Research Article
Open Access
CrossMark

Homologs of the tumor suppressor protein p53: A bioinformatics study for drug design

Abstract

Sequence and structure of proteins related to the tumor suppressor protein p53 were studied from the perspective of gaining insight for the development of therapeutic drugs. Our study addresses two major issues that encumber bringing novel drugs to market: side effects and artifacts from animal models. In the first phase of our study, we performed a genome-wide search to identify potentially similar proteins to p53 that may be susceptible to off target effects. In the second phase, we chose a selection of common model organisms that could potentially be available to undergraduate researchers in the university setting to assess which ones utilize p53 most similar to humans on the basis of sequence homology and structural similarity from predicted structures. Our results confirm the proteins in significantly similar to p53 are known paralogs within the p53 family. In considering model organisms, murine p53 bore great similarity to human p53 in terms of both sequence and structure, but others performed similarly well. We discuss the findings against the background of other structural benchmarks and point out potential benefits and drawbacks of various alternatives for use in future drug design pilot studies.

Keywords: tumor suppressor protein p53, cancer, homology modeling, structure prediction, drug design, human tumors

Volume 9 Issue 1 - 2020

Kelly M Thayer, Claudia Carcamo
Department of Chemistry, Vassar College, USA

Correspondence: Kelly M Thayer, Department of Chemistry, Vassar College, 124 Raymond Avenue, Poughkeepsie, NY 12604, USA, Email kethayer@vassar.edu

Received: January 15, 2020 | **Published:** February 05, 2020

Introduction

The tumor suppressor protein p53 plays the vital role in the cell of preventing tumor formation. Known as the "Guardian of the Genome,"[1] p53 initiates one of two major pathways when a cell experiences insults, such as ionizing radiation, hypoxia, carcinogens, and oxidative stress, to the integrity of the DNA:[2] apoptosis or DNA repair.[3–8] These options are directed by post translational modifications to the N and C terminal regions of p53.[9–13] As a transcription factor, p53 recognizes an ensemble of related but non-identical recognition sequences,[6] and the post translational modifications influence the affinity for sequences, thereby serving as a complex switching mechanism to transmit a variety of cellular signals to the activation of the appropriate response to cell damage.[1,12] Both alternatives prevent tumor development, thereby protecting organisms from the development of cancer. However, work conducted in mouse models strongly suggests that in the absence of p53, tumors can readily develop and corroborates the idea that mutated p53 can allow tumor formation.[14–16] In fact, more than 50% of all human cancers involve mutant p53,[17] and sequence analysis of p53 extracted from human tumors indicates a series of "hotspot" mutations.[18,19] Although much progress has been made, the National Cancer Institute and the American Cancer Institute continue to cite cancer among the leading causes of death worldwide; in the United States alone, some one million new cases of cancer have been projected to occur in 2019 accompanied by an estimated six hundred thousand deaths based on projections from the National Center for Health statistics database current through 2016 at the time of writing.[20–22]

The 393 amino acid sequence of p53 is comprised of three domains: the transactivation domain, the DNA binding domain (DBD) which interacts directly with the DNA, and the carboxyterminal domain involved in tetramerization.[23–25] While the full length p53 structure has not been determined crystallographically, the core domain has been crystallized,[26–28] and we have reported on the modelled full length structure.[29,30] Although the N- and C- terminal domains are intrinsically disordered,[31] they are not without secondary structure; they both contain alpha helices connected to the DBD via a highly flexible loop. The central DBD contains a series of β-sheets in the β-sandwich configuration. Notable key features, localized in this domain, include four loops: L1 to L3 and a strand loop helix involved in the recognition of the cognate DNA sequences. The structure also features a zinc ion tetrahedrally coordinated between Cys176 and His179 in L2, and Cys238 and Cys 242 in L3; this provides structural stability to the protein and orients the eight residues mediating interaction with the DNA.[26,32,33]

A molecular level approach to developing therapeutics has shown promise for restoring native p53 functionality. The R175H gain of function mutation exhibits a variety of aberrant behaviors, among them failure to activate major targets.[34,35] Rescue of p53 R175H wild type activity was suggested by restoration of upregulation of three target p53 genes, Bax, Puma, and Nox by PRIMA-1 and the closely related PRIMA-1^{MET} analog.[36] It is metabolically converted to methylene quinuclidinone, which is capable of reacting with cysteine thoils,[37] which is believed to be the mechanism of reactivation of p53. The compound, also known as APR46, has advanced to clinical trials[38] and cancer.gov reports ongoing Phase III trials. Encouraged by this example of restoration of function of p53 by a small molecule, our lab is working towards the integration of computational chemistry, structural bioinformatics, and mathematical modeling to rationally design such small molecules for p53.

Two new molecular dynamics based methods developed in our lab are providing insight on the structure and function of p53, thereby facilitating inroads for allosteric rational drug design. The structure of macromolecules is known to be dynamic due to thermal

fluctuations. While X-ray crystallography, NMR spectroscopy, and most recently cryo-electron microscopy are useful in gaining overall insight on the structure, Molecular Dynamics (MD) simulations on structures starting from these and related engineered structures has provided a more complete understanding of molecular regulation.[39–41] MD simulations enumerate specific snapshots commensurate with the Boltzmann ensemble of structures accessible at a given temperature. MD reports on the free energy landscape of the molecule and the interchange between substates as mathematically described by classical statistical mechanics. Within the free energy landscape, the snapshots tend to populate substates. Markov state models[42,43] based on MD simulations (MD-MSMs) provide a mathematical framework for capturing these dynamics.[44,45] Our recent MD-MSMs have raised the idea that perturbations to molecules such as by ligand binding or mutation can be cast as changes to the relative populations in the free energy landscape.[43] A possible means of restoring native functionality to p53 mutants entails developing small molecules to bind to p53 that will redistribute the substates to more closely match the native distribution.

To complement our use of MD-MSMs for drug discovery, we have also developed MD sectors.[46] This method identifies key residues involved in network signaling with a protein or other macromolecule. The work borrows many ideas from classic allosteric signaling, a well-known phenomenon in biology in which a perturbation occurs to a molecule at one point, and manifests itself as a change at a distal location. Allosteric regulators operate in this way, binding to one site to modulate affinity in the active site. However, in the case of restoring functionality, the allosteric site is not necessarily known *a priori*, and no specific protocol exists to identify it. Our method works to identify a group of residues having concerted motions in MD simulations as predictors of residues involved in a control network. While mutations to p53 on the near the surface of the molecule may be reversible by small molecules such as PRIMA1,[36] a more generalized practice will be necessary to deal with more deeply buried changes. This offers a possible way of identifying a network of residues which can be screened for residues on the surface. This subset can then be computationally verified, and used for *in silico* drug design.

