

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/376618747>

# Optimizing Targeted Inhibition of the hnRNPA2/B1 Splicing Factor in Cancer Cells through Analysis of Single Nucleotide Polymorphism Effects

Article in ACS Omega · December 2023

DOI: 10.1021/acsomega.3c07195

CITATIONS

0

READS

75

6 authors, including:



Viacheslav Kravtsov

ITMO University

77 PUBLICATIONS 124 CITATIONS

[SEE PROFILE](#)



Sergey Shityakov

University of Wuerzburg

165 PUBLICATIONS 1,841 CITATIONS

[SEE PROFILE](#)

# Exploring the Impact of Single Nucleotide Polymorphisms on the Stability and Function of the hnRNPA2/B1 Protein: Implications for the Design of Anticancer Therapeutics

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



Cite This: [https://doi.org/10.1021/acosomega.3c07195](https://doi.org/10.1021/acsomega.3c07195)



Read Online

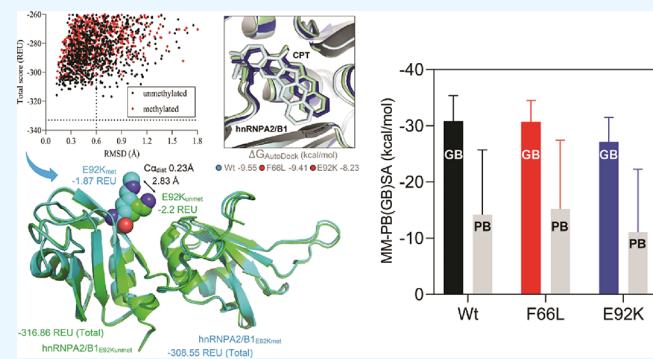
ACCESS |

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 HNRNPA2B1, 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 HNRNPA2B1'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

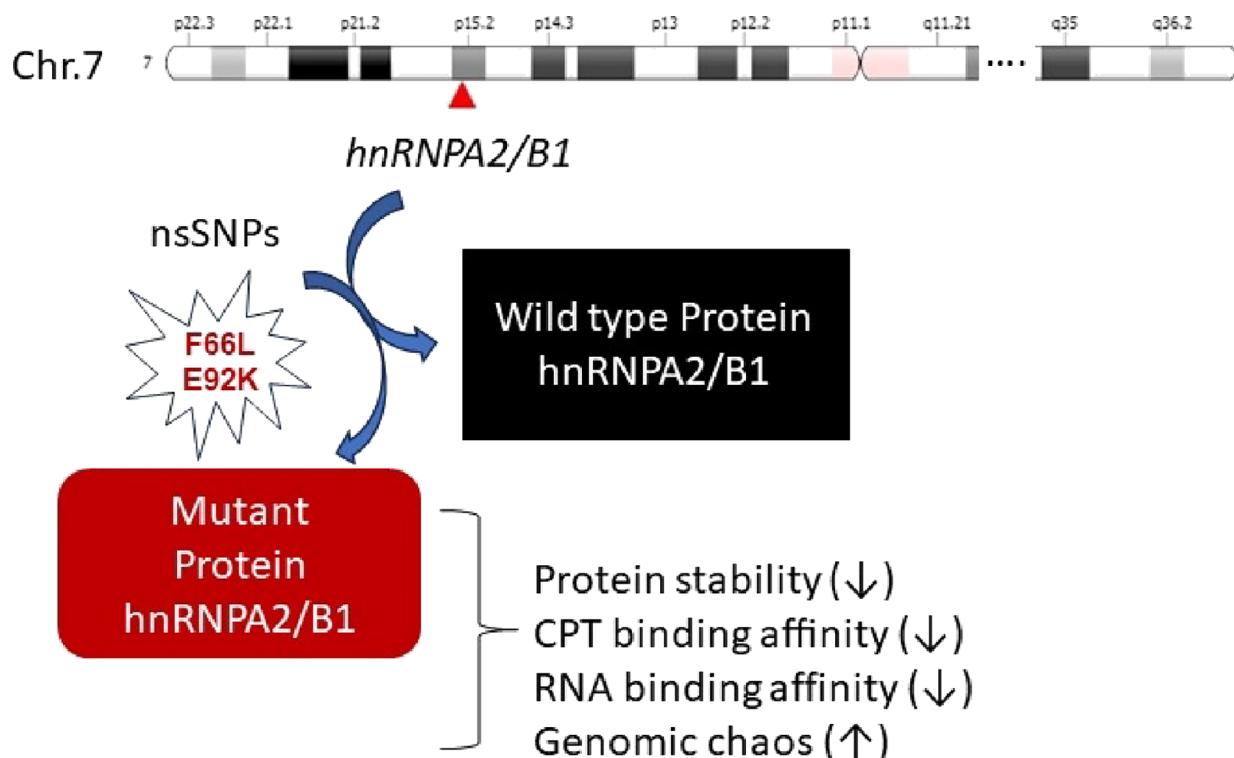
Heterogeneous nuclear ribonucleoproteins (hnRNPs) are RNA-binding proteins<sup>1</sup>; they interact with heterogeneous nuclear RNA (hnRNA).<sup>2</sup> In general, hnRNPs reside in the nucleus, but some of them move back and forth between the nucleus and the cytoplasm.<sup>3</sup> The hnRNPs have unique 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 subfamily A2/B1 of the hnRNP, i.e., hnRNPA2/B1, plays an important role in controlling mRNA transcripts for proper cellular activities.<sup>5</sup> Proteins A2 and B1 are structurally quite similar, apart from 12 amino acids in the B1's N-terminal domain.<sup>6</sup> Conversely, the RRMs of A1 and B1 share fewer than 30% amino acid sequence similarity in their respective glycine-rich domain (GRD). However, A1 and B1 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 IFNG (IFN- $\gamma$ ) signaling in macrophages through alternative splicing of the IFNG 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

