Combining Ramachandran plot and molecular dynamics simulation for structural-based variant classification: Using TP53 variants as model

Benjamin Tam, Siddharth Sinha, San Ming Wang *

Cancer Centre and Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Macau

ARTICLE INFO

Article history:
Received 29 September 2020
Received in revised form 23 November 2020
Accepted 23 November 2020
Available online 02 December 2020

Keywords:
Ramachandran plot
Molecular Dynamic Simulation
Protein structure
Variant of Uncertain Significance
Pathogenic TP53

ABSTRACT

The wide application of new DNA sequencing technologies is generating vast quantities of genetic variation data at unprecedented speed. Developing methodologies to decode the pathogenicity of the variants is imperatively demanding. We hypothesized that as deleterious variants may function through disturbing structural stability of their affected proteins, information from structural change caused by genetic variants can be used to identify the variants with deleterious effects. In order to measure the structural change for proteins with large size, we designed a method named RP-MDS composed of Ramachandran plot (RP) and Molecular Dynamics Simulation (MDS). Ramachandran plot captures the variant-caused secondary structural change, whereas MDS provides a quantitative measure for the variant-caused globular structural change. We tested the method using variants in TP53 DNA binding domain of 219 residues as the model. In total, RP-MDS identified 23 of 38 (60.5%) TP53 known Pathogenic variants and 17 of 42 (41%) TP53 VUS that caused significant changes of P53 structure. Our study demonstrates that RP-MDS method provides a powerful protein structure-based tool to screen deleterious genetic variants affecting large-size proteins.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Clarification of the pathogenic impact of genetic variants is very challenging, as it relies on the combinational evidence derived from clinic, biostatistics, molecules and experiments [1,2]. Recent application of new DNA sequencing technologies has drastically increased the power for genetic study, resulting in the accumulation of massive genetic variation data at population level. The vast quantity of accumulated variation data has far surpassed the capacity of the current annotation system [3]. The situation is well exemplified by the genetic variants collected from the cancer predisposition gene BRCA1 and BRCA2: 80% of the over 40,000 genetic variants identified from these two genes remain uncharacterized (https://brcaexchange.org/factsheet); of the characterized ones, over 30% of the BRCA1 and 40% of BRCA2 variants are classified as Variant of Uncertain Significance (VUS) due to the lack of functional evidence to determine their pathogenicity (https://www.ncbi.nlm.nih.gov/clinvar/). Therefore, developing new approaches to ease the challenge is urgently demanding.

Protein structure is stabilized by intramolecular interactions of hydrophobic, electrostatic, hydrogen bonding, and Van der Waals interactions. Depending on the position in a protein, a residue modified by a genetic variant can have no, mild, or severe influences on protein structure till inactivation of the affected protein [4–7]. Therefore, we reasoned that protein structure could be used to identify the variants with deleterious effects [8]. Here, we defined deleterious variants as single amino acid substitution that causes overall structural deviation and impedes functionality. We also postulated that the structure-based methodologies should have these essential features: the targeted protein should have known protein structure in order to be used as the reference to judge the structural change caused by the variants, the methods should have high-throughput capacity, therefore, should be computational-based, in order to characterize a large number of variants simultaneously at low cost, and the results should be verifiable by existing well-classified variants to confirm their reliability.