We thus turn our attention to common issues that hinder potential therapeutics from advancing into FDA approved drugs. One of the most pressing problems is off target effects;[47–49] when a drug molecule binds to an unintentional target, unexpected negative results may ensue. Homologs within the same family[50] by their very nature of similarity to the target are especially susceptible to off target effects. This problem is well illustrated by protein kinases[51,52] which share considerable sequence and structure yet provide high level control of a variety of signal transduction cascades with far reaching physiological effects. p53 has membership in a family including p63 and p73;[53–56] these homologs play roles in developmental pathways in addition to having some overlapping functionality with p53. The role of p63 is primarily implicated in epithelial cell development and germline protection.[57,58] However, it also interacts with the anti-apoptotic protein Bcl-xL,[59] linking it to a role in apoptosis. Differentiating between homologs requires including structural and functional knowledge of family members during the development phase and capitalizing on features unique to the intended target, thereby mitigating off target effects. Therefore, identifying new homologs and exploring their structural similarity to the target is explored for p53 in this study.

Another issue that arises during drug discovery occurs at the level of animal model selection; with a signaling protein as complex as p53, relying on trials in alternate organisms' native system can potentially introduce extraneous information for human use. Such an example in developing MDM2 inhibitors based on the fragments in mouse models is the possible reason for failure; success in mouse models was followed by failure at the clinical trials when the compound was introduced into humans.[7,60,61]

In our present research, the issues of off target activity and potential extraneous results from animal models are addressed using computational bioinformatics approaches. First we perform bioinformatic analysis using a BLAST search against the human genome to investigate whether any homologs to p53 in addition to the known p63 and p73 exist based on sequence similarity. Our second area of focus is on structural analysis of homologs in a set of commonly chosen model organisms available for undergraduate researchers. We aim to characterize the sequence similarity and structural similarity of p53 in these organisms with the human form. Conformational changes conceivably could be affected by ligand binding; they might be sensitive to sterics added or subtracted with subtle changes to the amino acid sequence. In the absence of structural data for model organisms, the coordinates are predicted using the Phyre2 protein structure prediction suite. The structure and sequence information was then used to generate suggestions as to which model organisms may lead to the most predictive outcomes for human clinical trials.

Materials and methods

Protein sequence accession

Sequences used in the multiple sequence alignments were obtained from the UniProt[62,63] database via the web interface. For the *homo sapiens* p53 sequence, isoform one of entry P04637 was used for all comparisons.[10] Sequences were obtained for both the starting search for homologs of p53 in humans, and for obtaining sequences of p53 in other model organisms.

Homolog search

The BLAST[62,64,65] search algorithm was used to search for p53 homologs within the human genome. Results of blasting the protein sequence of human p53 isoform one from UniProt entry P04637 were evaluated for homology based on Max Score, Total Score, Query Cover, E value, and Percent Identity. Proteins with >80% Percent Identity were sought for further analysis. Because the initial search did not yield any new homologs, alternate search methods were also employed. As an alternative, a BLASTn search was run on the human genome selecting "somewhat similar sequences". A third method entailed using protein BLAST to search for the protein sequence of p53. Search results from the nucleotide search were converted to the corresponding protein sequence after the search and prior to alignment with p53.

Multiple sequence alignments (MSA) and phylogenetic trees

Clustal Omega[64,66,68] was used to generate an MSA and percent identity matrix for homology comparison across model organisms. The color align program (http://www.bioinformatics.org/sms2/) was used to annotate positions in the MSA exhibiting greater than or equal to 60% identity across all organisms. The phylogenetic tree feature generated the dendogram from the multiple sequence representation. The rooted phylogenetic tree representation displays the relationship

between the organisms in the multiple sequence alignment. The branch lengths were calculated by setting the number of evolutionary substitutions proportional to the length of the alignment, excluding gaps.

Selection of model organisms

We sought to study sequences of p53 in organisms that are available or would be reasonably available to undergraduate student researchers for follow up work in the drug design pipeline. The organisms chosen for study include *Ratticus Norwegicus* (rat), *Mus musculus* (mouse), *Danio rerio* (zebrafish), *Xenopus* (frog), *Gallus gallus* (chicken), *Bos Taurus* (cow), *Oryzias latipes* (Japanese rice fish), *Drosophila melanogaster* (fruit fly), and *Caenorhabditis elegans* (nematode). Protein sequences were taken from UniProt and structures were built using Phyre2.

Protein structure prediction

Phyre2[69] was used to build protein structures. Sequences were entered, and folding was carried out using the intense mode. While the Protein Data Bank (PDB) contains structures of p53 and p63, the Phyre folded versions were used in this study because the full length version is not available. We have previously carried out detailed analysis of folding of p53 as compared to the crystal structure, and the DNA binding domain assumes a nearly identical structure as the crystallographic form (RMSD=0.442, 175 atoms).[30] We therefore have opted to work with the full length Phyre predicted structures for p53 and p63 in this study.

Molecular visualization

PyMol[70] was used to visualize the protein structures and perform structural alignments. This software was used to overlay human p53 and p63 as compared to that of the chosen model organism p53 sequences. Root Mean Square Deviation (RMSD) values were recorded.

Results

Search for homologs within the human genome

Blast searches using the sequence of p53, both at the protein and nucleotide level, were used to search for paralogous homologs within the human genome. The hits were analyzed on the basis of several statistics: Max Score, Total Score, Query Cover, E-value, and Percent Identity. Maximum Score reports the highest alignment score of a set of aligned segments from the same query; it is computed as the sum of the match score and the penalties for mismatch and gap open and extend penalties computed independently for each segment. The total score reports the sum of alignment scores of all segments again from the same query, but it can differ from the MaxScore if several parts of the database sequence match different parts of the query sequence. The BLAST E-value ranks the sequences the same as the Max Score; it reports the number of expected hits of similar integrity that could be found by chance. Thus a low E-score indicates a better match, and an E-score of 0 indicates identity between the query sequence and the database match. The Query Cover tells how much of the sequence was aligned, and the Percent Identity tells the extent to which the matching positions were identical.

