scientific reports

OPEN

Impact of frequent ARID1A mutations on protein stability provides insights into cancer pathogenesis

Rajen K. Goutam[1], Gangtong Huang[1], Exequiel Medina[2], Feng Ding[1], William J. Edenfield[3] & Hugo Sanabria[1]✉

The ARID1A gene, frequently mutated in cancer, encodes the AT-rich interactive domain-containing protein 1 A, a key component of the chromatin remodeling SWI/SNF complex. The ARID1A protein features a conserved DNA-binding domain (ARID domain) of approximately 100 residues crucial for its function. Despite the frequency of mutations, the impact on ARID1A's stability and contribution to cancer progression remains unclear. We analyzed five frequent missense mutations R1020S, M1022K, K1047Q, G1063V, and A1089T identified in The Cancer Genome Atlas (TCGA) to assess their effects on the stability of the ARID domain using a hybrid experimental and computational approach. By combining computational stability from web server tools, the structural dynamics from replica exchange discrete molecular simulation (rexDMD), and thermal and chemical denaturation experiments, we found that the R1020S mutation severely decreases structural stability, making it the most impactful, while M1022K has minimal effect, and others lie in between. These findings enhance our understanding of the structural-functional relationship of ARID1A missense mutations at the molecular levels and their role in cancer pathogenesis. This research paves the way for identifying and categorizing which ARID1A mutations are most pathogenic, potentially guiding the development of targeted therapies tailored to specific mutation profiles in cancer treatment.

Protein structural integrity and stability are essential for their proper function[1,2]. This is particularly important for proteins involved in chromatin dynamics and maintenance, as they play a critical role in regulating gene expression[3]. Within this context, the human ARID (AT-Rich Interaction Domain) family represents a physiologically significant group of proteins that participate in processes such as tissue-specific gene expression and cell growth regulation[4]. Among these, ARID1A stands out as a ubiquitous and vital component of the SWI/SNF chromatin remodeler complex[5–7], which modulates chromatin accessibility through ATP hydrolysis. Known by different nomenclatures, including B120, BAF250, BM029, and CSS2, ARID1A is crucial for organizing genomic architecture via its multidomain structure[8], influencing gene expression, DNA replication, and repair[9].

ARID family members share a highly conserved DNA-binding domain, known as the ARID domain, which is approximately 100 amino acids in length[10] (Fig. 1A). Structurally, this domain features a helix-turn-helix motif composed of six helices[11] (Fig. 1B). The DNA-binding interface is formed by three loops near the N-terminal, along with helices H5 and H6 and their connecting loops. However, mutations within the ARID1A DNA-binding domain, common in various cancers and malignancies, pose significant challenges, including therapeutic resistance and genomic instability[12,13]. Notably, these mutations are present in 6% of all cancers and approximately 46% of ovarian cancer cases[14,15]. Within the ARID domain, 37% of these mutations are missense or frameshift, as documented in The Cancer Genome Atlas (TCGA)[16]. While stop-gain and frameshift mutations clearly result in loss of function[17], the pathogenic impact of missense mutations remains difficult to interpret[8]. The most frequently observed ARID missense mutations in humans are R1020S, M1022K, K1047Q, G1063V, and A1089T (Fig. 1C), highlighting their potential pathogenicity. Structurally, R1020S and M1022K are located on helix H2, A1089T is found on helix H6, while the others are positioned on the loops flanking helix H3 (Fig. 1B). Disruption of these regions' structure and dynamics can impair function, potentially compromising tumor suppressor activity[18].

[1]Department of Physics and Astronomy, Clemson University, Clemson, SC, USA. [2]Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile. [3]Institute for Translational Oncology Research, Prisma Health, Greenville, SC, USA. ✉email: hsanabr@clemson.edu
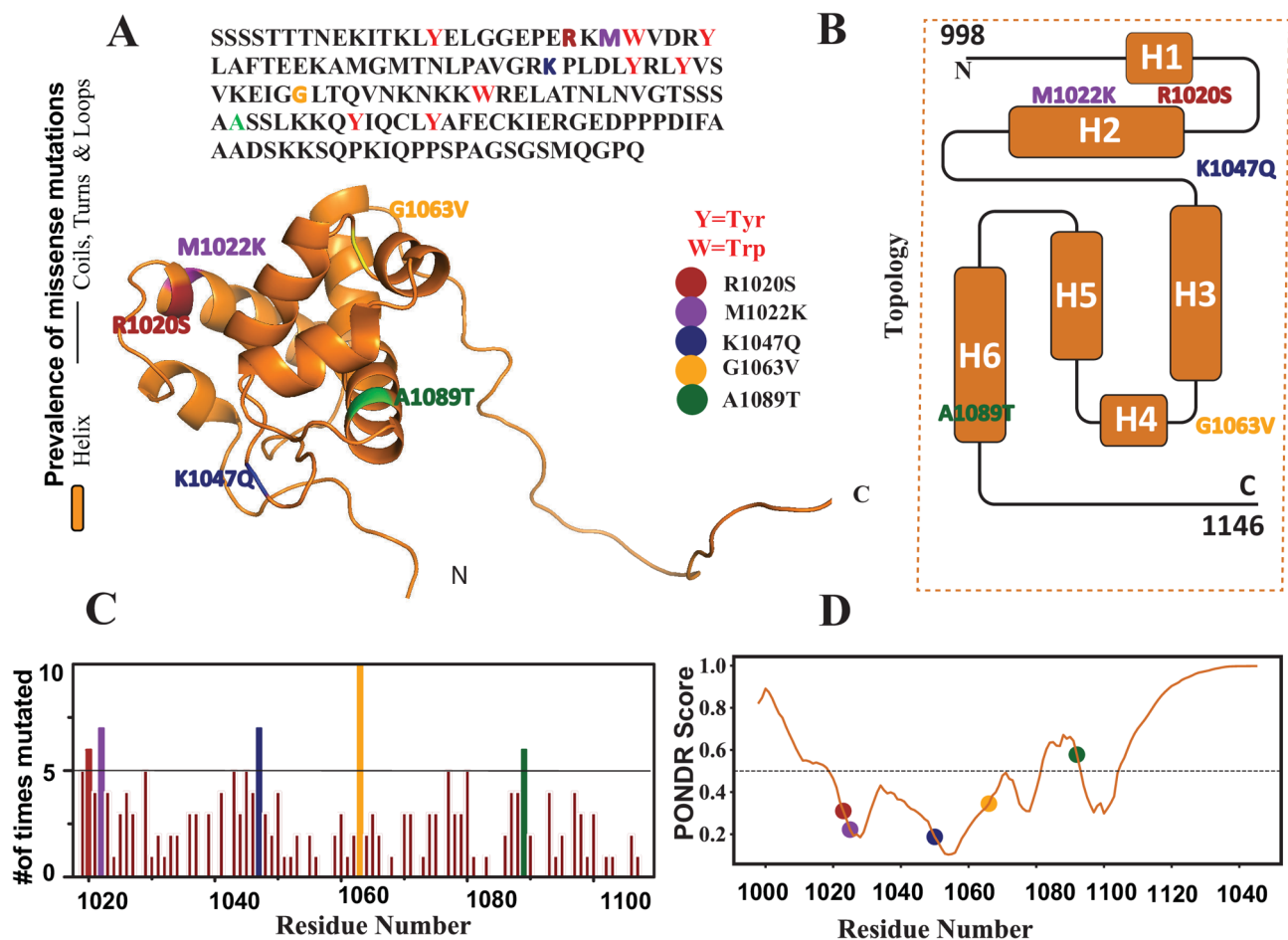
Fig. 1. Structural and Functional Motif of the ARID Domain in ARID1A. (A) Sequence and predicted three-dimensional ARID1A DNA binding domain, showing the positions of mutations and intrinsic residues as shown in color. (B) Secondary structure topology of the domain, showing the position of each mutant with the same color scheme as (A). (C) Frequency of every missense mutation in the ARID domain with respect to the position, using TCGA data and setting a threshold of at least five occurrences. (D) Disorder prediction using the PONDR score for the ARID1A's DNA binding domain, showing in color the position of each mutant.

