Article

Identifying Cancer-Relevant Mutations in the DLC START Domain Using Evolutionary and Structure-Function Analyses

Ashton S. Holub [1,2], Renee A. Bouley [3] , Ruben C. Petreaca [4] and Aman Y. Husbands [1,2],*

[1] Department of Molecular Genetics, The Ohio State University, Columbus, OH 43215, USA; holub.28@osu.edu
[2] Center for Applied Plant Sciences (CAPS), The Ohio State University, Columbus, OH 43215, USA
[3] Department of Chemistry and Biochemistry, The Ohio State University, Marion, OH 43302, USA; bouley.8@osu.edu
[4] Department of Molecular Genetics, The Ohio State University, Marion, OH 43302, USA; petreaca.1@osu.edu
* Correspondence: husbands.6@osu.edu

check for updates

Abstract: Rho GTPase signaling promotes proliferation, invasion, and metastasis in a broad spectrum of cancers. Rho GTPase activity is regulated by the deleted in liver cancer (DLC) family of bona fide tumor suppressors which directly inactivate Rho GTPases by stimulating GTP hydrolysis. In addition to a RhoGAP domain, DLC proteins contain a StAR-related lipid transfer (START) domain. START domains in other organisms bind hydrophobic small molecules and can regulate interacting partners or co-occurring domains through a variety of mechanisms. In the case of DLC proteins, their START domain appears to contribute to tumor suppressive activity. However, the nature of this START-directed mechanism, as well as the identities of relevant functional residues, remain virtually unknown. Using the Catalogue of Somatic Mutations in Cancer (COSMIC) dataset and evolutionary and structure-function analyses, we identify several conserved residues likely to be required for START-directed regulation of DLC-1 and DLC-2 tumor-suppressive capabilities. This pan-cancer analysis shows that conserved residues of both START domains are highly overrepresented in cancer cells from a wide range tissues. Interestingly, in DLC-1 and DLC-2, three of these residues form multiple interactions at the tertiary structural level. Furthermore, mutation of any of these residues is predicted to disrupt interactions and thus destabilize the START domain. As such, these mutations would not have emerged from traditional hotspot scans of COSMIC. We propose that evolutionary and structure-function analyses are an underutilized strategy which could be used to unmask cancer-relevant mutations within COSMIC. Our data also suggest DLC-1 and DLC-2 as high-priority candidates for development of novel therapeutics that target their START domain.

Keywords: evolution; structure-function analysis; ligand-binding domain; tumor suppressor; computational genomics; COSMIC; novel druggable therapeutic targets

1. Introduction

Rho GTPases are a subfamily of G proteins involved in signal transduction. These proteins regulate multiple pathways, including the actin cytoskeleton, cell polarity, cell cycle progression, and microtubule dynamics [1]. Signaling is activated by GTP binding at their GTPase domain and is inactivated by its hydrolysis into GDP. This balance is mediated by the opposing activities of Rho guanine nucleotide exchange factors (RhoGEFs) and Rho GTPase-activating proteins (RhoGAPs), which promote GTP replacement and GTP hydrolysis, respectively (Figure 1). Upregulated signaling of Rho GTPases, such as RhoA, Cdc42, and Rac, or their respective RhoGEFs leads to tumor proliferation, migration, invasion, and metastasis in a number of cancers [2–4].
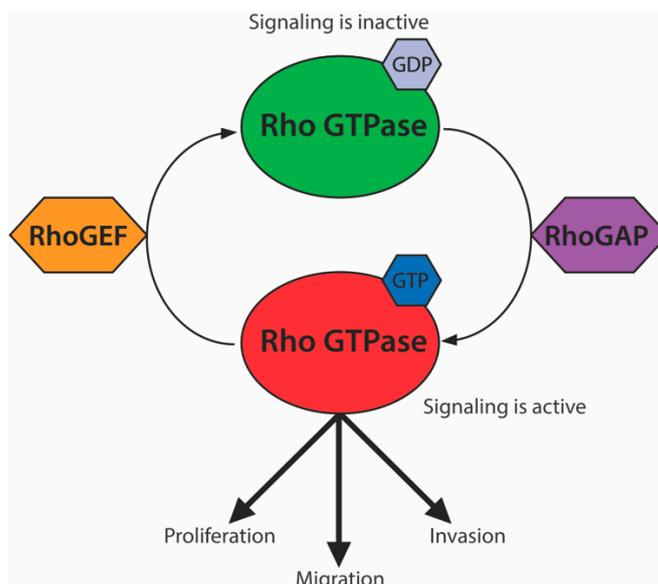
Figure 1. The Rho GTPase signaling module. Rho GTPases (red) are active while bound to GTP. Rho GTPase-activating proteins (RhoGAPs) (purple) stimulate hydrolysis of bound GTP, switching Rho GTPases to an inactive GDP-bound state. Rho guanine nucleotide exchange factors (RhoGEFs) (orange) exchange bound GDP for GTP, reactivating Rho GTPase signaling, which promotes cellular proliferation, migration, and invasion.

Rho GTPases have a globular structure with a limited druggable surface, making them particularly challenging drug targets [5–7]. Efforts to mitigate the effects of Rho GTPases in cancer have therefore focused primarily on disrupting aspects of Rho GTPase signaling (reviewed in [7]). These include studies of RhoGAPs, which have specific consistent inhibitory effects on Rho GTPase signaling and are considered particularly attractive components to target within the Rho GTPase signaling module [1,3,5,7–14] (Figure 1). The deleted in liver cancer 1 (DLC-1 or STARD12), DLC-2 (STARD13), and DLC-3 (STARD8) RhoGAPs are frequently down-regulated in cancer and associated with poor prognosis [12,15]. The Cancer Genome Atlas (TCGA) data indicate that of the three DLC genes, DLC-1 is usually the most dramatically down-regulated, followed by DLC-2 [12]. However, unlike DLC-2, DLC-1 is embryonic lethal when knocked out in mouse models [16,17]. A recent pan-cancer analysis also showed that DLC-1 missense mutations are common in cancer cells [15]. This suggests a larger role for DLC-1 in the control of disease and development. DLC-3, on the other hand, does not show a correlation between copy number loss and cancer progression in the TCGA dataset [12]. Despite making differential contributions to cancer progression, all three DLC proteins are considered to be bona fide tumor suppressors that inhibit cancer cell growth [16,18–24].

