

Structural analysis of *MIAP* variants associated with severely impaired spermatogenesis causing male infertility

Umut Gerlevik^{1,2}, Mahmut Cerkez Ergoren^{3,4},
Osman Uğur Sezerman^{1,5} and Sehime Gulsun Temel^{6,7,8}

¹ Department of Biostatistics and Bioinformatics, Institute of Health Sciences, Acibadem Mehmet Ali Aydınlar University, Istanbul, Turkey

² Department of Biochemistry, University of Oxford, Oxford, United Kingdom

³ Department of Medical Genetics, Faculty of Medicine, Near East University, Nicosia, Cyprus

⁴ DESAM Institute, Near East University, Nicosia, Cyprus

⁵ Department of Biostatistics and Medical Informatics, School of Medicine, Acibadem Mehmet Ali Aydınlar University, Istanbul, Turkey

⁶ Department of Medical Genetics, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey

⁷ Department of Histology & Embryology, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey

⁸ Department of Translational Medicine, Health Sciences Institute, Bursa Uludag University, Bursa, Turkey

ABSTRACT

Background: Impaired meiosis can result in absence of sperm in the seminal fluid. This condition, namely non-obstructive azoospermia (NOA), is one of the reasons of male infertility. Despite the low number of studies on meiosis 1-associated protein (*MIAP*) in the literature, *MIAP* is known to be crucial for spermatogenesis. Recently, seven variants (five missense, one frameshift, one splice-site) have been reported in the *MIAP* gene as associated with NOA, cryptozoospermia and oligozoospermia in two separate studies. However, all missense variants were evaluated as variant of uncertain significance by these studies. Therefore, we aimed to analyze their structural impacts on the *MIAP* protein that could lead to NOA.

Methods: We firstly performed an evolutionary conservation analysis for the variant positions. Afterwards, a comprehensive molecular modelling study was performed for the *MIAP* structure. By utilizing this model, protein dynamics were sampled for the wild-type and variants by performing molecular dynamics (MD) simulations.

Results: All variant positions are highly conserved, indicating that they are potentially important for function. In MD simulations, none of the variants led to a general misfolding or loss of stability in the protein structure, but they did cause severe modifications in the conformational dynamics of *MIAP*, particularly through changes in local interactions affecting flexibility, hinge and secondary structure.

Conclusions: Due to critical perturbations in protein dynamics, we propose that these variants may cause NOA by affecting important interactions regulating meiosis, particularly in wild-type *MIAP* deficiency since the variants are reported to be homozygous or bi-allelic in the infertile individuals. Our results provided reasonable insights about the *MIAP* structure and the effects of the variants to the structure and dynamics, which should be further investigated by experimental studies to validate.

Submitted 12 March 2021

Accepted 25 January 2022

Published 21 March 2022

Corresponding authors

Osman Uğur Sezerman,
ugur.sezerman@acibadem.edu.tr

Sehime Gulsun Temel,
sehime@uludag.edu.tr

Academic editor

Vladimir Uversky

Additional Information and
Declarations can be found on
page 18

DOI 10.7717/peerj.12947

© Copyright

2022 Gerlevik et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Computational Biology, Molecular Biology, Andrology, Medical Genetics

Keywords Molecular modelling, Molecular dynamics simulations, Male infertility, Meiosis 1-associated protein (MIAP), Non-obstructive azoospermia (NOA), Cryptozoospermia, Variant effect on protein structure

INTRODUCTION

Male infertility is a widespread problem in human reproduction, affecting many people in the world (Agarwal et al., 2015; Olszewska et al., 2020). There are numerous conditions that cause male infertility, and two of them are azoospermia and cryptozoospermia (World Health Organization, 2010). Azoospermia is the condition of lacking spermatozoa in seminal fluid. Moreover, there are two main types of azoospermia: obstructive (normal spermatogenesis) and non-obstructive (damaged spermatogenesis) (Gudeloglu & Parekattil, 2013; Hwang et al., 2018). In other words, there is no problem in sperm formation in obstructive azoospermia but there are other problems after the formation. Contrarily, the main problem in non-obstructive azoospermia (NOA) is during spermatogenesis. On the other hand, cryptozoospermia is the condition where there are no spermatozoa in fresh preparations but very few spermatozoa in the pellet of the centrifuged semen sample (Zhu et al., 2016; Karabulut et al., 2018). Besides the clinical similarities between NOA and cryptozoospermia, they may also have common genetic basis based on problems of spermatogenesis (Arango et al., 2006, 2013; World Health Organization, 2010; Wyrwoll et al., 2020). Distorted spermatogenesis is mostly relied on a genetic origin, but the affected individuals are routinely screened for chromosomal abnormalities and Y chromosome azoospermia factor microdeletions (Lee et al., 2011). These approaches cover only up to 20% of the cases affected by azoospermia alone (Tüttelmann, Ruckert & Röpke, 2018). Recent studies indicated that human male germ cell arrest is caused by monogenic variations (Gershoni et al., 2017) such as testis-expressed gene 11 (TEX11, MIM 300311) gene on the X chromosome (Yatsenko et al., 2015). Furthermore, bi-allelic genetic variants on meiosis 1-associated protein (MIAP, MIM 619098) gene were recently associated with both of NOA and cryptozoospermia in different patients (Wyrwoll et al., 2020). Thus, in these cases, NOA and cryptozoospermia could be closely associated in terms of the underlying mechanism (Wyrwoll et al., 2020).

MIAP is a gene expressed in germ cells in both males and females, and it encodes a 530 amino acid long protein (MIAP). Arango et al. (2006) found that this gene is expressed at the beginning of meiosis in female germ cells and at the final stages of spermatogenesis in male germ cells. Therefore, they suggest that the evolutionarily well-conserved MIAP protein may have a role in gametogenesis. Furthermore, Arango et al. (2013) showed that MIAP-deficient male mice have severe oligozoospermia and infertility with many testicular defects. Therefore, they suggest that mutations in MIAP might be reason for nonobstructive oligozoospermia in men. In addition, they localized MIAP as a cytosolic protein. Unfortunately, no further knowledge is available in the literature due to the lack of focus on this protein, which has the potential to be highly critical for meiosis and reproduction.

Table 1 Bi-allelic *M1AP* variants reported in infertile males and predictions for their functional effects by Wyrwoll et al. (2020).

Individual	Bi-allelic <i>M1AP</i> variants		Variant impact predictions (PolyPhen-2/SIFT/MutationTaster)		Diagnosis
	Allele 1	Allele 2	Allele 1	Allele 2	
P86	p.Ser50Pro	p.Leu430Pro	Tolerated/Polymorphism/Damaging	Damaging/Damaging/Damaging	Azoospermia
M1943	p.Arg266Gln	p.Trp226LeufsTer4	Damaging/Damaging/Damaging	–	Cryptozoospermia/ azoospermia [†]
Y126	p.Gly317Arg	p.Trp226LeufsTer4	Damaging/Damaging/Damaging	–	Azoospermia
T1024	p.Pro389Leu	p.Pro389Leu	Damaging/Damaging/Damaging	Damaging/Damaging/Damaging	Azoospermia

Notes:

The missense variants in the table (*i.e.*, p.Ser50Pro, p.Arg266Gln, p.Gly317Arg, p.Pro389Leu and p.Leu430Pro) were examined in this study.

All data in the table were reported in Wyrwoll et al. (2020). Although there were more individuals in the reference study, only bi-allelic variants with at least one missense change were included in this study since our focus is on the impacts of missense variants on the *M1AP* structure.

[†] Semen contained none or below 10 spermatozoa/sample on repeated analyses.

To the best of our knowledge, 1 frameshift, 1 splice site and 5 missense variants found in *M1AP* have so far been associated with male infertility conditions (*i.e.*, NOA, cryptozoospermia and oligozoospermia) in previous studies (Tu et al., 2020; Wyrwoll et al., 2020). As shown in Table 1, all these variants except the homozygous splice site variant were found in different bi-allelic combinations in the individuals rather than single variants (Wyrwoll et al., 2020). In addition, a fertile person was found despite having the heterozygous p.Pro389Leu variant (Wyrwoll et al., 2020). On the other hand, since the variants are on different alleles (compound heterozygous), any two of them are not expected to be in a protein product at the same time although they could be found in the same cell as separate proteins. Even though the algorithms predict the missense variants as mostly damaging (Table 1), Wyrwoll et al. concluded these variants individually as variants of uncertain significance based on a classification according to the ACMG-AMP Guidelines (Richards et al., 2015).

Although there is almost no information about *M1AP* structure and function in the literature, the effects of the missense variants might be revealed by molecular modeling and molecular dynamics (MD) simulations. These approaches allow to assess the impacts on protein structure and dynamics. Owing to its robustness, computational modeling is well-suited for studying missense variants associated with genetic diseases (Zimmermann et al., 2017). In this study, we investigated the structural mechanisms underlying NOA and cryptozoospermia by analyzing the impacts of five missense variants on the structure and dynamics of *M1AP* via MD simulations.

METHODS

A comprehensive whole-exome sequencing and variant prioritization study (Wyrwoll et al., 2020) has previously shown that some biallelic missense variants in the meiosis 1-associated protein (*M1AP*) gene were associated with male infertility via non-obstructive azoospermia (NOA) or cryptozoospermia. Each missense variant shown in Table 1 was analyzed individually in this study for their impacts on *M1AP* structure and dynamics.

