR (clustered regularly interspaced short palindromic repeats) – Cas (CRISPR-associated protein) systems have emerged as revolutionary tools in the field of molecular biology. This system, originally a part of bacterial immune defenses, has been adapted to perform a myriad of tasks in genetic engineering, due to its precision and versatility.^{1,2} One of the latest advances in CRISPR/Cas involves using Cas proteins as components in molecular diagnostics, with applications ranging from infectious disease detection to

genetic disorder diagnosis and cancer detection.^{3,4} Cas9 and dCas9 have been used to establish diagnostic systems for the detection of single nucleotide polymorphisms (SNPs) and featured nucleic acid sequences in pathogenic strains.^{5,6} The discovery of indiscriminate cleavage of other non-specific single-stranded RNAs upon target binding, also referred to as *trans*-cleavage or collateral cleavage activity of CRISPR-Cas13a (C2c2) in 2016 brought a new era of CRISPR/Cas-based biosensing.⁷ Cas13a was utilized to establish a comprehensive diagnostic platform, SHERLOCK, which was further developed for rapid diagnosis of infectious diseases.^{8–10}

In addition to Cas13, Cas12a (Cpf1) was found to have collateral cleavage activity on other single-stranded DNAs. Specifically, upon binding to the intended nucleotide target, a non-target or collateral cleavage is triggered in the Cas protein which is distinct from specific cleavage activity.^{11,12} Collateral activity can be utilized to cleave DNA reporter,

^a Department of Chemical Engineering, University of Florida, Gainesville, FL, USA.
E-mail: jsampath@ufl.edu

^b Department of Chemical and Biomolecular Engineering, University of Connecticut, Storrs, CT, USA

^c Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL, USA

^d Health Cancer Center, University of Florida, Gainesville, FL, USA



thereby magnifying the fluorescence signal and the sensitivity of the detection platform; this Cas12a-based nucleic acid detection platform, DETECTR, was reported by Chen *et al.*¹¹ At the same time, Li *et al.* discovered a similar collateral cleavage effect in Cas12a and developed the nucleic acid sensing assay HOLMES.¹³ While the specificity and the simplicity of these nucleic acid detection assays are advantages, the requirement of separate steps to amplify the sample before the detection to increase sensitivity poses significant challenges, including possible carryover contamination, longer processing time, complexity in handling, often an increase in the overall cost of the assay. To address this, the integration of Cas proteins with a pre-amplification step, such as reverse transcription loop-mediated isothermal amplification (RT-LAMP), has shown promise, as the integration allows for a sensitive nucleic acid detection method.^{14,15} However, the high-temperature nature of RT-LAMP, typically conducted at 65 °C, denatures most native Cas proteins, hindering proper integration to create a simple one-step process.

With the utilization of the Cas12b (C2c1), a thermophilic RNA-guided endonuclease from class II type V-B CRISPR/Cas, Wang and others showed that with the same cleavage activity, pairing with LAMP (loop-mediated isothermal amplification) results in a one-step assay for convenient target detection.^{16,17} More recent efforts including the usage of AapCas12b, which is derived from *Alicyclobacillus acidiphilus*, and BrCas12b, which derived from *Brevibacillus sp.*, as part of the one-step diagnostic assay for the detection of nucleic acids including SARS-CoV-2. Based on SHERLOCK, Joung *et al.* integrated RT-LAMP with AapCas12b, thereby achieving a one-step detection that exceeds the sensitivity of CDC RT-qPCR.¹⁸ Nguyen *et al.* simplified the detection process and reduced the carryover contamination by a one-pot assay developed using the newly identified thermostable Cas12b ortholog BrCas12b (with a melting temperature of 62 °C).¹⁹ Further optimization of the single-pot assay was achieved by the combination of *de novo* design through point mutations and experimental validation, resulting in a thermostable BrCas12b with a melting temperature of 67 °C (referred to as eBrCas12b) which does not degrade at the high operating temperatures (65 °C) of RT-LAMP.²⁰ While these modifications have shown promising results, the underlying mechanisms of how these mutations contribute to increased thermal stability is not fully understood, nor do we know its influence on the overall dynamics of the protein, which could affect its cleavage process.

Molecular dynamics (MD) simulations is a powerful tool to gain insights into the structure and dynamics of proteins, including the effects of mutations at the atomic level.^{21,22} MD simulations have been implemented previously to understand fundamental activation mechanism of CRISPR-Cas9 systems, revealing high conformational flexibility of the HNH domain and providing insight into the dynamic behavior of Cas9. Palermo *et al.* performed a series of MD simulations on CRISPR-Cas9 systems to explore the

conformational plasticity of the Cas9 protein and its interactions with nucleic acid during DNA binding.²³ These studies revealed that the Cas9 protein undergoes significant conformational changes during the process of nucleic acid binding and processing. Meanwhile, the 'closure' of the protein relies on the highly coupled and specific motion between different domains. Saha *et al.* investigated the structural dynamics of different Cas12a orthologs and found that DNA binding induces changes in Cas12a conformational dynamics, which leads to the activation of REC2 and Nuc domains that enable the cleavage of nucleic acids, particularly the large amplitude motion of Nuc that allows the protein to open and close. The dynamics of REC2 and Nuc domain are highly coupled, and REC2 can act as a regulator of Nuc, which is similar to HNH in Cas9 system.²⁴ While CRISPR-Cas9 has been extensively analyzed using MD simulations,²⁵ there remains a significant gap in understanding other Cas proteins, such as Cas12. Considering the unique collateral cleavage mechanism of Cas12 and the difference in cleavage mechanism between Cas12b and Cas12a,^{26,27} a thorough analysis of Cas12b is needed for rationally designing variants of the protein. Particularly, MD simulations can provide insights into BrCas12b and eBrCas12b, shedding light on how mutation improves the stability of BrCas12b at high temperatures and how the new point mutations alter the essential dynamics of BrCas12b. This information is important to design superior CRISPR analogs for future diagnostic and therapeutic applications.

We focus on the BrCas12b ortholog of Cas12b and employ all-atom MD simulations to probe its structure and dynamics before and after mutation. By conducting a series of high-temperature unfolding simulations, we reveal the unfolding process of wild type and mutant BrCas12b, highlighting the structure and flexibility change of specific domains containing mutations, which substantially improve the stability and activity at high temperatures. Additionally, we explore the essential dynamics of wild-type and mutated-type BrCas12b at both ambient and elevated temperatures together with cross-correlation analysis, examining the internal movements within the protein and interactions between its domains, revealing the change of essential dynamics as well as decreased correlation between domains upon mutation. These results contribute to a fundamental understanding of structural dynamics of BrCas12b and the influence of mutations on the protein function.

Methods

System setup

MD simulations were performed on two systems, wild-type (WT) and mutant (MT) BrCas12b, both in apo form (no nucleotide bound to the protein). The structure of WT BrCas12b was modeled using AlphaFold v2.3.1.²⁸ MT BrCas12b was also modeled using AlphaFold, with four point mutations (R160E/F208W/N524V/D868V) reported in a



previous study by Nguyen *et al.*²⁰ which outperformed other tested mutation combinations and displayed high-temperature stability and activity. Domain assignment (Fig. 1a and S1) was made based on sequence alignment on the existing crystal structure of BthC2c1/BthCas12b (PDB: 5WTI) available in the RCSB PDB databank and domain organization proposed by Wu *et al.*²⁹

All simulations were conducted using GROMACS 2021.3 software³⁰ with GPU acceleration. CHARMM36m³¹ forcefield and TIP3P model was used to represent the protein and waters, respectively. Protein was solvated in a cubic box with the edge at least 2.0 nm from protein, resulting in box sizes ranging 15.6 to 17 nm with a total of ~380 000 to ~520 000 atoms based on the conformation of protein adopted. Cl⁻ ions were introduced to neutralize each system. Hydrogen bonds were constrained during the simulation using the LINCS³² algorithm. The cutoff for short-range electrostatic and van der Waals interactions was set to 1.2 nm. The particle-mesh Ewald³³ method was used to compute long-range electrostatic interactions; fast Fourier transform grid

spacing was 0.15 nm. van der Waals interactions were smoothed between 1.0 to 1.2 nm using a force-base switching function.³⁴ Constant temperature was maintained using the Nosé–hoover³⁵ algorithm. For the constant-pressure simulations, the Parrinello–Rahman barostat (1 bar, 2.0 ps coupling constant) was used. The time step for numerical integration was 2 fs.