Adjacent to the DLC RhoGAP domain is a StAR-related lipid transfer (START) domain. START domains adopt a helix-grip fold structure, forming deep hydrophobic binding pockets that undergo conformational changes upon binding of lipophilic ligands [25–27]. START-containing proteins are distributed throughout Eubacteria and Eukaryota and are important regulators of many biological processes. For instance, in humans, START-containing proteins are associated with numerous diseases, including metabolic, oncogenic, and autoimmune disorders [28–30]. Structurally, these proteins can be divided into two classes: minimal START proteins and multidomain START-containing proteins which possess a wide variety of additional functional domains [31]. START domains employ a range of complex regulatory mechanisms [19,20,32–34]. Of particular interest, START domains in both minimal and multidomain START-containing proteins are able to stimulate regulatory outputs of interacting partners or co-occurring domains in a ligand-dependent manner [32,35]. In the context of DLC proteins, this would be predicted to manifest as START-dependent promotive effects on tumor suppressor activity.

The presence of a START domain suggests DLC proteins may be regulated by a lipophilic ligand, rendering them more easily targetable by natural or synthetic small molecules. However, while there is considerable in vivo and in vitro evidence for them as bona fide tumor suppressors, the role of their START domain remains poorly understood. In fact, the residues mediating START function remain virtually unknown. This is a key bottleneck for targeted modulation of DLC tumor suppressive activity. Here, we use an evolutionary approach to identify conserved residues in DLC START domains that are also mutated in the Catalogue of Somatic Mutations in Cancer (COSMIC) dataset. We find that conserved residues of DLC-1 and DLC-2 START domains are highly overrepresented in tumor samples from a broad spectrum of cancers. Moreover, using comparative structural modeling, we identify a set of three residues, present in both DLC-1 and DLC-2 START domains, that cluster and interact at the tertiary level. Mutations identified in COSMIC are predicted to break these non-covalent interactions and disrupt START tertiary structure and function. Our results support the notion that the START domain is involved in DLC-mediated tumor suppression and that DLC proteins may, thus, be particularly tractable therapeutic targets. Importantly, these residues would not have emerged from traditional scans for mutational hotspots. As such, we propose that tandem evolutionary and structural analyses could help unmask hidden cancer-relevant mutations within the COSMIC dataset.

2. Results and Discussion

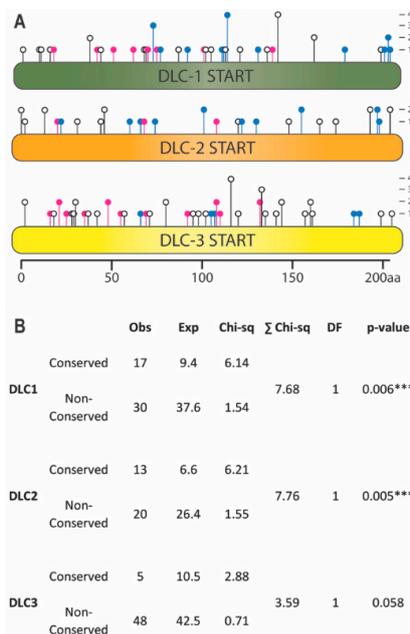### 2.1. Mutations in Conserved Residues of DLC-1 and DLC-2 START Domains Are Overrepresented in Tumors

COSMIC contains numerous missense, nonsense, and frameshift mutations in the START domains of DLC-1, DLC-2, and DLC-3 [36]. We focused on missense mutations, as the latter two are more likely to yield strong effects that confound specific contributions of the DLC START domain. We found 47, 33, and 53 missense mutations falling within the DLC-1, DLC-2, and DLC-3 START domains, respectively (Tables S1–S3). To begin to identify potentially clinically-relevant mutations, we utilized two computational predictors of pathogenicity: combined annotation-dependent depletion (CADD) [37] and the rare exome variant ensemble learner (REVEL) [38]. Both tools scored a large number of mutations as potentially disease-causing, with the meta-predictor REVEL returning fewer candidates than CADD, as expected (Tables S1–S3) [38]. To narrow this list of candidates further, we performed a Kolmogorov–Smirnov (K–S) test for uniformity, commonly used to identify mutational hotspots in cancer [39]. As the K–S test failed to reject the null hypothesis, these mutations do not seem to cluster into obvious hotspots at the primary amino acid sequence level (Table S4). Instead, we hypothesized that clustering occurs at the level of functional residues, which are distributed along the length of the DLC START domain.

To identify functional residues, we performed a multiple sequence alignment (MSA), as evolutionary conservation is a strong indicator of critical residues within protein domains [40]. START domains were selected from 123 DLC-1, DLC-2, and DLC-3 orthologs from 46 vertebrate species spanning 450 million years of divergent evolution. Using a stringent 98% threshold, we identified 41 residues (~20%) within the 206-amino acid DLC START domain that are identical in ≥98% of sequences (Figure 2; see Figure S1 for full species MSA). Missense mutations from COSMIC were placed onto the MSA to score whether they preferentially affected these conserved residues (Figure 3A). DLC-1 had 17 mutations falling in 11 conserved residues, DLC-2 had 13 mutations falling in 10 conserved residues, and DLC-3 had 5 mutations falling in 5 conserved residues (Figure 3; Tables S1–S3). This corresponds to 36.2%, 39.3%, and 9.3% of total mutations in DLC-1, DLC-2, and DLC-3 START domains, respectively (Table S5). We next tested whether these frequencies deviate from the approximately 20% likelihood of mutations occurring in conserved residues by chance (Table S5). Indeed, mutations are significantly overrepresented in conserved residues of both DLC-1 (p-value = 0.006) and DLC-2 (p-value = 0.005) START domains (Figure 3B; Tables S5–S7). By contrast, mutations of the DLC-3 START domain do not appear to be overrepresented in conserved residues (p-value = 0.058). These findings are consistent with the frequent downregulation of DLC-1 and DLC-2, but not DLC-3, in many cancers [12]. Taken together,

