

# Analyzing the Effects of Single Nucleotide Polymorphisms on hnRNP A2/B1 Protein Stability and Function: Insights for Anticancer Therapeutic Design

Kunal Dutta, Viacheslav Kravtsov, Katerina Oleynikova, Alexey Ruzov, Ekaterina V. Skorb, and Sergey Shityakov\*



Cite This: *ACS Omega* 2024, 9, 5485–5495



Read Online

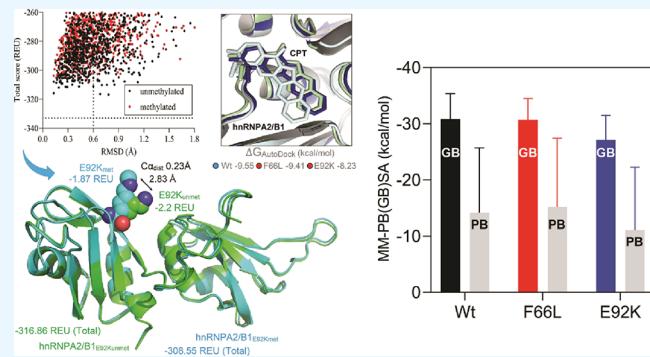
**ACCESS |**

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNPA2/B1) is a pivotal player in m6A recognition, RNA metabolism, and antiviral responses. In the context of cancer, overexpression of hnRNPA2/B1, abnormal RNA levels, and m6A depositions are evident. This study focuses on two significant nonsynonymous single nucleotide polymorphisms (nsSNPs) within hnRNPA2/B1, namely, F66L and E92K. Our structural analyses reveal decreased stability in these mutants, with E92K being predicted to undergo destabilizing post-translational methylation. Furthermore, our extensive analysis of 44,239 tumor samples from the COSMIC database uncovers that amino acid position 92 exhibits the second-highest mutation frequency within hnRNPA2/B1, particularly associated with breast and lung cancers. This experimental data aligns with our theoretical studies, highlighting the substantial impact of the nsSNP at position 92 on hnRNPA2/B1's stability and functionality. Given the critical role of pre-mRNA splicing, transcription, and translation regulation in cellular function, it is important to assess the impact of these nsSNPs on the stability and function of the hnRNPA2/B1 protein to design more efficient anticancer therapeutics.



## 1. INTRODUCTION

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are RNA-binding proteins;<sup>1</sup> they interact with heterogeneous nuclear RNA (hnRNA).<sup>2</sup> In general, hnRNPs reside in the nucleus, but some of them move back and forth between the nucleus and the cytoplasm.<sup>3</sup> The hnRNPs have unique nucleic acid binding capabilities, thanks to two quasi-RNA recognition motifs (RRMs) that bind to RNA.<sup>4</sup> In the nucleus, hnRNPs are associated with pre-mRNA processing, packaging, and other mRNA metabolism and transport-related functions.<sup>5</sup> The hnRNP A/B subfamily, i.e., the hnRNPA2/B1 protein, plays an important role in controlling mRNA transcripts for proper cellular activities.<sup>5</sup> The hnRNP A/B proteins are structurally quite similar, apart from 12 amino acids in the B1's N-terminal domain.<sup>6</sup> Conversely, the RRMs of A/B share fewer than 30% amino acid sequence similarity in their respective glycine-rich domain (GRD). However, hnRNP A/B have overall about 80% amino acid homology. hnRNPA2/B1 primarily resides in the nucleus, and trafficking of hnRNPA2/B1 to the cytoplasm is mediated by nuclear localization signals in the GRD.<sup>6</sup> In medical research, the importance of hnRNPA2/B1 is significant as many diseases are associated with hnRNPA2/B1, for instance, breast, lung, liver, and pancreatic cancers, neurodegenerative diseases, such as Alzheimer's disease, ataxia syndrome, and autoimmune

diseases.<sup>6,7</sup> Hence, drug discovery that targets hnRNPA2/B1 could represent a novel strategy as overexpression of the hnRNPA2/B1 has been manifested in cancer.<sup>8</sup> Heterogeneous nuclear ribonucleoprotein A2B1 (hnRNPA2B1) has been found to bind to m6A-bearing RNAs in vivo and in vitro.<sup>9</sup> However, there is a lack of information on the nonsynonymous single nucleotide polymorphisms (nsSNPs) in hnRNPA2B1 and their association with RNA binding sites. To our knowledge, this is the first study trying to find and analyze the nonsynonymous SNPs at the protein–RNA interface, leading to a disruption of the cell's vital functions, such as alternative splicing. Indeed, A2B1 has been found to regulate IFN- $\gamma$  signaling in macrophages through alternative splicing of the IFN- $\gamma$  receptor.<sup>10</sup> A2B1 also regulates the alternative splicing of BIRC5 to promote gastric cancer progression.<sup>11</sup> If hnRNPA2B1 does not bind to RNA, then it could potentially affect the splicing of BIRC5 and