Table 1 shows the top 40 results from the protein BLAST search using the human p53 isoform 1 sequence as the query sequence. Applying the criterion of hits with >80% similarity, matches for variants of p53, p63 and p73 were identified. The top 10 hits had E-scores of 0, indicating an excellent match, while having high sequence identity (>97%). This was at the expense of decreasing coverage of the hits, attributable to the splicing variations present in the various isoforms of p53 dominating the highest ranking sequences. The next group of matches exhibited relatively low E-scores, albeit with substantially lower percent identity with coverage at around the 75% rate. This regime identified variants of p63 and p73 and accounts for 28 sequences. The final 2 matches reported from this search exhibited poor E-scores with poor sequence identity and poor coverage; we did not consider them further as matches. While the results did find the known homologs p63 and p73, no new homologs were discovered by this process. To ensure the result was not an artifact of the search parameters, variants were carried out using the nucleotide based search, which similarly identified the known homologs but no new ones were unearthed by this method. These results therefore confirm p63 and p73 as the most important homologs to consider while developing small molecule therapeutics to target p53.

Table 1 Top 40 pBLAST hits for p53 sequence in *Homo sapiens* genome. Corresponding maximum score, total score, E-value, query cover and percent identity are given for matching UniProt accession numbers of matches and their descriptions. See text for explanations of terms

Accession number	Description	Max score	Total score	Query cover E value	Percent identity
|NP_000537.3|	cellular tumor antigen p53 isoform a [Homo sapiens] 813	813	100%	0	100.00%
|NP_001119590.1|	cellular tumor antigen p53 isoform g [Homo sapiens] 737	737	90%	0	100.00%
|NP_001119585.1|	cellular tumor antigen p53 isoform c [Homo sapiens] 689	689	86%	0	97.65%
|NP_001119586.1|	cellular tumor antigen p53 isoform b [Homo sapiens] 688	688	84%	0	100.00%
|NP_001263624.1|	cellular tumor antigen p53 isoform h [Homo sapiens] 612	612	76%	0	97.35%
|NP_001263625.1|	cellular tumor antigen p53 isoform i [Homo sapiens] 612	612	74%	0	100.00%
|NP_001119587.1|	cellular tumor antigen p53 isoform d [Homo sapiens] 549	549	66%	0	100.00%
|NP_001263626.1|	cellular tumor antigen p53 isoform j [Homo sapiens] 491	491	59%	3E-176	100.00%
|NP_001119588.1|	cellular tumor antigen p53 isoform e [Homo sapiens] 422	422	50%	2E-149	100.00%
|NP_001119589.1|	cellular tumor antigen p53 isoform f [Homo sapiens] 422	422	53%	3E-149	96.17%

Table continue

Accession number	Description	Max score	Total score	Query cover E value	Percent identity
|NP_001263627.1|	cellular tumor antigen p53 isoform k [Homo sapiens] 365	365	43%	3E-127	100.00%
|NP_001263628.1|	cellular tumor antigen p53 isoform l [Homo sapiens] 364	364	46%	7E-127	95.60%
|NP_001191118.1|	tumor protein p73 isoform e [Homo sapiens]	273	74%	1E-88	47.68%
|NP_001191115.1|	tumor protein p73 isoform j [Homo sapiens]	275	74%	2E-88	47.68%
|NP_001119714.1|	tumor protein p73 isoform d [Homo sapiens] 273	273	74%	1E-87	47.68%
|NP_001191114.1|	tumor protein p73 isoform i [Homo sapiens] 274	274	74%	2E-87	47.68%
|NP_001316074.1|	tumor protein 63 isoform 8 [Homo sapiens] 271	271	70%	3E-87	48.59%
|NP_001191120.1|	tumor protein p73 isoform g [Homo sapiens] 274	274	74%	3E-87	47.68%
|NP_001119713.1|	tumor protein p73 isoform c [Homo sapiens] 273	273	74%	4E-87	47.68%
|NP_001108454.1|	tumor protein 63 isoform 6 [Homo sapiens] 270	270	70%	5E-87	48.59%
|NP_001191113.1|	tumor protein p73 isoform h [Homo sapiens] 274	274	74%	5E-87	47.68%
|NP_001191119.1|	tumor protein p73 isoform f [Homo sapiens] 274	274	74%	5E-87	47.68%
|NP_001191117.1|	tumor protein p73 isoform l [Homo sapiens] 275	275	74%	6E-87	47.68%
|NP_001191116.1|	tumor protein p73 isoform k [Homo sapiens] 275	275	74%	1E-86	47.68%
|NP_001108453.1|	tumor protein 63 isoform 5 [Homo sapiens] 271	271	70%	2E-86	48.59%
|NP_001316078.1|	tumor protein 63 isoform 11 [Homo sapiens] 269	269	70%	3E-86	48.04%
|NP_001191121.1|	tumor protein p73 isoform m [Homo sapiens] 273	273	74%	5E-86	47.68%
|NP_001316073.1|	tumor protein 63 isoform 7 [Homo sapiens] 271	271	70%	6E-86	48.59%
|NP_001119712.1|	tumor protein p73 isoform b [Homo sapiens] 273	273	74%	7E-86	47.68%
|NP_001108451.1|	tumor protein 63 isoform 3 [Homo sapiens] 270	270	70%	8E-86	48.59%
|NP_005418.1|	tumor protein p73 isoform a [Homo sapiens] 274	274	74%	1E-85	47.68%
|NP_001108450.1|	tumor protein 63 isoform 2 [Homo sapiens] 270	270	70%	5E-85	48.59%
|NP_001108452.1|	tumor protein 63 isoform 4 [Homo sapiens] 271	271	70%	7E-85	48.59%
|NP_001316893.1|	tumor protein 63 isoform 13 [Homo sapiens] 271	271	70%	5E-84	48.59%
|NP_003713.3|	tumor protein 63 isoform 1 [Homo sapiens] 271	271	70%	6E-84	48.59%
|NP_001316077.1|	tumor protein 63 isoform 10 [Homo sapiens] 268	268	70%	9E-83	48.04%
|NP_001316079.1|	tumor protein 63 isoform 12 [Homo sapiens] 242	242	57%	5E-77	50.22%
|NP_001316075.1|	tumor protein 63 isoform 9 [Homo sapiens] 245	245	59%	8E-76	50.00%
|XP_024304272.1|	phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2 isoform X3 [Homo sapiens]30.8	30.8	24%	7.5	32.35%
|NP_001558.3|	phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2 [Homo sapiens] 30.8	30.8	18%	7.5	38.46%