these evolutionary analyses identify conserved residues in DLC-1 and DLC-2 START domains which may be linked to cancer phenotypes.
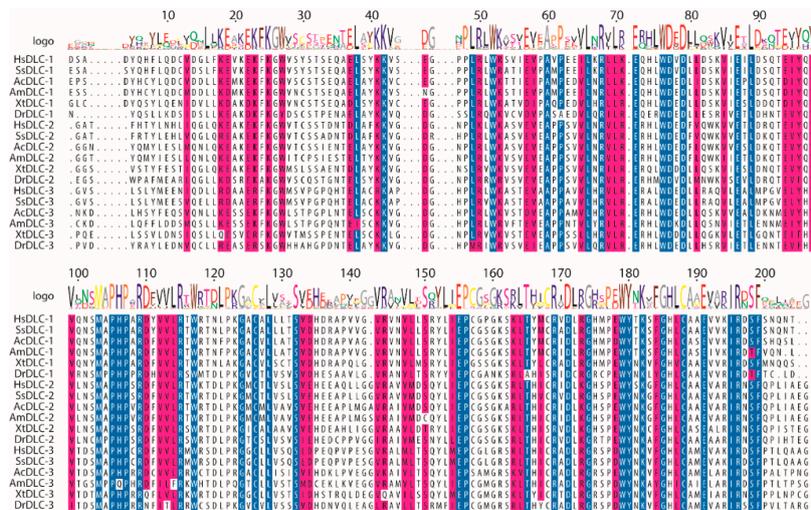
Figure 2. Representative subset of the multiple sequence alignment (MSA) of StAR-related lipid transfer (START) domains from DLC-1, DLC-2, and DLC-3 orthologs. Species include *Homo sapiens* (Hs), *Sus scrofa* (Ss), *Aquila chrysaetos* (Ac), *Alligator mississippiensis* (Am), *Xenopus tropicalis* (Xt), and *Danio rerio* (Dr). Blue indicates residues that are identical in ≥98% of sequences. Magenta indicates residues with similar physiochemical properties in ≥98% of sequences. White represents non-conserved residues. Logo represents the consensus residue(s) of the 123 sequences used in the full alignment (see Figure S1).


Figure 3. Catalogue of Somatic Mutations in Cancer (COSMIC) missense mutations localizing to conserved residues of DLC-1, DLC-2, and DLC-3 START domains. (**A**) Lollipop plot displaying the location and frequency of missense mutations. Mutations occur in identically conserved (blue), physiochemically conserved (magenta), or non-conserved residues (white). (**B**) COSMIC missense mutations are overrepresented in the evolutionarily-conserved residues of the START domains of DLC-1 and DLC-2 but not DLC-3. Content is identical to Table S7. *** $p < 0.001$, chi-square test.

2.2. Conserved Residues form Multiple Bonds within the START Tertiary Structure

Conservation of residues implies critical roles in function [40]. For example, an arginine residue near the opening of the START cavity is highly-conserved and known to be required for START activity [41]; this residue also accrued mutations in multiple independent patient samples (DLC-1 R988; Table S1). Furthermore, mutation of the conserved E966 residue of DLC-1 identified in our analysis impairs tumor suppressor activity [15] (Table S1). Note that this E966K substitution was predicted to be benign by REVEL (Table S1), despite having a clear effect on DLC-1 activity. In addition, a 29-amino acid deletion of the DLC-1 START domain severely reduces Caveolin-1 interaction and DLC-1 tumor suppressor activity [20]. This region contains two conserved residues (R947 and E951) and three residues with evolutionarily-conserved physicochemical properties (E936, I944, L949) (Table S1). The recovery of multiple residues with known roles in either cancer or START activity supports the biological relevance of the residues identified by our analyses. Most conserved DLC START residues are uncharacterized, however, and it is formally possible that amino acid substitutions have minimal effects on tertiary structure [42]. We thus sought to predict their impact on structure and function of DLC START domains using comparative structural modeling. As mutations in the DLC-3 START domain were not overrepresented in COSMIC, it was excluded from these structural analyses.

Informed by the crystal structure of truncated DLC-2 (2PSO; [43]), we generated homology models for both DLC-1 and DLC-2 START domains (Figure 4). B factors indicate several loops in the 2PSO structure are mobile, however we do not expect these disordered loops to affect results. To gauge the quality of our models, we examined interactions for the DLC-1 R988 residue described above, and its DLC-2 analog R1010, using the Crystallographic Object-Oriented Toolkit (CooT; [44]). DLC-1 R988 and DLC-2 R1010 occupy similar positions within their respective models, closely align with the 2PSO structure (RMSD—0.402 Å), and form hydrogen bonds with three residues: P986, R1052, and H1054 of DLC-1 and P1088, K1075, and H1077 of DLC-2 (Figure S2). Furthermore, these residues are either conserved (P986/P1088), have evolutionarily-conserved physicochemical properties (R1052/K1075), or are highly overrepresented in vertebrate DLC-1 and DLC-2 sequences (H1054/H1077) (Figure S1). These data suggest that our models are likely to reflect biologically-relevant conformations of DLC-1 and DLC-2 START domains.
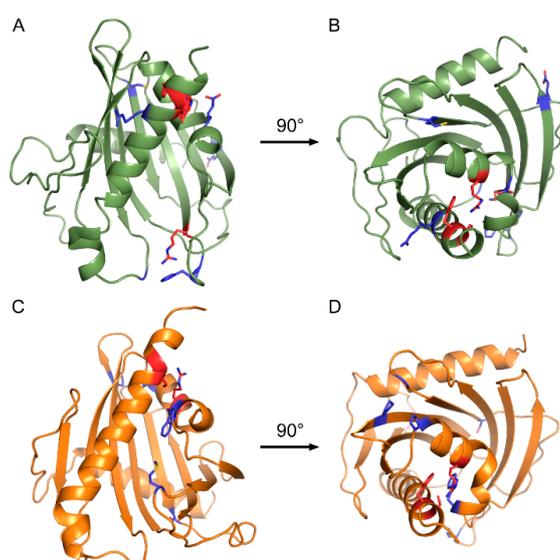

Figure 4. Homology models of DLC-1 and DLC-2 START domains. (**A**,**C**) Side and (**B**,**D**) top-down views of (**A**,**B**) DLC-1 (green) and (**C**,**D**) DLC-2 (orange) START domains. Blue indicates conserved residues with missense mutations in cancers. Red indicates conserved arginine, serine, and phenylalanine residues mutated in both DLC-1 and DLC-2, as well as the highly-conserved arginine in DLC-1 (R988) mutated in numerous cancers.