The systems, prepared as described above, were first subjected to steepest descent energy minimization with a maximum force of 1000 kJ mol⁻¹ nm⁻¹, followed by two stages of 100 ps NVT equilibration. In the first stage, heavy atoms of the protein were restrained, and the solvent molecules were equilibrated. Initial velocities were randomly assigned according to Maxwell–Boltzmann distribution at 300 K. In the second stage, all atoms were equilibrated with the velocity adopted from the first stage and their position restraints were removed. Subsequently, a 1000 ps constant pressure and temperature (NPT) simulation was conducted at 300 K. The final outputs were used as the starting point for the production simulation.

With the protein structure obtained from AlphaFold, a 200 ns long simulation was first set up at 300 K to equilibrate the system further. The sampling of the simulation was confirmed with all-to-all root mean square deviation (RMSD) matrix (Fig. S2) and was observed to have two distinct stages. To increase reproducibility and reliability, four replicates of each system were performed. The starting configuration of each replica was randomly selected from this trajectory between 100–200 ns. Following equilibration, a high-temperature unfolding simulation protocol was implemented, details can be found in our previous study.^{22,36} Simulations were conducted at both 300 K and 400 K, for 200 ns. To further sample the configuration space and capture the essential dynamics of BrCas12b, one replicate of WT and MT BrCas12b simulations at both 300 K and 400 K was extended to 1 μ s.

Analysis

Simulation trajectories were analyzed using GROMACS tools and visualized using VMD. The secondary structure of the protein and its evolution along the simulation were assigned using the inbuilt *timeline* tool of VMD with the STRIDE algorithm. The predicted circular dichroism (CD) spectra was generated by the PDBMD2CD web server using the resulting structure from each simulation trajectory.³⁷ Root mean square deviation (RMSD) of the whole protein was calculated by fitting to the backbone of the initial structure, whereas the per domain RMSD was calculated by fitting to the domains' initial backbone structure during the production run.

Principal component analysis (PCA) was done on the 1 μ s simulation to capture the essential dynamics of the simulated systems. The analysis was based on the atomic coordinates of the C α atoms, which serve as the input features for the analysis. The workflow involved two key steps. First, to remove global rotational and translational

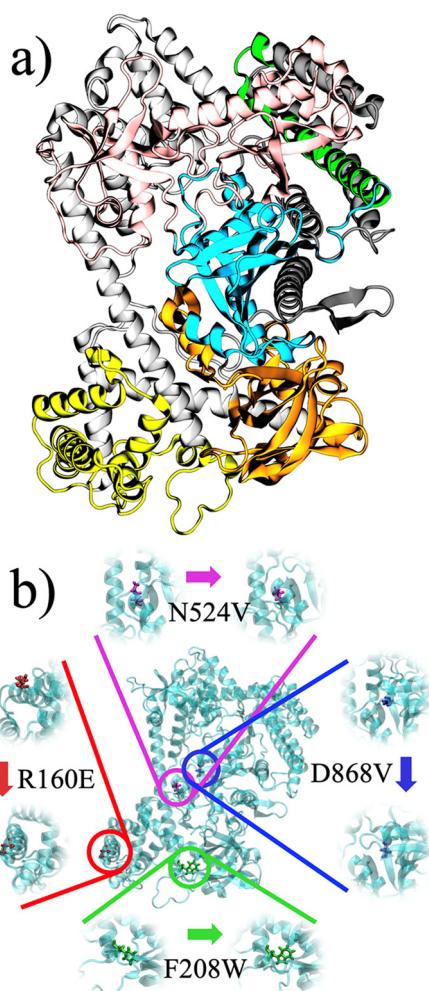


Fig. 1 Schematic of (a) domain assignment of BrCas12b with residues colored according to domains; detailed legend can be found in the SI (Fig. S1) and (b) location of mutated residues within the protein.



motion, each trajectory was aligned to the backbone of the initial structure using a least-squares fitting procedure. Second, a covariance matrix was constructed from the 3 N -dimensional vector of the Ca atomic coordinates from this aligned trajectory. The resulting covariance matrix was diagonalized to obtain the eigenvectors as well as the corresponding eigenvalues. Each eigenvector represents a principal component (PC) that depicts a mode of movement, with the corresponding eigenvalue indicating the extent of the motion along the eigenvector. The process of arranging eigenvectors based on their eigenvalues reveals that the first principal component (PC1) aligns with the greatest amplitude motion of the system. The movement of the system along PC1 is commonly known as “essential dynamics”.³⁸ In this work, we project the 1 μs long trajectories in 2D defined by the first two principal components (PC1 and PC2) to characterize the conformational space sampled by simulation. The covariance matrix was constructed by using *gmx covar* and the projection was done using *gmx ana eig*, both commands inbuilt in GROMACS. Visualization of the protein motion was done using VMD.

The dynamic cross-correlation matrix (DCCM) was constructed to capture correlated and anti-correlated motions within BrCas12b. The covariance matrix constructed during PCA analysis was post-processed to generate the DCCM according to eqn (1).³⁹

$$\text{CC}_{ij} = \frac{\langle \Delta\vec{r}_i(t) \cdot \Delta\vec{r}_j(t) \rangle}{(\langle \Delta\vec{r}_i(t)^2 \rangle \langle \Delta\vec{r}_j(t)^2 \rangle)^{1/2}} \quad (1)$$

where $\Delta\vec{r}_i(t)$ represents the position of Ca atom at time t . The cross-correlation score (Cs_i) is calculated for each residue according to eqn (2). The intra-domain score $\text{Cs}_i^{\text{intra}}$ is calculated when residues i and j belong to the same domain, and the inter-domain score $\text{Cs}_i^{\text{inter}}$ is calculated when they belong to different domains. Specifically, when residue i and j belong to different domains a and b , respectively, the calculated Cs_i is the cross-correlation score coefficient of residue i with respect to domain b . In this way, the cumulative Cs_i of all residues i belonging to domain a will give the cross-correlation between domains a and b .^{23,40}

$$\text{Cs}_i = \sum_{i \neq j}^N \text{CC}_{ij} \quad (2)$$

To determine the significance of changes in the root mean square fluctuation (RMSF) for each residue, a two-way analysis of variance (ANOVA) was performed. This test allowed for the simultaneous assessment of the effects of Genotype (WT *vs.* MT), Temperature (300 K *vs.* 400 K), and the interaction between genotype and temperature. Given that this analysis involved performing a statistical test for every residue in the protein, a false discovery rate (FDR) correction (Benjamini-Hochberg procedure) was applied to the resulting p -values to control for the multiple comparisons problem. A residue was considered to have a statistically significant change if its FDR-adjusted p -value was less than 0.05.

To compare the average root mean square deviation (RMSD) of specific protein domains, an independent two-sample *t*-test was used. The input for this test was the mean RMSD value calculated from the stable, equilibrated portion (the final 100 ns) of each of the four independent replicate simulations for both the WT and MT systems. The resulting p -value was used to determine if the difference in the mean domain RMSD between the WT and MT groups was statistically significant, with a significance threshold of $p < 0.05$.