Using the Predictor of Natural Disordered Regions (PONDR) web server[19], we identified regions within the ARID1A protein sequence that lack a fixed tertiary structure, known as disordered regions (DR) (Fig. 1D). Mutations within disordered regions could lead to dysfunction by disrupting chromatin dynamics[20]. And mutations within the structured regions could impact stability[21]. Therefore, investigating the structural consequences of these mutations is essential for understanding their physiological impact on ARID1A function.

In this study, we explored the impact of specific ARID1A mutations (Fig. 1) on protein structural stability using a hybrid approach that combined computational and experimental methods. Sequence and structure-based machine learning predictions, along with replica exchange molecular dynamics simulations (rexDMD)[22], identified R1020S as the most destabilizing mutation. Experimentally, thermal and chemical denaturation assays using intrinsic fluorescence showed that R1020S significantly reduced the melting temperature (T_m) and denaturation midpoint ($D_{1/2}$), corroborating our computational findings. Overall, our approach provides valuable insights into how specific amino acid changes affect protein stability, suggesting these variations may be pathogenic for ARID1A.

Materials and methods
Pathogenic predictor tools
We employed several webserver tools to analyze the predicted impact of mutations on ARID1A. We used SAAFEC-SEQ[23], which applies the Pseudo-Position Specific Scoring Matrix (PsePSSM) algorithm to predict changes in protein thermodynamic stability resulting from single mutations by encoding physicochemical properties, sequence features, and evolutionary information to compute changes in free energy. INPS[24] was utilized to predict the effects of mutations on protein stability directly from the protein sequence. I-Mutant2.0Field[25] also predicted stability changes using a Support Vector Machine algorithm trained on data from the ProTherm database[26]. We also employed mCSM[27] to evaluate mutation impacts by capturing distance patterns between atoms, representing protein residue environments, and training predictive models, providing

insights into protein stability and disease progression. Similarly, DDMut[28], a fast and accurate Siamese network, was used to predict changes in Gibbs Free Energy (ΔG) upon single and multiple point mutations by integrating graph-based representations with deep learning models, offering high predictive performance and scalability for understanding mutation effects and guiding rational protein engineering.

rexDMD simulation

In our in-silico investigation, all simulations were performed using all-atom discrete molecular dynamics (DMD) simulations with an implicit solvent model[22,29] DMD is a unique molecular dynamics algorithm with significantly enhanced sampling efficiency. The interatomic interactions were modeled with discretized step functions in the Medusa force field, which has been widely discussed in previous studies[30–35]. Anderson's thermostat was used to maintain the temperature of the simulations. The structure of wild-type ARID1A was predicted using AlphaFold 2[36]. The mutant structures were further prepared by introducing mutations at corresponding residues with the Mutagenesis tool in PyMOL[37]. We relaxed the predicted structures by performing 400 ps of DMD simulation under 300 K with a heat exchange factor of 10 to eliminate the unphysical defects. The resultant structures after relaxation were used as the starting structures for subsequent simulations.

In rexDMD simulations, multiple trajectories (replicas) run parallel under different temperatures. Temperature between replicas was exchanged every 50 ps according to the Metropolis criteria. Therefore, the simulations trapped at local energy minima could overcome the energy barrier by running at higher temperatures, thus increasing conformation sampling efficiency. In rexDMD simulations of all ARID1A mutants in this study, with a heat exchange factor of 0.1. A total of 20 trajectories with temperatures evenly distributed in the range 275–400 K lasted for ~0.65 μs, resulting in a cumulative total simulation time of ~13 μs. The minimum and maximum exchange rates between each two replicas in all simulations were 0.389 and 0.802, respectively.

The thermodynamic properties of the ARID1A variants were calculated using the weighted histogram analysis method (WHAM)[38] by self-consistently combining simulations under different temperatures. The potential of mean force (PMF) or effective free energy was calculated according to $PMF = -k_B T ln P (R_{PMF})$ where k_B is the Boltzmann constant, T is the simulation temperature, and R_{PMF} is the multiple dimensional parameter such as radius of gyration, R_g, and $P(R_{PMF})$ is the probability density obtained through WHAM.

Protein expression and purification

Codon-optimized DNA sequence encoding the ARID domain of human ARID1A, and its mutants were cloned into a pET-28a (+) vector containing a His- tag, with a thrombin cleavage site(Figure S1). Mutants R1020S, M1022K, K1047Q, G1063V and A1089T mutants were generated by PCR mutagenesis using Q5 Site-Directed Mutagenesis Kit (New England Biolabs, Inc) and verified by Eton Biosciences Inc. Protein overexpression was induced in *E. coli* C41(DE3) with 0.5 mM of Isopropyl-β-D-thiogalactopyranoside (IPTG) at an optical density measured at 600 nm between 0.6 and 0.8. Induced cells were harvested by centrifugation and lysed under sonication and treated with DNase I from Sigma- Aldrich in 10mM PBS buffer pH 7.5 supplemented with 1mM CaCl₂. Proteins were purified using HisPur™ Ni-NTA Spin Purification Kit (Thermo Scientific™) and eluted using the PBS buffer containing 250mM imidazole. The removal of the His tag was performed by incubating the respective protein without GuHCl with 5U of Thrombin (Cytiva) per mg of protein for two hours. This protein was further concentrated up to approximately 1–3 mg/ml and stored in -20 C after verifying their size and purity through Gel Electrophoresis.

Fluorescence measurements

The intrinsic fluorescence measurements were done in a Fluorolog by Horiba system. For all cases, 4 μM of protein concentration was excited at 275 nm and fluorescence emission was recorded between wavelengths of 290 nm to 360 nm (Excitation bandwidth 5 nm, Emission bandwidth 2.5 nm, Light source Xe lamp, response 0.5 s).

For thermal unfolding experiments, samples were heated and constantly stirred from 293 to 363 K using a water bath at a rate of 1 °C/min, where the emission spectra are recorded at intervals of 5 °C. To ensure the proteins were at equilibrium during the measurements, we closely monitored and measured the kinetics between each interval of recordings of the spectra. Similarly, for the refolding experiment, each variant was cooled from 363 K to 293 K at the same rate for heating with the same condition applied for heating. The buffer's spectra were recorded and subtracted as background for all measurements. Also, a magnetic stir bar is always on to ensure the samples do not aggregate. We utilized GFP (Tm: 341 K, typical range 338–343 K, Figure S2) as control proteins to validate our experimental setup[39]. For thermal denaturation, kinetic equilibration was ensured at each heating interval, and for chemical denaturation, a 4-hour equilibration period was implemented at each GuHCl concentration. Based on similar proteins, the expected Tm for our protein ranges from 310 to 340 K. All experiments were performed in triplicate to minimize variability and ensure the reliability of our findings.