Mutations in COSMIC falling within conserved residues were then mapped to these homology models (Figure 4). Strikingly, three conserved residues with missense mutations in COSMIC colocalized to the same positions within the DLC-1 and DLC-2 START domains: R947, S1077, and F1078 of DLC-1 (Figure 4A,B) and R969, S1100, and F1101 of DLC-2 (Figure 4C,D). The conserved R947/R969 residues are part of helix-2 and are in close proximity to the S1077/S1100 and F1078/F1101 residues, which sit near the end of the C-terminal helix-3. Given the degree of conservation and close proximity in both DLC-1 and DLC-2, we focused our subsequent analyses on these arginine, serine, and phenylalanine residues.

The close proximity of these residues in three-dimensional space presents the possibility that they may interact at the tertiary level. We examined this idea using CooT [44] and identified multiple predicted interactions between these residues, as well as between these residues and their neighbors. For instance, the R947 residue of DLC-1 forms a cation–π interaction with the aromatic ring of F1078 (3.3 Å; Figure 5A–C). Cation–π interactions in proteins generally occur between a positively charged amino acid and the π face of an aromatic ring and are key contributors to protein structure [45,46]. In addition to this cation–π interaction, R947 also forms a hydrogen bond with Q1081 (3.1 Å). Finally, S1077 does not directly interact with either R947 or F1078 but instead forms a hydrogen bond with K1073 within helix-3. In the case of DLC-2, R969 and F1101 in DLC-2 similarly form a cation–π interaction, although the strength of this interaction is predicted to be weaker (3.8 Å; Figure 6A–C). In contrast to S1077 of DLC-1, S1100 in DLC-2 interacts with R969 via hydrogen bonding at both its carboxy backbone (2.9 Å) and side chain (3.2 Å) of S1100. Thus, while not evident from the primary sequence, three-dimensional homology models reveal that these residues appear to function together at the tertiary level.
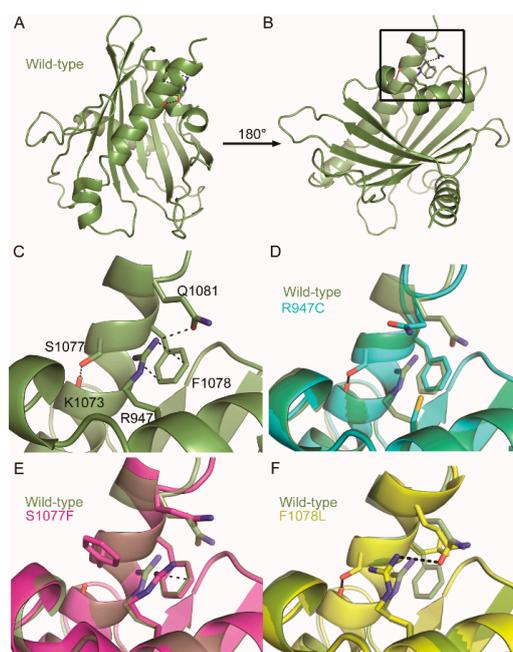
Figure 5. Consequences of COSMIC missense mutations on the tertiary structure of the DLC-1 START domain. (**A–C**) Conserved residues colocalizing near the C-terminal helix form multiple tertiary-level interactions in the wildtype DLC-1 START domain (green). (**D**) R947C disrupts the cation–π interaction of F1078 and hydrogen bond of Q1081. (**E**) F1078L disrupts the R947 cation–π interaction but retains the hydrogen bond between R947 and Q1081. (**F**) S1077F retains the cation–π interaction and disrupts all other bonds. Dotted lines indicate hydrogen bonding and cation–π interactions. Structures carrying the R947C, F1078L, and S1077F mutations are blue, pink, and yellow, respectively. (**C–F**) are zoom-ins of black box in (**B**).
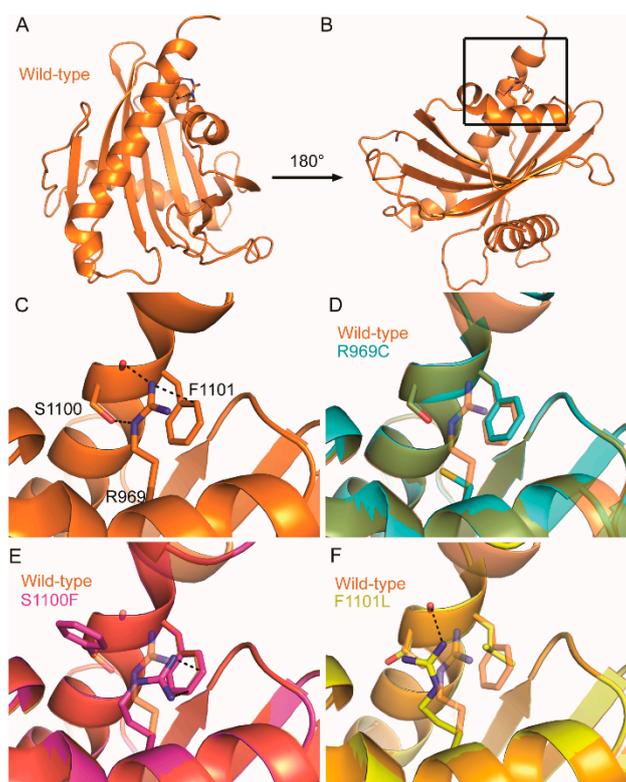
Figure 6. Consequences of COSMIC missense mutations on the tertiary structure of the DLC-2 START domain. (**A–C**) Conserved residues colocalizing near the C-terminal helix form multiple tertiary-level interactions in the wildtype DLC-2 START domain (orange). (**D**) R969C mutation disrupts all interactions formed between these residues. (**E**) F1101L disrupts the cation–π interaction of R969 and the hydrogen bond with the side chain of S1100. Dotted lines indicate hydrogen bonding and cation–π interactions. Structures carrying the R969C, S1100F, and F1101L mutations are blue, pink, and yellow, respectively. (**C–F**) are zoom-ins of black box in (**B**).

