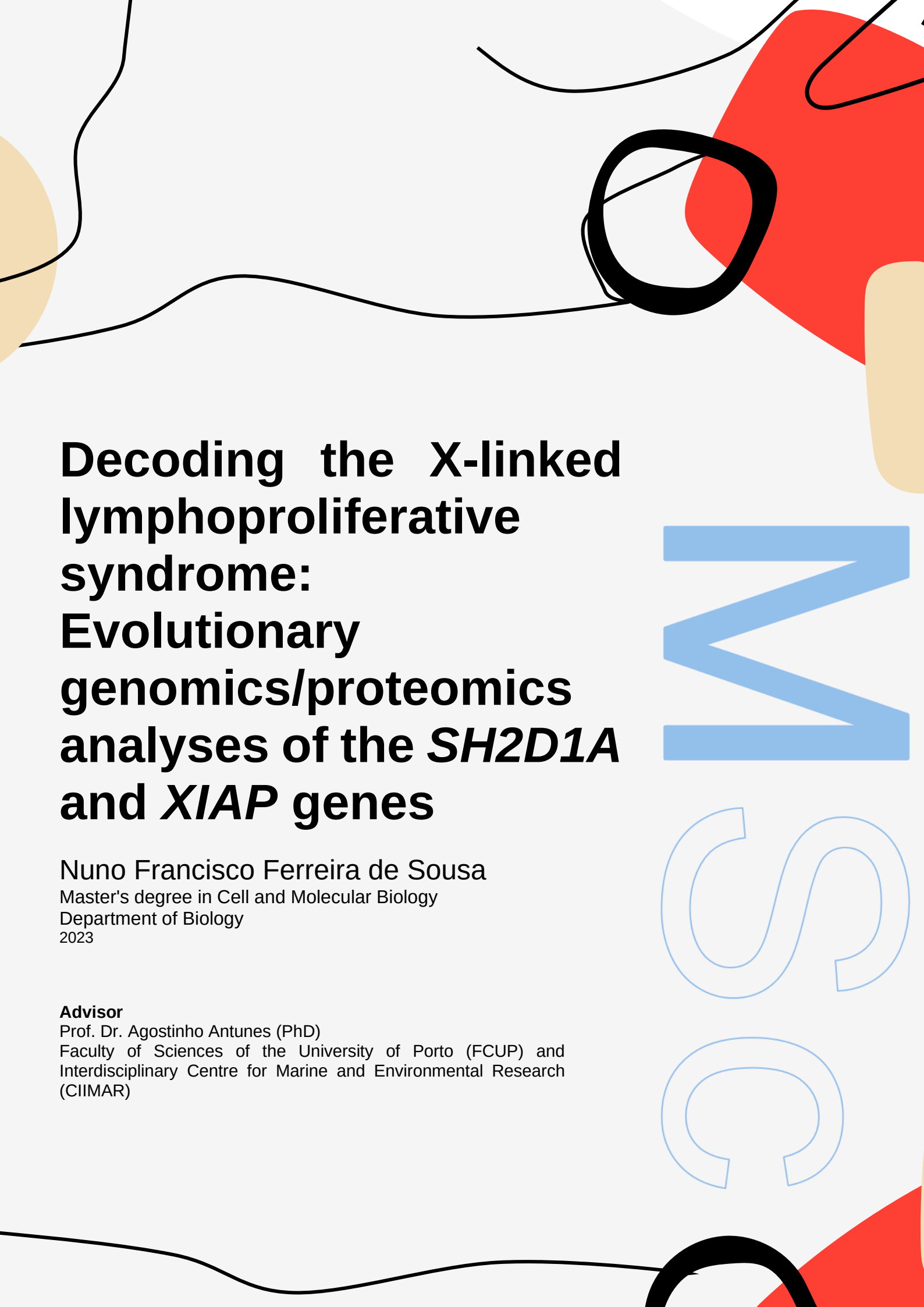




Decoding the X-linked lymphoproliferative syndrome: Evolutionary genomics/proteomics analyses of the *SH2D1A* and *XIAP* genes

Nuno Francisco Ferreira de Sousa

Master's degree in Cell and Molecular Biology

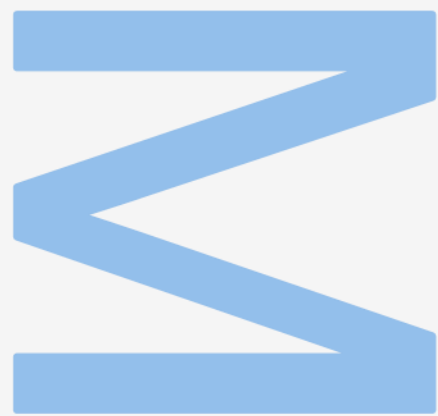
Department of Biology

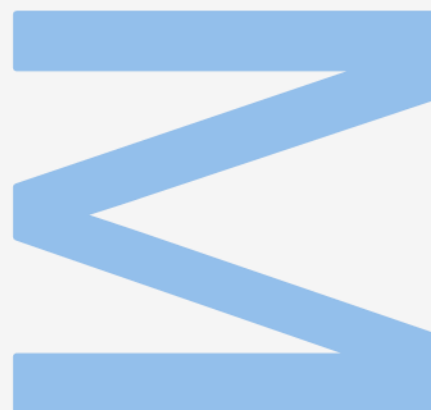
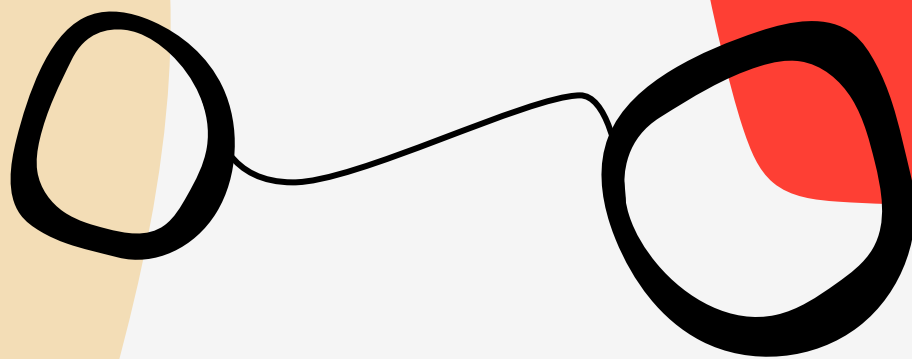
2023

Advisor

Prof. Dr. Agostinho Antunes (PhD)

Faculty of Sciences of the University of Porto (FCUP) and
Interdisciplinary Centre for Marine and Environmental Research
(CIIMAR)





“I think...”

– **Charles Darwin**

Acknowledgements

Gostava de começar por agradecer a Faculdade de Ciências da Universidade do Porto (FCUP) e ao Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), na pessoa do meu orientador, professor doutor Agostinho Antunes, pela disponibilidade em aceitar-me para desenvolver o meu projeto e trabalho de tese de mestrado no seu grupo de investigação. Quero agradecer especialmente a minha supervisora Luana Ramos, pela disponibilidade e acessibilidade pela qual teve comigo e com o meu trabalho. Sem ti, o meu trabalho não existia! Não existem obrigados nem abraços suficientes que consigam expressar o sentimento de gratidão, apreço e amizade que sinto para contigo. Desejo que tenhas uma vida repleta de alegrias e sucesso!

Um grande e muito especial agradecimento a minha família, por me ter ajudado e apoiado em tudo o que podia nestes últimos 5 anos para conseguir concretizar o meu sonho, de me tornar alguém na área das ciências da vida e da saúde.

Um obrigado gigantesco ao meu grupo de amigos Bi(o)Godes & BioPorcas pelos bons momentos, amor e apoio que me deram ao longo destes 5 anos de faculdade. Desejo profundamente que nossa amizade e animação nunca acabem! Obrigado por terem estado lá sempre a apoiar, principalmente nos piores momentos da minha vida. Apesar de gostarmos de estar juntos para a diversão, nunca “abandonamos o barco quando havia tempestade”, ajudando-nos sempre uns aos outros nos tempos de maiores necessidades e fraqueza. Agradeço também de igual forma a minha afilhada Nádia, que juntamente com o Quim, o Joca e o Silva acompanharam, literalmente, o desenvolvimento do meu trabalho de tese durante o ano inteiro, seguindo-me pelas intermináveis horas do dia e da noite, ouvindo muitas vezes as minhas lamúrias e preocupações do mestrado e da vida. Irei estar ao vosso lado sempre, tal como tiveram ao meu, quando for também a vossa vez de escrever a tese! Agradeço também as minhas afilhadas, em especial o meu afilhado Gustavo, por ter estado sempre ao meu lado, apesar de todas as circunstâncias. Espero que tenhas orgulho de mim sempre, mesmo quando eu falho, tal como eu tenho por ti! Agradecimentos igualmente aos meus amigos do secundário, por nunca termos esquecido uns dos outros e continuarmos sempre juntos. Estar com vocês é sentir-me em casa! Quero também agradecer a ti André, a minha pessoa favorita, por me apoiares todos os dias e incentivares a nunca desistir dos meus sonhos. Onde muitos outros viraram as costas, apesar de tudo o que tinha sido construído, tu continuaste lá. Obrigado por nunca teres desistido mim!

“O verdadeiro amor nasce em tempos difíceis” - **John Green**

Resumo

As doenças associadas ao cromossoma X causadas por mutações genéticas que afetam as funções das proteínas, em particular nos genes relacionados com o sistema imunitário, têm efeitos consideráveis em diversos sistemas de órgãos. A síndrome linfoproliferativa ligada ao cromossoma X (XLP), uma doença imunitária que afeta predominantemente jovens do sexo masculino, é causada por mutações nos genes *SH2D1A* e *XIAP*, localizados no cromossoma X humano. O primeiro capítulo desta tese inclui uma revisão detalhada centrada na síndrome XLP, detalhando os seus aspetos clínicos, fisiopatológicos e moleculares e destacando futuras direções de investigação.

Os avanços na sequenciação genómica revolucionaram o estudo da evolução dos genes, tornando a genómica comparativa crucial para identificar forças seletivas e desvendar interações biológicas. Por conseguinte, o segundo capítulo desta tese centra-se na análise da evolução molecular dos genes *SH2D1A*, *SH2D1B* e *XIAP* em várias espécies de primatas e de outros mamíferos, através de reconstruções filogenéticas de “Maximum Likelihood”, análises de seleção nos softwares CODEML e Datamonkey, e análises sinténicas. Os resultados indicam a presença de seleção purificadora robusta, sugerindo uma forte conservação dos três genes, particularmente do *SH2D1A* e do *XIAP*. Esta conservação realça a função imunitária crucial mantida ao longo da evolução dos mamíferos e a conservação das configurações estruturais das proteínas, sugerindo que as variações deletérias observadas na XLP estão sujeitas a seleção negativa, contribuindo para a sua eliminação na população.

No terceiro capítulo, analisamos o impacto de mutações na estrutura da proteína *SH2D1A*. Utilizando ferramentas de previsão *in silico*, identificámos 23 polimorfismos de nucleótido único (SNP) altamente deletérios e instáveis. Estas mutações afetam predominantemente aminoácidos conservados, induzindo alterações que afetam a estabilidade da proteína. As simulações de dinâmica molecular revelam vários efeitos sobre a estabilidade, flexibilidade e estrutura das proteínas, destacando a complexidade das relações entre os diferentes parâmetros obtidos a partir da simulação. Embora a validação experimental seja crucial, estes resultados abrem portas a potenciais vias de intervenção da medicina personalizada com base em perfis genéticos.

Palavras-chave: Síndrome linfoproliferativa ligada ao cromossoma X, *SH2D1A*, *SH2D1B*, *XIAP*, Análises de seleção, Simulações de Dinâmica Molecular, Mutações deletérias

Abstract

X-linked diseases caused by gene mutations that affect the function of proteins, particularly in immune-related genes, have far-reaching effects on organ systems. The X-linked lymphoproliferative syndrome (XLP), a X-linked immune disorder which predominantly affects young males, is caused by mutations in the *SH2D1A* and *XIAP* genes. The first chapter of this thesis includes a comprehensive review focused on the XLP syndrome, examining its clinical, pathophysiological and molecular aspects, and highlighting future research directions.

Recent advances in sequencing techniques have revolutionised our ability to obtain comprehensive whole-genome data, enabling a deeper understanding of gene evolution. Thus, comparative genomics is an important tool for identifying selective forces and unravelling biological interactions that influence gene evolution. The second chapter of this study therefore focuses on the molecular evolutionary analysis of the *SH2D1A*, *SH2D1B* and *XIAP* genes. We analyzed coding sequences from various primates and other mammalian species, conducting Maximum Likelihood phylogenetic reconstructions using IQ-TREE, selection analyses with CODEML and Datamonkey, and syntenic analyses. Our results indicate robust purifying selection, suggesting strong conservation of all three genes, particularly *SH2D1A* and *XIAP*. This conservation highlights the crucial immune function maintained throughout mammalian evolution and the conservation of protein structural configurations, suggesting that deleterious variations observed in XLP are subject to negative selection, contributing to their elimination in the population.

In the third chapter, we analyse the impact of certain mutations on the protein structure of SH2D1A. Using *in silico* prediction tools, we identified 23 highly deleterious and unstable single nucleotide polymorphisms (SNPs). These mutations predominantly affect conserved amino acids, inducing changes in physicochemical properties that affect protein stability. Molecular dynamics simulations (MDS) reveal various effects on the stability, flexibility and structure of the protein, highlighting the complexity of the relationships between the different parameters obtained from the simulation. Although experimental validation is crucial, these findings open doors to potential avenues of personalized medicine intervention based on genetic profiles.

Keywords: X-linked lymphoproliferative syndrome, *SH2D1A*, *SH2D1B*, *XIAP*, Selection analyses, Deleterious mutations

Contents

List of Charts	XI
List of Figures	XVII
List of Tables	XIX
List of Abbreviations	XXIII

Chapter I: General Introduction – Exploring X-linked immune disorders: A clinical and molecular overview of X-linked lymphoproliferative syndrome (XLP)	1
--	---

Subchapter 1.1.: Abstract	3
--	---

Subchapter 1.2.: Introduction	4
--	---

The X chromosome and the inheritance modes.....	4
---	---

The X-linked immune disorders	5
-------------------------------------	---

Subchapter 1.3.: The clinical landscape of X-linked lymphoproliferative syndrome	7
---	---

What is the X-linked lymphoproliferative syndrome?.....	7
---	---

Characterization of clinical manifestations of XLP	8
--	---

Therapies and Treatments: What exists and future prospects	10
--	----

Subchapter 1.4.: The pathophysiological landscape of X-linked lymphoproliferative syndrome	11
---	----

The Epstein-Barr virus	11
------------------------------	----

EBV infection and immune response	13
---	----

T and NK cells: The crucial role in the immune response	16
---	----

Subchapter 1.5.: The molecular landscape of X-linked lymphoproliferative syndrome	17
--	----

<i>SH2D1A</i> and <i>XIAP</i>	17
<i>SH2D1A</i> signalling pathways.....	19
<i>XIAP</i> signalling pathways	22
Subchapter 1.6.: Concluding remarks and future perspectives	23
References	24
Chapter II: Evolutionary genomic analyses of <i>SH2D1A</i> , <i>SH2D1B</i> and <i>XIAP</i> genes in primates: Implications on immune functionality	45
Subchapter 2.1.: Abstract	47
Subchapter 2.2.: Introduction	48
Subchapter 2.3.: Material and Methods	50
Genomic data compilation	50
Dataset preparation and multiple sequence alignment.....	50
Recombination and saturation analyses	51
Phylogenetic inference	51
Selection analyses.....	52
Synteny analyses	54
Subchapter 2.4.: Results.....	54
Sequence collection and MSA construction	54
Phylogenetic inference based on the mammalian <i>SH2D1A</i> , <i>SH2D1B</i> and <i>XIAP</i> genes	55
Inspecting the selective forces.....	55
Mapping the genes in syntenic blocks	58
Subchapter 2.5.: Discussion	58
Subchapter 2.6.: Conclusion and Future Perspectives.....	61
References	62

Appendix.....	74
Chapter III: <i>In silico</i> analysis of non-synonymous SNPs and their implications on X-linked lymphoproliferative syndrome type 1 (XLP1)	99
Subchapter 3.1.: Abstract	101
Subchapter 3.2.: Introduction	102
Subchapter 3.3.: Material and Methods	104
Retrieval of SNPs and SH2D1A protein sequence.....	104
Missense SNPs analysis	104
Loop modelling, validation and mutagenesis	106
Molecular dynamics simulation	107
Subchapter 3.4.: Results.....	108
Missense SNPs collection.....	108
Analyses and selection of the most deleterious SNPs	108
Molecular Dynamics of native and mutant SH2D1A protein.....	109
Subchapter 3.5.: Discussion	113
Subchapter 3.6.: Conclusion and Future Perspectives.....	118
References	119
Appendix.....	134
Chapter IV: General Discussion, Conclusion and Future Perspectives	177
Subchapter 4.1.: General Discussion.....	179
Subchapter 4.2.: General Conclusion and Future Perspectives	180
References	181

List of Charts

Chapter II: Evolutionary genomic analyses of *SH2D1A*, *SH2D1B* and *XIAP* genes in primates: Implications on immune functionality

Figure S4 - Graphs showing the PhyloP and PhastCons scores for the *SH2D1A* gene in the primate (A) and Mammalia (B) datasets **94**

Figure S5 - Graphs showing the PhyloP and PhastCons scores for the *SH2D1B* gene in the primate (A) and Mammalia (B) datasets **94**

Figure S6 - Graphs showing the PhyloP and PhastCons scores for the *XIAP* gene in the primate (A) and Mammalia (B) datasets **95**

Chapter III: *In silico* analysis of non-synonymous SNPs and their implications on X-linked lymphoproliferative syndrome type 1 (XLP1)

Figure 3 - Analysis of the evolution of backbone RMSD of all mutants and native (WT) SH2D1A protein at 25 ns **108**

Figure 4 - Analysis RMS fluctuation of carbon alpha on all mutants and native (WT) SH2D1A protein **108**

Figure 5 - Analysis of radius gyration (Rg) on the backbone of all mutants and native (WT) SH2D1A protein **109**

Figure 6 - Analysis of intramolecular hydrogen bonds on all mutants and native (WT) SH2D1A protein **110**

Figure 7 - Analysis of solvent accessible surface area (SASA) on all mutants and native (WT) SH2D1A protein **110**

Figure 8 - Graphical summary of the average frequency of protein secondary structures in the native protein and by mutant **111**

Figure S1 - Analysis of the evolution of backbone RMSD of native and mutants SH2D1A protein at 25 ns. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{x}$) Average of the RMSD values and ($\Delta\bar{x}$)

difference between the average of the RMSD values between the mutant and the wild type..... **141**

Figure S2 - Analysis of the evolution of backbone RMSD of mutants SH2D1A protein at 25 ns. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{x}$) Average of the RMSD values and ($\Delta\bar{x}$) difference between the average of the RMSD values between the mutant and the wild type **142**

Figure S3 - Analysis of the evolution of backbone RMSD of mutants SH2D1A protein at 25 ns. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{x}$) Average of the RMSD values and ($\Delta\bar{x}$) difference between the average of the RMSD values between the mutant and the wild type **143**

Figure S4 - Analysis of the evolution of backbone RMSD of mutants SH2D1A protein at 25 ns. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G. ($\bar{x}$) Average of the RMSD values and ($\Delta\bar{x}$) difference between the average of the RMSD values between the mutant and the wild type **144**

Figure S5 - Analysis RMS fluctuation of carbon alpha on native and mutants SH2D1A protein. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{x}$) Average of the RMSF values and ($\Delta\bar{x}$) difference between the average of the RMSF values between the mutant and the wild type **145**

Figure S6 - Analysis RMS fluctuation of carbon alpha on mutants SH2D1A protein. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{x}$) Average of the RMSF values and ($\Delta\bar{x}$) difference between the average of the RMSF values between the mutant and the wild type **146**

Figure S7 - Analysis RMS fluctuation of carbon alpha on mutants SH2D1A protein. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{x}$) Average of the RMSF values and ($\Delta\bar{x}$) difference between the average of the RMSF values between the mutant and the wild type **147**

Figure S8 - Analysis RMS fluctuation of carbon alpha on mutants SH2D1A protein. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G. ($\bar{x}$) Average of the RMSF values and ($\Delta\bar{x}$) difference between the average of the RMSF values between the mutant and the wild type **148**

Figure S9 - Analysis of radius gyration (Rg) on the backbone of native and mutants SH2D1A protein. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant

H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{x}$) Average of the Rg values and ($\Delta\bar{x}$) difference between the average of the Rg values between the mutant and the wild type..... **149**

Figure S10 - Analysis of radius gyration (Rg) on the backbone of mutants SH2D1A protein. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{x}$) Average of the Rg values and ($\Delta\bar{x}$) difference between the average of the Rg values between the mutant and the wild type **150**

Figure S11 - Analysis of radius gyration (Rg) on the backbone of mutants SH2D1A protein. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{x}$) Average of the Rg values and ($\Delta\bar{x}$) difference between the average of the Rg values between the mutant and the wild type **151**

Figure S12 - Analysis of radius gyration (Rg) on the backbone of mutants SH2D1A protein. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G. ($\bar{x}$) Average of the Rg values and ($\Delta\bar{x}$) difference between the average of the Rg values between the mutant and the wild type..... **152**

Figure S13 - Analysis of intramolecular hydrogen bonds on the native and mutants SH2D1A protein. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{x}$) Average of bonds number and ($\Delta\bar{x}$) difference between the average of bond numbers between the mutant and the wild type..... **153**

Figure S14 - Analysis of intramolecular hydrogen bonds on the mutants SH2D1A protein. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{x}$) Average of bonds number and ($\Delta\bar{x}$) difference between the average of bond numbers between the mutant and the wild type **154**

Figure S15 - Analysis of intramolecular hydrogen bonds on the mutants SH2D1A protein. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{x}$) Average of bonds number and ($\Delta\bar{x}$) difference between the average of bond numbers between the mutant and the wild type **155**

Figure S16 - Analysis of intramolecular hydrogen bonds on the mutants SH2D1A protein. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G. ($\bar{x}$) Average of bonds number and ($\Delta\bar{x}$) difference between the average of bond numbers between the mutant and the wild type..... **156**

Figure S17 - Analysis of solvent accessible surface area (SASA) on the native and mutants SH2D1A protein. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{x}$) Average of SASA value and ($\Delta\bar{x}$) difference between the average of SASA value between the mutant and the wild type..... **157**

Figure S18 - Analysis of solvent accessible surface area (SASA) on the mutants SH2D1A protein. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{x}$) Average of SASA value and ($\Delta\bar{x}$) difference between the average of SASA value between the mutant and the wild type **158**

Figure S19 - Analysis of solvent accessible surface area (SASA) on the mutants SH2D1A protein. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{x}$) Average of SASA value and ($\Delta\bar{x}$) difference between the average of SASA value between the mutant and the wild type **159**

Figure S20 - Analysis of solvent accessible surface area (SASA) on the mutants SH2D1A protein. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G. ($\bar{x}$) Average of SASA value and ($\Delta\bar{x}$) difference between the average of SASA value between the mutant and the wild type **160**

Figure S21 - Frequency of secondary structures of wild-type and mutants SH2D1A proteins during simulation, by residue. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S **161**

Figure S22 - Frequency of secondary structures of wild-type and mutants SH2D1A proteins during simulation, by residue. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N..... **162**

Figure S23 - Frequency of secondary structures of wild-type and mutants SH2D1A proteins during simulation, by residue. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q **163**

Figure S24 - Frequency of secondary structures of wild-type and mutants SH2D1A proteins during simulation, by residue. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G **164**

Figure S25 - Secondary structures of the native protein (WT) per residue during the simulation period. Obtained using the GROMACS DSSP algorithm..... **165**

Figure S26 - Secondary structures of the M1K mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	165
Figure S27 - Secondary structures of the Y7C mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	165
Figure S28 - Secondary structures of the H8D mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	166
Figure S29 - Secondary structures of the G27D mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	166
Figure S30 - Secondary structures of the G27S mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	166
Figure S31 - Secondary structures of the L30S mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	167
Figure S32 - Secondary structures of the R32S mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	167
Figure S33 - Secondary structures of the R32T mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	167
Figure S34 - Secondary structures of the S34G mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	168
Figure S35 - Secondary structures of the S34R mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	168
Figure S36 - Secondary structures of the Y41N mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	168
Figure S37 - Secondary structures of the C42Y mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	169
Figure S38 - Secondary structures of the T53I mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	169
Figure S39 - Secondary structures of the Y54N mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	169

Figure S40 - Secondary structures of the Y54C mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	170
Figure S41 - Secondary structures of the R55P mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	170
Figure S42 - Secondary structures of the R55Q mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	170
Figure S43 - Secondary structures of the I80R mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	171
Figure S44 - Secondary structures of the I84T mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	171
Figure S45 - Secondary structures of the F87S mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	171
Figure S46 - Secondary structures of the G93D mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	172
Figure S47 - Secondary structures of the I98P mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	172
Figure S48 - Secondary structures of the V102G mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.....	172

List of Figures

Chapter I: Exploring X-linked immune disorders: A clinical and molecular overview of X-linked lymphoproliferative syndrome (XLP)

Figure 1 - Graphic summary of the contents present in this literary review.....	3
Figure 2 - Illustrative scheme of the four major clinical manifestations of XLP	7
Figure 3 - Taxonomy of the nine herpesviruses capable of human infection	11
Figure 4 - Illustrative scheme of the Epstein-Barr Virus (EBV)	12
Figure 5 - Representative scheme of the immune response to EBV.....	15
Figure 6 - Cellular interactions of the iNKT cell.....	16
Figure 7 - Summary of XLP-related mutations in the <i>SH2D1A</i> (A) and <i>XIAP</i> (B) genes.....	18
Figure 8 - Representation of ERK, AKT, NF- κ B and NFAT molecular pathways activated by SAP in T and NK cells.....	21

Chapter II: Evolutionary genomic analyses of *SH2D1A*, *SH2D1B* and *XIAP* genes in primates: Implications on immune functionality

Figure 1 - Graphic summary of the work carried out in this study	45
Figure 2 - Predicted protein structures of <i>XIAP</i> (A), <i>SH2D1A</i> (B) and <i>SH2D1B</i> (C) genes Illustrative scheme of the four major clinical manifestations of XLP	47
Figure 3 - Cladogram representation of the phylogenetic tree based on the concatenated <i>XIAP</i> , <i>SH2D1A</i> and <i>SH2D1B</i> genes, created using IQ-TREE 2.0.7 with the Maximum Likelihood method.....	54
Figure 4 - Schematic representation of the structure of the syntenic block of the <i>SH2D1A</i> and <i>XIAP</i> (A) genes, and <i>SH2D1B</i> gene (B)	55

Figure S1 - Cladogram representation of the phylogenetic tree based on the concatenated *XIAP*, *SH2D1A* and *SH2D1B* genes of great apes dataset, created using IQ-TREE 2.0.7 with the Maximum Likelihood method **91**

Figure S2 - Cladogram representation of the phylogenetic tree based on the concatenated *XIAP*, *SH2D1A* and *SH2D1B* genes of Mammalia dataset, created using IQ-TREE 2.0.7 with the Maximum Likelihood method **91**

Figure S3 - Cladogram representation of the phylogenetic tree based on the concatenated *XIAP*, *SH2D1A* and *SH2D1B* genes of primates dataset, created using IQ-TREE 2.0.7 with the Maximum Likelihood method **91**

Chapter III: *In silico* analysis of non-synonymous SNPs and their implications on X-linked lymphoproliferative syndrome type 1 (XLP1)

Figure 1 - Crystal structure of SH2D1A/FYN/SLAM in ternary complex (ID: 1M27). In orange the SLAM protein, in green the SH2D1A protein and in purple the FYN protein. Extracted from the Protein Data Bank (PDB) **100**

Figure 2 - Protein structure of SH2D1A modelled with the sites affected by the SNPs highlighted in red **106**

List of Tables

Chapter I: Exploring X-linked immune disorders: A clinical and molecular overview of X-linked lymphoproliferative syndrome (XLP)

Table 1 - Phenotypic classification for Human Inborn Errors of Immunity (IEI) associated with the X chromosome.....	5
Table 2 - Other clinical conditions associated with XLP	10
Table 3 - Latent proteins present by phase of the EBV latency cycle.....	14

Chapter II: Evolutionary genomic analyses of *SH2D1A*, *SH2D1B* and *XIAP* genes in primates: Implications on immune functionality