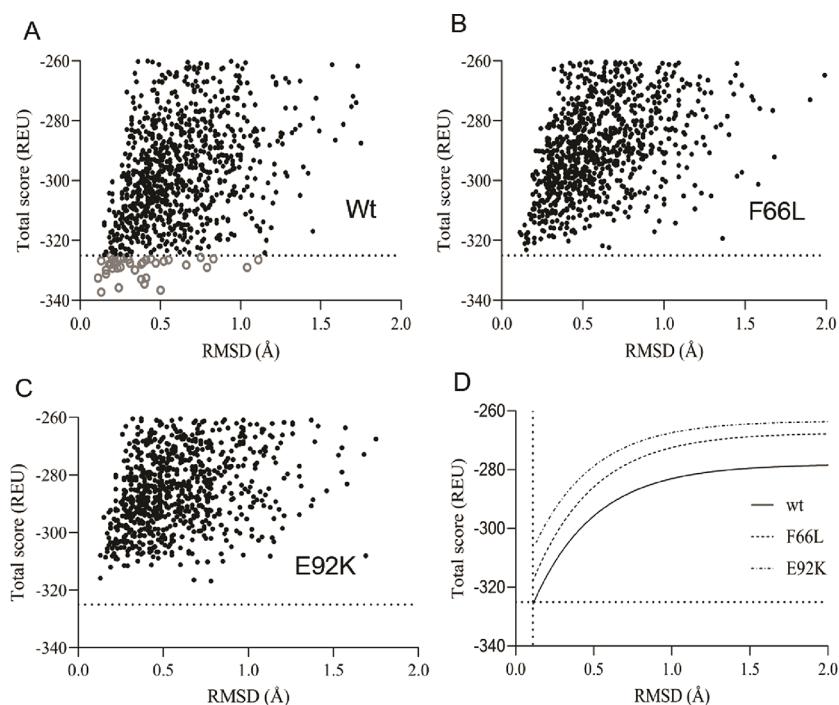
Received: September 19, 2023

Revised: November 29, 2023

Accepted: December 8, 2023



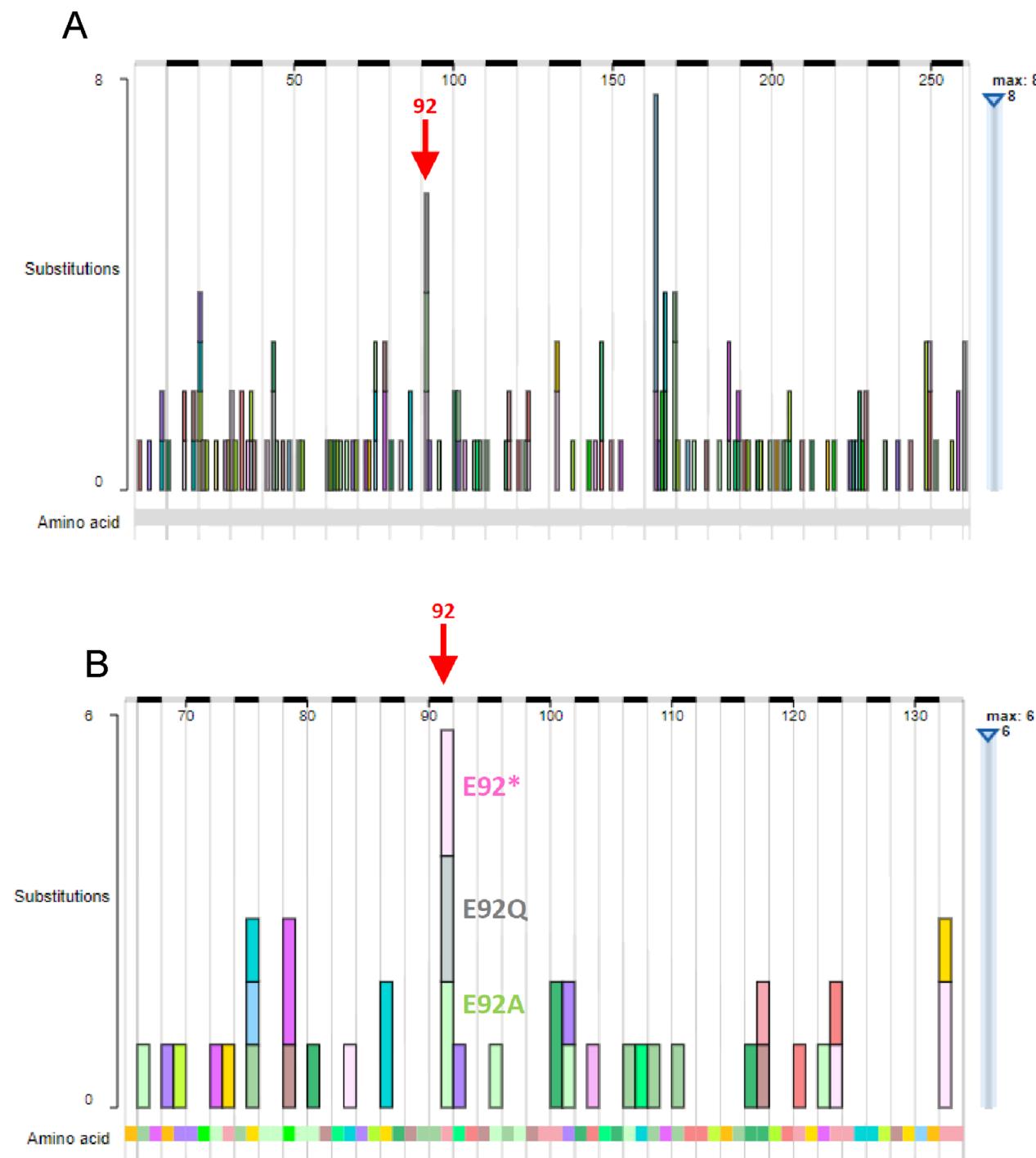
**Figure 1.** Biophysical impacts of nsSNPs on hnRNPA2B1.



**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.

affect the splicing of BIRC5 and 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, RNA, and camptothecin (CPT, BDBMS0008923) binding affinities using multiple innovative bioinformatics tools. The results showed that nsSNPs, i.e., F66L and E92K, are pathogenic, with E92K



**Figure 3.** Two histograms displaying the positions of mutations (A) and single base substitutions (B) across the HNRNPA2B1 (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.

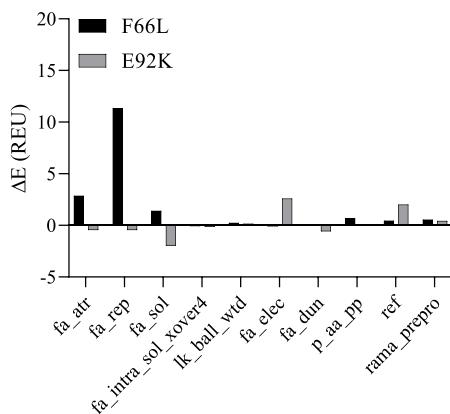
73 being modified by methylation, and have lower RNA and CPT  
74 binding affinities than that of the wild-type hnRNPA2B1.

## 2. COMPUTATIONAL METHODS

75 **2.1. Data Collection, Prediction, and Search Analyses**  
76 of SNPs. Prediction and search analyses of SNPs of the human  
77 hnRNPA2/B1 were conducted, as described by our previous  
78 work.<sup>14</sup> In brief, the NCBI-SNP database and UniProt database  
79 were used for further analyses of nsSNPs.

80 **2.2. Identification of Deleterious nsSNPs.** Deleterious  
81 nonsynonymous mutations were identified using the methods  
82 described in our previous work.<sup>14</sup> In brief, the SIFT tool and  
83 PolyPhenV2 algorithms were used to determine any deleterious  
84 nsSNPs in hnRNPA2/B1. SNP variants were predicted by using  
85 the SNP predictor tool.

86 **2.3. nsSNP Search Analyses in Various Databases.** Data  
87 regarding the occurrence of these nsSNPs in real-world cases  
88 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\_preprom, 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 | $\Delta E_{\text{mol}}$ | $\Delta E_{\text{ros}}$ |
|---------|-------|---------|-----------|--------|-------------------------|-------------------------|
| Wt      | A     | 66      | Phe       | Leu    | 1.33                    | 14.18                   |
| Wt      | A     | 92      | Glu       | Lys    | 0.18                    | 21.4                    |

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

93 **2.4. Protein Stability Prediction.** The Molsoft ( $\Delta E_{\text{mol}}$ )  
94 and Rosetta ( $\Delta E_{\text{ros}}$ ) protocols were used to predict the protein  
95 stability via measuring the energy differences between mutated  
96 ( $E_{\text{mut}}$ ) and wild-type forms ( $E_{\text{wt}}$ ) and were calculated using the  
97 following equation:

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

98 **2.5. Protein Stability and Interface Analysis.** The  
99 protein structure of hnRNPA2/B1 in complex with 10-mer  
100 RNA (SHO4)<sup>15</sup> was obtained from RCSB-Protein Data Bank.  
101 Protein stability was analyzed using Molsoft ICM-Pro (Molsoft  
102 LLC, San Diego, CA, USA).