Ramachandran plot is a graphical illustration for visualizing protein backbone energetic position in terms of torsion angles [9]. Ramachandran plot is one of the best theories in protein structure study with minimal discrepancy between experiments and simulations. The concept is based on the rigidity of the N-C peptide bond, in which the torsion angle $\Phi$ and $\Psi$, representing...
X – N – Cx – X and X – Cx – C – O bonded atoms, are restricted by sterically unfavorable structure conformation due to collusion between non-bonded atoms. These physical limits are embedded with conformation information and deciphering the data provides essential insight for the protein structure. The essence of Ramachandran plot remains unchanged since it was developed but its reliability has been significantly improved in recent years [9–11]. Through capturing the distortion caused by genetic variants, we reasoned that the Ramachandran plot can be applied to study the influence of genetic variants on protein structure although this has not been tested in analyzing genetic variants [9,10,12–14]. Molecular Dynamics Simulation (MDS) is a computation-based atomistic simulation method. It analyzes physical movement of atoms and molecules after interacting for a fixed time period, and the trajectories are used to determine macroscopic thermodynamics properties of the targeted molecular structure. MDS has been widely used to analyze protein structural dynamics [15–17], and we also successfully applied MDS to characterize the genetic variants in BRCA1 BRCT domain [18]. Although MDS or Ramachandran plot alone provides an independent measure in protein structure, we reasoned that combination of Ramachandran Plot and MDS, we named it RP-MDS, could enhance the capacity of detecting the impact of genetic variants on protein structure.

**TP53** is a tumor suppressor gene. It plays a key role in maintaining genome stability. Germline mutation in TP53 predisposes to a wide spectrum of early-onset cancers as exemplified by Li-Fraumeni syndrome [19,20]. Three decades’ studies have identified over 1845 germline variants in TP53, 66% are located at the DNA binding domain (DBD) [21–25]. Despite extensive efforts made so far, over 60% of the germline variants in TP53 still remain as VUS due to the lack of functional evidence [24,26]. In this study, we used TP53 germline variants as the model to test the use of RP-MDS method for genetic variant analysis. We observed that RP-MDS was able to identify 23 of the 38 known Pathogenic variants, and 17 of the 42 coding-change VUS, demonstrating that RP-MDS can effectively identify the deleterious genetic variants.

2. Materials and methods

2.1. Source of variants and modelling P53 mutant structure

We selected a total of 88 TP53 variants from ClinVar database, consisting of 38 Pathogenic, 8 Benign/Likely Benign, and 42 VUS variants (Supplementary Table 1). Single crystal DBD structure (native) of P53 were retrieved from the PDB database (PDB ID:2OCJ, at 2.05 Å) and the sequence numbering starts from 94 to 313 [27]. The structure was used as the template to build each P53 mutant structure using the UCSF Chimera software [28] and Modeller package [29]. The process was illustrated in Fig. 1.

2.2. Molecular dynamics simulations

Each mutant P53 DBD and wild-type P53 DBD structure was simulated using GROMACS molecular dynamics software, version 2020 [30]. A forcefield comparative simulation was performed between OPLS/AA and AMBER03, showing a comparable intramolecular number of Hydrogen bond (H bond) and Solvent-Accessible Surface Area (Supplementary Fig. S1). Thus, AMBER03 was chosen to model the protein complex. Zinc ion was described by a non-bonded model, which mimics the 4s4p3 vacant orbitals [31]. The protein structure was situated in the 10 × 10 × 10 nm simulation box, solvated with SPC/E water and neutralized with Cl– ions. The system was optimized with steep descent algorithm before 1 ns equilibration run at 298 K and 1 bar in the NPT ensemble using Berendsen thermostat and barostat. Forty ns production run was simulated for the system at 298 K and 1 bar in the NPT ensemble using V-rescale thermostat and Parrinello-
3.2. MDS measurements

The dynamic effects of the variants on P53 DBD were investigated by MDS. The results showed that the RMSD of Cα atoms value was 0.313 ± 0.015 nm and Root Mean Square Fluctuation (RMSF) of the Cα atoms have a similar fluctuation at the residue Cα atoms 112–124, 178–190, 206–214, 221–232, 239–250, 289–301, which were consistent with the literature [15] which R, which used RMSD, RMSF, and H bond to analyze the impact of Pathogenic, Benign/Likely Benign and VUS variants on P53 structure [15].