Search for orthologs in model organisms

In the second phase of our study, we sought to identify which model organisms reasonable for use by undergraduate researchers have orthologs most similar to human p53 by comparison of both sequence and structure. The protein sequences for the model organisms obtained from UniProt were compared with a multiple sequence alignment (Figure 1). A striking feature of the MSA is the extent of similarity of p53 across most of the organisms studied, yet only 10 of the positions contain sequence identity. Stars of the 393 amino sequences of human p53, the central region encompassing the core domain and portions of the N- and C- terminal tails exhibited the greatest extent of sequence similarity. The phylogenetic tree representation of the MSA was generated using Clustal Omega (Figure 2). The tree shows that the most closely related p53 sequences to humans are cow, rat, and mouse, the three mammals represented by the sampling. The most distant organisms were nematode and fly, which were found at a similar distance as the human p63, included as a control.

In addition to the sequence analysis, structural similarity was also

investigated. The three dimensional structures of the model organism sequences generated via Phyre protein folding were compared to the human p53 structure, and to p63 as a control (Figure 3). The RMSD values for the alignments categorized them as being more similar to either p53 or p63 (Table 2). A scatter plot of the RMSD values (Figure 4) indicates the relative distance of each structure to each of the reference structures. Rat, mouse, and zebrafish have structures identical to p53 and similar to p63. Cow, frog, and chicken also exhibit small RMSD values less than 1 Angstrom and appear approximately equidistant from p53 and p63. Japanese rice fish is slightly closer to p53, but bears a larger RMSD. The most structurally divergent are nematode and fruit fly, consistent with them also having the most deviant sequence as observed in the dendogram.

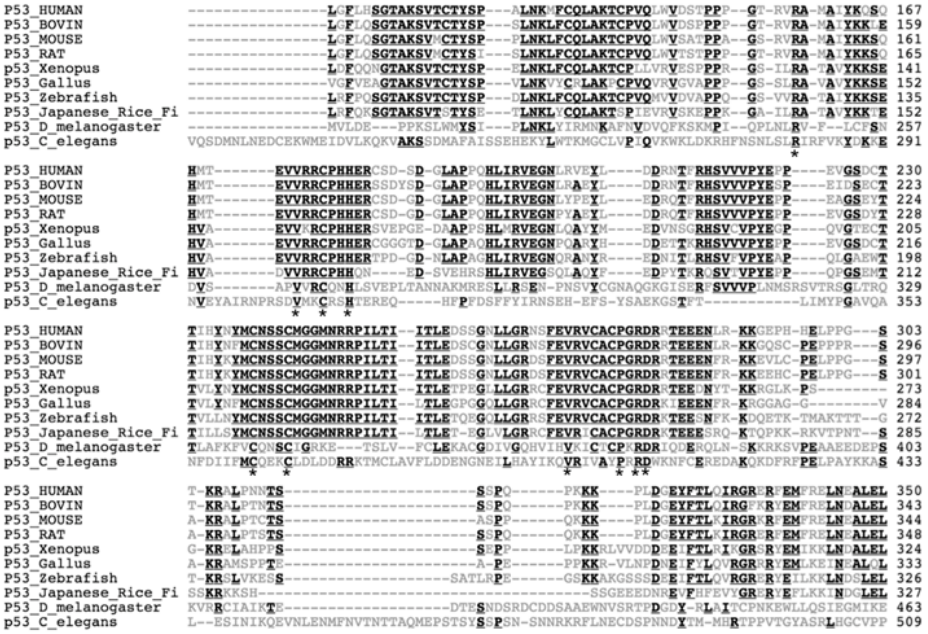

Figure 1 Sequence alignments of p53 between the selected model organisms. A sequence alignment was performed where ≥60% of the sequences were required to match for the residues to be bolded. The alignment selected focuses on the core domain. Asterisks indicate residues that are 100% conserved across the organisms selected.

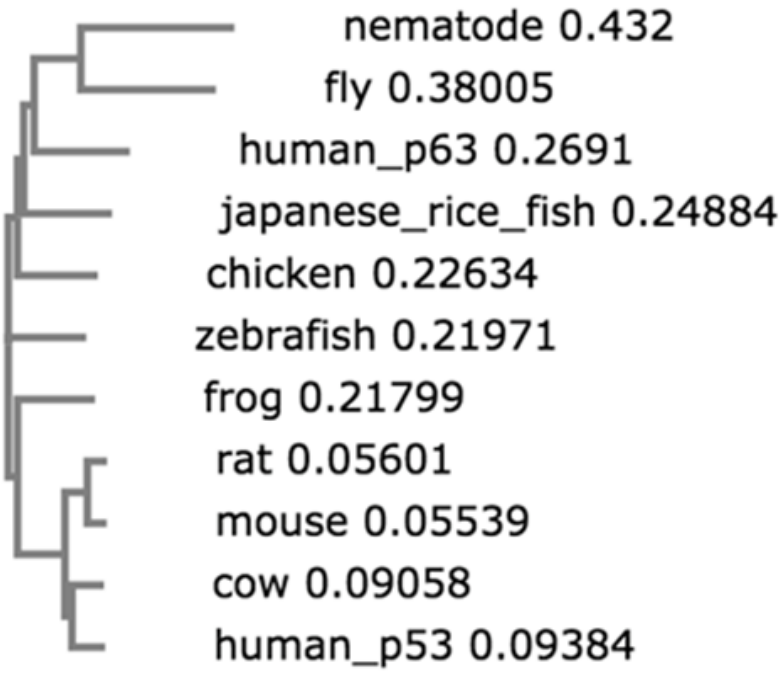

Figure 2 Phylogenetic tree of p53 sequences of model organisms. Human p53 and p63 are shown for reference. The dendogram was rendered with real branch lengths based on the MSA (Figure 1). The branch lengths are displayed to the right of the model organism names.