Results and discussion

Global structure assessment

To understand how point mutations enhance the thermostability of BrCas12b, we carry out MD simulations at 300 K and 400 K to capture the unfolding process of the WT and MT BrCas12b. Fig. 2 shows the backbone root mean square deviation (RMSD). The RMSD was calculated by fitting the backbone carbon atoms and using the first frame of the simulation trajectory as the reference structure. At 300 K, we observe a relatively small increase in RMSD (<0.5 nm) within 200 ns for both WT and MT, indicating a minor deviation from their initial configuration, considering the relatively large structure of the protein (Fig. 2a and b). We also compute the change in hydrogen bonds and solvent accessible surface area (SASA). Both WT and MT have a relatively stable hydrogen bond network, with an average total hydrogen bond number (HBN) of 840 (± 15) at 300 K (Fig. S3). At this temperature, the total solvent accessible surface area (SASA) of both WT and MT fluctuates at around 600 nm² (Fig. S3). Secondary structure analysis (Fig. S4) shows a small fluctuation of the helix and sheet components of both proteins, in agreement with the predicted circular dichroism (CD) spectra (Fig. S5), and shows comparable secondary structure content calculated from the available crystal structure of Cas12b family (Table S1). The overall structural features of the apo BrCas12b for both WT and MT at room temperature are comparable, with a steady hydrogen bond network and corresponding SASA, and the structure deviates marginally from its starting conformation.

At 400 K, we see that the RMSD of both WT and MT increases with time compared to 300 K, as expected. This indicates a change in the protein structure at 400 K compared to its initial structure, with both proteins showing approximately the same RMSD of 1.0 nm over the last 100 ns of simulation time. Experimentally, it was observed that the MT protein's melting temperature is higher than that of the WT protein, and thus more thermostable.²⁰ However, the overall RMSD does not indicate a major difference in the thermostability of the two proteins. This could be because the experiments were performed at around 60–65 °C (~ 335 K) while the simulations are conducted at 400 K for efficiency. Hydrogen bonds were observed to decrease for WT and MT but with a small change of 30 fewer hydrogen bonds at 400 K for both (Fig. S3). Interestingly, SASA showed a



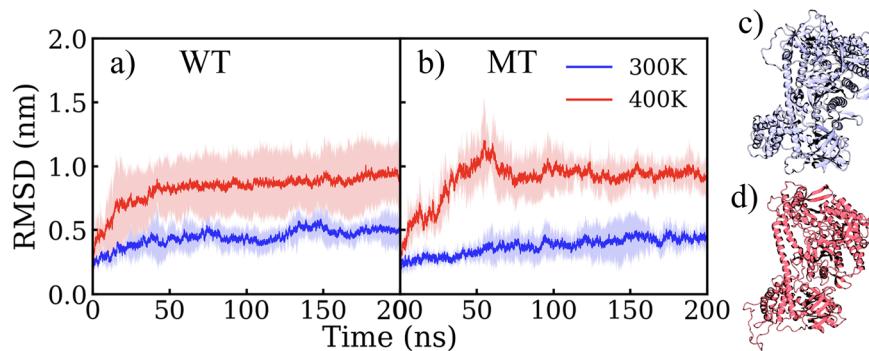


Fig. 2 Backbone root mean square deviation (RMSD) of BrCas12b (a) wild type (WT) and (b) mutant (MT) as a function of simulation time, and snapshots of BrCas12b wild type at 200 ns from (c) 300 K and (d) 400 K. The solid lines indicate the mean RMSD value averaged over four independent trials, while the shaded region indicates the standard deviation.

decreasing trend at elevated temperatures for both WT and MT, with MT showing a slower decrease at the beginning, and a lower overall surface area at the end (Fig. S3). The decrease in SASA indicates the collapse and shrinking of the protein structure at high temperatures, which differs from other proteins that we have investigated before, like insulin and lysozyme, that show an increase in SASA at elevated temperatures. Considering the thermophilic nature of the BrCas12b, high temperatures trigger hydrophobic residues to come closer to prevent their exposure at the protein-solvent interface. The decreased SASA at high temperatures has been observed in other thermophilic proteins.^{41–43} Secondary structure analysis (Fig. S4) shows a decrease in helix and sheet components in both WT and MT, while predicted circular dichroism spectra (Fig. S5) also indicate the loss of secondary structure, suggesting that parts of the protein are unfolding.

To understand the effect that the four point mutations (R160E, F208W, N524V, D868V) have on the flexibility of the residues, we analyzed the root mean square fluctuation (RMSF) on the backbone atoms of each residue; this is shown in Fig. 3. To distinguish changes from expected simulation variance, we performed a per-residue two-way ANOVA with a false discovery rate (FDR) correction, assessing the impact of genotype (WT vs. MT) and temperature (300 K vs. 400 K).

First, we examined the local effects at the mutation sites themselves. At 300 K, the charge-reversal mutation of residue R160E leads to a significant increase in the fluctuations of the protein residues close to it, with an RMSF value of 1.2 nm for the mutated protein, compared to that of 0.6 nm for the wild-type protein. We ascribe this to the change of the residue charge from positive to negative after mutation. When comparing snapshots from regions containing mutated residues (Fig. S6), we noticed that the introduction of mutation R160E resulted in the extension of the α -helix (formed by residue 151 to 161) into a slightly longer α -helix (formed by residue 151 to 165) and thus led to a conformational change relative to the wild-type, with the formation of more extended helix also confirmed by secondary structure analysis. This newly formed longer

α -helix, together with a nearby α -helix (residue 140 to 147), adopts an open conformation and extends towards the outside of the protein regarding the initial inward position. Such an open conformation in MT presents a more significant structural fluctuation overall than in WT. The other three mutations to hydrophobic residues show a slight trend towards lower RMSF. However, our ANOVA reveals that none of these local changes at the four mutation sites were statistically significant after FDR correction. At 400 K, the difference of RMSF between WT and MT was lower, with residue R160E having a marginally larger value in MT compared to WT.

Interestingly, the primary impact of the mutations was observed at distal locations. The statistical analysis identified significant dynamic changes in regions far from the mutation sites, such as the 600–800 amino acid region, where 4 residues (628–621) exhibit a statistically significant difference

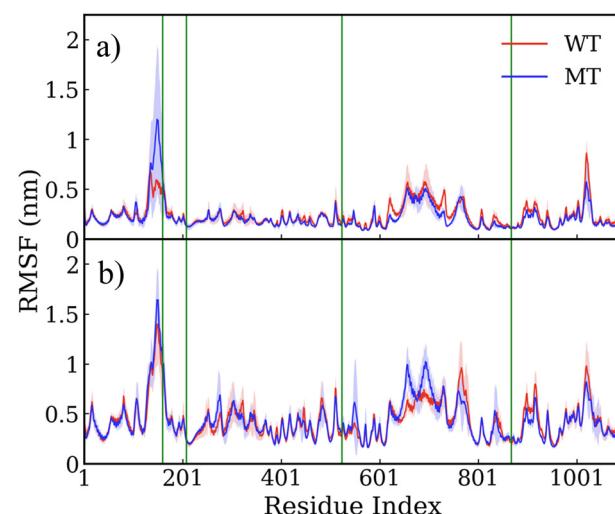


Fig. 3 Root mean square fluctuation (RMSF) of BrCas12b wild type (WT) and mutant (MT) from (a) 300 K and (b) 400 K, with four green vertical lines highlighting the mutation points R160E, F208W, N524V, and D868V. The solid lines indicate the mean value averaged over four independent trials, while the shaded region indicates the standard deviation.



in flexibility between the WT and MT systems (adjusted *p*-value < 0.05). This finding suggests an allosteric-type mechanism, where the engineered mutations collaboratively induce measurable changes in the protein's long-range dynamics.

Given the relatively large protein size, it is hard to discern structural changes brought about by just four mutations (effectively 0.4% of the entire protein) using these metrics alone. Detailed analyses at a local residue level can shed more light on the underlying mechanism of stabilization stemming from the mutations, which we discuss below.