To analyze the thermal unfolding, we estimated the apparent Tm values using Eq. 1

$$fluorescence\,(T) = \frac{Y_n + Y_d\,exp\,\left(\frac{\Delta H_{app}}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right)\right)}{1 + exp\,\left(\frac{\Delta H_{app}}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right)\right)} \tag{1}$$

where ΔH_{app} is the apparent enthalpy change upon denaturation, T_m is the apparent melting point, and Y_n and Y_d are the pre and post transition fluorescence signal and T is the temperature of measurements.

We employed guanidine chloride (GuHCl) as a denaturant for chemical unfolding. For unfolding reactions, proteins were maintained and incubated with different GuHCl concentrations between 0 and 4 M at room

temperature for at least 4 h to reach equilibrium. For refolding, 40 µM protein samples were incubated with 4 M of GuHCl for at least 4 h at room temperature. Next, the protein was diluted 10 times, whereas the GuHCl concentration was changed between 0 and 4 M, incubating for at least another 4 h before taking the measurements. We measured each sample and subtracted the background using the buffer's spectra. Also, a magnetic stir bar is always on to ensure the samples do not aggregate.

We analyzed the chemical unfolding data using Eq. 2.

$$fluorescence\left([GuHCl]\right) = \frac{Y_n + Y_d exp\left(-\frac{\Delta G_{app} - m[GuHCl]}{R.T}\right)}{1 + exp\left(-\frac{\Delta G_{app} - m[GuHCl]}{R.T}\right)}, \tag{2}$$

where ΔG_{app} is the apparent unfolding free energy change, and m is the slope of denaturation transition.

Overall impact

We calculated the overall impact by first converting all effects into the same unit, kcal/mol, using Eqs. (3), (4) and (5). The web server tools provided the change in folding free energy, which we denote as ($\Delta \Delta G_{PT}$) (predictor tool). To add the effect of ΔT_m values from rexDMD and thermal denaturation, we used the following equations to convert it to kcal/mol:

$$\Delta \Delta G_{rexDMD} \approx \Delta H_m \left(\frac{\Delta T_m}{T_m}\right) \tag{3}$$

and for thermal denaturation (TD), $\Delta \Delta G_{TD} \approx \Delta H_m \left(\frac{\Delta T_m}{T_m}\right)$ (4)

where ΔH_m is the change in enthalpy, T_m is the melting point.

For chemical denaturation (CD), we evaluated the change in melting concentration ($\Delta D_{1/2}$) and used Eq. 5

$$\Delta \Delta G_{CD} \approx m \cdot \Delta D_{1/2} \tag{5}$$

where m is the slope of denaturation curve and $D_{1/2}$ is the denaturant concentration at which half of native signal is lost. Finally, we summed all the ΔΔG values and normalized the result by dividing by the reference value (the largest individual change).

$$\text{Overall Impact} = \frac{|\Delta \Delta G_{PT}| + |\Delta \Delta G_{rexDMD}| + |\Delta \Delta G_{TD}| + |\Delta \Delta G_{CD}|}{\text{scaling factor}} \tag{6}$$

Circular dichroism experiments

Measurements were done in a J-1500 spectropolarimeter instrument using 10 µM of protein concentration, recording the spectra from 200 to 260 nm at different 2,2,2-trifluoroethanol (TFE) concentrations (0–50% v/v), incubating at room temperature. The scanning rate was 50 nm/min.

Results

R1020S is the most destabilizing mutation by pathogenicity prediction tools

We evaluated the effects of five missense mutations (R1020S, M1022K, K1047Q, G1063V, and A1089T) on ARID1A's stability by employing sequence-based predictors (SAAFEC-SEQ, INPS, and I-Mutant2.0) and structure-based predictors (mCSM and DDMut) (see Materials and Methods). We compared the predicted stability and the change in folding free energy ($\Delta \Delta G$) of these mutations against the wild type (ΔΔG=ΔG_{WT}-ΔG_{MUTANT}), interpreting negative values as destabilizing and positive values as stabilizing. Most mutations predominantly resulted in destabilization, although a few exhibited minor stabilizing effects (Fig. 2 and Figure S3). Remarkably, the R1020S mutation had the most significant destabilizing impact across both sequence-based and structure-based predictors. However, the heterogeneity in predicted effects emphasizes the complexity of the mutational landscape and the challenges of relying solely on web server tools for accurate predictions. The accuracy of DDMut (Pearson 0.70–0.74, RMSE 1.37–1.67 kcal/mol) and mCSM (0.67–0.75) depends on their training models, with structure-based methods outperforming sequence-based tools achieves moderate accuracy (70–80%) but simplifies complex proteins. These predictors are biased toward simpler, single-domain proteins, making experimental validation necessary.

Discrete molecular dynamics simulations showed the R1020S as the most destabilizing mutant

To illustrate the impact of mutations on protein stability and dynamics at the molecular level, we conducted replica-exchange discrete molecular dynamics (rexDMD) simulations of both wild-type (WT) and mutant proteins. We ran 20 replicas at consecutive temperatures (300–380 K) over a cumulative time. To assess the changes in structural dynamics and impact introduced by the mutations induced by heating, we calculated the specific heat (C_v) and the average radius of gyration (R_g) within the simulated temperature range for all proteins (WT and MUTANT). The C_v curves revealed that, except for the R1020S mutant, all amino acid substitutions were stabilizing compared to the WT protein (Fig. 3A, Figure S4A). Next, we determined the temperature dependence of R_g across the specified temperature range (see Materials and Methods). The unfolding process exhibited cooperativity in all cases, indicated by the sigmoidal curves (Figure S4B). We
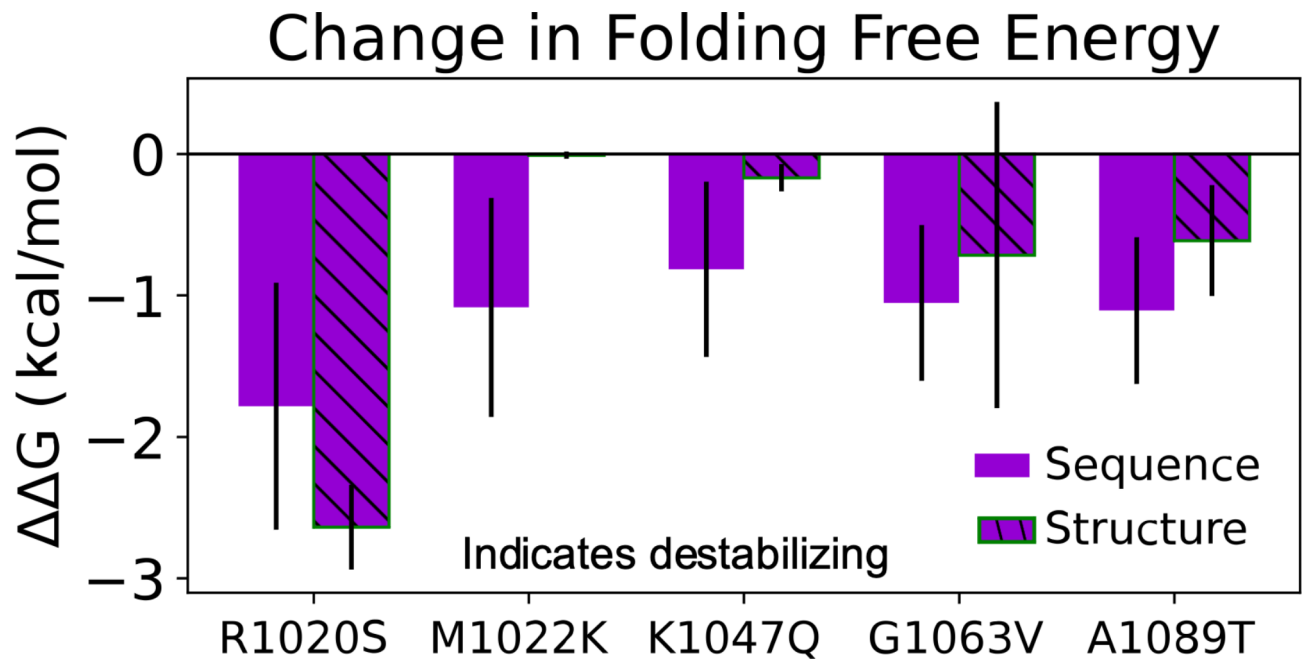
Fig. 2. Stability changes in ARID1A predicted by both sequence-based and structure-based models with respect to the selected mutations. The position and type of mutation are shown by the x-axis along with the unhatched from sequence-based and hatched from structure-based predictors. The evaluation of stability ($\Delta\Delta G_{\text{folding}}$) was made by calculating ΔG_{WT}-ΔG_{MUTANT}. Most of the mutations of interest show a negative $\Delta\Delta G_{\text{folding}}$ value, suggesting that the mutations are destabilizing.