2.3. COSMIC Missense Mutations Disrupt Tertiary-Level Interactions in DLC-1 and DLC-2 START Domains

Their conservation, proximity, and extensive intermolecular interactions suggest that missense mutants in the arginine, serine, or phenylalanine residues would disrupt START structure and function. To test this, we separately replaced each residue with one of the missense mutations identified in COSMIC (Tables S1–S3). We then generated new homology models for DLC-1 and DLC-2 START domains and used CoOT to determine the likely effect of these mutations on the previously identified noncovalent network.

In DLC-1, the R947C mutation abolishes its cation–π interaction with F1078 and hydrogen bond with Q1081 but leaves the S1077–K1073 hydrogen bond intact (Figure 5D). S1077F retains the R947 and F1078 cation–π interaction but abolishes the S1077F–K1073 hydrogen bond, as well as the hydrogen bond between R947 and Q1081 (Figure 5E). Finally, F1078L disrupts its cation–π interaction with R947 but retains the hydrogen bond between R947 and Q1081, as well as the S1077–K1073 hydrogen bond (Figure 5F).

In the case of DLC-2, the R969C mutation similarly abolishes its cation–π interaction with F1101, as well as its hydrogen bond to S1100 (Figure 6D). S1100F retains the cation–π interaction with R969 and F1101 while abolishing all hydrogen bonds formed with R969 (Figure 6E). Finally, F1101L abolishes its cation–π interaction with R969 and disrupts the hydrogen bond between R969 and the side chain of S1100 but retains the hydrogen bond between R969 and the carboxyl backbone of S1100 (Figure 6F).

The arginine-to-cysteine mutations in both DLC-1 (R947C) and DLC-2 (R969C) disrupt cation–π interactions, as well as multiple hydrogen bonds. Thus, we predict this substitution to be the most

destabilizing in both START domains. This is also in line with global analyses of the COSMIC dataset, which found that arginine-to-glutamine and arginine-to-cysteine substitutions occur with high frequency in driver mutations [47]. The serine-to-phenylalanine mutations (S1077F in DLC-1 and S1100F in DLC-2) also abolish multiple hydrogen bonds and replace a small polar side chain with a large hydrophobic side chain. Thus, this substitution could conceivably destabilize START domains to a similar extent as the arginine-to-cysteine mutations. By contrast, the phenylalanine-to-leucine mutations (F1078L and F1101L) are predicted to be the least destabilizing, as several hydrogen bonds are left intact and both amino acids have similar physicochemical properties [42]. These structural analyses, focusing on a subset of mutations overrepresented in patient tumors, support a role for the START domain in regulating the tumor suppressor activity of DLC-1 and DLC-2.

3. Conclusions

Functional residues mediating the tumor suppressive activity of the DLC START domain remain virtually unknown. Here, we undertook a pan-cancer analysis and identified residues within DLC-1 and DLC-2 START domains likely to affect the function of these genes. We also show that COSMIC missense mutations in DLC-1 and DLC-2 START domains are significantly overrepresented in evolutionarily-conserved residues predicted to be critical for proper folding, interaction, and/or activity [40,41]. Importantly, the relevance of these potentially clinically-relevant residues to START and DLC tumor suppressive activities is supported by multiple previous findings [15,20]. Our analyses thus provide an evolutionary rationale to begin drug targeting of key residues within the conserved START domains of these genes.

Three of these substitutions, occurring in both DLC-1 and DLC-2, disrupt multiple tertiary-level interactions and are likely to impact START structure. The remaining overrepresented COSMIC missense mutations may similarly impact START activity or affect other properties, such as ligand binding or conformational change. A potential DLC START regulatory mechanism is suggested by studies of both minimal and multidomain START-containing proteins, demonstrating that START domains can stimulate regulatory outputs in an allosteric manner [32,35]. If the DLC START domain functions similarly, it could promote RhoGAP tumor suppressor activity, possibly via allosteric regulation of intra- or inter-molecular interactions. As mutations and deletions in the DLC-1 START domain do not seem to impact the ability of DLC-1 to reduce RhoA [15,20], START-mediated effects may be independent of the DLC RhoGAP domain.

These data also suggest that mutation hotspots might not capture the full spectrum of cancer-relevant mutations in COSMIC. For instance, the conserved arginine, serine, and phenylalanine residues in DLC-1 and DLC-2 appear to work together to stabilize START structure (Figures 5 and 6). Single substitutions in any of these residues would have the same structural and functional consequences for DLC START activity, thereby diluting its signal in COSMIC. We propose that, in addition to traditional analyses of mutation frequency, COSMIC should be queried using coupled evolutionary and structure-function analyses. These tandem approaches may help unmask hidden cancer-relevant mutations.

DLC-1 and DLC-2 are frequently down-regulated, as opposed to mutated, in cancer [12,48]. For instance, DLC-1 expression is reduced 10-fold in lung adenocarcinoma, 3-fold in hepatocellular carcinoma, 4-fold in breast cancer, and 3-fold in colon cancer [12]. Stimulating the activity of residual DLC-1 and DLC-2 proteins through an agonistic START ligand could be a means to improve patient outcomes, particularly if used in combination with current therapeutic approaches. Importantly, agonistic START ligands have already been developed. In plants, signaling of the hormone abscisic acid (ABA) is mediated by its minimal START domain receptor [49]. Opabactin, a rationally designed agonistic START ligand, has a seven-fold stronger affinity than the native ligand and increases ABA-signaling outputs in vivo by a factor of 10 [50]. Agonistic ligands designed for the DLC START domain could compensate for reduced outputs and would be relevant for a number of cancer types. Rational design of these agonists would be greatly aided by the identification of natural DLC START ligands, which remain unknown. The DLC-2 START cavity contains more polar side chains than the

cavities of cholesterol-binding START domains, and DLC START domains were therefore proposed to bind fatty acid, rather than sterol, ligands [43]. Supporting this, the STARD14 START domain, which shares characteristics of the DLC START binding pocket, was recently shown to bind several species of fatty acids [32]. Taken together, our data identify several conserved residues likely to underlie the START-directed regulation of DLC-1 and DLC-2 tumor-suppressive capabilities. DLC-1 and DLC-2 may, thus, be high-priority candidates for development of novel therapeutics targeting their START domain.