Benign/Likely Benign mutants had average numbers of intramolecular H bond and RMSD of the protein backbone 135 ± 4.37 and 0.348 ± 0.045 nm, respectively (Supplementary Figs. S2a and S3a). Pathogenic mutants were separated into 2 regions: the lower bound regions occupied by Y163C, R175H, Y220C, G245D, G245S, R248Q, R273C, R282W, and the higher bound regions occupied by the remaining variants (Supplementary Figs. S2b and S3b) [21], with the average H bond in the lower and higher regions 98 ± 3.88 and 134 ± 4.66, respectively (Supplementary Fig. 2b). Similarly, RMSD showed two regions of displacement. VUS maintained the dynamics alike the native protein and Benign/Likely Benign mutants, with an average 133.8 ± 0.335 nm for H bond and RMSD, except R249S with deviates of 103.9 and 0.668 nm for H bond and RMSD (Supplementary Figs. S2c and S3c). RMSF showed no significant structural deviation in comparison to Benign/Likely Benign (Supplementary Fig. S4c). The majority of variants (inclusive of Pathogenic, Benign/Likely Benign and VUS) fluctuated at 0.348 ± 0.045 nm and 135 ± 0.65 for RMSD and H bond. The results showed that the majority of P53 variants had similar structure dynamics for VUS. Using the cut-offs (H bond < 300, RMSD > 0.3 nm, RMSF > 0.25 nm, Gyration (Rg) > 1.7, SASA > 100 nm³) that successfully differentiated between deleterious and non-deleterious variants in BRCA1 BRCT domain [18], we observed that all except RMSF classified P53 variants into deleterious variants (Supplementary Table S2). In total, MDS (H Bond and RMSD) was only able to identify 8 Pathogenic variants with structural impact. Thus, MDS alone composed of H bond, RMSD, Gyration (Rg) and SASA was insufficient to differentiate a clear boundary between non-deleterious and deleterious VUS, reflecting the limited power of MDS for P53 DBD due likely to its much larger size (198 residues) than BRCA1 BRCT (95 residues) analyzed by MDS in our previous study.

3.3. Ramachandran plot

For each step of simulations, the torsion angle φ and Ψ for each residue were calculated and plotted (Fig. 3a) and converted into relative density graphs through kernel density estimation (Fig. 3b). Native P53 had a high-density peak occupied at P-II region, β sheet region and α-helical region and a gentle peak at the γ (Fig. 3b). The bridge at γ indicated that there was a collectively strong NH$_2$ to O, backbone hydrogen bond. Notably, these deviations were unique to each variant. Analyzing the overall distribution of Pathogenic variants showed that the distribution of structural deviation lay at 28.0%, 38.2% and 33.4% for the 1st, 3rd quartiles, and the mean, respectively. These data were fitted

| Pathogenic |
|-----------|
| 3.452     |
| 0.241     |
| 3.376     |
| 3.526     |
| 1         |
| Can't reject Lognormal |

*K-S: Kolmogorov-Smirnov test (36); A-D: Anderson-Darling test (37).
Fig. 2. Examples of native P53 DBD structure and the variant-affected structure after 40 ns simulation. The well-determined Pathogenic variant R175H unbounded L1, L2 and L3 loops, in contrast to the intact core in the well-determined Benign variant N235S. Red colour shows the variants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 3. Ramachandran plots for native p53. a) Ramachandran scatter plot for native P53. A torsional angle $\Psi$ and $\Phi$ were plotted for all residues. The fluctuation densities were concentrated at $\alpha$ helix (red), $\beta$ strand (blue), $\gamma$ (teal), $\delta$ (green), $\phi$ (purple) and PII strand (orange) regions. There was a minor fluctuation concentration at $\phi'$ regions. The axis is represented at the top right side of the figure. b) 2D Kernel density plot for native P53 transformed from Ramachandran plot. Red to purple colours represent the degree of intensity from high to low. c) Ramachandran plots of native (black), Pathogenic (R175H, G245D, G245S, R248Q, R273C) and Benign (N235S) variants (Red). d) 2D Kernel density plot for the Pathogenic variants. The table shows the structural deviation from Ramachandran plot and the H bond for each variant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
against log-normal distribution and the bottom 95% tiles (3.376) were used to categorize deleterious regions. Deviation by pathogenic variants <95% tiles was considered as independent of conformation alteration, such as these associated with the disruption of protein-DNA or protein-protein interaction but not caused by P53 conformation change, therefore were excluded from further examination. Testing in the 38 Pathogenic variants showed that besides the 8 Pathogenic variants identified by MDS, Ramachandran plot was able to detect additional 15 variants with milder structural deviation that H bond, RMSD, and RMSF were unable to detect (Table 2).