Structural analysis of mutated domains

To understand the effect of high temperature on the stability of MT and WT BrCas12b, we decompose the protein into domains as discussed in the methods section. Particularly, the BthC2c1 structure features a two-lobed composition, with an α -helical recognition (REC) lobe and a nuclease (NUC) lobe. The REC lobe is formed by a PAM-interacting (PI) domain, REC1 and REC2 domains, and an extended α helix known as the bridge helix (BH). In the NUC lobe, there is an OBD domain, a RuvC domain, and a domain with undetermined functions, referred to as the "UK" domain. The RuvC domain in the NUC lobe, which is composed of three separate RuvC motifs (RuvC I-III), interacts with the REC2 domain in the REC lobe. The connection between the RuvC domain and the REC1 domain is primarily facilitated through the UK domain.²⁹ The domains with mutated residues are PI, REC1-II, OBD-II, and RuvC-II.

RMSD analysis was carried out for each decomposed domain. Domains with mutated residues are highlighted in Fig. 4, and the RMSD for all domains are presented in Fig. S7. RMSD is calculated by first performing root-mean-square fit on backbone atoms of a specific domain to its initial structure. At 300 K, the RMSD of most domains in both WT and MT are low, indicating small deviations from their initial structure. Exceptions to this are domains PI (Fig. 4), REC2, and UK-II (Fig. S7), in which the RMSD increases over time

compared to the other domains. The higher RMSD for these three domains even at 300 K indicates that they are more likely to undergo conformational changes compared to the other domains, especially PI, which shows an RMSD near 0.5 nm in the last 50 ns of simulation time. The higher RMSD of these domains also coincides with the higher RMSF, indicating higher structural flexibility. Previous studies on other CRISPR-Cas proteins like Cas9 and Cas12a suggest the high flexibility of PI domain and the C-term (part of UK-II here), is substantially reduced after being bound to nucleic acid.^{23,24} At 400 K, the trend of RMSD is different across multiple domains. At 400 K in WT, there are several domains, including REC1-I, BH, and RuvC-III (Fig. S7), which do not undergo dramatic conformational change and show a comparable RMSD to 300 K. The RMSD of domains OBD-I (WED-1), REC1-II, OBD-II, RuvC-I, RuvC-II, and UK-I show a slight increase but do not exceed 0.5 nm. Domains PI, REC2, and UK-II, which show high conformational change at even 300 K have the highest change at 400 K as well. Of particular importance is domain PI, which shows a dramatic structural change at the beginning of the high-temperature simulation, indicated by the steep slope of the RMSD, before 20 ns. After the mutation, no apparent difference was observed in most domains except for domain PI (Fig. 4). Though there is a sharp rise in RMSD at the start of the simulation, the RMSD plateaus at \sim 0.9 nm for MT, lower than the value for WT which is \sim 1.3 nm, suggesting a smaller change in the MT structure compared to the WT structure. With further statistical analysis (Table S2) we confirm that the stabilizing effect of the mutations becomes most apparent under thermal stress. At 300 K, the mutant and wild-type behave very similarly. At 400 K, however, the mutations significantly stabilize the PI domain. For the other domains which contain mutated residues (REC1-II, RuvC-II, and OBD-II) we observed modest trends towards lower RMSD values or slower RMSD increases in the mutant compared to the wild-type. However, as detailed in Table S2, these differences were not statistically significant. This suggests that the primary stabilizing effect of the mutations is concentrated in the PI domain at elevated

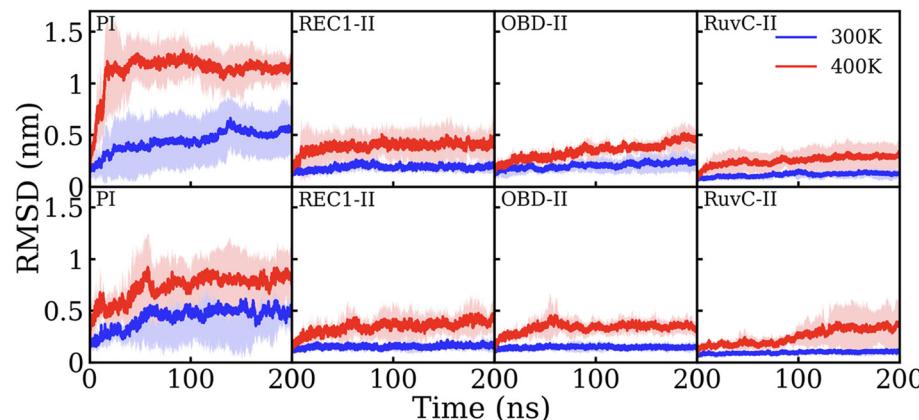


Fig. 4 RMSD of four domains which contain point mutations in the wildtype (top) and mutant (bottom) protein, as labeled. The solid lines indicate mean value averaged over four independent trials, while the shaded region indicates the standard deviation.



Table 1 Hydrogen bond occupancy of hydrogen bonds formed between mutated residues to other residues. Mutated residues are highlighted in bold. Hydrogen bond pairs with >10% occupancy are shown in the form of donor(atom)–acceptor(atom), with occupancy averaged over four 200 ns trajectories

Donor(atom)–acceptor(atom) (WT)	300 K	400 K	Donor(atom)–acceptor(atom) (MT)	300 K	400 K
R160(HN)–R156(O)	89.5	41.2	E160(HN)–R156(O)	83.5	49.6
R160(H11)–E157(OE2)	11.8	5.4	A164(HN)–E160(O)	70.1	10.6
R160(H11)–E157(OE1)	11.9	5.8			
M212(HN)–F208(O)	99.2	93.0	M212(HN)–W208(O)	99.5	97.5
H394(HE2)–F208(N)	11.1	2.2	W208(HE1)–E422(OE1)	94.2	48.0
N524(HN)–K530(O)	98.7	95.8	V524(HN)–K530(O)	98.9	93.3
K530(HN)–N524(O)	78.9	77.0	K530(HN)–V524(O)	94.8	60.8
K526(HN)–N524(OD1)	66.5	31.2			
K530(HZ1)–D868(OD1)	54.5	54.7	V868(HN)–V531(O)	84.8	83.1
D868(HN)–V531(O)	56.9	45.6			
K530(HZ1)–D868(OD2)	67.0	51.2			

temperatures, while the impacts on other domains are more subtle.

To further understand intra-protein structural changes, we investigated the formation of hydrogen bonds, to capture how mutation alters the original hydrogen bond network of the protein. Table 1 and Fig. S8 show the formation of hydrogen bonds between the mutated residues and other protein residues before and after mutation. As the mutations contribute both to the formation or disappearance of hydrogen bonds, a comparable number of total hydrogen bonds before and after mutation at 300 K is observed. We then investigate residues pairs that give high occupation values (*i.e.*, hydrogen bonds observed in more than 10% of simulation time). For R160, the hydrogen bonds are formed mainly with R156 (backbone oxygen) and E157 (side chain oxygen). Specifically, R160–R156 was found to be the dominant hydrogen bond with a high occupation (89%) on average across all four replicates. This is an example of a hydrogen bond between pairs that are less than four residues apart and is integral in maintaining the alpha-helical structure of the PI domain. After mutation R160E, A164 takes the place of E157 as an acceptor to form a hydrogen bond with E160 and a larger helix bundle is observed. For F208, residue pairs forming hydrogen bonds are M212–F208 and H394–F208, where F208 is the acceptor. M212–F208 has a relatively high occupation, 99% on average. After the mutation F208W, the hydrogen bond previously formed between H394 and F208 is no longer observed between H394 and W208. W208 becomes the donor to form a hydrogen bond with E422, which possesses a steadier interaction with an occupation of ~94%. A closer examination of the switch of hydrogen bond involving F208W suggests the change of interaction with H398, located in a loop, to E422, located in a bundle of beta sheets. The formation of W208–E422 thus strengthens the interaction between the helix bundle of REC1-II and the beta-sheet of OBD-II. Unlike R160E and F208W, N524V and D868V lead to a decrease in the number of hydrogen bonds formed to other residues due to the change of the long side chains to relatively short and hydrophobic methyl side chains. Post mutation of N524V, its hydrogen bond with K526 disappears, which was located on

the loop connecting two anti-parallel beta-sheets. D868 also forms hydrogen bonds with K530 and V531. Mutation D868V results in the loss of hydrogen bonds with K530 but hydrogen bonds between hydrophobic residues such as V531 is maintained.