Citation: Thayer KM, Carcamo C. Homologs of the tumor suppressor protein p53: A bioinformatics study for drug design. *MOJ Proteomics Bioinform.* 2020;9(1):5–14. DOI: 10.15406/mojpb.2020.09.00270

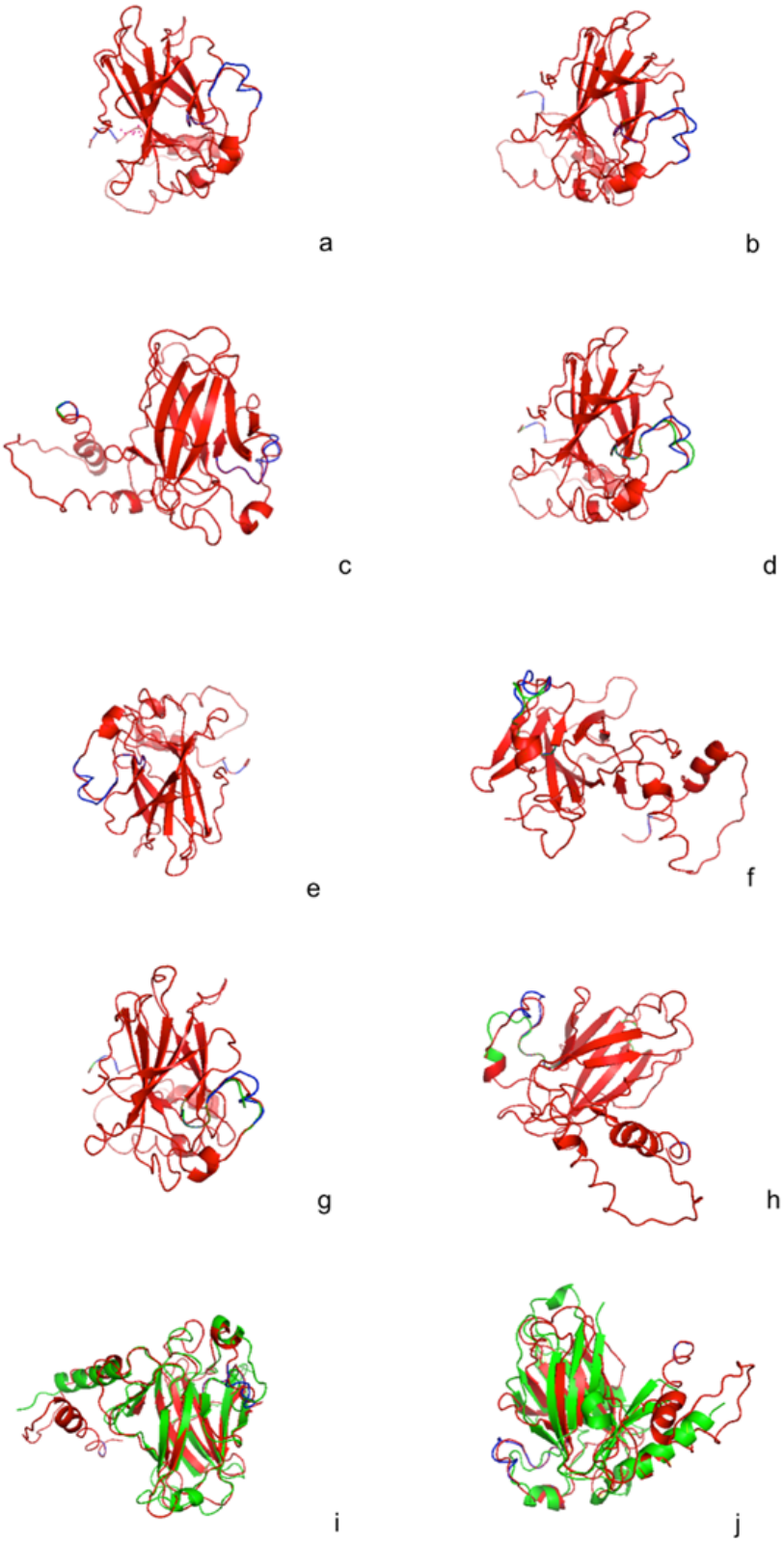

Figure 3 Structural alignment of aligned structures of human p53, shown in red, p53 for select model organisms, shown in green, and human p63, shown in blue. The panels depict the following: (a) rat, (b) mouse, (c) zebrafish, (d) frog, (e) human p63, (f) chicken, (g) cow, (h) Japanese rice fish, (i) fruit fly, and (j) nematode. Structure visualization and alignment was performed using PyMol.

Citation: Thayer KM, Carcamo C. Homologs of the tumor suppressor protein p53: A bioinformatics study for drug design. *MOJ Proteomics Bioinform.* 2020;9(1):5–14. DOI: 10.15406/mojpb.2020.09.00270

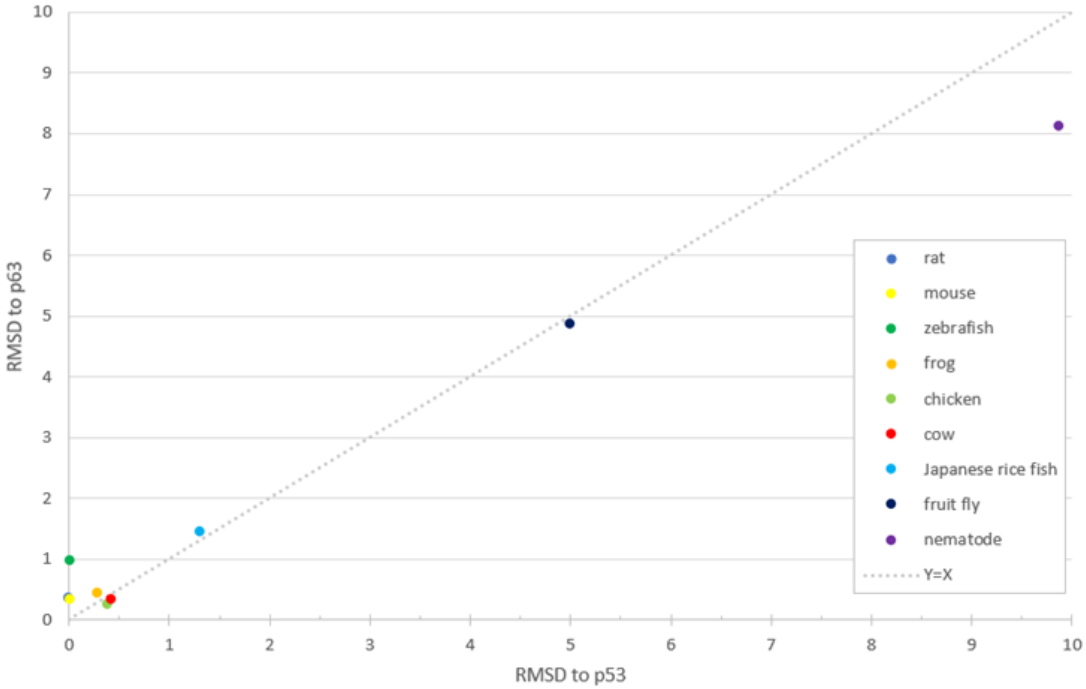

Figure 4 Distances of Phyre folded structures of model organisms to human p53 and p63. The x-axis indicates RMSD distance from p53, and y-axis displacement indicates RMSD from p63. The line Y=X is shown in grey for reference.