Fig. 3c, d exhibited examples of Ramachandra plots for the well-determined Pathogenic mutations. The Pathogenic variants had shifted from the native position (Fig. 3d) The polyproline-II (P-II) and β region peaks showed a noticeable modification in density implying a structural change. In particular, the torsional density for R175H, G245S and R248Q dissipated at β region and higher density at the P-II region, suggesting a weaken β strand formation and the strengthen P-II formation at the protein backbone. The substituted residue (red) fluctuated at a different torsional angle in comparison to the native residue (black) (Fig. 3c). This torsional angular change illustrated that the substituted residue interacted with a different part of the protein. Fig. 3c and d listed examples of the global torsional density deviation for the strong Pathogenic variants (R175H, G245D, G245S, R248Q, R273C), with the deviation of 43.8%, 38.7%, 49.4%, 43.6% and 46.3%, respectively, consistent with the results in H bond. Thus, Ramachandran plot provided a sensitive measure for the variants with deleterious effects on P53 structure.

Table 2

| Genome position | Change* Amino Acid | H bond (nm) | RMSD Structure deviation (%) | Impact |
|-----------------|--------------------|------------|-----------------------------|--------|
| **Pathogenic**  |                    |            |                             |        |
| Chr7:7578407    | c.523C>G           | p.R175G    | 125.9 0.251                 | 32.98  Deleterious |
| Chr7:7578403    | c.527G>A           | p.C176Y    | 140.3 0.265                 | 35.35  Deleterious |
| Chr7:7578271    | c.578A>C           | p.H193P    | 129.1 0.325                 | 33.40  Deleterious |
| Chr7:7578265    | c.584T>C           | p.I195T    | 135.4 0.257                 | 35.25  Deleterious |
| Chr7:7578211    | c.638G>A           | p.R213Q    | 131.0 0.307                 | 33.89  Deleterious |
| Chr7:7577556    | c.725G>A           | p.C242Y    | 135.8 0.337                 | 30.96  Deleterious |
| Chr7:7577551    | c.730G>A           | p.G244S    | 138.1 0.231                 | 36.52  Deleterious |
| Chr7:7577548    | c.733G>A           | p.G245V    | 131.9 0.320                 | 32.03  Deleterious |
| Chr7:7577538    | c.743G>T           | p.R248L    | 126.5 0.309                 | 33.01  Deleterious |
| Chr7:7577121    | c.817G>T           | p.R273S    | 141.2 0.285                 | 31.54  Deleterious |
| Chr7:7577121    | c.817C>G           | p.R273L    | 130.4 0.380                 | 29.30  Deleterious |
| Chr7:7577120    | c.818G>A           | p.R273G    | 134.3 0.284                 | 42.87  Deleterious |
| Chr7:7577096    | c.842A>G           | p.D281G    | 128.7 0.346                 | 33.30  Deleterious |
| Chr7:7577096    | c.842A>T           | p.D281N    | 133.9 0.295                 | 33.79  Deleterious |
| Chr7:7577084    | c.854A>T           | p.E285V    | 136.4 0.362                 | 38.48  Deleterious |
| Chr7:7578442    | c.488A>G           | p.Y163C    | 90.6 0.639                  | 38.57  Deleterious |
| Chr7:7578406    | c.524G>A           | p.R175H    | 101.6 0.720                 | 43.75  Deleterious |
| Chr7:7578190    | c.659G>A           | p.Y220C    | 84.4 0.649                  | 43.85  Deleterious |
| Chr7:7577550    | c.731G>A           | p.G245D    | 104.8 0.580                 | 38.67  Deleterious |
| Chr7:7577548    | c.733G>T           | p.G245S    | 107.4 0.650                 | 49.41  Deleterious |
| Chr7:7577538    | c.743G>A           | p.R248Q    | 103.5 0.582                 | 43.55  Deleterious |
| Chr7:7577120    | c.818G>C           | p.R273C    | 97.9 0.685                  | 46.29  Deleterious |
| Chr7:7577093    | c.845C>T           | p.R282W    | 87.6 0.775                  | 44.43  Deleterious |

