



Article

STAG2: Computational Analysis of Missense Variants Involved in Disease

David Ros-Pardo , Paulino Gómez-Puertas * and Íñigo Marcos-Alcalde

Centro de Biología Molecular Severo Ochoa, Consejo Superior de Investigaciones Científicas-Universidad Autónoma de Madrid (CSIC-UAM), C/Nicolás Cabrera, 1, 28049 Madrid, Spain; davidrp@cbm.csic.es (D.R.-P.); imarcos@cbm.csic.es (Í.M.-A.)

* Correspondence: pagomez@cbm.csic.es; Tel.: +34-91-196-4663

Abstract: The human STAG2 protein is an essential component of the cohesin complex involved in cellular processes of gene expression, DNA repair, and genomic integrity. Somatic mutations in the STAG2 sequence have been associated with various types of cancer, while congenital variants have been linked to developmental disorders such as Mullegama–Klein–Martinez syndrome, X-linked holoprosencephaly-13, and Cornelia de Lange syndrome. In the cohesin complex, the direct interaction of STAG2 with DNA and with NIPBL, RAD21, and CTCF proteins has been described. The function of STAG2 within the complex is still unknown, but it is related to its DNA binding capacity and is modulated by its binding to the other three proteins. Every missense variant described for STAG2 is located in regions involved in one of these interactions. In the present work, we model the structure of 12 missense variants described for STAG2, as well as two other variants of NIPBL and two of RAD21 located at STAG2 interaction zone, and then analyze their behavior through molecular dynamic simulations, comparing them with the same simulation of the wild-type protein. This will allow the effects of variants to be rationalized at the atomic level and provide clues as to how STAG2 functions in the cohesin complex.

Keywords: STAG2; NIPBL; RAD21; molecular modeling; Cornelia de Lange syndrome; variant rationalization



Citation: Ros-Pardo, D.; Gómez-Puertas, P.; Marcos-Alcalde, Í. STAG2: Computational Analysis of Missense Variants Involved in Disease. *Int. J. Mol. Sci.* **2024**, *25*, 1280. <https://doi.org/10.3390/ijms25021280>

Academic Editor: Wolfgang Sippl

Received: 20 December 2023

Revised: 16 January 2024

Accepted: 18 January 2024

Published: 20 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human cohesin complex is a multimeric protein complex involved in key cellular processes such as the regulation of gene expression, DNA repair, and maintenance of 3D structure and genomic integrity. Its major components are the proteins SMC1, SMC3, RAD21 (also known as SCC1), and STAG1/2 (also known as SA-1/2), as well as other associated proteins including NIPBL, PDS5, and WAPL [1–6].

Somatic mutations in coding sequences and regulatory elements of the cohesin genes have been associated with several types of cancer. These include glioblastoma, Ewing's sarcoma, urothelial carcinoma, and myeloid neoplasms. Among the cohesin genes, STAG2 is the most frequently mutated in cancer [7]. Congenital variants in STAG2 have also been described in association with developmental disorders, most notably Mullegama–Klein–Martinez syndrome [8,9], X-linked holoprosencephaly-13 [10], and Cornelia de Lange syndrome [11].

From a structural point of view, X-ray crystallography and cryomicroscopy studies have described the interaction of the STAG1/2 protein surface with RAD21, NIPBL, and DNA, as well as with a segment of the amino-terminal end of the transcriptional regulator CTCF, within a macromolecular complex that also includes SMC1A and SMC3 [12,13].

Although the role of STAG2 as a DNA-binding subunit is well established [14,15], its specific function within the cohesin complex is still unknown. Disease-causing mutations usually exert subtle structural effects that impair the protein function while still allowing some degree of viability. Hence, mechanistic understanding by experimental means is often challenging. In the past, we have successfully used molecular dynamic

(MD) simulations to functionally rationalize variants associated with Cornelia de Lange syndrome in the context of the ATP hydrolysis mechanism at the head of the cohesin complex [16]. In this work, using information from missense mutations that affect the normal function of the protein, we rationalize the effect of such mutations on the interaction of STAG2 with each of the molecules it physically contacts and for which structural information is available: NIPBL, DNA, RAD21, and CTCF. The study of the effect of these mutations on the local environment will help to understand, on the one hand, the possible origin of the disease and, on the other, which of the specific functions are affected, helping to unveil which of these interactions are key for the function of STAG2.

2. Results

The structure of STAG2 and the models of its interaction with RAD21, DNA, and CCTF (Figure 1) were derived from experimental structures 4PJW [17], which contains the crystallized structure of human STAG2 in contact with the human RAD21 structure; 6WG3 [13], which contains the cryo-EM structure of human NIPBL in contact with human RAD21; and 6QNX [12], with the crystallized structure of human STAG2 and RAD21 proteins in contact with a segment of human CCTF. The relative position of the STAG2, RAD21, CCTF, and DNA moieties was inferred through modeling techniques based on the structural information available. The relative positions of NIPBL and STAG2 were located using the 6WG3 structure [13] as a template, which determines the relative position of NIPBL and STAG1. Due to the high similarity between the sequence corresponding to the experimental structures of human STAG1 and STAG2, with an identity percentage of 79.7% and a similarity percentage of 90.4% (Figure 2, orange line), STAG1 scaffold could be used as a template to reliably superimpose STAG2. Finally, to determine the relative position of the DNA molecule in contact with STAG2 and RAD21, the structure 6H8Q [14], which contains the structural homologs of STAG2 and RAD21 in yeast (Scc3 and Scc1, respectively) co-crystallized with a DNA segment, was used as reference, as this is the only available structure of a STAG2 homolog in contact with DNA.

In the present work, using MD techniques, we have simulated STAG2 in a complex with the macromolecules for which structural information on their physical interaction is available: NIPBL, DNA, RAD21, and a segment of the N-terminal end of CTCF. These simulations were performed for wild-type proteins, as well as for a total of 16 missense mutations (Table 1) that were located at the interface between STAG2 and the other complex components: 12 STAG2, 2 NIPBL, and 2 RAD21 variants. Figure 1 shows a model of the interaction of STAG2 with the different molecules in its environment: NIPBL, DNA, RAD21, and CTCF, as well as the position of the amino acids, the variants of which will be studied. All STAG2 variants are located in highly conserved segments (Figure 2).

The variants have been divided into four groups according to the interface they are closest to (Figure 1): STAG2/NIPBL interaction, STAG2/DNA interaction, STAG2/RAD21 interaction, and STAG2/CTCF interaction. Each of the variants, as well as the corresponding controls, have been subjected to a 200 ns long molecular dynamics simulation, the stability of which has been assessed by measuring the Root Mean Square Deviation (RMSD, Supplementary Figures S1–S3).

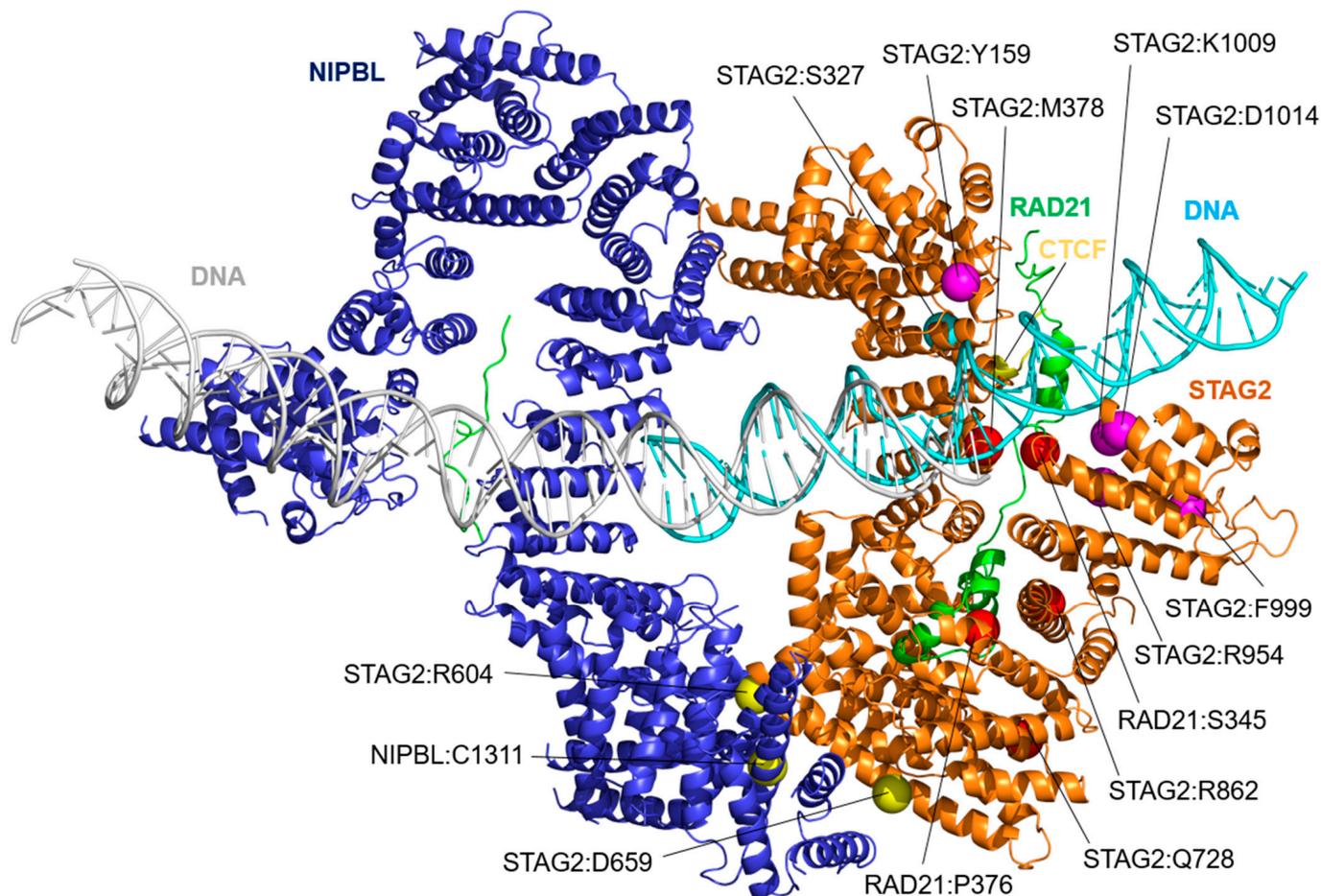


Figure 1. Three-dimensional model showing the relative position of STAG2 (orange) interacting with NIPBL (blue), RAD21 (green), a segment of the N-terminal domain of CTCF (yellow), and a fragment of a DNA molecule (cyan). The model is based on the Protein Data Bank structures 6WG3 [13], 4PJW [17], 6H8Q [14], and 6QNX [12]. The DNA molecule in gray corresponds to the one included in the 6WG3 structure in contact with the NIPBL molecule. Although this interaction was not analyzed in the present work, it has been included in the figure to illustrate its position. The amino acid positions of STAG2, NIPBL, and RAD21, the variants of which are analyzed, are indicated by solid spheres. The interface group they have been assigned to is indicated by the sphere color: STAG2/NIPBL interaction (yellow spheres): STAG2:Arg604Gln (R604), STAG2:Asp659His (D659), NIPBL:Cys1311Arg (C1311), and NIPBL:Cys1311His (C1311); STAG2/DNA interaction (magenta spheres): STAG2:Tyr159Cys (Y159), STAG2:Tyr159His (Y159), STAG2:Arg954Cys (R954), STAG2:Phe999Cys (F999), STAG2:Lys1009Asn (K1009), and STAG2:Asp1014Asn (D1014); STAG2/RAD21 interaction (red spheres): STAG2:Met378Val (M378), STAG2:Gln728His (Q728), STAG2:Arg862Gly (R862), RAD21:Ser345Pro (S345), and RAD21:Pro376Arg (P376); and STAG2/CTCF interaction (cyan sphere): STAG2:Ser327Asn (S327).

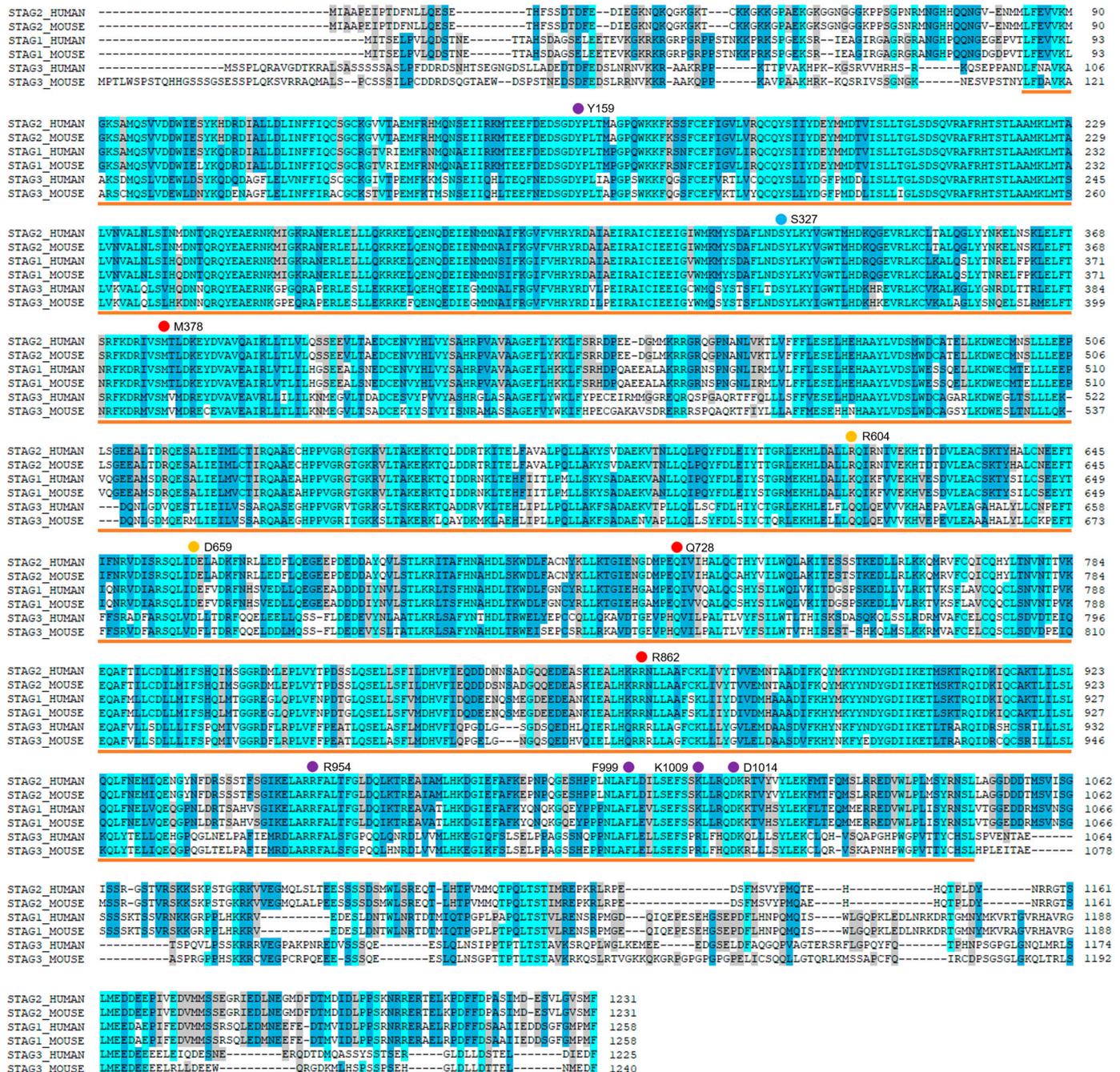


Figure 2. Multiple sequence alignment of mouse and human STAG1, STAG2, and STAG3 proteins (Uniprot entries: Q8WVM7, Q8N3U4, Q9UJ98, Q9D3E6, O35638, and O70576, respectively). The dots indicate the position of the human STAG2 amino acids that were analyzed, colored according to the interface to which each is associated: STAG2/NIPBL interaction (yellow dots): STAG2:Arg604Gln (R604) and STAG2:Asp659His (D659); STAG2/DNA interaction (magenta dots): STAG2:Tyr159Cys (Y159), STAG2:Tyr159His (Y159), STAG2:Arg954Cys (R954), STAG2:Phe999Cys (F999), STAG2:Lys1009Asn (K1009), and STAG2:Asp1014Asn (D1014); STAG2/RAD21 interaction (red dots): STAG2:Met378Val (M378), STAG2:Gln728His (Q728), and STAG2:Arg862Gly (R862); and STAG2/CTCF interaction (cyan dots): STAG2:Ser327Asn (S327). Sequences are colored according to their evolutionary conservation. The STAG2 sequence segment corresponding to the 3D structure is marked with an orange line.

Table 1. Missense variants analyzed.

Interaction	Variant	References
STAG2/NIPBL	STAG2:Arg604Gln	[8,18]
	STAG2:Asp659His	[11]
	NIPBL:Cys1311Arg	[19]
	NIPBL:Cys1311His	ClinVar VCV001697856.1 [20]
STAG2/DNA	STAG2:Tyr159Cys	[18]
	STAG2:Tyr159His	[21]
	STAG2:Arg954Cys	[22]
	STAG2:Phe999Cys	ClinVar VCV001395977.2 [20]
	STAG2:Lys1009Asn	[9]
	STAG2:Asp1014Asn	[23]
STAG2/RAD21	STAG2:Met378Val	[11]
	STAG2:Gln728His	ClinVar VCV001299204.12 [20]
	STAG2:Arg862Gly	ClinVar VCV000521058.3 [20]
	RAD21:Ser345Pro	[24]
	RAD21:Pro376Arg	[24,25]
STAG2/CTCF	STAG2:Ser327Asn	[26]

Variants analyzed, grouped according to their proximity to each interacting molecule.

2.1. STAG2/NIPBL Interaction

Four variants are included in this group, two in the STAG2 sequence (STAG2:Arg604Gln and STAG2:Asp659His), and two others in the NIPBL sequence (NIPBL:Cys1311Arg and NIPBL:Cys1311Tyr), all four of which are located at the interface between the two proteins. In the case of the STAG2:Asp659His variant, Figure 3 shows an image of the local structure after 200 nanoseconds (ns) of productive MD simulation of both the wild-type protein (STAG2:Asp659) and the STAG2:Asp659His variant. In the wild-type protein, the negatively charged amino acid STAG2:Asp659 was observed to interact stably with STAG2:Lys705 and, somewhat less stably throughout the simulation, with NIPBL:Lys1247. This interaction contributes to the correct localization of the STAG2 and NIPBL surfaces in contact with each other (Figure 3A). In the case of the STAG2:Asp659His variant, the substitution of Asp659 by a positively charged amino acid, such as His, causes a number of structural changes (Figure 3B; Supplementary Figure S3). Due to their positive charge, NIPBL:Lys1247 and STAG2:Lys705 no longer interact with STAG2:Asp659His and move to the surrounding area. This fact allows an amino acid that was previously far away from this zone, NIPBL:Glu1240, to change its position and come into contact with STAG2:Asp659His. This maintains the local interaction between the surfaces of both proteins but at the cost of a disorganization of the alpha-helix containing STAG2:Asp659His, which is split into two smaller alpha-helices, and a change in the local relative position of NIPBL with respect to STAG2. Furthermore, the replacement of two salt bridges by a single polar interaction can be expected to significantly weaken the interaction between both proteins. The average distance between the amino acids in the NIPBL helix containing the NIPBL:Lys1247 residue and the STAG2 helix containing the STAG2:Asp659His variant changes from a value of 11.06 ± 1.24 Angstrom (\AA) for the wild-type protein over the 200 ns MD trajectory to a value of 27.71 ± 1.32 \AA in the last 50 ns of the MD trajectory for the STAG2:Asp659His variant. Therefore, the described effect may explain, on the one hand, the deleterious effect of the variant by the noticeable interface disruption, but, on the other, may also explain the partial retention of its function due to a new contact between STAG2:Asp659His and NIPBL:Glu1240.

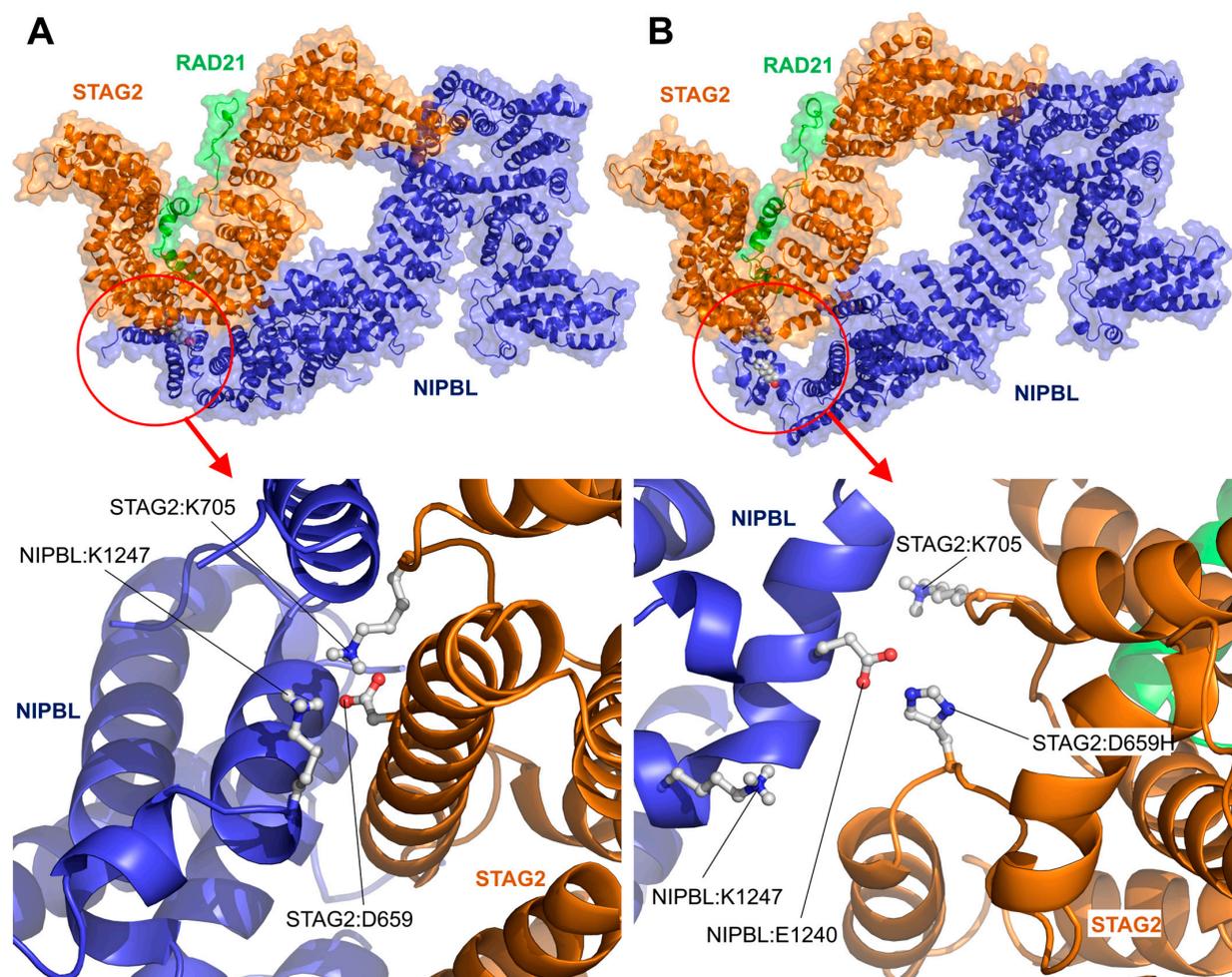


Figure 3. STAG2/NIPBL interaction: STAG2:Asp659His variant. Structure of the interaction interface between NIPBL and STAG2 in presence of the wild-type STAG2:Asp659 residue (**A**) and variant STAG2:Asp659His (**B**) after a 200 ns long MD simulation. (**Upper panels**) Structure of the STAG2-NIPBL complex after MD simulation. (**Bottom panels**) Zoom on the area where the STAG2:Asp659 residue is located on the interaction surface between the two proteins. (**A**) The negatively charged residue STAG2:Asp659 (D659) forms a stable interaction with the positively charged amino acids STAG2:Lys705 (K705) and NIPBL:Lys1247 (K1247). (**B**) The presence of a histidine at the position corresponding to the Asp659 residue in the STAG2:Asp659His (D659H) variant causes the separation of the positive amino acids STAG2:Lys705 (K705) and NIPBL:Lys1247 (K1247) from their original position. Instead, STAG2:Asp659His binds to the residue NIPBL:Glu1240 (E1240). This structural change causes the noticeable disorganization of the neighboring structures and replaces two salt-bridges with a single polar interaction, which can be expected to significantly weaken the interaction between NIPBL and STAG2.

In the STAG2:Arg604Gln variant, no significant changes in the local environment were observed during the MD simulation, with respect to the wild-type condition (Supplementary Figures S1 and S2). STAG2:Arg604 does not seem to be involved in the formation of salt bridges with charged amino acids of NIPBL, so the exchange of STAG2:Arg604 for Gln (a polar but uncharged amino acid) does not seem to promote significant changes. Both STAG2:Arg604 and STAG2:Arg604Gln appear to interact with the same amino acids and in a similar geometry. The deleterious effect might be caused by the non-ionic character of glutamine in comparison with aspartic acid, resulting in weaker interactions with the neighboring polar residues, by possible problems during protein folding, by effects that

involve other proteins not described in the model, or by any other effect beyond the scope of this work.

Two variants have been described in the NIPBL:Cys1311 position: NIPBL:Cys1311Arg and NIPBL:Cys1311Tyr. In the wild-type protein, the amino acid NIPBL:Cys1311 is localized in one of the HEAT repeat structures of NIPBL that is close to but not in contact with STAG2 (Figure 4A). Near NIPBL:Cys1311, the residue NIPBL:Lys1389 is located in direct contact with STAG2:Asn648 on the surface of STAG2, contributing to the stability of the interaction between the two proteins. In the NIPBL:Cys1311Arg variant (Figure 4B), after MD simulation, the amino acid NIPBL:Lys1389 is slightly shifted due to the positioning of the mutated amino acid NIPBL:Cys1311Arg in its place, which also interacts with the STAG2:Asn648 residue. The new localization of NIPBL:Cys1311Arg facilitates the approach of an amino acid that was not previously in the environment: the negatively charged STAG2:Asp651. Thus, the variant seems to exacerbate the interaction between NIPBL and STAG2 due to the electrostatic attraction between the face of the alpha helices containing the positively charged amino acids NIPBL:Lys1389 and NIPBL:Cys1311Arg and the negatively charged amino acid STAG2:Asp651. This change is indeed accompanied by a shift of the alpha-helices containing NIPBL:Lys1389 and NIPBL:Cys1311Arg towards the position occupied by STAG2:Asp651. In the case of the NIPBL:Cys1311Tyr variant (Figure 4C), although the interaction between NIPBL:Cys1311Tyr (polar but uncharged) and STAG2:Asp651 is not established, a large displacement is observed, not only of the helices containing the amino acids NIPBL:Lys1389 and NIPBL:Cys1311Tyr, but also of a third NIPBL alpha-helix (named a3 in Figure 4C), which shifts out of the local environment during the simulation, causing a large structural change at the contact surface between the two proteins. The average distance between the amino acids in helices a1 and a2, containing residues NIPBL:Cys1311 and NIPBL:Lys1389, the STAG2 helix, and the amino acid STAG2:Asp651, varies from a mean of $18.18 \pm 0.62 \text{ \AA}$ over the 200 ns MD trajectory of the wild-type protein to mean values of $12.23 \pm 1.83 \text{ \AA}$ and $13.85 \pm 1.44 \text{ \AA}$ over the last 50 ns of the MD trajectory for the NIPBL:Cys1311Arg and NIPBL:Cys1311Tyr variants, respectively. The structural changes induced by the STAG2:Asp659His, NIPBL:Cys1311Arg, and NIPBL:Cys1311Tyr variants are associated with high RMSD values measured in the corresponding regions (Supplementary Figure S3).

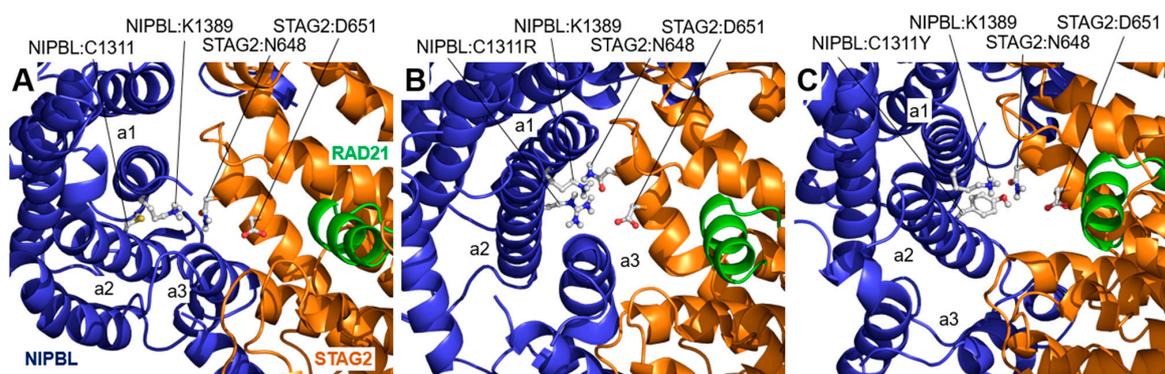


Figure 4. STAG2/NIPBL interaction: NIPBL:Cys1311Arg and NIPBL:Cys1311Tyr variants. Interface between wild type NIPBL and STAG2 in the vicinity of the amino acid NIPBL:Cys1311 (C1311) after a 200 ns long MD simulation. (A) In the wild-type interface, close to the NIPBL:Cys1311 residue (located in an alpha-helix in the HEAT repeat structure, marked (a2)), the amino acid NIPBL:Lys1389 (K1389) (located in the adjacent helix, marked (a1)) interacts with the amino acid STAG2:Asn648 (N648), contributing to the stability of the local structure of the interface between NIPBL and STAG2. (B) In the NIPBL:Cys1311Arg (C1311R) variant, the amino acid NIPBL:Arg1311 interacts with the amino acids STAG2:Asn648 (N648) and STAG2:Asp651 (D651), which causes the displacement of the (a1)- and (a2)-labeled helices of NIPBL during the MD simulation, locally disorganizing the interface structure. (C) A similar effect is observed for the NIPBL:Cys1311Tyr (C1311Y) variant, where the structural change is enhanced by the large displacement of a third helix of the HEAT repeat structure of NIPBL, labeled (a3).

2.2. STAG2/DNA Interaction

Six STAG2 variants are included in this group. Two of them correspond to a mutation in the same residue: STAG2:Tyr159Cys and STAG2:Tyr159His (Figure 5), while the other four are grouped on the other side of the DNA: STAG2:Arg954Cys, STAG2:Phe999Cys, STAG2:Lys1009Asn, and STAG2:Asp1014Asn (Figure 6). The amino acid STAG2:Tyr159 localizes to a region of STAG2 in contact with DNA, although the amino acid itself has no direct interaction with the nucleotide chain (Figure 5A). In the STAG2:Tyr159Cys and STAG2:Tyr159His variants, the amino acid changes are not particularly drastic because both the original amino acid (Tyr) and the variants (Cys and His) are all polar residues. However, the small changes associated with these substitutions seem to correspond to a greater flexibility in the local interaction between the protein moiety located close to the mutated amino acid and the DNA molecule. Figure 5B,C show an example of one of the separation moments between the DNA and the STAG2 domain during the simulation corresponding to the STAG2:Tyr159Cys and STAG2:Tyr159His variants, respectively. Figure 5D shows the variation, along the MD trajectory, of the distance between the alpha carbon of the STAG2:Tyr159 amino acid (or of the STAG2:Tyr159Cys and STAG2:Tyr159His variants) and the nearest phosphate group of the DNA chain at the beginning of the simulation. The graph shows how this distance remains within the range of 9 to 17 Angstrom for the case of the wild-type amino acid. On the other hand, in the two mutants, a much larger oscillation is observed, reaching values of up to 30 Å at some moments of the simulations (mean values of 19.69 ± 3.05 and 17.00 ± 3.44 Å for STAG2:Tyr159Cys and STAG2:Tyr159His variants, respectively, in the last 50 ns of the trajectory).

The four variants STAG2:Arg954Cys, STAG2:Phe999Cys, STAG2:Lys1009Asn, and STAG2:Asp1014Asn are located in region of STAG2 that is in close contact with the DNA double strand (Figure 6A). As in the case of the STAG2:Tyr159Cys and STAG2:Tyr159His mutants, no significant structural change in the surrounding domain is observed in these four variants. In the case of the STAG2:Arg954Cys, STAG2:Lys1009Asn, and STAG2:Asp1014Asn variants, the subtlety of the effect may be due to the resulting amino acid substitutions retaining the polar character (Cys, Asn, and Asn). Nevertheless, none of these substitutions retain the ionic nature of the wild-type amino acids, which might significantly weaken the interaction with the DNA. On the other hand, STAG2:Phe999 is buried in the domain structure within a hydrophobic cluster, yet the switch to a polar amino acid (Cys) does not seem to cause a large local change in the 3D structure of the backbone, probably because of the smaller size of STAG2:Phe999Cys; however, some effects during protein folding can be expected given the change from a non-polar to polar amino acid in the context of a hydrophobic pocket.

In contrast, STAG2:Arg954Cys, STAG2:Phe999Cys, STAG2:Lys1009Asn, and STAG2:Asp1014Asn exhibit an increased variability in the distance between the STAG2 region and the DNA molecule (Figure 6), similarly to what was observed on STAG2:Tyr159His. In the wild-type protein, the contact between the domain and the DNA is stable, moving around values close to 5 Å towards the end of the MD trajectory. In contrast, in the case of the four mutants, this distance is much more unstable, oscillating between values of 10 and 20 Å (mean values of 13.60 ± 1.36 , 12.75 ± 1.42 , 11.86 ± 2.15 , and 13.70 ± 1.88 Å for STAG2:Arg954Cys, STAG2:Phe999Cys, STAG2:Lys1009Asn, and STAG2:Asp1014Asn variants, respectively).

Considering that the main function described for STAG2 is specifically to bind to the DNA molecule, these variations in the interaction between the two molecules observed in the six variants studied are very likely to affect the stability and geometry of the interaction and its associated function.

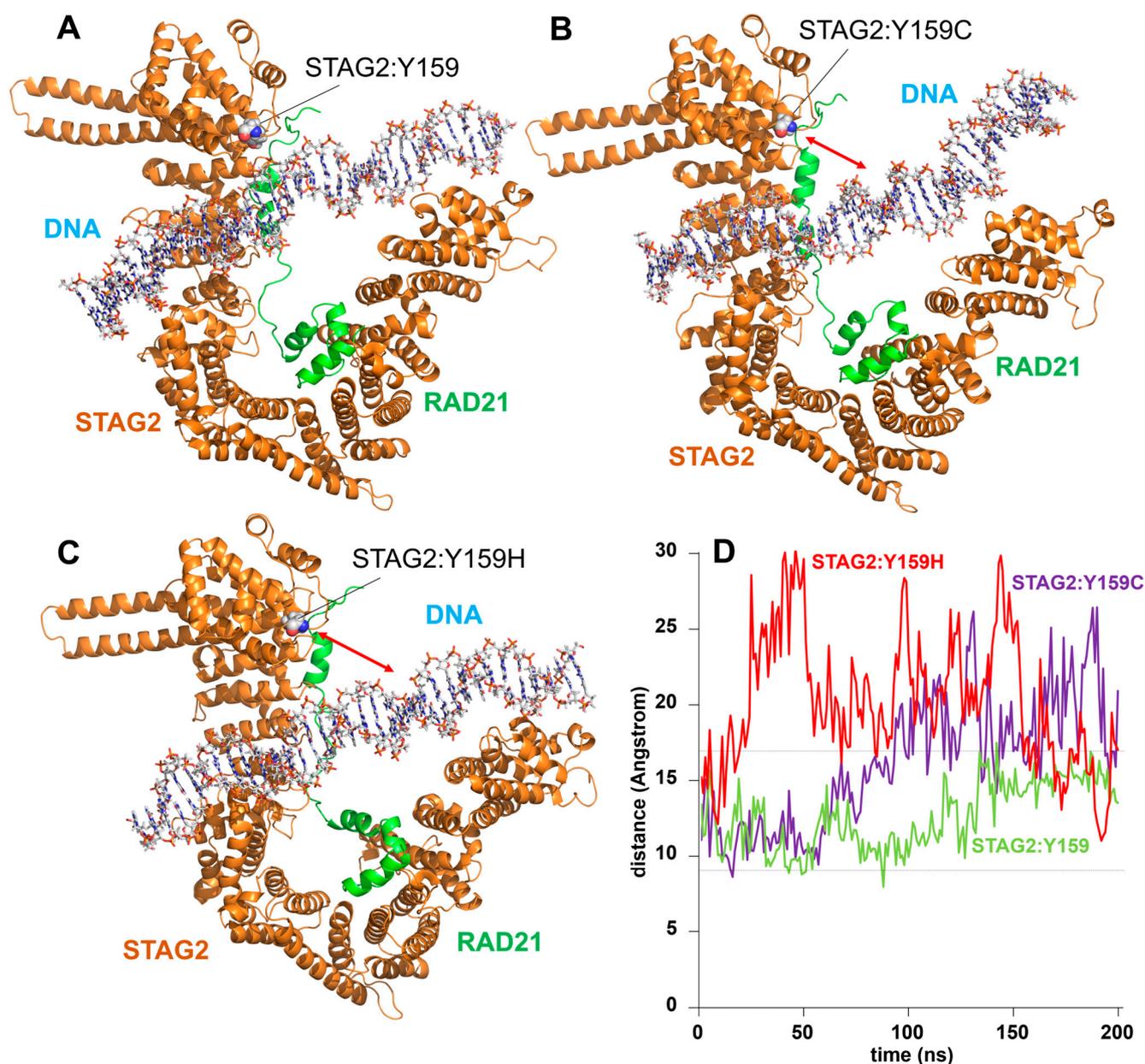


Figure 5. STAG2/DNA interaction: STAG2:Tyr159Cys and STAG2:Tyr159His variants. Model of the interaction of the STAG2-RAD21 dimer with a DNA molecule after 200 ns of MD simulation. (A) In the wild-type dimer, the position of the STAG2:Tyr159 (Y159) amino acid in a domain of STAG2 contacting the DNA molecule is indicated. (B) Model of the interaction of STAG2-RAD21 dimer with DNA in the case of the STAG2:Tyr159Cys (Y159C) variant. The red arrow indicates the separation between the DNA molecule and the STAG2 domain. (C) Model of the interaction of STAG2-RAD21 dimer with DNA in the case of the STAG2:Tyr159His (Y159H) variant. (D) Plot showing the variation of the distance from the alpha carbon of STAG2:Tyr159 (Y159, green line), STAG2:Tyr159Cys (Y159C, purple line), and STAG2:Tyr159His (Y159H, red line) to the nearest DNA phosphate group over 200 ns of unrestricted MD simulation. The distance in the case of the wild-type molecule ranges between 9 and 17 Å (dashed lines), whereas, in the case of both variants, the oscillation is much larger, reaching separation values of up to 30 Å.

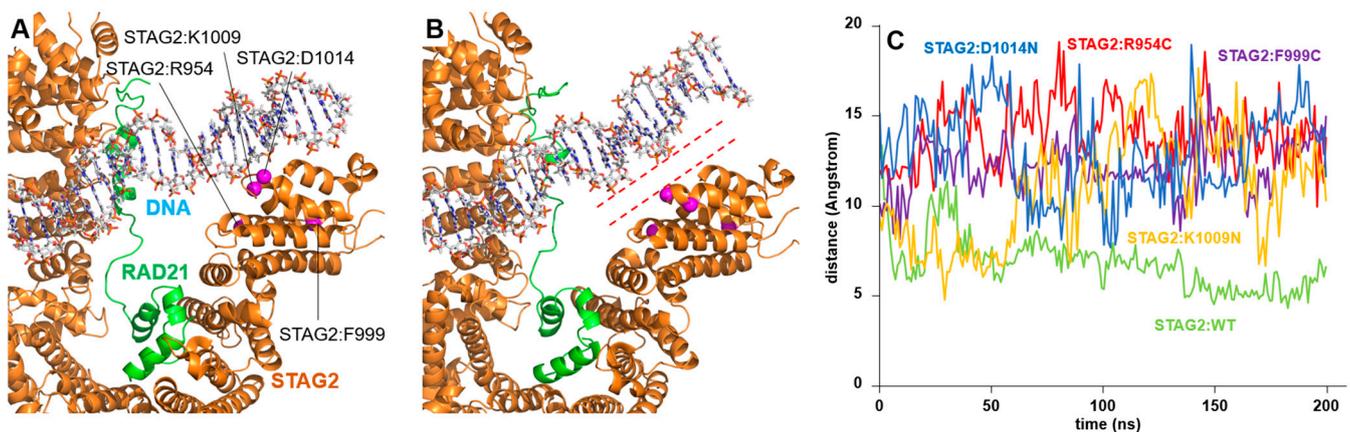


Figure 6. STAG2/DNA interaction: STAG2:Arg954Cys, STAG2:Phe999Cys, STAG2: Lys1009Asn, and STAG2:Asp1014Asn variants. Model of the interaction of the STAG2-RAD21 dimer with a DNA molecule after 200 ns of MD simulation. (A) The position of residues STAG2:Arg954 (R954), STAG2:Phe999 (F999), STAG2: Lys1009 (K1009), and STAG2:Asp1014 (D1014) located in a DNA contacting domain of wild-type STAG2-RAD21 dimer are shown. (B) Model of the interaction between DNA and the STAG2-RAD21 dimer in the case of the STAG2:Arg954Cys (R954C) variant during MD simulation. The dashed lines indicate the separation between the DNA strand and the STAG2 domain where the variants are located. (C) Plot showing the variation of the distance between the alpha carbon of a residue in the contact surface of STAG2 and the nearest DNA phosphate group over 200 ns of unrestricted MD simulation in the following conditions: wild-type STAG2 (green line), STAG2:Arg954Cys (R954C, red line), STAG2:Phe999Cys (F999C, purple line), STAG2: Lys1009Asn (K1009N, yellow line), and STAG2:Asp1014Asn (D1014N, blue line). The distance in the case of the wild-type molecule stabilizes at a position close to 5 Å; the analysis of the four variants shows that the distance oscillates strongly and reaches values of almost 20 Å at certain times.

2.3. STAG2/RAD21 Interaction

Three STAG2 variants (STAG2:Met378Val, STAG2:Gln728His, and STAG2:Arg862Gly) and two RAD21 variants (RAD21:Ser345Pro and RAD21:Pro376Arg) belong to this group. Two of them (STAG2:Met378Val and RAD21:Ser345Pro) localize very close to each other on the interaction surface between RAD21 and STAG2 (Figure 7A). Figure 7B shows the effect of the STAG2:Met378Val variant on the interaction between STAG2 and RAD21. Although the change from STAG2:Met378 to Val is conservative, as are both hydrophobic amino acids, the difference in size between the two is likely to induce a change in local structure that is transmitted to the neighboring RAD21 protein. In RAD21, the alpha helix contacting STAG2 at this site is displaced and undergoes a change in morphology, resulting in a separation between the two surfaces, from a mean value of 6.82 ± 0.49 Å in the wild-type trajectory to a mean distance of 10.65 ± 2.04 in the last 50 ns of the STAG2:Met378Val variant trajectory, as shown in Figure 7B (bottom). In the case of the RAD21:Ser345Pro variant, located in a position close to the residue STAG2:Met378, the non-conservative substitution of RAD21:Ser345 by Pro causes a conformational change in the RAD21 alpha-helix where the mutated amino acid is located (Figure 7C, Supplementary Figure S3). As in the case of the STAG2:Met378Val variant, the RAD21:Ser345Pro mutation results in the surface separation of the two proteins (mean distance of 9.02 ± 1.09 Å in the last 50 ns of the trajectory). In both cases, this separation leads to a change in the stability of the coupling between the two molecules, affecting both the formation of the RAD21–STAG2 dimer, which is necessary for the complex to work properly, and the potential interaction of STAG2 with the DNA molecule.