To understand the interplay between polar interactions and hydrophobic effects, we complemented the analysis of hydrogen bond network by examining the SASA of mutation point to quantify how these hydrogen bond rearrangements affect the overall exposure of the mutation site to the solvent. The changes in SASA (Fig. S9) provide insight into how these rearrangements affect local packing and solvent exposure. For instance, the significant structural rearrangement described for the R160E mutation, which results in the formation of a larger helical bundle, corresponds to a dramatic decrease in SASA. This indicates that the new hydrogen bonding pattern causes the residue at position 160 to become significantly more buried, likely benefiting from improved hydrophobic packing within the protein core. Conversely, for mutations like F208W, where the hydrogen bond partner is switched rather than lost, the effect on solvent accessibility is minimal, reflecting a subtler structural adjustment rather than a large-scale refolding. Notably, for the N524V and D868V mutations, the loss of hydrogen bonds does not lead to increased solvent exposure; instead, the mutated residues remain buried, suggesting a compensatory mechanism where stability is maintained or enhanced through more favorable hydrophobic interactions within the core. Together, these analyses illustrate complex interplay where both the remodeling of hydrogen bond networks and the optimization of hydrophobic burial contribute to local protein packing and solvent interface.

We posit that this combination of effects presents a key mechanism for enhancing protein function at higher temperature. The general trend across mutations like F208W is the replacement of transient interactions with more persistent hydrogen bonds, coupled with the enhanced burial of key residues. We infer that this enhanced structural stability, driven by both polar and nonpolar forces, contributes to the protein's increased catalytic activity at

higher temperatures by maintaining a more functionally optimal conformation.

Protein essential dynamics

The mutations bring structural change to BrCas12b, potentially responsible for the improved stability and activity. To decouple the complex motion of the apo-BrCas12b protein and to understand the essential degree of freedom for both WT and MT, we performed PCA, as described in the Methods section. The dynamics of the protein along the first principal component is usually referred to as essential dynamics.³⁸

Upon projecting the simulation trajectory on the top two principal components (PC1 and PC2), we examine the projected motion of each principal component (Fig. 5, S10 and S11). For WT BrCas12b at 300 K (Fig. 5a), the essential dynamics presented along PC1 represent a large-scale “breathing” motion within and between the REC and nuclease NUC lobes of the protein. Specifically, this involves the movement of domains BH and REC2 (part of the REC lobe) toward domain REC1-II (Fig. S11a), alongside relative motions between domain PI and domain OBD-II with a smaller amplitude. This motion is analogous to the “open” and “closed” conformations described for other Cas enzymes, which are characterized to be functionally important for nucleic acid binding.^{23,24} Our simulations capture the intrinsic flexibility of the apo-protein, and the observed motions represent the dominant conformational changes accessible to the enzyme in its unbound state. Establishing a direct link between these specific apo-state dynamics and the mechanism of nucleic acid binding remains a hypothesis for future bound-state studies. The projected motion on PC2

(Fig. S10a) highlights the orthogonal motion of domain BH together with REC2 compared to PC1, presenting a twisting motion making the front part of domain RuvC-I act as a fulcrum. A significant amplitude motion is also observed for domain PI, with an alpha-helix bundle moving inward and outward around the hollow part of the apo-BrCas12b. With the four RFND mutations, a smaller ‘open’ and ‘close’ motion is observed for MT BrCas12b from the projection of PC1 (Fig. 5c). The previously observed significant amplitude movement between domain BH, REC2, and domain REC1-II was much smaller than WT BrCas12b (Fig. S11c). The initially small-scale motion from domain PI and domain OBD-II is enhanced in MT BrCas12b, with PI presenting a relative open conformation, where the point mutation was introduced. For PC2, MT presents a similar motion for domain BH and REC2, as seen in WT. The motion of PI, however, toward the outside of the protein is different due to the open conformation adopted.

We also performed PCA on high-temperature simulations of the WT and MT BrCas12b, to understand how the mutation alters the essential dynamics at high temperatures and whether the mutation changes the mode of motion that unfolds the protein. For both WT and MT, an inward motion is observed for most of the domains from PC1 (Fig. 5b and d), primarily domains REC2 and BH, moving together with domains OBD-I and II toward each other. This motion results in a lower exposed surface area, which also explains the downward trend of the SASA as the simulation proceeds. The large central cleft between the protein’s lobes also collapses during this motion. As for PC2 (Fig. S10b and S10d), the domain BH and domain REC2 complex move towards domain REC1-II, however, domain REC1-II moves

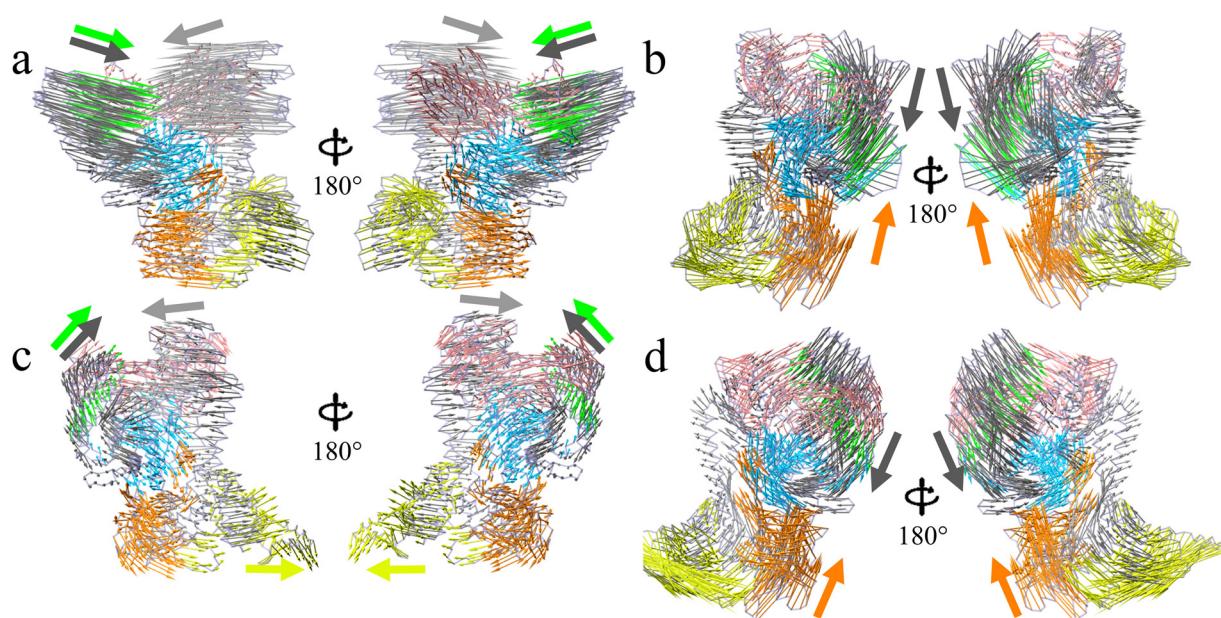


Fig. 5 Essential dynamics derived from the first principal component (PC1) of wild-type BrCas12b at (a) 300 K, (b) 400 K, and mutant BrCas12b at (c) 300 K, (d) 400 K shown using arrows of sizes equivalent to the amplitude of motions, with colors adapted from Fig. 1 to distinguish motions of different domains.



away from BH and REC2, since the extent of this motion is small, thus preserves the open and close motion observed at 300 K. The same motion was observed with the MT between domain REC2, BH, and domain OBD-I, II. Domain PI preserves the outward trend at 300 K and undergoes a large-scale motion toward the OBD-II than WT. The “open” motion is also observed for MT at 400 K, together with the inward motion of domain OBD, which differs from the WT protein. The domain PI then undergoes a significant outward expansion, which is not observed for in WT. Compared with the counterpart at 300 K, the motion from OBD domain is altered both from direction and amplitude, while REC2 also shown different motion.