**VUS**

| Chr7:7578503    | c.427G>T           | p.V143L    | 137.4 0.317                 | 36.33  Deleterious |
| Chr7:7578487    | c.443A>C           | p.D148A    | 138.9 0.338                 | 31.84  Deleterious |
| Chr7:7578469    | c.461G>A           | p.G154D    | 137.2 0.329                 | 36.52  Deleterious |
| Chr7:7578461    | c.469G>A           | p.V157I    | 139.7 0.315                 | 33.11  Deleterious |
| Chr7:7578274    | c.575A>G           | p.Q192R    | 137.8 0.407                 | 31.45  Deleterious |
| Chr7:7578196    | c.653T>G           | p.V218G    | 135.3 0.283                 | 36.13  Deleterious |
| Chr7:7577595    | c.686G>A           | p.C229Y    | 136.4 0.322                 | 31.84  Deleterious |
| Chr7:7577521    | c.760A>G           | p.I254V    | 134.1 0.333                 | 32.71  Deleterious |
| Chr7:7577517    | c.764T>A           | p.I255N    | 132.7 0.280                 | 30.37  Deleterious |
| Chr7:7577147    | c.791T>C           | p.L264P    | 138.7 0.279                 | 31.15  Deleterious |
| Chr7:7577124    | c.814G>A           | p.V272M    | 137.2 0.311                 | 29.98  Deleterious |
| Chr7:7577105    | c.833C>A           | p.P278R    | 131.3 0.367                 | 40.53  Deleterious |
| Chr7:7577081    | c.877G>A           | p.G293R    | 137.9 0.333                 | 32.52  Deleterious |
| Chr7:7577081    | c.877G>T           | p.G293W    | 131.1 0.362                 | 29.59  Deleterious |
| Chr7:7577052    | c.886C>T           | p.H296Y    | 132.2 0.360                 | 30.37  Deleterious |
| Chr7:7577033    | c.905G>A           | p.G302E    | 134.9 0.348                 | 28.79  Deleterious |
| Chr7:7577534    | c.747G>C           | p.R249S    | 103.9 0.688                 | 43.85  Deleterious |

*The ones identified by H bond and RMSD highlighted in grey, by RP- MDS but not by H bond and RMSD in bold.
3.4. Searching VUS with deleterious effects

Using the parameters determined from known Benign/Likely Benign and Pathogenic variants, we tested RP-MDS in 42 coding-varying VUS variants to identify these with deleterious impact. Supplementary Fig. S5 illustrated the Ramachandran density plot for each of the 42 VUS variants. Under the deleterious >3.376 generated from known Pathogenic variants, Ramachandran plot identified 17 of 42 (41%) VUS (V143L, D148A, G154D, V157I, Q192R, V218G, C229Y, R249S, I254V, I255N, L264P, V272M, P278R, G293R, G293W, H296Y, G302E) causing apparent structural deviation.

Benign and Pathogenic variants, we tested RP-MDS in 42 coding-varying VUS variants and 42 VUS with missense 3D and SuSPect programs [38,39]. Missense 3D was able to detect 13 pathogenic variants (R175G, R175H, C176Y, H193P, R213Q, V220C, C238R, C242Y, G245D, G245S, G245V, L265P, R273P) and 8 VUS (C176W, G187D, S215R, V218G, L252P, I255N, P278R, G279R) with potential structural damage and all benign variants with no structural damage. SuSPect was able to detect all pathogenic variants, however, it failed to differentiate Benign/Likely Benign variants and classified all VUS as disease-associated variants. Here, our RP-MDS method was able to significantly increase the detection of deleterious variants.

4. Discussion

The conventional approaches for determining the pathogenicity of genetic variants rely on the evidence from experiment-based functional assays, bio-statistics-based methods, evolution conservation-based algorithms, and clinical data. Previously, we developed an protein structure-based MDS approach to identify the deleterious variants and successfully applied it in classifying the variants in smaller-size BRCA1 BRCT domain. In this study, we further developed the MDS-based approach into R-MDS approach in order to identify the deleterious variants in large-size functional domains. By using P53 DNA binding domain as the model, our study showed that RP-MDS fulfills our expectation. Data from our previous and current studies demonstrate that protein structure can be used for identifying the deleterious variants. This is particularly meaningful in current status of genetic variant annotation, considering the fact that protein structure for many disease genes have been well determined, lack of functional evidence for vast quantity of unclassified variants such as VUS, and advanced computational power allowing large-scale performance.

MDS is proficient to identify the deleterious variants with strong detrimental nature as demonstrated by the lower H bond and higher RMSD for the eight TP53 Pathogenic variants (Supplementary Fig. 2 and 3). For the Pathogenic variants with lesser severity, however, MDS showed inadequate sensitivity as their fluctuations are closer to the native P53. Similar situation was present in VUS that 41 of the 42 VUS showed no significant structural deviation (Supplementary Fig. 2c–4c). Our previous study demonstrated that MDS was sensitive in classifying the VUS in the BRCA1 BRCT domain [18]. The discrepancy is likely due to the size-difference between P53 DBD (213 residues) and BRCA1 BRCT domain (95 residues) that MDS has limited power to differentiate the structural changes in a larger protein structure. By using the information of backbone torsional angle from Ramachandran plot, this limitation is largely overcome allowing to expose the differences hid within the larger protein structure. This is well reflected by the increasingly identified TP53 structural-changing Pathogenic variants and VUS by RP-MDS.

The Ramachandran density plot demonstrated a dissimilar structural affinity for Pathogenic variants with the deleterious structural attitude in comparison to native P53. Fig. 3c showed the local residue torsional angle difference in comparison to the native structure and Fig. 3d showed that the Ramachandran density plot of its respective protein. P-II, α, β and δ regions (which are predominantly populated by folded proteins) were notably altered by the residue substitution. In this case, a dissimilar torsional angle in respect to the native residue inferred that the variant residue had positionally interacted with another part of the protein. Conversely, the substituted residue that fluctuated at a similar torsional angle had different interaction characteristics and consequently affected the global structure. Thus, Ramachandran plot can effectively detect the deleterious characteristics of deleterious variants. Although Benign/Likely Benign variants had no noticeable disordered fluctuation, this information cannot be directly applied to assign the VUS variants with insignificant impact on the structure as non-deleterious one, as possibilities exist that these variants may still have deleterious effects through non-structural factors, e.g. post-translational modification, which may or may not disturb protein structure.

While many genetic variants can cause “loss of function” in the affected protein, it is often the case that many genetic variants can also cause “gain of function” consequence. The gain of function is often present in P53. For example, P53 R175, R248, and R273 are typical gain of function mutations, contributing to carcinogenesis [40–42], and P53 R249S is a typical conditional gain of function mutation [43,44]. While the current structure-based approach used in our study can not distinguish between loss of function and gain of function mutations, it will be interesting to further explore the means to obtain the ability. These could include the selection of specific parameters, the use of known gain of function and loss of function p53 mutation as the control, or combination of structural changes with experimental methodologies.

Our study has limitations by the lack of Benign/Likely Benign variants in TP53 as the control. Comparing to a large number of Pathogenic and VUS variants, there were only 8 Benign/Likely Benign variants in ClinVar and International Agency for Research on Cancer (IARC) TP53 databases. This could relate with the nature of P53 that most of the variants could be deleterious because of its low thermodynamic and kinetic stability. In our analysis, we disassociated Zinc ion from P53. Polarization could be an attributing factor for the zinc ion dissociation, but the polarized force field is yet to be determined for P53. We cautiously disregarded to use the bonded model for the simulation as it artificially pull the L2 and L3 loops together. As such, it might become an artificial structure and the stability observed might not be real. Study showed that the zinc ion contributes to (ΔG ≈ 4.6 kJ mol⁻¹) stability at 10 °C [45]. We reasoned that thermal contribution and entropy at 25 °C could easily delocalise zinc ion. Our reasoning is supported by the observation that “a significant fraction of p53 may exist in the zinc-free state under physiological conditions” [46], and zinc-free (apo) P53 is “both thermodynamically stable and kinetically accessible. Therefore, to understand how variants shift the balance of p53 conformational states, it is necessary to generate a complete thermodynamic model that includes parameters for the zinc-free form”. Thus we reasoned that choosing zinc-free structure can reflect better the impact of altered residues on the structure under physiological conditions. Lastly, our study excluded the interference by external organic molecule, i.e. DNA and other proteins, to avoid adverse impact on the speed of the simulation and increased complexity of the system.

In summary, the RP-MDS method provides a structural-based means to effectively identify deleterious variants in larger-size proteins. Its computational nature allows large-scale application for characterizing genetic variants in disease genes.
Funding

This work was funded by Macau Science and Technology Development Fund (085/2017/A2), Macau Science and Technology Development – Ministry of Science and Technology of People’s Republic of China fund (007/2019/AM), grants from the University of Macau (SRG2017-00097-FHS, MYRG2019-0018-FHS), the Faculty of Health Sciences, University of Macau (Start up fund, FHSIG/SW/0007/2020P, FHS Innovation grant) (SMW). BT is the recipient of University of Macau Postdoctoral Fellowship Class A of the Macao Talent Program.

CRediT authorship contribution statement

Benjamin Tam: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. Siddharth Sinha: Data curation, Resources, Software, Methodology, Writing - original draft. San Ming Wang: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Validation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was performed at the high-performance computing cluster supported by Information and Communication Technology Office of the University of Macau.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cbsj.2020.11.041.

References

[1] He KY, Ge D, He MM. Big data analytics for genomic medicine. Int J Mol Sci 2017;18(2):412.
[2] Dankar FK, Pitsyn A, Dankar SK. The development of large-scale de-identified biomedical databases in the age of genomics-principles and challenges. Hum Genomics 2018;12(1):19–19.
[3] Federici G, Soddu S. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. J Exp Clin Cancer Res 2020;39(1).
[4] Cho Y, Gorina S, Jeffrey P, et al. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 1994;265(5170):346–55.
[5] Joerger AC, Ang HC, Fersht AR. Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. Proc Natl Acad Sci USA 2006;103(41):15056–61.
[6] Eldar A, Rozenberg H, Diskin-Posner Y, et al. Structural studies of p53 inactivation by DNA-contact mutations and its rescue by suppressor mutations via alternative protein-DNA interactions. Nucleic Acids Res 2013;41(18):8748–59.
[7] Joerger AC, Fersht AR. The p53 pathway: origins, inactivation in cancer, and emerging therapeutic approaches. Annu Rev Biochem 2010;8(1):375–404.
[8] Kumar S, Clarke D, Gerstein MB. Leveraging protein dynamics to identify cancer mutational hotspots using 3D structures. Proc Natl Acad Sci USA 2019;116(38):18962–70.
[9] Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol 1963;7(1):95–9.
[10] Vega MC, Serrano L, Martinez JC. Thermodynamic and structural characterization of Asn and Ala residues in the disallowed II region of the Ramachandran plot. Protein Sci 2000;9(12):2322–8.
[11] Wishart DS, Nip AM. Protein chemical shift analysis: a practical guide. Biochim Biophys Acta 1998;76(2-3):153–63.
[12] Gromiha MM, Obohatake M, Kono H, Uedaia H, Sarai A. Importance of mutant position in Ramachandran plot for predicting protein stability of surface mutations. Biopolymers 2002;64(4):210–20.
[13] Hollingsworth SA, Karpus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomol Concepts 2010;1(3–4):271–83.