4. Materials and Methods

4.1. Mutation Identification and Kolmogorov-Smirnov Test of Uniformity

Mutation data were obtained from the COSMIC database (v91 GRCh38; https://cancer.sanger.ac.uk/cosmic) for the dominant isoforms of DLC-1 (DLC1_ENST00000358919.6), DLC-2 (ENST00000336934.9), and DLC-3 (ENST00000374599.7). DLC START domains were identified using HHpred [51]: DLC-1 (880aa–1083aa), DLC-2 (903aa–1108aa), and DLC-3 (893aa–1098aa). Mutations outside of the START domain, as well as nonsense, frameshift, and intronic mutations within the START domain, were excluded from our analyses. To ensure only independent mutation events were measured, duplicate samples from the same patient were removed before analysis. Pathogenicity scores were obtained from REVEL (from dbNSFP v.4.0a) [38] and CADD (v1.6) [37] using the ENSEMBLE interface (v. 94; https://www.ensembl.org/Tools/VEP) [52]. Missense mutations were counted for individual residues and a one-sample nonparametric test was done using the Kolmogorov–Smirnov test in IBM SPSS Statistics 26. The observed distribution of missense mutations was tested against the null hypothesis that mutations are uniformly distributed along the sample data. Test option settings were set to a significance value of 0.05, a confidence value of 95%, and to exclude cases test-by-test. User-Missing Values were set to "Exclude".

4.2. Multiple Sequence Alignment (MSA)

To build an MSA and identify conserved residues in DLC START domains, orthologs of DLC-1, DLC-2, and DLC-3 were identified using NCBI's Blastp program against the reference proteins (refseq_protein) database [53]. In total, 123 full-length sequences were found from 46 vertebrate species. Sequences were aligned using the ClustalW algorithm [54] in the MEGA X software (v.10.1.7; https://www.megasoftware.net/) [55] and trimmed to exclude residues falling outside the START domain. Conservation of residues was analyzed using R and the "msa" package from Bioconductor [56]. A stringent 98% consensus threshold was used, allowing for only 2 of the 123 sequences to contain a non-similar residue at a given position. This threshold was selected to account for rare lineage specific mutations that may be permissive to changes in critical residues but not representative of most sequences [57]. Residues were then shaded according to similarity and a consensus logo was added using R and the "msa" package from Bioconductor. Finally, this file was exported in .tex format and compiled using TeXworks (v.0.6.5; Free Software Foundation, Inc. Boston, MA, USA; Figure S1).

4.3. Mutation Counts and Chi-Square Analysis

The expected number of mutations in conserved residues assuming a random distribution was calculated by using the percentage of identical residues multiplied by the total number of mutations in the START domain of each paralog.

The expected number of mutations in non-conserved residues was determined by subtracting the expected number of mutations in conserved residues from the total number of mutations. Chi-square values were calculated using the formula (Equations (1) and (2)):

$$\textit{Expected mutations in conserved residues} \quad = \textit{\# of mutations} \times \frac{\textit{\# of conserved residues}}{\textit{Domain length}} \quad (1)$$

$$Chi\ square\ value\ =\ \frac{(Expected - Observed)^2}{Expected} \tag{2}$$

These summed chi-square values were then calculated using IBM SPSS Statistics 26.

4.4. Homology Modeling and Identification of Tertiary-Level Interactions

Homology modeling of each START domain was performed using the online Max Planck Institute Bioinformatics Toolkit (https://toolkit.tuebingen.mpg.de/) [51]. START domain sequences of DLC-1 and DLC-2 were separately queried using HHpred against the Protein Data Bank (PDB_mmCIF70_20_May) structural database to identify related protein structures for homology modelling. The top 10 structures were identical for both START domains and included the crystal structure of DLC-2 (2PSO). These 10 protein structures (chain A of 2PSO, chain A 2R55, chain A of 2MOU, chain A of 5I9J, chain A of 6L1M, chain A of 2E3P, chain C of 3P0L, chain A of 1LN1, chain B of 3FO5, and chain B of 3QSZ) were forwarded to HHpred-TemplateSelection to generate the MSA. Template alignment was subsequently forwarded to MODELLER [58] to predict the tertiary structures of the DLC-1 and DLC-2 START domains. Structural models were examined using the Crystallographic Object-Oriented Toolkit (v.0.8.7; https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/) to identify predicted interactions formed with the conserved residues using the Environmental Distances tool with a cut-off of 4.0 Å [44]. Mutated DLC-1 and DLC-2 START domains were modeled using the same template alignment as their respective wild-type models and analyzed for predicted interactions in CoOT using the same cut-off value.

References

1. Etienne-Manneville, S.; Hall, A. Rho GTPases in cell biology. *Nature* **2002**, *420*, 629–635. [CrossRef] [PubMed]
2. Vega, F.M.; Ridley, A.J. Rho GTPases in cancer cell biology. *FEBS Lett.* **2008**, *582*, 2093–2101. [CrossRef] [PubMed]
3. Haga, R.B.; Ridley, A.J. Rho GTPases: Regulation and roles in cancer cell biology. *Small GTPases* **2016**, *7*, 207–221. [CrossRef] [PubMed]
4. Jung, H.; Yoon, S.R.; Lim, J.; Cho, H.J.; Lee, H.G. Dysregulation of Rho GTPases in Human Cancers. *Cancers* **2020**, *12*, 1179. [CrossRef]
5. Maldonado, M.D.M.; Dharmawardhane, S. Targeting Rac and Cdc42 GTPases in Cancer. *Cancer Res.* **2018**, *78*, 3101–3111. [CrossRef] [PubMed]
6. Shang, X.; Marchioni, F.; Sipes, N.; Evelyn, C.R.; Jerabek-Willemsen, M.; Duhr, S.; Seibel, W.; Wortman, M.; Zheng, Y. Rational Design of Small Molecule Inhibitors Targeting RhoA Subfamily Rho GTPases. *Chem. Biol.* **2012**, *19*, 699–710. [CrossRef] [PubMed]
7. Lin, Y.; Zheng, Y. Approaches of targeting Rho GTPases in cancer drug discovery. *Expert Opin. Drug Discov.* **2015**, *10*, 991–1010. [CrossRef] [PubMed]

8. Prieto-Dominguez, N.; Parnell, C.; Teng, Y. Drugging the Small GTPase Pathways in Cancer Treatment: Promises and Challenges. *Cells* **2019**, *8*, 255. [CrossRef]