Evolutionary conservations

It is known that structurally and functionally important amino acids are mostly conserved during evolution (Ashkenazy et al., 2016). To reveal the importance of variant regions, we aligned the UniProt (The UniProt Consortium, 2019) protein sequences of M1AP orthologs from human (*Homo sapiens*, UniProt ID: Q8TC57), chimpanzee (*Pan troglodytes*, UniProt ID: H2R3U8), cat (*Felis catus*, UniProt ID: A0A337RXR8), wild boar (*Sus scrofa*, UniProt ID: A0A5G2QVF2), bovine (*Bos taurus*, UniProt ID: E1BF42) and house mouse (*Mus musculus*, UniProt ID: Q9Z0E1) by using T-Coffee with its default parameters (Notredame, Higgins & Heringa, 2000; Di Tommaso et al., 2011). The alignment was visualized via ESPript (v3.0) (Robert & Gouet, 2014).

Molecular modeling and validation

Sequence of the canonical isoform of M1AP protein (UniProt ID: Q8TC57-1) was used to build models. GalaxyWeb (Ko et al., 2012), I-TASSER (Roy, Kucukural & Zhang, 2010), Phyre2 (Kelley et al., 2015), PRIMO (Hatherley et al., 2016), RaptorX (Peng & Xu, 2011) and Robetta (Kim, Chivian & Baker, 2004) were used to model different structures with their default parameters. Regular comparative algorithm (we called Robetta model) and domain prediction algorithm (we called Robetta-domain model), which are part of Robetta webserver (<https://robetta.bakerlab.org>), were applied separately (Kim, Chivian & Baker, 2004). Robetta algorithms also applied an *ab initio* prediction for part of the structure and compared their reliability with its template-based predictions.

A list of the templates automatically selected by the modeling algorithms, along with their experimental methodologies, resolution values, and Ramachandran outliers are available in Table S1. Since all structure predictors do not list the identity, similarity, and coverage details, we aligned the templates with M1AP via NCBI's Protein BLAST (<https://blast.ncbi.nlm.nih.gov/>) by applying default parameters except the removed E-value threshold (Table S2).

Quality of the structures were assessed by Verify3D (Eisenberg, Lüthy & Bowie, 1997), ERRAT (Colovos & Yeates, 1993), Prove (Pontius, Richelle & Wodak, 1996), PROCHECK (Laskowski et al., 1993) and WHATCHECK (Hoofst et al., 1996) algorithms on SAVES v6.0 webserver (<https://saves.mbi.ucla.edu/>). Using these assessments, we firstly selected the best model from each algorithm and then compared them with each other. Since the qualities of the two best models were so close, they were further evaluated for folding stability via molecular dynamics (MD) simulations. Of note, before the quality assessments, the models subjected to MD simulations (see "further models" and "variant models" in Table 2) were processed with "RepairPDB" command of FoldX (v5.0) (Delgado et al., 2019; van Durme et al., 2011), which converts the saved coordinates to the PDB format of SAVES and applies a basic minimization procedure.

During the revision process of this article, a revolution in protein structure prediction was carried out in the Critical Assessment of protein Structure Prediction (CASP) (Pereira et al., 2021) by Google DeepMind. They developed a deep learning-based algorithm called AlphaFold2 (AF2) which makes *de novo* structure prediction with high accuracy (Jumper et al., 2021; Thornton, Laskowski & Borkakoti, 2021; Varadi et al., 2021).

Table 2 Quality assessments of M1AP models from five different algorithms. Initial and comparison models were not assessed after being processed with molecular dynamics (MD) simulations whereas further and variant models were assessed after being subjected to MD simulations for a time period specified at the end of the model name (e.g., 450ns).

Model	VERIFY3D ²	ERRAT ³	PROVE ⁴	PROCHECK ⁵					WHATCHECK ⁶			Overall report
	Aa ¹ % with score ≥ 0.2	AOQF ¹ %	BOPA ¹ %	Aa% in MFR ¹	Aa% in AAR ¹	Aa% in GAR ¹	Aa% in DR ¹	G-Factor	# of SCP ¹ with pass	# of SCP with warning	# of SCP with error	
Initial Models												
GalaxyWeb	69.25	74.07	8.1	89.1	8.4	1.7	0.8	-0.05	26	15	4	Pass
I-TASSER	47.36	90.47	7.5	60.8	31.7	4.7	2.8	-0.77	19	18	10	Pass
Phyre2	56.98	22.99	10.4	76.0	14.6	5.1	4.3	-0.60	27	12	7	Pass
PRIMO	39.66	1.35	33.4	58.4	25.1	9.2	7.3	-1.41	23	15	8	Pass
RaptorX	41.32	43.72	9.1	78.6	17.8	1.9	1.7	-0.42	23	16	7	Pass
Robetta	76.23	95.22	4.8	83.9	15.6	0.3	0.2	0.11	27	14	6	Pass
Robetta-domain	89.06	88.24	4.4	85.2	13.9	0.5	0.4	0.20	27	15	5	Pass
Comparison Model												
AlphaFold2	89.55	63.21	4.4	80.9	14.3	3.5	1.3	-0.14	22	17	6	Pass
Further Models												
GalaxyWeb_100ns	88.30	78.35	7.1	84.6	13.7	1.1	0.6	-1.11	17	21	8	Pass
Robetta-domain_100ns	83.89	87.53	6.6	87.1	10.7	1.3	0.9	-1.11	17	21	8	Pass
GalaxyWeb_450ns	80.75	85.38	5.9	89.1	9.4	1.3	0.2	-1.05	17	22	7	Pass
Variant Models												
GalaxyWeb_Ser50Pro_450ns	79.62	87.45	8.5	88.0	9.2	2.4	0.4	-1.10	16	22	8	Pass
GalaxyWeb_Arg266Gln_450ns	86.60	77.26	6.0	84.6	12.6	1.9	0.9	-1.14	16	20	10	Pass
GalaxyWeb_Gly317Arg_450ns	84.91	74.90	7.1	85.5	12.4	1.1	1.0	-1.09	18	20	8	Pass
GalaxyWeb_Pro389Leu_450ns	77.36	70.50	8.2	84.2	13.0	1.7	1.1	-1.11	16	21	9	Pass
GalaxyWeb_Leu430Pro_450ns	79.06	78.46	6.3	85.6	12.0	1.7	0.7	-1.09	18	20	8	Pass

Notes:

¹ Aa, amino acid; AOQF, average overall quality factor; BOPA, buried outlier protein atoms; MFR, the most favored regions; AAR, additional allowed regions; GAR, generously allowed regions; DR, disallowed regions; SCP, stereochemical properties.

² In VERIFY3D, at least 80% of aa in a protein should have a score ≥ 0.2 in the 3D-1D profile for a good quality.

³ In ERRAT, AOQF $\geq 95\%$ indicates a good quality, AOQF around 91% indicates an average quality, and lower AOQF indicates a bad quality.

⁴ In PROVE, BOPA $\leq 1\%$ indicates a good quality, BOPA between 1% and 5% indicates an average quality, and BOPA $> 5\%$ indicates a bad quality.

⁵ In the Ramachandran plot analysis of PROCHECK, model has a good quality if overall G-factor is > -0.5 and $\geq 90\%$ of aa in protein are in the most favored regions.

⁶ In WHATCHECK, SCP with pass refers to a good quality, SCP with warning indicates an average quality, and SCP with error refers to a bad quality. Of note, total number of SCP might vary from structure to structure.

Because of its success, we compared our model with the AF2 model as an additional validation. AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk/entry/Q8TC57>) (Varadi et al., 2021) was used to obtain AF2's M1AP model.

We used PyMOL's iterative superposition algorithm (Schrödinger & DeLano, 2020) for the structural alignment of the AF2 model with our model (see Fig. S1) and to compare the models generated by the same homology modeling algorithm (see Fig. S2).

Visual Molecular Dynamics (VMD v1.9.3) (Humphrey, Dalke & Schulten, 1996) was used to model the structures with the variants via the VMD Mutator Plugin.

System preparation

PDB2PQR (Dolinsky et al., 2004) was used to calculate the protonation states of amino acids at pH 7. Moreover, Glu43, Glu436 and Asp118 in the GalaxyWeb models (i.e.,

wild-type and all mutants) were protonated whereas His185, His238 and His375 were protonated in the Robetta-domain model. Models were solvated in a TIP3P water (Jorgensen *et al.*, 1983) box. The systems were neutralized and ionized by adding 0.15 M KCl. The dimensions of the system were approximately $112.6 \times 108.8 \times 125.2 \text{ \AA}^3$ in the xyz axes and the system contained $\sim 146,000$ atoms. VMD was used for the entire preparation procedure.

Simulation setup

NAMD 2.13-multicore-CUDA (Phillips *et al.*, 2005) was used for MD simulations with the CHARMM36m all-atom force field (Huang *et al.*, 2016). Simulations were performed under NPT ensemble with Langevin thermostat and Nosé-Hoover Langevin piston at 310 K and 1 atm (Feller *et al.*, 1995; Davidchack, Handel & Tretyakov, 2009). The damping coefficient (γ) of the Langevin thermostat was 1/ps. The oscillation period of the piston was around 100 fs whereas the oscillation decay time was 50 fs. A cutoff distance of 12 \AA was used for the van der Waals interactions. The switching function starts at 10 \AA and reaches zero at 14 \AA . Long-range Coulomb interactions were calculated with particle-mesh Ewald (Kolafa & Perram, 1992). A 2 fs integration time-step was used. The minimization and equilibration procedure was performed as follows: (1) 5,000 steps conjugate gradient algorithm and 1 ns equilibration of the solution by fixing the protein; (2) 5,000 steps of conjugate gradient minimization and 1 ns equilibration of the whole system by releasing all constraints except the ShakeH algorithm. This procedure provided sufficient relaxation for the system (Fig. S3). ShakeH of NAMD was applied during the entire simulation procedure to keep hydrogens constrained. Production simulations (Repeat 1) for each system, including wild-type and the variants, were performed for 500 ns. Production simulations were re-performed for 500 ns (Repeat 2) with different random number generator seeds to validate the reproducibility of the results and/or to obtain different possible conformations independent of the original simulations. In this way, the total production simulation time for each system was 1,000 ns. All simulation results are available at Zenodo (<https://doi.org/10.5281/zenodo.5811977>).

Structural and dynamical analyses

The “Generate” function of PDBsum (Laskowski, 2009) was used to predict the 2-D structure topology, clefts and tunnels of the model.

Alterations in a protein’s compactness, folding, solvent accessibility, stability, flexibility, hinges, local interactions (*i.e.*, hydrogen bonds (H-bonds) and salt bridges within 10 \AA of variant sites) and essential motions might influence functions, interactions and regulations of that protein. Therefore, we used in-house tcl scripts (available at https://github.com/ugerlevik/M1AP_analysis) and built-in VMD plugins to calculate backbone root-mean-square deviation (RMSD, related to folding), average C_α root-mean-square fluctuation (RMSF, indicates the flexibility), radius of gyration (R_g , inversely associated with compactness), solvent accessible surface area (SASA) and distances.

FoldX (v5.0) (Delgado *et al.*, 2019) was used for stability analysis. ProDy (Bakan, Meireles & Bahar, 2011) was used for principal component analysis (PCA), which refers to

conformational dynamics or essential motions related to the free energy landscape observed in the MD simulations. ProDy was also used for Gaussian network modeling to predict the hinge residues that play a key role in the essential motions of proteins.

Secondary structure analysis and related visualizations were performed by using the ggstride package (<https://github.com/ugerlevik/ggstride/>), which was developed for the first time to be used in this study. ggstride uses STRIDE (Frishman & Argos, 1995) for secondary structure assignment, bio3d (Grant *et al.*, 2006) for structure and trajectory evaluations, and ggplot2 (Wickham, 2016) and ggpubr (Kassambara, 2020) for visualizations. This whole procedure was performed in R (v4.0.3) (R Development Core Team, 2020).

In PCA and secondary structure analysis, repeat 1 and 2 trajectories were merged while obtaining the observed conformations and secondary structure percentages. Only the last 75 ns of each trajectory was used in RMSF, PCA, secondary structure, and hinge analyzes as these parts had the most stable RMSD levels (Fig. 1) indicating that the structures had reached the energetic minimum.

Analysis results were visualized using the ggplot2, ggpubr and gg.gap (<https://github.com/ChrisLou-bioinfo/gg.gap>) packages in R, and the matplotlib package (Hunter, 2007) in python (v3.7.6) (van Rossum & Drake, 2009). Structural visualizations were performed by using VMD.

RESULTS

Evolutionary conservation

We observed that all variant sites except Ser50 were conserved among all orthologs examined (Fig. S4). Ser50 was also highly preserved by changing to a Cys only in the cat (*Felis catus*, UniProt ID: A0A337RXR8). This suggests that all these variants may have significant effects on the structure and function of meiosis 1-associated protein (M1AP). Therefore, all of them were further investigated using molecular dynamics (MD) simulations.

Molecular modeling and validation

Twenty-six models for the M1AP structure were built *via* the following algorithms with automatically selected templates: GalaxyWeb (five models), I-TASSER (five models), Phyre2 (one model), PRIMO (four models), RaptorX (one model), Robetta (five models) and Robetta-domain (five models). Although the properties regarding to the templates such as identity and coverage (Tables S1 and S2) indicate that the M1AP structure is difficult to model, it was possible to model the entire structure as the templates covered different parts of the M1AP and as utilizing the *ab initio* modeling. However, we required a comprehensive validation procedure as follows to proceed with the simulations reliably.

Firstly, we selected the best model from each algorithm by comparing their quality assessment scores (Table S3) to simplify benchmarking the models of the different algorithms. Remarkably, the models generated by the same algorithm were mostly similar to each other, except for the Robetta and Robetta domain models, which were more diverse than the models of the other tools (Fig. S2). As a result, model 1 from GalaxyWeb,

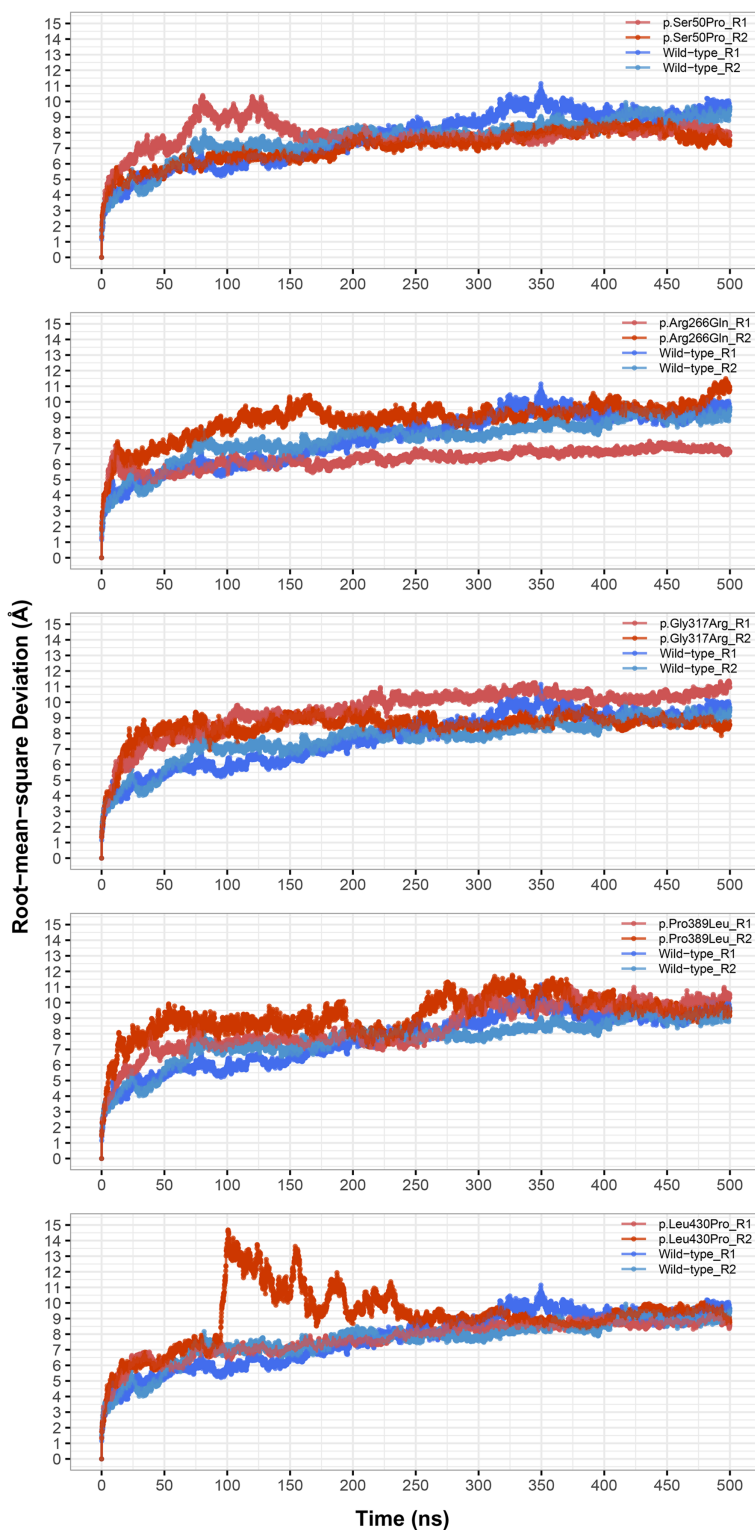


Figure 1 Root-mean-square deviation (RMSD) analysis. Backbone RMSD comparison between molecular dynamics simulations of wild-type and changed M1AP structures. R1 and R2 indicates “Repeat 1” and “Repeat 2” trajectories, respectively. [Full-size !\[\]\(5fd6ef84f97f42d7f8b34275f1b65312_img.jpg\) DOI: 10.7717/peerj.12947/fig-1](https://doi.org/10.7717/peerj.12947/fig-1)

I-TASSER, PRIMO and Robetta-domain, and model 5 from Robetta came forward for further benchmarking. Furthermore, among the single models of RaptorX and Phyre2 and the selected models of other tools, the top three models were from GalaxyWeb, Robetta and Robetta-domain (see “Initial Models” in Table 2). Since the GalaxyWeb model had the highest Ramachandran plot (PROCHECK) scores and the Robetta-domain model had slightly better scores than the Robetta model, we retained the GalaxyWeb and Robetta-domain models to continue with them as the two best models.

Instead of basic and static algorithms, we compared the two best models based on the observation of their folding stability over 100 ns of molecular dynamics (MD) simulations. We used the backbone root-mean-square deviation (RMSD) as an indicator of the stability of predicted protein folding. As shown in Fig. 2A, the GalaxyWeb model was quite stable after a short time whereas the Robetta-domain model continued to deviate throughout the entire simulation period. In addition, the quality scores of the models at the end of the 100 ns simulations differed only slightly from the initial structures, as expected (see “Further Models” in Table 2). Based on the RMSD results, we finally chose the GalaxyWeb model to examine the effects of variants. This M1AP model is available in the PMDB Protein Model DataBase (<http://srv00.recas.ba.infn.it/PMDB/>) (Castrignanò *et al.*, 2006) under the accession identifier PM0083500.

Since the AlphaFold2 (AF2) revolutionized the field of *de novo* protein modeling during the revision process of this article (Jumper *et al.*, 2021; Thornton, Laskowski & Borkakoti, 2021; Varadi *et al.*, 2021), we compared our final model with the AF2 model as an additional validation. As shown in Fig. S1, our model and AF2 model were quite similar (RMSD = 8.42 Å). In addition, the AF2 model has similar level of quality (*i.e.*, better VERIFY3D and PROVE but lower ERRAT, PROCHECK and WHATCHECK scores) with our final model (see “Comparison Model” in Table 2). Because of this independently obtained resemblance of our GalaxyWeb model to this revolutionary *de novo* structure, we might infer that it is one of the best structures that can be computationally obtained by present knowledge.

Variant models were evaluated to see if they retained quality. As shown in “Variant Models” in Table 2, their qualities were quite similar to those of the wild-type near the end of the simulations (*i.e.*, 450th ns), pointing out the validity of the simulation results.

Structural details

The final M1AP model has a globular shape (Fig. 2B) and involves 13 α -helices and 11 β -sheets. The detailed topology including the secondary structures generated by PDBsum is shown in Fig. 2C. Moreover, the model has a large cleft (volume: 11,448 Å³) shown in Fig. 2D. Upon observations of the absence of M1AP in the nucleus and impaired meiosis in the M1AP deficiency, Arango *et al.* (2013) theorized that M1AP might indirectly regulate meiosis progression by binding and activating a cytoplasmic protein that would later translocate into the nucleus. If this suggestion is correct, this large cleft in M1AP would be the most probable binding site for this theorized protein. Other clefts are not shown because they were significantly smaller (*i.e.*, 3272 Å³, 2207 Å³, and 1507 Å³) and less likely to be suitable for ligand binding (Laskowski *et al.*, 1996).

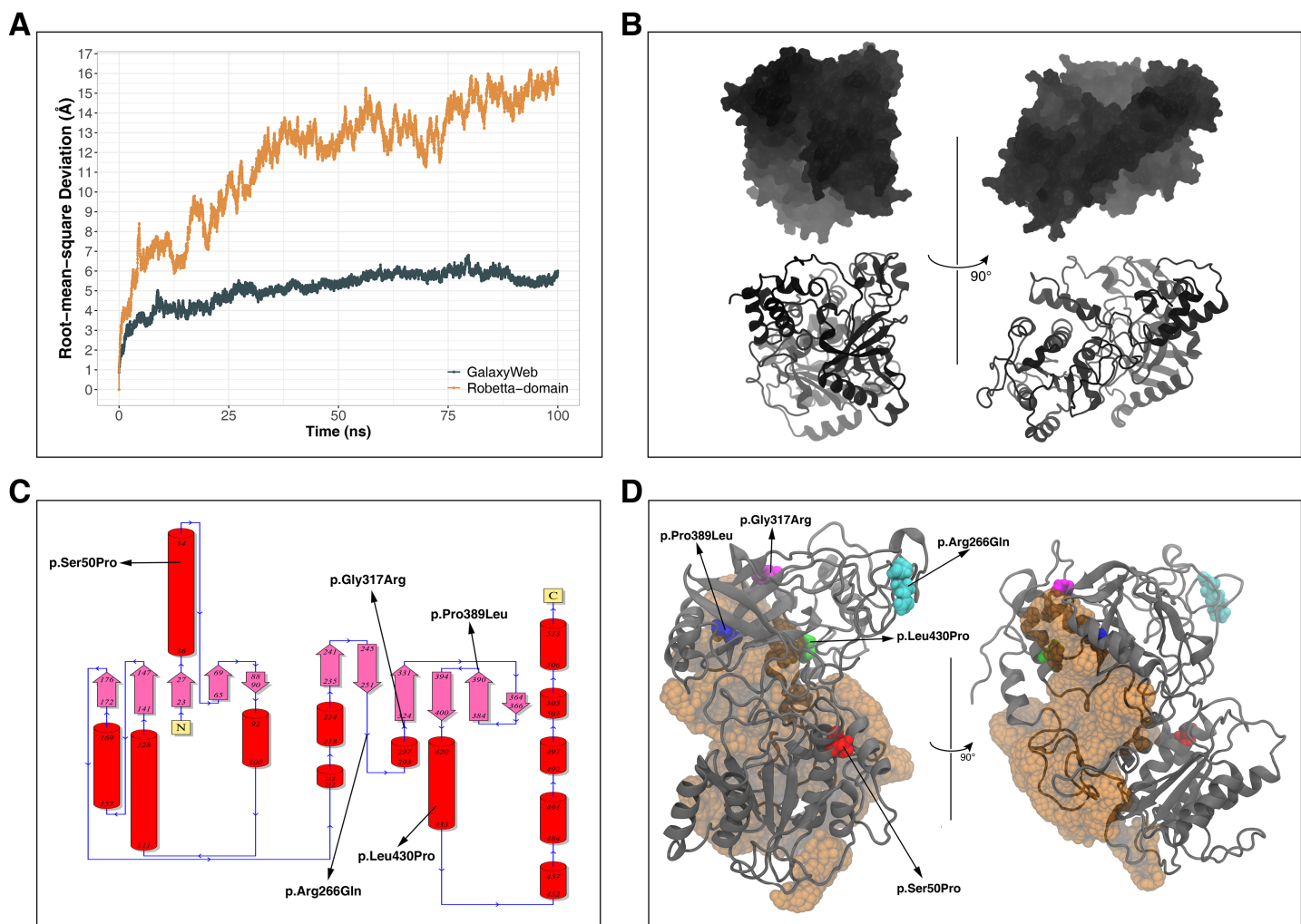


Figure 2 Model details. (A) Backbone root-mean-square deviation (RMSD) comparison between molecular dynamics simulations of GalaxyWeb and Robetta-domain models. (B) Overall view of the M1AP model. (C) The 2D topology of the M1AP model, including the localizations of variants. (D) The largest cleft (orange) in the M1AP structure and localizations of variants on the 3D model. [Full-size !\[\]\(1663bb69f307a960345edb0e712f8c02_img.jpg\) DOI: 10.7717/peerj.12947/fig-2](https://doi.org/10.7717/peerj.12947/fig-2)

The locations of missense variants are shown on the tertiary structure of M1AP in [Fig. 2D](#) (also visible in the secondary structure topology of M1AP in [Fig. 2C](#)). Moreover, Ser50 and Leu430 are in α -helices; Gly317 is in a loop close to an α -helix; Pro389 is at the edge of a β -sheet; and Arg266 is in a surface-exposed loop. On the other hand, Pro389 and Leu430 are part of the largest cleft while Gly317 is close to it ([Fig. 2D](#)).

Impacts of the variants

Folding, compactness and solvent accessibility

The most stable part of the RMSD was the last 75 ns of each 500 ns trajectory, indicating that the simulations reached equilibrium ([Fig. 1](#)). Taking this into account, backbone deviation was lower in variants p.Ser50Pro (both repeats) and p.Arg266Gln (only repeat 1) and slightly higher in p.Gly317Arg (only repeat 1) than in wild-type whereas there were no remarkable changes in other variants ([Fig. 1](#)). Of note, a rapid and sharp increase was

observed in the p.Leu430Pro variant before gradually decreasing and stabilizing. Combined with the radius of gyration (Rg) pattern (Fig. 3), it can be inferred that the structure partially unfolded and refolded during this time.

Rg was slightly lower in variants p.Arg266Gln (both repeats), p.Gly317Arg (only repeat 2) and p.Leu430Pro (both repeats), indicating that the protein was more compact upon these amino acid changes (Fig. 3). Rg of other variants did not differ from the wild-type. In any variant, we did not observe any critical solvent accessible solvent area (SASA) differences with wild-type (Fig. 4), which might be expected upon compactness changes.

Stability and H-bonds

Although loss or gain of structural stability could be expected in line with RMSD and Rg changes, there were no significant differences in stability of any variant protein (Fig. S5). The number of H-bonds was also not different in any variant (Fig. S6), supporting the stability of the proteins to remain unchanged.

Local interactions

To understand how the local environment is affected by the amino acid changes, we analyzed the alterations in the H-bonds and salt bridges around 10 Å of each variant site. There were minor differences in the number of local H-bonds from the wild-type only in p.Ser50Pro (as decrease in both repeats), p.Arg266Gln (as increase in only repeat 2) and p.Leu430Pro (as decrease in only repeat 2) variants (Fig. S7). On the other hand, there were significant alterations in local salt bridges as mostly loss in the p.Arg266Gln (both repeats) and p.Pro389Leu (mostly repeat 1) whilst mostly gain in the p.Gly317Arg (mostly repeat 1) and Leu430Pro (both repeats) variants (Fig. S8). In addition, there were both type of salt bridge changes in the p.Ser50Pro variant.

Flexibility, hinge and essential dynamics

The conformational spaces observed in all variants were significantly different from the wild-type (Fig. 5), indicating altered essential protein dynamics and energetic minima of backbone conformations. In line with this observation, there were critical root-mean-square fluctuation (RMSF, indicating residue flexibility) changes in at least one repeat of each variant as pointed out by the yellow stars in Fig. 6, and hinge site (zero flexibility regions) differences as indicated by yellow and red colors in Table S4.

Secondary structures

Ser50 and Leu430 were found in α -helices (Figs. 2C and 2D) and changed to Pro, which is one of the amino acids known as secondary structure breakers. Consistent with this expectation, both variants disrupted their α -helices, making them turn and 3(10)-helix structures (Fig. 7), which are less stable than α -helices. In p.Arg266Gln, the variant region was more stable as α -helix and turn rather than a coil observed in wild-type (Fig. 7). Furthermore, these were the most outstanding secondary structure changes at variant sites. However, there were less significant changes in other parts of the protein as well (Fig. S9).