Table S1 - List of primate species chosen by family for the study and respective assemblies.....	72
Table S2 - List of mammalian species chosen by order for the study and respective assemblies.....	75
Table S3 - List of species by dataset for each gene. Green blocks mean the presence of the species in the dataset	76
Table S4 - Homo sapiens protein sequences from the UniProt database used as queries for synteny analyses	80
Table S5 - Saturation signature values resulting from Xia's test for each gene in each dataset.....	80
Table S6 - Statistical hypothesis testing (SHT) of the branch-site model analysis of great apes dataset, performed by the PAML program package	80
Table S7 - Statistical hypothesis testing (SHT) of the branch-site model analysis of Mammalia dataset, performed by the PAML program package	81
Table S8 - Statistical hypothesis testing (SHT) of the branch-site model analysis of primates dataset, performed by the PAML program package.....	82

Table S9 - Statistical hypothesis testing (SHT) of the site model analysis for the great apes dataset, performed by the PAML program package	83
Table S10 - Statistical hypothesis testing (SHT) of the site model analysis for the Mammalia dataset, performed by the PAML program package.....	83
Table S11 - Statistical hypothesis testing (SHT) of the site model analysis for the primates dataset, performed by the PAML program package.....	84
Table S12 - Positive sites determined by Bayes Empirical Bayes (BEB) of site model analysis for mammalia dataset,performed by the PAML program package	85
Table S13 - Positive sites determined by Bayes Empirical Bayes (BEB) of site model analysis for primates dataset, performed by the PAML program package.....	85
Table S14 - FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the <i>SH2D1A</i> gene in Mammalia dataset	86
Table S15 - FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the <i>XIAP</i> gene in Mammalia dataset.....	86
Table S16 - FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the <i>SH2D1B</i> gene in Mammalia dataset	89
Table S17 - FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the <i>XIAP</i> gene in primates dataset.....	90
Table S18 - FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the <i>SH2D1B</i> gene in primates dataset.....	90
Table S17 - FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the <i>SH2D1A</i> gene in primates dataset.....	91

Chapter III: *In silico* analysis of non-synonymous SNPs and their implications on X-linked lymphoproliferative syndrome type 1 (XLP1)

Table S1 - List of primate species chosen by family for the study and respective assemblies.....	131
---	------------

Table S2 - Results of the analysis of missense SNPs by the PredictSNP1.0 web server. Mutations marked in orange correspond to SNPs with at least six tests showing deleterious results.....	132
Table S3 - Results of the structural stability analysis of proteins with missense SNPs using the I-Mutant2.0 and MUpro web servers. Mutations marked in orange correspond to SNPs with at least three tests showing deleterious results.....	134
Table S4 - Degree of conservation, molecular pathological mechanism and physicochemical characteristics of selected missense SNPs. Results obtained through ConSurf, MutPred2 and HOPE web servers.	135
Table S5 - Average RMSD values of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, p-values are provided for the entire simulation duration and at specific time intervals, offering insights into structural changes and significance. P-values above 0,05 marked in italic	139
Table S6 - Average RMSF values of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, p-values are provided for the entire structure and at specific amino acid (aa) intervals, offering insights into structural changes and significance. P-values above 0,05 marked in italic	139
Table S7 - Average Rg values of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, p-values are provided for the entire simulation duration and at specific time intervals, offering insights into structural changes and significance. P-values above 0,05 marked in italic	139
Table S8 - Average hydrogen bonds number of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, p-values are provided for the entire simulation duration and at specific time intervals, offering insights into structural changes and significance. P-values above 0,05 marked in italic.....	140
Table S9 - Average SASA value of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, pvalues are provided for the entire simulation duration and	

at specific time intervals, offering insights into structural changes and significance. P-values above 0,05 marked in italic.	140
--	------------

List of Abbreviations

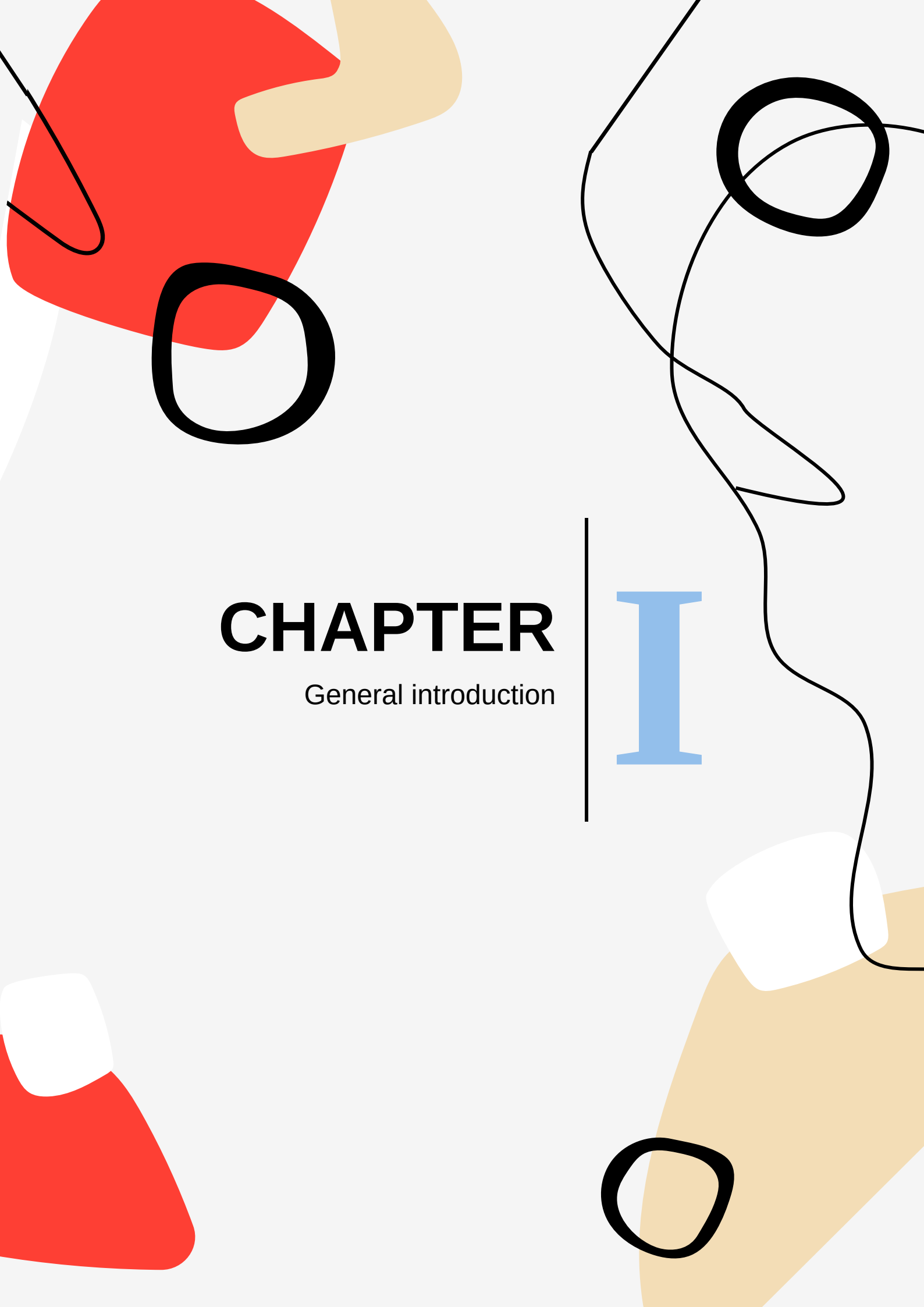
3D	THREE-DIMENSIONAL
AICD	ACTIVATION-INDUCED CELL DEATH
AKT	PROTEIN KINASE B
APC	ANTIGEN-PRESENTING
bb	BOOTSTRAPS
BCL10	B-CELL LEUKEMIA/LYMPHOMA 10
BEB	BAYES EMPIRICAL BAYES
bg	BACKGROUND
BIR	BACULOVIRUS INHIBITOR REPEAT
bp	BASE PAIR
BUSTED	BRANCH-SITE UNRESTRICTED STATISTICAL TEST FOR EPISODIC DIVERSIFICATION
CD	CLUSTER OF DIFFERENTIATION
CDS	CODING SEQUENCE
CNS	CENTRAL NERVOUS SYSTEM
Cα	CARBON ALPHA ATOMS
DAG	DIACYLGLYCEROL
DAMBE5	DATA ANALYSES FOR MOLECULAR BIOLOGY AND EVOLUTION 5
DAS	DISTRIBUTED ANNOTATION SYSTEM
DF	DEGREES OF FREEDOM
DKAα	DIACYLGLYCEROL KINASE ALPHA
<i>dN</i>	NON-SYNONYMOUS MUTATION RATE
DNAM1	DNAX ACCESSORY MOLECULE 1
DNASP	DNA SEQUENCE POLYMORPHISM
<i>dS</i>	SYNONYMOUS MUTATION RATE
DSSP	DEFINE SECONDARY STRUCTURE OF PROTEINS
E(ω X)	POSTERIOR MEAN OF
EAT2	EWING'S SARCOMA ASSOCIATED TRANSCRIPT 2
EBV	EPSTEIN-BARR VIRUS
ERK	MITOGEN-ACTIVATED PROTEIN KINASE
FEL	FIXED EFFECTS LIKELIHOOD
fg	FOREGROUND

fHLH	FAMILIAL HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS
FIM	FULMINANT INFECTIOUS MONONUCLEOSIS
FOXO1	FORKHEAD BOX PROTEIN O1
FUBAR	FAST UNCONSTRAINED BAYESIAN APPROXIMATION
FYN	PROTO-ONCOGENE TYROSINE PROTEIN KINASE
GARD	GENETIC ALGORITHM FOR RECOMBINATION DETECTION
GC	GERMINAL CENTRE
GC	GUANINE-CYTOKINE
gCF	GENE CONCORDANCE FACTOR
GRB2	GROWTH FACTOR RECEPTOR-BOUND PROTEIN 2
HGMD	HUMAN GENE MUTATION DATABASE
HHV-4	<i>HUMAN GAMMAHERPESVIRUS 4</i>
HiFi	HIGH-FIDELITY READS
Hipoy	HIPOGAMMAGLOBULINEMIA
HLH	HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS
HOPE	HAVE YOUR PROTEIN EXPLAINED
HSCT	HEMATOPOIETIC STEM CELL TRANSPLANTATION
IBD	INFLAMMATORY BOWEL DISEASE
iDC	IMMATURE DENDRITIC CELL
IEI	INBORN ERRORS OF IMMUNITY
IFN	INTERFERON
IKK	IKB KINASE COMPLEX
IL	INTERLEUKIN
iNKT	INVARIABLE NATURAL KILLER T CELL
IP₃	INOSITOL 1,2,4 TRISFOSFATO
IR	INTERNAL REPEAT
ITAM	IMMUNORECEPTOR TYROSINE-BASED ACTIVATE MOTIF
ITSM	IMMUNORECEPTOR TYROSINE-BASED SWITCH MOTIF
IUIS	INTERNATIONAL UNION OF IMMUNOLOGICAL SOCIETIES
k	KAPPA
k_i	INICIAL KAPPA VALUE
KIR	KILLER IMMUNOGLOBULIN-LIKE RECEPTOR
LAT	LINKER OF ACTIVATED T CELLS
LCK	LYMPHOCYTE-SPECIFIC TYROSINE PROTEIN KINASE
LFA1	LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 1
LLPC	LONG-LIVED PLASMA CELLS

LMAP	LIGHTWEIGHT MULTIGENE ANALYSES IN PAML
LRT	LIKELIHOOD-RATIO TEST
LUBAC	LINEAR UBIQUITIN ASSEMBLY COMPLEX
LYG	LYMPHOMATOID GRANULOMATOSIS
MA	ALTERNATIVE TWO-RATIO MODEL
MA1	NULL ONE-RATIO MODEL
MAC	MYELOABLATIVE CONDITIONING
MBC	MEMORY B CELL
MD	MOLECULAR DYNAMICS
mDC	MATURE DENDRITIC CELL
MDS	MD SIMULATION
MDSC	MYELOID-DERIVED SUPPRESSOR CELL
MEGA7	MOLECULAR EVOLUTIONARY GENETICS ANALYSIS 7.0
MEME	MIXED EFFECTS MODEL OF EVOLUTION
MHC	MAJOR HISTOCOMPATIBILITY COMPLEX
ML	MAXIMUM LIKELIHOOD
MSA	MULTIPLE SEQUENCE ALIGNMENT
MΦ	MACROPHAGE
NBC	NAÏVE B CELL
NCBI	NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION
NFAT	NUCLEAR FACTOR OF ACTIVATED T-CELLS
NF-κB	NUCLEAR FACTOR KAPPA-LIGHT-CHAIN-ENHANCER OF ACTIVATED B CELLS
NK	NATURAL KILLER
NMR	NUCLEAR MAGNETIC RESONANCE
NOD1	NUCLEOTIDE-BINDING OLIGOMERIZATION DOMAIN- CONTAINING PROTEIN 1
NOD2	NUCLEOTIDE-BINDING OLIGOMERIZATION DOMAIN- CONTAINING PROTEIN 2
np	NUMBER OF PARAMETERS
NPT	CONSTANT NUMBER OF PARTICLES, PRESSURE AND TEMPERATURE
nsSNP	NONSYNONYMOUS SNP
NVT	CONSTANT NUMBER OF PARTICLES, VOLUME AND TEMPERATURE
ONT	OXFORD NANOPORE

oriL	LYTIC ORIGIN OF REPLICATION
p	PROPORTION
PacBio	PACIFIC BIOSCIENCES
PAML	PHYLOGENETIC ANALYSIS BY MAXIMUM LIKELIHOOD
PDB	PROTEIN DATA BANK
PI2K	PHOSPHOINOSITIDE 3-KINASE
PID	PRIMARY IMMUNODEFICIENCY
PIP₂	PHOSPHATIDYLINOSITOL 4,5-BISPHOSPHATE
PIP₃	PHOSPHATIDYLINOSITOL (3,4,5)-TRIPHOSPHATE
PKCθ	PROTEIN KINASE C THETA
PLCγ1	PHOSPHOLIPASE C GAMMA 1
PP	POSTERIOR PROBABILITY
Pr	PROBABILITY
Pr[$\alpha < \beta$]	POSTERIOR PROBABILITY OF POSITIVE SELECTION
Pr[$\alpha > \beta$]	POSTERIOR PROBABILITY OF NEGATIVE SELECTION
pre-T_{FH}	IMMATURE FOLLICULAR T HELPER CELLS
PRR	PATTERN RECOGNITION RECEPTOR
PS	PROBABILITY OF SUCCESS
PSS	POSITIVELY SELECTED SITES
PTK	PROTEIN TYROSINE KINASE
RAC1	RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1
Rg	RADIUS OF GYRATION
RIC	REDUCED-INTENSITY CONDITIONING
RICD	RESTIMULATION-INDUCED CELL DEATH
RING	REALLY INTERESTING NEW GENE
RIPK2	RECEPTOR-INTERACTING SERINE/THREONINE-PROTEIN KINASE 2
RMSD	ROOT MEAN SQUARE DEVIATION
RMSF	ROOT MEAN SQUARE FLUCTUATION
RP	RECOMBINATION POINT
SAP	SLAM-ASSOCIATED PROTEIN
SASA	SOLVENT ACCESSIBLE SURFACE AREA
SC	SITE CLASS
sCF	SITE CONCORDANCE FACTOR
SE	STANDARD ERROR
SH2D1A	SH2 DOMAIN CONTAINING PROTEIN 1A

SH2D1B	SH2 DOMAIN CONTAINING PROTEIN 1B
SHIP1	SH2 DOMAIN CONTAINING INOSITOL POLYPHOSPHATE 5-PHOSPHATASE 1
SHP1	SH2 DOMAIN CONTAINING PHOSPHATASE 1
SHP2	SH2 DOMAIN CONTAINING PHOSPHATASE 2
SHT	STATISTICAL HYPOTHESIS TESTING
SID	SECONDARY IMMUNODEFICIENCY
SLAC	SINGLE-LIKELIHOOD ANCESTOR COUNTING
SLAM	SIGNALING LYMPHOCYTIC ACTIVATION MOLECULES
SLAMF	SLAM FAMILY
SLP76	SH2 DOMAIN CONTAINING LEUKOCYTE PROTEIN OF 76KDA
SLPC	SHORT-LIVED PLASMA CELLS
SNP	SINGLE NUCLEOTIDE POLIMORPHISM
SNP	SINGLE NUCLEOTIDE POLYMORPHISM
SVM	SUPPORT VECTOR MACHINE
TBM	TINGIBLE BODY MACROPHAGES
TCR	T CELL RECEPTOR
TR	TERMINAL REPEAT
TTol5	TIMETREE OF LIFE RESOURCE
U	UNIQUE DOMAIN
UBA	UBIQUITIN-BINDING DOMAIN
VAV1	PROTO-ONCOGENE VAV
WHO	WORLD HEALTH ORGANIZATION
WT	WILD-TYPE
XIAP	X-LINKED INHIBITOR OF APOPTOSIS
XLP	X-LINKED LYMPHOPROLIFERATIVE SYNDROME
XLP1	X-LINKED LYMPHOPROLIFERATIVE SYNDROME TYPE 1
XLP2	X-LINKED LYMPHOPROLIFERATIVE SYNDROME TYPE 2
ZAP70	ZETA-CHAIN-ASSOCIATED PROTEIN KINASE 70
α	MEAN POSTERIOR SYNONYMOUS SUBSTITUTION RATE
β	MEAN POSTERIOR NON-SYNONYMOUS SUBSTITUTION RATE
$\Delta\Delta G$	VARIATIONS IN FREE ENERGY
ω	RATIO OF F NON-SYNONYMOUS TO SYNONYMOUS MUTATION RATES



CHAPTER

General introduction

I

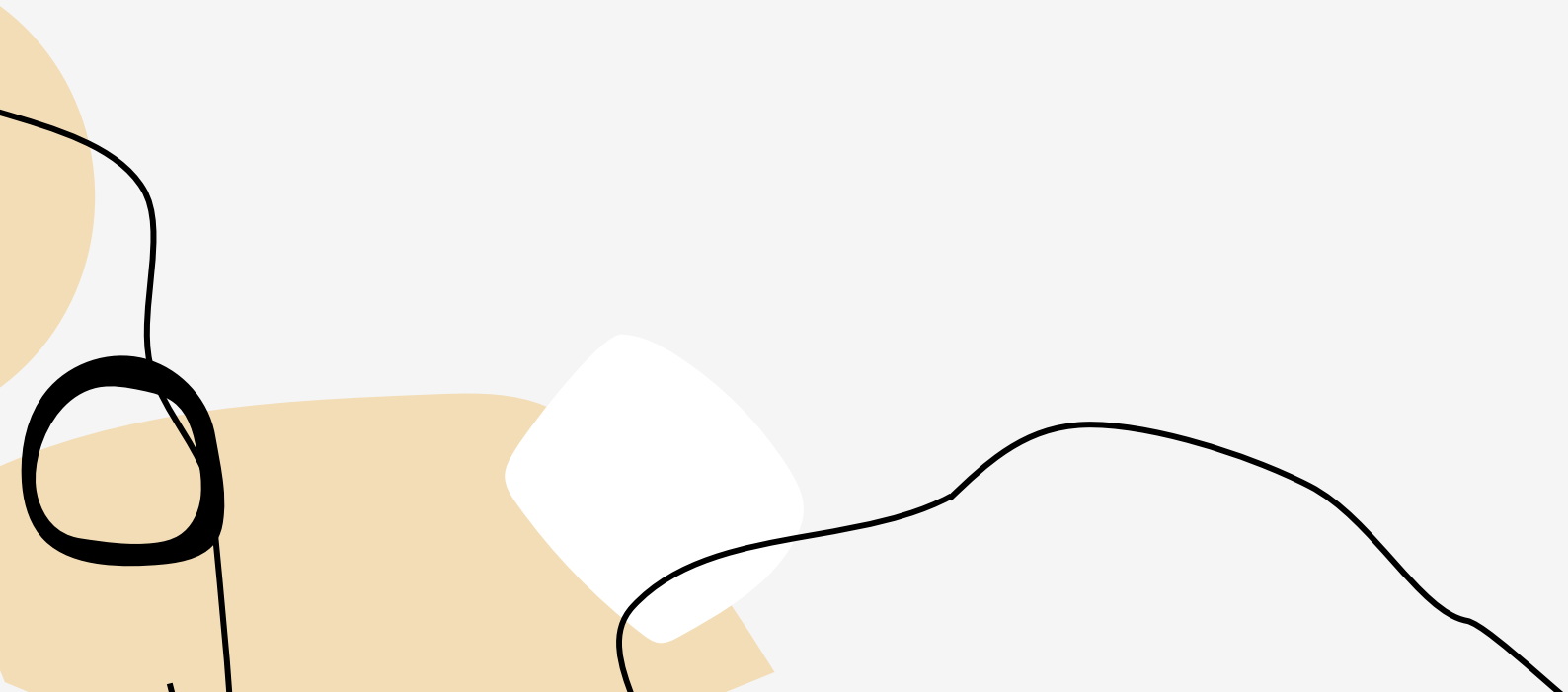


Exploring X-linked immune disorders: A clinical and molecular overview of X-linked lymphoproliferative syndrome (XLP)

Nuno Sousa^{1,2}, Luana Ramos^{1,2}, Agostinho Antunes^{1,2}

¹ Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre, 4169-007, Porto, Portugal

² CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, 4450-208, Porto, Portugal



1.1. Abstract

Most X-linked diseases result from mutations in genes that lead to alterations in the normal functionality and/or expression of the proteins they encode, affecting various organ systems. The X chromosome is enriched in genes associated with the immune system, rendering major relevance to understand immune-related diseases. X-linked immune disorders comprise a wide variety of autoimmune diseases and immunodeficiencies, such as the X-linked lymphoproliferative (XLP) syndrome.

XLP is a rare disease characterised by an uncontrolled and ineffective response to infection by the Epstein-Barr virus, affecting men in the first years of life. This disorder has two subtypes, XLP1 and XLP2, caused by mutations in the *SH2D1A* and *XIAP* genes, respectively. However, XLP1 has a higher prevalence compared to the number of individuals affected by XLP2.

Here we compile the latest information on XLP and evaluate the disease in detail from a segmented and systematic clinical, pathophysiological and molecular perspective. Key points discussed in the review include the symptomatology, available therapies, the effects at the cellular and molecular level on the immune system and its increased susceptibility to Epstein-Barr virus due to XLP. The better understanding of XLP and its effects on the immune system, could assist the development of more effective treatments and therapies for this disease.

Keywords: XLP, X-linked immune disorders, *XIAP* gene, *SH2D1A* gene, deleterious mutations

Graphical abstract

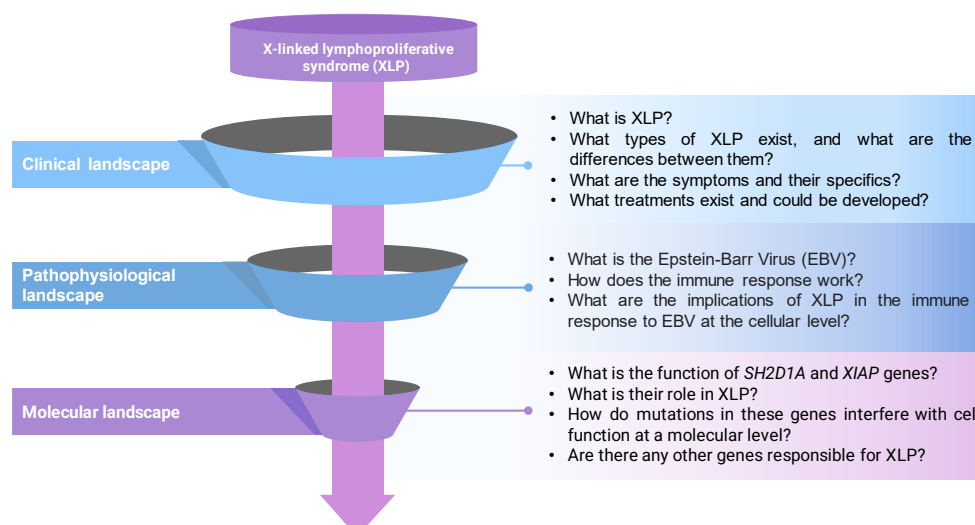


Figure 1: Graphic summary of the contents addressed in this review assessment.

1.2. Introduction

The X chromosome and the inheritance modes

In the Animalia kingdom, several mechanisms are responsible for determining biological sex [1,2]. In eutherian mammals, sexual determination relies on the XY system, a pair of heteromorphic sex chromosomes [2–4]. In males, the sex chromosome pair is composed of one X chromosome (a situation known as genetic unisomy) and one Y chromosome – the heterogametic sex, while the female genome contains two X chromosomes – the homogametic sex [5,6].

The eutherian X chromosome is the longer of the two sex chromosomes and maintains a much higher number of coding genes than the Y [2]. Due to the cytogenetic differences between them, as well as the differentiated chromosomal composition between the two biological sexes, several genes are represented with one X-linked locus in the heterogametic sex (XY) and two loci in the homogametic sex (XX) [2,7], resulting in different doses of transcripts for these genes [2]. To balance the gene dosage in different sexes, eutherian females epigenetically silence one of the X chromosomes randomly (functional unisomy) in each cell of the late blastocyst [2,8,9]. This process is irreversible, forming a cell mosaic in the female body depending on the active X (X_a), which may be the paternal (X_P) or the maternal one (X_M) [10,11].

The inactivation mosaic provides protection against possible pathogenic mutations linked to the genes of the X chromosome in females [12], since they are expressed only in cells in which the active chromosome carries the mutated allele [10,13]. In X-linked recessive mutations, heterozygous females are carriers, while in X-linked dominant they present the phenotype, but with lower severity compared to males [14]. In addition, they present different degrees of symptoms associated with the great variation in genetic expression caused by dosage compensation, X inactivation, cell autonomy, and mosaicism, leading some authors to argue that the terms "X-linked dominant" and "X-linked recessive" should be replaced by the general term "X-linked" [15]. Unlike females, in XY individuals, all cells express the mutations due to the hemizyosity of the alleles, exhibiting the abnormal/diseased phenotype regardless of the inheritance mode [8,13,16].

Therefore, the mosaic contributes to a sexual bias, presenting a defensive role against X-linked diseases in females, contrasting with males that are the most affected [13,17]. Nevertheless, besides the importance of the type of mutation and its inheritance in the

occurrence and severity of the pathogenic phenotypes, other aspects can be involved in this process, such as environmental factors or the remaining genetic background [18,19].

An illustrative example of sexual bias influenced by the distinctive status of the X chromosome in the chromosome set are the variations observed in the progression of cancer between males and females [20]. Due to the genetic and functional unisomy in males and females, respectively, both of the main types of genetic alterations that lead to cancer — tumour-suppressor inactivation and oncogene activation — act dominantly when they affect the single active copy of an X-linked gene [21]. The same alterations remain silent when they affect the inactivated X chromosome in female cells [22,23]. This can lead to differences in cancer progression between males and females [24,25].

Some interesting cases of sexual bias in the inheritance and development of certain cancers related to the genetics of the X chromosome include Klinefelter syndrome, which affects males with an extra X chromosome and is associated with an increased risk of breast cancer, which is usually more common in females [26,27]. These observations suggest that the dosage and inactivation of X chromosome genes may have sex-specific effects on the risk and development of cancer [28]. Another fascinating example is the gene coding for the gastrin-releasing peptide receptor (GRPR), which resists inactivation on the X chromosome. This gene has been identified as a growth factor for lung epithelial cells. Increased GRPR expression is associated with an increased risk of lung cancer in women. The presence of two active copies of the GRPR gene in women may contribute to their increased susceptibility to tobacco induced lung cancer [29–32].

The X-linked immune disorders

Several X-linked diseases are associated with the nervous system (e.g., X-linked adrenoleukodystrophy and Menkes disease) [33,34], muscular system (e.g., Duchenne muscular dystrophy and Becker muscular dystrophy) [35], reproductive system (e.g., hypogonadism and infertility) [36], circulatory system (e.g., hemophilia A/B and X-linked thrombocytopenia) [37,38] and immune system (e.g., X-linked-lymphoproliferative syndrome and X-linked agammaglobulinemia) [39,40]. The majority of these disorders occur mostly due to the occurrence of pathogenic mutations [41] in the gene or the non-coding regions (e.g. promoters or enhancers), leading to alterations in the functionality and expression of the transcripts, respectively [42,43]. In humans, the X chromosome is characterized by having the largest number of genes associated with the immune system [44,45]. Consequently, this chromosome is particularly important for understanding diseases related to immunity [46].

X-linked immune disorders comprise a wide variety of diseases [47]. These diseases are divided into two large groups: *i)* autoimmune diseases, characterized by an atypical immune response in which the immune system reacts against the body's own cells [48–50] and *ii)* immunodeficiencies, defined as an inefficient response of the immune system to pathogens [50,51].

Table 1: Phenotypic classification for Human Inborn Errors of Immunity (IEI) developed by the International Union of Immunological Societies (IUIS), associated with the X chromosome. Information retrieved from [52].

Class	Subgroup	Disease name	Gene
Immunodeficiencies affecting cellular and humoral immunity	Severe combined immunodeficiencies (SCID)	γc deficiency	<i>IL2RG</i>
		SASH3 deficiency *	<i>SASH3</i>
	Combined immunodeficiencies (CID)	Moesin deficiency *	<i>MSN</i>
		CD40 ligand deficiency	<i>CD40LG</i>
CID with Associated Syndrome Features	Congenital thrombocytopenia	Wiskott Aldrich syndrome	<i>WAS</i>
	Anhidrotic ectodermal dysplasia with immunodeficiency	NEMO deficiency syndrome	<i>IKBKG</i>
Predominantly antibody deficiencies		X-linked Agammaglobulinemia	<i>BTK</i>
	Hypogammaglobulinemia	ATP6AP1 deficiency	<i>ATP6AP1</i>
		SH3KBP1 deficiency **	<i>SH3KBP1</i>
Diseases of immune dysregulation		XL magnesium EBV and neoplasia (XNEM)	<i>MAGT1</i>
	Hemophagocytic lymphohistiocytosis and EBV susceptibility	SAP deficiency (XLP1)	<i>SH2D1A</i>
		XIAP deficiency (XLP2)	<i>XIAP</i>
	Syndromes with autoimmunity and others	Immune dysregulation, polyendocrinopathy, enteropathy X-linked (IPEX)	<i>FOXP3</i>
		ELF4 deficiency *	<i>ELF4</i>
Congenital defects of phagocyte		Barth syndrome	<i>TAZ</i>
	Neutropenia	TLR8 gain-of-function *	<i>TLR8</i>
		Pulmonary alveolar proteinosis *	<i>CSF2RB</i>
	Functional defects	XL Chronic granulomatous disease	<i>CYBB</i>
		G6PD deficiency class I	<i>G6PD</i>
Defects in intrinsic and innate immunity	Predisposition to viral infection	Severe COVID19	<i>TLR7</i>
		IRAK-1 deficiency **	<i>IRAK1</i>
	Predisposition to bacterial, fungal and parasitic infections	Macrophage gp91 phox deficiency	<i>CYBB</i>
Complement deficiencies	High level of infection susceptibility	Properdin deficiency	<i>PFC</i>
Bone marrow failure	Fanconi anemia	XL Fanconi anemia	<i>FANCB</i>
	Dyskeratosis congenita	Dyskeratosis congenita progressive decrease	<i>DKC1</i>

* Less than ten cases reported. ** Only one case or kindred has been reported.

Immunodeficiencies are categorized as primary or secondary depending on the origin of the disease [51]. Secondary immunodeficiencies (SIDs) are the result of external causes (i.e., diseases, environmental factors or drug consumption) [53]. On the other hand, primary immunodeficiencies (PIDs) are rare congenital clinical conditions, mostly hereditary [54,55]. PIDs are caused by monogenic germline mutations that result in a gain or loss of function/expression of the protein encoded by the mutated gene [56]. These abnormalities can occur in autosomal chromosomes or in sex chromosomes and are inherited in different ways, with recessive inheritance being the most typical [52,56]. Regarding X-linked PIDs, most are diagnosed in the primordial stages of human life, with a high incidence in males [54]. According to the 2022 updated phenotypic classification for Human Inborn Errors of Immunity (IEI) developed by the International Union of Immunological Societies (IUIS) [52], there are 24 different X-linked inheritance PIDs (Table 1).

Here, we critically revised and compiled available information regarding one of the most severe PIDs, the X-linked lymphoproliferative syndrome (XLP), starting with a comprehensive approach at the clinical level and providing an overview of the current knowledge from a pathophysiological perspective, ending with a molecular characterization of the pathology (Figure 1).

1.3. The clinical landscape of X-linked lymphoproliferative syndrome

What is the X-linked lymphoproliferative syndrome?

The X-linked lymphoproliferative syndrome (XLP) [57,58] is a rare deficiency in the immune system, [59] typified by an uncontrolled immune reaction to the Epstein-Barr virus (EBV) infection [58,60–62]. It has a worldwide prevalence of less than 0.0001% [61], affecting individuals in their first years of life [63]. Since it is an inherited recessive genetic disorder linked to the X chromosome, male individuals are usually most severely affected [58], while female individuals who carry the abnormality are asymptomatic or present a less severe phenotype, depending on whether the X chromosome with the mutation is inactivated or not [64,65]. Affected individuals have severe mononucleosis, hypogammaglobulinemia, hemophagocytic lymphohistiocytosis (HLH) [59,61,66,67], and may present other clinical manifestations more specific to each disease variant, such as the presence of lymphomas (Figure 2 and Table 2) [57]. Patients with XLP typically

show clinical manifestations only after EBV infection. However, there are cases of patients with XLP who develop lymphoma and HLH even without being infected with the virus [68,69].

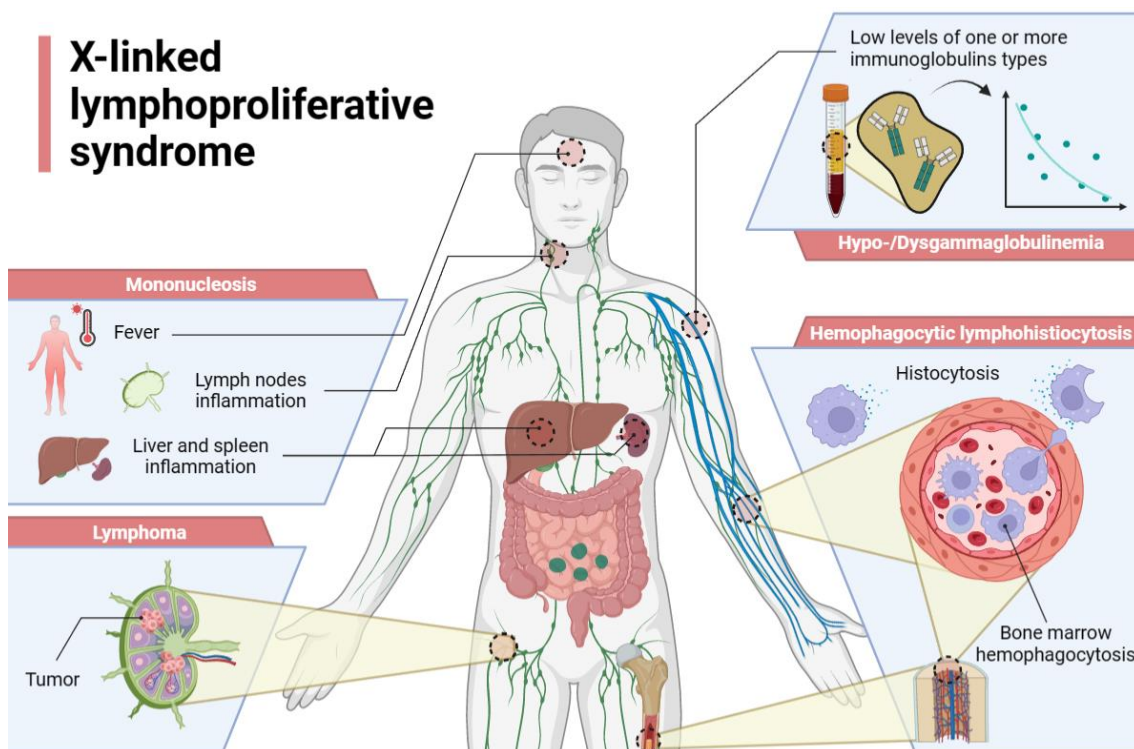


Figure 2: Illustrative scheme of the four major clinical manifestations of XLP. Created with [BioRender.com](https://www.biorender.com)

There are two different forms of this disease depending on the affected gene [61]: i) Type 1 (*SH2D1A*-deficiency), the most common form caused by a deficiency of the *SH2D1A* (SH2 Domain Containing Protein 1A) gene (XLP1, OMIM #308240) [59,61,70], and ii) Type 2 (*XIAP*-deficiency), which is caused by a deficiency of the *XIAP* (X-linked Inhibitor of Apoptosis) gene (XLP2, OMIM #300635) [61,66]. The molecular mechanism responsible for the pathogenicity of XLP1 and XLP2 remains mostly unclear [61,71].

Characterization of clinical manifestations of XLP

Infectious mononucleosis (IM)

IM, also known as kissing disease [72,73], is a viral disease caused in most cases by the Epstein-Barr virus (EBV) [74]. It has an incubation period of 1 to 2 months and is rare in older adults, affecting mostly young people in the age group of 15 to 24 years old [75]. People with this disease can show various symptoms such as fever, fatigue, inflamed lymph nodes (lymphadenopathy) and pharyngitis [76,77]. Due to the importance of the spleen in the filtration and purification of blood [78], splenomegaly and subsequent rupture may occur in certain cases [72]. In infected individuals with immune disorders

such as XLP, fulminant infectious mononucleosis (FIM) may occur, resulting in a hemophagocytic reaction and uncontrolled lymphocyte proliferation [76].

Hemophagocytic lymphohistiocytosis (HLH)

HLH is a rare disorder characterized by unregulated hyperactivity of the immune system, being potentially fatal if untreated [79]. It results from the unbalanced production and activation of immune cells, such as lymphocytes, leading to the release and accumulation of high levels of cytokines (cytokine storm) that inflame various tissues, resulting in multiple organ damage [80]. This clinical condition exists in two forms: *i)* Genetic HLH, also called primary HLH, which occurs due to genetic mutations, as in familial hemophagocytic lymphohistiocytosis (fHLH) or in XLP, appearing in the first years of life; and *ii)* Acquired HLH, called secondary HLH, derived from pathogens or malignancies, which tends to occur more in adolescents and young adults [81,82]. In general, both types have very similar symptomatology, such as the occurrence of fever, splenomegaly, cytopenia, hemophagocytosis in the bone marrow, spleen, or lymph nodes, and low/no activity of Natural Killer (NK) cells [83,84].

Hypogammaglobulinemia (hipoy)

Hipoy is an immunodeficiency defined by insufficiency of antibodies (immunoglobulins) in the blood [85]. It can be classified as a PID or SID when it results from genetic factors or medical treatments, such as the use of immunosuppressant drugs, respectively [86]. The main symptoms of this clinical manifestation are recurrent infections, lymphadenopathy, splenomegaly, allergic diseases, malignancies, cardiovascular, otorhinolaryngological and developmental abnormalities [87]. In some cases, dysgammaglobulinemia (another type of antibody insufficiency) may occur instead of hypogammaglobulinemia, which does not affect all forms of immunoglobulins, but only one or more types [58].

Lymphoma

Lymphoma is a tumour that originates in the lymphatic system and can diverge into any organ [88]. Depending on the type of cells that originate them, and according to the 5th edition of the World Health Organization (WHO) Classification of Haematolymphoid Tumours, lymphatic neoplasms have been categorized into three major groups: *i)* B-cell lymphomas, *ii)* T-cell and NK cell lymphomas and *iii)* Stroma-derived neoplasms of lymphoid tissues [89]. In patients with XLP, the most common form is B-cell lymphomas, mostly of the Burkitt type, formed in germinal centres, usually localized in the ileocecal area [58,68]. Hodgkin lymphomas (a B-cell lymphoma type) and T-cell lymphomas,

although rare, have also been observed [58]. The main symptoms include lymphadenopathy, fever, night sweats, fatigue, weight loss and rashes [88,90].

Table 2: Other clinical conditions associated with XLP.

Splenomegaly	Increase in the volume of the spleen, resulting from the process of hyperinflammation [72].
Vasculitis	Restriction of blood flow by thickening of the walls of the blood vessels, caused by inflammation, occurs normally in the central nervous system (CNS) [91].
Colitis	One of the most common types of intestinal inflammation marked by persistent inflammation in the colon, causing rectal bleeding and diarrhea [92]. It's part of inflammatory bowel disease (IBD) along with Crohn's disease. In some XLP cases, symptoms similar to the second type of IBD have been observed [93], emphasizing the link between genetics and gastrointestinal disorders.
Aplastic anemia (AA)	A clinical manifestation characterized by insufficient production of blood cells by the bone marrow [94].
Lymphomatoid granulomatosis (LYG)	Characterized by benign proliferation of lymphocytes after EBV infection, which may progress to lymphoma, occurring in patients with XLP1 [95].

Therapies and Treatments: What exists and future prospects

Allogeneic hematopoietic stem cell transplantation (HSCT) is known as the only curative treatment for X-linked lymphoproliferative disease (XLP) [57]. However, this treatment approach is associated with diverse complications, being performed in two different forms [96]. The first is the myeloablative conditioning (MAC) protocol, which involves the complete destruction of the patient's bone marrow using high doses of radiation or chemotherapy, followed by the infusion of the donor's bone marrow [97]. The second is the reduced intensity conditioning (RIC) protocol, in which the patient's bone marrow activity is suppressed using low doses of radiation or chemotherapy, without complete elimination [98,99]. Clinical studies have shown

that RIC is the most effective form of HSCT for XLP patients, resulting in improved post-operative outcomes and higher survival rates of 70% and 86% for *SH2D1A*-deficiency and *XIAP*-deficiency, respectively [96].

In addition to HSCT, other therapeutic approaches are used to manage the clinical manifestations of XLP. These include immunological and chemotherapeutic therapies aimed at antiviral prophylaxis, immunoglobulin restoration, and chemotherapy to prevent viral infections, restore antibody levels, and combat cancer cells [57]. Emerging possibilities for future curative treatments of XLP include gene therapy and targeted modulation of specific immune response pathways. Gene therapy involving lentiviral vectors carrying the functional *SH2D1A* gene has shown promising results in laboratory mice with XLP-like symptoms [100,101]. Another avenue of research explores the use of compounds that inhibit or modulate specific immune response pathways. For instance, the discovery of a potent inhibitor for diacylglycerol kinase alpha (DKA α) has shown the restoration of restimulation-induced cell death (RICD) in *SH2D1A*-deficient cells [102,103].

Given the phenotypic complexity of XLP, finding a universal treatment that is effective for all cases remains challenging [57]. Further unravelling and understanding of the molecular and pathophysiological mechanisms underlying the disease are necessary to identify precise therapeutic targets, enabling personalized treatments based on the individual patient's XLP clinical condition in the future.

1.4. The pathophysiological landscape of X-linked lymphoproliferative syndrome

The Epstein–Barr virus

Herpesviridae is a family comprising over 130 different enveloped viruses that belong to Group I in the Baltimore classification (viruses with double-stranded DNA as their genetic material) [104,105]. This family is one of the largest among DNA viruses and has a broad host range that includes mollusks and various vertebrates. Of the numerous herpes viruses, only 9 are known to infect humans (Figure 3) [106]. The *Herpesviridae* family is further classified into three major groups [104,107,108].

All three subdivisions are known to establish latency periods in different types of tissues [109]. During latency, the viruses can remain dormant within the host, with the potential to reactivate and cause recurrent infections in the future [110].

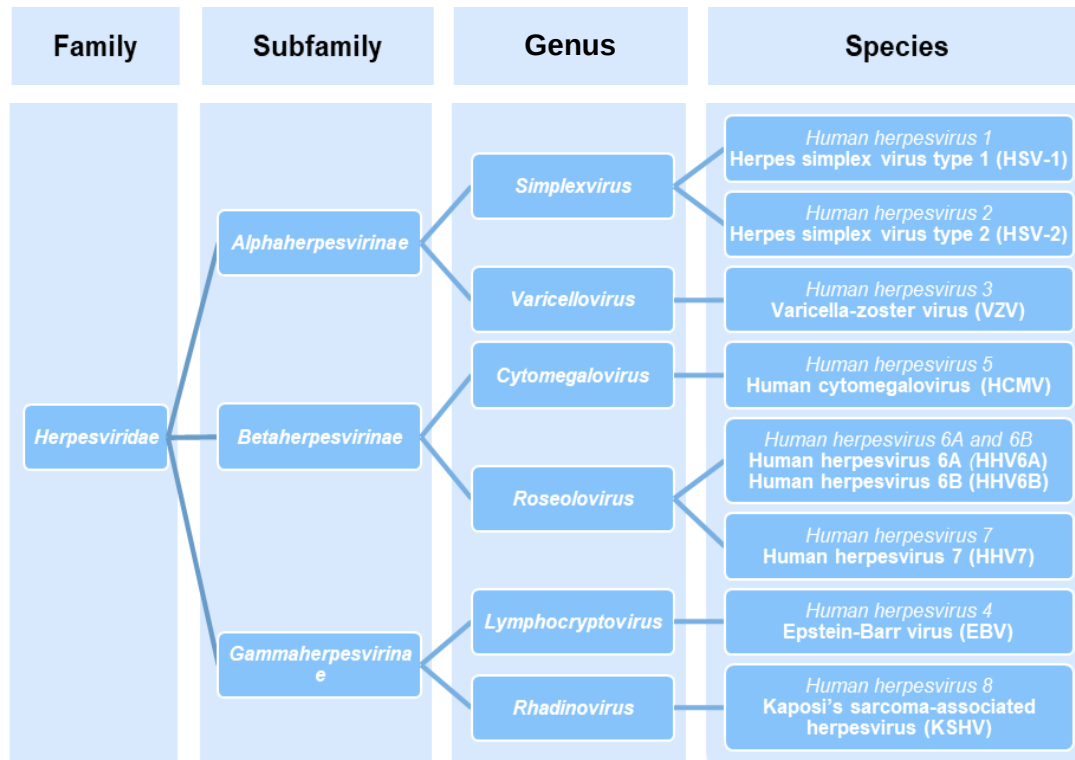


Figure 3: Taxonomy of the nine herpesviruses capable of human infection. Adapted from [111].

The *Human gammaherpesvirus 4* (HHV-4), commonly known as Epstein-Barr virus (EBV), belongs to the *Lymphocryptovirus* genus within the *Gammaherpesvirinae* subfamily, and the human species is its only known host [112]. As its genus name implies, EBV not only causes infectious mononucleosis but is also implicated in various lymphoproliferative disorders, including different types of B-cell lymphomas and other related conditions (about 1,5% of cancers worldwide, with most cases occurring in immunocompromised patients) [113].

Structurally, this virus features a viral envelope composed of a bilipid layer derived from the host cell (Figure 4A). Embedded within this envelope are vital glycoproteins (Figure 4B) that play a crucial role in the binding process to target cells [114]. Inside the virus, the genetic material is enclosed within an icosahedral nucleocapsid [115], a protein shell bonded to nucleic acids [116]. This nucleocapsid possesses a single protein portal vertex responsible for injecting the viral DNA into the host cell [117]. Situated between the envelope and the nucleocapsid is the viral tegument, which

consists of a diverse array of proteins that not only suppress the host immune response but also facilitate viral DNA replication [115,118].

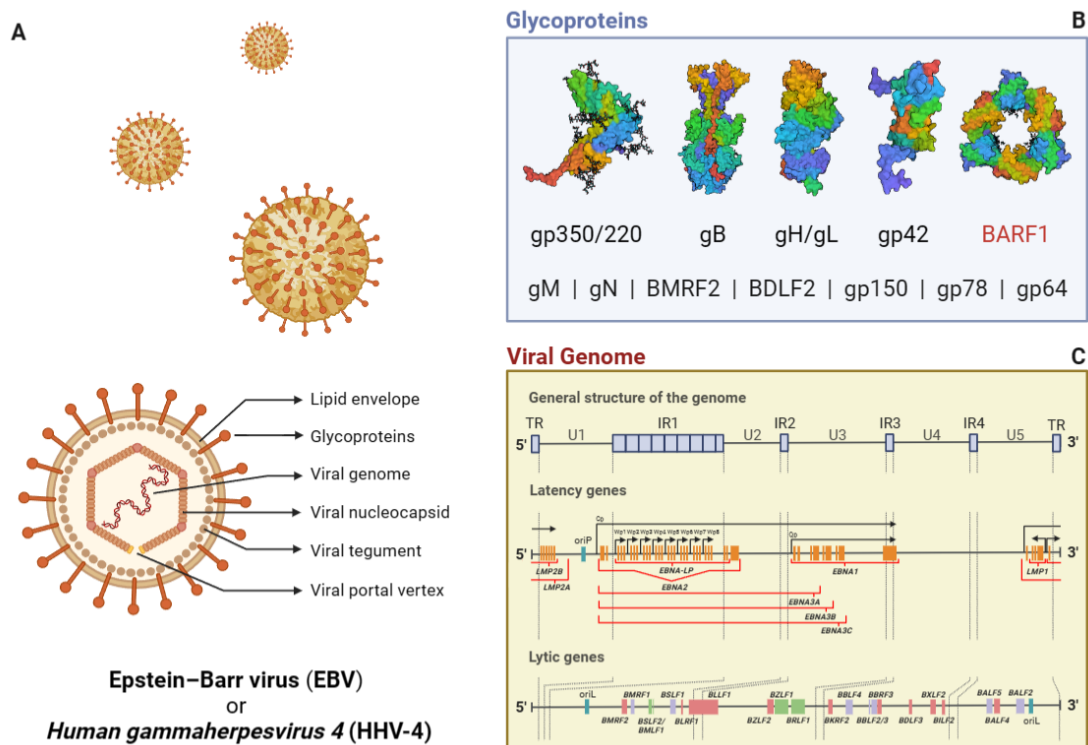


Figure 4: Illustrative scheme of the Epstein-Barr Virus (EBV). **(A)** Illustration depicting the structure of EBV. **(B)** Glycoproteins present in the viral envelope of EBV, excluding BARF1 (highlighted in red), which is secreted. **(C)** Representation of the EBV viral genome. The lytic genes include immediate early protein-coding genes (shown in green), main early protein-coding genes responsible for DNA replication through the lytic origin of replication (oriL) (in lilac), and key genes encoding glycoproteins classified as late proteins (in red). Created with [BioRender.com](https://www.biorender.com)

The genome of EBV (Figure 4C) comprises over 85 genes, with a total length of approximately 172 kb [119]. Typically, the viral genome can be divided into five unique domains (U1-U5), four large internal repeats (IR1-IR4), and the terminal repeats (TR) [120].

EBV infection and immune response

The virus is transmitted through saliva, initiating the infection process by infecting B cells in the lymphatic tissues, after invading lytically the epithelial cells of the oropharyngeal area. The life cycle of EBV typically occurs in epithelial cells and B cells, but in some cases, it can also involve NK cells and T cells. This cycle is divided into two phases: the lytic cycle and the latent cycle, each activated by different genetic programs (Figure 4C).

During the latent cycle, there are four phases characterized by the expression of specific latent genes (Table 3). In latency phase 3, all latent genes are expressed, leading to the proliferation of infected B cells. In phase 2, only some genes are expressed to facilitate the differentiation of B cells into memory B cells (MBC). Phase 1 involves the expression of the *EBNA1* gene, which facilitates the replication of the EBV genome during cell division in MBCs. Latency 0 is marked by the absence of viral protein expression, enabling the virus to persist in the body without detection.

Table 3: Latent proteins present by phase of the EBV latency cycle. Cells painted in green correspond to the positive presence of the protein in the respective phase.

Fase of latency	Latent proteins								
	EBNA-LP	EBNA1	EBNA2	EBNA3A	EBNA3B	EBNA3C	LMP1	LMP2A	LMP2B
Latency 0									
Latency 1									
Latency 2									
Latency 3									

The infection can occur via two routes, depending on whether the virus targets MBCs or Naive B cells (NBC). When infecting NBCs (Figure 5I), these cells become activated (primed), process the antigen, and present it to immature follicular T helper cells (pre-T_{FH}) at the border between the B/T cells zones (Figure 5II) [121]. Within the B follicle, the pre-T_{FH} cells stimulate the proliferation of primed B cells [122]. As these cells are in latency phase 3 and express all viral proteins, the majority of proliferating cells are eliminated by T cells and NK cells [123]. The surviving cells have three potential paths: *i*) becoming Germinal Center (GC)-dependent B cells; *ii*) becoming GC-independent B cells; or *iii*) differentiating into plasmablast cells, which further mature into short-lived plasma cells (SLPC) (Figure 5III) [124,125]. The likelihood of entering in the GC is higher when there is a stronger interaction between pre-T_{FH} cells and primed B cells.

The interaction between pre-T_{FH} cells and GC-dependent memory B cells allows for the development of GC T_{FH} and centroblast cells, respectively [122]. Centroblasts proliferate and experience somatic hypermutation in the dark zone of the GC (latency 2), generating variability in antigen affinity [126]. Following this process, centroblasts migrate to the light zone and transition into centrocytes, initiating the selection process by follicular dendritic cells (FDC) [122]. The FDC present antigens on their surface to select centrocytes with higher affinity. Additionally, centrocytes are selected by GC T_{FH} cells, which present processed antigens via the major histocompatibility complex II (MHC-II) to the T cell receptor (TCR), with the

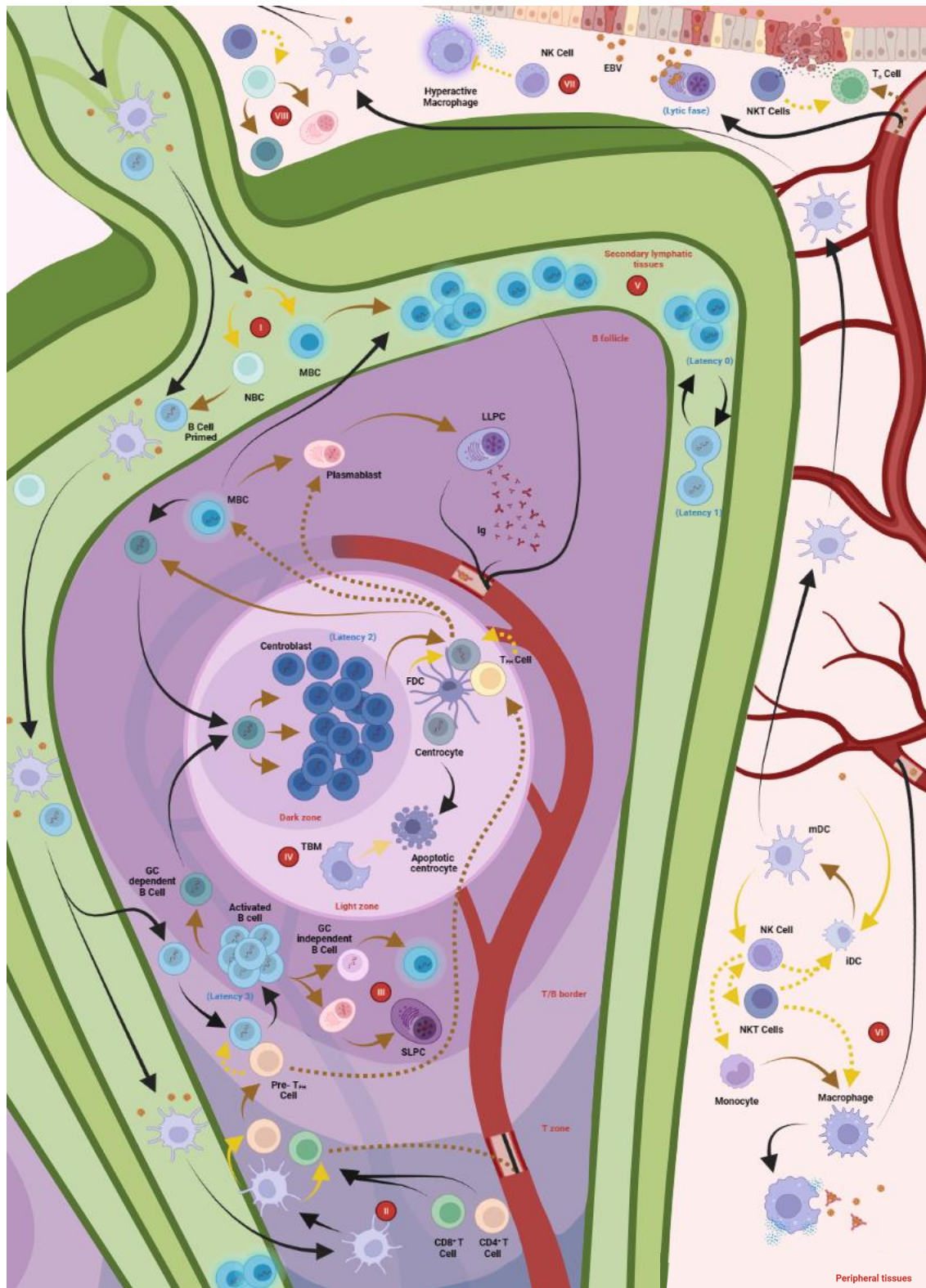


Figure 5: Representative scheme of the immune response to EBV. Brown arrows represent differentiation processes, while the yellow ones represent cell interactions. Dashed arrows show possible altered immune processes in patients with XLP. (I) EBV infection of B cells in lymphatic tissues. (II) Maturation and differentiation of T cells in the T zone of the lymph node. (III) MBC reservoir. (IV) Differentiation and maturation process of GC-dependent B cells. (V) Production of specific antibodies against EBV. (VI) Differentiation and maturation of dendritic cells and macrophages. (VII) Suppression of hyperactivity of macrophages. (VIII) Differentiation of B cells into GC B cells and plasmablast. Created with [BioRender.com](https://www.biorender.com)

assistance of various co-stimulatory molecules [127]. Accepted centrocytes receive signals for selection that promote their differentiation into GC memory B cells and long-lived plasma cells (LLPC). Cells that do not receive these signals undergo apoptosis and are cleared from the GC by tingible body macrophages (TBM) (Figure 5IV) [128]. GC MBCs transition to the MBC reservoir, in the peripheral blood and tissues (Figure 5V) (latency 0).

When EBV targets MBCs (Figure 5I), it is believed that these cells also proliferate within the B follicle (latency 3) and directly enter in the MBC reservoir, where they remain in a resting state (latency 0) [112].

In both paths, resting MBCs carrying the viral genome in the reservoir have the potential to enter latency phase 1 and divide. Both GC MBCs and GC-independent MBCs can differentiate into LLPCs, initiating the lytic cycle [121]. The virions produced during the lytic cycle have the ability to infect other naive/memory B cells and epithelial cells, ultimately reaching saliva and transmitting the virus to other individuals [123].

T and NK cells: The crucial role in the immune response

Although B cells are the most predominant during the EBV infection, other cells, such as the NK and T cells, are equally important, especially for the defense process against the virus.

NK cells play a crucial role in fighting EBV, being the primary defense against the initial infection [132]. These cells not only recognize and destroy infected B cells but also eliminate hyperactive macrophages (MΦ) (Figure 5VII). Furthermore, NK cells promote the maturation and activation of both these cells and immature dendritic cells (iDC) to mature dendritic cells (mDC) (Figure 5VI) [133,134].

In case of EBV infection, cytotoxic T cells (Tc), follicular helper T cells (TFH), and invariable (type I) natural killer T cells (iNKT) are the key T cell subtypes with crucial roles in the immune response. iNKT cells have a significant impact, contributing to the activation of NK cells and assisting in the maturation of immature dendritic cells (iDCs) and other T cells [135]. Additionally, iNKT cells can induce the differentiation of B cells into plasmablasts or GC-dependent B cells and can inhibit myeloid-derived suppressor cells (MDSC) which, in turn, are responsible for inhibiting T cells, macrophages and NK cells (Figure 6) [129,136].

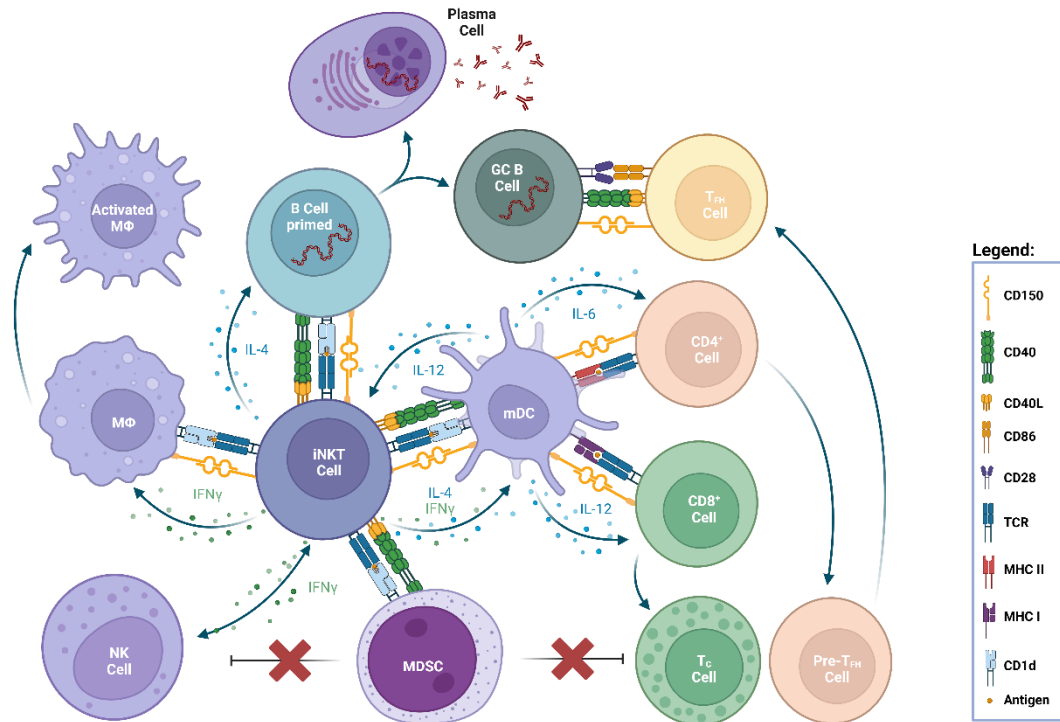


Figure 6: Cellular interactions of the iNKT cell. Adapted from [129–131]. Created with [BioRender.com](https://www.biorender.com)

All these cell types are primarily activated by antigen-presenting cells (APCs), such as mature dendritic cells (mDCs), through diverse types of chemical signals like the great spectrum of interleukins (IL) and interferon (IFN) [130,137]. Cell recognition and signaling are crucial for the immune response. Cell-surface proteins facilitate attachment and binding, initiating intracellular messenger cascades. These cascades trigger specific responses, such as the release of chemical messengers, essential for immune activation. Examples of such proteins include SH2D1A and XIAP, which are involved in intracellular cascades triggered by intercellular binding [58]. Mutations in these *SH2D1A* and *XIAP* genes often lead to dysfunctions, mainly of the different types of T and NK cells, or even resulting in lack of cells (iNKT in XLP1), compromising the various cellular pathways of the immune response and leading to different phenotypes [58,138].

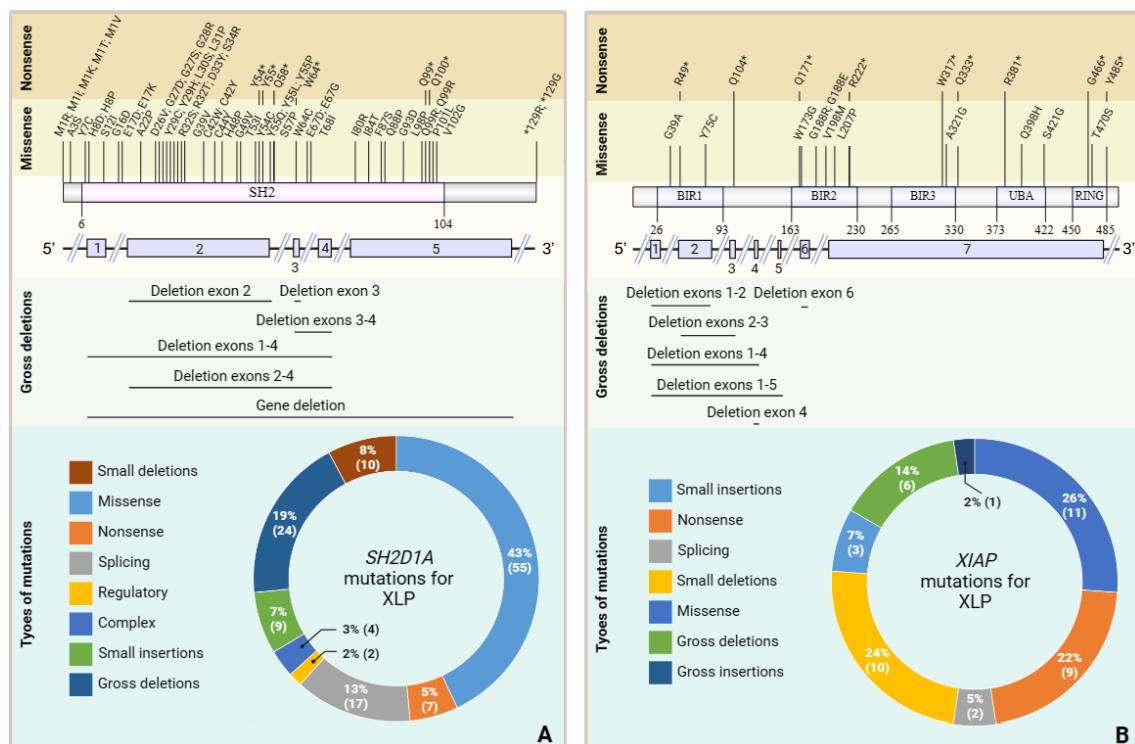
1.5. The molecular landscape of X-linked lymphoproliferative syndrome

SH2D1A and *XIAP*

The human SH2 Domain Containing 1A (*SH2D1A*) encodes the signaling lymphocyte activation molecule (SLAM) associated protein, which consists of 128 amino acids [139]. The *SH2D1A* protein contains a single SH2 (Src Homology 2) domain, allowing it to bind

to proteins with phosphorylated tyrosines (pTyr), such as in the immunoreceptor tyrosine-based switch motifs (ITSM), in the SLAMs. During the interaction with SLAMs, the SH2D1A protein interacts with the protein encoded by the paralogous gene *SH2D1B* (SH2 Domain Containing Protein 1B) [140]. The gene is primarily expressed in several types of T and NK cells, although it is also present in B cells, for instance, in reduced quantities. Its expression is limited to a specific range of tissues (peripheral blood leukocytes, thymus and lung) [141].

The human X-linked Inhibitor of Apoptosis (*XIAP*) encodes a protein that inhibits apoptosis and modulates immune signaling [139,142]. It consists of 497 amino acids and has five domains: one ubiquitin-binding domain (UBA), three baculovirus inhibitor repeats (BIR) domains and a RING domain [143]. *XIAP* is expressed in multiple tissues and is predominantly found in T and NK cells [144].



According to the HGMD Professional 2023.1, a total of 128 mutations related to XLP have been identified in the *SH2D1A* gene, while 42 mutations have been found in the *XIAP* gene (Figure 7) [145]. Nearly half of these mutations involve amino acid changes (missense and nonsense).

SH2D1A signaling pathways

As previously mentioned, T cells (T helper, follicular helper, cytotoxic and NKT cells) and NK cells are the main groups of cells affected in *SH2D1A* deficiency (XLP1), mainly affecting their development and activation and, consequently, the immune response.

In the case of the *SH2D1A* protein signaling, the development processes and cell activation of T and NK cells depend on the SLAM family (SLAMF), a group of transmembrane glycoproteins. They possess extracellular IgV and IgC2 domains, which facilitate interactions with other molecules, and intracellular ITSMs, which are involved in signaling events [146]. The ITSMs are conserved amino acid sequences with which the *SH2D1A* protein has a high affinity [147].

When the *SH2D1A* protein binds to ITSM, it recruits protein tyrosine kinases (PTKs) such as lymphocyte-specific protein tyrosine kinase (LCK) and proto-oncogene tyrosine protein kinase (FYN), which possess SH3 domains capable of binding to the SLAM-*SH2D1A* complex [148]. Although it does not require the tyrosine of ITSM to be phosphorylated, the *SH2D1A* protein exhibits high affinity and binding power for the phosphorylated residue [140]. Without *SH2D1A* protein binding to the ITSMs, SH2 domain-containing phosphatases (SH2 domain containing inositol polyphosphate 5-phosphatase 1 – SHIP1, SH2 domain containing phosphatase 1 – SHP1, and SH2 domain containing phosphatase 2 – SHP2) bind to the phosphorylated tyrosine (pTyr) of ITSM motifs, resulting in SLAM dephosphorylation and subsequent inhibition of the signaling cascade [146].

Besides the SLAMF proteins, there are several membrane receptors not associated with *SH2D1A* protein that are also responsible for the development and activation of T and NK cells, triggered by reception of extracellular stimuli, such as the TCR, DNAX accessory molecule 1 (DNAM1), killer immunoglobulin-like receptor (KIR) or NKG2A receptors [149].

All these receptors play important roles, triggering several internal signaling pathways that lead to diverse cellular responses. Within these paths, there are four major pathways responsible for most cellular immune processes in the T and NK cells: i) the AKT

pathway; *ii*) The MAPK/ERK pathway, *iii*) the Calcium-Calcieneurin-NFAT pathway and *iv*) the NF- κ B pathway [150]. Although these signaling pathways exist in different types of T and NK cells, some receptors and intermediate proteins in the signaling pathways, as well as the product of signal transduction, change with the cell type (Figure 8).

T cells signaling pathways

In T cells, the SH2D1A protein has the ability to bind directly to the immunoreceptor tyrosine-based activate motif (ITAM) of the cluster of differentiation (CD) 3 zeta (CD3 ζ) chains of the TCR. This link allows for LCK recruitment and activation. Activated LCK will phosphorylate tyrosine residues on the CD3 ζ chains of the TCR, creating binding sites for the zeta-chain-associated protein kinase 70 (ZAP70). LCK also phosphorylates and activates ZAP70, which in turn phosphorylates the linker of activated T cells (LAT) adapter protein and other substrates. LAT phosphorylation allows the formation of a signaling complex that involves other proteins such as SH2 domain containing leukocyte protein of 76kDa (SLP76), growth factor receptor-bound protein 2 (GRB2) and phospholipase C gamma 1 (PLC γ 1) [151]. This complex leads to the production of second messengers such as diacylglycerol (DAG) and inositol 1,2,4 trisphosphate (IP $_3$), which activate pathways such as MAPK/ERK, NF- κ B and Calcium-Calcieneurin-NFAT [152–155]. In addition, SH2D1A protein eventually influences the AKT pathway, via the phosphorylation of phosphoinositide 3-kinase (PI3K), which converts phosphatidylinositol 4,5-bisphosphate (PIP $_2$) into phosphatidylinositol (3,4,5)-triphosphate (PIP $_3$) [156]. When SH2D1A is mutated, these pathways eventually weaken, potentially resulting in the T cell dysfunctional phenotypes that we find in XLP1 patients.

Among these pathways, it is important to highlight the NF- κ B pathway, which plays a fundamental role in the ontology of NKT cells and in the development of thymocytes into CD4 $^+$ T cells, which largely explains the absence of NKT cells and the dysfunctionality of T cells in SH2D1A-deficiency. FYN is recruited via the SLAMF/SH2D1A complex, which phosphorylates and activates the protein kinase C theta (PKC θ), recruiting it to the TCR stimulated by the immune synapse. PKC θ in turn activates B-cell leukemia/lymphoma 10 (BCL10), which activates the I κ B kinase (IKK) complex. Phosphorylation of the IKK complex signals its degradation, releasing and allowing NF- κ B dimers to migrate into the nucleus and activate transcription of genes essential for the development of T cells, leading to the release of cytokines, particularly Th2 type cytokines such as IL-4 [157].

Although this pathway is similar and important in the diverse types of T cells, it plays an important role in NKT cells. In contrast, SH2D1A protein in CD4 $^+$ T cells does not play an

essential role in cytokine production but, instead, optimizes the production process. This is demonstrated by the absence of a significant impact on cytokine production in the case of FYN mutations, suggesting the existence of another pathway independent of FYN [157]. Recent studies show that in mice the development of conventional T cells is inhibited in the absence of LCK but remains normal in the absence of FYN [153], which could be a possible justification for the situation described above.

NK cells signaling pathways

Despite being involved in all four major molecular pathways controlling T and NK cell immune responses, the SH2D1A protein has a predominant function in T cell activation and development, with particular emphasis on NKT cells [158]. In the case of NK cells, the signaling pathways may proceed independently of SH2D1A activation, indicating that, once again, this protein does not play a critical role in the developmental and activation processes of these cells, but rather in its optimization.

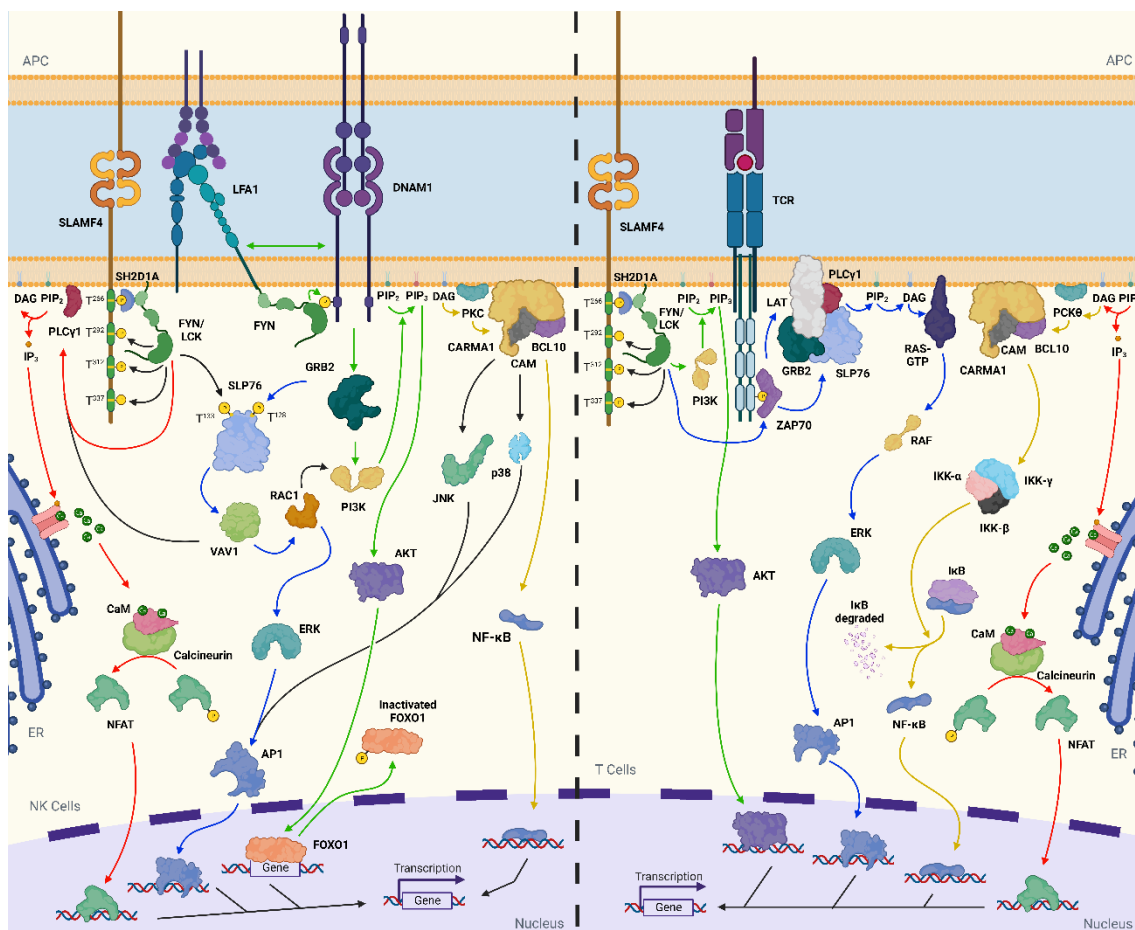


Figure 8: Representation of MAPK/ERK (blue arrows), AKT (green arrows), NF-κB (yellow arrows) and Calcium-Calcineurin-NFAT (red arrows) molecular pathways activated by the SH2D1A protein in T and NK cells. Black arrows represent accessory interactions that influence the pathway. Created with [BioRender.com](https://www.biorender.com)

In NK cells, the SLAMF/SH2D1A protein complex recruits FYN that phosphorylates the proto-oncogene vav (VAV1), a guanine nucleotide exchange factor. VAV1 can activate ras-related C3 botulinum toxin substrate 1 (RAC1), a small GTPase that activates PI3K, which produces PIP₃ from PIP₂. PIP₃ can recruit AKT that mediates various cellular processes, such as survival, proliferation and metabolism [159]. AKT phosphorylates the forkhead box O1 (FOXO1) protein present in the nucleus, which is transcriptionally inhibiting the genetic expression. Upon phosphorylation, FOXO1 disassociates from the DNA molecule and exits the nucleus, allowing the transcription of inhibited genes responsible for cell growth and survival. However, unlike DNAM1 and lymphocyte function-associated antigen 1 (LFA1) integrin, SLAM is not one of the main receptors responsible for activating the AKT pathway and the other pathways [160–163].

Therefore, the SLAM/SH2D1A protein complex can influence the AKT activation pathway in NK cells indirectly through FYN and VAV1. FYN triggers phosphorylation of VAV1 and PLCγ, as previously mentioned. VAV1, in turn, has the ability to activate PLCγ as well. This phospholipase can activate the NFAT, NF-κB, and ERK pathways by itself, leading to cytotoxicity, degranulation, and cytokine release [162,164–166].

XIAP signaling pathways

The XIAP protein can activate NF-κB and ERK signalling pathways through nucleotide-binding oligomerization domain-containing protein 1/2 (NOD1/2) and Dectin-1 signalling, which are intracellular pattern recognition receptors (PRR) involved in the innate immune response. XIAP mediates ubiquitination of the receptor-interacting serine/threonine-protein kinase 2 (RIPK2) and recruits the linear ubiquitin assembly (LUBAC) complex to the NOD receptor. XIAP also binds and ubiquitinates BCL10, an essential step for NF-κB and ERK activation through Dectin-1 signalling. Thus, XIAP activates both pathways but depends on the type of PRR stimulated [142]. Nevertheless, NK and T cell development and immune responses, particularly their cytotoxic capacity, do not appear to be impaired in individuals with XLP2. This indicates that XIAP does not play a critical role in cytokine production and release. However, in XIAP-deficient T cells, especially NKT cells, caspase levels are high. These proteins are inhibited by the BIR domain of XIAP and are essential for apoptosis, which may explain the susceptibility of cells with mutant XIAP to activation-induced cell death (AICD). As with SH2D1A, pathogens survive longer in the body due to an impaired immune response. This results in a chronic inflammatory response due to the continuous release of pro-inflammatory cytokines [58].

1.6. Concluding remarks and future perspectives

XLP is a complex disease caused by mutations in the *SH2D1A* and *XIAP* genes with the pathophysiological and molecular mechanisms not entirely understood. Diagnosis of XLP is often delayed or confused with other clinical conditions, highlighting the need to improve diagnostic methods and develop guidelines for patient management. Biomarker identification is emerging as a promising line of research to address these issues [167]. Biomarkers are molecules that can facilitate early diagnosis, assisting the selection of personalized therapies, monitor disease progression, and predict potential outcomes. Investigation in this area can improve the treatment of XLP and the understanding of its pathophysiology.

Currently, treatment of XLP is mainly aimed at controlling symptoms and preventing or treating lymphoma and EBV infection. Bone marrow transplantation is the only curative option, but it is an invasive procedure, especially for children, and the lack of compatible donors poses a challenge. Therefore, it is critical to develop new, more effective, and immediate therapies that specifically target XLP. To do so, it is critical to understand the specific mechanisms by which immune dysfunction occurs in XLP, including the impact of mutations in molecular pathways and variability in clinical severity. Evolutionary analyses could elucidate the function role of mutations in genes and their relation with regions of the protein evolving under purifying (conserved zones) or positive selection (site diversification). Moreover, *in silico* proteomics can contribute to understand the effects of mutations on the tridimensional structure of the proteins. Identifying the protein domain or region affected by the mutations could provide cues to recognize why a particular molecular pathway is suppressed, especially when influencing a region responsible for a protein interaction. Discovery of the suppressed cellular signaling pathways and responses is an important step toward deciphering the molecular mechanisms of this disease and providing specific targets for *in vitro/in vivo* studies to validate the findings of computational analyses. In addition, molecular docking analyses and virtual screening in small molecule databases can identify potential drugs for *in vitro* testing.

References

- [1] B. Vyskot, Y Chromosome Evolution, in: Brenner's Encyclopedia of Genetics, Elsevier, 2013: pp. 372–375. <https://doi.org/10.1016/B978-0-12-374984-0.01659-4>.
- [2] L. Ramos, A. Antunes, Decoding sex: Elucidating sex determination and how high-quality genome assemblies are untangling the evolutionary dynamics of sex chromosomes, *Genomics*. 114 (2022) 110277. <https://doi.org/10.1016/J.YGENO.2022.110277>.
- [3] M. Scharl, A. Herpin, Sex Determination in Vertebrates, *Encyclopedia of Reproduction*. (2018) 159–167. <https://doi.org/10.1016/B978-0-12-809633-8.20551-7>.
- [4] J. Kim, M. Farré, L. Auvil, B. Capitanu, D.M. Larkin, J. Ma, H.A. Lewin, Reconstruction and evolutionary history of eutherian chromosomes, *Proc Natl Acad Sci U S A*. 114 (2017) E5379–E5388. https://doi.org/10.1073/PNAS.1702012114/SUPPL_FILE/PNAS.1702012114.SD04.PDF.
- [5] L. Ross, H. Blackmon, Sex Determination, in: *Encyclopedia of Evolutionary Biology*, Elsevier, 2016: pp. 81–88. <https://doi.org/10.1016/B978-0-12-800049-6.00146-3>.
- [6] C.S. dos Santos, T. Mendes, A. Antunes, The genes from the pseudoautosomal region 1 (PAR1) of the mammalian sex chromosomes: Synteny, phylogeny and selection, *Genomics*. 114 (2022). <https://doi.org/10.1016/J.YGENO.2022.110419>.
- [7] J.A.M. Graves, Evolution of vertebrate sex chromosomes and dosage compensation, *Nature Reviews Genetics* 2015 17:1. 17 (2015) 33–46. <https://doi.org/10.1038/nrg.2015.2>.
- [8] A. Ballabio, D. Nelson, S. Rozen, Genetics of disease. The sex chromosomes and human disease, *Curr Opin Genet Dev*. 16 (2006) 209–212. <https://doi.org/10.1016/J.GDE.2006.04.016>.
- [9] A.P. Arnold, Y chromosome's roles in sex differences in disease, *Proceedings of the National Academy of Sciences*. 114 (2017) 3787–3789. <https://doi.org/10.1073/PNAS.1702161114>.

- [10] B.R. Migeon, Why females are mosaics, x-chromosome inactivation, and sex differences in disease, *Gend Med.* 4 (2007) 97–105. [https://doi.org/10.1016/S1550-8579\(07\)80024-6](https://doi.org/10.1016/S1550-8579(07)80024-6).
- [11] C.M. Valley, H.F. Willard, Genomic and epigenomic approaches to the study of X chromosome inactivation, *Curr Opin Genet Dev.* 16 (2006) 240–245. <https://doi.org/10.1016/J.GDE.2006.04.008>.
- [12] B.R. Migeon, The Role of X Inactivation and Cellular Mosaicism in Women's Health and Sex-Specific Diseases, *JAMA.* 295 (2006) 1428–1433. <https://doi.org/10.1001/JAMA.295.12.1428>.
- [13] A.P. Arnold, K. Reue, M. Eghbali, E. Vilain, X. Chen, N. Ghahramani, Y. Itoh, J. Li, J.C. Link, T. Ngun, S.M. Williams-Burris, The importance of having two X chromosomes, *Philosophical Transactions of the Royal Society B: Biological Sciences.* 371 (2016). <https://doi.org/10.1098/RSTB.2015.0113>.
- [14] H.C. Martin, E.J. Gardner, K.E. Samocha, J. Kaplanis, N. Akawi, A. Sifrim, R.Y. Eberhardt, A.L.T. Tavares, M.D.C. Neville, M.E.K. Niemi, G. Gallone, J. McRae, S. Borrás, C. Clark, J. Dean, Z. Miedzybrodzka, A. Ross, S. Tennant, T. Dabir, D. Donnelly, M. Humphreys, A. Magee, V. McConnell, S. McKee, S. McNerlan, P.J. Morrison, G. Rea, F. Stewart, T. Cole, N. Cooper, L. Cooper-Charles, H. Cox, L. Islam, J. Jarvis, R. Keelagher, D. Lim, D. McMullan, J. Morton, S. Naik, M. O'Driscoll, K.R. Ong, D. Osio, N. Ragge, S. Turton, J. Vogt, D. Williams, S. Bodek, A. Donaldson, A. Hills, K. Low, R. Newbury-Ecob, A.M. Norman, E. Roberts, I. Scurr, S. Smithson, M. Tooley, S. Abbs, R. Armstrong, C. Dunn, S. Holden, S.M. Park, J. Paterson, L. Raymond, E. Reid, R. Sandford, I. Simoncic, M. Tischkowitz, G. Woods, L. Bradley, J. Comerford, A. Green, S. Lynch, S. McQuaid, B. Mullaney, J. Berg, D. Goudie, E. Mavrak, J. McLean, C. McWilliam, E. Reavey, T. Azam, E. Cleary, A. Jackson, W. Lam, A. Lampe, D. Moore, M. Porteous, E. Baple, J. Baptista, C. Brewer, B. Castle, E. Kivuva, M. Owens, J. Rankin, C. Shaw-Smith, C. Turner, P. Turnpenny, C. Tysoe, T. Bradley, R. Davidson, C. Gardiner, S. Joss, E. Kinning, C. Longman, R. McGowan, V. Murday, D. Pilz, E. Tobias, M. Whiteford, N. Williams, A. Barnicoat, E. Clement, F. Faravelli, J. Hurst, L. Jenkins, W. Jones, V.K.A. Kumar, M. Lees, S. Loughlin, A. Male, D. Morrogh, E. Rosser, R. Scott, L. Wilson, A. Beleza, C. Deshpande, F. Flinter, M. Holder, M. Irving, L. Izatt, D. Josifova, S. Mohammed, A. Molenda, L. Robert, W. Roworth, D. Ruddy, M. Ryten, S. Yau, C. Bennett, M. Blyth, J. Campbell, A. Coates, A. Dobbie, S. Hewitt, E. Hobson, E. Jackson, R. Jewell, A. Kraus, K. Prescott, E. Sheridan, J. Thomson,

- K. Bradshaw, A. Dixit, J. Eason, R. Haines, R. Harrison, S. Mutch, A. Sarkar, C. Searle, N. Shannon, A. Sharif, M. Suri, P. Vasudevan, N. Canham, I. Ellis, L. Greenhalgh, E. Howard, V. Stinton, A. Swale, A. Weber, S. Banka, C. Breen, T. Briggs, E. Burkitt-Wright, K. Chandler, J. Clayton-Smith, D. Donnai, S. Douzgou, L. Gaunt, E. Jones, B. Kerr, C. Langley, K. Metcalfe, A. Smith, R. Wright, D. Bourn, J. Burn, R. Fisher, S. Hellens, A. Henderson, T. Montgomery, M. Splitt, V. Straub, M. Wright, S. Zwolinski, Z. Allen, B. Bernhard, A. Brady, C. Brooks, L. Busby, V. Clowes, N. Ghali, S. Holder, R. Ibitoye, E. Wakeling, E. Blair, J. Carmichael, D. Cilliers, S. Clasper, R. Gibbons, U. Kini, T. Lester, A. Nemeth, J. Poulton, S. Price, D. Shears, H. Stewart, A. Wilkie, S. Albaba, D. Baker, M. Balasubramanian, D. Johnson, M. Parker, O. Quarrell, A. Stewart, J. Willoughby, C. Crosby, F. Elmslie, T. Homfray, H. Jin, N. Lahiri, S. Mansour, K. Marks, M. McEntagart, A. Saggart, K. Tatton-Brown, R. Butler, A. Clarke, S. Corrin, A. Fry, A. Kamath, E. McCann, H. Mugalaasi, C. Pottinger, A. Procter, J. Sampson, F. Sansbury, V. Varghese, D. Baralle, A. Callaway, E.J. Cassidy, S. Daniels, A. Douglas, N. Foulds, D. Hunt, M. Kharbanda, K. Lachlan, C. Mercer, L. Side, I.K. Temple, D. Wellesley, C.F. Wright, D.R. FitzPatrick, H. v. Firth, M.E. Hurles, The contribution of X-linked coding variation to severe developmental disorders, *Nature Communications* 2021 12:1. 12 (2021) 1–13. <https://doi.org/10.1038/s41467-020-20852-3>.
- [15] W.B. Dobyns, A. Filauro, B.N. Tomson, A.S. Chan, A.W. Ho, N.T. Ting, J.C. Oosterwijk, C. Ober, Inheritance of most X-linked traits is not dominant or recessive, just X-linked, *Am J Med Genet A*. 129A (2004) 136–143. <https://doi.org/10.1002/AJMG.A.30123>.
- [16] W.S. Liu, Mammalian Sex Chromosome Structure, Gene Content, and Function in Male Fertility, <https://doi.org/10.1146/Annurev-Animal-020518-115332>. 7 (2019) 103–124. <https://doi.org/10.1146/ANNUREV-ANIMAL-020518-115332>.
- [17] S.C. Credendino, C. Neumayer, I. Cantone, Genetics and Epigenetics of Sex Bias: Insights from Human Cancer and Autoimmunity, *Trends in Genetics*. 36 (2020) 650–663. <https://doi.org/10.1016/J.TIG.2020.06.016>.
- [18] A.M. Tukker, C.D. Royal, A.B. Bowman, K.A. McAllister, The Impact of Environmental Factors on Monogenic Mendelian Diseases, *Toxicol Sci*. 181 (2021) 3–12. <https://doi.org/10.1093/TOXSCI/KFAB022>.

- [19] R. Kingdom, C.F. Wright, Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts, *Front Genet.* 13 (2022). <https://doi.org/10.3389/FGENE.2022.920390>.
- [20] C.J. Brown, Role of the X Chromosome in Cancer, *J Natl Cancer Inst.* 88 (1996) 480–482. <https://academic.oup.com/jnci/article/88/8/480/914974> (accessed November 23, 2023).
- [21] A. Spatz, C. Borg, J. Feunteun, X-chromosome genetics and human cancer, *Nat Rev Cancer.* 4 (2004) 617–629. <https://doi.org/10.1038/NRC1413>.
- [22] C. Achilla, T. Papavramidis, L. Angelis, A. Chatzikyriakidou, The Implication of X-Linked Genetic Polymorphisms in Susceptibility and Sexual Dimorphism of Cancer, *Anticancer Res.* 42 (2022) 2261–2276. <https://doi.org/10.21873/ANTICANRES.15706>.
- [23] F. Caramia, T.P. Speed, H. Shen, Y. Haupt, S. Haupt, Establishing the Link between X-Chromosome Aberrations and TP53 Status, with Breast Cancer Patient Outcomes, *Cells.* 12 (2023) 2245. <https://doi.org/10.3390/CELLS12182245/S1>.
- [24] S.J. Winham, N.B. Larson, S.M. Armasu, Z.C. Fogarty, M.C. Larson, B.M. McCauley, C. Wang, K. Lawrenson, S. Gayther, J.M. Cunningham, B.L. Fridley, E.L. Goode, Molecular signatures of X chromosome inactivation and associations with clinical outcomes in epithelial ovarian cancer, *Hum Mol Genet.* 28 (2019) 1331. <https://doi.org/10.1093/HMG/DDY444>.
- [25] D. Wang, L. Tang, Y. Wu, C. Fan, S. Zhang, B. Xiang, M. Zhou, X. Li, Y. Li, G. Li, W. Xiong, Z. Zeng, C. Guo, Abnormal X chromosome inactivation and tumor development, *Cellular and Molecular Life Sciences.* 77 (2020) 2949–2958. <https://doi.org/10.1007/S00018-020-03469-Z/FIGURES/2>.
- [26] T. Kawakami, C. Zhang, T. Taniguchi, J.K. Chul, Y. Okada, H. Sugihara, T. Hattori, A.E. Reeve, O. Ogawa, K. Okamoto, Characterization of loss-of-inactive X in Klinefelter syndrome and female-derived cancer cells, *Oncogene.* 23 (2004) 6163–6169. <https://doi.org/10.1038/SJ.ONC.1207808>.
- [27] A. Giammarini, M. Rocchi, W. Zennaro, G. Filippi, XX male with breast cancer, *Clin Genet.* 18 (1980) 103–108. <https://doi.org/10.1111/J.1399-0004.1980.TB01019.X>.

- [28] S.M. Sirchia, L. Ramoscelli, F.R. Grati, F. Barbera, D. Coradini, F. Rossella, G. Porta, E. Lesma, A. Ruggeri, P. Radice, G. Simoni, M. Miozzo, Loss of the inactive X chromosome and replication of the active X in BRCA1-defective and wild-type breast cancer cells, *Cancer Res.* 65 (2005) 2139–2146. <https://doi.org/10.1158/0008-5472.CAN-04-3465>.
- [29] S.P. Shriver, H.A. Bourdeau, C.T. Gubish, D.L. Tirpak, A.L.G. Davis, J.D. Luketich, J.M. Siegfried, Sex-specific expression of gastrin-releasing peptide receptor: relationship to smoking history and risk of lung cancer, *J Natl Cancer Inst.* 92 (2000) 24–33. <https://doi.org/10.1093/JNCI/92.1.24>.
- [30] E.A. Zang, E.L. Wynder, Differences in lung cancer risk between men and women: examination of the evidence, *J Natl Cancer Inst.* 88 (1996) 183–192. <https://doi.org/10.1093/JNCI/88.3-4.183>.
- [31] L. May, K. Shows, P. Nana-Sinkam, H. Li, J.W. Landry, Sex Differences in Lung Cancer, *Cancers (Basel)*. 15 (2023). <https://doi.org/10.3390/CANCERS15123111>.
- [32] A. Haugen, Women who smoke: are women more susceptible to tobacco-induced lung cancer?, *Carcinogenesis*. 23 (2002) 227–229. <https://doi.org/10.1093/CARCIN/23.2.227>.
- [33] B.A. Brand, A.E. Blesson, C.L. Smith-Hicks, The Impact of X-Chromosome Inactivation on Phenotypic Expression of X-Linked Neurodevelopmental Disorders, *Brain Sci.* 11 (2021). <https://doi.org/10.3390/BRAINS11070904>.
- [34] M. Foschi, V. Vacchiano, P. Avoni, A. Incensi, S. Battaglia, V. Donadio, E. Panzeri, M.T. Bassi, R. Liguori, G. Rizzo, Broadening the spectrum of adulthood X-linked adrenoleukodystrophy: A report of two atypical cases, *Front Neurol.* 10 (2019) 70. <https://doi.org/10.3389/FNEUR.2019.00070/BIBTEX>.
- [35] C. Brinkmeyer-Langford, J.N. Kornegay, Comparative Genomics of X-linked Muscular Dystrophies: The Golden Retriever Model, *Curr Genomics*. 14 (2013) 330. <https://doi.org/10.2174/13892029113149990004>.
- [36] S. Benchikh, A. Bousfiha, L. Razoki, J. Aboulfaraj, L. Zarouf, C. Elbakay, L.L. Rifai, A. el Hamouchi, S. Nassereddine, Chromosome Abnormalities Related to Reproductive and Sexual Development Disorders: A 5-Year Retrospective Study, *Biomed Res Int.* 2021 (2021). <https://doi.org/10.1155/2021/8893467>.
- [37] P.H.B. Bolton-Maggs, K.J. Pasi, Haemophilias A and B, *Lancet*. 361 (2003) 1801–1809. [https://doi.org/10.1016/S0140-6736\(03\)13405-8](https://doi.org/10.1016/S0140-6736(03)13405-8).

- [38] M.H. Albert, T.C. Bittner, S. Nonoyama, L.D. Notarangelo, S. Burns, K. Imai, T. Espanol, A. Fasth, I. Pellier, G. Strauss, T. Morio, B. Gathmann, J.G. Noordzij, C. Fillat, M. Hoenig, M. Nathrath, A. Meindl, P. Pagel, U. Wintergerst, A. Fischer, A.J. Thrasher, B.H. Belohradsky, H.D. Ochs, X-linked thrombocytopenia (XLT) due to WAS mutations: clinical characteristics, long-term outcome, and treatment options, *Blood*. 115 (2010) 3231–3238. <https://doi.org/10.1182/BLOOD-2009-09-239087>.
- [39] H. Bassiri, W.C. Janice Yeo, J. Rothman, G.A. Koretzky, K.E. Nichols, X-linked lymphoproliferative disease (XLP): A model of impaired anti-viral, anti-tumor and humoral immune responses, *Immunol Res*. 42 (2008) 145–159. <https://doi.org/10.1007/S12026-008-8048-7/FIGURES/3>.
- [40] Z. Alizadeh, P. Dashti, M. Mazinani, M. Nourizadeh, L. Shakerian, S. Tajik, M. Movahedi, S. Mamishi, Z. Pourpak, M.R. Fazlollahi, Clinical and Genetic Study of X-linked Agammaglobulinemia Patients (The Benefit of Early Diagnosis), *Iran J Allergy Asthma Immunol*. 19 (2020) 305–309. <https://doi.org/10.18502/IJAAI.V19I3.3458>.
- [41] R.J. Trent, Mendelian Genetic Traits, *Molecular Medicine*. (2005) 43–II. <https://doi.org/10.1016/B978-012699057-7/50003-5>.
- [42] M. Francesconi, B. Lehner, The effects of genetic variation on gene expression dynamics during development, *Nature*. 505 (2014) 208–211. <https://doi.org/10.1038/NATURE12772>.
- [43] S. Malhotra, A.F. Alsulami, Y. Heiyun, B.M. Ochoa, H. Jubbs, S. Forbes, T.L. Blundell, Understanding the impacts of missense mutations on structures and functions of human cancer-related genes: A preliminary computational analysis of the COSMIC Cancer Gene Census, *PLoS One*. 14 (2019). <https://doi.org/10.1371/JOURNAL.PONE.0219935>.
- [44] I. Bianchi, A. Lleo, M.E. Gershwin, P. Invernizzi, The X chromosome and immune associated genes, *J Autoimmun*. 38 (2012). <https://doi.org/10.1016/J.JAUT.2011.11.012>.
- [45] W.H. Brooks, Y. Renaudineau, Epigenetics and autoimmune diseases: The X chromosome-nucleolus nexus, *Front Genet*. 6 (2015) 22. <https://doi.org/10.3389/FGENE.2015.00022/BIBTEX>.

- [46] C. Libert, L. Dejager, I. Pinheiro, The X chromosome in immune functions: when a chromosome makes the difference, *Nature Reviews Immunology* 2010 10:8. 10 (2010) 594–604. <https://doi.org/10.1038/nri2815>.
- [47] J.S. Amberger, C.A. Bocchini, F. Schiettecatte, A.F. Scott, A. Hamosh, OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders, *Nucleic Acids Res.* 43 (2015) D789–D798. <https://doi.org/10.1093/NAR/GKU1205>.
- [48] N.R. Rose, Autoimmune Diseases, *International Encyclopedia of Public Health*. (2017) 192–195. <https://doi.org/10.1016/B978-0-12-803678-5.00029-1>.
- [49] F. Angum, T. Khan, J. Kaler, L. Siddiqui, A. Hussain, F. Angum, T. Khan, J. Kaler, L. Siddiqui, A. Hussain, The Prevalence of Autoimmune Disorders in Women: A Narrative Review, *Cureus*. 12 (2020). <https://doi.org/10.7759/CUREUS.8094>.
- [50] L. Wang, F.-S. Wang, & M.E. Gershwin, A. Wang, L. Wang, G. Me, Human autoimmune diseases: a comprehensive update, *J Intern Med*. 278 (2015) 369–395. <https://doi.org/10.1111/JOIM.12395>.
- [51] M. Allez, L. Mayer, Immunodeficiency, *Encyclopedia of Gastroenterology*. (2004) 425–429. <https://doi.org/10.1016/B0-12-386860-2/00394-4>.
- [52] A. Bousfiha, A. Moundir, S.G. Tangye, C. Picard, L. Jeddane, W. Al-Herz, C.C. Rundles, J.L. Franco, S.M. Holland, C. Klein, T. Morio, E. Oksenhendler, A. Puel, J. Puck, M.R.J. Seppänen, R. Somech, H.C. Su, K.E. Sullivan, T.R. Torgerson, I. Meyts, The 2022 Update of IUIS Phenotypical Classification for Human Inborn Errors of Immunity, *J Clin Immunol.* 42 (2022) 1508–1520. <https://doi.org/10.1007/S10875-022-01352-Z/FIGURES/10>.
- [53] L. Schweitzer, V. Korsos, D.C. Vinh, Treatment of Secondary Immunodeficiencies, *Encyclopedia of Infection and Immunity*. (2022) 819–842. <https://doi.org/10.1016/B978-0-12-818731-9.00175-0>.
- [54] G. Aytekin, S. Vaqar, X-Linked Immunodeficiency, *StatPearls*. (2022). <https://www.ncbi.nlm.nih.gov/books/NBK562182/> (accessed February 22, 2023).
- [55] I. Meyts, A. Bousfiha, C. Duff, S. Singh, Y.L. Lau, A. Condino-Neto, L. Bezrodnik, A. Ali, M. Adeli, J. Drabwell, Primary Immunodeficiencies: A Decade of Progress and a Promising Future, *Front Immunol.* 11 (2021) 625753. <https://doi.org/10.3389/FIMMU.2020.625753/BIBTEX>.

- [56] S.G. Tangye, W. Al-Herz, A. Bousfiha, T. Chatila, C. Cunningham-Rundles, A. Etzioni, J.L. Franco, S.M. Holland, C. Klein, T. Morio, H.D. Ochs, E. Oksenhendler, C. Picard, J. Puck, T.R. Torgerson, J.L. Casanova, K.E. Sullivan, Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee, *Journal of Clinical Immunology* 2020 40:1. 40 (2020) 24–64. <https://doi.org/10.1007/S10875-019-00737-X>.
- [57] N. Panchal, C. Booth, J.L. Cannons, P.L. Schwartzberg, X-linked lymphoproliferative disease type 1: A clinical and molecular perspective, *Front Immunol.* 9 (2018). <https://doi.org/10.3389/FIMMU.2018.00666>.
- [58] K.E. Nichols, R.A. Marsh, The X-Linked Lymphoproliferative Syndromes Chapter Outline, Stiehm's Immune Deficiencies. (2014) 475–496. <https://doi.org/10.1016/B978-0-12-405546-9.00022-4>.
- [59] T. Xu, Q. Zhao, W. Li, X. Chen, X. Xue, Z. Chen, X. Du, X. Bai, Q. Zhao, L. Zhou, X. Tang, X. Yang, H. Kanegane, X. Zhao, X-linked lymphoproliferative syndrome in mainland China: review of clinical, genetic, and immunological characteristic, *Eur J Pediatr.* 179 (2020) 327. <https://doi.org/10.1007/S00431-019-03512-7>.
- [60] S. Ono, T. Okano, A. Hoshino, M. Yanagimachi, K. Hamamoto, Y. Nakazawa, T. Imamura, M. Onuma, H. Niizuma, Y. Sasahara, H. Tsujimoto, T. Wada, R. Kunisaki, M. Takagi, K. Imai, T. Morio, H. Kanegane, Hematopoietic stem cell transplantation for XIAP deficiency in Japan, *J Clin Immunol.* 37 (2017) 85–91. <https://doi.org/10.1007/s10875-016-0348-4>.
- [61] Y. Wang, Y. Wang, W. Lu, L. Tao, Y. Xiao, Y. Zhou, X. He, Y. Zhang, L. Li, Potential pathogenic mechanism of type 1 X-linked lymphoproliferative syndrome caused by a mutation of SH2D1A gene in an infant: A case report, *Medicine.* 101 (2022) E30951. <https://doi.org/10.1097/MD.00000000000030951>.
- [62] S. Chen, W. Shi, Y. Qian, L. Wang, J. Zhang, S. Li, Y. Chen, C. Chang, H. Fei, L. Zhang, H. Huang, C. Xu, Preimplantation Genetic Testing for a Chinese Family With X-Linked Lymphoproliferative Syndrome Type 1, *Front Genet.* 11 (2020) 550507. <https://doi.org/10.3389/FGENE.2020.550507>.
- [63] J.P. Schmid, D. Canioni, D. Moshous, F. Touzot, N. Mahlaoui, F. Hauck, H. Kanegane, E. Lopez-Granados, E. Mejstrikova, I. Pellier, L. Galicier, C. Galambrun, V. Barlogis, P. Bordigoni, A. Fourmaintraux, M. Hamidou, A. Dabadie, F. Le Deist, F. Haerynck, M. Ouachée-Chardin, P. Rohrlich, J.L. Stephan, C. Lenoir, S. Rigaud, N. Lambert, M. Milili, C. Schiff, H. Chapel, C. Picard, G. De

- Saint Basile, S. Blanche, A. Fischer, S. Latour, Clinical similarities and differences of patients with X-linked lymphoproliferative syndrome type 1 (XLP-1/SAP deficiency) versus type 2 (XLP-2/XIAP deficiency), *Blood*. 117 (2011) 1522–1529. <https://doi.org/10.1182/BLOOD-2010-07-298372>.
- [64] X. Yang, A. Hoshino, T. Taga, T. Kunitsu, Y. Ikeda, T. Yasumi, K. Yoshida, T. Wada, K. Miyake, T. Kubota, Y. Okuno, H. Muramatsu, Y. Adachi, S. Miyano, S. Ogawa, S. Kojima, H. Kanegane, A female patient with incomplete hemophagocytic lymphohistiocytosis caused by a heterozygous XIAP mutation associated with non-random X-chromosome inactivation skewed towards the wild-type XIAP allele, *J Clin Immunol*. 35 (2015) 244–248. <https://doi.org/10.1007/s10875-015-0144-6>.
- [65] B.R. Migeon, X-linked diseases: susceptible females, *Genetics in Medicine* 2020 22:7. 22 (2020) 1156–1174. <https://doi.org/10.1038/s41436-020-0779-4>.
- [66] Z. Sbihi, K. Tanita, C. Bachelet, C. Bole, F. Jabot-Hanin, F. Tores, M. le Loch, R. Khodr, A. Hoshino, C. Lenoir, M. Oleastro, M. Villa, L. Sposito, E. Prieto, S. Danielian, E. Brunet, C. Picard, T. Taga, S.S.M.A. Abdrabou, T. Isoda, M. Yamada, A. Palma, H. Kanegane, S. Latour, Identification of Germline Non-coding Deletions in XIAP Gene Causing XIAP Deficiency Reveals a Key Promoter Sequence, *J Clin Immunol*. 42 (2022) 559–571. <https://doi.org/10.1007/S10875-021-01188-Z/FIGURES/5>.
- [67] K. Zhang, E. Wakefield, R. Marsh, Lymphoproliferative Disease, X-Linked, *GeneReviews®*. (2016) 1–7. <https://www.ncbi.nlm.nih.gov/books/NBK1406/> (accessed April 7, 2023).
- [68] A. Majeranowski, Z. Biologii Komórki I Immunologii, W. Biotechnologii, N.A. Dułak, J. Juwa, R. Trzciński, N. Irga-Jaworska, Hematological manifestations of X-linked lymphoproliferative disease, *Hematology in Clinical Practice*. 13 (2022) 1–6. <https://doi.org/10.5603/HCP.A2022.0002>.
- [69] C. Booth, K.C. Gilmour, P. Veys, A.R. Gennery, M.A. Slatter, H. Chapel, P.T. Heath, C.G. Steward, O. Smith, A. O'Meara, H. Kerrigan, N. Mahlaoui, M. Cavazzana-Calvo, A. Fischer, D. Moshous, S. Blanche, J. Pachlopnick-Schmid, S. Latour, G. De Saint-Basile, M. Albert, G. Notheis, N. Rieber, B. Strahm, H. Ritterbusch, A. Lankester, N.G. Hartwig, I. Meyts, A. Plebani, A. Soresina, A. Finocchi, C. Pignata, E. Cirillo, S. Bonanomi, C. Peters, K. Kalwak, S. Pasic, P. Sedlacek, J. Jazbec, H. Kanegane, K.E. Nichols, I.C. Hanson, N. Kapoor, E. Haddad, M. Cowan, S. Choo,

- J. Smart, P.D. Arkwright, H.B. Gaspar, X-linked lymphoproliferative disease due to SAP/SH2D1A deficiency: a multicenter study on the manifestations, management and outcome of the disease, *Blood*. 117 (2011) 53–62. <https://doi.org/10.1182/BLOOD-2010-06-284935>.
- [70] W.K. Kwon, J.A. Kim, J.H. Park, D.R. Kim, S.E. Park, Y.J. Kim, K.H. Yoo, J.H. Jang, E.S. Kang, Case Report: Novel Splicing Variant in SH2D1A in a Patient With X-Linked Lymphoproliferative Syndrome Type 1, *Front Pediatr*. 10 (2022) 812590. <https://doi.org/10.3389/FPED.2022.812590>.
- [71] C.L. Ku, I.T. Chen, M.Z. Lai, Infection-induced inflammation from specific inborn errors of immunity to COVID-19, *FEBS J*. 288 (2021) 5021. <https://doi.org/10.1111/FEBS.15961>.
- [72] H. Pervez, A.T.U. Din, A. Khan, A Mysterious Case of an Infarcted Spleen due to Kissing Disease: A Rare Entity, *Cureus*. 12 (2020). <https://doi.org/10.7759/CUREUS.6700>.
- [73] L.J. Carbarry, Infectious mononucleosis: the kissing disease, *Nurs Care*. 8 (1975) 22–24. <https://pubmed.ncbi.nlm.nih.gov/163472/> (accessed May 2, 2023).
- [74] J. Schwartzkopf, Infectious mononucleosis, *JAAPA*. 31 (2018) 52–53. <https://doi.org/10.1097/01.JAA.0000546488.73851.DD>.
- [75] M. Mohseni, M.P. Boniface, C. Graham, Mononucleosis, *Prim Care Update Ob Gyns*. 2 (2023) 152–156. [https://doi.org/10.1016/1068-607X\(95\)00034-5](https://doi.org/10.1016/1068-607X(95)00034-5).
- [76] J. Womack, M. Jimenez, Common Questions About Infectious Mononucleosis, *Am Fam Physician*. 91 (2015) 372–376. <https://www.aafp.org/pubs/afp/issues/2015/0315/p372.html> (accessed May 2, 2023).
- [77] J.E. Sylvester, B.K. Buchanan, T.W. Silva, Infectious Mononucleosis: Rapid Evidence Review, *Am Fam Physician*. 107 (2023) 71–78. <https://www.aafp.org/pubs/afp/issues/2023/0100/infectious-mononucleosis.html> (accessed May 2, 2023).
- [78] V. Kapila, C.J. Wehrle, F. Tuma, Physiology, Spleen, *StatPearls*. (2022). <https://www.ncbi.nlm.nih.gov/books/NBK537307/> (accessed May 2, 2023).
- [79] M.R. Hines, T.E. Knight, K.O. McNerney, M.B. Leick, T. Jain, S. Ahmed, M.J. Frigault, J.A. Hill, M.D. Jain, W.T. Johnson, Y. Lin, K.M. Mahadeo, G.M. Maron,

- R.A. Marsh, S.S. Neelapu, S. Nikiforow, A.K. Ombrello, N.N. Shah, A.C. Talleur, D. Turicek, A. Vatsayan, S.W. Wong, M. V. Maus, K. V. Komanduri, N. Berliner, J.-I. Henter, M.-A. Perales, N. V. Frey, D.T. Teachey, M.J. Frank, N.N. Shah, Immune Effector Cell-Associated Hemophagocytic Lymphohistiocytosis-Like Syndrome, *Transplant Cell Ther.* (2023). <https://doi.org/10.1016/J.JTCT.2023.03.006>.
- [80] L. Diaz, B. Jauzelon, A.C. Dillies, C. Le Souder, J.L. Faillie, A.T.J. Maria, P. Palassin, Hemophagocytic Lymphohistiocytosis Associated with Immunological Checkpoint Inhibitors: A Pharmacovigilance Study, *J Clin Med.* 12 (2023). <https://doi.org/10.3390/JCM12051985/S1>.
- [81] G.E. Janka, Familial and acquired hemophagocytic lymphohistiocytosis, *Eur J Pediatr.* 166 (2007) 95–109. <https://doi.org/10.1007/S00431-006-0258-1/TABLES/3>.
- [82] J. Lee, B. Pham, Z.E. Karanjawala, O. Adesina, Postpartum fevers, a rare presentation of secondary hemophagocytic lymphohistiocytosis, *Clin Case Rep.* 11 (2023). <https://doi.org/10.1002/CCR3.7070>.
- [83] K. Zhang, I. Astigarraga, Y. Bryceson, K. Lehmberg, R. Machowicz, R. Marsh, E. Sieni, Z. Wang, K.E. Nichols, Familial Hemophagocytic Lymphohistiocytosis, *GeneReviews®*. (2021). <https://www.ncbi.nlm.nih.gov/books/NBK1444/> (accessed May 2, 2023).
- [84] H. Trottestam, Clinical Studies of Hemophagocytic Lymphohistiocytosis, *PQDT - Global.* 85 (2011). <https://www.proquest.com/dissertations-theses/clinical-studies-hemophagocytic/docview/2565999517/se-2?accountid=26646>
http://link.periodicos.capes.gov.br/sfxclcl41?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=unknown&sid=ProQ:Pro (accessed May 2, 2023).
- [85] J.F. Viallard, Conduite à tenir devant une hypogammaglobulinémie, *Rev Med Interne.* 44 (2023) 133–138. <https://doi.org/10.1016/J.REVMED.2023.01.010>.
- [86] I.M. Otani, H.K. Lehman, A.M. Jongco, L.R. Tsao, A.E. Azar, T.K. Tarrant, E. Engel, J.E. Walter, T.Q. Truong, D.A. Khan, M. Ballow, C. Cunningham-Rundles, H. Lu, M. Kwan, S. Barmettler, Practical guidance for the diagnosis and management of secondary hypogammaglobulinemia: A Work Group Report of the AAAAI Primary Immunodeficiency and Altered Immune Response Committees, *Journal of Allergy and Clinical Immunology.* 149 (2022) 1525–1560. <https://doi.org/10.1016/J.JACI.2022.01.025>.

- [87] M. Huq, N.K. Bhatnagar, R.W. Hostoffer, Hypogammaglobulinemia, StatPearls. (2022). <https://www.ncbi.nlm.nih.gov/books/NBK563134/> (accessed May 2, 2023).
- [88] M. Jiang, N.N. Bennani, A.L. Feldman, Lymphoma classification update: T-cell lymphomas, Hodgkin lymphomas, and histiocytic/dendritic cell neoplasms, *Expert Rev Hematol.* 10 (2017) 239. <https://doi.org/10.1080/17474086.2017.1281122>.
- [89] R. Alaggio, C. Amador, I. Anagnostopoulos, A.D. Attygalle, I.B. de O. Araujo, E. Berti, G. Bhagat, A.M. Borges, D. Boyer, M. Calaminici, A. Chadburn, J.K.C. Chan, W. Cheuk, W.J. Chng, J.K. Choi, S.S. Chuang, S.E. Coupland, M. Czader, S.S. Dave, D. de Jong, M.Q. Du, K.S. Elenitoba-Johnson, J. Ferry, J. Geyer, D. Gratzinger, J. Guitart, S. Gujral, M. Harris, C.J. Harrison, S. Hartmann, A. Hochhaus, P.M. Jansen, K. Karube, W. Kempf, J. Khoury, H. Kimura, W. Klapper, A.E. Kovach, S. Kumar, A.J. Lazar, S. Lazzi, L. Leoncini, N. Leung, V. Leventaki, X.Q. Li, M.S. Lim, W.P. Liu, A. Louissaint, A. Marcogliese, L.J. Medeiros, M. Michal, R.N. Miranda, C. Mitteldorf, S. Montes-Moreno, W. Morice, V. Nardi, K.N. Naresh, Y. Natkunam, S.B. Ng, I. Oschlies, G. Ott, M. Parrens, M. Pulitzer, S.V. Rajkumar, A.C. Rawstron, K. Rech, A. Rosenwald, J. Said, C. Sarkozy, S. Sayed, C. Saygin, A. Schuh, W. Sewell, R. Siebert, A.R. Sohani, R. Tooze, A. Traverse-Glehen, F. Vega, B. Vergier, A.D. Wechalekar, B. Wood, L. Xerri, W. Xiao, The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms, *Leukemia.* 36 (2022) 1720. <https://doi.org/10.1038/S41375-022-01620-2>.
- [90] I.N. Sheikh, A. Elgehiny, D. Ragoonanan, K.M. Mahadeo, Y. Nieto, S. Khazal, Management of Aggressive Non-Hodgkin Lymphomas in the Pediatric, Adolescent, and Young Adult Population: An Adult vs. Pediatric Perspective, *Cancers (Basel).* 14 (2022). <https://doi.org/10.3390/CANCERS14122912>.
- [91] C.B. Geier, J.R. Farmer, Z. Foldvari, B. Ujhazi, J. Steininger, J.W. Sleasman, S. Parikh, M.A. Dilley, S.Y. Pai, L. Henderson, M. Hazen, B. Neven, D. Moshous, S.O. Sharapova, S. Mihailova, P. Yankova, E. Naumova, S. Özen, K. Byram, J. Fernandez, H.M. Wolf, M.M. Eibl, L.D. Notarangelo, L.H. Calabrese, J.E. Walter, Vasculitis as a Major Morbidity Factor in Patients With Partial RAG Deficiency, *Front Immunol.* 11 (2020) 20. <https://doi.org/10.3389/FIMMU.2020.574738>.

- [92] A. Pandey, C. Shen, S.M. Man, Focus: Organelles: Inflammasomes in Colitis and Colorectal Cancer: Mechanism of Action and Therapies, *Yale J Biol Med.* 92 (2019) 481. /pmc/articles/PMC6747943/ (accessed May 3, 2023).
- [93] O.H. Nielsen, E.C. LaCasse, How genetic testing can lead to targeted management of XIAP deficiency–related inflammatory bowel disease, *Genetics in Medicine.* 19 (2017) 133–143. <https://doi.org/10.1038/GIM.2016.82>.
- [94] S.A. Peslak, T. Olson, D. V. Babushok, Diagnosis and Treatment of Aplastic Anemia, *Curr Treat Options Oncol.* 18 (2017) 70. <https://doi.org/10.1007/S11864-017-0511-Z>.
- [95] D.L. Nelson, C. Terhorst, X–linked lymphoproliferative syndrome, *Clin Exp Immunol.* 122 (2000) 291. <https://doi.org/10.1046/J.1365-2249.2000.01400.X>.
- [96] R.A. Marsh, K. Rao, P. Satwani, K. Lehmberg, I. Müller, D. Li, M.O. Kim, A. Fischer, S. Latour, P. Sedlacek, V. Barlogis, K. Hamamoto, H. Kanegane, S. Milanovich, D.A. Margolis, D. Dimmock, J. Casper, D.N. Douglas, P.J. Amrolia, P. Veys, A.R. Kumar, M.B. Jordan, J.J. Bleesing, A.H. Filipovich, Allogeneic hematopoietic cell transplantation for XIAP deficiency: an international survey reveals poor outcomes, *Blood.* 121 (2013) 877–883. <https://doi.org/10.1182/BLOOD-2012-06-432500>.
- [97] K.G. Ross, L. Veltri, A.S. Kanate, Umbilical Cord Blood Transplantation, Hematopoietic Cell Transplantation for Malignant Conditions. (2019) 125–137. <https://doi.org/10.1016/B978-0-323-56802-9.00008-0>.
- [98] C.C. Pinnix, The Role of Radiation Therapy in Hematopoietic Stem Cell Transplantation, Hematopoietic Cell Transplantation for Malignant Conditions. (2019) 59–72. <https://doi.org/10.1016/B978-0-323-56802-9.00004-3>.
- [99] J. Yu, Y. Du, S. Ahmad, R.D. Patel, J.C. Varela, C.C. Chang, S. Mori, Comparison of Myeloablative versus Reduced-Intensity Conditioning Regimens in Allogeneic Stem Cell Transplantation Recipients with Acute Myelogenous Leukemia with Measurable Residual Disease-Negative Disease at the Time of Transplantation: A Retrospective Cohort Study, *Transplant Cell Ther.* 27 (2021) 663.e1-663.e6. <https://doi.org/10.1016/J.JTCT.2021.04.017>.
- [100] A.H. Filipovich, Gene therapy targets XLP, *Blood.* 121 (2013) 1066–1067. <https://doi.org/10.1182/BLOOD-2012-12-472670>.

- [101] C. Rivat, C. Booth, M. Alonso-Ferrero, M. Blundell, N.J. Sebire, A.J. Thrasher, H.B. Gaspar, SAP gene transfer restores cellular and humoral immune function in a murine model of X-linked lymphoproliferative disease, *Blood*. 121 (2013) 1073–1076. <https://doi.org/10.1182/BLOOD-2012-07-445858>.
- [102] M.J. Riese, E.K. Moon, B.D. Johnson, S.M. Albelda, Diacylglycerol kinases (DGKs): Novel targets for improving T cell activity in cancer, *Front Cell Dev Biol*. 4 (2016) 108. <https://doi.org/10.3389/FCELL.2016.00108/BIBTEX>.
- [103] S. Velnati, E. Ruffo, A. Massarotti, M. Talmon, K.S.S. Varma, A. Gesu, L.G. Fresu, A.L. Snow, A. Bertoni, D. Capello, G.C. Tron, A. Graziani, G. Baldanzi, Identification of a novel DGK α inhibitor for XLP-1 therapy by virtual screening, *Eur J Med Chem*. 164 (2019) 378–390. <https://doi.org/10.1016/J.EJMECH.2018.12.061>.
- [104] J. Luczkowiak, M. Álvarez, A. Sebastián-Martín, L. Menéndez-Arias, DNA-Dependent DNA Polymerases as Drug Targets in Herpesviruses and Poxviruses, *Viral Polymerases: Structures, Functions and Roles as Antiviral Drug Targets*. (2019) 95–134. <https://doi.org/10.1016/B978-0-12-815422-9.00004-8>.
- [105] A.M.Q. King, M.J. Adams, E.B. Carstens, E.J. (Ed.) Lefkowitz, Virus taxonomy: classification and nomenclature of viruses. Ninth report of the International Committee on Taxonomy of Viruses, in: Elsevier Academic Press, Amsterdam, 2012. <https://www.marinespecies.org/aphia.php?p=sourcedetails&id=163053>.
- [106] J. Louten, Herpesviruses, *Essential Human Virology*. (2023) 255–276. <https://doi.org/10.1016/B978-0-323-90565-7.00013-7>.
- [107] Herpesviridae, *Virus Taxonomy*. (2012) 111–122. <https://doi.org/10.1016/B978-0-12-384684-6.00007-0>.
- [108] S. Payne, Family Herpesviridae, *Viruses*. (2017) 269–278. <https://doi.org/10.1016/B978-0-12-803109-4.00034-9>.
- [109] J.I. Cohen, Herpesvirus latency, *J Clin Invest*. 130 (2020) 3361–3369. <https://doi.org/10.1172/JCI136225>.
- [110] J. Piret, G. Boivin, DNA polymerases of herpesviruses and their inhibitors, *Enzymes (Essen)*. 50 (2021) 79–132. <https://doi.org/10.1016/BS.ENZ.2021.07.003>.

- [111] J. Chodosh, L. Ung, Adoption of Innovation in Herpes Simplex Virus Keratitis, Cornea. 39 (2020) S7–S18. <https://doi.org/10.1097/ICO.0000000000002425>.
- [112] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, Epstein-Barr Virus, in: Monographs on the Evaluation of Carcinogenic Risks to Humans, International Agency for Research on Cancer, Lyon, France, 2012.
- [113] P.J. Farrell, Epstein–Barr Virus and Cancer, <https://doi.org/10.1146/Annurev-Pathmechdis-012418-013023>. 14 (2019) 29–53. <https://doi.org/10.1146/ANNUREV-PATHMECHDIS-012418-013023>.
- [114] L.M. Hutt-Fletcher, EBV glycoproteins: where are we now?, Future Virol. 10 (2015) 1155. <https://doi.org/10.2217/FVL.15.80>.
- [115] Z. Li, X. Zhang, L. Dong, J. Pang, M. Xu, Q. Zhong, M.S. Zeng, X. Yu, CryoEM structure of the tegumented capsid of Epstein-Barr virus, Cell Res. 30 (2020) 873. <https://doi.org/10.1038/S41422-020-0363-0>.
- [116] G.F. Rohrmann, BACULOVIRUSES (BACULOVIRIDAE) | NUCLEOPOLYHEDROVIRUS, Encyclopedia of Virology. (1999) 146–152. <https://doi.org/10.1006/RWVI.1999.0031>.
- [117] C. Machón, M. Fàbrega-Ferrer, D. Zhou, A. Cuervo, J.L. Carrascosa, D.I. Stuart, M. Coll, Atomic structure of the Epstein-Barr virus portal, Nat Commun. 10 (2019). <https://doi.org/10.1038/S41467-019-11706-8>.
- [118] T. Murata, Tegument proteins of Epstein-Barr virus: Diverse functions, complex networks, and oncogenesis, Tumour Virus Res. 15 (2023) 200260. <https://doi.org/10.1016/J.TVR.2023.200260>.
- [119] S. Vranic, F.S. Cyprian, S. Akhtar, A.E. Al Moustafa, The role of Epstein-Barr virus in cervical cancer: A brief update, Front Oncol. 8 (2018) 113. <https://doi.org/10.3389/FONC.2018.00113/BIBTEX>.
- [120] M.M.B. abdullah, R.D. Palermo, A.L. Palser, N.E. Grayson, P. Kellam, S. Correia, A. Szymula, R.E. White, Heterogeneity of the Epstein-Barr Virus (EBV) Major Internal Repeat Reveals Evolutionary Mechanisms of EBV and a Functional Defect in the Prototype EBV Strain B95-8, J Virol. 91 (2017) 920–937. <https://doi.org/10.1128/JVI.00920-17>.

- [121] X. Sagaert, B. Sprangers, C. De Wolf-Peeters, The dynamics of the B follicle: Understanding the normal counterpart of B-cell-derived malignancies, *Leukemia*. 21 (2007) 1378–1386. <https://doi.org/10.1038/SJ.LEU.2404737>.
- [122] S. Rodriguez, M. Roussel, K. Tarte, P. Amé-Thomas, Impact of chronic viral infection on T-cell dependent humoral immune response, *Front Immunol*. 8 (2017) 1434. <https://doi.org/10.3389/FIMMU.2017.01434/BIBTEX>.
- [123] L. Tonoyan, S. Vincent-Bugnas, C.-V. Olivieri, A. Doglio, New Viral Facets in Oral Diseases: The EBV Paradox, *Int J Mol Sci*. (2019). <https://doi.org/10.3390/ijms20235861>.
- [124] M. Akkaya, K. Kwak, S.K. Pierce, B cell memory: building two walls of protection against pathogens, *Nature Reviews Immunology* 2019 20:4. 20 (2019) 229–238. <https://doi.org/10.1038/s41577-019-0244-2>.
- [125] P. Liew, The Longevity of the Humoral Immune Response: Survival of Long-lived Plasma Cells, *Akademeia*. 2 (2012) ea0116.
- [126] T. Kurosaki, K. Kometani, W. Ise, Memory B cells, *Nat Rev Immunol*. 15 (2015) 149–159. <https://doi.org/10.1038/NRI3802>.
- [127] H.B. Shah, K. Smith, J.D. Wren, C.F. Webb, J.D. Ballard, R.L. Bourn, J.A. James, M.L. Lang, Insights from analysis of human antigen-specific memory B cell repertoires, *Front Immunol*. 10 (2019) 3064. <https://doi.org/10.3389/FIMMU.2018.03064/BIBTEX>.
- [128] C.G. Vinuesa, Í. Sanz, M.C. Cook, Dysregulation of germinal centres in autoimmune disease, *Nat Rev Immunol*. 9 (2009) 845–857. <https://doi.org/10.1038/NRI2637>.
- [129] V. Cerundolo, J.D. Silk, S.H. Masri, M. Salio, Harnessing invariant NKT cells in vaccination strategies, *Nature Reviews Immunology* 2008 9:1. 9 (2009) 28–38. <https://doi.org/10.1038/nri2451>.
- [130] E.E.S. Brettschneider, M. Terabe, The Role of NKT Cells in Glioblastoma, *Cells* 2021, Vol. 10, Page 1641. 10 (2021) 1641. <https://doi.org/10.3390/CELLS10071641>.
- [131] G. Varricchi, J. Harker, F. Borriello, G. Marone, S.R. Durham, M.H. Shamji, T follicular helper (Tfh) cells in normal immune responses and in allergic disorders, *Allergy*. 71 (2016) 1086–1094. <https://doi.org/10.1111/ALL.12878>.

- [132] Y.T. Png, A.Z.Y. Yang, M.Y. Lee, M.J.M. Chua, C.M. Lim, The Role of NK Cells in EBV Infection and EBV-Associated NPC, *Viruses*. 13 (2021). <https://doi.org/10.3390/V13020300>.
- [133] E. Vivier, E. Tomasello, M. Baratin, T. Walzer, S. Ugolini, Functions of natural killer cells, *Nature Immunology* 2008 9:5. 9 (2008) 503–510. <https://doi.org/10.1038/ni1582>.
- [134] M.G. Desimio, D.A. Covino, B. Rivalta, C. Cancrini, M. Doria, The Role of NK Cells in EBV Infection and Related Diseases: Current Understanding and Hints for Novel Therapies, *Cancers* 2023, Vol. 15, Page 1914. 15 (2023) 1914. <https://doi.org/10.3390/CANCERS15061914>.
- [135] S.I. Fujii, S. Yamasaki, Y. Sato, K. Shimizu, Vaccine designs utilizing invariant NKT-licensed antigen-presenting cells provide NKT or T cell help for B cell responses, *Front Immunol.* 9 (2018) 1267. <https://doi.org/10.3389/FIMMU.2018.01267/BIBTEX>.
- [136] E.E. Vomhof-DeKrey, J. Yates, E.A. Leadbetter, Invariant NKT cells provide innate and adaptive help for B cells, *Curr Opin Immunol.* 28 (2014) 12–17. <https://doi.org/10.1016/J.COI.2014.01.007>.
- [137] N.W. Zwirner, A. Ziblat, Regulation of NK cell activation and effector functions by the IL-12 family of cytokines: The case of IL-27, *Front Immunol.* 8 (2017) 25. <https://doi.org/10.3389/FIMMU.2017.00025/BIBTEX>.
- [138] R.A. Marsh, J. Villanueva, M.O. Kim, K. Zhang, D. Marmer, K.A. Risma, M.B. Jordan, J.J. Bleesing, A.H. Filipovich, Patients with X-linked lymphoproliferative disease due to BIRC4 mutation have normal invariant natural killer T-cell populations, *Clin Immunol.* 132 (2009) 116–123. <https://doi.org/10.1016/J.CLIM.2009.03.517>.
- [139] F. Cunningham, J.E. Allen, J. Allen, J. Alvarez-Jarreta, M.R. Amode, I.M. Armean, O. Austine-Orimoloye, A.G. Azov, I. Barnes, R. Bennett, A. Berry, J. Bhai, A. Bignell, K. Billis, S. Boddu, L. Brooks, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C.G. Giron, T. Genez, J.G. Martinez, C. Guijarro-Clarke, A. Gymer, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J.C. Marugán, S. Mohanan, A. Mushtaq, M. Naven, D.N. Ogeh, A. Parker, A. Parton, M. Perry, I. Pilizota, I. Prosovetskaia, M.P. Sakthivel, A.I.A. Salam, B.M. Schmitt, H. Schuilenburg, D. Sheppard, J.G. Perez-Silva, W. Stark,

- E. Steed, K. Sutinen, R. Sukumaran, D. Sumathipala, M.M. Suner, M. Szpak, A. Thormann, F.F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T.A. Walsh, B. Walts, N. Willhoft, A. Winterbottom, E. Wass, M. Chakiachvili, B. Flint, A. Frankish, S. Giorgetti, L. Haggerty, S.E. Hunt, G.R. Ilesley, J.E. Loveland, F.J. Martin, B. Moore, J.M. Mudge, M. Muffato, E. Perry, M. Ruffier, J. Tate, D. Thybert, S.J. Trevanion, S. Dyer, P.W. Harrison, K.L. Howe, A.D. Yates, D.R. Zerbino, P. Flicek, *Ensembl* 2022, *Nucleic Acids Res.* 50 (2022) D988–D995. <https://doi.org/10.1093/NAR/GKAB1049>.
- [140] J.L. Cannons, S.G. Tangye, P.L. Schwartzberg, SLAM Family Receptors and SAP Adaptors in Immunity, <https://doi.org/10.1146/Annurev-Immunol-030409-101302>. 29 (2011) 665–705. <https://doi.org/10.1146/ANNUREV-IMMUNOL-030409-101302>.
- [141] T. Schmidt, P. Samaras, M. Frejno, S. Gessulat, M. Barnert, H. Kienegger, H. Krcmar, J. Schlegl, H.C. Ehrlich, S. Aiche, B. Kuster, M. Wilhelm, *ProteomicsDB*, *Nucleic Acids Res.* 46 (2018) D1271–D1281. <https://doi.org/10.1093/NAR/GKX1029>.
- [142] A.C.A. Mudde, C. Booth, R.A. Marsh, Evolution of Our Understanding of XIAP Deficiency, *Front Pediatr.* 9 (2021) 557. <https://doi.org/10.3389/FPED.2021.660520/BIBTEX>.
- [143] M. Holcik, H. Gibson, R.G. Korneluk, XIAP: apoptotic brake and promising therapeutic target, *Apoptosis.* 6 (2001) 253–261. <https://doi.org/10.1023/A:1011379307472>.
- [144] P. Samaras, T. Schmidt, M. Frejno, S. Gessulat, M. Reinecke, A. Jarzab, J. Zecha, J. Mergner, P. Giansanti, H.C. Ehrlich, S. Aiche, J. Rank, H. Kienegger, H. Krcmar, B. Kuster, M. Wilhelm, *ProteomicsDB: a multi-omics and multi-organism resource for life science research*, *Nucleic Acids Res.* 48 (2020) D1153–D1163. <https://doi.org/10.1093/NAR/GKZ974>.
- [145] P.D. Stenson, M. Mort, E. V. Ball, K. Shaw, A.D. Phillips, D.N. Cooper, The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine, *Hum Genet.* 133 (2014) 1–9. <https://doi.org/10.1007/S00439-013-1358-4>.

- [146] P. Engel, M.J. Eck, C. Terhorst, The SAP and SLAM families in immune responses and X-linked lymphoproliferative disease, *Nat Rev Immunol.* 3 (2003) 813–821. <https://doi.org/10.1038/NRI1202>.
- [147] M. Ishibashi, R. Morita, H. Tamura, Immune Functions of Signaling Lymphocytic Activation Molecule Family Molecules in Multiple Myeloma, *Cancers* 2021, Vol. 13, Page 279. 13 (2021) 279. <https://doi.org/10.3390/CANCERS13020279>.
- [148] A. Veillette, Immune regulation by SLAM family receptors and SAP-related adaptors, *Nat Rev Immunol.* 6 (2006) 56–66. <https://doi.org/10.1038/NRI1761>.
- [149] B. Foley, M. Felices, F. Cichocki, S. Cooley, M.R. Verneris, J.S. Miller, The biology of NK cells and their receptors affects clinical outcomes after hematopoietic cell transplantation (HCT) Human NK cell development, *Immunol Rev.* 258 (2014) 45–63. <https://doi.org/10.1111/imr.12157>.
- [150] S. Latour, Inherited immunodeficiencies associated with proximal and distal defects in T cell receptor signaling and co-signaling, *Biomed J.* 45 (2022) 321–333. <https://doi.org/10.1016/J.BJ.2022.01.013>.
- [151] R. Proust, J. Bertoglio, F. Gesbert, The Adaptor Protein SAP Directly Associates with CD3 ζ Chain and Regulates T Cell Receptor Signaling, *PLoS One.* 7 (2012) 43200. <https://doi.org/10.1371/JOURNAL.PONE.0043200>.
- [152] K. Shah, A. Al-Haidari, J. Sun, J.U. Kazi, T cell receptor (TCR) signaling in health and disease, *Signal Transduction and Targeted Therapy* 2021 6:1. 6 (2021) 1–26. <https://doi.org/10.1038/s41392-021-00823-w>.
- [153] C. Borowski, A. Bendelac, Signaling for NKT cell development: the SAP-FynT connection, *Journal of Experimental Medicine.* 201 (2005) 833–836. <https://doi.org/10.1084/jem.20050339>.
- [154] C.W. Arendt, B. Albrecht, T.J. Soos, D.R. Littman, Protein kinase C- θ : signaling from the center of the T-cell synapse, *Curr Opin Immunol.* 14 (2002) 323–330. [https://doi.org/10.1016/S0952-7915\(02\)00346-1](https://doi.org/10.1016/S0952-7915(02)00346-1).
- [155] Y.J. Park, S.A. Yoo, M. Kim, W.U. Kim, The Role of Calcium–Calcineurin–NFAT Signaling Pathway in Health and Autoimmune Diseases, *Front Immunol.* 11 (2020) 455685. <https://doi.org/10.3389/FIMMU.2020.00195/BIBTEX>.
- [156] L. Abdullah, L.B. Hills, E.B. Winter, Y.H. Huang, Diverse Roles of Akt in T cells, *Immunometabolism.* 3 (2021). <https://doi.org/10.20900/immunometab20210007>.

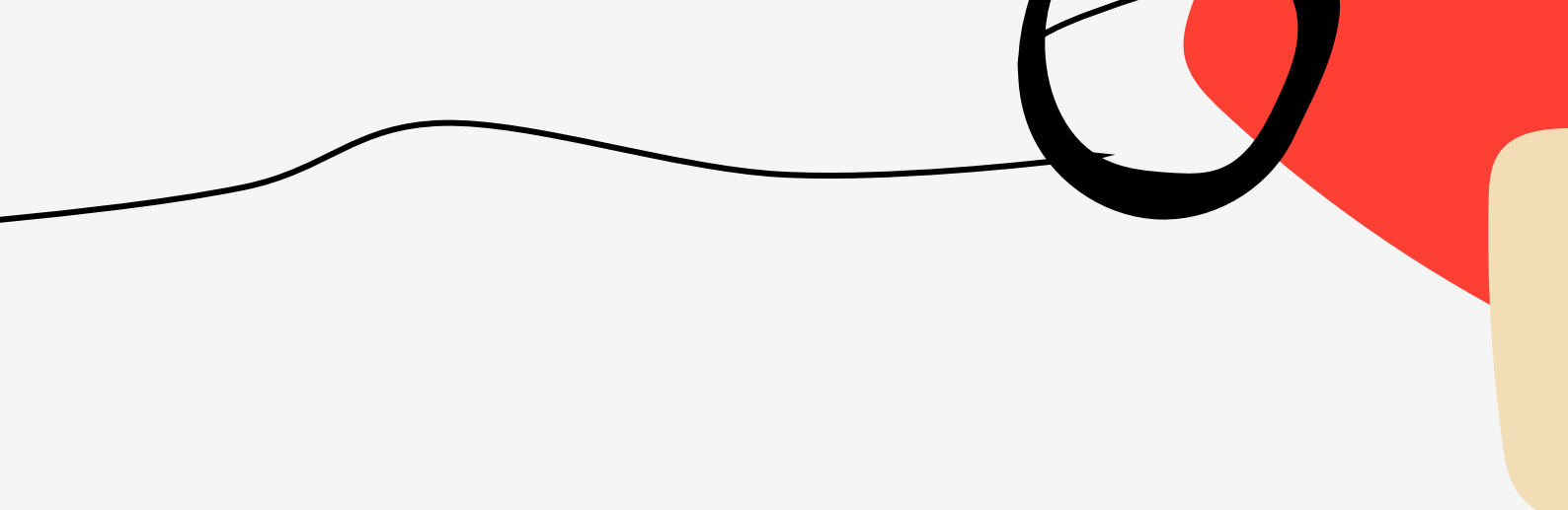
- [157] W. Li, M.H. Sofi, S. Rietdijk, N. Wang, C. Terhorst, C.H. Chang, The SLAM-Associated Protein (SAP)/Fyn/PKC θ Pathway is Required for Thymocyte-mediated CD4 T Cell Development, *Immunity*. 27 (2007) 763. <https://doi.org/10.1016/J.IMMUNI.2007.10.008>.
- [158] E. Bell, NKT-cell defects in XLP, *Nature Reviews Immunology* 2005 5:4. 5 (2005) 265–265. <https://doi.org/10.1038/nri1595>.
- [159] Z. Cao, Q. Liao, M. Su, K. Huang, J. Jin, D. Cao, AKT and ERK dual inhibitors: The way forward?, *Cancer Lett.* 459 (2019) 30–40. <https://doi.org/10.1016/J.CANLET.2019.05.025>.
- [160] L. Cifaldi, M. Doria, N. Cotugno, S. Zicari, C. Cancrini, P. Palma, P. Rossi, DNAM-1 Activating Receptor and Its Ligands: How Do Viruses Affect the NK Cell-Mediated Immune Surveillance during the Various Phases of Infection?, *International Journal of Molecular Sciences* 2019, Vol. 20, Page 3715. 20 (2019) 3715. <https://doi.org/10.3390/IJMS20153715>.
- [161] A.K. Ali, N. Nandagopal, S.H. Lee, IL-15–PI3K–AKT–mTOR: A Critical Pathway in the Life Journey of Natural Killer Cells, *Front Immunol.* 6 (2015) 1–22. <https://doi.org/10.3389/FIMMU.2015.00355>.
- [162] Y. Chen, D. Lu, A. Churov, R. Fu, Research Progress on NK Cell Receptors and Their Signaling Pathways, *Mediators Inflamm.* 2020 (2020). <https://doi.org/10.1155/2020/6437057>.
- [163] H.J. Kwon, N. Kim, H.S. Kim, Molecular checkpoints controlling natural killer cell activation and their modulation for cancer immunotherapy, *Experimental & Molecular Medicine* 2017 49:3. 49 (2017) e311–e311. <https://doi.org/10.1038/emm.2017.42>.
- [164] J. Cannons, L. Yu, B. Hill, L. Mijares, D. Dombroski, K. Nichols, A. Antonellis, G. Koretzky, K. Gardner, P. Schwartzberg, SAP Regulates T(H)2 Differentiation and PKC- θ -Mediated Activation of NF- κ B1, *Immunity*. (2004) 693–703. <https://doi.org/10.1016/j.immuni.2004.09.012>.
- [165] S. Paul, G. Lal, The molecular mechanism of natural killer cells function and its importance in cancer immunotherapy, *Front Immunol.* 8 (2017) 290728. <https://doi.org/10.3389/FIMMU.2017.01124/BIBTEX>.

- [166] S. Chen, Z. Dong, NK cell recognition of hematopoietic cells by SLAM-SAP families, *Cellular & Molecular Immunology* 2019 16:5. 16 (2019) 452–459. <https://doi.org/10.1038/s41423-019-0222-4>.
- [167] A. Bodaghi, N. Fattahi, A. Ramazani, Biomarkers: Promising and valuable tools towards diagnosis, prognosis and treatment of Covid-19 and other diseases, *Heliyon*. 9 (2023) e13323. <https://doi.org/10.1016/J.HELIYON.2023.E13323>.



CHAPTER

II



Evolutionary genomic analyses of *SH2D1A*, *SH2D1B* and *XIAP* genes in primates: Implications on immune functionality

Nuno Sousa^{1,2}, Luana Ramos^{1,2}, Agostinho Antunes^{1,2}

¹ Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre, 4169-007, Porto, Portugal

² CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, 4450-208, Porto, Portugal



2.1. Abstract

Advancements in sequencing techniques have allowed for the acquisition of better whole genome data, facilitating the understanding of gene evolution. Comparative genomics plays a crucial role in identifying the selective forces acting on genes, detecting positive selection, and revealing biological interactions that influence their evolution. Notably, immune system-related genes often exhibit positive selection, emphasizing the significance of host-pathogen dynamics in mammals. The *SH2D1A*, *SH2D1B* and *XIAP* proteins play important roles in the body's immune response to various types of pathogens. Mutations in the *SH2D1A* and *XIAP* genes are responsible for the development of the X-linked lymphoproliferative syndrome (XLP). In this study, we performed evolutionary molecular analyses of the *SH2D1A*, *SH2D1B* and *XIAP* genes with the aim of understanding the role of natural selection in the evolutionary history of these genes, with particular emphasis on the lineage of the great apes, and its putative influence on the development of their important immune-related function. Accordingly, the coding sequences of each gene were extracted from non-annotated genomes of representative species of great apes, and of the remaining primate families and mammalian orders. We performed phylogenetic maximum-likelihood reconstructions in IQ-TREE, which served as the basis for performing selection analyses using the CODEML program and the Datamonkey Adaptive Evolution Server. We also performed syntenic analyses for the three genes. Our results indicate that all genes, especially *SH2D1A* and *XIAP*, are highly conserved in mammals, with strong purifying selection acting on both. These results are in accordance with the maintenance of a crucial immune function in the mammalian lineage, as well as the conservation of the structural configuration of the proteins, showing that deleterious variations, as in the case of XLP, are negatively selected in all protein structures and can be removed from the population.

Keywords: X-linked lymphoproliferative syndrome, *SH2D1A*, *SH2D1B*, *XIAP*, Selection analyses, Molecular evolution

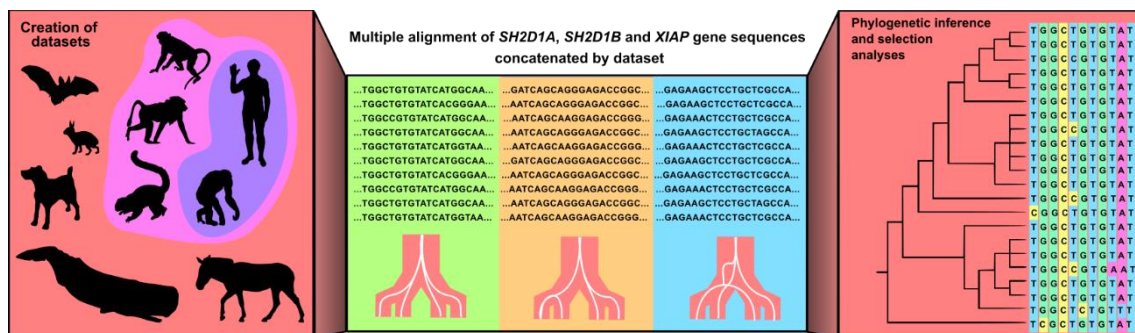


Figure 1: Graphic summary of the work carried out in this study.

2.2. Introduction

The improvement of sequencing techniques over the years has allowed scientists to obtain increasingly better and more complete whole genome assemblies for several species, contributing to deepen the general knowledge about gene evolution [1]. Comparative genomics thus plays an important role in understanding the evolution of genes and their associated phenotypic traits by enabling, for instance, the study of potential selective forces on a genomic level. Selective pressures play a key role in the evolution of the structure and complexity of protein domains. Understanding the evolutionary background behind specific candidate genes, by revealing the selective forces that have shaped their function and diversity over time, can help to understand their roles in underlying molecular and cellular mechanisms [2].

Detecting positive selection in genes, sites or lineages provides clues about evolutionary adaptation, helping to uncover biological interactions that influence their evolution, such as parasite-host coevolution. Strong signals of negative (purifying) selection, on the other hand, help identify functional constraint or evolutionary conservation [3]. This information is helpful to infer essential structural or functional properties of certain proteins or domains, as well as evolutionary similarities between different species or clades [4,5]. Previous studies using comprehensive analyses of whole genomes, with a particular focus on primates, have shown subsets of proteins under positive selection [6,7]. The vast majority of these are encoded by genes related to the immune system, especially proteins that interact with pathogens, highlighting the importance of host-pathogen interactions in protein adaptation in mammals [8–10]. Furthermore, a great similarity in immune mechanisms was observed, indicating that they are ancient and maintained throughout the evolution of vertebrates, such as the case of the NF- κ B pathway. This pathway is responsible for the activation of several different defence mechanisms, which is shown to be ubiquitous in almost all organisms. However, there are also differences and variations between organisms in how they express, regulate and utilise these mechanisms, resulting in a diversity of immune responses. This diversity may reflect the adaptations of organisms to their environments and the pathogens they encounter [11].

Studying the evolution of genes related to the immune response is a promising field in identifying host defence mechanisms that have been conserved or modified throughout the evolutionary history of organisms. These mechanisms are fundamental for human health as they may affect susceptibility or resistance to infectious or autoimmune diseases. For example, studying the evolution of the complement system, which is part of the innate immune response, can reveal how this system has diversified and

specialised to fight different types of pathogens [12,13]. In addition, studying the evolution of genes that interact with viruses can identify their molecular targets and the possible escape strategies they use to infect host cells [11]. Therefore, unravelling the evolution of certain candidate genes can help to understand diseases related to these genes and to develop new therapeutic or preventive interventions [14,15].

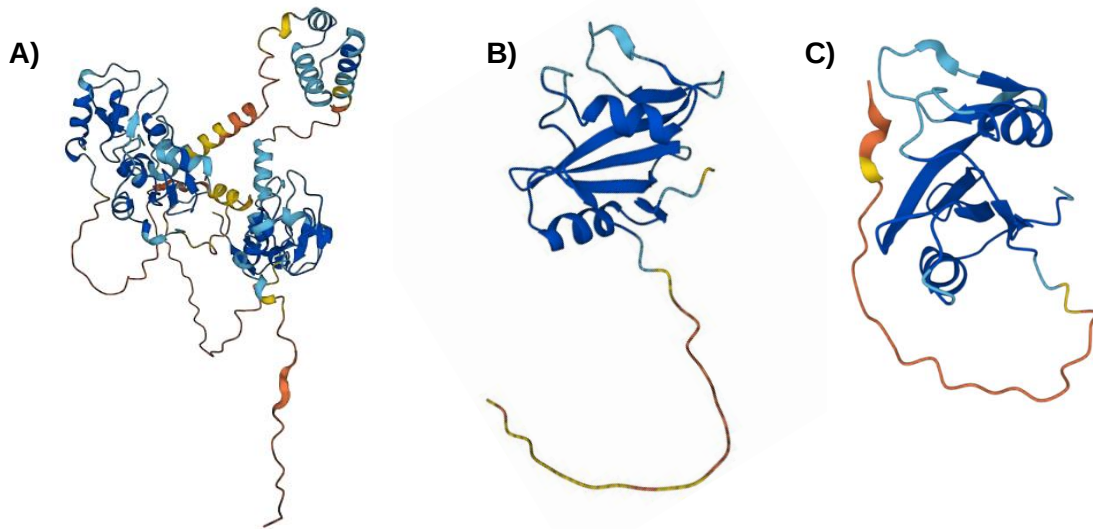


Figure 2: Predicted protein structures of *XIAP* (A), *SH2D1A* (B) and *SH2D1B* (C) genes. Extracted from the AlphaFold database [16].

Several genes are of particular interest due to their important role in T and NK cell activation and differentiation, mainly through the NF- κ B pathway [17–24]. Reported mutations in *SH2D1A* and *XIAP* genes have shown very negative impacts on the immune response of affected individuals. These seem to be involved in the X-linked lymphoproliferative syndrome (XLP), a rare immunodeficiency portrayed by an insufficient immune response to Epstein-Barr virus (EBV) infection [24–27]. The X chromosome contains a high density of immune-related genes and regulatory elements that are extensively involved in both the innate and adaptive immune responses, which can contribute to differences in immune responses between sexes (see Chapter I) [28].

In order to fill the gap of information regarding the molecular evolution of the immune-associated genes *XIAP*, *SH2D1A*, and its paralogous *SH2D1B* (which interacts with *SH2D1A*) (Figure 1), we aim to understand how selection shaped the immune response of these genes across the evolution of mammals, focusing on the primate order and the Human lineage in particular. Employing several selection detection methods, the selective forces behind the evolution of these genes can be evaluated using the ratio of non-synonymous (dN) to synonymous (dS) mutation rates ($\omega = dN/dS$) [29].

2.3. Material and Methods

Genomic Data Compilation

Representative species of each family of primates and of each of the remaining mammalian orders were selected from publicly available genomic data, allowing the study of evolutionary patterns from a broader to a more detailed scale (Table S1 and S2). To obtain the most reliable genomic information, a non-annotated genome assembly for each species was chosen from the NCBI Genome database [30] according to the following quality parameters – largest total genome size; lowest number of scaffolds; preference for chromosome-level assemblies that include the sex chromosomes; and the sequencing platform used, giving preference to long-read sequencing. The use of these third-generation technologies, often combined with short-read sequencing, allows a more robust assembly of highly repetitive regions, which are common in sex chromosomes [31]. The coding sequences (CDS) of the *XIAP*, *SH2D1A* and *SH2D1B* genes were extracted using the protein2genome tool of the Exonerate v2.4.0 package [32]. The human protein sequence of the most extensive transcriptional variant, available at the UniProt database [33], was used as query for each gene (table S4). Sequences with intrasequential stop codons or frameshifts were considered pseudogenes and discarded from analyses. Each Exonerate output was inspected for intron number, size and position in the obtained nucleotide sequence.

Dataset preparation and multiple sequence alignment

The nucleotide sequences of the different species were grouped into three different datasets for each gene: *i*) a general dataset, composed of all mammalian orders, *ii*) a primate dataset, compiling all families of this order, and *iii*) a Great apes dataset (detailed in table S3).

The basis of phylogenetic analyses is undertaken through multiple sequence alignments (MSA), obtaining the first inferences about the similarity of the orthologous sequences [34]. The CDS from each dataset were aligned on the GUIDANCE2 webserver [35], using the associated MAFFT algorithm [36] with 100 bootstrap replicates. Different alignment algorithms generate MSAs based on the similarity between sequences, assigning different penalties depending on the lack of parity or the existence of gaps between sequences [36–39]. In addition, the GUIDANCE2 server provides a phylogenetically informed confidence score for each residue of the alignment, by generating and comparing several alternative MSAs [35,38]. This characteristic allows an automated

refinement of the DNA sequences, eliminating unreliable columns [36]. Thus, columns with confidence score below the predefined value of 0.93 were removed. Subsequently, the sequences underwent manual refinement using the Molecular Evolutionary Genetics Analysis 7.0 (MEGA7) software [40] to eliminate columns or sequences presenting more than 20% of missing data.

Recombination and saturation analyses

Using the Genetic Algorithm for Recombination Detection (GARD) method [41], available at the Datamonkey Adaptive Evolution Server [42], we then tested for potential recombination points in each of the nine datasets, which allows further analyses to account for variable evolutionary patterns within each gene.

Substitution saturation in MSAs occurs when high mutation rates over time lead to sequence convergence in several residues after multiple substitutions, interfering with phylogenetic analyses by misestimating genetic distances [43]. Xia's test was applied to each MSA through the Data Analyses for Molecular Biology and Evolution 5 (DAMBE5) software [44], employing 60 replicates.

Phylogenetic inference

The phylogenetic gene trees provide a representation of the evolutionary history of each gene, allowing to obtain information about their evolution across the mammalian lineage by comparing them to phylogenomic species trees [34,45,46]. To ensure the best phylogenetic signal was attained, the obtained MSAs were concatenated for each dataset using Geneious 11.0.5 software [47].

The phylogenetic inferences were obtained for each dataset through the Maximum Likelihood (ML) method, implemented on the IQ-TREE 2.0.7 software [48], with 10000 ultrafast bootstraps (bb) [49]. The choice of an appropriate evolutionary model is an essential process for phylogenetic reconstruction, providing information regarding the mutation rates in the MSA datasets. Considering this, the ModelFinder [50] option of the IQ-TREE software was used as it provides an accurate approach for best evolutionary model and partitioning scheme selection. A general tree that includes all studied species was also created using the same methodology. Finally, the gene concordance (gCF) and the site concordance factors (sCF) [51] were also calculated using IQ-TREE 2.0.7, defining the percentage of gene trees (gCF) and percentage of aligned sites (sCF) within the dataset that support the final tree topology. In addition to these ML inferences, three species trees were created for each dataset and one species tree for the complete

dataset (all species used) using the fifth edition of the TimeTree of Life resource (TTOL5) [52]. All resulting phylogenetic trees were edited on MEGA7.

Selection analyses

Selection analyses detect signals of evolutionary pressures present in the CDS of each gene [53]. This calculation is based on the ratio of non-synonymous (dN) to synonymous (dS) mutation rates ($\omega=dN/dS$), being positive selection when $\omega>1$, purifying selection when $\omega<1$ and neutral selection when $\omega\approx 1$ [29]. Selection was evaluated between sequence sites and between phylogenetic branches using site and branch-site model tests from the Datamonkey Adaptive Evolution Server [42], and the CODEML program of the Phylogenetic Analysis by Maximum Likelihood (PAML) v4.6 package [54], supported by the Lightweight Multigene Analyses in PAML v1.0.2 (LMAP) software [55]. The existence of positively selected sites (PSS) might indicate a potential change in the structure, and consequently, on the function of the protein, while signals of negative selection point to restrictive pressures to maintain its function.

To ensure that the MSA for each gene is appropriate to be used for selection analyses, the pairwise dS values were calculated using the DNA Sequence Polymorphism (DNASP v6.12.03) [56]. To avoid compromising the estimation of ω , the ratio of synonymous substitutions should remain below the recommended value of $dS < 2$ [57].

PAML analyses

Branch-site models test for positive selection signals at specific codon sites across different lineages of interest marked as foreground branches. For this purpose, an Alternative two-ratio model (MA), which allows for positive selection, was used in comparison with a Null one-ratio model (MA1), which fixes the ω value as 1 [58]. For these analyses, in each dataset different foreground lineages were marked: *i*) the ancestral lineage and the apical branches (in separate analyses) of the primate order in the general mammalian dataset, to test for potential selection signals in these genes that occurred during the evolution of primates; *ii*) the ancestral lineage and the apical branches of the Great apes in the primate dataset; and *iii*) the *Homo sapiens* lineage in the Great apes' dataset.

Site models test for positive selection signals at different codon sites. Although several models have been developed, in this study M0, M7, M8, and M8a were used. Model M0 assumes a uniform selective pressure among sites, providing an overall estimation for the gene; M7 is a Null model that allows for variable selection over sites, without positive

selection; M8 is the Alternative model also based on variable selection but, in contrast with M7, allowing positive selection to occur; and M8a is an additional Null model that fixes the ω value as 1 [59]. In this way, model M8 is compared with Null models M7 and M8a [60]. The PAML site model analyses were conducted for the background species, in order to discard potential sites under selection on the foreground branches that are not exclusive to these lineages.

For both analyses, the Null and Alternative models were compared using the Likelihood-Ratio Test (LRT) [59], and the Alternative hypothesis was accepted when the LRT test was significant ($p < 0,05$). Furthermore, both analyses provide a Bayes Empirical Bayes (BEB) test [61] that calculates the posterior probability (PP) of a given site being under positive selection [60]. PSS were only considered if the PP were above 0.95.

HyPhy analyses

Four methods were used to compute selection analyses on a per-site basis in the HyPhy software, implemented on the Datamonkey webserver, for each of the complete datasets. The Fixed Effects Likelihood (FEL) tests whether positive selection has occurred at a set of sites in the alignment, assuming that selection pressure is constant along the entire phylogeny for each site. It uses a ML approach to infer the dN and dS rates per site. The Single-Likelihood Ancestor Counting (SLAC) is a similar method to FEL but combines ML and counting approaches [62].

The Fast Unconstrained Bayesian Approximation (FUBAR) uses a Bayesian approach to infer per-site substitution rates, also assuming that the selection pressure is constant. This test reports evidence of positive selection with posterior probabilities (ranging from 0-1), in which values above 0.9 are strongly suggestive of positive selection. It has advantages over FEL when alignments are larger and positive selection is present but weak [63]. Finally, the Mixed Effects Model of Evolution (MEME) uses a mixed-effects ML method to infer sites under positive selection not only on the whole phylogeny, but also occurring on a subset of lineages (episodic selection). As a final integrative approach, only the PSS detected by at least three of the four methods were considered, following previously published approaches [64].

The Branch-site Unrestricted Statistical Test for Episodic Diversification (BUSTED) is a statistical test that detects positive selection by testing if it has occurred in at least one site in at least one of the pre-defined branches of the genetic tree. It does not imply that the whole gene evolved under positive selection [65]. In contrast, the RELAX test allows us to understand whether there has been a relaxation or intensification of natural

selection in the foreground branches. A significant result of $k > 1$ indicates intensification of selection, while $k < 1$ suggests relaxation [66]. The foreground branches in the phylogenetic trees of each dataset were marked according to the described for the PAML branch-site model test, for both BUSTED and RELAX analyses. All analyses were run using default parameters.

Additionally, the calculation of the conservation and evolutionary acceleration of sites within the genes was performed using PhyloP on the PhastWeb webserver [67]. This tool provides a positive score to conserved sites and a negative score to sites in evolutionary acceleration [53].

Syntenic analyses

Syntenic reflects the chromosomal location of each gene and its corresponding flanking genes, which can be preserved or altered across different species during evolution [68]. Translocations or inversions have the potential to alter the physical framing of genes in the genome, impacting their transcription and regulation [69]. Syntenic analyses were carried out to clarify the genes' evolutionary history and identify genetic rearrangements in humans, which might be associated with specific disease phenotypes. The Genomicus database [70] was consulted to identify which genes are located in the flanking regions of each of the studied genes. The Exonerate software was then employed to search for the location of the flanking genes in several genomes, following the extraction methodology described above. The accession number of each protein taken from the UniProt database to be used as query is listed in table S4. The syntenic block was determined for each of the three genes in 13 species out of the total number of species used in the study. Of these 13 mammal species, seven belong to the primate order.

The guanine-cytosine (GC) content of the syntenic block, as well as the graphic visualization of the order of the genes, were calculated and created using the packages gggenomes version 0.9.9.9000 [71] and seqinr version 4.2-30 [72] from R v3.4.1 software.

2.4. Results

Sequence collection and MSA construction

The sequences of the *SH2D1A*, *SH2D1B* and *XIAP* genes were extracted from 51 primate species of 14 different families and 23 species from the remaining mammalian orders. The extracted sequences were grouped by dataset. Due to poor quality, some sequences were discarded, and the final composition of each dataset can be seen in

[table S3](#). The results of the GARD method confirmed the existence of recombination points (RPs). Two RPs were found in the *XIAP* gene (123 and 377 bp) and one in the *SH2D1A* gene (129 bp) in the Mammalia dataset. The last one is in the *XIAP* gene of the primate dataset (879 bp). Saturation analyses for each gene in the three MSA datasets revealed statistically significant low saturation signatures, confirming that all datasets are suitable for phylogenetic inference ([table S5](#)).

Phylogenetic inference based on the mammalian SH2D1A, SH2D1B and XIAP genes

The evolution of the immunity-associated genes *SH2D1A*, *SH2D1B* and *XIAP* was initially assessed by analysing the phylogenetic tree based on the concatenated alignment of all species selected for the study ([Figure 3](#)). In general, the bootstrap values of the tree with all species, as well as the trees for each dataset ([Figure 1S](#), [2S](#) and [3S](#)), show high support for the obtained topology. Almost all clades show support values above 70%, most above the 95% threshold considered highly reliable for ultrafast bootstraps [73]. When comparing with the phylogenomic tree obtained by Esselstyn *et al.* [74], the overall topology of the phylogenetic trees is similar when considering the mammalian orders. However, our results contrast with the species tree in several taxonomic lineages. In particular, the primate order appears as paraphyletic in our results ([Figure 3](#)). Within this order, several inconsistencies can be found. The great apes' lineage, for instance, does not appear as a monophyletic cluster, as it is possible to observe that the group composed by the branch of genus *Pongo* and the branch containing the remaining great apes does not contain all the descendants of their most recent common ancestor. Contrasting with the high bootstrap support, most branches reveal low gCF and sCF values.

Inspecting the selective forces

The search for signs of selection was carried out using PAML and Hyphy analyses. To identify sites under positive selection in the foreground species (*Homo sapiens*, great apes and primates, depending on the dataset, as described in Material and Methods), the CODEML branch-site model analysis was applied to both the apical branches and the ancestral branch (common ancestor). The statistical hypothesis testing (SHT) yielded non-significant (p-value for the LRT tests higher than 0.05) results across all datasets ([Tables S6](#), [S7](#) and [S8](#)). The site model analyses were applied to the background species (all species not selected as foreground in the branch-site analysis) of each

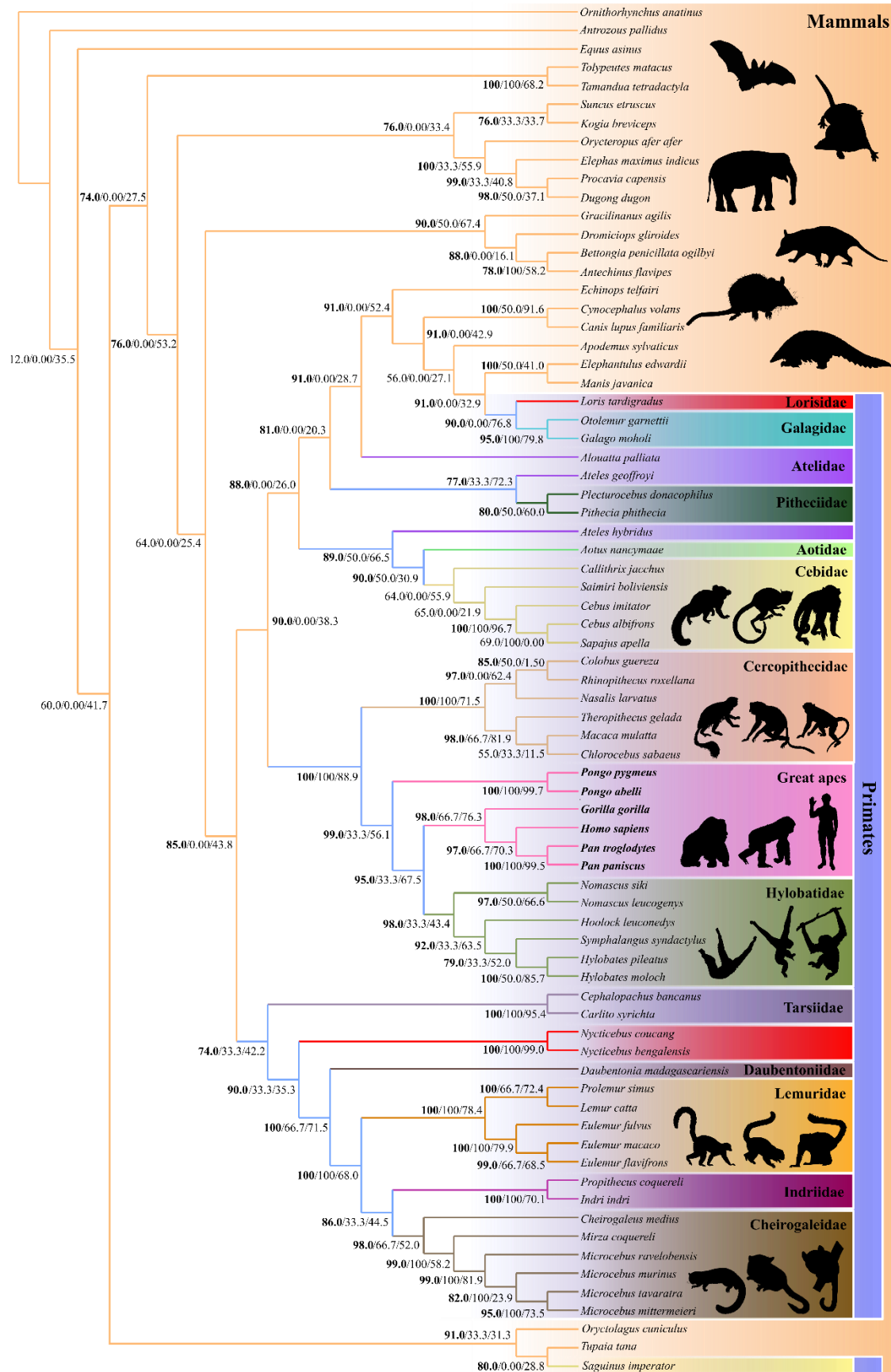


Figure 3: Cladogram representation of the phylogenetic tree based on the concatenated *XIAP*, *SH2D1A* and *SH2D1B* genes, created using IQ-TREE 2.0.7 with the Maximum Likelihood method. The support values follow the order: bb/gCF/sCF. Bootstrap values above 70% are marked in bold.

dataset. Although some results showed significant signs of positive selection (mammalian dataset for the *SH2D1B* and *XIAP* genes, primates dataset for the *SH2D1B* gene) (Tables S9, S10 and S11), with statistically significant PSS ($PP > 0.950$) in the Bayes Empirical Bayes (BEB) analyses (Tables S12 and S13), those were not relevant considering the lack of significant results in the branch-site analysis.

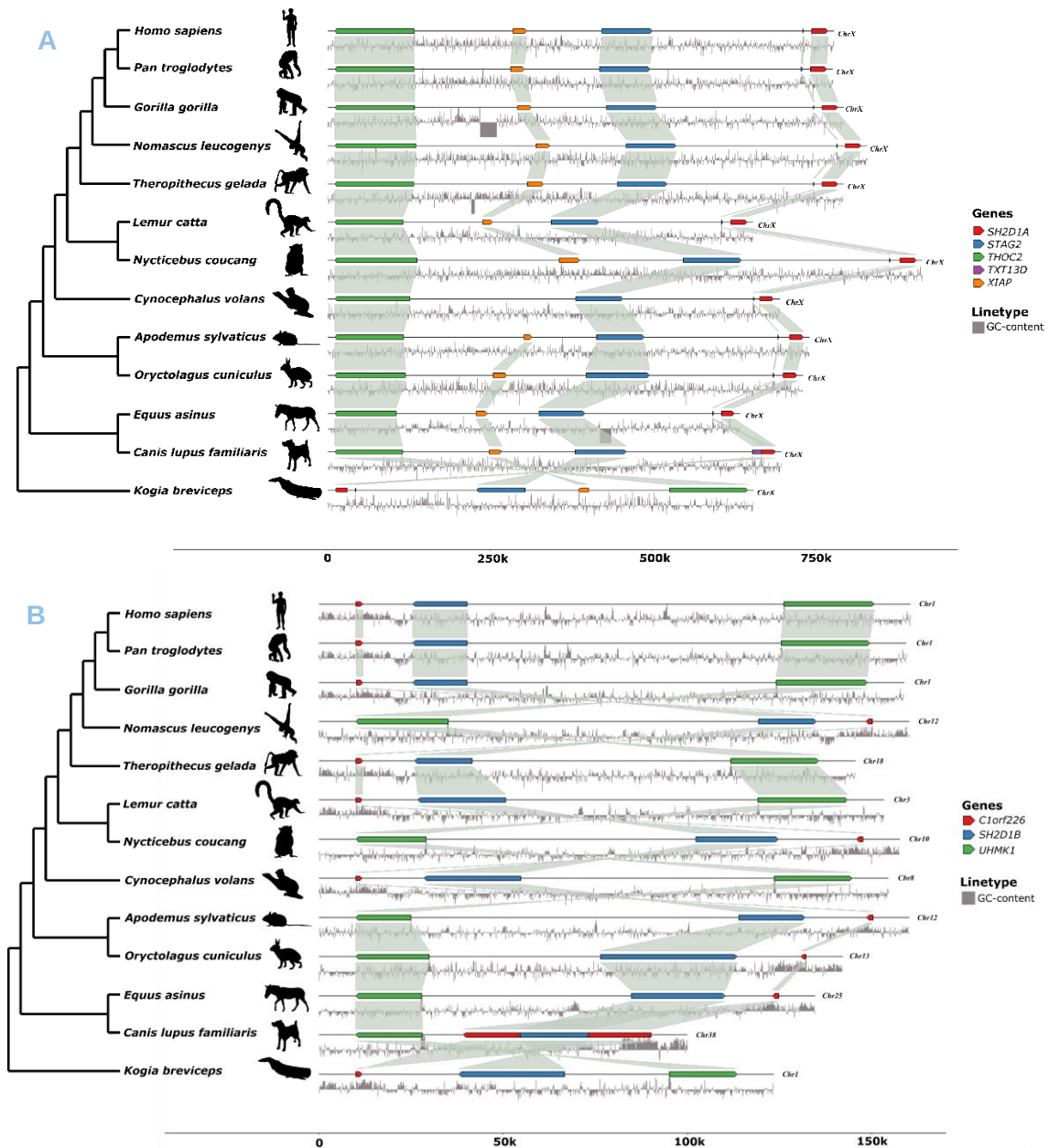


Figure 4: Schematic representation of the structure of the syntenic block of the *SH2D1A* and *XIAP* (A) genes, and *SH2D1B* gene (B). Each black line represents the block with the different genes (arrows) represented in different colours. Below each line is a graphical representation of the GC content along the block. In both syntenic maps, the wiggle bounds of the GC content are 0.2 (low), 0.4 (mid) and 0.6 (high). The numerical scale represents the number of base pairs (bp) in the syntenic blocks.

The Hyphy site-model analyses were conducted for all lineages, without marking foreground branches, in order to analyse the overall pressures acting on the residues of

each gene. The FEL, SLAC, FUBAR and MEME tests show no sign of PSS in the great apes dataset. In the primate dataset, some sites under negative selection were found in the three genes (Tables S17, S18 and S19). In contrast, sites under positive selection were found only for *SH2D1B* (Tables S18). A similar situation occurs in the Mammalia dataset, for which only the *XIAP* gene presents a site under positive selection (Table S15). In all datasets, the BUSTED and RELAX analyses revealed non-significant results.

The PhyloP and PhastCons results show a high degree of conservation of the genetic sequences of the three genes for the primate and Mammalia dataset (Figure S4, S5 and S6).

Mapping the genes in syntenic blocks

The syntenic block comprising the *SH2D1A* and *XIAP* genes appears to be highly conserved, considering the stable position of their flanking genes as well as their syntenic locations and orientation, especially in great apes (Figure 4A). The *XIAP* gene is always flanked by genes *THOC2* and *STAG2*, and the *SH2D1A* gene appears always after the *TXT13D* gene. We highlight, however, the translocation of the *XIAP* gene from chromosome X to chromosome 13 in the species *Cynocephalus volans* (not shown) and the inversion of the whole syntenic block in the whale species *Kogia breviceps*. In contrast, the synteny of the *SH2D1B* gene is marked by several events of inversion and translocation of the syntenic block (always within autosomes), but with conservation of the order and the flanking genes (Figure 4B). However, the great apes show once again a high conservation of the chromosomal location and of the synteny of genes in this block. Despite the processes of inversion and chromosomal translocation, the homologous blocks in both syntenies show the same flanking genes especially in for the *SH2D1A* and *XIAP* genes.

2.5. Discussion

Comparative genomics is crucial for studying gene evolution and its impact on phenotypic traits, including the examination of selective forces acting on a genomic level. Studying the selective pressures involved in genes related to T and NK cell activation and differentiation, such as the *SH2D1A* and *XIAP* genes, is essential since these forces are responsible for shaping the structure and complexity of protein domains, providing insights into their function and diversity over time.

The construction of phylogenetic trees is the basis of most comparative genomic studies, being of great importance in understanding the evolutionary processes that shape the

structure and functionality of the various genes and proteins throughout their evolutionary history [75,76]. As described in the results section, when we compare the obtained phylogenetic tree (Figure 3) with a reference species tree from Esselstyn *et al.* [74], several incongruencies are found in our topology. In our phylogenetic reconstruction, well defined mammalian orders and families do not appear as monophyletic clades. This shows that our phylogeny is unable to represent the mammalian evolutionary history. This phenomenon is particularly noticeable in the primate order, which is represented as a paraphyletic clade, sharing the common ancestor with other mammalian orders. In addition to the order itself, various families within Primates also show differences in topology, such as the great apes' group. Despite finding no evidence of saturation and obtaining high support values, these differences in topology are expected. While our reconstruction is based on the evolutionary history of just three genes, the phylogenomic species tree [74] uses a large amount of genomic information to generate its topology [77,78]. The concatenation of several genes enables an increase in the number of sites being analysed per species, consequently enhancing the phylogenetic signal, helping to overcome this problem [29]. However, each gene has its own evolutionary history, evolving at different rates of nucleotide substitutions and subjected to variations in selective pressures. The candidate genes *XIAP*, *SH2D1A* and *SH2D1B* have crucial roles in immune responses [14,26,79], which provided a particular evolutionary environment. In addition, *XIAP* and *SH2D1A* are located in the X chromosomes in most mammalian lineages [80,81] (Figure 4), which influences the evolution of the genes located in it due essentially to their hemizygous nature in male individuals (see Chapter I). The functional necessity of the X-linked genes in both sexes leads them to undergo stronger purifying selection (selection against deleterious mutations) than autosomal genes [82]. Unlike *SH2D1A* and *XIAP*, the *SH2D1B* gene is found on autosomes.

Considering this, we hypothesize that the divergent evolutionary frameworks of our genes are mostly responsible for the inconsistencies between our phylogenetic reconstruction and the species tree, which is consistent with the low concordance values we obtained. Since gCF and sCF measure the share of gene trees containing the same branch and the percentage of alignment sites that support that branch [51,83], respectively, we can see that in the vast majority of cases, although the bootstrap values are high, the sCF and gCF values are quite low. This phenomenon indicates the existence of multiple conflicting signals caused by insufficient information and stochastic errors associated with it. To guarantee that the phylogenetic trees used for further analyses accurately represent mammalian evolution, it was necessary to construct

separate trees for each dataset. These individual species trees, created using TimeTree, were then utilized for the subsequent selection analyses.

Given that each dataset represents a distinct clade, and each monophyletic clade is nested within the previous (specifically, mammals > primates > great apes), employing this type of “funnel” analysis for selection allows us to discern if any sites have undergone positive selection in the common ancestors of the lineages that lead to *Homo sapiens*, investigating putative neofunctionalizations that might have occurred in the evolution of Primates, Great apes or the *Homo* lineage.

Looking closely at the PAML results, the three genes seem to be overall under broad conservation, which can be seen by the absence of significant results in the branch-site model when we test for our branches of interest. These results are supported by the non-significant RELAX and BUSTED results, which seem to suggest that there were no remarkable changes in selective pressures in the evolution of these genes towards the Human lineage. Also in agreement, in the general analysis of site-models (hyphy) we have more signals of sites under purifying selection than PSS. In the mammalian dataset, each of the genes of interest has numerous sites subjected to purifying selection, while only a single site shows evidence of positive selection. A pattern of negative selection is also evident in the primate data set, with many of these sites overlapping with those in the mammal dataset, albeit to a lesser extent. In particular, three sites subject to positive selection stand out in the *SH2D1B* gene, located in the SH2 domain of the protein, responsible for binding to SLAMs, highlighting site 49 which interacts directly with CD150 [84]. In contrast, no significant signals of selective pressures were detected in the great ape dataset. These findings align with the hypothesis proposed by Sackton *et al.* [85], suggesting that immune genes responsible for pathogen recognition exhibit more pronounced signs of positive selection compared to signal transduction molecules, which are more likely to undergo purifying selection. The continued purifying selection observed both in a general mammalian dataset and in the primates group suggests that these genes must have extremely important functions, in all mammals, possibly even having similar immune-related functions in other vertebrate groups. This conservation is also supported by the results of the PhyloP and PhastCons analyses (Figure S4, S5 and S6). These results show that, as well as being important in human health, studying these genes in other organisms can also be relevant to animal conservation and veterinary science studies.

The genetic conservation suggested by the results of the selection analyses is also corroborated by the high similarity between the various primate and mammalian species

in the syntenic map of the *SH2D1A* and *XIAP* genes, indicating conservation of the genetic locations and direction of each of the genes. As for the syntenic map of the *SH2D1B* gene, although the flanking genes remain continuous in the different species, there is a clear occurrence of several syntenic block inversion events, altering the direction of the genes, as well as translocation events, clear by the changes in chromosome location between the different species. These evolutionary events are especially evident within the primate order. It may be necessary to carry out more extensive synteny studies, encompassing more species. To this end, genome sequencing using third-generation sequencing techniques, such as high-fidelity reads (HiFi) from Pacific Biosciences (PacBio) and ultra-long reads from Oxford Nanopore (ONT), which were recently used to completely sequence the human Y chromosome without gaps [86], would be very advantageous for comparative genomics studies. These technologies make it possible to obtain longer and more accurate reads, which is especially useful for assembling complex genomes and resolving regions that are difficult to sequence, such as the sex chromosomes [87–90].

Since the sex chromosomes do not pair with each other for most of their length, inversions and translocations in these inert genetic zones end up accumulating more frequently than in autosomal regions [91,92]. The degrees of conservation of each of the syntenic blocks under study expose the functional importance of the associated genes. Since these events can lead to functional disruption [93], natural selection can favour the stability of the location of genes with high importance for the functioning of the organism, conserving their function throughout evolutionary history.

2.6. Conclusion and Future Perspectives

Genes like *SH2D1A* and *XIAP* play crucial roles in immune responses and studying the selective pressures that acted on them throughout evolution provides insights into protein domain structure and function over time. Molecular evolution and comparative genomics studies are important for understanding the evolutionary context and selective forces acting on disease-associated genes, and can shed light on their ancestral functions, how these may have changed over time and their functional importance. This information can help identify conserved regions or domains critical to their function, providing potential therapeutic targets. The identification of closely conserved genes that cross species provides potential candidates for functional studies and may help to unravel the common mechanisms involved in disease.

In general, the analyses of selection and synteny suggest significant conservation in the studied genes, particularly in the *SH2D1A* and *XIAP* genes located on the X chromosome. These show signs of stronger purifying selection than the autosomal *SH2D1B* gene. The conservation of syntenic blocks, especially on the sex chromosomes, shows the functional importance of the associated genes since natural selection favors stability in the location of genes when they are critical for the functioning of the organism. Possibly, the functions of these genes are ancestral and transverse to all mammalian groups, and perhaps to many other vertebrates. This high degree of continuous conservation explains severe phenotypes caused by simple nucleotide changes (e.g. SNPs), such as XLP.

Although this study justifies the great severity of the mutant phenotypes in these genes, future studies in the area of proteomics could help to better understand the impact of different mutations on the development of the disease. Understanding the structural changes in the proteins encoded by these genes could help understand the underlying mechanisms of the disease in greater depth, from a molecular point of view, and could provide more specific therapeutic targets.

References

- [1] S.C. Watkinson, Molecular Ecology, The Fungi: Third Edition. (2016) 189–203. <https://doi.org/10.1016/B978-0-12-382034-1.00006-2>.
- [2] T. Kawashima, Comparative and Evolutionary Genomics, Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics. 1–3 (2019) 257–267. <https://doi.org/10.1016/B978-0-12-809633-8.20236-7>.
- [3] M. Somel, M.A.W. Sayres, G. Jordan, E. Huerta-Sanchez, M. Fumagalli, A. Ferrer-Admetlla, R. Nielsen, A Scan for Human-Specific Relaxation of Negative Selection Reveals Unexpected Polymorphism in Proteasome Genes, Mol Biol Evol. 30 (2013) 1808–1815. <https://doi.org/10.1093/MOLBEV/MST098>.
- [4] T.B. Sackton, Studying Natural Selection in the Era of Ubiquitous Genomes, Trends in Genetics. 36 (2020) 792–803. <https://doi.org/10.1016/J.TIG.2020.07.008>.
- [5] D. Weghorn, S. Sunyaev, Bayesian inference of negative and positive selection in human cancers, Nature Genetics 2017 49:12. 49 (2017) 1785–1788. <https://doi.org/10.1038/ng.3987>.

- [6] R. Nielsen, C. Bustamante, A.G. Clark, S. Glanowski, T.B. Sackton, M.J. Hubisz, A. Fledel-Alon, D.M. Tanenbaum, D. Civello, T.J. White, J.J. Sninsky, M.D. Adams, M. Cargill, A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees, *PLoS Biol.* 3 (2005) e170. <https://doi.org/10.1371/JOURNAL.PBIO.0030170>.
- [7] R. Van Der Lee, L. Wiel, T.J.P. Van Dam, M.A. Huynen, Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts, *Nucleic Acids Res.* 45 (2017) 10634–10648. <https://doi.org/10.1093/NAR/GKX704>.
- [8] M. Croze, D. Živković, W. Stephan, S. Hutter, Balancing selection on immunity genes: review of the current literature and new analysis in *Drosophila melanogaster*, *Zoology.* 119 (2016) 322–329. <https://doi.org/10.1016/J.ZOOL.2016.03.004>.
- [9] M. Fumagalli, M. Sironi, U. Pozzoli, A. Ferrer-Admettla, L. Pattini, R. Nielsen, Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution, *PLoS Genet.* 7 (2011) e1002355. <https://doi.org/10.1371/JOURNAL.PGEN.1002355>.
- [10] H. Quach, M. Rotival, J. Pothlichet, Y.H.E. Loh, M. Dannemann, N. Zidane, G. Laval, E. Patin, C. Harmant, M. Lopez, M. Deschamps, N. Naffakh, D. Duffy, A. Coen, G. Leroux-Roels, F. Clément, A. Boland, J.F. Deleuze, J. Kelso, M.L. Albert, L. Quintana-Murci, Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations, *Cell.* 167 (2016) 643-656.e17. <https://doi.org/10.1016/j.cell.2016.09.024>.
- [11] K. Murphy, M. Walport, P. Travers, *Imunobiologia de Janeway*, 7th ed., Artmed Editora, 2010.
- [12] M. Elvington, M.K. Liszewski, J.P. Atkinson, Evolution of the complement system: from defense of the single cell to guardian of the intravascular space, *Immunol Rev.* 274 (2016) 9. <https://doi.org/10.1111/IMR.12474>.
- [13] M. Nonaka, Evolution of the complement system, *Subcell Biochem.* 80 (2014) 31–43. https://doi.org/10.1007/978-94-017-8881-6_3/COVER.
- [14] M. Taha, E. Nezerwa, H.-J. Nam, The X-ray Crystallographic Structure of Human EAT2 (SH2D1B), *Protein Pept Lett.* 23 (2016) 862–866. <https://doi.org/10.2174/0929866523666160831162239>.

- [15] J.L. Cannons, S.G. Tangye, P.L. Schwartzberg, SLAM Family Receptors and SAP Adaptors in Immunity, <https://doi.org/10.1146/Annurev-Immunol-030409-101302>. 29 (2011) 665–705. <https://doi.org/10.1146/ANNUREV-IMMUNOL-030409-101302>.
- [16] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature* 2021 596:7873. 596 (2021) 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- [17] K. Shah, A. Al-Haidari, J. Sun, J.U. Kazi, T cell receptor (TCR) signaling in health and disease, *Signal Transduction and Targeted Therapy* 2021 6:1. 6 (2021) 1–26. <https://doi.org/10.1038/s41392-021-00823-w>.
- [18] C. Borowski, A. Bendelac, Signaling for NKT cell development: the SAP-FynT connection, *Journal of Experimental Medicine*. 201 (2005) 833–836. <https://doi.org/10.1084/jem.20050339>.
- [19] C.W. Arendt, B. Albrecht, T.J. Soos, D.R. Littman, Protein kinase C- θ : signaling from the center of the T-cell synapse, *Curr Opin Immunol*. 14 (2002) 323–330. [https://doi.org/10.1016/S0952-7915\(02\)00346-1](https://doi.org/10.1016/S0952-7915(02)00346-1).
- [20] S. Paul, G. Lal, The molecular mechanism of natural killer cells function and its importance in cancer immunotherapy, *Front Immunol*. 8 (2017) 290728. <https://doi.org/10.3389/FIMMU.2017.01124/BIBTEX>.
- [21] Y. Chen, D. Lu, A. Churov, R. Fu, Research Progress on NK Cell Receptors and Their Signaling Pathways, *Mediators Inflamm*. 2020 (2020). <https://doi.org/10.1155/2020/6437057>.
- [22] S. Chen, Z. Dong, NK cell recognition of hematopoietic cells by SLAM-SAP families, *Cellular & Molecular Immunology* 2019 16:5. 16 (2019) 452–459. <https://doi.org/10.1038/s41423-019-0222-4>.
- [23] J. Cannons, L. Yu, B. Hill, L. Mijares, D. Dombroski, K. Nichols, A. Antonellis, G. Koretzky, K. Gardner, P. Schwartzberg, SAP Regulates T(H)2 Differentiation and

- PKC- θ -Mediated Activation of NF- κ B1, Immunity. (2004) 693–703.
<https://doi.org/10.1016/j.immuni.2004.09.012>.
- [24] K.E. Nichols, R.A. Marsh, The X-Linked Lymphoproliferative Syndromes Chapter Outline, Stiehm's Immune Deficiencies. (2014) 475–496.
<https://doi.org/10.1016/B978-0-12-405546-9.00022-4>.
- [25] S. Ono, T. Okano, A. Hoshino, M. Yanagimachi, K. Hamamoto, Y. Nakazawa, T. Imamura, M. Onuma, H. Niizuma, Y. Sasahara, H. Tsujimoto, T. Wada, R. Kunisaki, M. Takagi, K. Imai, T. Morio, H. Kanegane, Hematopoietic stem cell transplantation for XIAP deficiency in Japan, J Clin Immunol. 37 (2017) 85–91.
<https://doi.org/10.1007/s10875-016-0348-4>.
- [26] Y. Wang, Y. Wang, W. Lu, L. Tao, Y. Xiao, Y. Zhou, X. He, Y. Zhang, L. Li, Potential pathogenic mechanism of type 1 X-linked lymphoproliferative syndrome caused by a mutation of SH2D1A gene in an infant: A case report, Medicine. 101 (2022) E30951. <https://doi.org/10.1097/MD.00000000000030951>.
- [27] S. Chen, W. Shi, Y. Qian, L. Wang, J. Zhang, S. Li, Y. Chen, C. Chang, H. Fei, L. Zhang, H. Huang, C. Xu, Preimplantation Genetic Testing for a Chinese Family With X-Linked Lymphoproliferative Syndrome Type 1, Front Genet. 11 (2020) 550507. <https://doi.org/10.3389/FGENE.2020.550507>.
- [28] R.G. Lahita, Sex and gender influence on immunity and autoimmunity, Front Immunol. 14 (2023) 1142723. <https://doi.org/10.3389/FIMMU.2023.1142723/BIBTEX>.
- [29] C.S. dos Santos, T. Mendes, A. Antunes, The genes from the pseudoautosomal region 1 (PAR1) of the mammalian sex chromosomes: Synteny, phylogeny and selection, Genomics. 114 (2022) 110419. <https://doi.org/10.1016/J.YGENO.2022.110419>.
- [30] E.W. Sayers, E.E. Bolton, J.R. Brister, K. Canese, J. Chan, D.C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B.W. Trawick, K.D. Pruitt, S.T. Sherry, Database resources of the national center for biotechnology information, Nucleic Acids Res. 50 (2022) D20–D26. <https://doi.org/10.1093/NAR/GKAB1112>.
- [31] L. Ramos, A. Antunes, Decoding sex: Elucidating sex determination and how high-quality genome assemblies are untangling the evolutionary dynamics of sex

- p chromosomes, Genomics. 114 (2022) 110277.
-
- <https://doi.org/10.1016/J.YGENO.2022.110277>
- .
-
- [32] G.S.C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison, BMC Bioinformatics. 6 (2005) 1–11.
<https://doi.org/10.1186/1471-2105-6-31/FIGURES/6>.
- [33] A. Bateman, M.J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-AJee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L.G. Castro, L. Figueira, P. Garmiri, G. Georgiou, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M.C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nospikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A.L. Veuthey, C.H. Wu, C.N. Arighi, L. Arminski, C. Chen, Y. Chen, J.S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D.A. Natale, K. Ross, C.R. Vinayaka, Q. Wang, Y. Wang, L.S. Yeh, J. Zhang, UniProt: the universal protein knowledgebase, Nucleic Acids Res. 45 (2017) D158–D169. <https://doi.org/10.1093/NAR/GKW1099>.
- [34] E.T. Caldart, H. Mata, C. Wageck Canal, A.P. Ravazzolo, Phylogenetic Analysis: Basic Concepts and Its Use as a Tool for Virology and Molecular epidemiology, Acta Sci Vet. 44 (2016) 1392.
- [35] I. Sela, H. Ashkenazy, K. Katoh, T. Pupko, GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters, Nucleic Acids Res. 43 (2015) W7. <https://doi.org/10.1093/NAR/GKV318>.
- [36] K. Katoh, K.I. Kuma, H. Toh, T. Miyata, MAFFT version 5: improvement in accuracy of multiple sequence alignment, Nucleic Acids Res. 33 (2005) 511–518. <https://doi.org/10.1093/NAR/GKI198>.

- [37] M.P. Berger, P.J. Munson, A novel randomized iterative strategy for aligning multiple protein sequences, *Bioinformatics*. 7 (1991) 479–484. <https://doi.org/10.1093/BIOINFORMATICS/7.4.479>.
- [38] O. Penn, E. Privman, H. Ashkenazy, G. Landan, D. Graur, T. Pupko, GUIDANCE: a web server for assessing alignment confidence scores, *Nucleic Acids Res.* 38 (2010) W23–W28. <https://doi.org/10.1093/NAR/GKQ443>.
- [39] O. Penn, E. Privman, G. Landan, D. Graur, T. Pupko, An Alignment Confidence Score Capturing Robustness to Guide Tree Uncertainty, *Mol Biol Evol.* 27 (2010) 1759–1767. <https://doi.org/10.1093/MOLBEV/MSQ066>.
- [40] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets, *Mol Biol Evol.* 33 (2016) 1870–1874. <https://doi.org/10.1093/MOLBEV/MSW054>.
- [41] S.L.K. Pond, D. Posada, M.B. Gravenor, C.H. Woelk, S.D.W. Frost, Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm, *Mol Biol Evol.* 23 (2006) 1891–1901. <https://doi.org/10.1093/MOLBEV/MSL051>.
- [42] S. Weaver, S.D. Shank, S.J. Spielman, M. Li, S. V. Muse, S.L. Kosakovsky Pond, Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes, *Mol Biol Evol.* 35 (2018) 773–777. <https://doi.org/10.1093/MOLBEV/MSX335>.
- [43] H. Philippe, P. Forterre, The rooting of the universal tree of life is not reliable, *J Mol Evol.* 49 (1999) 509–523. <https://doi.org/10.1007/PL00006573/METRICS>.
- [44] X. Xia, DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution, *Mol Biol Evol.* 30 (2013) 1720–1728. <https://doi.org/10.1093/MOLBEV/MST064>.
- [45] B. Hall, M. Barlow, Phylogenetic Analysis as a Tool in Molecular Epidemiology of Infectious Diseases, *Ann Epidemiol.* 16 (2006) 157–169. <https://doi.org/10.1016/j.annepidem.2005.04.010>.
- [46] D. Baum, Reading a phylogenetic tree: The meaning of monophyletic groups, *Nature Education*. 1 (2008). <https://www.nature.com/scitable/topicpage/reading-a-phylogenetic-tree-the-meaning-of-41956/> (accessed January 5, 2023).
- [47] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: An integrated and extendable desktop software

- platform for the organization and analysis of sequence data, *Bioinformatics*. 28 (2012) 1647. <https://doi.org/10.1093/BIOINFORMATICS/BTS199>.
- [48] B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, R. Lanfear, E. Teeling, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era, *Mol Biol Evol.* 37 (2020) 1530–1534. <https://doi.org/10.1093/MOLBEV/MSAA015>.
- [49] B.Q. Minh, M.A.T. Nguyen, A. Von Haeseler, Ultrafast Approximation for Phylogenetic Bootstrap, *Mol Biol Evol.* 30 (2013) 1188–1195. <https://doi.org/10.1093/MOLBEV/MST024>.
- [50] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. Von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, *Nature Methods* 2017 14:6. 14 (2017) 587–589. <https://doi.org/10.1038/nmeth.4285>.
- [51] B.Q. Minh, M.W. Hahn, R. Lanfear, New Methods to Calculate Concordance Factors for Phylogenomic Datasets, *Mol Biol Evol.* 37 (2020) 2727–2733. <https://doi.org/10.1093/MOLBEV/MSAA106>.
- [52] S. Kumar, M. Suleski, J.M. Craig, A.E. Kasprowitz, M. Sanderford, M. Li, G. Stecher, S.B. Hedges, TimeTree 5: An Expanded Resource for Species Divergence Times, *Mol Biol Evol.* 39 (2022). <https://doi.org/10.1093/MOLBEV/MSAC174>.
- [53] M. Emam, J.P. Machado, A. Antunes, Evolutionary genomics of mammalian lung cancer genes reveals signatures of positive selection in APC, RB1 and TP53, *Genomics*. 112 (2020) 4722–4731. <https://doi.org/10.1016/J.YGENO.2020.08.020>.
- [54] Z. Yang, PAML 4: Phylogenetic Analysis by Maximum Likelihood, *Mol Biol Evol.* 24 (2007) 1586–1591. <https://doi.org/10.1093/MOLBEV/MSM088>.
- [55] E. Maldonado, D. Almeida, T. Escalona, I. Khan, V. Vasconcelos, A. