Structural analysis of *M1AP* variants associated with severely impaired spermatogenesis causing male infertility

Umut Gerlevik (ORCID: 0000-0003-4288-6340) [1, 2], Mahmut Çerkez Ergören (ORCID: 0000-0001-9593-9325) [3, 4], Osman Uğur Sezerman (ORCID: 0000-0003-0905-6783) [1, 5, *], Şehime Gülsün Temel (ORCID: 0000-0002-9802-0880) [6, 7, 8, *]

[1] Department of Biostatistics and Bioinformatics, Institute of Health Sciences, Acibadem Mehmet Ali Aydınlar University, Istanbul, 34752, Turkey.

[2] Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom.

[3] Department of Medical Genetics, Faculty of Medicine, Near East University, Nicosia, 99138, Cyprus.

[4] DESAM Institute, Near East University, Nicosia, 99138, Cyprus.

[5] Department of Biostatistics and Medical Informatics, School of Medicine, Acibadem Mehmet Ali Aydınlar University, Istanbul, 34752, Turkey.

[6] Department of Medical Genetics, Faculty of Medicine, Bursa Uludag University, Bursa, 16059, Turkey.

[7] Department of Histology & Embryology, Faculty of Medicine, Bursa Uludag University, Bursa, 16059, Turkey.

[8] Department of Translational Medicine, Health Sciences Institute, Bursa Uludag University, 16059, Bursa, Turkey.

* Corresponding authors:

Osman Uğur Sezerman
E-mail address: ugur.sezerman@acibadem.edu.tr

Şehime Gülsün Temel
E-mail address: sehime@uludag.edu.tr

Abstract

Background: Impaired meiosis can result in absence of sperm in the seminal fluid. This condition, namely non-obstructive azoospermia (NOA), is one the reasons of male infertility. Despite the low number of studies on meiosis 1-associated protein (*M1AP*) in the literature, *M1AP* is known to be crucial for spermatogenesis. Recently, seven variants (5 missense, 1 frameshift, 1 splice-site) have been reported in the *M1AP* 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 M1AP 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 M1AP 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 M1AP, 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 M1AP deficiency since the variants are reported to be homozygous or bi-allelic in the infertile individuals. Our results provided reasonable insights about the M1AP structure and the effects of the variants to the structure and dynamics, which should be further investigated by experimental studies to validate.

53 **Keywords:** Meiosis 1-associated protein (M1AP); non-obstructive azoospermia (NOA);
54 cryptozoospermia; male infertility; molecular dynamics simulations; molecular modeling; variant
55 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 (*M1AP*, 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).

M1AP is a gene expressed in germ cells in both males and females, and it encodes a 530 amino acid long protein (M1AP). 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

86   spermatogenesis in male germ cells. Therefore, they suggest that the evolutionarily
87   well-conserved M1AP protein may have a role in gametogenesis. Furthermore, Arango
88   *et al.* (2013) showed that *M1AP*-deficient male mice have severe oligozoospermia and
89   infertility with many testicular defects. Therefore, they suggest that mutations in *M1AP*
90   might be reason for nonobstructive oligozoospermia in men. In addition, they localized
91   M1AP as a cytosolic protein. Unfortunately, no further knowledge is available in the
92   literature due to the lack of focus on this protein, which has the potential to be highly
93   critical for meiosis and reproduction.

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

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

115   **Methods**

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

Evolutionary conservations

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

Molecular modeling and validation

132 Sequence of the canonical isoform of M1AP protein (UniProt ID: Q8TC57-1) was used
133 to build models. GalaxyWeb (Ko *et al.*, 2012), I-TASSER (Roy, Kucukural & Zhang,
134 2010), Phyre2 (Kelley *et al.*, 2015), PRIMO (Hatherley *et al.*, 2016), RaptorX (Peng &
135 Xu, 2011) and Robetta (Kim, Chivian & Baker, 2004) were used to model different
136 structures with their default parameters. Regular comparative algorithm (we called
137 Robetta model) and domain prediction algorithm (we called Robetta-domain model),
138 which are part of Robetta webserver (https://robetta.bakerlab.org), were applied
139 separately (Kim, Chivian & Baker, 2004). Robetta algorithms also applied an *ab initio*
140 prediction for part of the structure and compared their reliability with its template-based
141 predictions.

142 A list of the templates automatically selected by the modeling algorithms, along with
143 their experimental methodologies, resolution values, and Ramachandran outliers are
144 available in Table S1. Since all structure predictors do not list the identity, similarity, and

145 coverage details, we aligned the templates with M1AP via NCBI's Protein BLAST

146 (https://blast.ncbi.nlm.nih.gov/) by applying default parameters except the removed E-

147 value threshold (Table S2).

148 Quality of the structures were assessed by Verify3D (Eisenberg, Lüthy & Bowie, 1997),

149 ERRAT (Colovos & Yeates, 1993), Prove (Pontius, Richelle & Wodak, 1996),

150 PROCHECK (Laskowski *et al.*, 1993) and WHATCHECK (Hooft *et al.*, 1996) algorithms

151 on SAVES v6.0 webserver (https://saves.mbi.ucla.edu/). Using these assessments, we

152 firstly selected the best model from each algorithm and then compared them with each

153 other. Since the qualities of the two best models were so close, they were further

154 evaluated for folding stability via molecular dynamics (MD) simulations. Of note, before

155 the quality assessments, the models subjected to MD simulations (see "further models"

156 and "variant models" in Table 2) were processed with "RepairPDB" command of FoldX

157 (v5.0) (Delgado *et al.*, 2019; van Durme *et al.*, 2011), which converts the saved

158 coordinates to the PDB format of SAVES and applies a basic minimization procedure.

159 During the revision process of this article, a revolution in protein structure prediction was

160 carried out in the Critical Assessment of protein Structure Prediction (CASP) (Pereira *et*

161 *al.*, 2021) by Google DeepMind. They developed a deep learning-based algorithm

162 called AlphaFold2 (AF2) which makes *de novo* structure prediction with high accuracy

163 (Jumper *et al.*, 2021; Thornton, Laskowski & Borkakoti, 2021; Varadi *et al.*, 2021).

164 Because of its success, we compared our model with the AF2 model as an additional

165 validation.   AlphaFold   Protein   Structure   Database

166 (https://alphafold.ebi.ac.uk/entry/Q8TC57) (Varadi *et al.*, 2021) was used to obtain

167 AF2's M1AP model.

168 We used PyMOL's iterative superposition algorithm (Schrödinger & DeLano, 2020) for

169 the structural alignment of the AF2 model with our model (see Figure S1) and to

170 compare the models generated by the same homology modeling algorithm (see Figure

171 S2).

172 Visual Molecular Dynamics (VMD v1.9.3) (Humphrey, Dalke & Schulten, 1996) was

173 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 $112.6 \times 108.8 \times 125.2$ Å^3 in the xyz axes and the system contained ~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 (gamma) 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 Å was used for the van der Waals interactions. The switching function starts at 10 Å and reaches zero at 14 Å. Long-range Coulomb interactions were calculated with particle-mash 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 (Figure 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

204   simulation time for each system was 1,000 ns. All simulation results are available at
205   Zenodo (https://doi.org/10.5281/zenodo.5811977).

206   **Structural and dynamical analyses**

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

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

218   FoldX (v5.0) (Delgado *et al.*, 2019) was used for stability analysis. ProDy (Bakan,
219   Meireles & Bahar, 2011) was used for principal component analysis (PCA), which refers
220   to conformational dynamics or essential motions related to the free energy landscape
221   observed in the MD simulations. ProDy was also used for Gaussian network modeling
222   (GNM) to predict the hinge residues that play a key role in the essential motions of
223   proteins.