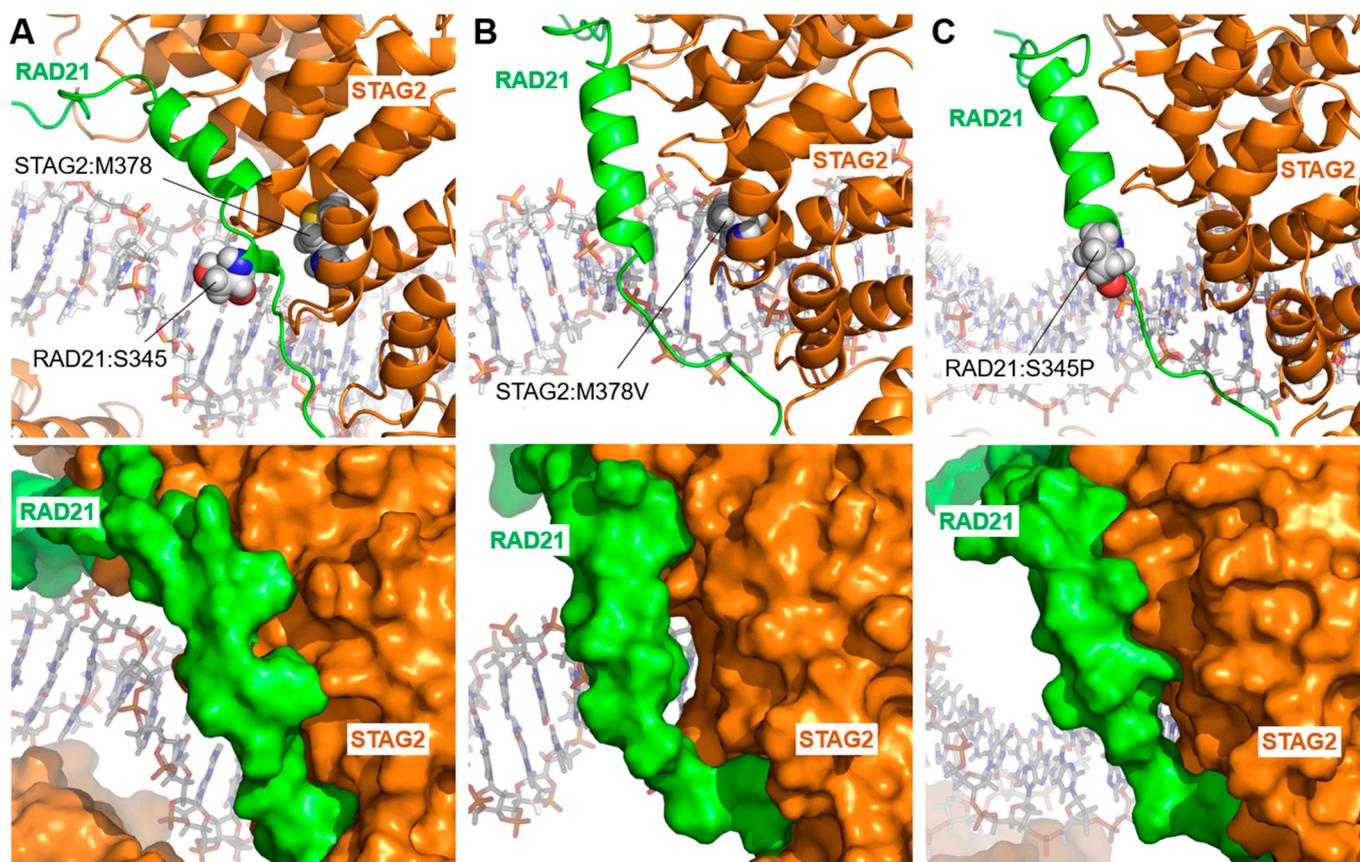


Figure 7. STAG2/RAD21 interaction: STAG2:Met378Val and RAD21:Ser345Pro variants. Interface between STAG2 and RAD21 after 200 ns of unconstrained MD simulation. (A) (Upper panel) position of the amino acids STAG2:Met378 (M378) and RAD21:Ser345 (S345) at the interface between wild-type STAG2 and RAD21. (Lower panel) the same structure as in the upper panel, illustrating the surface of the proteins, showing the close contact between the two molecules. (B) Structure of the STAG2-RAD21 interface for the STAG2:Met378Val (M378V) variant. A change in the structure of the adjacent RAD21 alpha-helix (upper panel) is observed, resulting in a separation between the surfaces of the two proteins (lower panel). (C) Structure of the STAG2-RAD21 interface in the case of the RAD21:Ser345Pro (S345P) variant. In this case, there is also a change in the structure of the RAD21 alpha-helix where the variant is localized (upper panel), forcing a separation at the interface of the two proteins (lower panel).

Figure 8A shows the position of the other three variants (STAG2:Gln728His, STAG2:Arg862Gly, and RAD21:Pro376Arg) in a position close to each other on the interaction surface between STAG2 and RAD21, although far from the area where the two variants discussed above are located. Figure 8B–D show the domain structure after the MD simulation of STAG2:Gln728His, STAG2:Arg862Gly, and RAD21:Pro376Arg, respectively. The figure shows how, although some of the changes are clearly non-conservative, no major shifts in the relative position of the two proteins are apparent. As in the case of the STAG2:Arg604Gln variant mentioned above, the effect of the mutations may have to be explained by the altered interaction with components not present in this model, possible problems during protein folding, or other effects out of the scope of the methodology applied in this work.

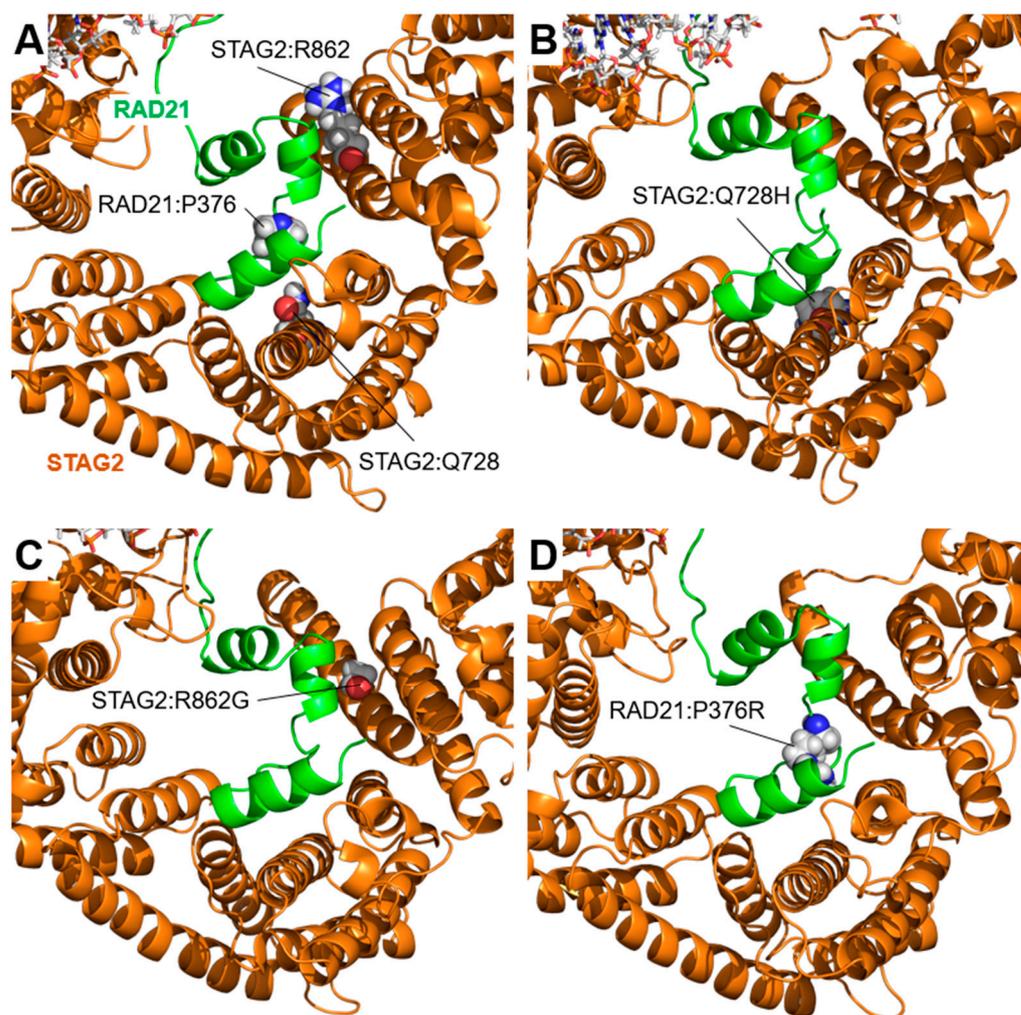


Figure 8. STAG2/RAD21 interaction: STAG2:Gln728His, STAG2:Arg862Gly, and RAD21:Pro376Arg variants. Model of the interaction between STAG2 and RAD21 after 200 ns of MD simulation. (A) The positions of the residues STAG2:Gln728 (Q728), STAG2:Arg862 (R862), and RAD21:Pro376 (P376) in the interface between wild-type STAG2 and RAD21 are indicated. (B–D). Models of the interaction between STAG2 and RAD21 corresponding to the variants STAG2:Gln728His (Q728H, panel (B)), STAG2:Arg862Gly (R862G, panel (C)), and RAD21:Pro376Arg (P376R, panel (D)). No major changes were observed in any of the three variants analyzed.

2.4. STAG2/CTCF Interaction

The interaction segment of CTCF with STAG2 consists of only nine amino acids located in the N-terminal domain of the protein [12]. Only one of the studied STAG2 variants, STAG2:Ser327Asn, is located near this binding site. Figure 9 shows the position of the STAG2-bound CTCF segment in the wild-type protein (left) and in the STAG2:Ser327Asn variant (right). Although no significant structural change is observed when replacing Ser with Asn (both are uncharged polar residues), the geometry of the CTCF segment interaction changes slightly during the MD simulation, moving away from the contact with RAD21.

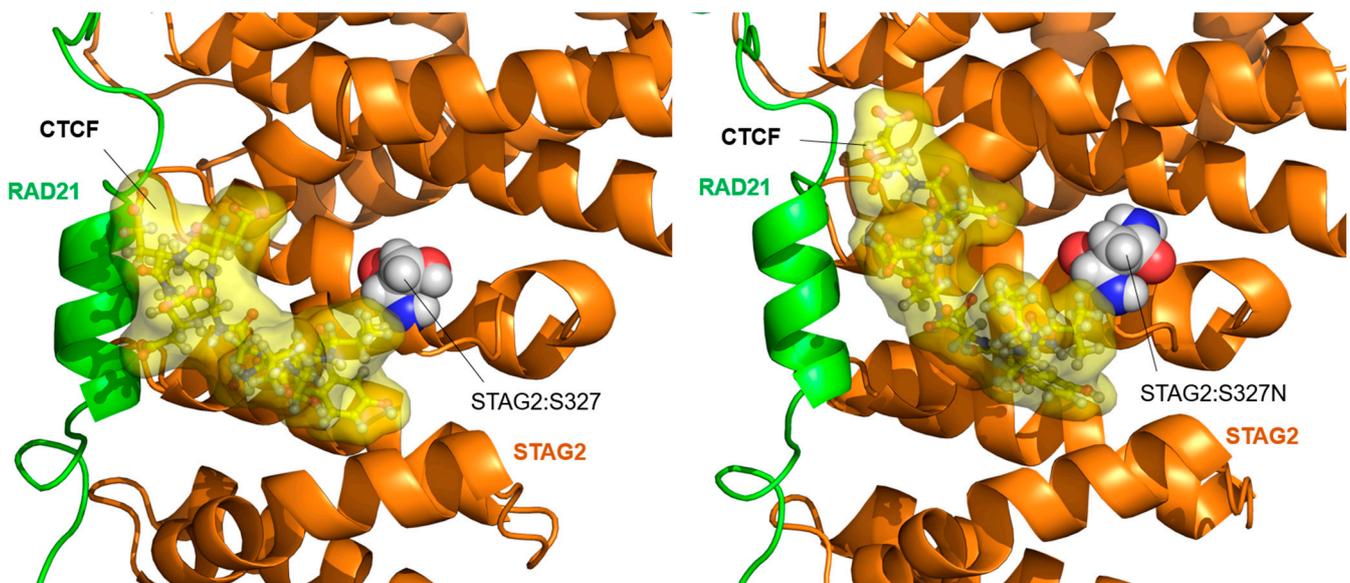


Figure 9. STAG2/CTCF interaction: STAG2:Ser327Asn variant. Model of the interaction between STAG2-RAD21 and a peptide located in the N-terminal domain of CTCF, after 200 ns of MD simulation. (Left) The position of the wild-type amino acid STAG2:Ser327 (S327) in an alpha-helix in contact with the position of the CTCF peptide is indicated. (Right) Model of the interaction in the case of the STAG2:Ser327Asn (S327N) variant. A slight shift in the position of the CTCF peptide can be observed, separating it from the contact with RAD21.