By looking at the conformational space presented by the projection onto PC1 and PC2, we can clearly identify the open and closed motion along the PC1 (Fig. S12). To ensure this result was robust and representative, we projected our four shorter (200 ns) replica trajectories onto this 2D subspace. The resulting projections, also shown in Fig. S13, confirm that the replicas sample the same conformational landscape as the main 1 μ s trajectory. This validates our sampling and the relevance of the identified essential dynamics. Covariance

analysis showed that trace values of WT and MT are 483 nm² and 162 nm², respectively, at 300 K. The lower trace value of MT indicates a generally lower flexibility post mutation. At 400 K, the trace values of WT and MT are 376 nm² and 280 nm², respectively.

Cross correlation changes upon mutation

Next, we performed a cross-correlation analysis to probe the correlation of individual residues and corresponding protein domains. The results are shown in Fig. 6, with the positive value of $C(i, j)$ implying positively correlated movement whereby the two Ca atoms moved in the same direction, indicating a coordinated or coupled motion. In contrast, the negative value indicates anti-correlated movements whereby the atoms move in the opposite direction.³⁹ As shown in Fig. 6a and c, at 300 K, the introduction of point mutations changes the correlated motions nearby. For R160E and F208W, the mutation slightly changed the correlated motion from positive to negative, involving part of domain PI and REC 1-II. The same correlation change is observed between domain RuvC-II, where D868V is located, and domain UK-I,

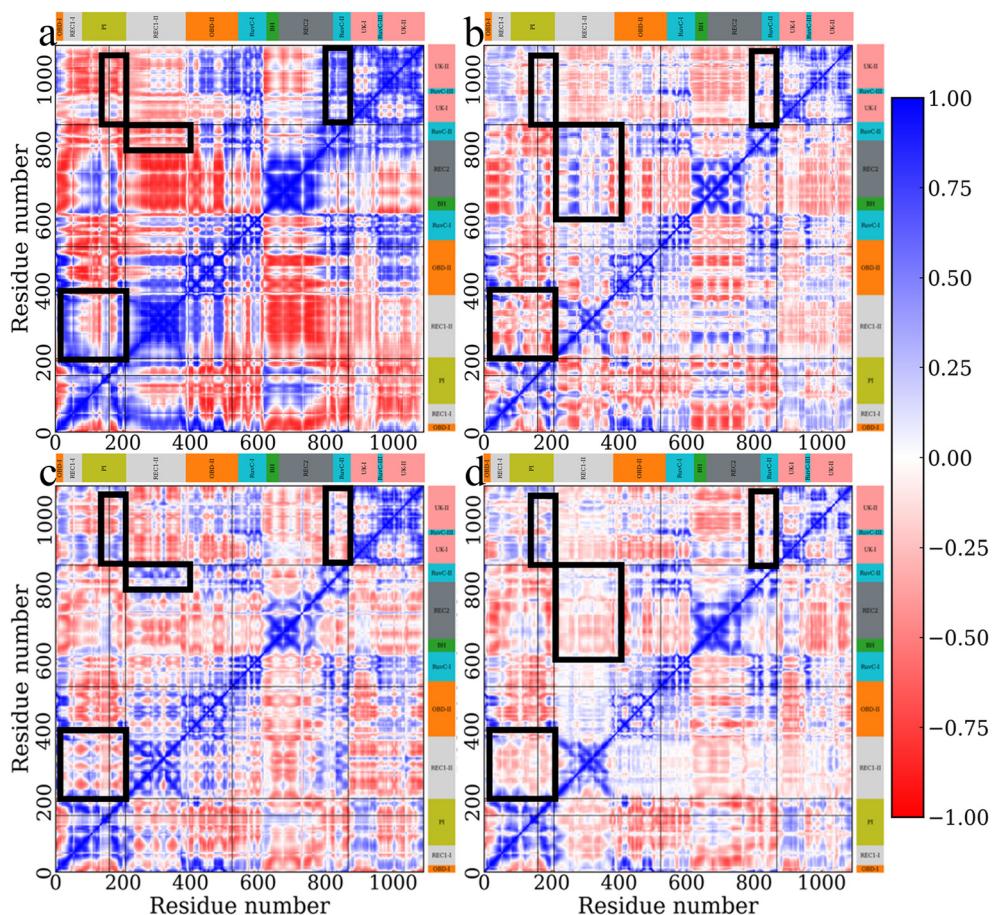


Fig. 6 Dynamic cross-correlation matrix (DCCM) calculated from 1 μ s simulation trajectory of wild-type BrCas12b at (a) 300 K, (b) 400 K, and mutant BrCas12b at (c) 300 K, (d) 400 K, with correlation between pairs from negative to positive colored red to blue. The protein sequence is shown along the x- and y-axis with the corresponding domain assignment highlighted with colors. The mutated residues are highlighted in the black boxes.



UK-II. The shift from negative to positive was then observed from the correlation between domain RuvC-II and domain REC1-II as well as domain PI and domain UK-I, UK-II. At 400 K, the initial high correlation was observed to decrease, especially for the off-diagonal component (Fig. 6b and d). The correlation between domain PI and domain UK shows a change from negative to positive, similar to that of WT 300 K. The same result was also observed for the correlation between REC1-II and RuvC-II, domain RuvC-II, and domain UK. The correlation change between domain REC1-II and PI was rather complicated, with the reverse of correlation mode happening differently. However, the correlation changes in MT at 400 K mainly decreased instead of completely changing. Thus, the introduced point mutations change the correlation mode at 300 K and reduces the amplitude of correlation change at 400 K.

Finally, per residue correlation was calculated to highlight the interdomain correlation. As shown in Fig. 7a, strong positive and negative correlations were observed across multiple domains for WT at 300 K. As mentioned earlier, large-scale 'open' and 'close' motions between domain REC1-II and domain BH, REC2, can also be identified here to show a strong negative correlation, indicating the opposite direction movement, together with the smaller amplitude motion between domain PI and domain OBD-II. Those motions become smaller after introducing the four mutations, as shown in Fig. 7c. At high temperatures, initially

strongly correlated motions between specific domains are observed to change (Fig. 7b and d), especially for domain REC1-I and REC1-II. The initially strong negative correlations are observed to decrease or even reverse to positive correlations. However, the correlation between REC1-I and REC1-II changed from positive to slightly negative, indicating the change from moving toward the same direction to the opposite direction. For MT, the high temperature does not change the pattern of cross-correlation much. However, it is worth noting that the dynamic cross correlation may only account for the linear correlation.

Conclusions

We conducted MD simulations on the wild type (WT) and mutated (MT) BrCas12b at low and high temperatures, proposing a mechanistic hypothesis for the unfolding behavior of both WT and MT as well as the change of essential dynamics upon mutation to understand the origins of thermostability in mutant BrCas12b. Our study began with a protein structure predicted by AlphaFold. While these models are highly accurate, they represent a single, static, low-energy conformation and do not capture the full thermodynamic ensemble of a protein in solution. Meanwhile, the structure predicted after the four-point mutation did not change significantly. To mitigate the potential bias of starting all simulations from a single,

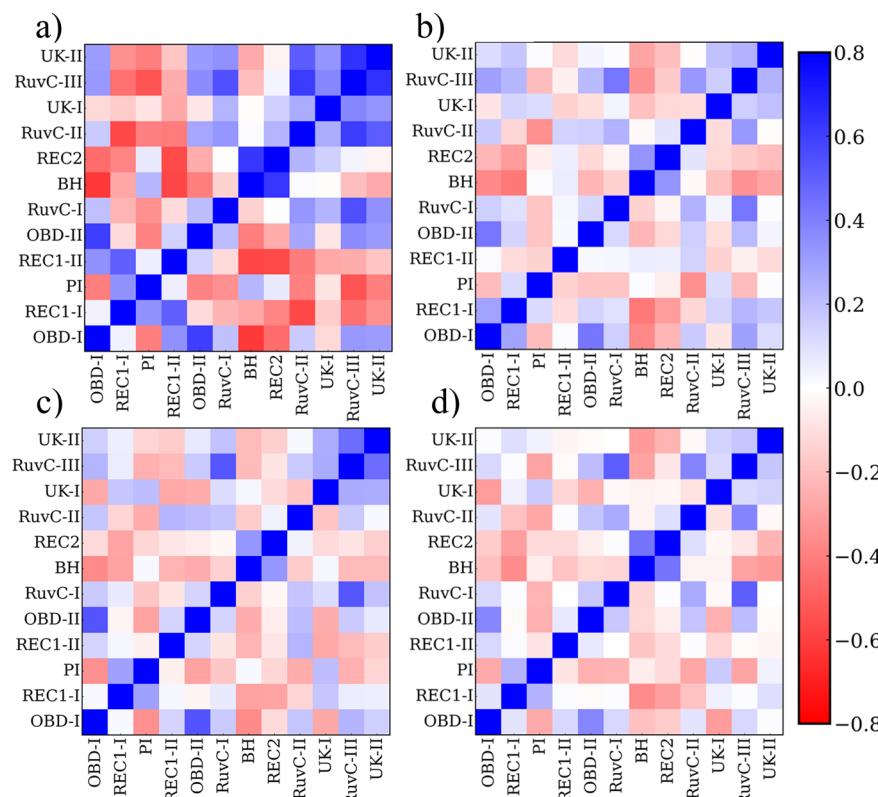


Fig. 7 Accumulated residue correlation between separate domains for wild-type BrCas12b at (a) 300 K, (b) 400 K, and mutated-type BrCas12b at (c) 300 K, (d) 400 K.



idealized structure, we first performed a 200 ns equilibration simulation at 300 K. The starting configurations for our four production replicates were then randomly selected from the stable latter half of this trajectory (100–200 ns). This protocol ensures that our production simulations began from a set of relaxed, structurally diverse conformations, thereby enhancing the sampling of the near-native state. Nonetheless, a limitation of this approach is that the entire simulation still originates from a single predicted fold. While our high-temperature simulations are designed to accelerate unfolding, the observed pathways are inherently linked to this initial structural basin. Therefore, the mechanistic insights presented should be interpreted as well-supported hypotheses that highlight accelerated destabilization trends, rather than a true depiction of the unfolding pathway at physiological temperatures.

Global property assessment on BrCas12b variants at ambient and elevated temperatures helped identify overall structural change, which indicates the unfolding of the protein at high temperatures. Further analysis on the decomposed domains based on sequence alignment of the protein highlighted the stabilization of the PI domain post mutation, while the other domains containing mutations show smaller structural change at the end of the simulation compared to the wild type. Hydrogen bond analysis shed light on the influence of mutations in generating new hydrogen bond networks, contributing to improved protein stability. The analysis of the essential dynamics from principal component analysis illustrates how wild type BrCas12b behaves at room temperature with the “open” and “close” motions observed in previous studies of other CRISPR-Cas proteins and how high temperature alters these dynamics to unfold the protein by first inducing the collapse of the initial structure followed by shrinking of the protein. Similar analysis of mutant BrCas12b suggests alterations in the protein dynamics after mutation, especially domain PI, as well as a change in the dominant motion during the unfolding at high temperatures. The cross-correlation between atom pairs and accumulated cross-correlation score between domains further corroborates the “open” and “close” motion and show how mutation decreases the correlation at both 300 K and 400 K. Overall, MD simulations of WT and MT BrCas12b at low and high temperature provide important details on how mutations improve the protein stability while altering the essential motions, which explains prior experimental observations. We note that all simulations were performed on the apo form of the protein, without RNA/DNA binding. We chose to simulate this state to align with the experimental study by Nguyen *et al.*,^{19,20} which was also done on the apo form of the protein. Our apo-state simulations are therefore the most direct and relevant method to propose a mechanistic hypothesis for those specific experimental findings. The “open–close” transitions and domain-level motions identified by our PCA represent the intrinsic conformational dynamics of the unbound enzyme. These pre-existing motions are what prepare the protein to accommodate its guide RNA and target DNA. Our results suggest that the mutations confer stability by modulating

these intrinsic dynamics, which in turn alter the conformational landscape of the unbound enzyme. Nevertheless, we acknowledge that the functional motions and stability are ultimately coupled to nucleic acid binding. Comparing the dynamics of the WT and mutant complexes will be essential to fully understand how these mutations impact the entire functional and thermodynamic cycle of the enzyme in a diagnostic context. Future experimental work would be invaluable for testing the specific predictions about secondary structure and domain stability generated by our simulations. This paves a clear path for future investigations. The most direct and compelling next step would be to perform a parallel set of simulations on the holo form of the protein. Comparing the dynamics of the WT and mutant proteins in their substrate-bound state would provide critical insights into how these mutations affect the functional catalytic cycle. Furthermore, to gain a more quantitative understanding of the protein’s stability, future work could employ enhanced sampling methods. Techniques such as metadynamics or replica exchange molecular dynamics (REMD) would allow for a more exhaustive exploration of the unfolding landscape and enable the calculation of the free energy differences ($\Delta\Delta G$) between the wild-type and mutant proteins. This would provide a definitive, quantitative measure of the thermodynamic impact of these stabilizing mutations.

Conflicts of interest

S. R. R. and P. K. J. are listed as inventors on the patent applications related to BrCas12b and their engineered variants. S. R. R. is a co-founder of CasNx, LLC. P. K. J. is a co-founder of CRISPR, LLC and CasNx, LLC.

Data availability

Data for this article, including GROMACS .mdp files used to execute the simulations, PDB files for the wild-type and mutated BrCas12b, and Jupyter notebook that contains analysis of cross-correlation matrix are available at GitHub at https://github.com/UFSRG/published-work/tree/main/2025_MSDE.

Supplementary information: domain assignment details. Additional protein structure analysis includes total hydrogen bond number, total SASA, predicted CD spectra, backbone RMSD for different protein domains, and number of hydrogen bonds formed before and after mutation. Additional protein dynamics analysis includes protein dynamics projection onto PC2, simulation trajectory projected onto PC1–PC2 phase space, and cumulative contribution of all principal components. See DOI: <https://doi.org/10.1039/d5me00140d>.

Acknowledgements

J. S. acknowledges startup funds provided by the Department of Chemical Engineering and the Herbert Wertheim College of Engineering at the University of Florida (<https://www.che.ufl.edu>).



ufl.edu). J. S. and Y. J. gratefully acknowledge funding provided by the Oak Ridge Associated Universities (ORAU) Ralph E. Powe Junior Faculty Enhancement Award. P. K. J. and S. R. R. acknowledge support through the National Institute of General Medical Sciences (R35GM147788). K. H. acknowledges support by the National Science Foundation under Grant No. 1852111. Y. J. and K. H. acknowledge University of Florida Research Computing for providing computational resources and support (<https://www.rc.ufl.edu/>).

References

- 1 M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier, *Science*, 2012, **337**, 816–821.
- 2 J. A. Doudna and E. Charpentier, *Science*, 2014, **346**, 1258096.
- 3 Y. Li, S. Li, J. Wang and G. Liu, *Trends Biotechnol.*, 2019, **37**, 730–743.
- 4 R. Aman, A. Mahas and M. Mahfouz, *ACS Synth. Biol.*, 2020, **9**, 1226–1233.
- 5 K. Pardee, A. A. Green, M. K. Takahashi, D. Braff, G. Lambert, J. W. Lee, T. Ferrante, D. Ma, N. Donghia, M. Fan, N. M. Daringer, I. Bosch, D. M. Dudley, D. H. O'Connor, L. Gehrke and J. J. Collins, *Cell*, 2016, **165**, 1255–1266.
- 6 Y. Zhang, L. Qian, W. Wei, Y. Wang, B. Wang, P. Lin, W. Liu, L. Xu, X. Li, D. Liu, S. Cheng, J. Li, Y. Ye, H. Li, X. Zhang, Y. Dong, X. Zhao, C. Liu, H. M. Zhang, Q. Ouyang and C. Lou, *ACS Synth. Biol.*, 2017, **6**, 211–216.
- 7 O. O. Abudayyeh, J. S. Gootenberg, S. Konermann, J. Joung, I. M. Slaymaker, D. B. T. Cox, S. Shmakov, K. S. Makarova, E. Semenova, L. Minakhin, K. Severinov, A. Regev, E. S. Lander, E. V. Koonin and F. Zhang, *Science*, 2016, **353**, aaf5573.
- 8 J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, P. Essletzbichler, A. J. Dy, J. Joung, V. Verdine, N. Donghia, N. M. Daringer, C. A. Freije, C. Myhrvold, R. P. Bhattacharyya, J. Livny, A. Regev, E. V. Koonin, D. T. Hung, P. C. Sabeti, J. J. Collins and F. Zhang, *Science*, 2017, **356**, 438–442.
- 9 J. S. Gootenberg, O. O. Abudayyeh, M. J. Kellner, J. Joung, J. J. Collins and F. Zhang, *Science*, 2018, **360**, 439–444.
- 10 C. Myhrvold, C. A. Freije, J. S. Gootenberg, O. O. Abudayyeh, H. C. Metsky, A. F. Durbin, M. J. Kellner, A. L. Tan, L. M. Paul, L. A. Parham, K. F. Garcia, K. G. Barnes, B. Chak, A. Mondini, M. L. Nogueira, S. Isern, S. F. Michael, I. Lorenzana, N. L. Yozwiak, B. L. MacInnis, I. Bosch, L. Gehrke, F. Zhang and P. C. Sabeti, *Science*, 2018, **360**, 444–448.
- 11 J. S. Chen, E. Ma, L. B. Harrington, M. Da Costa, X. Tian, J. M. Palefsky and J. A. Doudna, *Science*, 2018, **360**, 436–439.
- 12 S.-Y. Li, Q.-X. Cheng, J.-K. Liu, X.-Q. Nie, G.-P. Zhao and J. Wang, *Cell Res.*, 2018, **28**, 491–493.
- 13 S.-Y. Li, Q.-X. Cheng, J.-M. Wang, X.-Y. Li, Z.-L. Zhang, S. Gao, R.-B. Cao, G.-P. Zhao and J. Wang, *Cell Discovery*, 2018, **4**, 20.
- 14 J. P. Broughton, X. Deng, G. Yu, C. L. Fasching, V. Servellita, J. Singh, X. Miao, J. A. Streithorst, A. Granados, A. Sotomayor-Gonzalez, K. Zorn, A. Gopez, E. Hsu, W. Gu, S. Miller, C.-Y. Pan, H. Guevara, D. A. Wadford, J. S. Chen and C. Y. Chiu, *Nat. Biotechnol.*, 2020, **38**, 870–874.
- 15 L. T. Nguyen, B. M. Smith and P. K. Jain, *Nat. Commun.*, 2020, **11**, 4906.
- 16 L. Liu, P. Chen, M. Wang, X. Li, J. Wang, M. Yin and Y. Wang, *Mol. Cell*, 2017, **65**, 310–322.
- 17 L. Li, S. Li, N. Wu, J. Wu, G. Wang, G. Zhao and J. Wang, *ACS Synth. Biol.*, 2019, **8**, 2228–2237.
- 18 J. Joung, A. Ladha, M. Saito, N.-G. Kim, A. E. Woolley, M. Segel, R. P. J. Barretto, A. Ranu, R. K. Macrae, G. Faure, E. I. Ioannidi, R. N. Krajeski, R. Bruneau, M.-L. W. Huang, X. G. Yu, J. Z. Li, B. D. Walker, D. T. Hung, A. L. Greninger, K. R. Jerome, J. S. Gootenberg, O. O. Abudayyeh and F. Zhang, *N. Engl. J. Med.*, 2020, **383**, 1492–1494.
- 19 L. T. Nguyen, N. C. Macaluso, B. L. M. Pizzano, M. N. Cash, J. Spacek, J. Karasek, M. R. Miller, J. A. Lednický, R. R. Dinglasan, M. Salemi and P. K. Jain, *EBioMedicine*, 2022, **77**, 103926.
- 20 L. T. Nguyen, S. R. Rananaware, L. G. Yang, N. C. Macaluso, J. E. Ocana-Ortiz, K. S. Meister, B. L. M. Pizzano, L. S. W. Sandoval, R. C. Hautamaki, Z. R. Fang, S. M. Joseph, G. M. Shoemaker, D. R. Carman, L. Chang, N. R. Rakestraw, J. F. Zachary, S. Guerra, A. Perez and P. K. Jain, *Cell Rep. Med.*, 2023, **4**, 101037.
- 21 N. Nagasundaram, H. Zhu, J. Liu, V. Karthick, C. G. P. Doss, C. Chakraborty and L. Chen, *PLoS One*, 2015, **10**, 1–21.
- 22 Y. Jia, A. Fernandez and J. Sampath, *J. Phys. Chem. B*, 2023, **127**, 6856–6866.
- 23 G. Palermo, Y. Miao, R. C. Walker, M. Jinek and J. A. McCammon, *ACS Cent. Sci.*, 2016, **2**, 756–763.
- 24 A. Saha, P. R. Arantes, R. V. Hsu, Y. B. Narkhede, M. Jinek and G. Palermo, *J. Chem. Inf. Model.*, 2020, **60**, 6427–6437.
- 25 S. Bhattacharya and P. Satpati, *ACS Omega*, 2022, **8**, 1817–1837.
- 26 F. Teng, T. Cui, G. Feng, L. Guo, K. Xu, Q. Gao, T. Li, J. Li, Q. Zhou and W. Li, *Cell Discovery*, 2018, **4**, 63.
- 27 I. Fonfara, H. Richter, M. Bratović, A. Le Rhun and E. Charpentier, *Nature*, 2016, **532**, 517–521.
- 28 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583–589.
- 29 D. Wu, X. Guan, Y. Zhu, K. Ren and Z. Huang, *Cell Res.*, 2017, **27**, 705–708.
- 30 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1–2**, 19–25.
- 31 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller and A. D. MacKerell, *Nat. Methods*, 2016, **14**, 71–73.



32 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

33 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

34 P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess and E. Lindahl, *J. Chem. Theory Comput.*, 2010, **6**, 459–466.

35 D. J. Evans and B. L. Holian, *J. Chem. Phys.*, 1985, **83**, 4069–4074.

36 Y. Jia, C. Cocker and J. Sampath, *Biomacromolecules*, 2025, **26**, 2095–2105.

37 E. D. Drew and R. W. Janes, *Nucleic Acids Res.*, 2020, **48**, W17–W24.

38 A. Amadei, A. B. M. Linssen and H. J. C. Berendsen, *Proteins: Struct., Funct., Bioinf.*, 1993, **17**, 412–425.

39 H. Yu and P. A. Dalby, in *Methods in Enzymology*, Elsevier, 2020, vol. 643, pp. 15–49.

40 C. G. Ricci, R. L. Silveira, I. Rivalta, V. S. Batista and M. S. Skaf, *Sci. Rep.*, 2016, **6**, 19940.

41 Y. Mazola, O. Guirola, S. Palomares, G. Chinea, C. Menéndez, L. Hernández and A. Musacchio, *J. Mol. Model.*, 2015, **21**, 228.

42 S. Kumar and P. A. Deshpande, *PLoS One*, 2021, **16**, e0249866.

43 Z. Niu, K. Hasegawa, Y. Deng, Z. Zhang, M. Rafailovich, M. Simon and P. Zhang, *Front. Mol. Biosci.*, 2022, **9**, 953064.