Table 2 RMSD values for the aligned structures of human p53, and human p63 to select model organisms. The table includes RMS values for each structural alignment. The bolded numbers indicate to which structure, p53 or p63, the structural alignment is more closely matched

Organism	Panel	Human p53 RMS	Human p63 RMS
	panel letter	RMSD to p53	RMSD to p63
Ratticus Norwegicus (rat)	A	**0.00**	0.36
Mus musculus (mouse)	B	**0.00**	0.36
Danio rerio (zebrafish)	C	**0.00**	0.99
Xenopus (frog)	D	**0.28**	0.45
Homo sapiens p63 (human p63)	E	**0.36**	NA
Gallus gallus (chicken)	F	0.38	**0.27**
Bos taurus (cow)	G	0.41	**0.35**
Oryzias latipes (japanese rice fish)	H	**1.30**	1.47
Drosophila melanogaster (fruit fly)	I	4.99	**4.89**
Caenorhabditis elegans (nematode)	J	9.87	**8.13**

Discussion

In this study we sought to examine homologs of p53 from the perspective of potential issues we may encounter while developing therapeutics to selectively target p53 tumor suppressor protein and potentially carry out trials of any such molecules in model organisms. Our study shed light on both the potential for cross reacting proteins that could cause side effects, as well as the suitability of common model organisms for testing the effectiveness of drugs on p53. First we considered the issue of off target hits, which could cause undesired side effects. Operating on the hypothesis that other proteins in the genome having a similar sequence may share a similar structure and therefore be susceptible to modulation by any small molecules developed in the lab, we adopted a bioinformatic approach to search the human genome to discover potentially similar proteins. Our BLAST searches uncovered significant matches only for variants of p53, p63,

and p73 (Table 1), all characterized members of the p53 family.[2,52,53,71] Drug development will thus focus on developing small molecules that either benefit all three or that capitalize on the structural differences between them to selectively interact with the desired target. The second phase of our study focused on examining the suitability of p53 in model organisms for studying the effects of treating drugs. A potential issue is that p53 may evolve divergently in organisms, particularly because it is a hub of signaling and different interactions may have evolved in different systems. We have examined several model organisms potentially available for modeling p53 interaction with drug candidates by university students to identify which would be most similar to the human system on the basis of structure and sequence similarity. We were encouraged by the ability of our approach to distinguish between p53 and p63, albeit by an RMSD value of 0.36 Angstroms (Table 2). As a point of comparison, MD simulations typically deviate around 2 Angstroms from x-ray crystallographically determined starting structures due to thermal motions.[41] Therefore the difference in structure between the two distinct proteins is relatively small compared to thermal fluctuations. Keeping these figures in mind as estimations of precision of our methods, we use the results to guide our quantitative interpretation. We note that the precision of Phyre folding method may reflect limited resolution because, as a template based threading method, is will only be as good as the structural data available. The structures we are working with here also do not reflect thermal dispersion.

On the basis of structure, we found that the organisms most similar to humans were the other mammals in the study (Figure 3&4): cows, rats, mice, frogs, and chickens. Interestingly, zebrafish p53 also had a structure surprisingly close to the human version. Rice fish, nematode, and fruit fly had structures more distant (>2 Angstroms) from both human p53 and p63 and therefore would not be preferred choices for model organisms. Of the top organisms from the structural point of view, cows share the closest sequence similarity (Figure 1&2), followed by mouse and rat. This is encouraging because much work on p53 has been carried out on mouse models.[16,72,73] However, the drawback remains that even the small differences between humans and mice may be accountable for unpredictability seen in clinical trials related to the p53-MDM2 peptidomimetics (see Introduction). According to predictions based on sequence and structure from this study, the murine model would only be superseded by the bovine model, which suggests collaboration with veterinarians could be a venue of exploration. For many researchers, however, mice will likely remain the most feasible and accessible model organisms on which to perform pilot animal model testing because the potential gains of the bovine model are small, and the short generation time, ease of manipulating genetics, and modest requirements to house the animals favor the mouse model. Humanized murine p53 knock-in models[35,74,75] may perhaps be the best alternative despite potentially having mismatched interactions with some signaling factors. Our findings suggest that the rat model would likely work equally well as the murine model. Similarities of human p53 to that of chickens, frogs and zebrafish suggest they could also work as models. While the structures of chicken and cow may more closely resemble p63, their sequences more closely resemble p53. Chimeric knock-in models in these organisms could be an alternative, but further studies would need to be done to determine the viability of such mutants, and benefits above the knock-in murine models would need to be established. Zebrafish have been used as a model of choice for developmental pathways,[76] and therefore could potentially be a good choice for those interested in focusing on p63 and its function in development.

Structure based drug discovery uses the technique of building and comparing protein structures.[1] The structures presented are approximations; however, researchers might extrapolate from these finds when determining which model organisms would prove best for drug discovery. In many of the alignments where sequences appear to be very similar, the structures exhibit instances of exposed loops, which add flexibility to the protein that might render it ineffective for drug development.[1] For example, this is the case with structures for frog, chicken, cow, and Japanese rice fish. Others, such as fly and nematode, align very poorly with p53 and would likely make poor model organisms for drug discovery.