103 **2.6. Conservation Analyses.** Conservation analysis was  
104 performed by the protocol described by our previous  
105 studies.<sup>14,16</sup> The normalized consensus hydrophobicity scale<sup>17</sup>  
106 was used for the prediction of the conservation profile of the  
107 hnRNPA2/B1. Additionally, the ConSurf web server was used for  
108 such analyses.<sup>18</sup>

109 **2.7. Prediction of Post-Translational Modifications.**  
110 The post-translational modifications in hnRNPA2/B1 (SHO4),  
111 F66L, and E92K mutants were predicted by using MusiteDeep  
112 available at <https://www.musite.net>.<sup>19</sup> The MutPred2 web  
113 server (<http://mutpred.mutdb.org/index.html>)<sup>20</sup> was used to  
114 predict the impact of F66L and E92K substitutions on  
115 hnRNPA2/B1, as described by our previous work.<sup>14</sup> Addition-

ally, the PyTM plugin of PyMOL was used for E92K<sup>21</sup> methylation.<sup>116</sup>

117 **2.8. Gene Association Network Analysis.** The gene–<sup>118</sup>  
gene interaction network was studied using the GeneMANIA<sup>119</sup>  
tool, as described by our previous work.<sup>14</sup>

120 **2.9. Interference Analysis and Receptor–Drug Affinity**<sup>121</sup>  
**Predictions.** The analyses of molecular interactions were<sup>122</sup>  
performed using in-house PyMOL scripts as described by our<sup>123</sup>  
previous study with the normalized consensus hydrophobicity<sup>124</sup>  
scale.<sup>14</sup> The CASTp server<sup>22</sup> was used for binding site<sup>125</sup>  
identification (job id: j\_63ca8c75e54d8). AutoDock<sup>23</sup> molec-<sup>126</sup>  
ular docking was used for the molecular interaction between<sup>127</sup>  
CPT and hnRNPA2/B1. The center grid dimensions were set to<sup>128</sup>  
 $-30.110 \times 5.710 \times 7.930$  with a grid spacing of  $0.375 \text{ \AA}$ .<sup>129</sup>  
Additionally, FoldX was used for molecular docking between<sup>130</sup>  
RNA and human hnRNPA2/B1.<sup>24</sup> And structure-truncated<sup>131</sup>  
MM/PB(GB)SA (molecular mechanics/Poisson–Boltzmann/<sup>132</sup>  
generalized Born surface area) rescoring and hotspot predictions<sup>133</sup>  
were performed using the fastDRH server.<sup>25</sup>

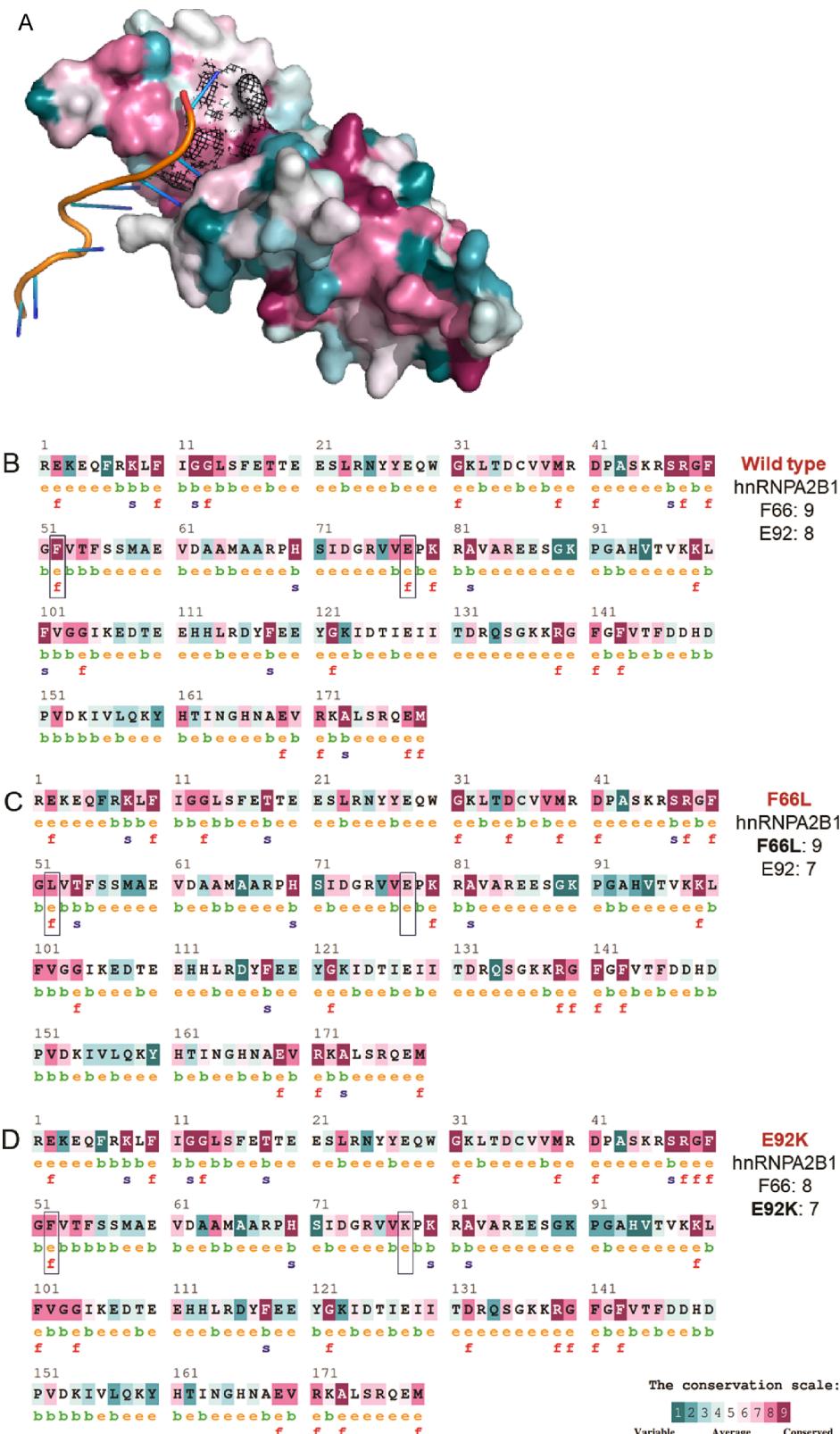
### 3. RESULTS

135 **3.1. Identification of SNPs.** We identified a total of 8494<sup>135</sup>  
SNPs in human hnRNPA2/B1. While 1303 were found in the<sup>136</sup>  
3'-UTR region, 858 of them were found in the 5'-UTR region.<sup>137</sup>  
In addition, further examination of SNPs using the SIFT tool<sup>138</sup>  
revealed 157 nonsynonymous and 204 synonymous SNPs and 6<sup>139</sup>  
PhD-SNPs (supplementary file, Table S1). Two important<sup>140</sup>  
nsSNPs, F66L and E92K (Figure 1), among PhD-SNPs were<sup>141</sup>  
selected for a detailed study based on their probability (>70%)<sup>142</sup>  
(Table S2).