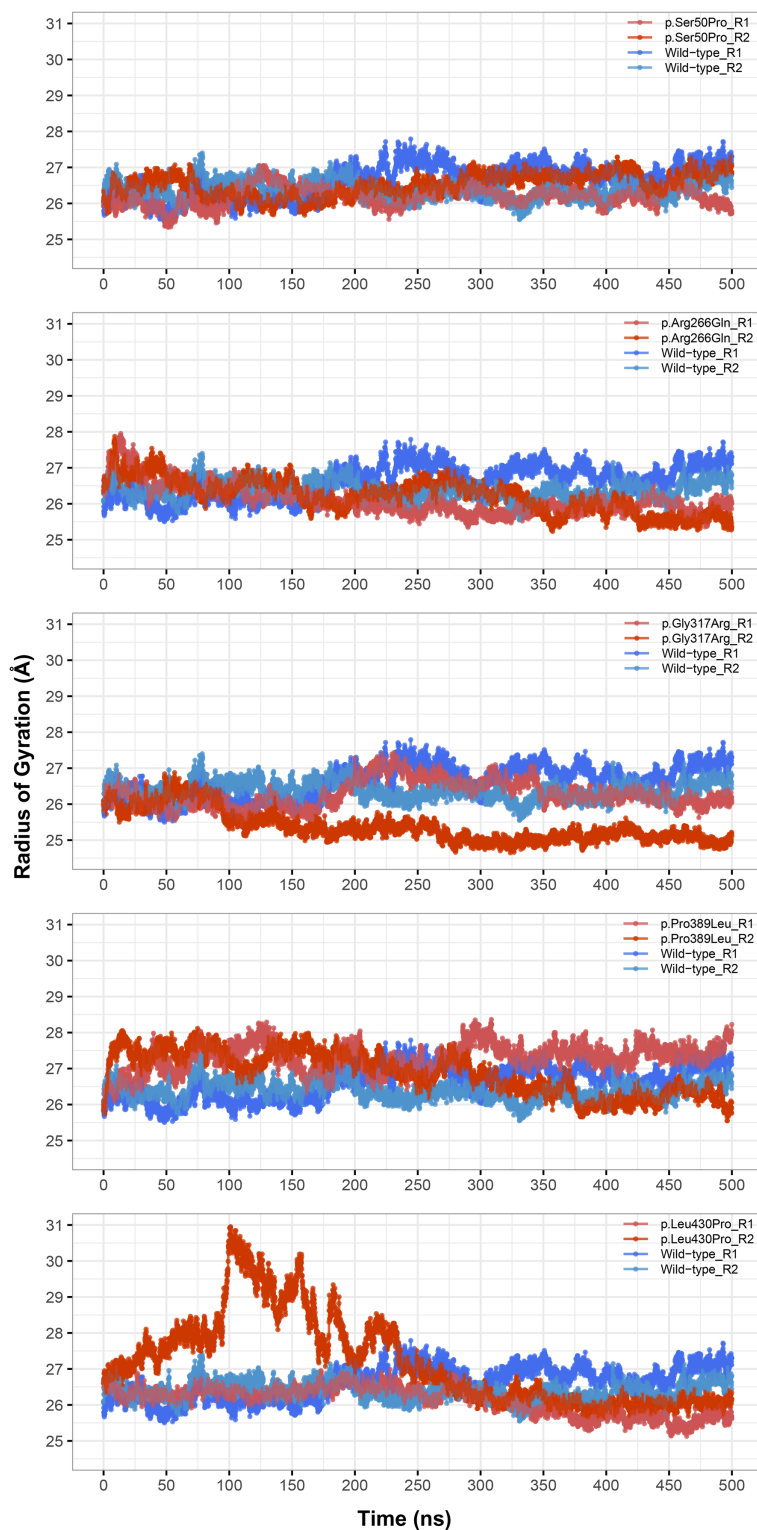


Figure 3 Radius of gyration (Rg) analysis. Rg comparison between molecular dynamics simulations of wild-type and changed M1AP structures. R1 and R2 indicates “Repeat 1” and “Repeat 2” trajectories, respectively. [Full-size !\[\]\(5f471a71b78d7676bc356df190b88ab4_img.jpg\) DOI: 10.7717/peerj.12947/fig-3](https://doi.org/10.7717/peerj.12947/fig-3)

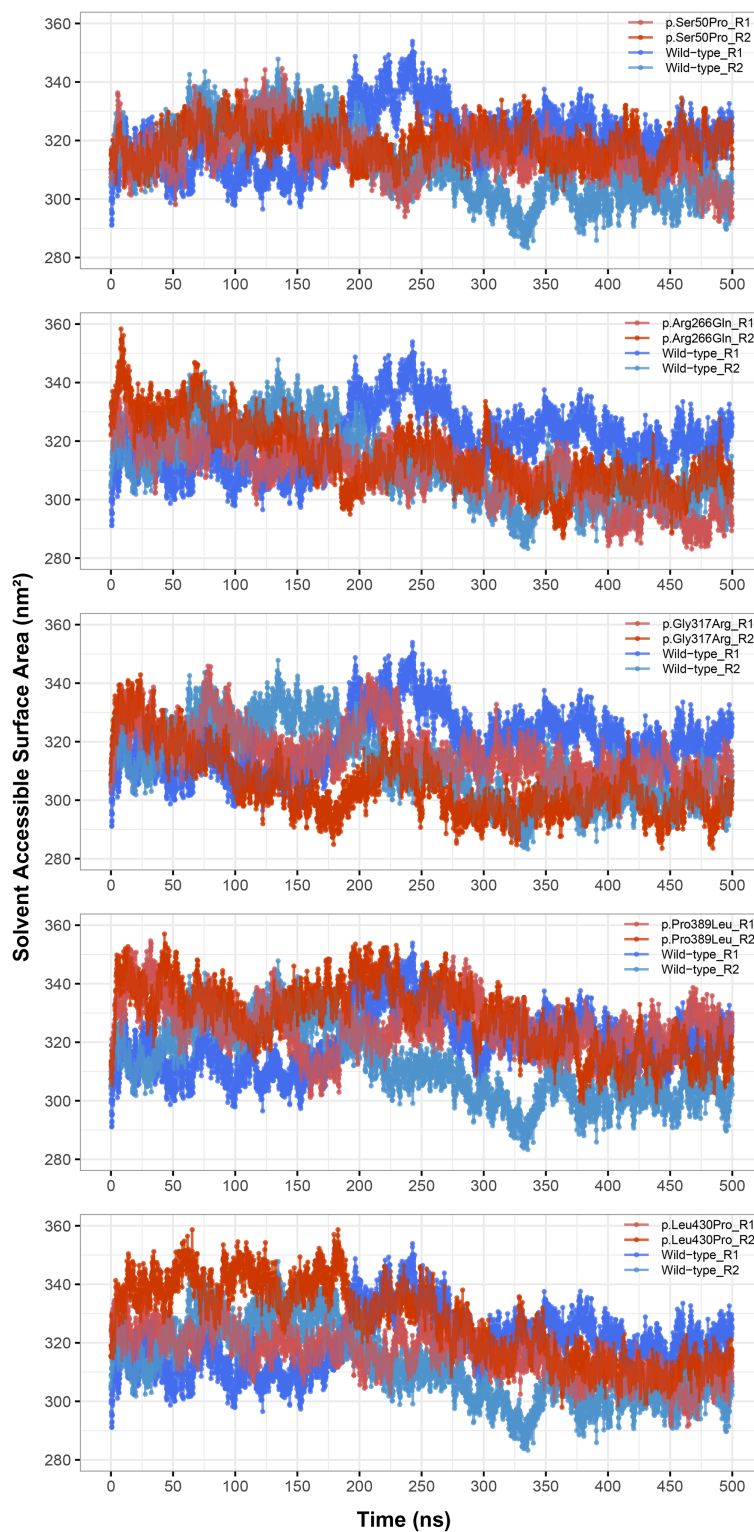


Figure 4 Solvent accessible surface area (SASA) analysis. SASA comparison between molecular dynamics simulations of wild-type and changed M1AP structures. R1 and R2 indicates “Repeat 1” and “Repeat 2” trajectories, respectively. [Full-size !\[\]\(b345a1c4255362eec3746050dd71ccac_img.jpg\) DOI: 10.7717/peerj.12947/fig-4](https://doi.org/10.7717/peerj.12947/fig-4)

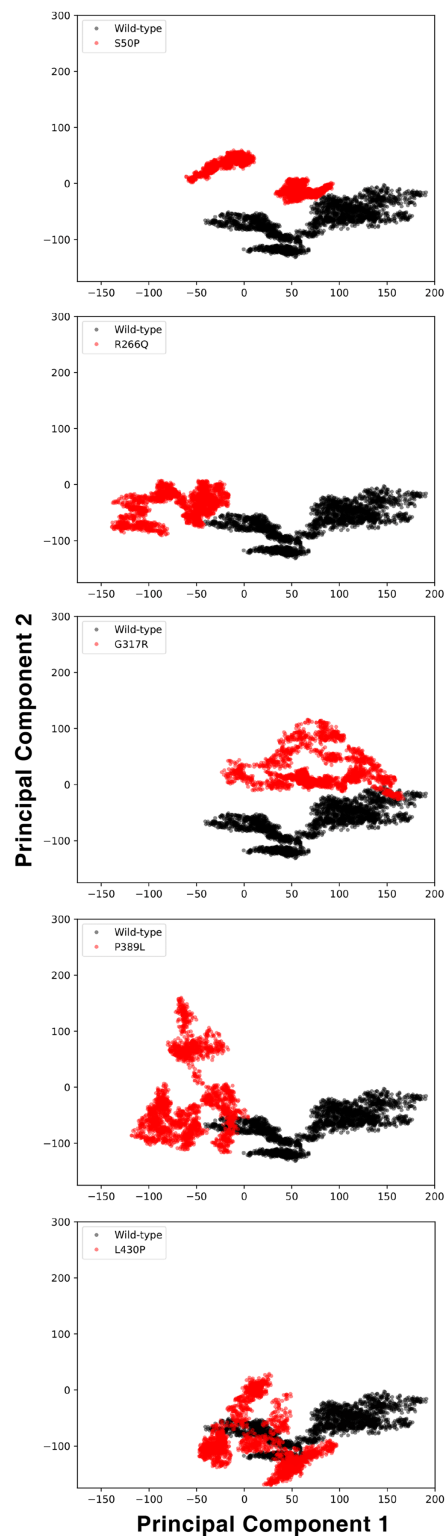


Figure 5 Principal component analysis (PCA). PCA comparison between molecular dynamics simulations of wild-type and changed M1AP structures. The first two principal components are presented.

Full-size  DOI: [10.7717/peerj.12947/fig-5](https://doi.org/10.7717/peerj.12947/fig-5)

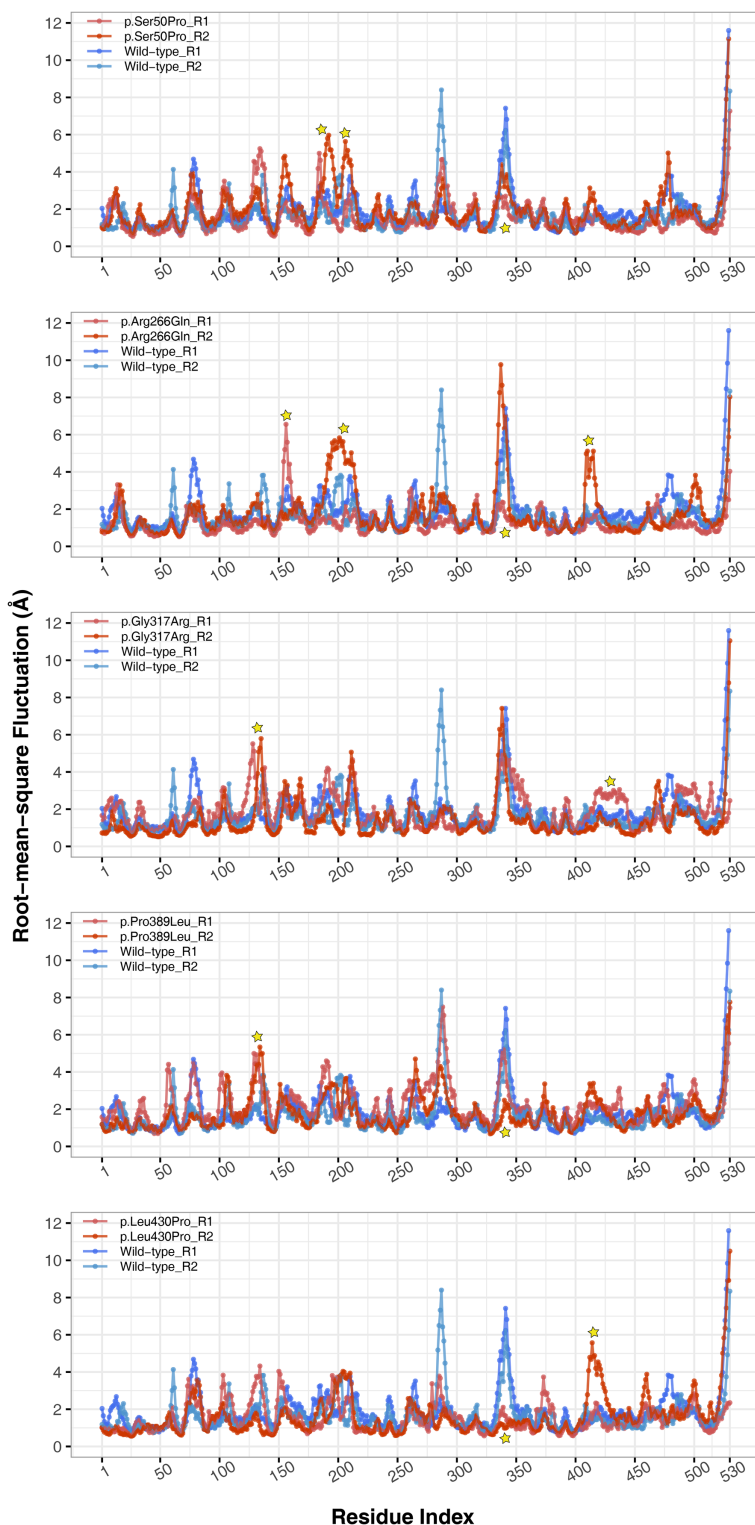


Figure 6 Root-mean-square fluctuations (RMSF) analysis. α RMSF comparison between molecular dynamics simulations of wild-type and changed M1AP structures. Yellow stars indicate regions that differ from the wild-type. R1 and R2 indicates “Repeat 1” and “Repeat 2” trajectories, respectively.

Full-size DOI: [10.7717/peerj.12947/fig-6](https://doi.org/10.7717/peerj.12947/fig-6)

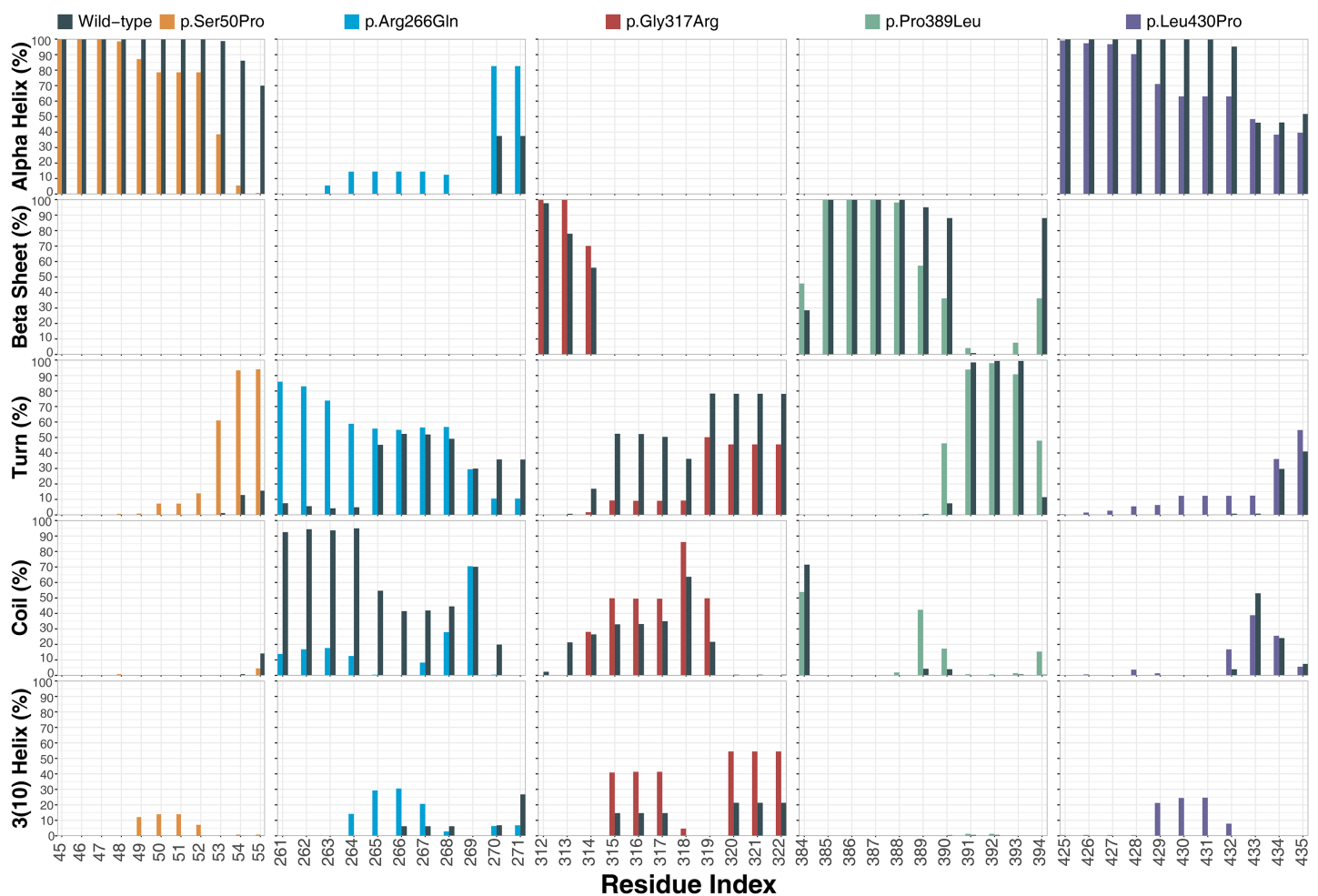


Figure 7 Secondary structure analysis. Secondary structure comparison of the variant sites between molecular dynamics simulations of wild-type and changed M1AP structures. Percentage refers to the percentage of the last 75 ns of simulation during which the amino acids were adopting each one of the secondary structure conformations.

Full-size DOI: [10.7717/peerj.12947/fig-7](https://doi.org/10.7717/peerj.12947/fig-7)

DISCUSSION

In this study, five non-obstructive azoospermia (NOA)-associated missense variants were individually analyzed for their impact on the meiosis 1-associated protein (M1AP) structure by performing molecular dynamics (MD) simulations. Moreover, none of the variants led to general misfolding or loss of stability of the M1AP structure. However, these amino acid changes severely disrupt the conformational dynamics of the protein as a result of altered flexibility, hinge and secondary structure properties due to the local interactions, particularly salt bridges near variant sites. These results were in line with our expectations based on the literature knowledge we had. Specifically, findings by *Wyrwoll et al. (2020)* and *Arango et al. (2013)* helped to shape these expectations and to propose the following hypothesis. Firstly, *Arango et al. (2013)* hypothesized an interaction partner for M1AP that goes later into the nucleus to regulate meiosis since they observed that M1AP deficiency cause disrupted meiosis but there is no M1AP presence in the nucleus. Considering our MD results on the conformational dynamics, these variants have a huge

potential to disrupt such protein–protein interactions, which might relate them with the underlying mechanism of NOA. On the other hand, [Wyrwoll et al. \(2020\)](#) observed that these variants were bi-allelic or homozygous in the infertile individuals suffered from NOA whereas a heterozygous carrier was fertile and healthy. Thus, such a pathogenicity mechanism seems only possible when the wild-type M1AP is deficient in the cell. This indicates that one copy of the variant M1AP might not be adequate to cause NOA by sufficiently disrupting the related cellular function. Consequently, instead of being variant of uncertain significance as previously concluded in [Wyrwoll et al. \(2020\)](#), these variants might cause NOA when homozygous or bi-allelic. Of note, we propose this hypothesis based on our results and poor literature knowledge. Therefore, all these inferences should be tested with experimental structural and functional studies.

This study has limitations as described in the following. Even though M1AP is known to play an essential role in spermatogenesis ([Arango et al., 2006, 2013](#)), there is no further knowledge in the literature about its structure, functions, interactions, regulation and modifications. Therefore, we are currently unable to interpret the impacts of the alterations observed in MD simulations on upstream/downstream cellular functions, particularly through interactions of M1AP with others that could associate our results with the pathogenic mechanisms of NOA. However, there are many other studies that show similar structural distortions observed in different proteins (*e.g.*, in secondary structures, flexibility of regions, hinge residues and essential dynamics) are frequently associated with diseases in a causative manner ([Zheng, Hitchcock-DeGregori & Barua, 2016](#); [Amir et al., 2019](#); [Elghobashi-Meinhardt, 2020](#); [Martínez-Archundia et al., 2020](#); [Mohammad et al., 2020](#); [Shuaib et al., 2020](#)). Therefore, these *M1AP* variants may give rise to NOA considering their effects on protein structure.

CONCLUSIONS

Our aim in this study was to understand how non-obstructive azoospermia (NOA)-related missense variants previously reported in meiosis 1-associated protein (*M1AP*) gene affect the protein structure of M1AP. For this purpose, we firstly modeled the protein structure of M1AP, which has not been studied sufficiently in the literature despite its known importance in meiosis. We tried to obtain the best model we could by applying comprehensive validation procedures to existing methodologies. Next, we utilized molecular dynamics simulations, a robust method to find out how M1AP structure and dynamics are affected by the variants. Our results show that these variants significantly alter the conformational dynamics of M1AP by influencing local interactions and so flexibility, hinge and secondary structure patterns. Therefore, these variants might be causative for NOA when present homozygously or bi-allelically in individuals instead of being heterozygous, as supported by observations in [Wyrwoll et al. \(2020\)](#). However, more studies focusing on M1AP structure, function, regulation and interactions are needed to understand how these variants may affect cellular processes to cause NOA. Although we provide crucial insights about the pathogenicity mechanisms in the *M1AP*-associated NOA causing male infertility, our results and inferences must be tested by *in vitro* and *in vivo* studies.

ACKNOWLEDGEMENTS

Molecular dynamics calculations reported in this study were performed partly in TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA), and partly in Department of Biostatistics and Bioinformatics, Institute of Health Sciences, Acıbadem Mehmet Ali Aydınlar University.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Umut Gerlevik conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Mahmut Cerkez Ergoren conceived and designed the experiments, authored or reviewed drafts of the paper, co-author of *Wyrwoll et al. (2020)*, started this study, and approved the final draft.
- Osman Uğur Sezerman conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Sehime Gulsun Temel conceived and designed the experiments, authored or reviewed drafts of the paper, co-author of *Wyrwoll et al. (2020)*, started this study, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The generated and validated structural model of M1AP is available in the Protein Model Data Base (<http://srv00.recas.ba.infn.it/PMDB/>): PM0083500.

The data generated by molecular dynamics simulations and the related configuration files are available at Zenodo: Gerlevik, Umut, Ergoren, Mahmut Cerkez, Sezerman, Osman Ugur, & Temel, Sehime Gulsun. (2021). All-atom Molecular Dynamics Simulations of Meiosis 1-associated protein (M1AP) to Investigate the Impact of Known Missense Mutations Associated with Male Infertility through Non-obstructive Azoospermia [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5811977>.

All codes used in the analysis are available at GitHub: https://github.com/ugerlevik/M1AP_analysis.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12947#supplemental-information>.

REFERENCES

- Agarwal A, Mulgund A, Hamada A, Chyatte MR. 2015. A unique view on male infertility around the globe. *Reproductive Biology and Endocrinology* **13**(1):873 DOI [10.1186/s12958-015-0032-1](https://doi.org/10.1186/s12958-015-0032-1).
- Amir M, Mohammad T, Kumar V, Alajmi MF, Rehman MT, Hussain A, Alam P, Dohare R, Islam A, Ahmad F, Hassan MI. 2019. Structural analysis and conformational dynamics of STN1 gene mutations involved in coat plus syndrome. *Frontiers in Molecular Biosciences* **6**:41 DOI [10.3389/fmolb.2019.00041](https://doi.org/10.3389/fmolb.2019.00041).
- Arango NA, Huang TT, Fujino A, Pieretti-Vanmarcke R, Donahoe PK. 2006. Expression analysis and evolutionary conservation of the mouse germ cell-specific D6Mm5e gene. *Developmental Dynamics* **235**(9):2613–2619 DOI [10.1002/dvdy.20907](https://doi.org/10.1002/dvdy.20907).
- Arango NA, Li L, Dabir D, Nicolau F, Pieretti-Vanmarcke R, Koehler C, McCarrey JR, Lu N, Donahoe PK. 2013. Meiosis I arrest abnormalities lead to severe oligozoospermia in meiosis 1 arresting protein (M1ap)-deficient mice. *Biology of Reproduction* **88**(3):1–11 DOI [10.1095/biolreprod.111.098673](https://doi.org/10.1095/biolreprod.111.098673).
- Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research* **44**:W344–W350 DOI [10.1093/NAR/GKW408](https://doi.org/10.1093/NAR/GKW408).
- Bakan A, Meireles LM, Bahar I. 2011. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**(11):1575–1577 DOI [10.1093/bioinformatics/btr168](https://doi.org/10.1093/bioinformatics/btr168).
- Castrignanò T, De Meo PDO, Cozzetto D, Talamo IG, Tramontano A. 2006. The PMDB Protein Model Database. *Nucleic Acids Research* **34**:D306–D309 DOI [10.1093/nar/gkj105](https://doi.org/10.1093/nar/gkj105).
- Colovos C, Yeates TO. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science* **2**(9):1511–1519 DOI [10.1002/pro.5560020916](https://doi.org/10.1002/pro.5560020916).
- Davidchack RL, Handel R, Tretyakov MV. 2009. Langevin thermostat for rigid body dynamics. *Journal of Chemical Physics* **130**(23):234101 DOI [10.1063/1.3149788](https://doi.org/10.1063/1.3149788).
- Delgado J, Radusky LG, Cianferoni D, Serrano L, Valencia A. 2019. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics* **35**(20):4168–4169 DOI [10.1093/bioinformatics/btz184](https://doi.org/10.1093/bioinformatics/btz184).
- Di Tommaso P, Moretti S, Xenarios I, Orobittg M, Montanyola A, Chang JM, Taly JF, Notredame C. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research* **39**:W13–W17 DOI [10.1093/nar/gkr245](https://doi.org/10.1093/nar/gkr245).
- Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* **32**:W665–W667 DOI [10.1093/nar/gkh381](https://doi.org/10.1093/nar/gkh381).
- Eisenberg D, Lüthy R, Bowie JU. 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods in Enzymology* **277**:396–404 DOI [10.1016/S0076-6879\(97\)77022-8](https://doi.org/10.1016/S0076-6879(97)77022-8).
- Elghobashi-Meinhardt N. 2020. Cholesterol transport in wild-type NPC1 and P691S: molecular dynamics simulations reveal changes in dynamical behavior. *International Journal of Molecular Sciences* **21**(8):2962 DOI [10.3390/ijms21082962](https://doi.org/10.3390/ijms21082962).
- Feller SE, Zhang Y, Pastor RW, Brooks BR. 1995. Constant pressure molecular dynamics simulation: the Langevin piston method. *Journal of Chemical Physics* **103**(11):4613–4621 DOI [10.1063/1.470648](https://doi.org/10.1063/1.470648).
- Frishman D, Argos P. 1995. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Genetics* **23**(4):566–579 DOI [10.1002/prot.340230412](https://doi.org/10.1002/prot.340230412).

- Gershoni M, Hauser R, Yogev L, Lehavi O, Azem F, Yavetz H, Pietrokovski S, Kleiman SE. 2017. A familial study of azoospermic men identifies three novel causative mutations in three new human azoospermia genes. *Genetics in Medicine* 19:998–1006 DOI 10.1038/gim.2016.225.
- Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. 2006. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 22(21):2695–2696 DOI 10.1093/bioinformatics/btl461.
- Gudeloglu A, Parekattil SJ. 2013. Update in the evaluation of the azoospermic male. *Clinics* 68(S1):27–34 DOI 10.6061/clinics/2013(Sup01)04.
- Hatherley R, Brown DK, Glenister M, Tastan Bishop Ö. 2016. PRIMO: an interactive homology modeling pipeline. *PLOS ONE* 11(11):e0166698 DOI 10.1371/journal.pone.0166698.
- Hooft RWW, Vriend G, Sander C, Abola EE. 1996. Errors in protein structures. *Nature* 381(6580):272–272 DOI 10.1038/381272a0.
- Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, Grubmüller H, MacKerell AD. 2016. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* 14(1):71–73 DOI 10.1038/nmeth.4067.
- Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of Molecular Graphics* 14:33–38 DOI 10.1016/0263-7855(96)00018-5.
- Hunter JD. 2007. Matplotlib: a 2D graphics environment. *Computing in Science & Engineering* 9(3):90–95 DOI 10.1109/MCSE.2007.55.
- Hwang K, Smith JF, Coward RM, Penzias A, Bendikson K, Butts S, Coutifaris C, Falcone T, Fossum G, Gitlin S, Gracia C, Hansen K, Jindal S, La Barbera A, Mersereau J, Odem R, Paulson R, Pfeifer S, Pisarska M, Rebar R, Reindollar R, Rosen M, Sandlow J, Stovall D, Vernon M. 2018. Evaluation of the azoospermic male: a committee opinion. *Fertility and Sterility* 109(5):777–782 DOI 10.1016/j.fertnstert.2018.01.043.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* 79(2):926–935 DOI 10.1063/1.445869.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589 DOI 10.1038/s41586-021-03819-2.
- Karabulut S, Keskin İ, Kutlu P, Delikara N, Atvar Ö, Öztürk M. 2018. Male infertility, azoospermia and cryptozoospermia incidence among three infertility clinics in Turkey. *Turkish Journal of Urology* 44(2):109–113 DOI 10.5152/tud.2018.59196.
- Kassambara A. 2020. Package ‘ggpubr’: ggplot2 based publication ready plots. R package version 0.4.0. Available at <https://cran.r-project.org/package=ggpubr>.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10(6):845–858 DOI 10.1038/nprot.2015.053.
- Kim DE, Chivian D, Baker D. 2004. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* 32:W526–W531 DOI 10.1093/nar/gkh468.
- Ko J, Park H, Heo L, Seok C. 2012. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Research* 40(W1):W294–W297 DOI 10.1093/nar/gks493.
- Kolafa J, Perram JW. 1992. Cutoff errors in the Ewald summation formulae for point charge systems. *Molecular Simulation* 9(5):351–368 DOI 10.1080/08927029208049126.

- Laskowski RA. 2009. PDBsum new things. *Nucleic Acids Research* 37:D355–D359
DOI 10.1093/nar/gkn860.
- Laskowski RA, Luscombe NM, Swindells MB, Thornton JM. 1996. Protein clefts in molecular recognition and function. *Protein Science: A Publication of the Protein Society* 5(12):2438
DOI 10.1002/pro.5560051206.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26(2):283–291 DOI 10.1107/s0021889892009944.
- Lee JY, Dada R, Sabanegh E, Carpi A, Agarwal A. 2011. Role of genetics in azoospermia. *Urology* 77(3):598–601 DOI 10.1016/j.urology.2010.10.001.
- Martínez-Archundia M, Hernández Mojica TG, Correa-Basurto J, Montaña S, Camacho-Molina A. 2020. Molecular dynamics simulations reveal structural differences among wild-type NPC1 protein and its mutant forms. *Journal of Biomolecular Structure and Dynamics* 38(12):3527–3532 DOI 10.1080/07391102.2019.1664324.
- Mohammad T, Amir M, Prasad K, Batra S, Kumar V, Hussain A, Rehman MT, AlAjmi MF, Hassan MI. 2020. Impact of amino acid substitution in the kinase domain of Bruton tyrosine kinase and its association with X-linked agammaglobulinemia. *International Journal of Biological Macromolecules* 164(5):2399–2408 DOI 10.1016/j.ijbiomac.2020.08.057.
- Notredame C, Higgins DG, Heringa J. 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 302(1):205–217
DOI 10.1006/jmbi.2000.4042.
- Olszewska M, Stokowy T, Pollock N, Huleyuk N, Georgiadis A, Yatsenko S, Zastavna D, Yatsenko AN, Kurpisz M. 2020. Familial infertility (Azoospermia and cryptozoospermia) in two brothers—carriers of t(1;7) complex chromosomal rearrangement (ccr): molecular cytogenetic analysis. *International Journal of Molecular Sciences* 21(12):1–18
DOI 10.3390/ijms21124559.
- Peng J, Xu J. 2011. RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins: Structure Function and Bioinformatics* 79(S10):161–171
DOI 10.1002/prot.23175.
- Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. 2021. High-accuracy protein structure prediction in CASP14. *Proteins: Structure, Function, and Bioinformatics* 89(12):1687–1699 DOI 10.1002/prot.26171.
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. 2005. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26:1781–1802 DOI 10.1002/(ISSN)1096-987X.
- Pontius J, Richelle J, Wodak SJ. 1996. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *Journal of Molecular Biology* 264(1):121–136
DOI 10.1006/jmbi.1996.0628.
- R Development Core Team. 2020. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at <http://www.R-project.org/>.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* 17:405–423 DOI 10.1038/gim.2015.30.
- Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 42:W320–W324 DOI 10.1093/nar/gku316.

- Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 5(4):725–738 DOI 10.1038/nprot.2010.5.
- Schrödinger L, DeLano W. 2020. PyMOL. Available at <https://www.pymol.org/>.
- Shuaib S, Saini RK, Goyal D, Goyal B. 2020. Impact of K16A and K28A mutation on the structure and dynamics of amyloid- β 42 peptide in Alzheimer's disease: key insights from molecular dynamics simulations. *Journal of Biomolecular Structure and Dynamics* 38(3):708–721 DOI 10.1080/07391102.2019.1586587.
- The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* 47(D1):D506–D515 DOI 10.1093/nar/gky1049.
- Thornton JM, Laskowski RA, Borkakoti N. 2021. AlphaFold heralds a data-driven revolution in biology and medicine. *Nature Medicine* 27(10):1666–1669 DOI 10.1038/s41591-021-01533-0.
- Tu C, Wang Y, Nie H, Meng L, Wang W, Li Y, Li D, Zhang H, Lu G, Lin G, Tan YQ, Du J. 2020. An M1AP homozygous splice-site mutation associated with severe oligozoospermia in a consanguineous family. *Clinical Genetics* 97(5):741–746 DOI 10.1111/cge.13712.
- Tüttelmann F, Ruckert C, Röpke A. 2018. Disorders of spermatogenesis: perspectives for novel genetic diagnostics after 20 years of unchanged routine. *Medizinische Genetik* 30(1):12–20 DOI 10.1007/s11825-018-0181-7.
- van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. 2011. A graphical interface for the FoldX forcefield. *Bioinformatics* 27(12):1711–1712 DOI 10.1093/bioinformatics/btr254.
- van Rossum G, Drake FL. 2009. Python 3 reference manual. Available at <https://www.python.org/>.
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Židek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S. 2021. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50(D1):D439–D444 DOI 10.1093/NAR/GKAB1061.
- Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag.
- World Health Organization. 2010. *WHO laboratory manual for the examination and processing of human semen*. Geneva: WHO.
- Wyrwoll MJ, Temel ŞG, Nagirnaja L, Oud MS, Lopes AM, van der Heijden GW, Heald JS, Rotte N, Wistuba J, Wöste M, Ledig S, Krenz H, Smits RM, Carvalho F, Gonçalves J, Fietz D, Türkgenç B, Ergören MC, Çetinkaya M, Başar M, Kahraman S, McEleny K, Xavier MJ, Turner H, Pilatz A, Röpke A, Dugas M, Kliesch S, Neuhaus N, Aston KI, Conrad DF, Veltman JA, Friedrich C, Tüttelmann F. 2020. Bi-allelic mutations in M1AP are a frequent cause of meiotic arrest and severely impaired spermatogenesis leading to male infertility. *American Journal of Human Genetics* 107(2):342–351 DOI 10.1016/j.ajhg.2020.06.010.
- Yatsenko AN, Georgiadis AP, Röpke A, Berman AJ, Jaffe T, Olszewska M, Westernströer B, Sanfilippo J, Kurpisz M, Rajkovic A, Yatsenko SA, Kliesch S, Schlatt S, Tüttelmann F. 2015. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *New England Journal of Medicine* 372(22):2097–2107 DOI 10.1056/NEJMoa1406192.
- Zheng W, Hitchcock-DeGregori SE, Barua B. 2016. Investigating the effects of tropomyosin mutations on its flexibility and interactions with filamentous actin using molecular dynamics simulation. *Journal of Muscle Research and Cell Motility* 37(4–5):131–147 DOI 10.1007/s10974-016-9447-3.

- Zhu YT, Luo C, Li Y, Li H, Quan S, Deng YJ, Yang Y, Hu YH, Tan WL, Chu QJ. 2016.** Differences and similarities between extremely severe oligozoospermia and cryptozoospermia in intracytoplasmic sperm injection. *Asian Journal of Andrology* **18(0)**:904–907 DOI [10.4103/1008-682X.165948](https://doi.org/10.4103/1008-682X.165948).
- Zimmermann MT, Urrutia R, Oliver GR, Blackburn PR, Cousin MA, Bozeck NJ, Klee EW. 2017.** Molecular modeling and molecular dynamic simulation of the effects of variants in the TGFBR2 kinase domain as a paradigm for interpretation of variants obtained by next generation sequencing. *PLOS ONE* **12(2)**:1–21 DOI [10.1371/journal.pone.0170822](https://doi.org/10.1371/journal.pone.0170822).