estimated the melting temperature as the temperature corresponding to the peak of C_v and calculated the difference in melting temperatures ($\Delta T_m = T_{m,WT} - T_{m,MUTANT}$) for all proteins. Only the R1020S mutation showed a destabilizing effect (Fig. 3B). Given the consistent destabilizing effect of the R1020S mutant, we aimed to use a high-resolution approach to further understand its impact on stability.

We further examined the free energy landscape by analyzing the potential mean force in relation to α-helix content and radius of gyration (R_g) for the wild type (WT) and the R1020S mutant (Fig. 3, C-H, Figure S5). At 300 K, the WT protein exhibited an energy minimum with an α-helix content of approximately 0.32 and an R_g of 17 Å. In contrast, the R1020S mutant showed a much broader distribution in α-helix content, indicating greater structural heterogeneity. As the temperature increased, both variants demonstrated a decrease in α-helix content and an increase in R_g, reflecting increased structural flexibility due to partial unfolding. However, thermal denaturation occurred differently among the mutants at 320 K. Notably, the R1020S mutant exhibited a wider range of R_g, consistent with its reduced thermal stability. At 380 K, we observed a complete loss of α-helices and a significant increase in R_g, indicating the full unfolding of the WT and mutant ARID1A. We calculated the intra-chain contact maps of the WT and R1020S Arid1A (Figure S7) and calculated the difference in residue contact probabilities (Fig. 3I). The R1020S mutation induced the changes in α-helix and β-sheet structures across the entire Arid1A sequence. Compared with WT, R1020S Arid1A exhibited the most significant loss of α-helical contact patterns in residues 1120 to 1140 (red boxes in Fig. 3I), close to the C-terminal region, suggesting allosteric disruption of α-helix induced by R1020S. The change from positively charged R to uncharged S also induced changes in nearby anti-parallel β-sheet structures (residues 1000 to 1030), which is likely due to the rearrangements of the hydrophobic core in that region. Overall, our simulations confirmed that the R1020S mutation was the most destabilizing, significantly increasing structural heterogeneity and reducing the thermal stability of ARID1A.

Thermal stability showed irreversible process with R1020S impactful mutation

Bioinformatic approaches and molecular dynamics simulations suggest that the R1020S destabilizes the protein the most compared to the other mutations, likely due to the increased structural heterogeneity. To validate these predictions, we experimentally assessed the protein's stability through thermal denaturation using intrinsic fluorescence measurement. We monitored changes in fluorescence across a temperature range of 298–360 K.

We began with unfolding experiments, which revealed a red shift in the emission wavelength at maximum intensity and a decrease in fluorescence intensity in all cases (Fig. 4A and B, Figure S8). Next, we assessed the reversibility of the process by conducting refolding experiments over the same temperature range, starting with protein previously incubated at 360–298 K. During subsequent refolding, the variants exhibited a minimal blue shift in wavelength at maximum intensity, indicating irreversibility, and only partial recovery of fluorescence intensity (Fig. 4C and D, Figure S9). Notably, the R1020S mutant showed less fluorescence recovery compared to WT, and the native maximum wavelength was not restored in any case.
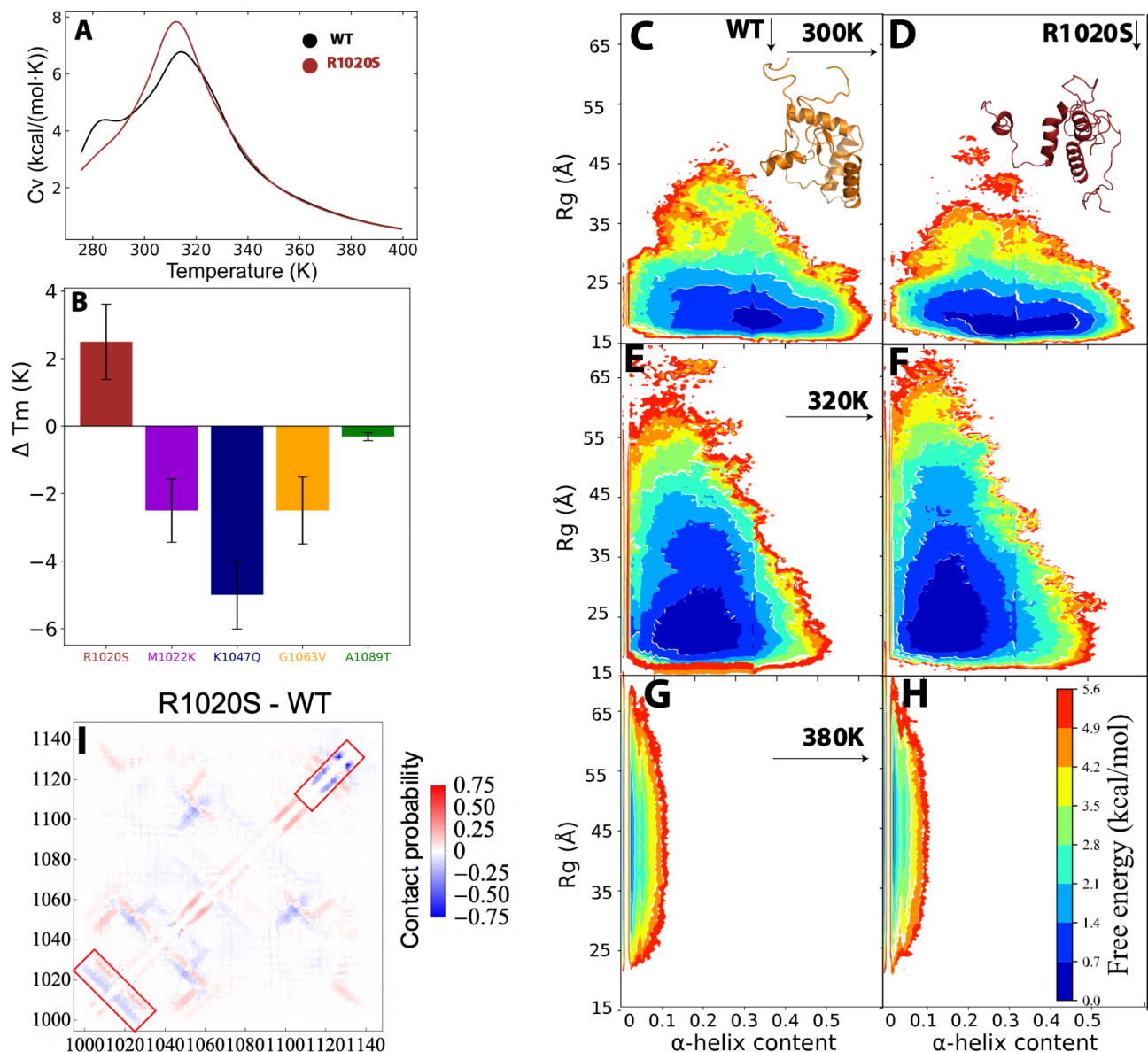
Fig. 3. rexDMD simulations for wild-type and mutants. (A) variation of C_v and temperature for the WT and R1020S, (B) Change in melting temperature (ΔT_m) for all mutants obtained from C_v calculation, C-H) 2D-PMF plot of WT (C, E and G) and R1020S (D, F and H) at different temperatures (top panel 300 K, middle panel 320 K and bottom panel 380 K) of R_g versus the α-helical contents. The structures inset of C and D are the native representation (basin at ~ 0.32 α-helical content and ~ 17 Å of R_g). I) Change in pair-wise residue contact probability between R1020S and WT, the rearrangements of β-sheets near R1020S mutation site and the most significant reduction in α-helix content are highlighted with red boxes.