3. Discussion

The computational simulation by the MD of the motions associated with macromolecular structures is a tool that allows us to rationalize, at the atomic level, the effects that certain variants in the sequence of a protein can have on its function and on its relationship with other molecules. For example, in a previous work of our group focusing on Cornelia de Lange syndrome, the computational modeling of ATP hydrolysis in the head domain of the SMC1A and SMC3 proteins was able to illustrate the changes associated with a pathogenic variant in SMC3 that functionally alters this mechanism [16].

To our knowledge, no previous work has been published that has studied the effect of the different STAG2 variants or the proteins that interact with it at the biochemical or biophysical level. For this reason, we believe that the use of molecular dynamics simulation techniques is one of the few ways available to approach an atomic or molecular rationalization of the phenotypic effect of each of the variants.

STAG2 is an essential protein for the structure and function of the cohesin complex, although its specific role beyond its ability to bind DNA is unknown. In the present work, we have used information from pathogenic variants of STAG2 and its interacting proteins, all in evolutionary conserved positions, to advance our understanding of the function of STAG2 and the causes of disease, using computational modeling and MD simulation techniques. The variants studied were divided into four groups according to the closest neighboring molecule in the structure: STAG2/NIPBL interaction, STAG2/DNA interaction, STAG2/RAD21 interaction, and STAG2/CTCF interaction.

Of the four variants close to the STAG2/NIPBL interaction, three of them have shown similar behavior, strongly modifying the contact structure between the two proteins. Some works have shown how the relationship of NIPBL with STAG2 is different from the relationship with STAG1 [5,27], which opens the field to study in the future, by computational modeling, the interaction of STAG1 with NIPBL, and its differences with the interaction observed with STAG2 in the current work.

In the case of the variants affecting the STAG2/DNA interaction, all of them showed a similar behavior, destabilizing the coupling between the two molecules, with large

variations in the distance between the DNA and the nearby domains. Considering the major role of STAG2 in DNA binding, these observations open the way to a more detailed study of the interaction, using advanced modeling and dynamic simulation techniques. In our group, we are currently working in this direction, which will hopefully contribute to a better understanding of the role of STAG2 in the cohesin complex machinery.

Among the variants related to the STAG2/RAD21 interaction, two of them, one in RAD21 (RAD21:Ser345Pro) and another in STAG2 (STAG2:Met378Val), have shown very similar behavior, partially disrupting the interaction zone between the two proteins. The effect of RAD21:Ser345Pro described in this work is very similar to that previously described by our group, although this work did center around a simulation performed in the absence of DNA [24]. The consistency between the two observations in two different simulation conditions boosts the confidence in these results. Further studies are needed to fully understand the role that the interaction between STAG2 and RAD21 plays, not only for the co-localization of both, but also as part of their structural mechanism of action.

In the area of the STAG2/CTCF interaction, the changes observed in the simulation should be taken with caution. The small size of the CTCF peptide bound to STAG2 makes it very sensitive to small perturbations of the MD environment. It will be necessary to generate structures that include a larger region of CTCF, including its C-terminal domain, in order to advance the modeling of the interaction between the two proteins.

One of the points to be considered when studying the effect of variants in the proteins of the cohesin complex is the effect of dominance. Most of the variants found in patients are either heterozygous or, for the genes located on the X chromosome (SMC1A and HDAC8), correspond to amino acid changes with limited effect on function [8,11,18–26]. A reasonable hypothesis would be that, in the case of drastic mutations substantially affecting protein folding and, thus, profoundly impairing the cohesin complex function, the effect on cell division and gene regulation in embryonic stages would be so severe that it would not allow progression. In the case of the variants analyzed in this work, all but four positions (STAG2:Tyr159, STAG2:Phe999, STAG2:Met378, and NIPBL:Cys1311) correspond to surface residues exposed to the solvent and are therefore more involved in interactions with neighboring molecules than in internal interactions related to protein folding. Thus, STAG2 Arg604Gln and Asp659His variants would affect NIPBL binding; STAG2 Arg954Cys, Lys1009Asn, and Asp1014Asn variants would affect DNA interaction; STAG2 Gln728His, Arg862Gly, RAD21 Ser345Pro, and Pro376Arg variants would be involved in RAD21–STAG2 interaction; and the STAG2 Ser327Asn variant would be associated with STAG2 binding to CTCF. In addition to these interactions, their possible role in undescribed contacts with other macromolecules in the environment cannot be excluded. In the case of STAG2 variants of residues buried inside the protein (Tyr159Cys, Tyr159His, Phe999Cys, and Met378Val), the change does not seem to be related to changes in the local folding of the corresponding domains, but to local modifications transmitted to the surface amino acids close to them, which are involved in the interaction with other macromolecules (NIPBL, DNA, or RAD21). The case of the NIPBL variants Cys1311Arg and Cys1311His is somewhat mixed, as the electrostatic charge of the variants seems to push these residues from the hidden position of the original amino acid to a more surface-exposed position in contact with STAG2. In any case, as shown in Supplementary Figures S1 and S2, most of the variants have little effect on the overall structure of the protein regions. The only exceptions would be the STAG2:Asp659His, NIPBL:Cys1311Arg, NIPBL:Cys1311His, and RAD21:Ser345Pro variants (Supplementary Figure S3), the presence of which induces significant changes in the regions in which they are located. Nevertheless, even in these cases, the effect of these mutations is local and does not seem to profoundly disrupt the folding of the protein.

4. Materials and Methods

The structure of STAG2 and the models of its interaction with RAD21 and DNA were obtained by homology modeling, using the PDB entry 4PJW [17] as a template, which

describes the interaction of STAG2 with RAD21. The DNA position was modeled using the 3D structure of the SCC3–SCC1 complex bound to DNA, solved by X-ray crystallography (PDB entry 6H8Q [14]), as a reference. The DNA model was constructed using x3DNA [28]. Gaps were modeled by combining predictions from Phyre 2 [29] and SwissModel [30]. The NIPBL interaction model was obtained by homology modeling from the cryo-EM structure of the human SMC1A-SMC3-RAD21-STAG1-NIPBL-DNA complex (PDB id: 6WG3 [13]). The interaction of STAG2 with CTCF was modeled from PDB ID: 6QNX [12], which contains the structure of a segment of the N-terminal domain of CTCF bound to STAG2.

The 3D coordinates of STAG2 and the models of its interaction with RAD21, DNA, and CCTF were obtained from experimental structures: 4PJW [17] (crystallized structure of human STAG2 and human RAD21); 6WG3 [13] (cryo-EM structure of human NIPBL, RAD21, SMC1A, SMC3, STAG1, and DNA); and 6QNX [12] (crystallized structure of human STAG2, RAD21, and a segment of CCTF). The relative positions of NIPBL and STAG2 were determined by directly superimposing the STAG2 structure on the position of STAG1 in the 6WG3 structure [13]. The initial position of the DNA molecule in contact with STAG2 and RAD21 was determined using, as a guide, the 6H8Q structure [14], which contains the structural homologues of STAG2 and RAD21 in yeast (Scc3 and Scc1, respectively), co-crystallized with a DNA segment.

The wild-type models were parameterized with the LEaP module of Amber Tools [31], using the Amber ff14SB force field [32] for the protein moieties and the Amber OL15 force field [33] for the DNA chain. The structures were solvated with the TIP3P solvent model [34], adding Na⁺ or Cl⁻ ions to neutralize the charge on the system.

Molecular dynamics simulations were conducted using the Amber18 software package (University of California, San Francisco, CA, USA), employing the Particle Mesh Ewald (PME) method for non-bonded interactions with a cutoff distance of 8 Å. The temperature was regulated through the Langevin thermostat, maintaining a fixed temperature of 297 K with a collision frequency of 1 ps. Hydrogen bond constraints were implemented using the SHAKE algorithm, enabling a simulation timestep of 2 fs. To uphold NPT conditions (constant number of particles, pressure, and temperature), the pressure coupling was managed by a Monte Carlo Barostat set to 297 K and 1 bar.

The initial model structures underwent 10,000 cycles of energy minimization, followed by a 1 ns restrained equilibration phase, smoothly raising the temperature to 297 K, after which restraints were gradually removed over 10 ns. Subsequently, each system was subjected to a 200 ns long free MD production phase.

Simulations of the different missense variants were performed by substituting the corresponding wild-type amino acid for the variant, following the same modeling and MD simulation steps previously described. The positions of amino acids and DNA strands throughout the MD trajectories were analyzed using cpptraj [35] and VMD [36]. Figures were generated with Pymol (Schrödinger, New York, NY, USA).

Multiple sequence alignments of human and mouse STAG1/2/3 proteins were generated using MUSCLE [37] and visualized and colored according to their evolutionary conservation using belvu [38].

Supplementary Materials: The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25021280/s1>.

Author Contributions: Conceptualization, methodology, and writing: D.R.-P., P.G.-P. and Í.M.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financed by the following grants: Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (MCIU-AEI, RTI2018-094434-B-I00 and PID2021-126625OB-I00/10.13039/501100011033/FEDER, EU), and Instituto de Salud Carlos III (DTS20-00024).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data analyzed during this study are included in this published article. Additional supporting data are available from the corresponding authors upon reasonable request.