**Received:** September 19, 2023

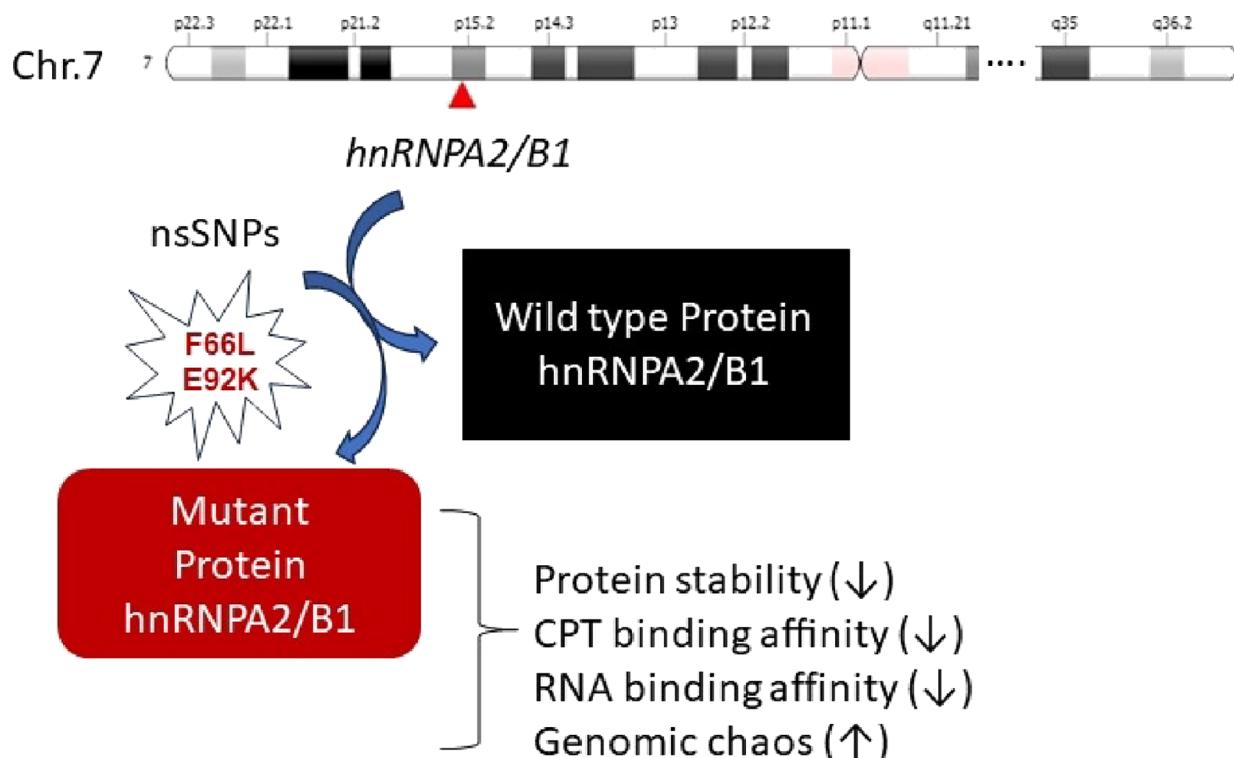
**Revised:** November 29, 2023

**Accepted:** December 8, 2023

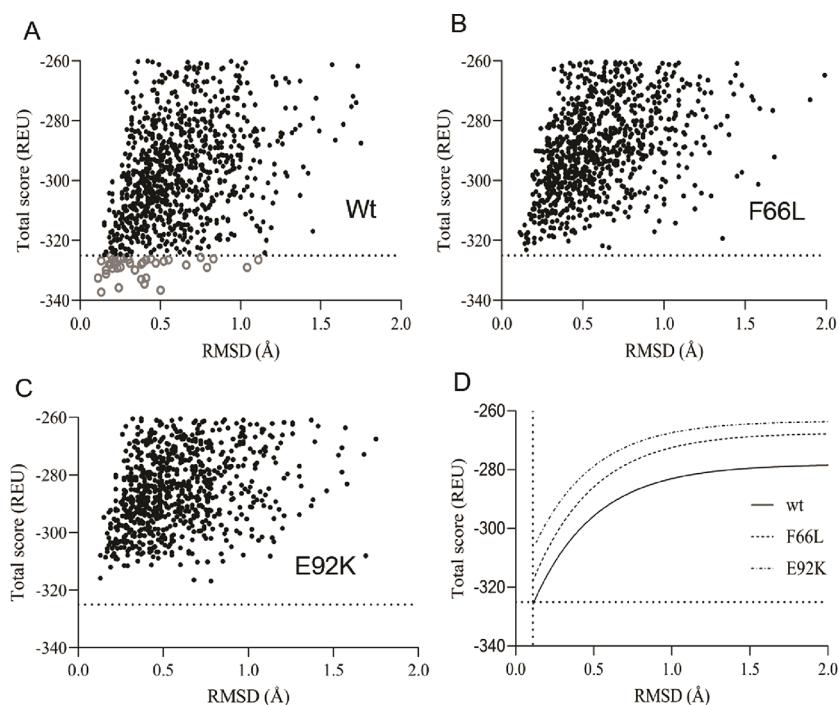
**Published:** January 26, 2024



ACS Publications



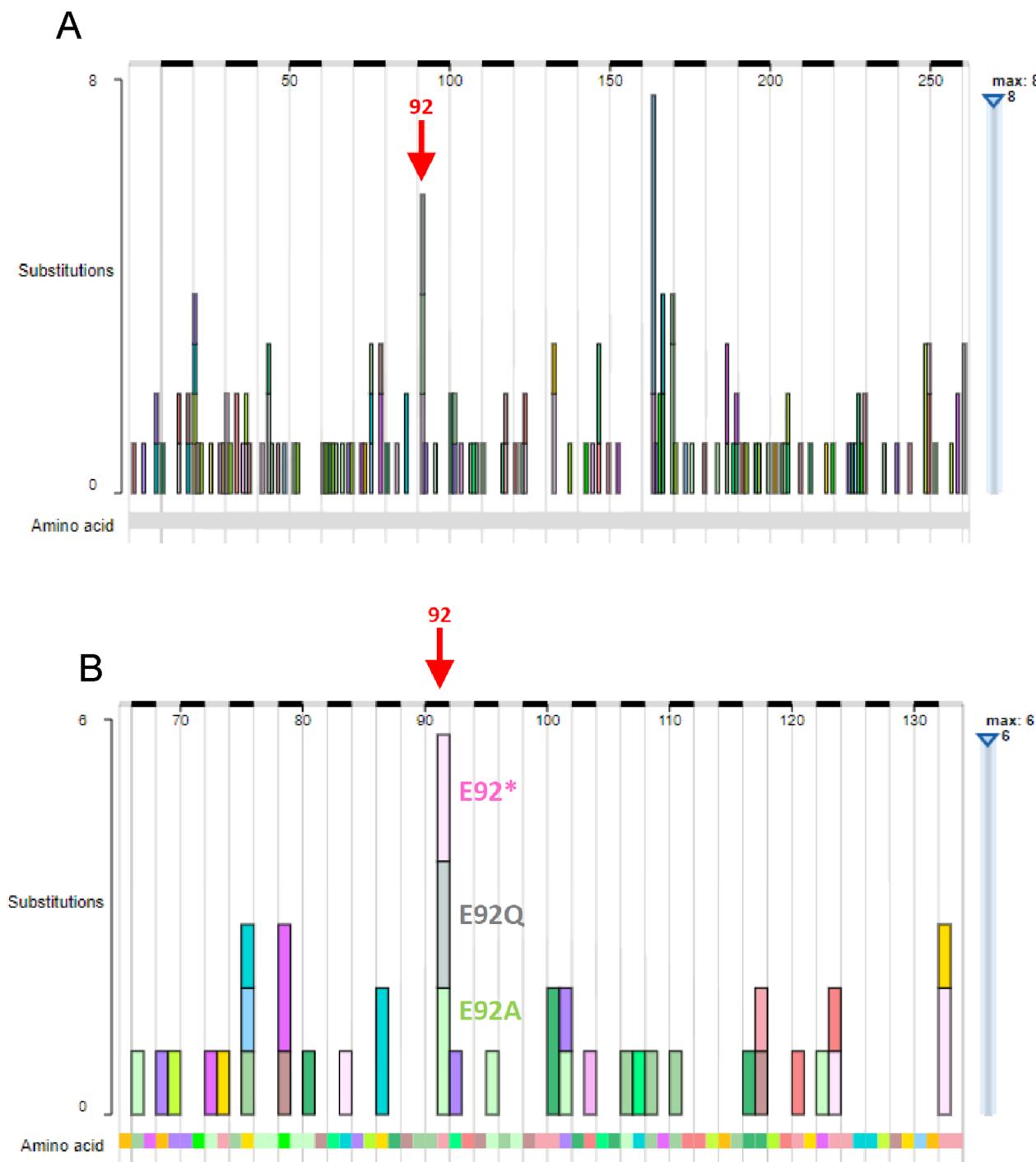
**Figure 1.** Biophysical impacts of nsSNPs on hnRNPA2B1.



**Figure 2.** Structural stability of the wild-type and mutant-type hnRNPA2/B1. Scatter plots of RMSD versus Rosetta total score for the protein (A) wild-type, (B) F66L, (C) E92K, and (D) exponential curve fitting; the wild-type hnRNPA2B1 is depicted by a black line, and F66L and E92K are depicted by a different pattern of dotted lines. RMSD, root-mean-square deviation; REU, Rosetta energy units.

other genes, which could have implications for disease progression. Nonetheless, there are a few studies on the effects of mutations in hnRNPA2/B1.<sup>12,13</sup> And it is unknown how single nucleotide polymorphisms (SNPs), particularly nsSNPs, affect the stability of the hnRNPA2/B1 and binding sites for nascent mRNA and inhibitory drugs. Hence, we investigated the

effects of nsSNPs on the stability of hnRNPA2/B1 and its binding to RNA, and camptothecin (CPT, BDBMS0008923) using multiple innovative bioinformatics tools. The results showed that nsSNPs, i.e., F66L and E92K, could be pathogenic, with E92K being modified by methylation, and could lead to a



**Figure 3.** Two histograms displaying the positions of mutations (A) and single base substitutions (B) across the hnRNPA2/B1 (ENST00000618183) gene from COSMIC (the Catalogue of Somatic Mutations in Cancer). These mutations are shown at the amino acid level at two different resolutions. Amino acid position 92 is highlighted with a red arrow in both diagrams. The substitutions are color-coded by residue according to the scheme used in Ensembl. The visualization was performed by using COSMIC online tools.

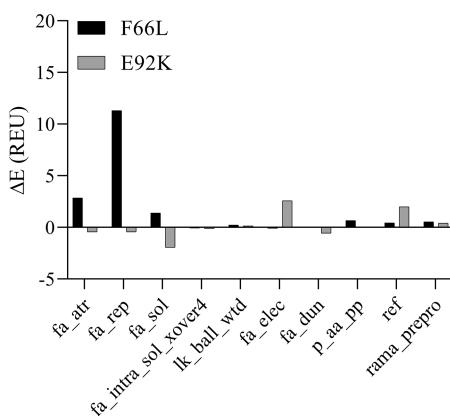
decrease in the RNA and CPT binding affinities than that of the wild-type hnRNPA2/B1.

## 2. COMPUTATIONAL METHODS

**2.1. Data Collection, Prediction, and Search Analyses of SNPs.** Prediction and search analyses of SNPs of the human hnRNPA2/B1 were conducted, as described by our previous work.<sup>14</sup> In brief, the NCBI-SNP database and UniProt database were used for further analyses of nsSNPs.

**2.2. Identification of Deleterious nsSNPs.** Deleterious nonsynonymous mutations were identified using the methods described in our previous work.<sup>14</sup> In brief, the SIFT tool and PolyPhenV2 algorithms were used to determine any deleterious nsSNPs in hnRNPA2/B1. SNP variants were predicted by using the SNP predictor tool.

**2.3. nsSNP Search Analyses in Various Databases.** Data regarding the occurrence of these nsSNPs in real-world cases were incorporated. For that, we conducted an exhaustive



**Figure 4.** Rosetta energy decomposition analyses for substitution mutants, i.e., F66L and E92K, of hnRNPA2/B1. fa\_atr, Lennard-Jones attractive between atoms in different residues; fa\_rep, Lennard-Jones repulsive between atoms in different residues; fa\_sol, Lazaridis-Karplus solvation energy; fa\_intra\_sol\_xover4, intraresidue Lazaridis-Karplus solvation, counted for the atom pairs beyond the torsion relationship; lk\_ball\_wtd, a weighted sum of anisotropic contribution to the solvation; fa\_elec, Coulombic electrostatic potential with a distance-dependent dielectric; fa\_dun, internal energy of side chain rotamers as derived from Dunbrack's statistics; p\_aa\_pp, the probability of amino acid at  $\Phi/\Psi$  angles; ref, reference energy for each amino acid, balancing internal energy of amino acid terms; rama\_prep, backbone torsion preference term that takes into account of whether the preceding amino acid is proline or not; REU, Rosetta energy units.

**Table 1. Molsoft Protein Stability Prediction**

protein	chain	residue	wild-type	mutant	$\Delta E_{\text{mol}}$	$\Delta E_{\text{ros}}$
hnRNPA2/B1 <sub>F66L</sub>	A	66	Phe	Leu	1.33	14.18
hnRNPA2/B1 <sub>E92K</sub>	A	92	Glu	Lys	0.18	21.4

analysis of various public databases, including TCGA (The Cancer Genome Atlas), COSMIC (Catalogue of Somatic Mutations in Cancer), and ClinVar (the genomic variation database).

**2.4. Protein Stability Prediction.** The Molsoft ( $\Delta E_{\text{mol}}$ ) and Rosetta ( $\Delta E_{\text{ros}}$ ) protocols were used to predict the protein stability via measuring the energy differences between mutated ( $E_{\text{mut}}$ ) and wild-type forms ( $E_{\text{wt}}$ ) and were calculated using the following equation:

$$\Delta E = E_{\text{mut}} - E_{\text{wt}}$$

**2.5. Protein Stability and Interface Analysis.** The protein structure of hnRNPA2/B1 in complex with 10-mer RNA (SHO4)<sup>15</sup> was obtained from RCSB-Protein Data Bank. Protein stability was analyzed using Molsoft ICM-Pro (Molsoft LLC, San Diego, CA, USA).

**2.6. Conservation Analyses.** Conservation analysis was performed by the protocol described by our previous studies.<sup>14,16</sup> The normalized consensus hydrophobicity scale<sup>17</sup> was used for the prediction of the conservation profile of the hnRNPA2/B1. Additionally, the ConSurf web server was used for such analyses.<sup>18</sup>

**2.7. Prediction of Post-Translational Modifications.** The post-translational modifications in the wild-type hnRNPA2/B1 structure, F66L, and E92K mutants were predicted by using MusiteDeep available at <https://www.musite.net>.<sup>19</sup> The MutPred2 web server (<http://mutpred.mutdb.org/index.html>)<sup>20</sup> was used to predict the impact of

F66L and E92K substitutions on hnRNPA2/B1, as described by our previous work.<sup>14</sup> Additionally, the PyTM plugin of PyMOL was used for E92K methylation.<sup>21</sup>

**2.8. Gene Association Network Analysis.** The gene–gene interaction network was studied using the GeneMANIA tool, as described by our previous work.<sup>14</sup>

### 2.9. Interference Analysis and Receptor–Drug Affinity Predictions.

The analyses of molecular interactions were performed using in-house PyMOL scripts as described by our previous study with the normalized consensus hydrophobicity scale.<sup>14</sup> The CASTp server<sup>22</sup> was used for binding site identification (job id: j\_63ca8c75e54d8). AutoDock<sup>23</sup> molecular docking was used for the molecular interaction between CPT and hnRNPA2/B1. The center grid dimensions were set to  $-30.110 \times 5.710 \times 7.930 \text{ \AA}$  with a grid spacing of  $0.375 \text{ \AA}$ . Additionally, the FoldX software was used for molecular docking between RNA and human hnRNPA2/B1.<sup>24</sup> And structure-truncated MM/PB(GB)SA (molecular mechanics/Poisson–Boltzmann/generalized Born surface area) rescoring and hotspot predictions were performed using the fastDRH server.<sup>25</sup>

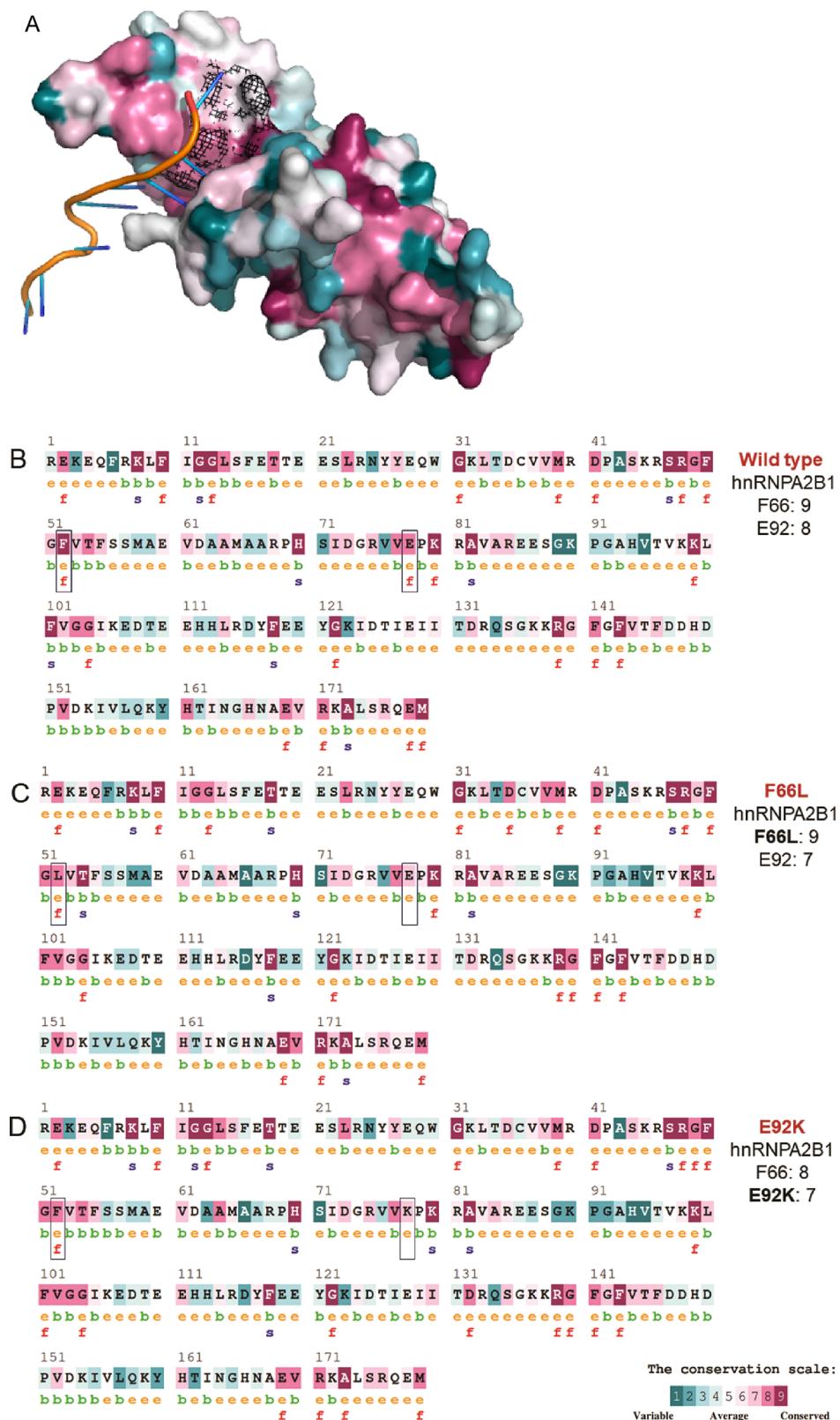
## 3. RESULTS

**3.1. Identification of SNPs.** We identified a total of 8494 SNPs in human hnRNPA2/B1. While 1303 were found in the 3'-UTR region, 858 of them were found in the 5'-UTR region. In addition, further examination of SNPs using the SIFT tool revealed 157 nonsynonymous and 204 synonymous SNPs and 6 PhD-SNPs (supplementary file, Table S1). Two important nsSNPs, F66L and E92K (Figure 1), among PhD-SNPs were selected for a detailed study based on their probability (>70%) (Table S2).

**3.2. Analysis of nsSNP in Various Databases.** Analyzing a comprehensive data set of 44,239 unique tumor samples from COSMIC, we observed that 469 unique samples contained mutations. Intriguingly, amino acid position 92 exhibited the second-highest mutation frequency within hnRNPA2/B1 (Figure 3). Specifically, we identified the following substitutions: E92A (275A > C, observed in two samples), E92Q (274G > C, observed in two samples), and E92\* (274G > T, observed in two samples), as visually represented in Figure 3B. These substitutions appear to be notably associated with breast and lung cancers. Overall, the experimental data presented here align with our theoretical studies, where the nsSNP at position 92 exerts the most substantial influence on the stability and functionality of HNRNPA2B1 (Figure 4). Overall, the experimental data presented here aligns with our theoretical studies, where the nsSNP at position 92 exerts the most substantial influence on the stability and functionality of hnRNPA2/B1.

**3.3. Protein Stability Prediction.** The stability of hnRNPA2/B1<sub>F66L</sub> and hnRNPA2/B1<sub>E92K</sub> mutants were significantly diminished according to the Rosetta standard energy function (Figure 2) and MolSoft stability prediction (Table 1). The nonsynonymous SNPs F66L and E92K of hnRNPA2/B1 showed  $\Delta E_{\text{mol}}$  of 1.33 and 0.1804 kcal/mol and  $\Delta E_{\text{ros}}$  of 14.18 and 21.4 REU, respectively. Moreover, it is important to note that the F66L substitution mutant showed a higher Lennard-Jones attractive as well as repulsive potential between atoms in different residues, but these parameters were negative for E92K (Figure 4).

The Lazaridis-Karplus solvation energy was found positive for F66L substitution but not for E92K. Nonetheless, electrostatic potential with a distance-dependent dielectric was found



**Figure 5.** Evolutionary conservation profiles of hnRNP A2/B1. (A) Color protein molecule according to the Eisenberg hydrophobicity scale. (B) Wild-type, (C) F66L, (D) E92K, and their respective conservation scores (right); F66, E92, and substitutions are highlighted by black rectangles. Conservation profile scale depicted by a color-coded map (bottom) according to the NACCESS algorithm.

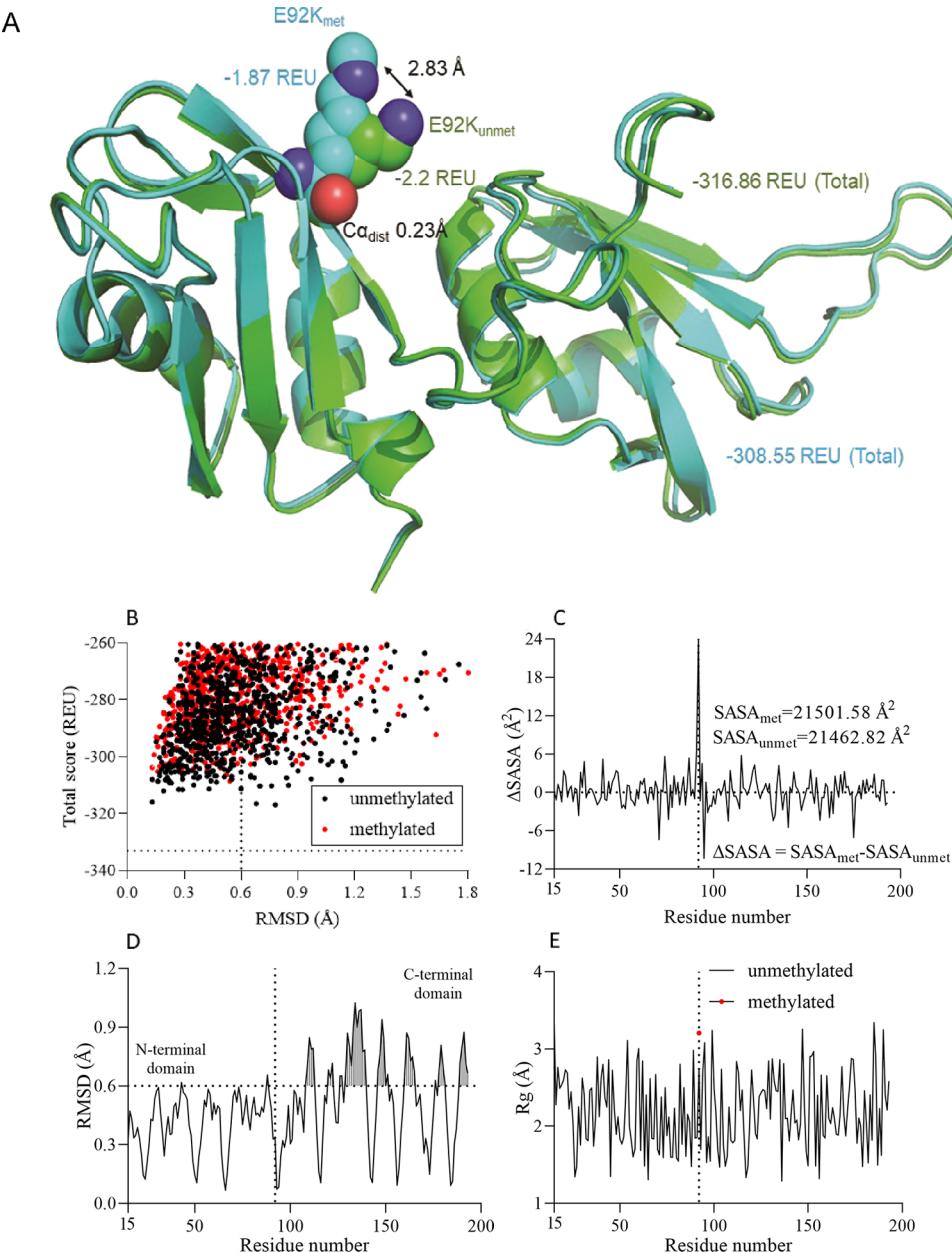
positive for E92K substitution. The probability term of amino acid at  $\Phi/\Psi$  angles was positive for F66L. Finally, the backbone torsion preference term was positive for both F66L and E92K.

**3.4. Conservation Analysis.** The analysis of evolutionary conservation profiles reveals notable conservation and exposure of amino acids F66 and E92 in the wild-type hnRNP A2/B1, with ConSurf prediction scores of eight and nine, respectively (Figure

**Table 2. Predicted Post-Translational Modifications and Pathogenicity of the Wild-Type and Substitution Mutants of hnRNPA2/B1<sup>a</sup>**

position	PTM scores	SNP type	MutPred2 score	SNP effect	PTMs
F66L		NS	0.88	pathogenic	
E92K	0.68	NS	0.78	pathogenic	methylation

<sup>a</sup>NS, nonsynonymous.

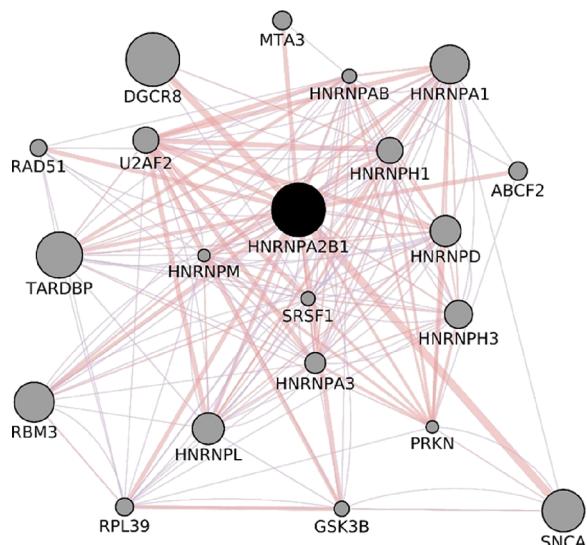


**Figure 6.** Structural stability analyses before and after post-translational modification E92K of hnRNPA2/B1. (A) Cartoon model of hnRNPA2/B1, E92K methylated (cyan), and unmethylated (green). Ca<sub>dist</sub>, a difference of distance between the  $\alpha$  carbon of E92K<sub>met</sub> and E92K<sub>unmet</sub>. REUs of E92K methylated (cyan) and unmethylated (green) are shown next protein structures, (B) scatter plots of RMSD versus Rosetta total score, (C) SASA versus amino acid residue number, (D) RMSD versus amino acid residue number, (E) R<sub>g</sub> versus amino acid residue number of methylated, unmethylated E92K of hnRNPA2/B1. REU, Rosetta energy units; RMSD, root-mean-square deviation; R<sub>g</sub>, radius of gyration. A threshold is depicted by horizontal dotted lines; E92K<sub>met</sub> and E92K<sub>unmet</sub> are marked by vertical dotted lines.

5). However, when substitutions such as F66L and E92K are considered, a subtle alteration in the conservation profile is observed. In the case of the F66L substitution, the ConSurf prediction score at E92 decreases to seven. Despite this, the ConSurf prediction still categorizes F66L as a functionally

significant, highly conserved, and exposed residue. Conversely, for the E92K substitution, the ConSurf prediction score also drops to seven, aligning with the NACCESS algorithm in predicting it as an exposed residue.

**3.5. Post-Translational Modification Protein Stability and Network Analyses.** The MusiteDeep server did not predict post-translational modifications at F66 and E92 in the wild-type hnRNPA2/B1. Interestingly, the substitution E92K was forecasted to undergo methylation, with a noteworthy PTM score of 0.68. In contrast, the F66L substitution did not exhibit predictions for post-translational modification. Notably, both substitutions, F66L and E92K, were anticipated to be pathogenic, given their MutPred2 scores exceeding 0.5 (Table 2). Furthermore, according to the standard Rosetta energy function, the structural stability of hnRNPA2/B1, E92K<sub>met</sub>, was diminished compared to that of hnRNPA2/B1, E92K<sub>unmet</sub>. And the SASA of the E92K<sub>met</sub> was 38.76 Å<sup>2</sup> higher than E92K<sub>unmet</sub> (Figure 6). The molecular interaction network of human hnRNPA2/B1 showed that few proteins have participated in physical interactions (bold line) with the hnRNPA2/B1 such as DGCR8, SNCA, MTA3, and U2AF2. Conversely, many nodes are coexpressed (normal line) with hnRNPA2/B1 (Figure 7).



**Figure 7.** Molecular interaction network of human hnRNPA2/B1. The black node at the center, hnRNPA2/B1, physical interactions (bold line) and coexpression (line), and degree of interactions are depicted proportionally by the size of the nodes.

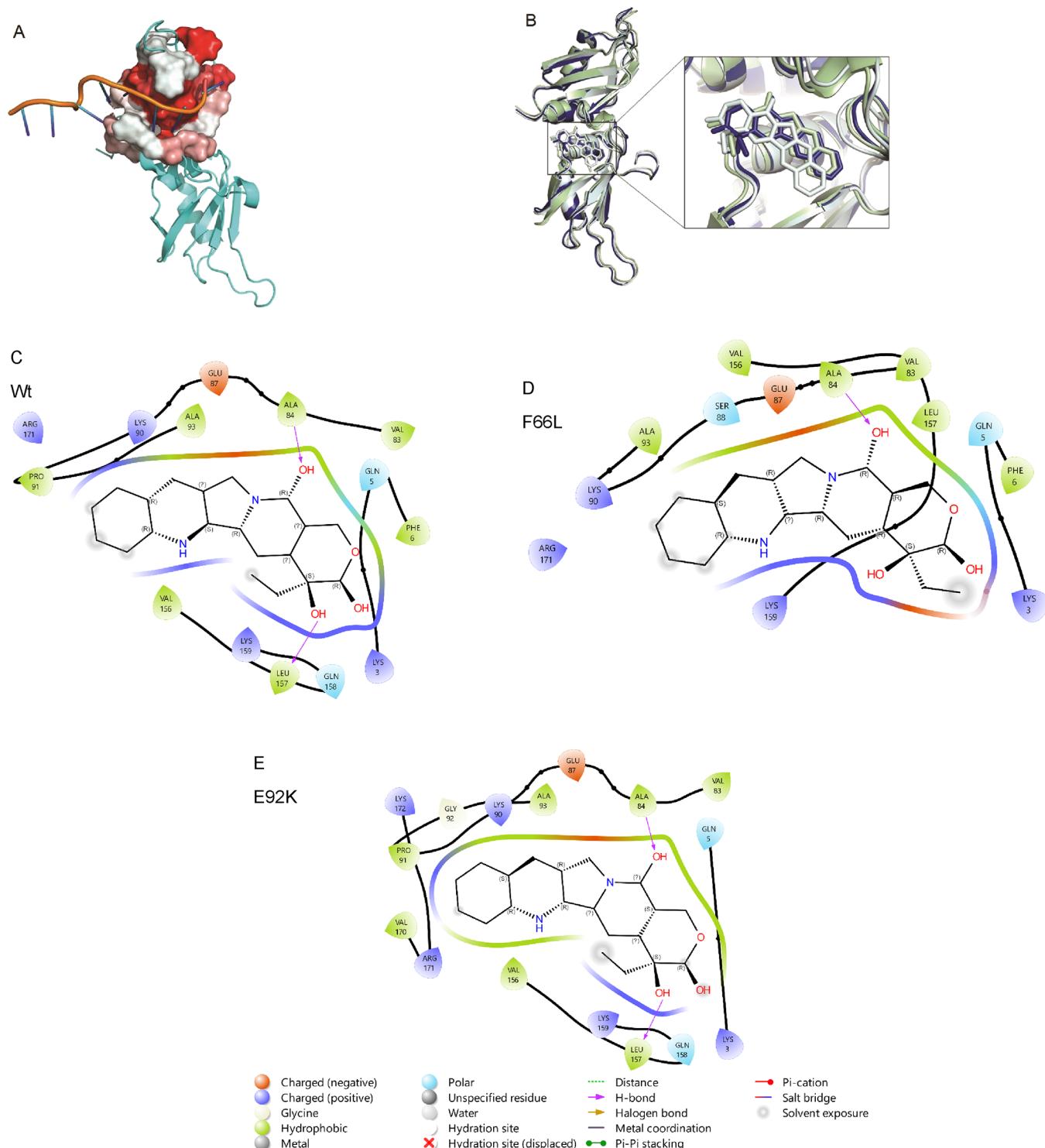
**3.6. Binding Site Analyses.** The RNA binding site is situated near the CPT binding site. The results depicted distinct CPT binding poses for the wild-type hnRNPA2/B1 and the substitution mutants E92K and F66L (Figure 8). Notably, the nsSNPs F66L and E92K were found to be distantly located from the CPT binding site of hnRNPA2/B1. The 2D protein–ligand interaction plot revealed that A84 was a shared amino acid engaged in hydrogen bonding with the exposed hydroxyl group of CPT bound to the wild-type, F66L, and E92K substitution mutants. Additionally, L157 of wild-type and E92K participated in hydrogen bonding with the exposed hydroxyl group of CPT (Figure 8C–E).

**3.7. Molecular Docking.** The results obtained from the protein–RNA molecular docking among the wild-type and substitution mutants F66L and E92K of hnRNPA2/B1 showed a significant reduction of RNA binding affinity of the mutants (Table 4). In addition, a molecular docking study between hnRNPA2/B1<sub>wt</sub> and CPT showed a  $\Delta G_{\text{AutoDock}}$  of -9.55 kcal/mol. On the other hand, the  $\Delta G_{\text{AutoDock}}$  values of the substitution mutants F66L and E92K were -9.41 and -8.23 kcal/mol,

respectively. Consistent trends were noted even after MM/PB(GB)SA rescoring, indicating a reduced affinity of the mutants to bind the ligand. However, it's worth noting that the F66L mutant exhibited an exception to this pattern when utilizing the MM-PBSA solvation model.