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

230   In PCA and secondary structure analysis, repeat 1 and 2 trajectories were merged while
231   obtaining the secondary structure percentages. Only the last 75 ns of each trajectory
232   was used in RMSF, PCA, secondary structure, and hinge analyzes as these parts had

233 the most stable RMSD levels (Fig. 1) indicating that the proteins had reached the
234 energetic minimum.

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

Results

Evolutionary conservation

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

Molecular modeling and validation

248 Twenty-six models for the M1AP structure were built via the following algorithms with
249 automatically selected templates: GalaxyWeb (5 models), I-TASSER (5 models),
250 Phyre2 (1 model), PRIMO (4 models), RaptorX (1 model), Robetta (5 models) and
251 Robetta-domain (5 models). Although the properties regarding to the templates such as
252 identity and coverage (Table S1 and S2) indicate that the M1AP structure is difficult to
253 model, it was possible to model the entire structure as the templates covered different
254 parts of the M1AP and as utilizing the *ab initio* modeling. However, we required a
255 comprehensive validation procedure as follows to proceed with the simulations reliably.

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

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

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

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

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

Structural details

295 The final M1AP model has a globular shape (Fig. 2B) and involves 13 α-helices and 11
296 β-sheets. The detailed topology including the secondary structures generated by
297 PDBsum is shown in Fig. 2C. Moreover, the model has a large cleft (volume: 11,448 Å^3)
298 shown in Fig. 2D. Upon observations of the absence of M1AP in the nucleus and
299 impaired meiosis in the M1AP deficiency, Arango *et al.* (2013) theorized that M1AP
300 might indirectly regulate meiosis progression by binding and activating a cytoplasmic
301 protein that would later translocate into the nucleus. If this suggestion is correct, this
302 large cleft in M1AP would be the most probable binding site for this theorized protein.
303 Other clefts are not shown because they were significantly smaller (i.e., 3272 Å^3, 2207
304 Å^3, and 1507 Å^3) and less likely to be suitable for ligand binding (Laskowski *et al.*,
305 1996).

306 The locations of missense variants are shown on the tertiary structure of M1AP in Fig.
307 2D (also visible in the secondary structure topology of M1AP in Fig. 2C). Moreover,
308 Ser50 and Leu430 are in α-helices; Gly317 is in a loop close to an α-helix; Pro389 is at
309 the edge of a β-sheet; and Arg266 is in a surface-exposed loop. On the other hand,
310 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

313 The most stable part of the RMSD was the last 75 ns of each 500 ns trajectory,
314 indicating that the simulations reached equilibrium (Fig. 1). Taking this into account,
315 backbone deviation was lower in variants p.Ser50Pro (both repeats) and p.Arg266Gln
316 (only repeat 1) and slightly higher in p.Gly317Arg (only repeat 1) than in wild-type
317 whereas there were no remarkable changes in other variants (Fig. 1). Of note, a rapid
318 and sharp increase was observed in the p.Leu430Pro variant before gradually

319 decreasing and stabilizing. Combined with the radius of gyration (Rg) pattern (Fig. 3), it
320 can be inferred that the structure partially unfolded and refolded during this time.

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

Stability and H-bonds

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

Local interactions

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

Flexibility, hinge and essential dynamics

343 The conformational spaces observed in all variants were significantly different from the
344 wild-type (Fig. 5), indicating altered essential protein dynamics and energetic minima of
345 backbone conformations. In line with this observation, there were critical root-mean-
346 square fluctuation (RMSF, indicating residue flexibility) changes in at least one repeat of

347 each variant as pointed out by the yellow stars in Fig. 6, and hinge site (zero flexibility

348 regions) differences as indicated by yellow and red colors in Table S4.

349 *Secondary structures*

350 Ser50 and Leu430 were found in α-helices (Fig. 2C and 2D) and changed to Pro, which

351 is one of the amino acids known as secondary structure breakers. Consistent with this

352 expectation, both variants disrupted their α-helices, making them turn and 3(10)-helix

353 structures (Figure 7), which are less stable than α-helices. In p.Arg266Gln, the variant

354 region was more stable as α-helix and turn rather than a coil observed in wild-type

355 (Fig.7). Furthermore, these were the most outstanding secondary structure changes at

356 variant sites. However, there were less significant changes in other parts of the protein

357 as well (Fig. S9).

Discussion

359 In this study, five non-obstructive azoospermia (NOA)-associated missense variants

360 were individually analyzed for their impact on the meiosis 1-associated protein (M1AP)

361 structure by performing molecular dynamics (MD) simulations. Moreover, none of the

362 variants led to general misfolding or loss of stability of the M1AP structure. However,

363 these amino acid changes severely disrupt the conformational dynamics of the protein

364 as a result of altered flexibility, hinge and secondary structure properties due to the local

365 interactions, particularly salt bridges near variant sites. These results were in line with

366 our expectations based on the literature knowledge we had. Specifically, findings by

367 Wyrwoll *et al*. (2020) and Arango *et al*. (2013) helped to shape these expectations and

368 to propose the following hypothesis. Firstly, Arango *et al.* (2013) hypothesized an

369 interaction partner for M1AP that goes later into the nucleus to regulate meiosis since

370 they observed that M1AP deficiency cause disrupted meiosis but there is no M1AP

371 presence in the nucleus. Considering our MD results on the conformational dynamics,

372 these variants have a huge potential to disrupt such protein-protein interactions, which

373 might relate them with the underlying mechanism of NOA. On the other hand, Wyrwoll

374 *et al*. (2020) observed that these variants were bi-allelic or homozygous in the infertile

375 individuals suffered from NOA whereas a heterozygous carrier was fertile and healthy.

376 Thus, such a pathogenicity mechanism seems only possible when the wild-type M1AP
377 is deficient in the cell. This indicates that one copy of the variant M1AP might not be
378 adequate to cause NOA by sufficiently disrupting the related cellular function.
379 Consequently, instead of being variant of uncertain significance as previously concluded
380 in Wyrwoll *et al.* (2020), these variants might cause NOA when homozygous or bi-
381 allelic. Of note, we propose this hypothesis based on our results and poor literature
382 knowledge. Therefore, all these inferences should be tested with experimental structural
383 and functional studies.

384 This study has limitations as described in the following. Even though M1AP is known to
385 play an essential role in spermatogenesis (Arango *et al.*, 2006, 2013), there is no further
386 knowledge in the literature about its structure, functions, interactions, regulation and
387 modifications. Therefore, we are currently unable to interpret the impacts of the
388 alterations observed in MD simulations on upstream/downstream cellular functions,
389 particularly through interactions of M1AP with others that could associate our results
390 with the pathogenic mechanisms of NOA. However, there are many other studies that
391 show similar structural distortions observed in different proteins (e.g., in secondary
392 structures, flexibility of regions, hinge residues and essential dynamics) are frequently
393 associated with diseases in a causative manner (Zheng, Hitchcock-DeGregori & Barua,
394 2016; Amir *et al.*, 2019; Elghobashi-Meinhardt, 2020; Martínez-Archundia *et al.*, 2020;
395 Mohammad *et al.*, 2020; Shuaib *et al.*, 2020). Therefore, these *M1AP* variants may give
396 rise to NOA considering their effects on protein structure.

397 **Conclusions**

398 Our aim in this study was to understand how non-obstructive azoospermia (NOA)-
399 related missense variants previously reported in meiosis 1-associated protein (*M1AP*)
400 gene affect the protein structure of M1AP. For this purpose, we firstly modeled the
401 protein structure of M1AP, which has not been studied sufficiently in the literature
402 despite its known importance in meiosis. We tried to obtain the best model we could by
403 applying comprehensive validation procedures to existing methodologies. Next, we
404 utilized molecular dynamics simulations, a robust method to find out how M1AP
405 structure and dynamics are affected by the variants. Our results show that these

406 variants significantly alter the conformational dynamics of M1AP by influencing local
407 interactions and so flexibility, hinge and secondary structure patterns. Therefore, these
408 variants might be causative for NOA when present homozygously or bi-allelically in
409 individuals instead of being heterozygous, as supported by observations in Wyrwoll *et*
410 *al.* (2020). However, more studies focusing on M1AP structure, function, regulation and
411 interactions are needed to understand how these variants may affect cellular processes
412 to cause NOA. Although we provide crucial insights about the pathogenicity
413 mechanisms in the *M1AP*-associated NOA causing male infertility, our results and
414 inferences must be tested by *in vitro* and *in vivo* studies.

Supplemental materials

415

416 **Figure S1: AlphaFold2 (AF2) model comparison.** (A) M1AP model of AF2. Blue color
417 refers to ≥ 70% confidence while others indicate lower confidence. (B) Structural
418 alignment of GalaxyWeb (grey) and AF2 (orange) M1AP models, including root-mean-
419 square deviation (RMSD) value.

420 **Figure S2: Structural alignments of each model of multiple model generating**
421 **algorithms to its model 1.** Root-mean-square deviation (RMSD) values are included.
422 Grey color indicate model 1.

423 **Figure S3: Potential energy analysis.** The energy pattern of each system during the
424 entire simulation procedure.

425 **Figure S4: Evolutionary conservation analysis.** Alignment of the human M1AP
426 protein to its five orthologs. Yellow stars with labels indicate the positions of the variants
427 examined in this study.

428 **Figure S5: FoldX folding stability analysis.** Stability comparison between molecular
429 dynamics simulations of wild-type and changed M1AP structures. R1 and R2 indicates
430 "Repeat 1" and "Repeat 2" trajectories, respectively.

431 **Figure S6: Overall H-bond analysis.** Comparison of the number of all H-bonds
432 between molecular dynamics simulations of wild-type and changed M1AP structures.
433 R1 and R2 indicates "Repeat 1" and "Repeat 2" trajectories, respectively.

434 **Figure S7: Local (within 10 Å of the variant site) H-bond analysis.** Comparison of

435 the number of local H-bonds between molecular dynamics simulations of wild-type and

436 changed M1AP structures. R1 and R2 indicates "Repeat 1" and "Repeat 2" trajectories,

437 respectively.

438 **Figure S8: Local (within 10 Å of the variant site) salt bridge analysis.** Comparison

439 of the local salt bridges between molecular dynamics simulations of wild-type and

440 changed M1AP structures.

441 **Figure S9: Secondary structure analysis.** Secondary structure comparison between

442 molecular dynamics simulations of wild-type and changed M1AP structures. Percentage

443 refers to the percentage of the last 75 ns of simulation during which the amino acids

444 were adopting each one of the secondary structure conformations.

445 **Table S1: List of templates assigned for M1AP by the structural modeling**

446 **algorithms.** The methods used to obtain the structures are given with resolution (the

447 smaller the better) and Ramachandran outliers (the smaller the better) indicating the

448 quality of the structures.

449 **Table S2: BLAST alignment details of the templates automatically selected by the**

450 **structural modeling algorithms.**

451 **Table S3: Quality assessments of the models built by multiple model generating**

452 **algorithms from five different algorithms.** (A) GalaxyWeb, (B) I-TASSER, (C)

453 PRIMO, (D) Robetta and (E) Robetta-domain.

454 **Table S4: Hinge site analysis.** Hinge site comparison between molecular dynamics

455 simulations of wild-type and changed M1AP structures. Same colors refer to matching

456 regions. Red and yellow colors indicate the different regions from that found in the wild-

457 type structure.

458 ## Acknowledgments

459 Molecular dynamics calculations reported in this study were performed partly in

460 TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA), and

461 partly in Department of Biostatistics and Bioinformatics, Institute of Health Sciences,
462 Acıbadem Mehmet Ali Aydınlar University.

463 **References**

464 Agarwal A, Mulgund A, Hamada A, Chyatte MR. 2015. A unique view on male infertility
465   around the globe. *Reproductive Biology and Endocrinology* 13:1–9. DOI:
466   10.1186/s12958-015-0032-1.

467 Amir M, Mohammad T, Kumar V, Alajmi MF, Rehman MT, Hussain A, Alam P, Dohare
468   R, Islam A, Ahmad F, Hassan MI. 2019. Structural analysis and conformational
469   dynamics of STN1 gene mutations involved in coat plus syndrome. *Frontiers in*
470   *Molecular Biosciences*. DOI: 10.3389/fmolb.2019.00041.

471 Arango NA, Huang TT, Fujino A, Pieretti-Vanmarcke R, Donahoe PK. 2006. Expression
472   analysis and evolutionary conservation of the mouse germ cell-specific D6Mm5e
473   gene. *Developmental Dynamics* 235:2613–2619. DOI: 10.1002/dvdy.20907.

474 Arango NA, Li L, Dabir D, Nicolau F, Pieretti-Vanmarcke R, Koehler C, McCarrey JR, Lu
475   N, Donahoe PK. 2013. Meiosis I arrest abnormalities lead to severe
476   oligozoospermia in meiosis 1 arresting protein (M1ap)-deficient mice. *Biology of*
477   *Reproduction* 88:1–11. DOI: 10.1095/biolreprod.111.098673.

478 Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016.
479   ConSurf 2016: an improved methodology to estimate and visualize evolutionary
480   conservation in macromolecules. *Nucleic Acids Research* 44:W344–W350. DOI:
481   10.1093/NAR/GKW408.

482 Bakan A, Meireles LM, Bahar I. 2011. ProDy: Protein dynamics inferred from theory and

483   experiments. *Bioinformatics* 27:1575–1577. DOI: 10.1093/bioinformatics/btr168.

484   Castrignanò T, De Meo PDO, Cozzetto D, Talamo IG, Tramontano A. 2006. The PMDB

485   Protein Model Database. *Nucleic acids research*. DOI: 10.1093/nar/gkj105.

486   Colovos C, Yeates TO. 1993. Verification of protein structures: Patterns of nonbonded

487   atomic interactions. *Protein Science* 2:1511–1519. DOI: 10.1002/pro.5560020916.

488   Davidchack RL, Handel R, Tretyakov M V. 2009. Langevin thermostat for rigid body

489   dynamics. *Journal of Chemical Physics*. DOI: 10.1063/1.3149788.

490   Delgado J, Radusky LG, Cianferoni D, Serrano L, Valencia A. 2019. FoldX 5.0: Working

491   with RNA, small molecules and a new graphical interface. *Bioinformatics*. DOI:

492   10.1093/bioinformatics/btz184.

493   Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. 2004. PDB2PQR: An automated

494   pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic*

495   *Acids Research*. DOI: 10.1093/nar/gkh381.

496   van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. 2011. A

497   graphical interface for the FoldX forcefield. *Bioinformatics* 27:1711–1712. DOI:

498   10.1093/BIOINFORMATICS/BTR254.

499   Eisenberg D, Lüthy R, Bowie JU. 1997. VERIFY3D: Assessment of protein models with

500   three-dimensional profiles. *Methods in Enzymology* 277:396–404. DOI:

501   10.1016/S0076-6879(97)77022-8.

502   Elghobashi-Meinhardt N. 2020. Cholesterol transport in wild-type NPC1 and P691S:

503   Molecular dynamics simulations reveal changes in dynamical behavior.

504   *International Journal of Molecular Sciences*. DOI: 10.3390/ijms21082962.

505   Feller SE, Zhang Y, Pastor RW, Brooks BR. 1995. Constant pressure molecular

506   dynamics simulation: The Langevin piston method. *The Journal of Chemical*

507   *Physics*. DOI: 10.1063/1.470648.

508   Frishman D, Argos P. 1995. Knowledge-based protein secondary structure assignment.

509   *Proteins: Structure, Function, and Bioinformatics*. DOI: 10.1002/prot.340230412.

510   Gershoni M, Hauser R, Yogev L, Lehavi O, Azem F, Yavetz H, Pietrokovski S, Kleiman

511   SE. 2017. A familial study of azoospermic men identifies three novel causative

512   mutations in three new human azoospermia genes. *Genetics in Medicine*. DOI:

513   10.1038/gim.2016.225.

514   Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. 2006. Bio3d: An R

515   package for the comparative analysis of protein structures. *Bioinformatics*. DOI:

516   10.1093/bioinformatics/btl461.

517   Gudeloglu A, Parekattil SJ. 2013. Update in the evaluation of the azoospermic male.

518   *Clinics* 68:27–34. DOI: 10.6061/clinics/2013(Sup01)04.

519   Hatherley R, Brown DK, Glenister M, Tastan Bishop Ö. 2016. PRIMO: An Interactive

520   Homology   Modeling   Pipeline.   *PLOS*   *ONE*   11:e0166698.   DOI:

521   10.1371/journal.pone.0166698.

522   Hooft RWW, Vriend G, Sander C, Abola EE. 1996. Errors in protein structures. *Nature*

523   381:272–272. DOI: 10.1038/381272a0.

524   Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, Grubmüller H,

525   MacKerell AD. 2016. CHARMM36m: An improved force field for folded and

526   intrinsically disordered proteins. *Nature Methods* 14:71–73. DOI:

527   10.1038/nmeth.4067.

528   Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of*

529   *molecular graphics* 14:33–8, 27–8.

530   Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Computing in Science and*

531   *Engineering*. DOI: 10.1109/MCSE.2007.55.

532   Hwang K, Smith JF, Coward RM, Penzias A, Bendikson K, Butts S, Coutifaris C,

533   Falcone T, Fossum G, Gitlin S, Gracia C, Hansen K, Jindal S, La Barbera A,

534   Mersereau J, Odem R, Paulson R, Pfeifer S, Pisarska M, Rebar R, Reindollar R,

535   Rosen M, Sandlow J, Stovall D, Vernon M. 2018. Evaluation of the azoospermic

536   male: a committee opinion. *Fertility and Sterility* 109:777–782. DOI:

537   10.1016/j.fertnstert.2018.01.043.

538   Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison

539   of simple potential functions for simulating liquid water. *The Journal of Chemical*

540   *Physics*. DOI: 10.1063/1.445869.

541   Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool

542   K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ,

543   Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S,

544   Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T,

545   Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis

546   D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 2021

547   596:7873 596:583–589. DOI: 10.1038/s41586-021-03819-2.

548   Karabulut S, Keskin İ, Kutlu P, Delikara N, Atvar Ö, Öztürk M. 2018. Male infertility,

549   azoozpermia and cryptozoospermia incidence among three infertility clinics in

550   Turkey. *Turkish Journal of Urology* 44:109–113. DOI: 10.5152/tud.2018.59196.

551   Kassambara A. 2020. Package 'ggpubr': "ggplot2" Based Publication Ready Plots. *R*

552   *package version 0.4.0.*

553   Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web

554   portal for protein modeling, prediction and analysis. *Nature Protocols* 10:845–858.

555   DOI: 10.1038/nprot.2015.053.

556   Kim DE, Chivian D, Baker D. 2004. Protein structure prediction and analysis using the

557   Robetta server. *Nucleic acids research* 32:W526-31. DOI: 10.1093/nar/gkh468.

558   Ko J, Park H, Heo L, Seok C. 2012. GalaxyWEB server for protein structure prediction

559   and refinement. *Nucleic Acids Research*. DOI: 10.1093/nar/gks493.

560   Kolafa J, Perram JW. 1992. Cutoff Errors in the Ewald Summation Formulae for Point

561   Charge Systems. *Molecular Simulation*. DOI: 10.1080/08927029208049126.

562   Laskowski RA. 2009. PDBsum new things. *Nucleic acids research* 37:D355-9. DOI:

563   10.1093/nar/gkn860.

564   Laskowski RA, Luscombe NM, Swindells MB, Thornton JM. 1996. Protein clefts in

565   molecular recognition and function. Protein science : a publication of the Protein

566   Society. DOI: 10.1002/pro.5560051206.

567   Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program

568   to check the stereochemical quality of protein structures. *Journal of Applied*

569   *Crystallography* 26:283–291. DOI: 10.1107/s0021889892009944.

570   Lee JY, Dada R, Sabanegh E, Carpi A, Agarwal A. 2011. Role of genetics in

571   azoospermia. *Urology*. DOI: 10.1016/j.urology.2010.10.001.

572   Martínez-Archundia M, Hernández Mojica TG, Correa-Basurto J, Montaño S, Camacho-

573   Molina A. 2020. Molecular dynamics simulations reveal structural differences

574   among wild-type NPC1 protein and its mutant forms. *Journal of Biomolecular*

575   *Structure and Dynamics*. DOI: 10.1080/07391102.2019.1664324.

576   Mohammad T, Amir M, Prasad K, Batra S, Kumar V, Hussain A, Rehman MT, AlAjmi

577   MF, Hassan MI. 2020. Impact of amino acid substitution in the kinase domain of

578   Bruton tyrosine kinase and its association with X-linked agammaglobulinemia.

579   *International Journal of Biological Macromolecules*. DOI:

580   10.1016/j.ijbiomac.2020.08.057.

581   Notredame C, Higgins DG, Heringa J. 2000. T-coffee: A novel method for fast and

582   accurate multiple sequence alignment. *Journal of Molecular Biology*. DOI:

583   10.1006/jmbi.2000.4042.

584   Olszewska M, Stokowy T, Pollock N, Huleyuk N, Georgiadis A, Yatsenko S, Zastavna

585   D, Yatsenko AN, Kurpisz M. 2020. Familial infertility (Azoospermia and

586   cryptozoospermia) in two brothers—carriers of t(1;7) complex chromosomal

587   rearrangement (ccr): Molecular cytogenetic analysis. *International Journal of*

588   *Molecular Sciences* 21:1–18. DOI: 10.3390/ijms21124559.

589   Peng J, Xu J. 2011. RaptorX: exploiting structure information for protein alignment by

590   statistical inference. *Proteins* 79 Suppl 1:161–71. DOI: 10.1002/prot.23175.

591   Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. 2021. High-

592   accuracy protein structure prediction in CASP14. Proteins: Structure, Function, and

593   Bioinformatics 89:1687–1699. DOI: 10.1002/PROT.26171.

594   Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD,

595   Kalé L, Schulten K. 2005. Scalable molecular dynamics with NAMD. *Journal of*

596   *Computational Chemistry*. DOI: 10.1002/jcc.20289.

597   Pontius J, Richelle J, Wodak SJ. 1996. Deviations from standard atomic volumes as a

598   quality measure for protein crystal structures. *Journal of Molecular Biology*

599   264:121–136. DOI: 10.1006/jmbi.1996.0628.

600   R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna,

601   Austria: R Foundation for Statistical Computing.

602   Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M,

603   Lyon E, Spector E, Voelkerding K, Rehm HL. 2015. Standards and guidelines for

604   the interpretation of sequence variants: A joint consensus recommendation of the

605   American College of Medical Genetics and Genomics and the Association for

606   Molecular Pathology. Genetics in Medicine. DOI: 10.1038/gim.2015.30.

607   Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new

608   ENDscript server. *Nucleic Acids Research*. DOI: 10.1093/nar/gku316.

609   van Rossum G, Drake FL. 2009. *Python 3 Reference Manual.*

610   Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: A unified platform for automated

611   protein structure and function prediction. *Nature Protocols* 5:725–738. DOI:

612   10.1038/nprot.2010.5.

613 Schrödinger L, & DeLano, W. 2020. *PyMOL*. Available at https://www.pymol.org/.

614 Shuaib S, Saini RK, Goyal D, Goyal B. 2020. Impact of K16A and K28A mutation on the

615   structure and dynamics of amyloid-β42 peptide in Alzheimer's disease: key insights

616   from molecular dynamics simulations. *Journal of Biomolecular Structure and*

617   *Dynamics*. DOI: 10.1080/07391102.2019.1586587.

618 The UniProt Consortium. 2019. UniProt: A worldwide hub of protein knowledge. *Nucleic*

619   *Acids Research*. DOI: 10.1093/nar/gky1049.

620 Thornton JM, Laskowski RA, Borkakoti N. 2021. AlphaFold heralds a data-driven

621   revolution in biology and medicine. Nature Medicine 2021 27:10 27:1666–1669.

622   DOI: 10.1038/s41591-021-01533-0.

623 Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, Taly JF,

624   Notredame C. 2011. T-Coffee: A web server for the multiple sequence alignment of

625   protein and RNA sequences using structural information and homology extension.

626   *Nucleic Acids Research*. DOI: 10.1093/nar/gkr245.

627 Tu C, Wang Y, Nie H, Meng L, Wang W, Li Y, Li D, Zhang H, Lu G, Lin G, Tan YQ, Du

628   J. 2020. An M1AP homozygous splice-site mutation associated with severe

629   oligozoospermia in a consanguineous family. *Clinical Genetics* 97:741–746. DOI:

630   10.1111/cge.13712.

631 Tüttelmann F, Ruckert C, Röpke A. 2018. Disorders of spermatogenesis: Perspectives

632   for novel genetic diagnostics after 20years of unchanged routine. *medizinische*

633   *genetik*. DOI: 10.1007/s11825-018-0181-7.

634   Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe

635   O, Wood G, Laydon A, Žídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper

636   J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P,

637   Kleywegt G, Birney E, Hassabis D, Velankar S. 2021. AlphaFold Protein Structure

638   Database: massively expanding the structural coverage of protein-sequence space

639   with   high-accuracy   models.   Nucleic   Acids   Research.   DOI:

640   10.1093/NAR/GKAB1061.

641   Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New

642   York.

643   World Health Organization. 2010. WHO laboratory manual for the examination and

644   processing of human semen. :Previous editions had different title : WHO labora.

645   Wyrwoll MJ, Temel ŞG, Nagirnaja L, Oud MS, Lopes AM, van der Heijden GW, Heald

646   JS, Rotte N, Wistuba J, Wöste M, Ledig S, Krenz H, Smits RM, Carvalho F,

647   Gonçalves J, Fietz D, Türkgenç B, Ergören MC, Çetinkaya M, Başar M, Kahraman

648   S, McEleny K, Xavier MJ, Turner H, Pilatz A, Röpke A, Dugas M, Kliesch S,

649   Neuhaus N, Aston KI, Conrad DF, Veltman JA, Friedrich C, Tüttelmann F. 2020. Bi-

650   allelic Mutations in M1AP Are a Frequent Cause of Meiotic Arrest and Severely

651   Impaired Spermatogenesis Leading to Male Infertility. *American Journal of Human*

652   *Genetics* 107:342–351. DOI: 10.1016/j.ajhg.2020.06.010.

653   Yatsenko AN, Georgiadis AP, Röpke A, Berman AJ, Jaffe T, Olszewska M,

654   Westernströer B, Sanfilippo J, Kurpisz M, Rajkovic A, Yatsenko SA, Kliesch S,

655   Schlatt S, Tüttelmann F. 2015. X-Linked TEX11 Mutations, Meiotic Arrest, and

656   Azoospermia in Infertile Men . *New England Journal of Medicine*. DOI:

657   10.1056/nejmoa1406192.

658   Zheng W, Hitchcock-DeGregori SE, Barua B. 2016. Investigating the effects of

659   tropomyosin mutations on its flexibility and interactions with filamentous actin using

660   molecular dynamics simulation. *Journal of Muscle Research and Cell Motility*. DOI:

661   10.1007/s10974-016-9447-3.

662   Zhu YT, Luo C, Li Y, Li H, Quan S, Deng YJ, Yang Y, Hu YH, Tan WL, Chu QJ. 2016.

663   Differences and similarities between extremely severe oligozoospermia and

664   cryptozoospermia in intracytoplasmic sperm injection. *Asian Journal of Andrology*

665   18:904–907. DOI: 10.4103/1008-682X.165948.

666   Zimmermann MT, Urrutia R, Oliver GR, Blackburn PR, Cousin MA, Bozeck NJ, Klee

667   EW. 2017. Molecular modeling and molecular dynamic simulation of the effects of

668   variants in the TGFBR2 kinase domain as a paradigm for interpretation of variants

669   obtained by next generation sequencing. *PLoS ONE* 12:1–21. DOI:

670   10.1371/journal.pone.0170822.

671

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.
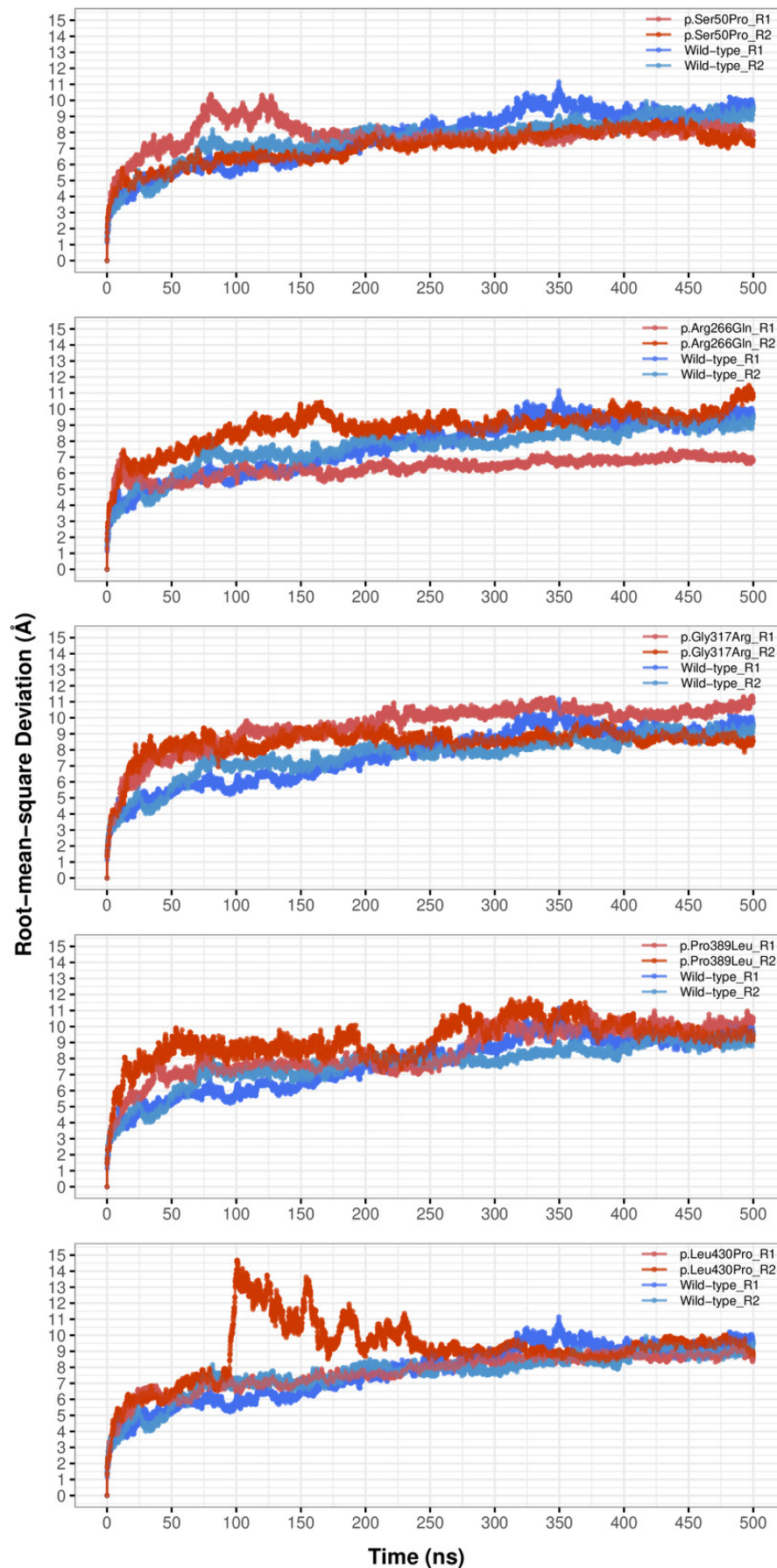
Figure 2

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.
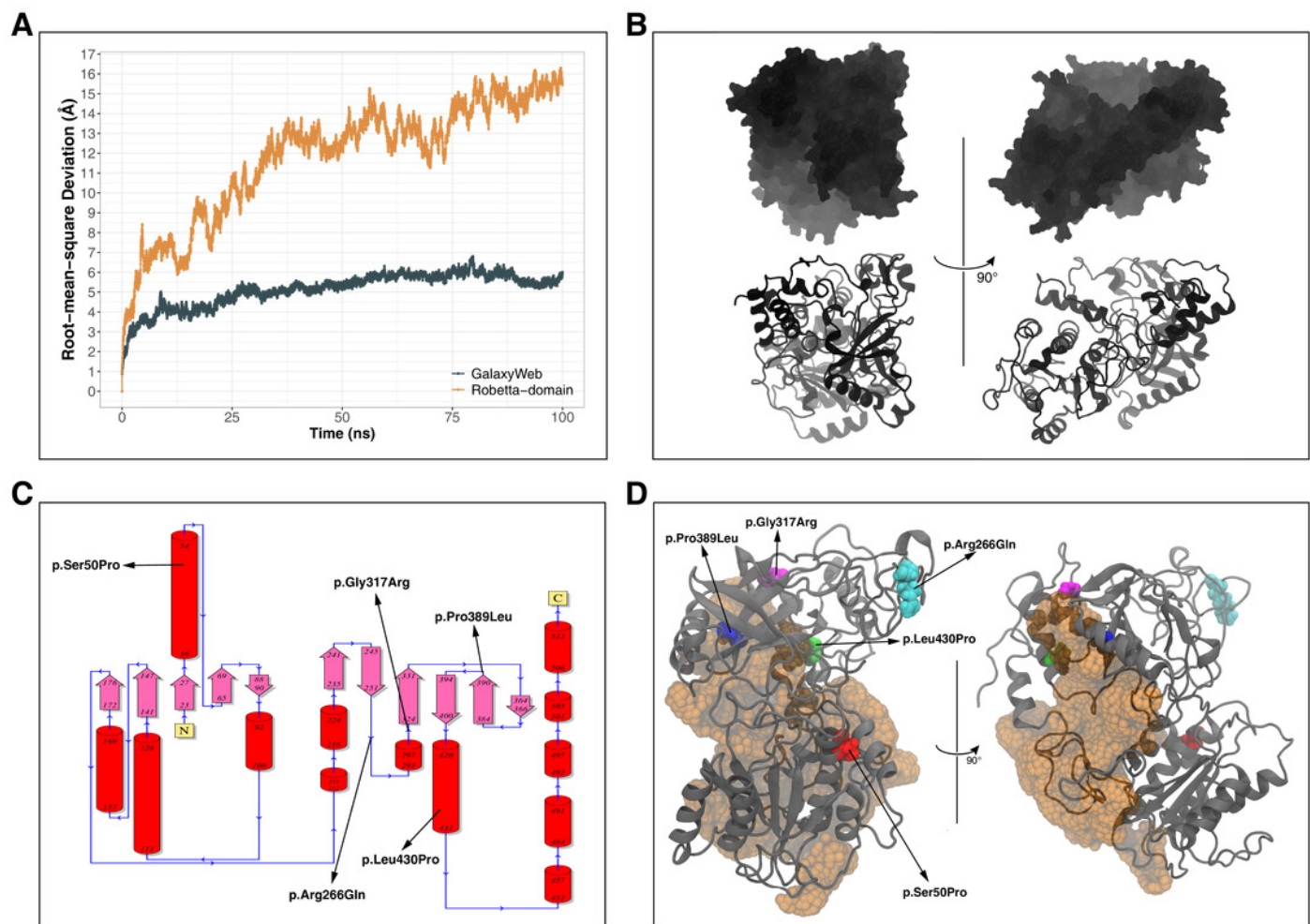
Figure 3

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.
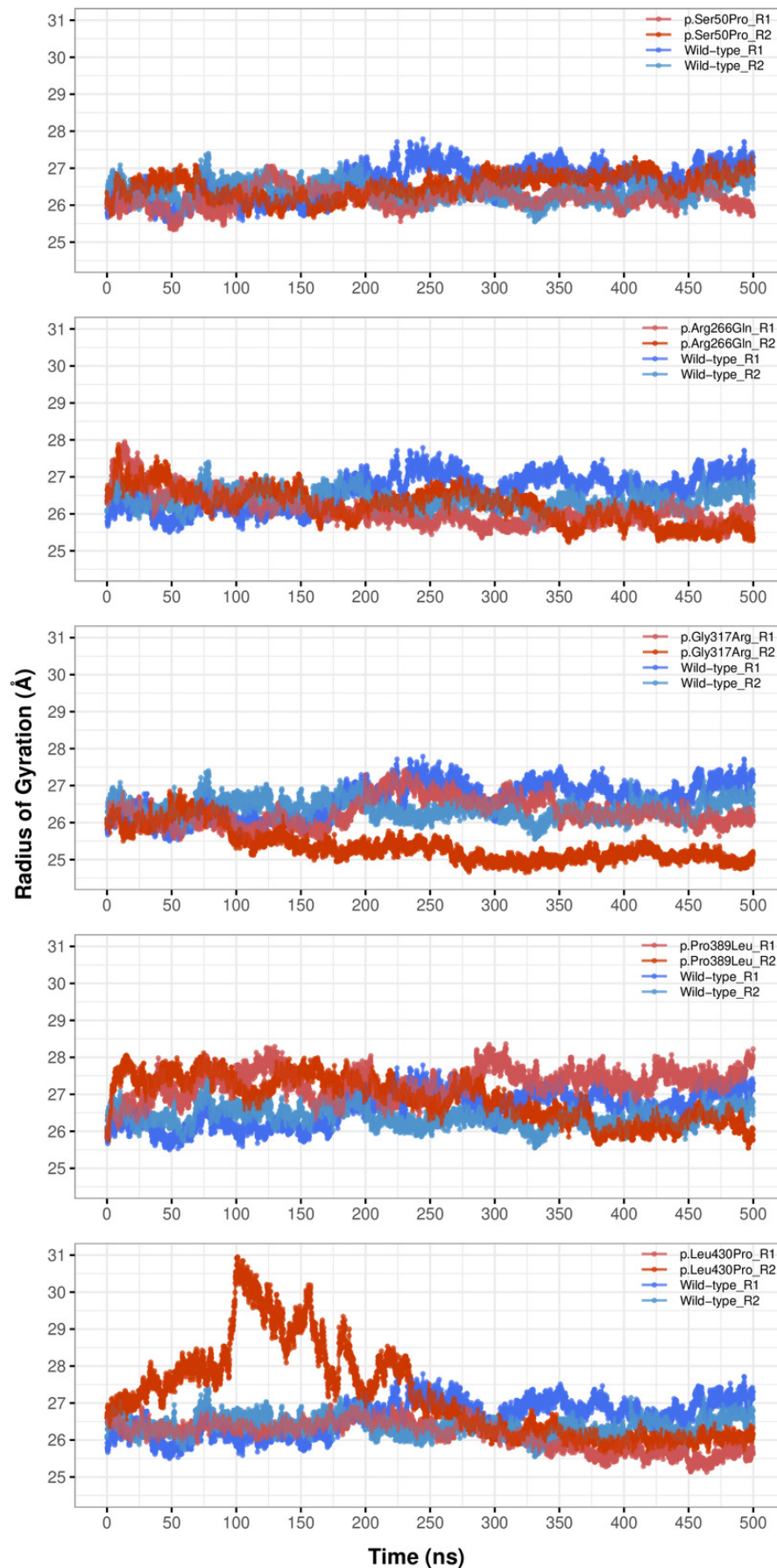
Figure 4

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.
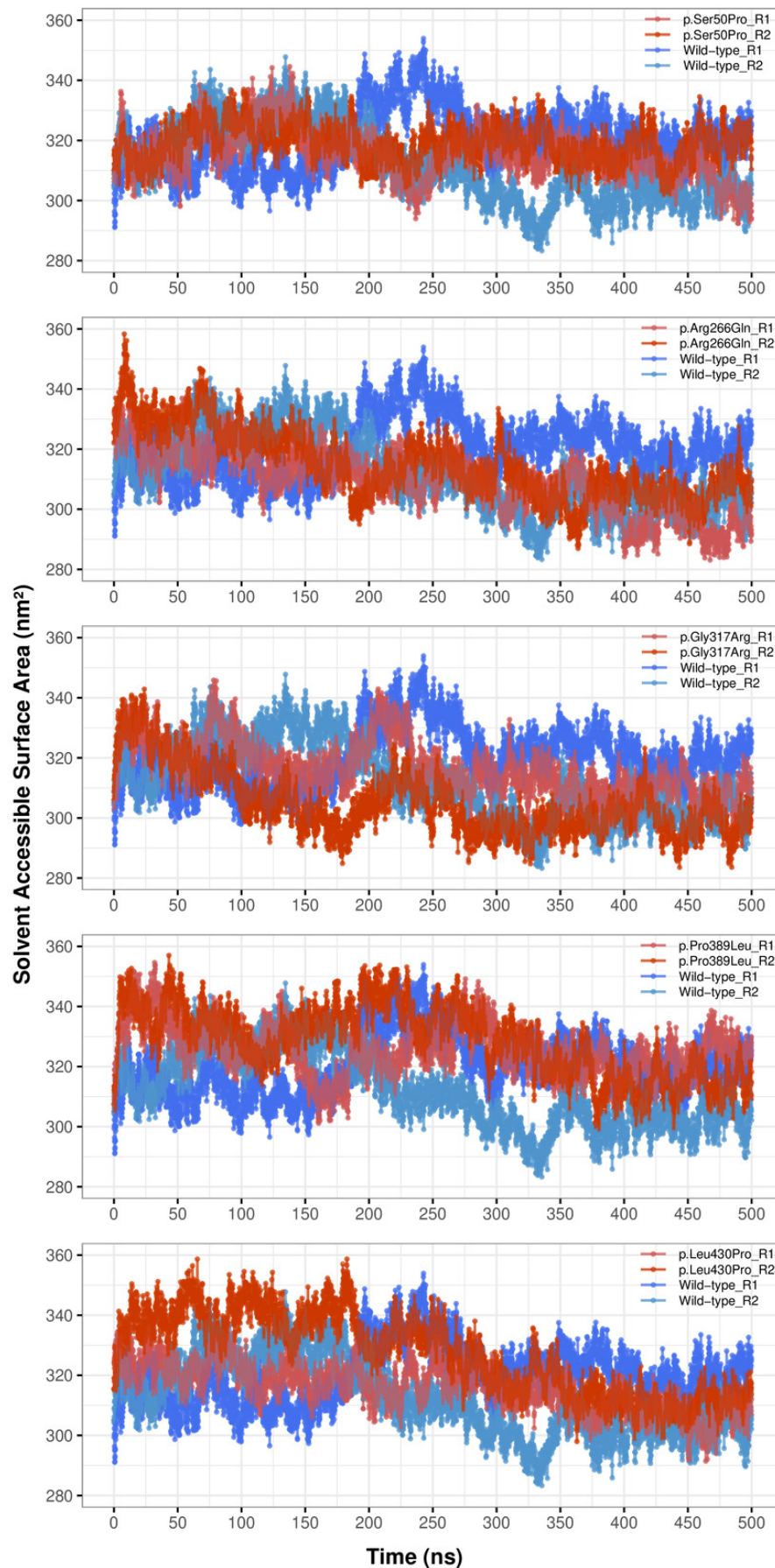
Figure 5

Principal component analysis (PCA).

PCA comparison between molecular dynamics simulations of wild-type and changed M1AP structures. The first two principal components are presented.

Principal Component 2

Principal Component 1
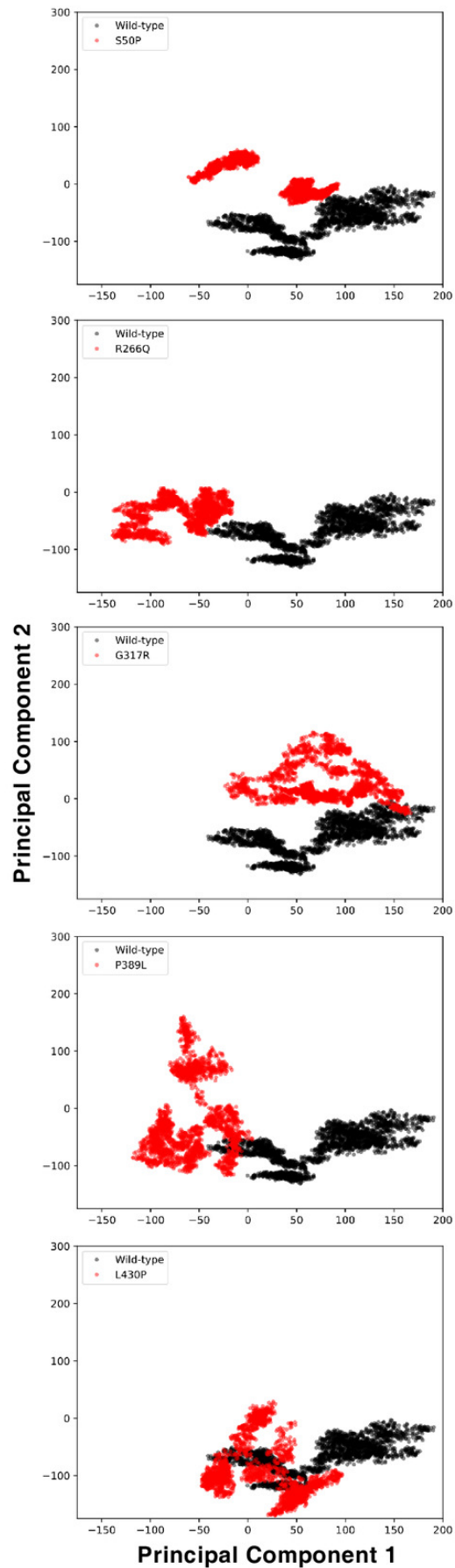
Figure 6

Root-mean-square fluctuations (RMSF) analysis.

Cα 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.
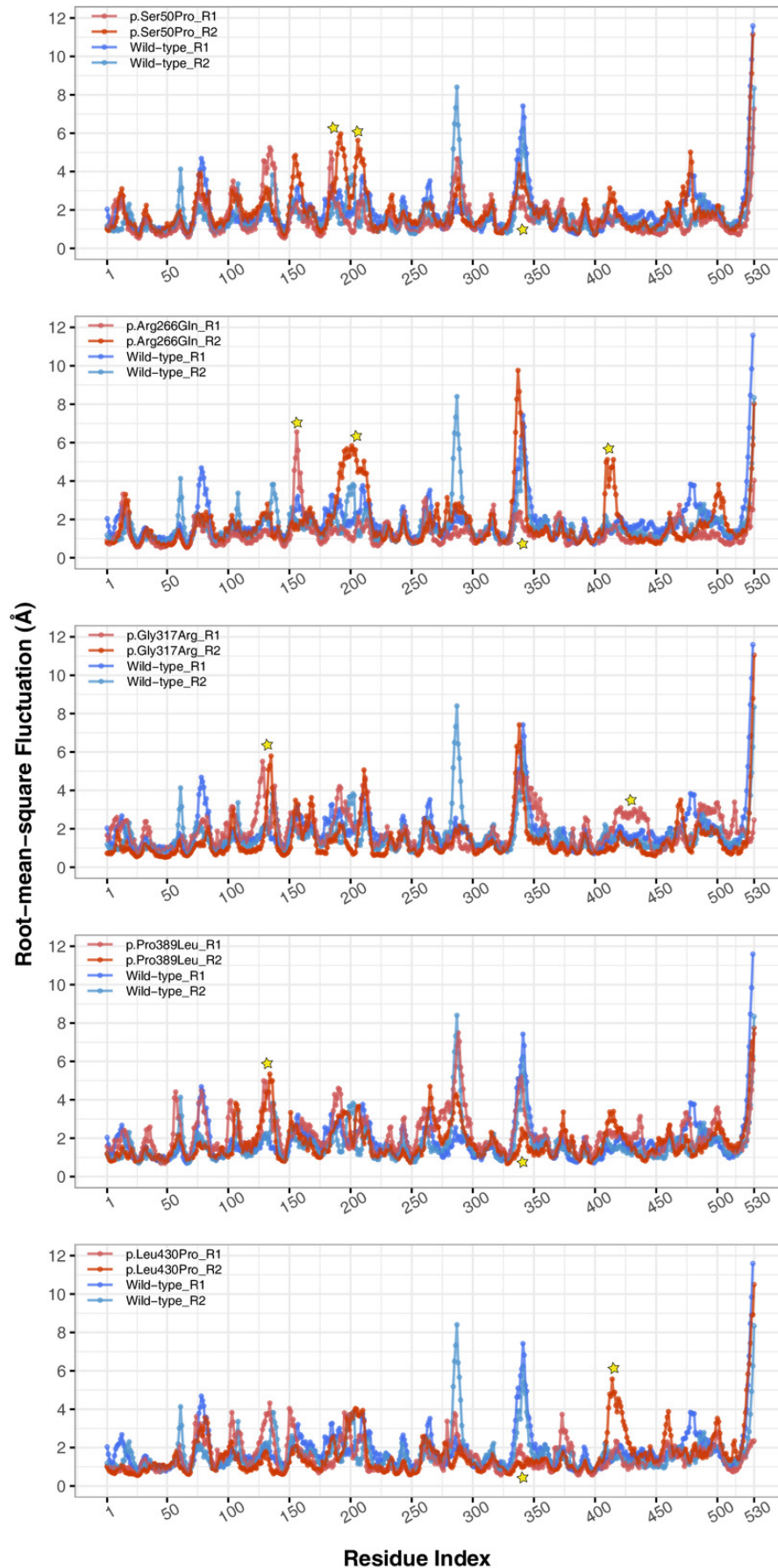
Figure 7

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.

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

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

1 **Table 1: Bi-allelic *M1AP* variants reported in infertile males and predictions for their functional effects by Wyrwoll**

2 ***et al.* (2020).** The missense variants in the table (i.e., p.Ser50Pro, p.Arg266Gln, p.Gly317Arg, p.Pro389Leu and p.Leu430Pro) were

3 examined in this study.

Individual	Bi-allelic M1AP 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

4 † 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

5 missense change were included in this study since our focus is on the impacts of missense variants on the M1AP structure.

6 ‡ Semen contained none or below 10 spermatozoa/sample on repeated analyses.

Quality assessments of M1AP models from five different algorithms.

1 **Table 2. Quality assessments of M1AP models from five different algorithms.**

Model	VERIFY3D Aa% with score ≥ 0.2	ERRAT AOQF%	PROVE BOPA%	PROCHECK Aa% in MFR	Aa% in AAR	Aa% in GAR	Aa% in DR	G-Factor	WHATCHECK # of SCP with Pass	# of SCP with Warning	# of SCP with Error	Overall report
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

2 [1] Aa, aa: amino acid. AOQF: average overall quality factor. BOPA: buried outlier protein atoms. MFR: the most favored regions. AAR: additional allowed regions. GAR: generously
3 allowed regions. DR: disallowed regions. SCP: stereochemical properties.
4 [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.
5 [3] In ERRAT, AOQF ≥ 95% indicates a good quality, AOQF around 91% indicates an average quality, and lower AOQF indicates a bad quality.
6 [4] In PROVE, BOPA ≤ 1% indicates a good quality, BOPA between 1% and 5% indicates an average quality, and BOPA > 5% indicates a bad quality.
7 [5] In the Ramachandran plot analysis of PROCHECK, model has a good quality if overall G-factor is > -0.5 and ≥ 90% of aa in protein are in the most favored regions.
8 [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
9 might vary from structure to structure.