Acknowledgments: The computational support of the “Centro de Computación Científica CCC-UAM” is gratefully recognized.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Peters, J.-M.; Tedeschi, A.; Schmitz, J. The Cohesin Complex and Its Roles in Chromosome Biology. *Genes. Dev.* **2008**, *22*, 3089–3114. [[CrossRef](#)] [[PubMed](#)]
- Gligoris, T.; Löwe, J. Structural Insights into Ring Formation of Cohesin and Related Smc Complexes. *Trends Cell Biol.* **2016**, *26*, 680–693. [[CrossRef](#)] [[PubMed](#)]
- Wutz, G.; Várnai, C.; Nagasaka, K.; Cisneros, D.A.; Stocsits, R.R.; Tang, W.; Schoenfelder, S.; Jessberger, G.; Muhar, M.; Hossain, M.J.; et al. Topologically Associating Domains and Chromatin Loops Depend on Cohesin and Are Regulated by CTCF, WAPL, and PDS5 Proteins. *EMBO J.* **2017**, *36*, 3573–3599. [[CrossRef](#)] [[PubMed](#)]
- Skibbens, R.V. Condensins and Cohesins—One of These Things Is Not like the Other! *J. Cell Sci.* **2019**, *132*, jcs220491. [[CrossRef](#)] [[PubMed](#)]
- Alonso-Gil, D.; Losada, A. NIPBL and Cohesin: New Take on a Classic Tale. *Trends Cell Biol.* **2023**, *33*, 860–871. [[CrossRef](#)]
- Horsfield, J.A. Full Circle: A Brief History of Cohesin and the Regulation of Gene Expression. *FEBS J.* **2023**, *290*, 1670–1687. [[CrossRef](#)] [[PubMed](#)]
- Antony, J.; Chin, C.V.; Horsfield, J.A. Cohesin Mutations in Cancer: Emerging Therapeutic Targets. *Int. J. Mol. Sci.* **2021**, *22*, 6788. [[CrossRef](#)]
- Mullegama, S.V.; Klein, S.D.; Mulatinho, M.V.; Senaratne, T.N.; Singh, K.; UCLA Clinical Genomics Center; Nguyen, D.C.; Gallant, N.M.; Strom, S.P.; Ghahremani, S.; et al. De Novo Loss-of-Function Variants in STAG2 Are Associated with Developmental Delay, Microcephaly, and Congenital Anomalies. *Am. J. Med. Genet. A* **2017**, *173*, 1319–1327. [[CrossRef](#)]
- Mullegama, S.V.; Klein, S.D.; Signer, R.H.; UCLA Clinical Genomics Center; Vilain, E.; Martinez-Agosto, J.A. Mutations in STAG2 Cause an X-Linked Cohesinopathy Associated with Undergrowth, Developmental Delay, and Dysmorphia: Expanding the Phenotype in Males. *Mol. Genet. Genom. Med.* **2019**, *7*, e00501. [[CrossRef](#)]
- Kruszka, P.; Berger, S.I.; Casa, V.; Dekker, M.R.; Gaesser, J.; Weiss, K.; Martinez, A.F.; Murdock, D.R.; Louie, R.J.; Prijoles, E.J.; et al. Cohesin Complex-Associated Holoprosencephaly. *Brain* **2019**, *142*, 2631–2643. [[CrossRef](#)]
- Kaur, M.; Blair, J.; Devkota, B.; Fortunato, S.; Clark, D.; Lawrence, A.; Kim, J.; Do, W.; Semeo, B.; Katz, O.; et al. Genomic Analyses in Cornelia de Lange Syndrome and Related Diagnoses: Novel Candidate Genes, Genotype-Phenotype Correlations and Common Mechanisms. *Am. J. Med. Genet. A* **2023**, *191*, 2113–2131. [[CrossRef](#)]
- Li, Y.; Haarhuis, J.H.I.; Sedeño Cacciatore, Á.; Oldenkamp, R.; van Ruiten, M.S.; Willems, L.; Teunissen, H.; Muir, K.W.; de Wit, E.; Rowland, B.D.; et al. The Structural Basis for Cohesin-CTCF-Anchored Loops. *Nature* **2020**, *578*, 472–476. [[CrossRef](#)] [[PubMed](#)]
- Shi, Z.; Gao, H.; Bai, X.-C.; Yu, H. Cryo-EM Structure of the Human Cohesin-NIPBL-DNA Complex. *Science* **2020**, *368*, 1454–1459. [[CrossRef](#)] [[PubMed](#)]
- Li, Y.; Muir, K.W.; Bowler, M.W.; Metz, J.; Haering, C.H.; Panne, D. Structural Basis for Scc3-Dependent Cohesin Recruitment to Chromatin. *Elife* **2018**, *7*, e38356. [[CrossRef](#)]
- Sun, Y.; Xu, X.; Zhao, W.; Zhang, Y.; Chen, K.; Li, Y.; Wang, X.; Zhang, M.; Xue, B.; Yu, W.; et al. RAD21 Is the Core Subunit of the Cohesin Complex Involved in Directing Genome Organization. *Genome Biol.* **2023**, *24*, 155. [[CrossRef](#)]
- Marcos-Alcalde, Í.; Mendieta-Moreno, J.I.; Puisac, B.; Gil-Rodríguez, M.C.; Hernández-Marcos, M.; Soler-Polo, D.; Ramos, F.J.; Ortega, J.; Pié, J.; Mendieta, J.; et al. Two-Step ATP-Driven Opening of Cohesin Head. *Sci. Rep.* **2017**, *7*, 3266. [[CrossRef](#)] [[PubMed](#)]
- Hara, K.; Zheng, G.; Qu, Q.; Liu, H.; Ouyang, Z.; Chen, Z.; Tomchick, D.R.; Yu, H. Structure of Cohesin Subcomplex Pinpoints Direct Shugoshin-Wapl Antagonism in Centromeric Cohesion. *Nat. Struct. Mol. Biol.* **2014**, *21*, 864–870. [[CrossRef](#)]
- Yuan, B.; Neira, J.; Pehlivan, D.; Santiago-Sim, T.; Song, X.; Rosenfeld, J.; Posey, J.E.; Patel, V.; Jin, W.; Adam, M.P.; et al. Clinical Exome Sequencing Reveals Locus Heterogeneity and Phenotypic Variability of Cohesinopathies. *Genet. Med.* **2019**, *21*, 663–675. [[CrossRef](#)]
- Tonkin, E.T.; Wang, T.-J.; Lisgo, S.; Bamshad, M.J.; Strachan, T. NIPBL, Encoding a Homolog of Fungal Scc2-Type Sister Chromatid Cohesion Proteins and Fly Nipped-B, Is Mutated in Cornelia de Lange Syndrome. *Nat. Genet.* **2004**, *36*, 636–641. [[CrossRef](#)]
- Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; et al. ClinVar: Improving Access to Variant Interpretations and Supporting Evidence. *Nucleic Acids Res.* **2018**, *46*, D1062–D1067. [[CrossRef](#)]
- Freyberger, F.; Kokotović, T.; Krnjak, G.; Frković, S.H.; Nagy, V. Expanding the Known Phenotype of Mullegama-Klein-Martinez Syndrome in Male Patients. *Hum. Genome Var.* **2021**, *8*, 37. [[CrossRef](#)] [[PubMed](#)]
- Brunet, T.; Jech, R.; Brugger, M.; Kovacs, R.; Alhaddad, B.; Leszinski, G.; Riedhammer, K.M.; Westphal, D.S.; Mahle, I.; Mayerhanser, K.; et al. De Novo Variants in Neurodevelopmental Disorders—Experiences from a Tertiary Care Center. *Clin. Genet.* **2021**, *100*, 14–28. [[CrossRef](#)] [[PubMed](#)]

23. Kosmicki, J.A.; Samocha, K.E.; Howrigan, D.P.; Sanders, S.J.; Slowikowski, K.; Lek, M.; Karczewski, K.J.; Cutler, D.J.; Devlin, B.; Roeder, K.; et al. Refining the Role of de Novo Protein-Truncating Variants in Neurodevelopmental Disorders by Using Population Reference Samples. *Nat. Genet.* **2017**, *49*, 504–510. [[CrossRef](#)] [[PubMed](#)]
24. Krab, L.C.; Marcos-Alcalde, I.; Assaf, M.; Balasubramanian, M.; Andersen, J.B.; Bisgaard, A.-M.; Fitzpatrick, D.R.; Gudmundsson, S.; Huisman, S.A.; Kalayci, T.; et al. Delineation of Phenotypes and Genotypes Related to Cohesin Structural Protein RAD21. *Hum. Genet.* **2020**, *139*, 575–592. [[CrossRef](#)] [[PubMed](#)]
25. Deardorff, M.A.; Wilde, J.J.; Albrecht, M.; Dickinson, E.; Tennstedt, S.; Braunholz, D.; Mönnich, M.; Yan, Y.; Xu, W.; Gil-Rodríguez, M.C.; et al. RAD21 Mutations Cause a Human Cohesinopathy. *Am. J. Hum. Genet.* **2012**, *90*, 1014–1027. [[CrossRef](#)]
26. Soardi, F.C.; Machado-Silva, A.; Linhares, N.D.; Zheng, G.; Qu, Q.; Pena, H.B.; Martins, T.M.M.; Vieira, H.G.S.; Pereira, N.B.; Melo-Minardi, R.C.; et al. Familial STAG2 Germline Mutation Defines a New Human Cohesinopathy. *NPJ Genom. Med.* **2017**, *2*, 7. [[CrossRef](#)]
27. Alonso-Gil, D.; Cuadrado, A.; Giménez-Llorente, D.; Rodríguez-Corsino, M.; Losada, A. Different NIPBL Requirements of Cohesin-STAG1 and Cohesin-STAG2. *Nat. Commun.* **2023**, *14*, 1326. [[CrossRef](#)]
28. Colasanti, A.V.; Lu, X.-J.; Olson, W.K. Analyzing and Building Nucleic Acid Structures with 3DNA. *J. Vis. Exp.* **2013**, *74*, e4401. [[CrossRef](#)]
29. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)]
30. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [[CrossRef](#)]
31. Case, D.A.; Aktulga, H.M.; Belfon, K.; Cerutti, D.S.; Cisneros, G.A.; Cruzeiro, V.W.D.; Forouzes, N.; Giese, T.J.; Götz, A.W.; Gohlke, H.; et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63*, 6183–6191. [[CrossRef](#)] [[PubMed](#)]
32. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)] [[PubMed](#)]
33. Love, O.; Galindo-Murillo, R.; Zgarbová, M.; Šponer, J.; Jurečka, P.; Cheatham, T.E. Assessing the Current State of Amber Force Field Modifications for DNA—2023 Edition. *J. Chem. Theory Comput.* **2023**, *19*, 4299–4307. [[CrossRef](#)] [[PubMed](#)]
34. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935. [[CrossRef](#)]
35. Roe, D.R.; Cheatham, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095. [[CrossRef](#)]
36. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
37. Edgar, R.C. MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity. *BMC Bioinform.* **2004**, *5*, 113. [[CrossRef](#)]
38. Sonnhammer, E.L.L.; Hollich, V. Scoredist: A Simple and Robust Protein Sequence Distance Estimator. *BMC Bioinform.* **2005**, *6*, 108. [[CrossRef](#)]

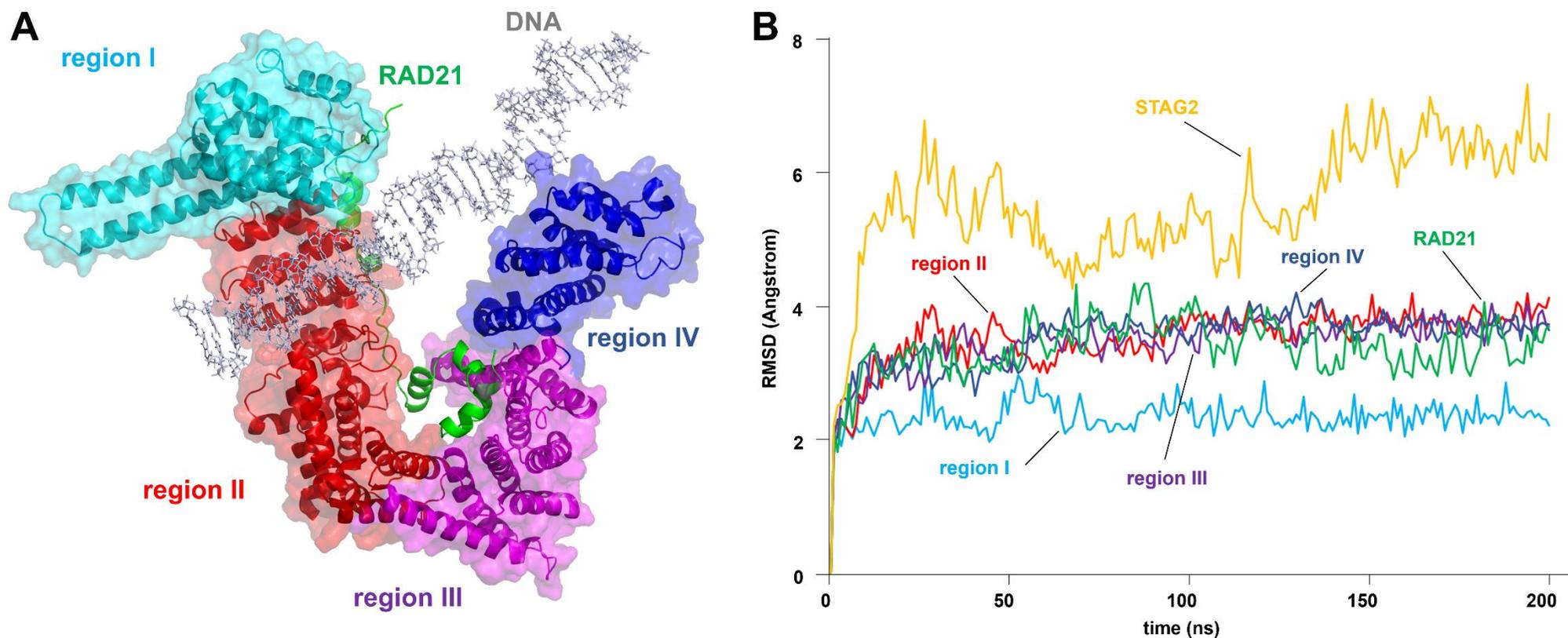
Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Supplementary Material to:

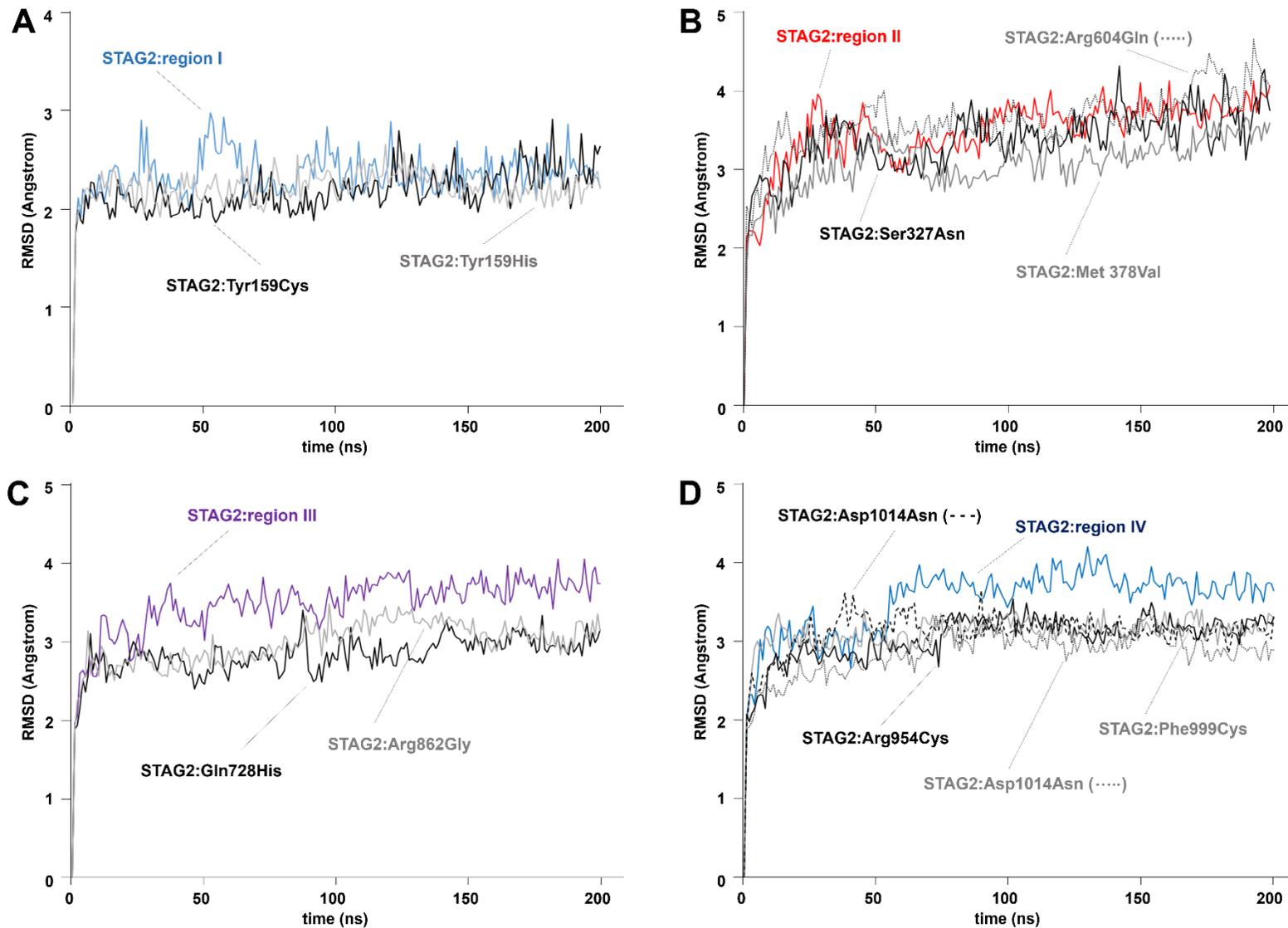
STAG2: Computational analysis of missense variants involved in disease.

Authors: David Ros-Pardo, Paulino Gómez-Puertas & Íñigo Marcos-Alcalde

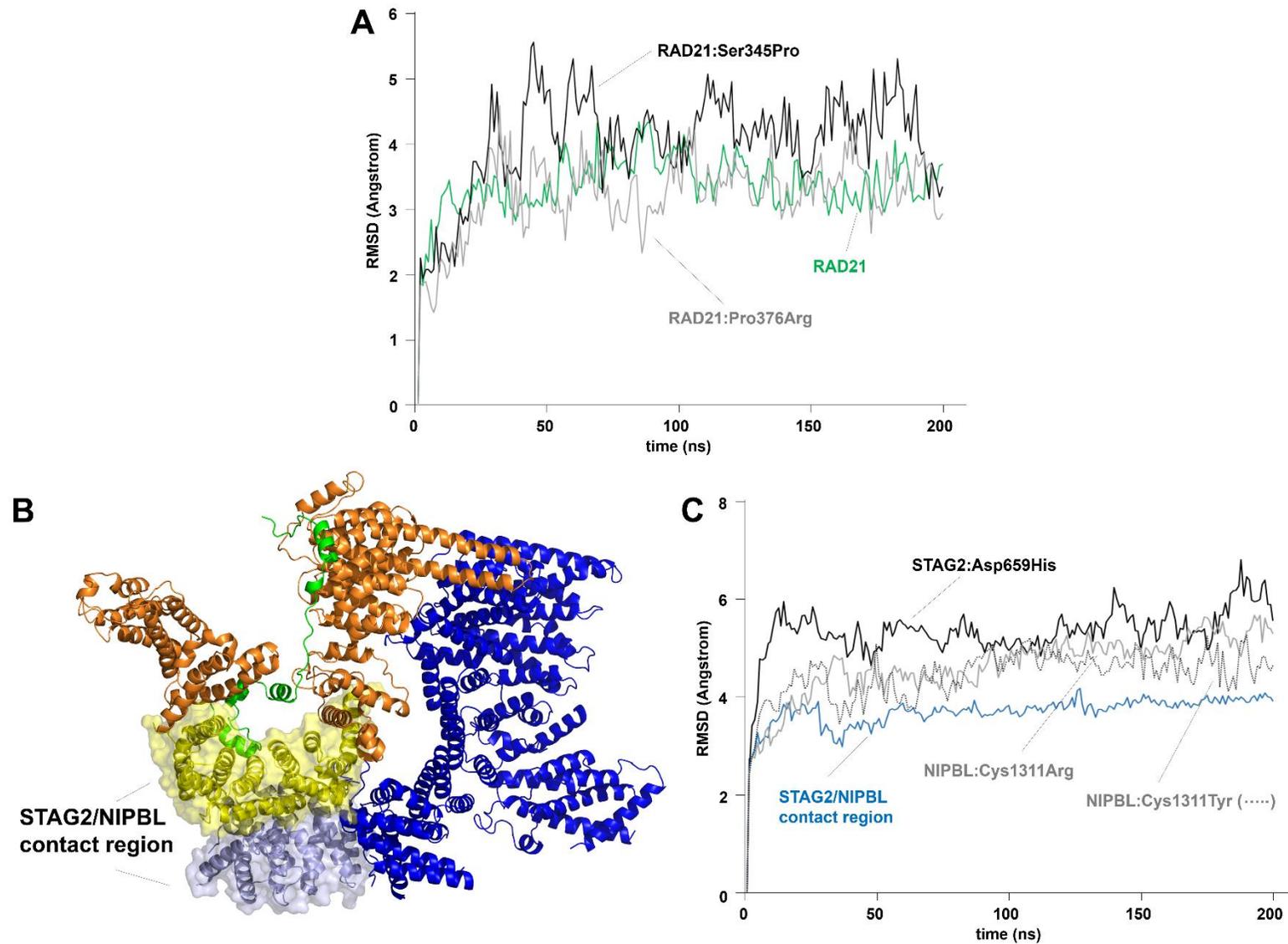
- Supplementary Figure S1
- Supplementary Figure S2
- Supplementary Figure S3



Supplementary Figure S1. Root Mean Square Deviation (RMSD) analysis of the structure of STAG2 and RAD21 during molecular dynamics simulations. A. Region partitioning of the STAG2 structure. During the molecular dynamics trajectory, due to the internal flexibility of its V-shaped structure, the entire STAG2 protein undergoes a series of oscillations that produce very high values in the RMSD measurement, masking any other variation of lesser amplitude. For this reason, the STAG2 protein has been subdivided into four regions the structure of which remains internally stable: region I (residues 84 to 302; cyan), region II (residues 303-647; red), region III (residues 648-880; purple) and region IV (residues 881-1048; blue). The segment corresponding to RAD21 is treated as an additional region (green color). B. Measured RMSD values (in Angstrom) during the 200 ns molecular dynamics simulation of the entire structure of STAG2 protein (orange line, showing large oscillations), the RAD21 segment (green line), and each of the four regions of STAG2 (I, II, III, and IV, cyan, red, purple, and blue lines, respectively).



Supplementary Figure S2. Comparison of RMSD values measured over 200 ns of the molecular dynamics trajectory of each of the four regions of STAG2 in the wild-type protein (line colors as in Supplementary Fig. S1) and in the variants located in each of the regions (in different shades of gray and trace types). A. RMSD values of region I of STAG2 wild type and the STAG2:Tyr159Cys and STAG2:Tyr159His variants. B. RMSD values of region II of STAG2 wild-type and the STAG2:Ser327Asn, STAG2:Met378Val and STAG2:Arg604Gln variants. C. RMSD values of region III of STAG2 wild-type and the STAG2:Gln728His and STAG2:Arg862Gly variants. D. RMSD values of region IV of STAG2 wild-type and the STAG2:Arg954Cys, STAG2:Phe999Cys, STAG2:Lys1009Asn and STAG2:Asp1014Asn variants. None of the simulated variants show RMSD values that are significantly higher than those for the wild-type regions, indicating that the amino acid changes analyzed do not generate large variations in the internal structure of the protein.



Supplementary Figure S3. A. Comparison of RMSD values measured over 200 ns of the molecular dynamics trajectory of RAD21 (green) and the RAD21:Ser345Pro and RAD21:Pro376Arg variants. The trajectory of the RAD21:Ser345Pro variant shows high RMSD values, indicating a significant change in the structure of the protein segment. B. Model of the interaction between STAG2 (orange) and NIPBL (blue). A region composed of the two contact structures of both proteins (in yellow and cyan) is highlighted. In this region, where the amino acids Asp659 of STAG2 and Cys1311 of NIPBL are located, the RMSD values along the molecular dynamics trajectory have been measured. C. Comparison of RMSD values measured over 200 ns of the molecular dynamics trajectory of the STAG/NIPBL contact region wild-type (blue) and the STAG2:Asp659His, NIPBL:Cys1311Arg and NIPBL:Cys1311His variants. The three variants lead to remarkable structural changes (especially the STAG2:Asp659His variant), resulting in high RMSD values along the trajectory.