#### 4. DISCUSSION

In the nucleus, hnRNPA2/B1 plays important roles in pre-mRNA processing, regulation of genes encoding long noncoding RNA (lncRNA), RNA packaging, mRNA metabolism, mRNA transport-related functions,<sup>15</sup> and also during tumorigenesis.<sup>26</sup> Moreover, hnRNPA2/B1 reads N6-methyladenosine (m6A) and facilitates mRNA processing by interacting with the DGCR8,<sup>9,27</sup> crucial for genomic stability. The stability of the genome and RNA:DNA hybrids are regulated by m6A depositions. Also, the formation of R-loops is common in DNA:RNA hybrids, and these are associated with genomic chaos in malignant cells.<sup>28</sup> Furthermore, recent work by Humphries and Fitzgerald<sup>29</sup> showed that hnRNPA2/B1 acts as a nuclear receptor for herpes virus (dsDNA) and via a cGAS-independent pathway, it activates a type-I interferon (IFN) response.<sup>30</sup> And hnRNPA2/B1 can amplify the antiviral response by controlling the exportations of cGAS and STING mRNAs.<sup>29</sup> Similarly, Zhang et al. (2019) reported that hnRNPA2/B1 acts as a DNA sensor for antiviral immunity.<sup>31</sup> These findings imply important functional roles of hnRNPA2/B1 in normal cellular processes. Fundamentally, functions of a protein are defined and confined by the protein primary structure (i.e., amino acid sequence) that determines its three-dimensional structure.<sup>32</sup> However, a nsSNP may impair the structure and functions of a protein; for instance, Lim et al. (2021) observed alteration of the DNA-binding specificities of MYB family proteins due to nsSNPs and PTMs.<sup>33</sup> The results obtained from the SNP search and analyses revealed two important nsSNPs, namely, F66L and E92K, and both were predicted as a pathogenic substitution (Table 2). Protein's stability is an indicator of how well a protein keeps a particular shape. A stable protein structure is associated with lower Rosetta energies.<sup>34</sup> However, structural stability analyses using the standard Rosetta energy function imply a significant reduction of stabilities of the F66L and E92K substitutions (Figure 2), which are also in agreement with the results obtained from Molsoft stability prediction (Table 1). It is noteworthy that E92K and F66L have been referenced in a paper about the identification of gene variants associated with SARS-CoV-2.<sup>35</sup> In addition, the X-ray model of hnRNPA2B1 showed better structural stability compared with the AlphaFold model of hnRNPA2B1 (Figure S1). The variances in structural stabilities observed between the X-ray and AlphaFold models may arise from the presence of an intrinsically disordered domain, given that protein folding is in close proximity to a threshold of 3 Å.<sup>36</sup> Nonetheless, close to the ideal 0.5 Å RMSD<sup>37</sup> was found for the Trp-Cage protein folded by the AlphaFold2 algorithm. Furthermore, the E92K substitution was predicted to be modified by methylation with a good PTM prediction score of 0.68 (Table 2). Post-translational modification of proteins plays a vital role in cell signaling.<sup>38</sup> For instance, phosphorylation of a protein leads to the activation of protein kinases and vice versa. Similarly, methylation of protein is linked with cell signaling, cell differentiation, and cell proliferation.<sup>38</sup> Notably, the stability of E92K<sub>met</sub> (-1.87 REU) was lower than that of E92K<sub>unmet</sub> (-2.2 REU). The distance between Cα of E92K<sub>met</sub> and E92K<sub>unmet</sub> was 0.23 Å, confirming a structural alteration (Figure 6a). Further,



**Figure 8.** Binding site analyses of hnRNP A2/B1. (A) hnRNP A2/B1 docked with RNA; the drug binding site is depicted by protein surface presentation. (B) Superimposed protein structure of hnRNP A2/B1 docked with CPT wild-type (blue), F66L (pale green), E92K (pale cyan), and 2D ligand interactions; (C) wild-type, (D) F66L, and (E) E92K.

the total REU of the hnRNP A2/B1<sub>E92K<sub>met</sub></sub> was also lower than that of hnRNP A2/B1<sub>E92K<sub>unmet</sub></sub> suggesting methylation-induced structural instability. Furthermore, structural alteration of E92K<sub>met</sub> and E92K<sub>unmet</sub> is supported by the dihedral angle analysis of the substitution mutants E92K<sub>met</sub> and E92K<sub>unmet</sub> where the  $\psi$  and  $\omega$  dihedral angles were affected (Table 3). These dihedral angles are crucial determinants of the protein's overall conformation and stability, and any modifications or

disturbances in the structural context may impact their values. Similarly, reductions of stabilities due to nsSNPs in BCL-2 were reported by Fareed et al. (2022),<sup>14</sup> in WFS1 by Zhao et al. (2023),<sup>39</sup> and in IL-8 by Dakal et al. (2017).<sup>40</sup> Moreover, the stability of a protein is also linked with the evolutionary rate of its amino acid sequence.<sup>41</sup> In other words, the evolutionary conservation rate of an amino acid in a particular position of a protein strongly relies on its structural and functional status and

**Table 3. Dihedral Angle Analysis of the Substitution Mutants E92K<sub>met</sub> and E92K<sub>unmet</sub>**

angle	unmethylated	methylated
$\varphi$	-131.98°	-131.95
$\psi$	87.08°	161.87
$\omega$	-176.31	87.13
$\chi_1$	-175.51	-175.51
$\chi_2$	177.99	177.98
$\chi_3$	178.49	178.49
$\chi_4$	-177.33	-177.33

thus is associated with its stability. The ConSurf server is an excellent bioinformatics tool that can predict the evolutionary conservation profile of a given protein structure.<sup>18</sup> The results obtained from the ConSurf server showed that F66 and F92 of hnRNP A2/B1<sub>wt</sub> and F66L of hnRNP A2/B1<sub>F66L</sub> are predicted functional, highly conserved, and exposed residue(s) (Figure 4). Notably, Kim et al. (2013) observed multisystem proteinopathy and amyotrophic lateral sclerosis owing to mutations in the porin-like domain of the hnRNP A2/B1 and hnRNP A1 proteins.<sup>12</sup> In addition, the results depicted by physical interactions and coexpression of nodes indicate essential functions of the human hnRNP A2/B1 protein (Figure 7). For instance, DiGeorge syndrome critical region 8 (DGCR8),<sup>42</sup> metastasis-associated protein 3 (MTA3),<sup>43</sup> and splicing factor U2AF2 (U2AF2),<sup>44</sup> all these proteins play essential roles in cancer, and the results showed that they have participated in physical intersections with hnRNP A2/B1. Computational chemistry is an established branch of science, and it can precisely predict the behaviors of molecules by using innovative computational tools. Notably, wild-type hnRNP A2/B1 showed molecular docking energies for CPT and RNA better than those of the substitution mutants F66L and E92K (Table 4).

**Table 4. Summary of Molecular Docking and Structure-Truncated MM/PB(GB)SA Rescoring<sup>a</sup>**

proteins	total energy (kcal/mol)				
	fastDRH*				
	FoldX protein– RNA docking	protein– CPT docking	GB	PB	
Wt	-18.52	-9.55	-30.83 ± 4.54	-14.21 ± 11.50	
F66L	-16.40	-9.41	-30.72 ± 3.77	-15.23 ± 12.17	
E92K	-15.42	-8.23	-27.13 ± 4.35	-11.07 ± 11.18	

<sup>a</sup>Wt, wild-type; GB, generalized Born model; PB, Poisson–Boltzmann model. \*Force field: ff99sb for receptor, gaff2 for ligand, and TIP3P water model.

Furthermore, predicted hotspot residues of wild-type hnRNP A2/B1 showed better per amino acid residue MM-GBSA energy (kcal/mol) than those of hnRNP A2/B1<sub>F66L</sub> and hnRNP A2/B1<sub>E92K</sub> (Figure S2). Hence, we speculate that nsSNPs, i.e., F66L, E92K, and PTM (E92K<sub>met</sub>), together may impair normal functions of the hnRNP A2/B1 as the sequence–structure–function paradigm is fundamental in molecular biology.<sup>45</sup>

## 5. CONCLUSIONS

Human hnRNP A2/B1 plays crucial roles in diverse biological processes, including carcinogenesis, viral immunity, and RNA metabolism. Serving as a molecular reader of N6-methyladeno-

sine (m6A), hnRNP A2/B1 contributes to mRNA processing, particularly in collaboration with DGCR8, vital for maintaining genomic stability. In the realm of cancer research, nsSNPs within hnRNP A2/B1 exert profound impacts, disrupting the intricate RNA life cycle and fostering genomic instability associated with R-loop formation and double-stranded DNA breaks. Our study identifies two pathogenic nsSNPs, F66L and E92K, with F66L considered more conserved and E92K prone to structural instability, especially after methylation, leading to severe functional impairment. Both substitutions exhibit reduced binding affinities for CPT and RNA compared to wild-type hnRNP A2/B1. Given the observed dysregulation in cancer contexts, particularly in m6A disposition, RNA levels, and hnRNP A2/B1 overexpression, scrutinizing F66L and E92K nsSNPs could offer valuable insights for designing epigenetic drugs aimed at altering RNA metabolism in cancer scenarios.

## ASSOCIATED CONTENT

### Data Availability Statement

All data is available on GitHub: <https://github.com/virtualscreenlab/hnRNPA2B1>.

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.3c07195>.

(Table S1) Characteristics of single nucleotide polymorphisms in hnRNP A2/B1. (Table S2) Prediction of single nucleotide polymorphisms in hnRNP A2/B1. (Figure S1) Structural stability of the hnRNP A2/B1 protein. (A) X-ray model of hnRNP A2/B1; (B) AlphaFold model; thresholds are depicted by dotted lines. (Figure S2) Predicted hotspot residues. (A) hnRNP A2/B1 wild-type, (B) F66L, and (C) E92K. Similar amino acid residues are depicted by the asterisk sign ([PDF](#))

## AUTHOR INFORMATION

### Corresponding Author

Sergey Shityakov — Infochemistry Scientific Center, ITMO University, St. Petersburg 191002, Russian Federation;  
[ORCID](https://orcid.org/0000-0002-6953-9771); Phone: +7 981-955-03-82; Email: shityakoff@hotmail.com

### Authors

Kunal Dutta — Infochemistry Scientific Center, ITMO University, St. Petersburg 191002, Russian Federation;  
[ORCID](https://orcid.org/0000-0002-0818-8787)

Viacheslav Kravtsov — Infochemistry Scientific Center, ITMO University, St. Petersburg 191002, Russian Federation

Katerina Olynikova — Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow 119071, Russian Federation

Alexey Ruzov — Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow 119071, Russian Federation

Ekaterina V. Skorb — Infochemistry Scientific Center, ITMO University, St. Petersburg 191002, Russian Federation;  
[ORCID](https://orcid.org/0000-0003-0888-1693)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acsomega.3c07195>

### Author Contributions

K.D.: software, writing—original draft; S.S.: data curation, investigation, resources, software, and validation; V.K., E.V.S.,

K.O., and A.R.: investigation, methodology, validation, and writing—review and editing; S.S.: conceptualization, investigation, methodology, project administration, supervision, validation, and writing—review and editing.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by grant from the Russian Science Foundation (RSF), Russia, no. 263 22-65-00022. This research received financial support from the Government of the Russian Federation through the ITMO Fellowship and Professorship Program. Additionally, partial funding for KO and AR was provided by the Ministry of Higher Education of the Russian Federation under Project number 1220411001149-7. Graphical abstract was prepared using an image by brgfx on Freepik.

## REFERENCES

- (1) Wang, J.; Sun, D.; Wang, M.; Cheng, A.; Zhu, Y.; Mao, S.; Ou, X.; Zhao, X.; Huang, J.; Gao, Q.; Zhang, S.; Yang, Q.; Wu, Y.; Zhu, D.; Jia, R.; Chen, S.; Liu, M. Multiple functions of heterogeneous nuclear ribonucleoproteins in the positive single-stranded RNA virus life cycle. *Front. Immunol.* **2022**, *13*, No. 989298.
- (2) Görlich, M.; Burd, C. G.; Portman, D. S.; Dreyfuss, G. The hnRNP proteins. *Molecular biology reports* **1993**, *18*, 73–78.
- (3) Kim, J. H.; Hähm, B.; Kim, Y. K.; Choi, M.; Jang, S. K. Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *Journal of molecular biology* **2000**, *298* (3), 395–405.
- (4) Caputi, M.; Zahler, A. M. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J. Biol. Chem.* **2001**, *276* (47), 43850–43859.
- (5) He, Y.; Smith, R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cellular and molecular life sciences* **2009**, *66*, 1239–1256.
- (6) Liu, Y.; Shi, S. L. The roles of hnRNP A2/B1 in RNA biology and disease. *Wiley Interdisc. Rev.: RNA* **2021**, *12* (2), No. e1612.
- (7) Zhou, J.; Allred, D.; Avis, I.; Martínez, A.; Vos, M. D.; Smith, L.; Treston, A. M.; Mulshine, J. L. Differential expression of the early lung cancer detection marker, heterogeneous nuclear ribonucleoprotein-A2/B1 (hnRNP-A2/B1) in normal breast and neoplastic breast cancer. *Breast cancer research and treatment* **2001**, *66*, 217–224. Cui, H.; Wu, F.; Sun, Y.; Fan, G.; Wang, Q. Up-regulation and subcellular localization of hnRNP A2/B1 in the development of hepatocellular carcinoma. *BMC Cancer* **2010**, *10*, 1–14. Yan-Sanders, Y.; Hammons, G. J.; Lyn-Cook, B. D. Increased expression of heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP) in pancreatic tissue from smokers and pancreatic tumor cells. *Cancer letters* **2002**, *183* (2), 215–220.
- (8) Xu, H.; Li, P.; Wang, X.; Zhuang, H.; Hua, Z.-C. Emerging roles of hnRNP A2B1 in cancer and inflammation. *Int. J. Biol. Macromol.* **2022**, *221*, 1077.
- (9) Alarcón, C. R.; Goodarzi, H.; Lee, H.; Liu, X.; Tavazoie, S.; Tavazoie, S. F. HNRNPA2B1 is a mediator of m6A-dependent nuclear RNA processing events. *Cell* **2015**, *162* (6), 1299–1308.
- (10) Salih, M. M.; Weindel, C. G.; Malekos, E.; Sudek, L.; Katzman, S.; Mabry, C. J.; Coleman, A. K.; Azam, S.; Watson, R.; Patrick, K. The RNA binding protein, HNRNPA2B1, regulates IFNG signaling in macrophages. *bioRxiv* **2023** *2023-10*, DOI: [10.1101/2023.10.12.562050](https://doi.org/10.1101/2023.10.12.562050).
- (11) Peng, W.-Z.; Zhao, J.; Liu, X.; Li, C.-F.; Si, S.; Ma, R. hnRNP A2B1 regulates the alternative splicing of BIRC5 to promote gastric cancer progression. *Cancer Cell Int.* **2021**, *21* (1), 1–14.
- (12) Kim, H. J.; Kim, N. C.; Wang, Y. D.; Scarborough, E. A.; Moore, J.; Diaz, Z.; MacLea, K. S.; Freibaum, B.; Li, S.; Molliex, A.; Kanagaraj, A. P.; Carter, R.; Boylan, K. B.; Wojtas, A. M.; Rademakers, R.; Pinkus, J. L.; Greenberg, S. A.; Trojanowski, J. Q.; Traynor, B. J.; Smith, B. N.; Topp, S.; Gkazi, A. S.; Miller, J.; Shaw, C. E.; Kottlors, M.; Kirschner, J.; Pestronk, A.; Li, Y. R.; Ford, A. F.; Gitler, A. D.; Benatar, M.; King, O. D.; Kimonis, V. E.; Ross, E. D.; Weihl, C. C.; Shorter, J.; Taylor, J. P. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **2013**, *495* (7442), 467–473.
- (13) Paul, K. R.; Molliex, A.; Cascarina, S.; Boncella, A. E.; Taylor, J. P.; Ross, E. D. Effects of mutations on the aggregation propensity of the human prion-like protein hnRNPA2B1. *Mol. Cell. Biol.* **2017**, *37* (8), e00652–00616. Qi, X.; Pang, Q.; Wang, J.; Zhao, Z.; Wang, O.; Xu, L.; Mao, J.; Jiang, Y.; Li, M.; Xing, X. Familial early-onset Paget's disease of bone associated with a novel hnRNPA2B1 mutation. *Calcif. Tissue Int.* **2017**, *101*, 159–169.
- (14) Fareed, M. M.; Dutta, K.; Dandekar, T.; Tarabonda, H.; Skorb, E. V.; Shityakov, S. In silico investigation of nonsynonymous single nucleotide polymorphisms in BCL2 apoptosis regulator gene to design novel protein-based drugs against cancer. *Journal of Cellular Biochemistry* **2022**, *123* (12), 2044–2056.
- (15) Wu, B.; Su, S.; Patil, D. P.; Liu, H.; Gan, J.; Jaffrey, S. R.; Ma, J. Molecular basis for the specific and multivariate recognitions of RNA substrates by human hnRNP A2/B1. *Nat. Commun.* **2018**, *9* (1), 420.
- (16) Shityakov, S.; Fischer, A.; Su, K.-P.; Hussein, A. A.; Dandekar, T.; Broscheit, J. Novel approach for characterizing propofol binding affinities to serum albumins from different species. *ACS omega* **2020**, *5* (40), 25543–25551.
- (17) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Amino acid scale: Normalized consensus hydrophobicity scale. *J. Mol. Biol.* **1984**, *179*, 125–142.
- (18) Glaser, F.; Pupko, T.; Paz, I.; Bell, R. E.; Bechor-Shental, D.; Martz, E.; Ben-Tal, N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **2003**, *19* (1), 163–164.
- (19) Wang, D.; Liu, D.; Yuchi, J.; He, F.; Jiang, Y.; Cai, S.; Li, J.; Xu, D. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Res.* **2020**, *48* (W1), W140–W146.
- (20) Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K. A.; Lin, G. N.; Nam, H. J.; Mort, M.; Cooper, D. N.; Sebat, J.; Iakoucheva, L. M.; Mooney, S. D.; Radivojac, P. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **2020**, *11* (1), 5918.
- (21) Warnecke, A.; Sandalova, T.; Achour, A.; Harris, R. A. PyTMs: a useful PyMOL plugin for modeling common post-translational modifications. *BMC Bioinf.* **2014**, *15* (1), 1–12.
- (22) Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research* **2018**, *46* (W1), W363–W367.
- (23) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* **2009**, *30* (16), 2785–2791.
- (24) Delgado, J.; Radusky, L. G.; Cianferoni, D.; Serrano, L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics* **2019**, *35* (20), 4168–4169.
- (25) Wang, Z.; Pan, H.; Sun, H.; Kang, Y.; Liu, H.; Cao, D.; Hou, T. fastDRH: a webserver to predict and analyze protein–ligand complexes based on molecular docking and MM/PB (GB) SA computation. *Briefings Bioinf.* **2022**, *23* (5), bbac201.
- (26) Tang, J.; Chen, Z.; Wang, Q.; Hao, W.; Gao, W.-Q.; Xu, H. hnRNPA2B1 promotes colon cancer progression via the MAPK pathway. *Frontiers in Genetics* **2021**, *12*, No. 666451. Chen, C.; Huang, L.; Sun, Q.; Yu, Z.; Wang, X.; Bu, L. HNRNPA2B1 Demonstrates Diagnostic and Prognostic Values Based on Pan-Cancer Analyses. *Comput. and Math. Methods Med.* **2022**, *2022*, No. 9867660.
- (27) Liu, N.; Pan, T. N 6-methyladenosine–encoded epitranscriptomics. *Nature structural & molecular biology* **2016**, *23* (2), 98–102.
- (28) Elsakrmy, N.; Cui, H. R-Loops and R-Loop-Binding Proteins in Cancer Progression and Drug Resistance. *International Journal of Molecular Sciences* **2023**, *24* (8), 7064.
- (29) Humphries, F.; Fitzgerald, K. A. hnRNPA2B1: fueling antiviral immunity from the nucleus. *Mol. Cell* **2019**, *76* (1), 8–10.

- (30) Wang, L.; Wen, M.; Cao, X. Nuclear hnRNPA2B1 initiates and amplifies the innate immune response to DNA viruses. *Science* **2019**, *365* (6454), No. eaav0758.
- (31) Zhang, X.; Flavell, R. A.; Li, H.-B. hnRNPA2B1: a nuclear DNA sensor in antiviral immunity. *Cell Research* **2019**, *29* (11), 879–880.
- (32) Serçinoğlu, O.; Ozbek, P. Sequence-structure-function relationships in class I MHC: A local frustration perspective. *PloS one* **2020**, *15* (5), No. e0232849.
- (33) Lim, S. W.; Tan, K. J.; Azuraidi, O. M.; Sathiya, M.; Lim, E. C.; Lai, K. S.; Yap, W.-S.; Afizan, N. A. R. N. M. Functional and structural analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in the MYB oncoproteins associated with human cancer. *Sci. Rep.* **2021**, *11* (1), 24206.
- (34) Maguire, J. B.; Haddox, H. K.; Strickland, D.; Halabiya, S. F.; Coventry, B.; Griffin, J. R.; Pulavarti, S. V. S. R. K.; Cummins, M.; Thieker, D. F.; Klavins, E.; Szyperski, T.; DiMaio, F.; Baker, D.; Kuhlman, B. Perturbing the energy landscape for improved packing during computational protein design. *Proteins: Struct., Funct., Bioinf.* **2021**, *89* (4), 436–449.
- (35) Holcomb, D.; Alexaki, A.; Hernandez, N.; Hunt, R.; Laurie, K.; Kames, J.; Hamasaki-Katagiri, N.; Komar, A. A.; DiCuccio, M.; Kimchi-Sarfaty, C. Gene variants of coagulation related proteins that interact with SARS-CoV-2. *PLoS computational biology* **2021**, *17* (3), No. e1008805.
- (36) Shityakov, S.; Skorb, E. V.; Nosonovsky, M. Topological bio-scaling analysis as a universal measure of protein folding. *Royal Society Open Science* **2022**, *9* (7), No. 220160.
- (37) Shityakov, S.; Skorb, E. V.; Nosonovsky, M. Folding–unfolding asymmetry and a RetroFold computational algorithm. *Royal Society Open Science* **2023**, *10* (5), No. 221594.
- (38) Kim, E.; Ahuja, A.; Kim, M.-Y.; Cho, J. Y. DNA or protein methylation-dependent regulation of activator protein-1 function. *Cells* **2021**, *10* (2), 461.
- (39) Zhao, J.; Zhang, S.; Jiang, Y.; Liu, Y.; Zhu, Q. Mutation analysis of pathogenic non-synonymous single nucleotide polymorphisms (nsSNPs) in WFS1 gene through computational approaches. *Sci. Rep.* **2023**, *13* (1), 6774.
- (40) Dakal, T. C.; Kala, D.; Dhiman, G.; Yadav, V.; Krokhotin, A.; Dokholyan, N. V. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene. *Sci. Rep.* **2017**, *7* (1), 6525.
- (41) Agozzino, L.; Dill, K. A. Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (37), 9092–9097.
- (42) Fardmanesh, H.; Shekari, M.; Movafagh, A.; Alizadeh Shargh, S.; Poursadegh Zonouzi, A. A.; Shakerizadeh, S.; Poursadegh Zonouzi, A.; Hosseinzadeh, A. Upregulation of the double-stranded RNA binding protein DGCR8 in invasive ductal breast carcinoma. *Gene* **2016**, *581* (2), 146–151.
- (43) Huang, Y.; Li, Y.; He, F.; Wang, S.; Li, Y.; Ji, G.; Liu, X.; Zhao, Q.; Li, J. Metastasis-associated protein 3 in colorectal cancer determines tumor recurrence and prognosis. *Oncotarget* **2017**, *8* (23), 37164.
- (44) Maji, D.; Glasser, E.; Henderson, S.; Galardi, J.; Pulvino, M. J.; Jenkins, J. L.; Kielkopf, C. L. Representative cancer-associated U2AF2 mutations alter RNA interactions and splicing. *J. Biol. Chem.* **2020**, *295* (50), 17148–17157.
- (45) Huang, T.; Li, Y. Current progress, challenges, and future perspectives of language models for protein representation and protein design. *Innovation* **2023**, *4* (4), No. 100446.