Given the irreversible nature of thermal unfolding[40] (Fig. 4E and F, Figure S9A and B), we analyzed the fluorescence data by fitting it to a two-state unfolding model to determine the apparent melting temperature ($T_{m,app}$) (Materials and Methods). We then compared these values to the WT ($\Delta T_m = T_{m,WT} - T_{m,MUTANT}$) to assess stability differences (Fig. 4G, Figure S9C). The extracted $T_{m,app}$ varied among the mutants (Figure S9D), with WT displaying the highest thermal stability at $T_{m,app}$ of 333±1 K, indicating that all mutations are destabilizing. The R1020S, K1047Q, and G1063V mutants exhibited the lowest stabilities, with average ΔT_m values decreasing by 10 K. Overall, these thermal unfolding results align with bioinformatics predictions and rexDMD findings, confirming the significant destabilizing impact of the R1020S mutation on the protein's stability.

Chemical stability showed reversibility and R1020S the most impactful mutation

Our thermal denaturation experiments showed that the protein underwent irreversible unfolding, yet we could obtain apparent melting values that supported the bioinformatic predictions. To determine whether the
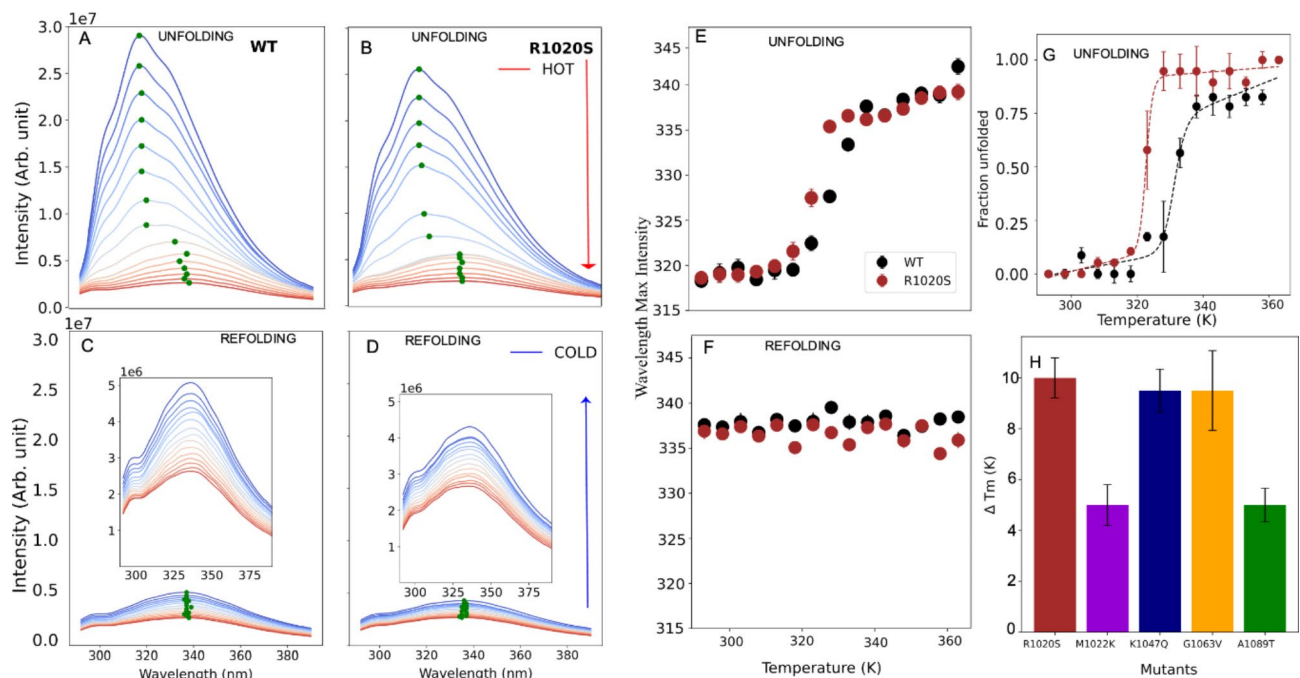
Fig. 4. Thermal denaturation experiment with wild-type ARID1A and mutants: (A and B) show unfolding, whereas (C and D) show refolding spectra plots obtained for WT and R1020S (the inset fig to zoom in view with different scale for clarity). (E and F) show the wavelength shift at max intensity for WT and R1020S as a function of temperature for unfolding and refolding, respectively. (G) normalized unfolding curve to estimate apparent melting point for wild-type and R1020S mutant, using a two-state model (Material and Methods). (H) Difference in melting temperature ($\Delta T_m = T_{m,\,WT} - T_{m,\,MUTANT}$) obtained from all mutants.

unfolding process is reversible under chemical denaturation, we conducted experiments using 0 to 4 M GuHCl (Materials and Methods) to induce protein unfolding and refolding, while monitoring the intrinsic fluorescence.

During unfolding, we initially observed a slight increase to no effect in the fluorescence intensity at low GuHCl concentrations, followed by a decrease as unfolding progressed. For all mutants, a red shift in the emission wavelength at maximum intensity was recorded with increasing GuHCl concentration (Fig. 5A and B, Figure S10). To assess the reversibility of this process, we conducted refolding experiments, observing a blue shift in the emission wavelength back to its original maximum (Fig. 5C and D, Figure S10), with over 80% of fluorescence intensity being restored. However, this partial recovery may be attributed to the extended exposure of the samples at room temperature, which nearly doubled the equilibration time for refolding (unfolding followed by refolding). We plotted the wavelength shift at maximum intensity against GuHCl concentration (Fig. 5E and F, Figure S11A and B) and normalized the data, fitting it to a two-state model (Materials and Methods) for both unfolding and refolding, clearly demonstrating reversibility (Fig. 5G and H, Figure S11C and D).

Interestingly, the unfolding and refolding pathways differed, as indicated by the $D_{1/2}$ values, which represent the GuHCl concentrations at which 50% of the protein is unfolded or refolded. We calculated the change in $\Delta D_{1/2} = D_{1/2,unfolding} - D_{1/2,refolding}$ (Fig. 5I) for all mutants, with WT showing a $\Delta D_{1/2}$ of 0.365 (not shown in the figure). The R1020S mutant exhibited the largest difference in its $\Delta D_{1/2}$ value, indicating high instability in response to chemical denaturation. We then calculated $\Delta \Delta D_{1/2} = \Delta D_{1/2,WT} - \Delta D_{1/2,MUTANT}$ (Fig. 5J), revealing that R1020S had the largest $\Delta \Delta D_{1/2}$ difference, consistent with results from other methods, highlighting it as the most affected by chemical denaturation. Surprisingly, the A1089T mutant had the smallest $\Delta D_{1/2}$ and $\Delta \Delta D_{1/2}$ values, suggesting a different response to chemical denaturation.

DISCUSSION

In this article, we evaluated the most prevalent ARID1A mutations to assess their impact on the protein's stability. By combining computational and experimental approaches, we found that the R1020S mutation causes the greatest destabilization. In contrast, the M1022K mutation exhibits the highest stability, as further supported by secondary structure content and CD spectra (Figure S4C and S12).

To evaluate the overall impact of the selected ARID1A mutations on protein stability and potential for pathogenicity, we computed a weighted average of all determined parameters (see Materials and Methods). This categorization agrees with the Pearson coefficient of 0.66 with the predictions made by the Alpha Missense (Fig. 6A, S13), a machine learning tool developed by Deep Learning, which assesses the pathogenicity of 216 million possible single amino acid changes across human proteins. AlphaMissense provides highly accurate classifications, supporting diagnostics and research on complex traits[41–43]. Our classification approach enhances the prediction of structural and functional impairments, which is essential for understanding disease mechanisms
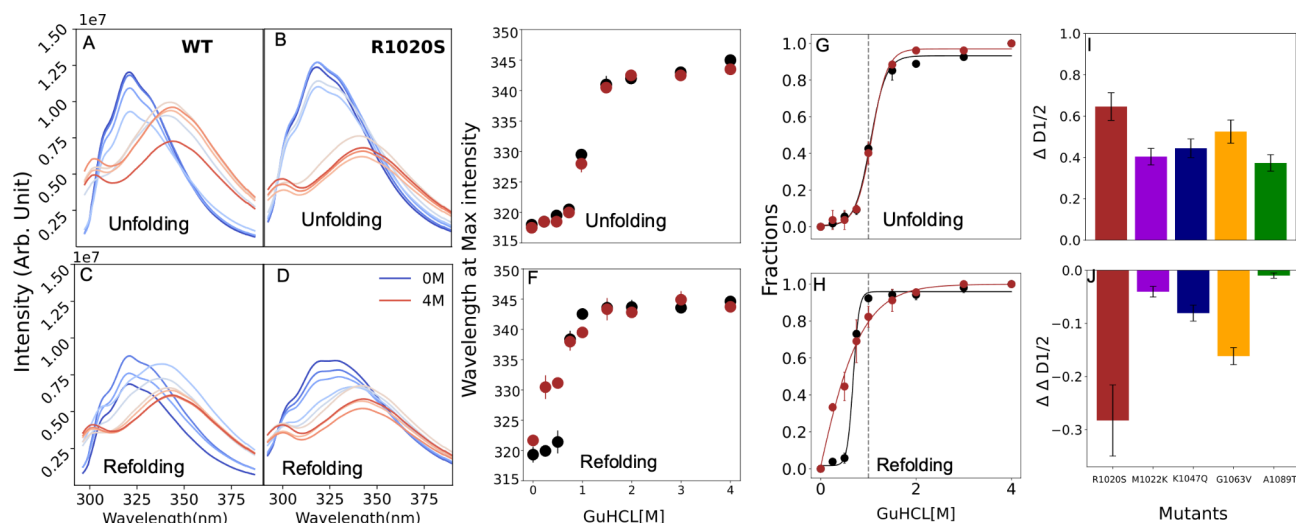
Fig. 5. Chemical denaturation experiment with wild-type ARID1A and mutants. (A and B) show unfolding, whereas (C and D) show refolding spectra plots obtained for WT and R1020S. (E and F) show the wavelength shift at max intensity for WT and R1020S as a function of GuHCl for unfolding and refolding, respectively. (G and H) normalized unfolding and refolding curve to estimate apparent $D_{1/2}$ for wild-type and R1020S mutant, using a two-state model (Material and Methods). (I and J) Difference in $D_{1/2}$ ($\Delta D_{1/2} = D_{1/2,unfolding} - D_{1/2,refolding}$) and Difference in $\Delta D_{1/2}$ ($\Delta \Delta D_{1/2} = \Delta D_{1/2,WT} - \Delta D_{1/2,MUTANT}$) obtained from all mutants.
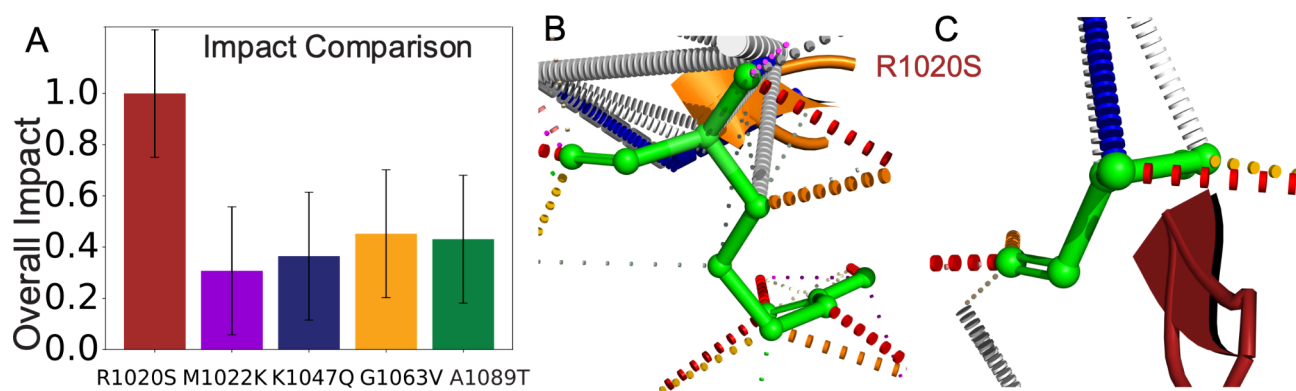
Fig. 6. The overall impact showed R1020S the most impacted. (A) The overall impact as weighted average of the computational and experimental outcomes. (B) WT interacting sites at R1020S (green) show a greater number of interaction sites, (C) The interacting sites in the mutant at R1020S (green) show fewer interacting sites.

associated with protein structural instability. The R1020S induced significant changes in local stabilizing contacts near the mutation site, as evidenced by the rexDMD simulations and AlphaMissense prediction (Figs. 3I and 6B). These insights can be translated into targeted therapeutic strategies, refining treatments tailored to individual genetic profiles, thereby advancing personalized medicine and promoting progress in drug discovery during clinical trials[44,45].

Given that most mutations occur in regions with high propensity for a-helical content, it's noteworthy that only the R1020S mutant significantly impacts stability, suggesting that inter amino acid interactions may drive this effect. We used the structures predicted by DDMut[28], which identifies detailed interaction sites within the three-dimensional context, to get insights into the number of interactions. These structural models (Fig. 6B and C) revealed that the R1020S mutation disrupts salt bridges with negatively charged residues within the same helix (H2) and helix H3, potentially altering bonding and leading to the observed instability. In contrast, the M1022K mutation introduces a positively charged residue that can form new salt bridges, enhancing stability and contributing to a more rigid structure. Although other mutations affect the a-helical content, they maintain their stability, suggesting that factors such as structural conservation of the DNA binding might be responsible for their pathogenicity. In a further study, we are exploring the direct impact on the binding free energy due to these missense mutations.

While predictive tools like web servers provide valuable initial insights, they often fall short in capturing the full biophysical and biochemical complexities of proteins, particularly in varied environmental conditions. For example, our experiments revealed significant differences between thermal and chemical denaturation, underscoring the biological relevance of these distinctions. This is critical because web servers typically assume reversible folding, a simplification that may not accurately represent the behavior of mutations susceptible to irreversible unfolding under thermal or chemical stress. Relying solely on these tools can lead to incomplete or misleading conclusions. Thus, integrating complementary experimental and computational approaches is essential (Figures S14, Improvement in Pearson correlation coefficient and validation of results with a p-value of 0.0231 using combined approaches) for a thorough understanding of protein stability and the biological impacts of mutations and paves the way for identifying and categorizing which mutations are most pathogenic, potentially guiding the development of targeted therapies tailored to specific mutation profiles in cancer treatment.

Building on these insights, the structural consequences of mutations can be further contextualized by their broader impact on critical pathways and cellular mechanisms. This is likely due to impaired chromatin remodeling and disrupted interactions with critical proteins like p53 and BRCA1[46]. These mutations may also contribute to chemoresistance, particularly against platinum-based therapies, and could affect the efficacy of emerging immunotherapies. By connecting structural instability with functional disruptions in the SWI/SNF complex, this study underscores the critical role of ARID1A in maintaining genomic integrity and regulating tumor suppressor pathways[47]. Future work that correlates specific mutations with clinical data will deepen our understanding of ARID1A's role in cancer biology and highlight potential therapeutic avenues, such as strategies to restore its function or target its downstream effects.

Our findings open several therapeutic avenues for targeting destabilizing ARID1A mutations like R1020S. Stabilizing compounds, such as small molecule chaperones or peptides, could be designed to restore structural integrity and enhance chromatin remodeling function disrupted by the R1020S mutation[48]. Alternatively, gene editing tools like CRISPR/Cas9 could correct this mutation at the genomic level, although challenges like off-target effects and efficient delivery must be addressed[49]. Small molecules or peptides that disrupt R1020S-mutant protein interactions could also offer therapeutic benefits. Combining these strategies with chemotherapy or immunotherapy may provide a comprehensive approach, while personalized treatment based on R1020S mutation profiles could optimize outcomes.

Data availability
Source files for rexDMD simulations and the intrinsic fluorescence spectra generated are available with the accession code doi:10.5281/zenodo.13712720.

References
1. Deller, M. C., Kong, L. & Rupp, B. Protein stability: a crystallographer's perspective. *Acta Crystallogr. Sect. F Struct. Biology Commun.* **72** (Pt 2), 72–95. https://doi.org/10.1107/S2053230X15024619 (2016).
2. Shirley, B. A. (ed) *Protein Stability and Folding: Theory and Practice* (Humana, 1995).
3. Bonifer, C. & Cockerill, P. N. Chromatin mechanisms regulating gene expression in health and disease. *Adv. Exp. Med. Biol.* **711**, 12–25. https://doi.org/10.1007/978-1-4419-8216-2_2 (2011).
4. Han, L. et al. Chromatin remodeling mediated by ARID1A is indispensable for normal hematopoiesis in mice. *Leukemia* **33** (9), 2291–2305. https://doi.org/10.1038/s41375-019-0438-4 (2019).
5. Mittal, P. & Roberts, C. W. M. The SWI/SNF complex in cancer — biology, biomarkers and therapy. *Nat. Reviews Clin. Oncol.* **17** (7), 435–448. https://doi.org/10.1038/s41571-020-0357-3 (2020).
6. Wilson, B. G. & Roberts, C. W. M. SWI/SNF nucleosome remodellers and cancer. *Nat. Rev. Cancer.* **11** (7), 481–492. https://doi.org/10.1038/nrc3068 (2011).
7. Kim, S., Zhang, Z., Upchurch, S., Isern, N. & Chen, Y. Structure and DNA-binding sites of the SWI1 AT-rich Interaction Domain (ARID) Suggest determinants for sequence-specific DNA recognition. *J. Biol. Chem.* **279** (16), 16670–16676. https://doi.org/10.1074/jbc.M312115200 (2004).
8. Nie, Z. et al. A specificity and targeting subunit of a human SWI/SNF family-related chromatin-remodeling complex. *Mol. Cell. Biol.* **20** (23), 8879–8888. https://doi.org/10.1128/MCB.20.23.8879-8888.2000 (2000).
9. Bakr, A. et al. ARID1A regulates DNA repair through chromatin organization and its deficiency triggers DNA damage-mediated anti-tumor immune response. *Nucleic Acids Res.* **52** (10), 5698–5719. https://doi.org/10.1093/nar/gkae233 (2024).
10. Wu, J. N. & Roberts, C. W. M. ARID1A mutations in Cancer: another epigenetic tumor suppressor? *Cancer Discov.* **3** (1), 35–43. https://doi.org/10.1158/2159-8290.CD-12-0361 (2013).
11. Wilsker, D., Patsialou, A., Dallas, P. B. & Moran, E. ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and Development1. *Cell Growth Differ.* **13** (3), 95–106 (2002).
12. Mullen, J., Kato, S., Sicklick, J. K. & Kurzrock, R. Targeting *ARID1A* mutations in cancer. *Cancer Treat. Rev.* **100**, 102287. https://doi.org/10.1016/j.ctrv.2021.102287 (2021).
13. Hein, K. Z., Stephen, B. & Fu, S. Therapeutic role of synthetic lethality in ARID1A-Deficient malignancies. *J. Immunotherapy Precision Oncol.* **7** (1), 41–52. https://doi.org/10.36401/JIPO-22-37 (2024).
14. TAKEDA, T. et al. ARID1A gene mutation in ovarian and endometrial cancers (review). *Oncol. Rep.* **35** (2), 607–613. https://doi.org/10.3892/or.2015.4421 (2016).
15. Tang, L., Nogales, E. & Ciferri, C. Structure and function of SWI/SNF Chromatin Remodeling Complexes and mechanistic implications for transcription. *Prog. Biophys. Mol. Biol.* **102** (2–3), 122–128. https://doi.org/10.1016/j.pbiomolbio.2010.05.001 (2010).
16. Bell, D. et al. Integrated genomic analyses of ovarian carcinoma. *Nature* **474** (7353), 609–615. https://doi.org/10.1038/nature10166 (2011). … Data coordination centre.
17. Rausell, A. et al. Analysis of stop-gain and frameshift variants in human innate immunity genes. *PLoS Comput. Biol.* **10** (7), e1003757. https://doi.org/10.1371/journal.pcbi.1003757 (2014).
18. Rivlin, N., Brosh, R., Oren, M. & Rotter, V. Mutations in the p53 tumor suppressor gene. *Genes Cancer.* **2** (4), 466–474. https://doi.org/10.1177/1947601911408889 (2011).

19. He, B. et al. Predicting intrinsic disorder in proteins: an overview. *Cell Res.* **19** (8), 929–949. https://doi.org/10.1038/cr.2009.87 (2009).
20. Patil, A. et al. A disordered region controls cBAF activity via condensation and partner recruitment. *Cell* **186** (22), 4936–4955e26. https://doi.org/10.1016/j.cell.2023.08.032 (2023).
21. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19** (5), 596–604. https://doi.org/10.1016/j.sbi.2009.08.003 (2009).
22. Ding, F., Tsao, D., Nie, H. & Dokholyan, N. V. Ab initio folding of proteins using all-atom discrete molecular dynamics. *Structure (London, England: 1993)*, 16(7), 1010–1018. (2008). https://doi.org/10.1016/j.str.2008.03.013
23. Li, G., Panday, S. K. & Alexov, E. SAAFEC-SEQ: a sequence-based Method for Predicting the Effect of single point mutations on protein Thermodynamic Stability. *Int. J. Mol. Sci.* **22** (2). https://doi.org/10.3390/ijms22020606 (2021).
24. Fariselli, P., Martelli, P. L., Savojardo, C. & Casadio, R. INPS: predicting the impact of non-synonymous variations on protein stability from sequence. *Bioinf. (Oxford England)*. **31** (17), 2816–2821. https://doi.org/10.1093/bioinformatics/btv291 (2015).
25. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33**(Web Server issue):W306–10. https://doi.org/10.1093/nar/gki375. PMID: 15980478; PMCID: PMC1160136 (2005).
26. Nikam, R., Kulandaisamy, A., Harini, K., Sharma, D. & Gromiha, M. M. ProThermDB: thermodynamic database for proteins and mutants revisited after 15 years. *Nucleic Acids Res.* **49** (D1), D420–D424. https://doi.org/10.1093/nar/gkaa1035 (2021).
27. Pires, D. E. V., Ascher, D. B. & Blundell, T. L. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **30** (3), 335–342. https://doi.org/10.1093/bioinformatics/btt691 (2014).
28. Zhou, Y., Pan, Q., Pires, D. E. V., Rodrigues, C. H. M. & Ascher, D. B. DDMut: predicting effects of mutations on protein stability using deep learning. *Nucleic Acids Res.* **51** (W1), W122–W128. https://doi.org/10.1093/nar/gkad472 (2023).
29. Ding, F., Tsao, D., Nie, H. & Dokholyan, N. V. Ab Initio Folding of Proteins with all-atom Discrete Molecular Dynamics. *Structure* **16** (7), 1010–1018. https://doi.org/10.1016/j.str.2008.03.013 (2008).
30. Cruz, P. et al. Domain tethering impacts dimerization and DNA-mediated allostery in the human transcription factor FoxP1. *J. Chem. Phys.* **158** (19), 195101. https://doi.org/10.1063/5.0138782 (2023).
31. He, J. et al. Enhanced label-free Nanoplasmonic Cytokine detection in SARS-CoV-2 Induced inflammation using rationally designed peptide aptamer. *ACS Appl. Mater. Interfaces.* **14** (43), 48464–48475. https://doi.org/10.1021/acsami.2c14748 (2022).
32. Huang, G. et al. Direct Observation of Seeded Conformational Conversion of hIAPP in Silico reveals the mechanisms for morphological dependence and asymmetry of Fibril Growth. *J. Chem. Inf. Model.* **63** (18), 5863–5873. https://doi.org/10.1021/acs.jcim.3c00898 (2023).
33. Song, Z., Gatch, A. J., Sun, Y. & Ding, F. Differential binding and Conformational Dynamics of Tau Microtubule-Binding Repeats with a preformed Amyloid-β Fibril seed. *ACS Chem. Neurosci.* **14** (7), 1321–1330. https://doi.org/10.1021/acschemneuro.3c00014 (2023).
34. Fuzzy supertertiary interactions within PSD-95 enable ligand binding | eLife. (n.d.). Retrieved December 2. from (2024). https://elifesciences.org/articles/77242#s4
35. Huang, G., Song, Z., Xu, Y., Sun, Y. & Ding, F. Deciphering the morphological difference of Amyloid-β fibrils in familial and sporadic Alzheimer's diseases. *J. Chem. Inf. Model.* **64** (20), 8024–8033. https://doi.org/10.1021/acs.jcim.4c01471 (2024).
36. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596** (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2 (2021).
37. Alexander, N., Woetzel, N., & Meiler, J. (2011). bcl::Cluster: A method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System.*IEEE … International Conference on Computational Advances in Bio and Medical Sciences:[proceedings]. IEEE International Conference on Computational Advances in Bio and Medical Sciences*, *2011*, 13–18. https://doi.org/10.1109/ICCABS.2011.5729867.
38. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. (n.d.). Citation for: THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Wiley Online Library*. Retrieved from https://onlinelibrary.wiley.com/action/showCitFormats?doi=10.1002%2Fjcc.540130812
39. Levine, L. D. & Ward, W. W. Isolation and characterization of a photoprotein, phialidin, and a spectrally unique green-fluorescent protein from the bioluminescent jellyfish *Phialidium gregarium*. *Comp. Biochem. Physiol. Part. B: Comp. Biochem.* **72** (1), 77–85. https://doi.org/10.1016/0305-0491(82)90013-X (1982).
40. Ibarra-Molero, B. & Sanchez-Ruiz, J. M. Irreversible protein denaturation. In (eds Roberts, G. & Watts, A.) Encyclopedia of Biophysics (1–5). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-35943-9_10067-1 (2018).
41. Sotomayor-Vivas, C., Hernández-Lemus, E. & Dorantes-Gilardi, R. Linking protein structural and functional change to mutation using amino acid networks. *PLoS ONE.* **17** (1), e0261829. https://doi.org/10.1371/journal.pone.0261829 (2022).
42. Petukh, M., Kucukkal, T. G. & Alexov, E. On human disease-causing amino acid variants: statistical study of sequence and structural patterns. *Hum. Mutat.* **36** (5), 524–534. https://doi.org/10.1002/humu.22770 (2015).
43. Cheng, J. et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381** (6664), eadg7492. https://doi.org/10.1126/science.adg7492 (2023).
44. Gerasimavicius, L., Livesey, B. J. & Marsh, J. A. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. *Nat. Commun.* **13** (1), 3895. https://doi.org/10.1038/s41467-022-31686-6 (2022).
45. Levenga, J., de Vrij, F. M. S., Oostra, B. A. & Willemsen, R. Potential therapeutic interventions for fragile X syndrome. *Trends Mol. Med.* **16** (11), 516–527. https://doi.org/10.1016/j.molmed.2010.08.005 (2010).
46. Jang, E. R. & Lee, J. S. DNA damage response mediated through BRCA1. *Cancer Res. Treatment: Official J. Korean Cancer Association.* **36** (4), 214–221. https://doi.org/10.4143/crt.2004.36.4.214 (2004).
47. Waarts, M. R., Stonestrom, A. J., Park, Y. C. & Levine, R. L. (n.d.). Targeting mutations in cancer. *J. Clin. Investig.*, **132**(8), e154943. https://doi.org/10.1172/JCI154943
48. Saibil, H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **14** (10), 630–642. https://doi.org/10.1038/nrm3658 (2013).
49. Abdelnour, S. A., Xie, L., Hassanin, A. A., Zuo, E. & Lu, Y. The potential of CRISPR/Cas9 gene editing as a treatment strategy for inherited diseases. *Front. Cell. Dev. Biology.* **9**, 699597. https://doi.org/10.3389/fcell.2021.699597 (2021).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-87103-7.

Correspondence and requests for materials should be addressed to H.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.