143 **3.2. Analysis of nsSNP in Various Databases.** Analyzing<sup>144</sup>  
a comprehensive data set of 44,239 unique tumor samples from<sup>145</sup>  
COSMIC, we observed that 469 unique samples contained<sup>146</sup>  
mutations. Intriguingly, amino acid position 92 exhibited the<sup>147</sup>  
second-highest mutation frequency within HNRNPA2B1<sup>148</sup>  
(Figure 3). Specifically, we identified the following substitu-<sup>149</sup>  
tions: E92A (275A > C, observed in two samples), E92Q (274G<sup>150</sup>  
> C, observed in two samples), and E92\* (274G > T, observed<sup>151</sup>  
in two samples), as visually represented in Figure 3B. These<sup>152</sup>  
substitutions appear to be notably associated with breast and<sup>153</sup>  
lung cancers. Overall, the experimental data presented here align<sup>154</sup>  
with our theoretical studies, where the nsSNP at position 92<sup>155</sup>  
exerts the most substantial influence on the stability and<sup>156</sup>  
functionality of HNRNPA2B1 (Figure 4). Overall, the<sup>157</sup>  
experimental data presented here aligns with our theoretical<sup>158</sup>  
studies, where the nsSNP at position 92 exerts the most<sup>159</sup>  
substantial influence on the stability and functionality of<sup>160</sup>  
HNRNPA2B1.

161 **3.3. Protein Stability Prediction.** The stabilities of<sup>162</sup>  
hnRNPA2/B1<sub>F66L</sub> and hnRNPA2/B1<sub>E92K</sub> mutants were sig-<sup>163</sup>  
nificantly diminished according to the results obtained from the<sup>164</sup>  
Rosetta standard energy function (Figure 2) and MolSoft<sup>165</sup>  
stability prediction (Table 1). The nonsynonymous SNPs F66L<sup>166</sup>  
and E92K of hnRNPA2/B1 showed  $\Delta E_{\text{mol}}$  of 1.33 and 0.1804<sup>167</sup>  
kcal/mol and  $\Delta E_{\text{ros}}$  of 14.18 and 21.4 REU, respectively.<sup>168</sup>  
Moreover, it is important to note that the F66L substitution<sup>169</sup>  
mutant showed a higher Lennard-Jones attractive as well as<sup>170</sup>  
repulsive potential between atoms in different residues, but<sup>171</sup>  
these parameters were negative for E92K (Figure 4).

172 Although, the Lazaridis-Karplus solvation energy was positive<sup>173</sup>  
for F66L substitution but not for E92K. Nonetheless, electro-<sup>174</sup>  
static potential with a distance-dependent dielectric was higher<sup>175</sup>  
for E92K substitution. In addition, the internal energy of side<sup>176</sup>  
177



**Figure 5.** Evolutionary conservation profiles of hnRNP A2B1. (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.

177 chain rotamers as derived from Dunbrack's statistics showed a  
178 negative value for E92K substitution. The probability of amino  
179 acid at  $\Phi/\Psi$  angles was positive for F66L. However, the  
180 backbone torsion preference term that takes into account

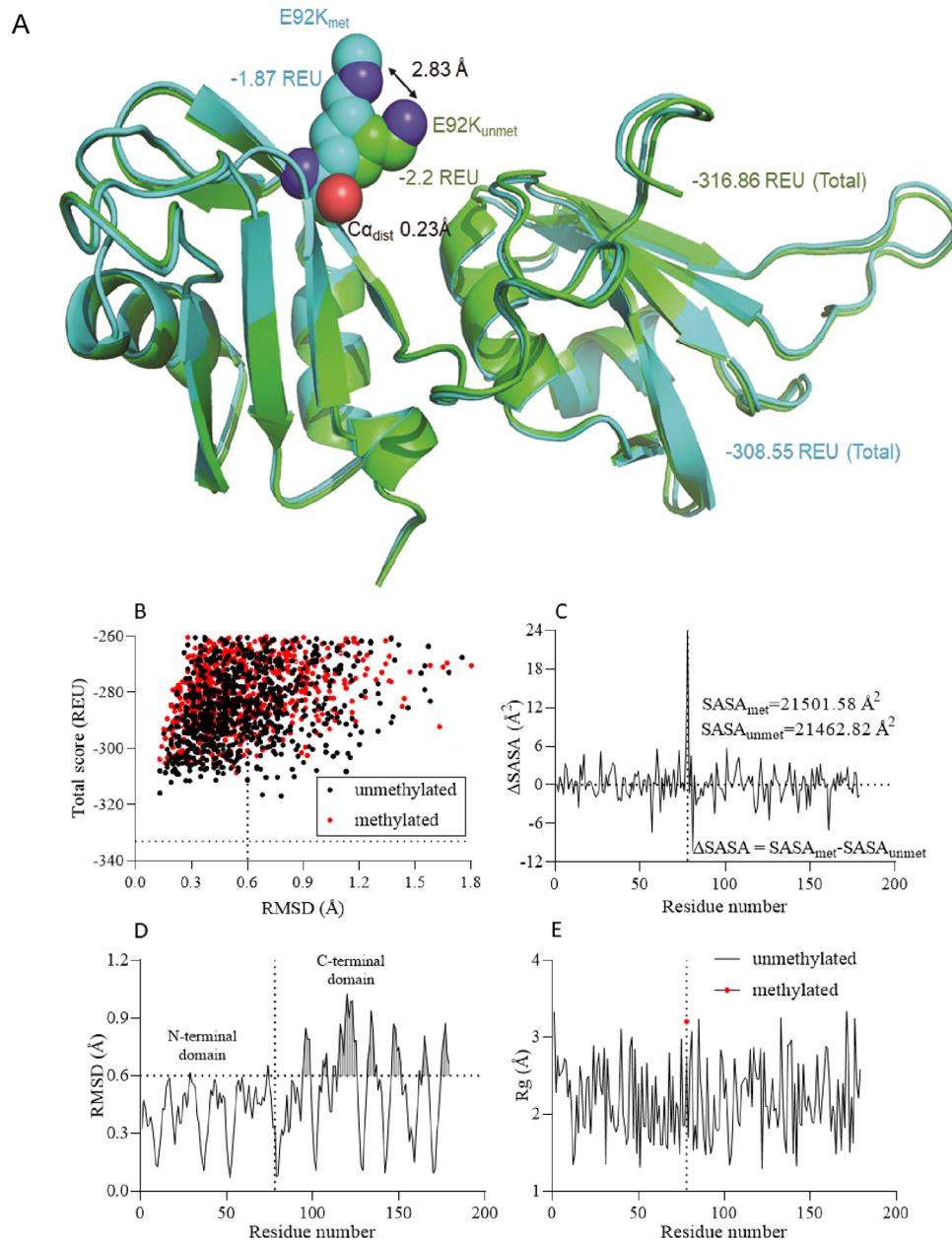
whether the preceding amino acid is proline was similar in the 181 case of both F66L and E92K.  
182

**3.4. Conservation Analysis.** The results derived from the 183 analysis of evolutionary conservation profiles highlight the 184

**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        |
|----------|------------|----------|----------------|------------|-------------|
| F66 (wt) |            |          |                |            |             |
| E92 (wt) |            |          |                |            |             |
| F66L     |            | NS       | 0.877          | pathogenic |             |
| E92K     | 0.677      | NS       | 0.779          | pathogenic | methylation |

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



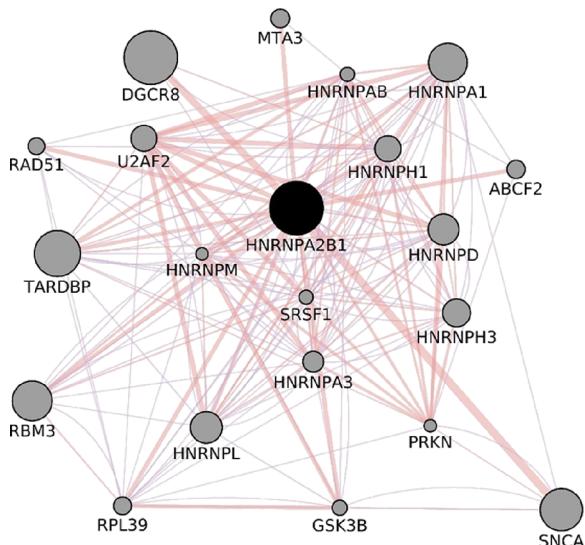
**Figure 6.** Structural stability analyses before and after post-translational modification E92K of hnRNPA2/B1. (A) Cartoon model of hnRNPA2/B1, E92K methylated (cyan), and unmethylated (green).  $\text{Ca}_{\text{dist}}$ , a difference of distance between the  $\alpha$  carbon of E92K<sub>met</sub> 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)  $R_g$  versus amino acid residue number of methylated, unmethylated E92K of hnRNPA2/B1. REU, Rosetta energy units; RMSD, root-mean-square deviation;  $R_g$ , 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.

185 conservation and exposure of amino acids F66 and E92 in wild-  
186 type hnRNPA2/B1, as indicated by ConSurf prediction scores  
187 of eight and nine (Figure 5). However, when considering

188 substitutions like F66L and E92K, a subtle alteration in the 188  
189 conservation profile is observed. For instance, in the case of the 189  
F66L substitution, the ConSurf prediction score at E92 190

decreases to seven. Nevertheless, according to the ConSurf prediction, the F66L substitution still classifies as a functionally significant, highly conserved, and exposed residue. Conversely, for the E92K substitution, the ConSurf prediction score also drops to seven, and it is predicted to be an exposed residue, in accordance with the NACCESS algorithm.

**3.5. Post-Translational Modification Protein Stability and Network Analyses.** Post-translational modifications were not predicted at the F66 and E92 of the wild-type hnRNPA2/B1 according to the MusiteDeep server. However, the substitution E92K was predicted to be modified by methylation with a good PTM score of 0.677. Conversely, post-translational modification was not predicted at the F66L substitution. However, both substitutions, i.e., F66L and E92K, were predicted to be pathogenic as these substitutions showed MutPred2 score higher than 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).



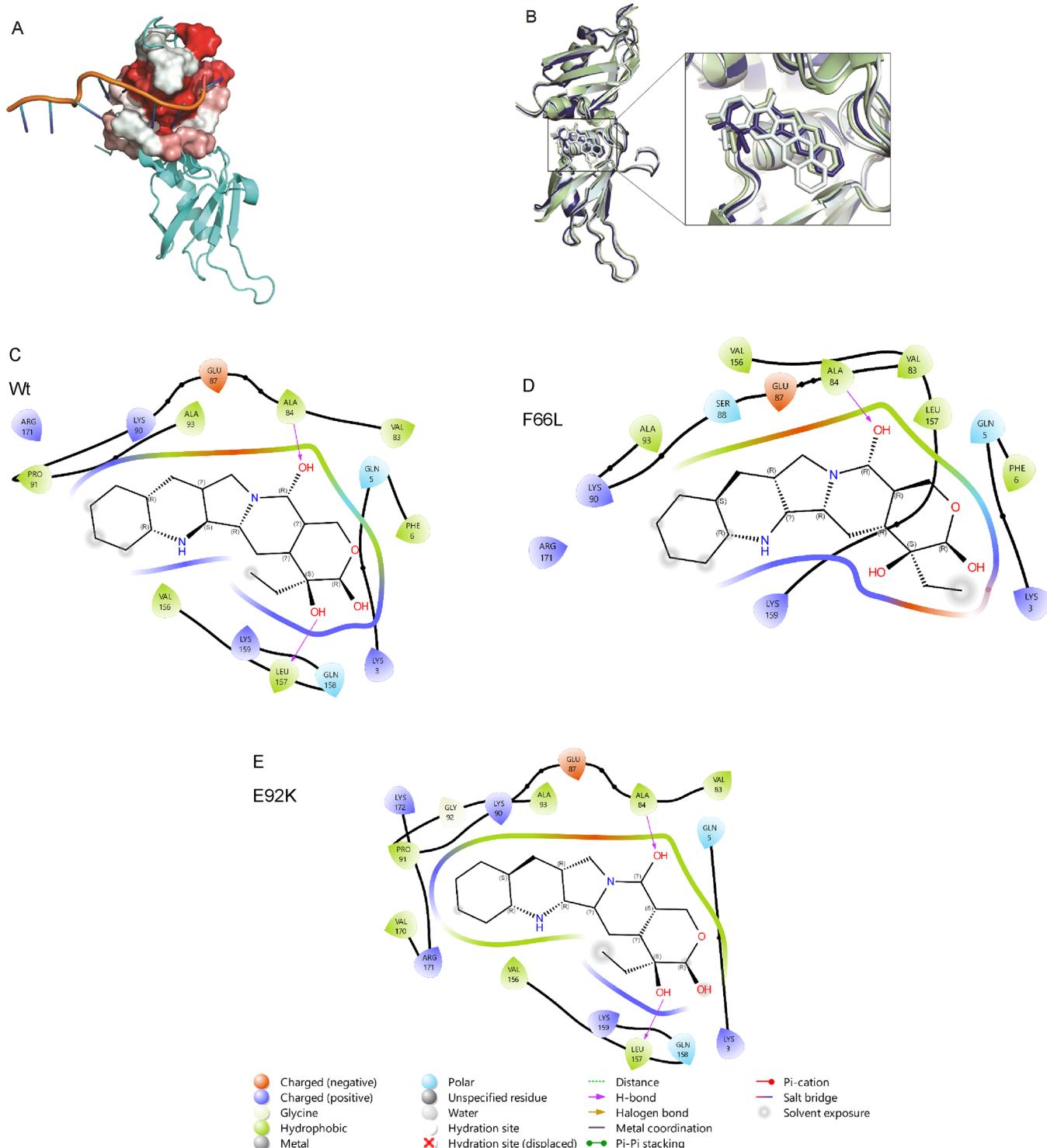
**Figure 7.** Molecular interaction network of human hnRNPA2/B1. The black node at the center, hnRNPA2/B1, physical interactions (bold line) and coexpression (line), and degree of interactions are depicted proportionally by the size of the nodes.

**3.6. Binding Site Analyses.** The RNA binding site is located in the proximity of the CPT binding site. The results showed CPT binding postures of the wild-type hnRNPA2/B1, and the substitution mutants E92K and F66L were nonidentical (Figure 8). However, the nsSNPs F66L and E92K were located far away from the CPT binding site of hnRNPA2/B1. And the 2D protein–ligand interaction plot showed that A84 was a common amino acid participating in hydrogen bonding with the exposed hydroxyl group of the CPT bound to hnRNPA2/B1 wild-type, F66L, and E92K substitution mutants. Additionally, Leucine 157 of the hnRNPA2/B1 wild-type and E92K were participated in hydrogen bonding with the exposed hydroxyl group of the 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 ΔG<sub>AutoDock</sub> of −9.55 kcal/mol. On the other hand, the ΔG<sub>AutoDock</sub> values of the substitution mutants F66L and E92K were −9.41 and −8.23 kcal/mol, respectively. Furthermore, hnRNPA2/B1<sub>wt</sub> showed better structure-truncated MM/PB(GB)SA rescoring of CPT’s pose than that of hnRNPA2/B1<sub>E92K</sub>. Conversely, these values were higher for the F66L substitution.

## 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 hydrides, 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). Furthermore, our research uncovered additional relevance, as we utilized the Mastermind platform, a comprehensive search engine that consolidates information from medical literature regarding diseases, genes, and variants. 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 differences in structural stabilities between X-ray and AlphaFold models may be due to the intrinsically disordered domain as protein folding is close to a threshold of 3 Å.<sup>36</sup> Nonetheless, close to the ideal 0.5 Å RMSD<sup>37</sup> was found by protein folded by the AlphaFold2 algorithm. Furthermore, the E92K substitution



**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.

290 was predicted to be modified by methylation with a good PTM  
291 prediction score of 0.68 (Table 2). Post-translational  
292 modification of protein plays a vital role in cell signaling.<sup>38</sup> For  
293 instance, phosphorylation of a protein leads to the activation of  
294 protein kinases and vice versa. Similarly, methylation of protein  
295 is linked with cell signaling, cell differentiation, and cell  
296 proliferation.<sup>38</sup> Notably, the stability of E92K<sub>met</sub> ( $-1.87$  REU)  
297 was lower than that of E92K<sub>unmet</sub> ( $-2.2$  REU). The distance

between  $\text{C}\alpha$  of E92K<sub>met</sub> and E92K<sub>unmet</sub> was  $0.23 \text{ \AA}$ , confirming a  
298 structural alteration (Figure 6a). Further, the total REU of the  
299 hnRNP A2/B1<sub>E92Kunmet</sub> was also lower than that of hnRNP A2/  
300 B1<sub>E92Kmet</sub> suggesting methylation-induced structural instability.  
301 Furthermore, structural alteration of E92K<sub>met</sub> and E92K<sub>unmet</sub> is  
302 supported by the dihedral angle analysis of the substitution  
303 mutants E92K<sub>met</sub> and E92K<sub>unmet</sub> (Table 3). Similarly, reductions  
304 of stabilities due to nsSNPs in BCL-2 were reported by Fareed et  
305

**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    |

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 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 hnRNPA2/B1<sub>wt</sub> and F66L of hnRNPA2/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 hnRNPA2/B1 and hnRNPA1 proteins.<sup>12</sup> In addition, the results depicted by physical interactions and coexpression of nodes indicate essential functions of the human hnRNPA2/B1 protein (Figure 7). For instance, DiGeorge syndrome critical region 8 (DGR8),<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 hnRNPA2/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 hnRNPA2/B1 showed molecular docking energies for CPT and RNA better than those of the substitution mutants F66L and E92K (Table 4). Additionally, hnRNPA2/B1<sub>wt</sub>

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

| proteins | total energy (kcal/mol)   |                     |               |                |          |
|----------|---------------------------|---------------------|---------------|----------------|----------|
|          | FoldX protein–RNA docking | protein–CPT docking | GB            | PB             | fastDRH* |
| 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.

performed better than hnRNPA2/B1<sub>E92K</sub> in terms of structure-truncated MM/PB(GB)SA rescoring of ligand's poses. Furthermore, predicted hotspot residues of wild-type hnRNPA2/B1 showed better per amino acid residue MM-GBSA energy (kcal/mol) than those of hnRNPA2/B1<sub>F66L</sub> and hnRNPA2/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 hnRNPA2/B1 as the sequence–structure–function paradigm is fundamental in molecular biology.<sup>45</sup>

## 5. CONCLUSIONS

Human hnRNPA2/B1 plays pivotal roles in various facets of biology, including carcinogenesis, viral immunity, and RNA metabolism. RNA, which governs a multitude of physiological activities in normal cells, relies on hnRNPA2/B1 as a molecular reader of N6-methyladenosine (m6A) to facilitate mRNA processing. This interaction, especially with DGCR8, is essential for maintaining the genomic stability. Intriguingly, in the context of cancer research, understanding RNA levels, dynamics, and their fate is an ever-evolving and critical component.

Of particular note, nsSNPs within hnRNPA2/B1 hold the potential to disrupt the intricate RNA life cycle associated with this protein, leading to genomic instability. This genomic chaos is closely linked to the formation of R-loops in DNA–RNA hybrids and double-stranded DNA breaks, significantly complicating translational therapeutic research.

In our study, we have identified two noteworthy nsSNPs, namely, F66L and E92K. Both of these nsSNPs are pathogenic, with F66L being predicted as a functional, highly conserved, and exposed residue. Furthermore, E92K is anticipated to be modified by methylation, resulting in a reduction in structural stability. Notably, both the F66L and E92K substitutions exhibit lower binding affinities for CPT and RNA compared to wild-type hnRNPA2/B1.

Significantly, the overexpression of hnRNPA2/B1, abnormal m6A disposition, and dysregulation of RNA levels have been observed in the context of cancer, underscoring their potential contributions to oncogenesis. Consequently, the consideration of the nsSNPs F66L and E92K within hnRNPA2/B1 could prove instrumental in the design of epigenetic drugs (Epi-drugs) aimed at restoring proper RNA metabolism in the context of cancer.

## 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 hnRNPA2/B1. (Table S2) Prediction of single nucleotide polymorphisms in hnRNPA2/B1. (Figure S1) Structural stability of the hnRNPA2/B1 protein. (A) X-ray model of hnRNPA2B1; (B) AlphaFold model; thresholds are depicted by dotted lines. (Figure S2) Predicted hotspot residues. (A) hnRNPA2/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.org/0000-0002-6953-9771; Phone: +7

9006371702; Email: shityakoff@itmo.ru

## 397 Authors

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

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

403 Katerina Oleynikova — Institute of Bioengineering, Research  
404 Center of Biotechnology of the Russian Academy of Sciences,  
405 Moscow 119071, Russian Federation

406 Alexey Ruzov — Institute of Bioengineering, Research Center of  
407 Biotechnology of the Russian Academy of Sciences, Moscow  
408 119071, Russian Federation

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

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

## 414 Author Contributions

415 K.D.: software, writing—original draft; S.S.: data curation,  
416 investigation, resources, software, and validation; K.O. and A.R.:  
417 investigation, methodology, validation, and writing—review and  
418 editing; V.K., E.V.S., and S.S.: conceptualization, investigation,  
419 methodology, project administration, supervision, validation,  
420 and writing—review and editing.

## 421 Notes

422 The authors declare no competing financial interest.

## 423 ■ ACKNOWLEDGMENTS

424 This work was supported by grant from the Russian Science  
425 Foundation (RSF), Russia, no. 263 22-65-00022. Graphical  
426 abstract was prepared using an image by brgfx on Freepik.

## 427 ■ REFERENCES

- 428 (1) Wang, J.; Sun, D.; Wang, M.; Cheng, A.; Zhu, Y.; Mao, S.; Ou, X.;  
429 Zhao, X.; Huang, J.; Gao, Q.; Zhang, S.; Yang, Q.; Wu, Y.; Zhu, D.; Jia,  
430 R.; Chen, S.; Liu, M. Multiple functions of heterogeneous nuclear  
431 ribonucleoproteins in the positive single-stranded RNA virus life cycle.  
432 *Front. Immunol.* **2022**, 13, No. 989298.
- 433 (2) Görlich, M.; Burd, C. G.; Portman, D. S.; Dreyfuss, G. The  
434 hnRNP proteins. *Molecular biology reports* **1993**, 18, 73–78.
- 435 (3) Kim, J. H.; Hahm, B.; Kim, Y. K.; Choi, M.; Jang, S. K. Protein-  
436 protein interaction among hnRNPs shuttling between nucleus and  
437 cytoplasm. *Journal of molecular biology* **2000**, 298 (3), 395–405.
- 438 (4) Caputi, M.; Zahler, A. M. Determination of the RNA binding  
439 specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP)  
440 H/H'/F/2H9 family. *J. Biol. Chem.* **2001**, 276 (47), 43850–43859.
- 441 (5) He, Y.; Smith, R. Nuclear functions of heterogeneous nuclear  
442 ribonucleoproteins A/B. *Cellular and molecular life sciences* **2009**, 66,  
443 1239–1256.
- 444 (6) Liu, Y.; Shi, S. L. The roles of hnRNP A2/B1 in RNA biology and  
445 disease. *Wiley Interdisc. Rev.: RNA* **2021**, 12 (2), No. e1612.
- 446 (7) Zhou, J.; Allred, D.; Avis, I.; Martínez, A.; Vos, M. D.; Smith, L.;  
447 Treston, A. M.; Mulshine, J. L. Differential expression of the early lung  
448 cancer detection marker, heterogeneous nuclear ribonucleoprotein-  
449 A2/B1 (hnRNP-A2/B1) in normal breast and neoplastic breast cancer.  
450 *Breast cancer research and treatment* **2001**, 66, 217–224. Cui, H.; Wu, F.;  
451 Sun, Y.; Fan, G.; Wang, Q. Up-regulation and subcellular localization of  
452 hnRNP A2/B1 in the development of hepatocellular carcinoma. *BMC*  
453 *Cancer* **2010**, 10, 1–14. Yan-Sanders, Y.; Hammons, G. J.; Lyn-Cook, B.  
454 D. Increased expression of heterogeneous nuclear ribonucleoprotein  
455 A2/B1 (hnRNP) in pancreatic tissue from smokers and pancreatic  
456 tumor cells. *Cancer letters* **2002**, 183 (2), 215–220.

- 457 (8) Xu, H.; Li, P.; Wang, X.; Zhuang, H.; Hua, Z.-C. Emerging roles of 457  
hnRNP A2B1 in cancer and inflammation. *Int. J. Biol. Macromol.* **2022**, 458  
221, 1077. 459
- 459 (9) Alarcón, C. R.; Goodarzi, H.; Lee, H.; Liu, X.; Tavazoie, S.; 460  
Tavazoie, S. F. HNRNPA2B1 is a mediator of m6A-dependent nuclear 461  
RNA processing events. *Cell* **2015**, 162 (6), 1299–1308. 462
- 462 (10) Salih, M. M.; Weindel, C. G.; Malekos, E.; Sudek, L.; Katzman, S.; 463  
Mabry, C. J.; Coleman, A. K.; Azam, S.; Watson, R.; Patrick, K. The 464  
RNA binding protein, HNRNPA2B1, regulates IFNG signaling in 465  
macrophages. *bioRxiv* **2023** 2023-10, DOI: 10.1101/ 466  
2023.10.12.562050. 467
- 467 (11) Peng, W.-Z.; Zhao, J.; Liu, X.; Li, C.-F.; Si, S.; Ma, R. 468  
hnRNPA2B1 regulates the alternative splicing of BIRC5 to promote 469  
gastric cancer progression. *Cancer Cell Int.* **2021**, 21 (1), 1–14. 470
- 470 (12) Kim, H. J.; Kim, N. C.; Wang, Y. D.; Scarborough, E. A.; Moore, 471  
J.; Diaz, Z.; MacLea, K. S.; Freibaum, B.; Li, S.; Molliex, A.; Kanagaraj, 472  
A. P.; Carter, R.; Boylan, K. B.; Wojtas, A. M.; Rademakers, R.; Pinkus, J. 473  
L.; Greenberg, S. A.; Trojanowski, J. Q.; Traynor, B. J.; Smith, B. N.; 474  
Topp, S.; Gkazi, A. S.; Miller, J.; Shaw, C. E.; Kottlors, M.; Kirschner, J.; 475  
Pestronk, A.; Li, Y. R.; Ford, A. F.; Gitler, A. D.; Benatar, M.; King, O. 476  
D.; Kimonis, V. E.; Ross, E. D.; Weihl, C. C.; Shorter, J.; Taylor, J. P. 477  
Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause 478  
multisystem proteinopathy and ALS. *Nature* **2013**, 495 (7442), 467– 479  
473. 480
- 480 (13) Paul, K. R.; Molliex, A.; Cascarina, S.; Boncella, A. E.; Taylor, J. 481  
P.; Ross, E. D. Effects of mutations on the aggregation propensity of the 482  
human prion-like protein hnRNPA2B1. *Mol. Cell. Biol.* **2017**, 37 (8), 483  
e00652–00616. Qi, X.; Pang, Q.; Wang, J.; Zhao, Z.; Wang, O.; Xu, L.; 484  
Mao, J.; Jiang, Y.; Li, M.; Xing, X. Familial early-onset Paget's disease of 485  
bone associated with a novel hnRNPA2B1 mutation. *Calcif. Tissue Int.* 486  
**2017**, 101, 159–169. 487
- 487 (14) Fareed, M. M.; Dutta, K.; Dandekar, T.; Tarabonda, H.; Skorb, E. 488  
V.; Shityakov, S. In silico investigation of nonsynonymous single 489  
nucleotide polymorphisms in BCL2 apoptosis regulator gene to design 490  
novel protein-based drugs against cancer. *Journal of Cellular* 491  
*Biochemistry* **2022**, 123 (12), 2044–2056. 492
- 492 (15) Wu, B.; Su, S.; Patil, D. P.; Liu, H.; Gan, J.; Jaffrey, S. R.; Ma, J. 493  
Molecular basis for the specific and multivarient recognitions of RNA 494  
substrates by human hnRNP A2/B1. *Nat. Commun.* **2018**, 9 (1), 420. 495
- 495 (16) Shityakov, S.; Fischer, A.; Su, K.-P.; Hussein, A. A.; Dandekar, T.; 496  
Broscheit, J. Novel approach for characterizing propofol binding 497  
affinities to serum albumins from different species. *ACS omega* **2020**, 5 498  
(40), 25543–25551. 499
- 499 (17) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Amino acid 500  
scale: Normalized consensus hydrophobicity scale. *J. Mol. Biol.* **1984**, 501  
179, 125–142. 502
- 502 (18) Glaser, F.; Pupko, T.; Paz, I.; Bell, R. E.; Bechor-Shental, D.; 503  
Martz, E.; Ben-Tal, N. ConSurf: identification of functional regions in 504  
proteins by surface-mapping of phylogenetic information. *Bioinforma-* 505  
*tatics* **2003**, 19 (1), 163–164. 506
- 506 (19) Wang, D.; Liu, D.; Yuchi, J.; He, F.; Jiang, Y.; Cai, S.; Li, J.; Xu, D. 507  
MusiteDeep: a deep-learning based webserver for protein post- 508  
translational modification site prediction and visualization. *Nucleic* 509  
*Acids Res.* **2020**, 48 (W1), W140–W146. 510
- 510 (20) Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K. A.; Lin, G. N.; 511  
Nam, H. J.; Mort, M.; Cooper, D. N.; Sebat, J.; Iakoucheva, L. M.; 512  
Mooney, S. D.; Radivojac, P. Inferring the molecular and phenotypic 513  
impact of amino acid variants with MutPred2. *Nat. Commun.* **2020**, 11 514  
(1), 5918. 515
- 515 (21) Warnecke, A.; Sandalova, T.; Achour, A.; Harris, R. A. PyTMs: a 516  
useful PyMOL plugin for modeling common post-translational 517  
modifications. *BMC Bioinf.* **2014**, 15 (1), 1–12. 518
- 518 (22) Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: 519  
computed atlas of surface topography of proteins. *Nucleic acids research* 520  
**2018**, 46 (W1), W363–W367. 521
- 521 (23) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. 522  
K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: 523  
Automated docking with selective receptor flexibility. *Journal of* 524  
*computational chemistry* **2009**, 30 (16), 2785–2791. 525

- 526 (24) Delgado, J.; Radusky, L. G.; Cianferoni, D.; Serrano, L. FoldX  
527 5.0: working with RNA, small molecules and a new graphical interface.  
528 *Bioinformatics* **2019**, *35* (20), 4168–4169.
- 529 (25) Wang, Z.; Pan, H.; Sun, H.; Kang, Y.; Liu, H.; Cao, D.; Hou, T.  
530 fastDRH: a webserver to predict and analyze protein–ligand complexes  
531 based on molecular docking and MM/PB (GB) SA computation.  
532 *Briefings Bioinf.* **2022**, *23* (5), bbac201.
- 533 (26) Tang, J.; Chen, Z.; Wang, Q.; Hao, W.; Gao, W.-Q.; Xu, H.  
534 hnRNPA2B1 promotes colon cancer progression via the MAPK  
535 pathway. *Frontiers in Genetics* **2021**, *12*, No. 666451. Chen, C.; Huang,  
536 L.; Sun, Q.; Yu, Z.; Wang, X.; Bu, L. HNRNPA2B1 Demonstrates  
537 Diagnostic and Prognostic Values Based on Pan-Cancer Analyses.  
538 *Comput. and Math. Methods Med.* **2022**, *2022*, No. 9867660.
- 539 (27) Liu, N.; Pan, T. N 6-methyladenosine–encoded epitranscrip-  
540 tomics. *Nature structural & molecular biology* **2016**, *23* (2), 98–102.
- 541 (28) Elsakrmy, N.; Cui, H. R-Loops and R-Loop-Binding Proteins in  
542 Cancer Progression and Drug Resistance. *International Journal of*  
543 *Molecular Sciences* **2023**, *24* (8), 7064.
- 544 (29) Humphries, F.; Fitzgerald, K. A. hnRNPA2B1: fueling antiviral  
545 immunity from the nucleus. *Mol. Cell* **2019**, *76* (1), 8–10.
- 546 (30) Wang, L.; Wen, M.; Cao, X. Nuclear hnRNPA2B1 initiates and  
547 amplifies the innate immune response to DNA viruses. *Science* **2019**,  
548 *365* (6454), No. eaav0758.
- 549 (31) Zhang, X.; Flavell, R. A.; Li, H.-B. hnRNPA2B1: a nuclear DNA  
550 sensor in antiviral immunity. *Cell Research* **2019**, *29* (11), 879–880.
- 551 (32) Serçinoğlu, O.; Ozbek, P. Sequence-structure-function relation-  
552 ships in class I MHC: A local frustration perspective. *PloS one* **2020**, *15*  
553 (5), No. e0232849.
- 554 (33) Lim, S. W.; Tan, K. J.; Azuraidi, O. M.; Sathiya, M.; Lim, E. C.;  
555 Lai, K. S.; Yap, W.-S.; Afizan, N. A. R. N. M. Functional and structural  
556 analysis of non-synonymous single nucleotide polymorphisms  
557 (nsSNPs) in the MYB oncoproteins associated with human cancer.  
558 *Sci. Rep.* **2021**, *11* (1), 24206.
- 559 (34) Maguire, J. B.; Haddox, H. K.; Strickland, D.; Halabiya, S. F.;  
560 Coventry, B.; Griffin, J. R.; Pulavarti, S. V. S. R. K.; Cummins, M.;  
561 Thieker, D. F.; Klavins, E.; Szyperski, T.; DiMaio, F.; Baker, D.;  
562 Kuhlman, B. Perturbing the energy landscape for improved packing  
563 during computational protein design. *Proteins: Struct., Funct., Bioinf.*  
564 **2021**, *89* (4), 436–449.
- 565 (35) Holcomb, D.; Alexaki, A.; Hernandez, N.; Hunt, R.; Laurie, K.;  
566 Kames, J.; Hamasaki-Katagiri, N.; Komar, A. A.; DiCuccio, M.; Kimchi-  
567 Sarfaty, C. Gene variants of coagulation related proteins that interact  
568 with SARS-CoV-2. *PLoS computational biology* **2021**, *17* (3),  
569 No. e1008805.
- 570 (36) Shityakov, S.; Skorb, E. V.; Nosonovsky, M. Topological bio-  
571 scaling analysis as a universal measure of protein folding. *Royal Society*  
572 *Open Science* **2022**, *9* (7), No. 220160.
- 573 (37) Shityakov, S.; Skorb, E. V.; Nosonovsky, M. Folding–unfolding  
574 asymmetry and a RetroFold computational algorithm. *Royal Society*  
575 *Open Science* **2023**, *10* (5), No. 221594.
- 576 (38) Kim, E.; Ahuja, A.; Kim, M.-Y.; Cho, J. Y. DNA or protein  
577 methylation-dependent regulation of activator protein-1 function. *Cells*  
578 **2021**, *10* (2), 461.
- 579 (39) Zhao, J.; Zhang, S.; Jiang, Y.; Liu, Y.; Zhu, Q. Mutation analysis of  
580 pathogenic non-synonymous single nucleotide polymorphisms  
581 (nsSNPs) in WFS1 gene through computational approaches. *Sci. Rep.*  
582 **2023**, *13* (1), 6774.
- 583 (40) Dakal, T. C.; Kala, D.; Dhiman, G.; Yadav, V.; Krokhotin, A.;  
584 Dokholyan, N. V. Predicting the functional consequences of non-  
585 synonymous single nucleotide polymorphisms in IL8 gene. *Sci. Rep.*  
586 **2017**, *7* (1), 6525.
- 587 (41) Agozzino, L.; Dill, K. A. Protein evolution speed depends on its  
588 stability and abundance and on chaperone concentrations. *Proc. Natl.*  
589 *Acad. Sci. U. S. A.* **2018**, *115* (37), 9092–9097.
- 590 (42) Fardmanesh, H.; Shekari, M.; Movafagh, A.; Alizadeh Shargh, S.;  
591 Poursadegh Zonouzi, A. A.; Shakerizadeh, S.; Poursadegh Zonouzi, A.;  
592 Hosseinzadeh, A. Upregulation of the double-stranded RNA binding  
593 protein DGCR8 in invasive ductal breast carcinoma. *Gene* **2016**, *581*  
594 (2), 146–151.
- 595 (43) Huang, Y.; Li, Y.; He, F.; Wang, S.; Li, Y.; Ji, G.; Liu, X.; Zhao, Q.;  
596 Li, J. Metastasis-associated protein 3 in colorectal cancer determines  
597 tumor recurrence and prognosis. *Oncotarget* **2017**, *8* (23), 37164.  
598 (44) Maji, D.; Glasser, E.; Henderson, S.; Galardi, J.; Puvino, M. J.;  
599 Jenkins, J. L.; Kielkopf, C. L. Representative cancer-associated U2AF2  
600 mutations alter RNA interactions and splicing. *J. Biol. Chem.* **2020**, *295*  
601 (50), 17148–17157.
- 602 (45) Huang, T.; Li, Y. Current progress, challenges, and future  
603 perspectives of language models for protein representation and protein  
604 design. *Innovation* **2023**, *4* (4), No. 100446.