The methodologies presented herein reflect our interest in p53, but could be adapted as a means of screening genomes and model organisms for other systems. While no new human homologs were discovered in this project, the method could uncover new knowledge about proteins of unknown function or those from less extensively studied families.[3]

In conclusion, we have successfully carried out a search for homologs of p53 in the human genome, and compared human p53 to orthologs in other organisms. The approach can be applied to other proteins of interest and to other organisms. From the vantage of looking forward to potential drug design projects, we have thus achieved valuable insight which we will apply in the next stages of our drug discovery work flow. We will use MD-MSMs and MD sectors to identify potential regions of interaction to allosterically control p53. Informed by this study, we will design our potential drugs so as to interact selectively with p53 and avoid binding to p63 and p73 to evade off target effects involving these family members. Furthermore, several potential model organisms have been identified for use on the basis of the similarity of their p53 to that of humans on the basis of sequence and predicted structure. The murine model in which much work has been carried out was confirmed as an excellent choice as an accessible and reasonably similar model system amenable to use in the undergraduate laboratory environment. This bioinformatic approach may be broadly applied to inform drug design studies.

Acknowledgments

The authors acknowledge technical assistance from the Vasar College IT department especially Shelly Johnson.

Conflicts of interest

Author declares there is no conflict of interest.

References

1. Lane DP. p53, guardian of the genome. *Nature*. 1992;358:15–16.
2. Pflaum J, Schlosser S, Müller M. P53 family and cellular stress responses in cancer. *Front Oncol*. 2014;4:285.
3. Polyak K, Xia Y, Zweier JL, et al. A model for p53-induced apoptosis. *Nature*. 1997;389(6648):300–305.
4. Chao C, Saito S, Kang J, et al. p53 transcriptional activity is essential for p53-dependent apoptosis following DNA damage. *EMBO J*. 2000;19(18):4967–75.
5. Knights CD, Catania J, Di Giovanni SD, et al. Distinct p53 acetylation cassettes differentially influence gene-expression patterns and cell fate. *J Cell Biol*. 2006;173(4):533–544.
6. Shaw PH. The role of p53 in cell cycle regulation. *Pathol Res Pract*. 1996;192:669–675.

7. Chen F, Wang W, El-Deiry WS. Current strategies to target p53 in cancer. *Biochem Pharmacol*. 2010;80(5):724–730.
8. Speidel D. The role of DNA damage responses in p53 biology. *Arch Toxicol*. 2015;89(4):501–517.
9. Bode AM, Dong Z. Post-Translational Modification of p53 in Tumorigenesis. *Nat Rev Cancer*. 2004;4(10):793–805.
10. Zhang J, Shen L, Sun LQ. The regulation of radiosensitivity by p53 and its acetylation. *Cancer Lett*. 2015;363(2):108–18.
11. Meek DW. Post-translational modification of p53. *Semin Cancer Biol*. 1994;5(3):203–210.
12. Gu B, Zhu WG. Surf the Post-translational Modification Network of p53 Regulation. *Int J Biol Sci*. 2012;8(5):672–684.
13. Olsson A, Manzl C, Strasser. How important are post-translational modifications in p53 for selectivity in target-gene transcription and tumour suppression? *Cell Death Differ*. 2007;14(9):1561–1575.
14. Donehower LA , Harvey M, Slagle BL, et al. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature*. 1992;356(6366):215–221.
15. Jacks T, Remington L, Williams BO, et al. Tumor spectrum analysis in p53-mutant mice. *Curr Biol*. 1994;4(1):1–7.
16. Purdie C, Purdie CA, Harrison DJ, et al. Tumour incidence, spectrum and ploidy in mice with a large deletion in the p53 gene. *Oncogene*. 1994;9(2):603–609.
17. Vogelstein B, Lane D, Levine. Surfing the p53 network. *Nature*. 2000;408:307–310.
18. Okorokov AL, Orlova EV. Structural biology of the p53 tumour suppressor. *Curr Opin Struct Biol*. 2009;19(2):197–202.
19. Okorokov AL, Sherman MB, Plisson C, et al. The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. *EMBO J*. 2006;25(21):5191–5200.
20. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34.
21. Thun MJ, DeLancey JO, Center MM, et al. The global burden of cancer: Priorities for prevention. *Carcinogenesis*. 2009;31(1):100–110.
22. Fitzmaurice C, Allen C, Barber RM, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol*. 2017;3(4):524–548.
23. Linzer DI, Levine J. Characterization Tumor Antigen and Uninfected of a 54K Dalton Cellular SV40 Present in SV40-Transformed Cells. *Cell*. 1979;17(1):43–52.
24. Zakut-Houri R, Bienz-Tadmor B, Givol D, et al. Human p53 cellular tumor antigen: cDNA sequence and expression in COS cells. *EMBO J*. 1985;4(5):1251–1255.
25. Pavletich NP, Chambers K, Pabo CO. The DNA-binding domain of 53 contains the four conserved regions the major mutation hot spots. *Genes Dev*. 1993;7(12):2556–2564.
26. Cho Y, Gorina S, Jeffrey PD, et al. Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science*. 1994;265(5170):346–355.
27. Kitayner M, Rozenberg H, Rohs R, et al. Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat Struct Mol Biol*. 2010;17(4):423–429.
28. Kitayner M, Rozenberg H, Kessler N, et al. Structural Basis of DNA Recognition by p53 Tetramers. *Mol Cell*. 2006;22(6):741–753.
29. Thayer KM, Galganov JC, Stein AJA. Dependence of prevalence of contiguous pathways in proteins on structural complexity. *PLoS One*. 2017;12:e0188616.
30. Thayer KM, Je L. p53 tumor suppressor protein mutants R175H, G245S and R282W: structural prediction and analysis of full length proteins. *MOJ Proteomics Bioinforma*. 2018;7(1):92–100.
31. Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu Rev Biochem*. 2014;83:553–584.
32. Butler JS, Loh SN. Kinetic partitioning during folding of the p53 DNA binding domain. *J Mol Biol*. 2005;350(5):906–918.
33. Loh SN. The missing zinc: p53 misfolding and cancer. *Metallomics*. 2010;2(7):442–449.
34. Vaughan CA, Singh S, Windle B, et al. Gain-of-Function Activity of Mutant p53 in Lung Cancer through Up-Regulation of Receptor Protein Tyrosine Kinase Axl. *Genes Cancer*. 2002;3(7):491–502.
35. Liu DP, Song H, Xu Y. A common gain of function of p53 cancer mutants in inducing genetic instability. *Oncogene*. 2010;29(7):949–956.
36. Bykov VJN, Issaeva N, Shilov A, et al. Restoration of the tumor suppressor function to mutant p53 by a low-molecular-weight compound. *Nat Med*. 2002;8(3):282–288.
37. Lambert JM, Gorzov P, Veprintsev DB, et al. PRIMA-1 Reactivates Mutant p53 by Covalent Binding to the Core Domain. *Cancer Cell*. 2009;15:376–388.
38. Saha T, Kar RK, Sa G. Structural and sequential context of p53: A review of experimental and theoretical evidence. *Prog Biophys Mol Biol*. 2014;117(2):250–263.
39. Acuner Ozbabacan SE, Gursoy A, Keskin O, et al. Conformational ensembles, signal transduction and residue hot spots: application to drug discovery. *Curr Opin Drug Discov Devel*. 2010;13(5):527-537.
40. Beveridge DL, Dixit SB, Barreiro G, et al. Molecular Dynamics Simulations of DNA Curvature and Flexibility: Helix Phasing and Premelting. *Biopolymers*. 2004;73(3):380–403.
41. Rueda M, Ferrer-Costa C, Meyer T, et al. A consensus view of protein dynamics. *Proc Natl Acad Sci U S A*. 2007;104(3):796–801.
42. Prinz JH, Hao Wu, Marco Sarich, et al. Markov models of molecular kinetics: Generation and validation. *J Chem Phys*. 2011;134.
43. Thayer KM, Lakhani B, Beveridge DL. Molecular Dynamics–Markov State Model of Protein Ligand Binding and Allostery in CRIB-PDZ: Conformational Selection and Induced Fit. *J Phys Chem B*. 2017;121:5509–5514.
44. Prinz JH, Keller B, Noe F, et al. Probing molecular kinetics with Markov models: metastable states, transition pathways and spectroscopic observables. *Phys Chem Chem Phys*. 2011;13:16912–16927.
45. Vitalis A, Caflisch A. Efficient Construction of Mesostate Networks from Molecular Dynamics Trajectories. *J Chem Theory Comput*. 2012;8:1008–1120.
46. Lakhani B, Thayer KM, Black E, et al. Spectral Analysis of molecular dynamics simulations in PDZ: MD sectors. *J Biomotec Stroc Dyn*. 2019;8:781–790.
47. Feng Z, Hu G, Ma S, et al. Computational Advances for the Development of Allosteric Modulators and Bitopic Ligands in G Protein-Coupled Receptors. *Aaps J*. 2015;17:1080–1095.

48. Ramsay RR., Popovic, Nikolic MR, et al. A perspective on multi-target drug discovery and design for complex diseases. *Clin Transl Med*. 2018;7(1):3.

49. Weaver RJ, Valentin JP. Today's Challenges to De-Risk and Predict Drug Safety in Human "Mind-the-Gap". *Toxicol Sci*. 2019;167:307–321.

50. Dietmann S, Holm L. Identification of homology in protein structure classification. *Nat Struct Biol*. 2001;8:953–957.

51. Bhullar KS, Lagarón NO, McGowan EM, et al. Kinase-targeted cancer therapies: Progress, challenges and future directions. *Molecular Cancer*. 2018;17(1):48.

52. Wynn ML, Ventura AC, Sepulchre JA, et al. Kinase inhibitors can produce off-target effects and activate linked pathways by retroactivity. *BMC Syst Biol*. 2011;5:156.

53. El Husseini N, Hales BF. The roles of p53 and its family proteins, p63 and p73, in the DNA damage stress response in organogenesis-stage mouse embryos. *Toxicol Sci*. 2018;162:439–449.

54. Billant O, Blondel M, Voisset C. p53, p63 and p73 in the wonderland of S. *cerevisiae*. *Oncotarget*. 2017;8(34):57855–57869.

55. Billant O, Léon A1, Le Guellec S, et al. The dominant-negative interplay between p53, p63 and p73: A family affair. *Oncotarget*. 2016;7:69549–69564.

56. Dötsch V, Bernassola F, Coutandin D, et al. p63 and p73, the ancestors of p53. *Cold Spring Harb Perspect Biol*. 2010;2(9):a004887.

57. Barbieri CE, Pietenpol JA. P63 and epithelial biology. *Experimental Cell Research*. 2006;312(6):695–706.

58. Botchkarev VA, Flores ER. p53/p63/p73 in the epidermis in health and disease. *Cold Spring Harb Perspect Med*. 2014;4(8):a015248.

59. Enthart A, Klein C, Dehner A, et al. Solution structure and binding specificity of the p63 DNA binding domain. *Sci Rep*. 2016;6:26707.

60. Yu X, Narayanan S, Vazquez A, et al. Small molecule compounds targeting the p53 pathway: Are we finally making progress? *Apoptosis*. 2014;19(7):1055–1068.

61. Khoo KH, Hoe KK, Verma CS, et al. Drugging the p53 pathway: understanding the route to clinical efficacy. *Nat Rev Drug Discov*. 2014;13(3):217–236.

62. Goujon M, McWilliam H, Li W, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*. 2010;38:695–699.

63. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res*. 2014;43:D204–D212.

64. McWilliam, H. et al. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res*. 2013;41:597–600.

65. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–410.

66. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.

67. Sievers F, Higgins DG. Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. *Methods in Molecular Biology*. 2014;1079:105–116.

68. Sievers F, Higgins DG. Clustal Omega. *Current Protocols in Bioinformatics*. 2014;48(1):1–16.

69. Kelley LA, Mezulis S, Yates CM, et al. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10:845–858.

70. DeLano WL. *The PyMOL Molecular Graphics System*. Schrodinger LLC: New York; 2010.

71. Strano S, Rossi M, Fontemaggi G, et al. From p63 to p53 across p73. *FEBS Lett*. 2001;490(3):163–170.

72. Garcia PB, Attardi LD. Illuminating p53 function in cancer with genetically engineered mouse models. *Semin Cell Dev Biol*. 2014;27:74–85.

73. Christophorou MA, Ringshausen I, Finch AJ, et al. The pathological response to DNA damage does not contribute to p53-mediated tumour suppression. *Nature*. 2006;443(7108):214–217.

74. Lu X, Liu DP, Xu Y. The Gain-of Function of p53 Cancer Mutant in Promoting Mammary Tumorigenesis. *Oncogene*. 2013;32(23):2900–2906.

75. Liu D, Song H, Xu Y. A common Gain of function of p53 cancer mutants in inducing genetic instability. *Oncogene*. 2010;29(7):949–956.

76. Meyers JR. Zebrafish: Development of a Vertebrate Model Organism. *Curr Protoc Essent Lab Tech*. 2018;16(1):e19.