9. Garcia-Mata, R.; Boulter, E.; Burridge, K. The invisible hand: Regulation of RHO GTPases by RHOGDIs. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 493–504. [CrossRef]

10. Zhao, L.; Wang, H.; Li, J.; Liu, Y.; Ding, Y. Overexpression of Rho GDP-Dissociation Inhibitor Alpha Is Associated with Tumor Progression and Poor Prognosis of Colorectal Cancer. *J. Proteome Res.* **2008**, *7*, 3994–4003. [CrossRef] [PubMed]

11. Jones, M.B.; Krutzsch, H.; Shu, H.; Zhao, Y.; Liotta, L.A.; Kohn, E.C.; Petricoin, E.F. Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer. *Proteomics* **2002**, *2*, 76–84. [CrossRef]

12. Wang, D.; Qian, X.; Rajaram, M.; Durkin, M.E.; Lowy, D.R. DLC1 is the principal biologically-relevant down-regulated DLC family member in several cancers. *Oncotarget* **2016**, *7*, 45144–45157. [CrossRef] [PubMed]

13. Peck, J.; Douglas, G.; Wu, C.H.; Burbelo, P.D. Human RhoGAP domain-containing proteins: Structure, function and evolutionary relationships. *FEBS Lett.* **2002**, *528*, 27–34. [CrossRef]

14. Hodge, R.G.; Ridley, A.J. Regulating Rho GTPases and their regulators. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 496–510. [CrossRef]

15. Wang, D.; Qian, X.; Sanchez-Solana, B.; Tripathi, B.K.; Durkin, M.E.; Lowy, D. Cancer-Associated Point Mutations in the DLC1 Tumor Suppressor and Other Rho-GAPs Occur Frequently and Are Associated with Decreased Function. *Cancer Res.* **2020**, *80*, 3568–3579. [CrossRef]

16. Durkin, M.E.; Avner, M.R.; Huh, C.-G.; Yuan, B.-Z.; Thorgeirsson, S.S.; Popescu, N.C. DLC-1, a Rho GTPase-activating protein with tumor suppressor function, is essential for embryonic development. *FEBS Lett.* **2005**, *579*, 1191–1196. [CrossRef] [PubMed]

17. Yau, T.O.; Leung, T.H.Y.; Lam, S.; Cheung, O.F.; Tung, E.K.K.; Khong, P.L.; Lam, A.; Chung, S.; Ng, I.O.-L. Deleted in Liver Cancer 2 (DLC2) Was Dispensable for Development and Its Deficiency Did Not Aggravate Hepatocarcinogenesis. *PLoS ONE* **2009**, *4*, e6566. [CrossRef]

18. Durkin, M.E.; Ullmannova, V.; Guan, M.; Popescu, N.C. Deleted in liver cancer 3 (DLC-3), a novel Rho GTPase-activating protein, is downregulated in cancer and inhibits tumor cell growth. *Oncogene* **2007**, *26*, 4580–4589. [CrossRef]

19. Sun, L.; Sun, J.; Song, J. High expression of DLC family proteins predicts better prognosis and inhibits tumor progression in NSCLC. *Mol. Med. Rep.* **2019**, *19*, 4881–4889. [CrossRef]

20. Du, X.; Qian, X.; Papageorge, A.; Schetter, A.J.; Vass, W.C.; Liu, X.; Braverman, R.; Robles, A.I.; Lowy, D. Functional interaction of tumor suppressor DLC1 and caveolin-1 in cancer cells. *Cancer Res.* **2012**, *72*, 4405–4416. [CrossRef]

21. Healy, K.D.; Hodgson, L.; Kim, T.-Y.; Shutes, A.; Maddileti, S.; Juliano, R.L.; Hahn, K.M.; Harden, T.K.; Bang, Y.-J.; Der, C.J. DLC-1 suppresses non-small cell lung cancer growth and invasion by RhoGAP-dependent and independent mechanisms. *Mol. Carcinog.* **2008**, *47*, 326–337. [CrossRef]

22. Lin, Y.; Chen, N.-T.; Shih, Y.-P.; Liao, Y.-C.; Xue, L.; Lo, S.H. DLC2 modulates angiogenic responses in vascular endothelial cells by regulating cell attachment and migration. *Oncogene* **2010**, *29*, 3010–3016. [CrossRef] [PubMed]

23. Leung, T.H.-Y.; Ching, Y.P.; Yam, J.W.P.; Wong, C.; Yau, T.-O.; Jin, N.-Y.; Ng, I.O.-L. Deleted in liver cancer 2 (DLC2) suppresses cell transformation by means of inhibition of RhoA activity. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15207–15212. [CrossRef]

24. Ching, Y.-P.; Wong, C.-M.; Chan, S.-F.; Leung, T.H.-Y.; Ng, D.C.-H.; Jin, D.-Y.; Ng, I.O.-L. Deleted in Liver Cancer (DLC) 2 Encodes a RhoGAP Protein with Growth Suppressor Function and Is Underexpressed in Hepatocellular Carcinoma. *J. Biol. Chem.* **2003**, *278*, 10824–10830. [CrossRef]

25. Iyer, L.; Koonin, E.; Aravind, L. Adaption of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins Struct. Funct. Genetics* **2001**, *43*, 134–144. [CrossRef]

26. Roderick, S.L.; Chan, W.W.; Agate, D.S.; Olsen, L.R.; Vetting, M.W.; Rajashankar, K.; Cohen, D.E. Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nat. Genet.* **2002**, *9*, 507–511. [CrossRef]

27. Murcia, M.; Faráldo-Gómez, J.D.; Maxfield, F.R.; Roux, B. Modeling the structure of the StART domains of MLN64 and StAR proteins in complex with cholesterol. *J. Lipid Res.* **2006**, *47*, 2614–2630. [CrossRef]

28. Soccio, R.E.; Breslow, J.L. StAR-related Lipid Transfer (START) Proteins: Mediators of Intracellular Lipid Metabolism. *J. Biol. Chem.* **2003**, *278*, 22183–22186. [CrossRef]

29. Alpy, F. Give lipids a START: The StAR-related lipid transfer (START) domain in mammals. *J. Cell Sci.* **2005**, *118*, 2791–2801. [CrossRef]

30. Raya, A.; Revert, F.; Navarro, S.; Saus, J. Characterization of a Novel Type of Serine/Threonine Kinase That Specifically Phosphorylates the Human Goodpasture Antigen. *J. Biol. Chem.* **1999**, *274*, 12642–12649. [CrossRef] [PubMed]

31. Schrick, K.; Nguyen, D.; Karlowski, W.M.; Mayer, K.F. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biol.* **2004**, *5*, R41. [CrossRef]

32. Tillman, M.C.; Imai, N.; Li, Y.; Khadka, M.; Okafor, C.D.; Juneja, P.; Adhiyaman, A.; Hagen, S.J.; Cohen, D.E.; Ortlund, E.A. Allosteric regulation of thioesterase superfamily member 1 by lipid sensor domain binding fatty acids and lysophosphatidylcholine. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 22080–22089. [CrossRef]

33. Wang, S.; Li, Q.-P.; Wang, J.; Yan, Y.; Zhang, G.-L.; Zhang, H.; Wu, J.; Chen, F.; Wang, X.; Kang, Z.; et al. YR36/WKS1-Mediated Phosphorylation of PsbO, an Extrinsic Member of Photosystem II, Inhibits Photosynthesis and Confers Stripe Rust Resistance in Wheat. *Mol. Plant* **2019**, *12*, 1639–1650. [CrossRef] [PubMed]

34. Khafif, M.; Balagué, C.; Huard-Chauveau, C.; Roby, D. An essential role for the VASt domain of the Arabidopsis VAD1 protein in the regulation of defense and cell death in response to pathogens. *PLoS ONE* **2017**, *12*, e0179782. [CrossRef]

35. Kanno, K.; Wu, M.K.; Agate, D.S.; Fanelli, B.J.; Wagle, N.; Scapa, E.F.; Ukomadu, C.; Cohen, D.E. Interacting Proteins Dictate Function of the Minimal START Domain Phosphatidylcholine Transfer Protein/StarD2. *J. Biol. Chem.* **2007**, *282*, 30728–30736. [CrossRef]

36. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **2018**, *47*, D941–D947. [CrossRef]

37. Rentzsch, P.; Witten, D.; Cooper, G.M.; Shendure, J.; Kircher, M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **2019**, *47*, D886–D894. [CrossRef]

38. Ioannidis, N.M.; Rothstein, J.H.; Pejaver, V.; Middha, S.; McDonnell, S.K.; Baheti, S.; Musolf, A.; Li, Q.; Holzinger, E.; Karyadi, D.; et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **2016**, *99*, 877–885. [CrossRef] [PubMed]

39. Marhuenda, Y.; Morales, D.; Pardo, M.C. A comparison of uniformity tests. *Statistics* **2005**, *39*, 315–327. [CrossRef]

40. Capra, J.A.; Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics* **2007**, *23*, 1875–1882. [CrossRef] [PubMed]

41. Baker, B.Y.; Epand, R.F.; Epand, R.M.; Miller, W.L. Cholesterol Binding Does Not Predict Activity of the Steroidogenic Acute Regulatory Protein, StAR. *J. Biol. Chem.* **2007**, *282*, 10223–10232. [CrossRef] [PubMed]

42. Castro-Chavez, F. The rule of variation: Amio acid exchange according to the rotating circular genetic code. *J. Theor. Biol.* **2010**, *264*, 711–721. [CrossRef] [PubMed]

43. Thorsell, A.-G.; Lee, W.H.; Persson, C.; Siponen, M.I.; Nilsson, M.; Busam, R.D.; Kotenyova, T.; Schüler, H.; Lehtiö, L. Comparative Structural Analysis of Lipid Binding START Domains. *PLoS ONE* **2011**, *6*, e19521. [CrossRef] [PubMed]

44. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501. [CrossRef]

45. Dougherty, D. Cation-π Interactions in Chemistry and Biology: A New View of Benen, Phe, Tyr, and Trp. *Science* **1996**, *271*, 163–168. [CrossRef]

46. Gallivan, J.P.; Dougherty, D.A. Cation-pi interactions in structural biology. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9459–9464. [CrossRef]

47. Anoosha, P.; Sakthivel, R.; Gromiha, M.M. Exploring preferred amino acid mutations in cancer genes: Applications to identify potential drug targets. *Biochim. Biophys. Acta BBA Mol. Basis Dis.* **2016**, *1862*, 155–165. [CrossRef]

48. Popescu, N.C.; Goodison, S. Deleted in liver cancer-1 (DLC1): An emerging metastasis suppressor gene. *Mol. Diagn. Ther.* **2014**, *18*, 293–302. [CrossRef]

49. Miyakawa, T.; Fujita, Y.; Yamaguchi-Shinozaki, K.; Tanokura, M. Structure and function of abscisic acid receptors. *Trends Plant Sci.* **2013**, *18*, 259–266. [CrossRef]

50. Vaidya, A.S.; Helander, J.D.M.; Peterson, F.C.; Elzinga, D.; Dejonghe, W.; Kaundal, A.; Park, S.-Y.; Xing, Z.; Mega, R.; Takeuchi, J.; et al. Dynamic control of plant water use using designed ABA receptor agonists. *Science* **2019**, *366*, eaaw8848. [CrossRef]

51. Zimmermann, L.; Stephens, A.; Nam, S.-Z.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A.N.; Alva, V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **2018**, *430*, 2237–2243. [CrossRef]

52. Yates, A.D.; Achuthan, P.; Akanni, W.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl. *Nucleic Acids Res.* **2019**, *48*, D682–D688. [CrossRef]

53. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8–D13. [CrossRef]

54. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [CrossRef] [PubMed]

55. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]

56. Bodenhofer, U.; Bonatesta, E.; Horejš-Kainrath, C.; Hochreiter, S. Msa: An R package for multiple sequence alignment. *Bioinformatics* **2015**, *31*, 3997–3999. [CrossRef] [PubMed]

57. Bridgham, J.T.; Carroll, S.M.; Thornton, J.W. Evolution of Hormone-Receptor Complexity by Molecular Exploitation. *Science* **2006**, *312*, 97–101. [CrossRef]

58. Šali, A.; Potterton, L.; Yuan, F.; Van Vlijmen, H.; Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins Struct. Funct. Bioinform.* **1995**, *23*, 318–326. [CrossRef]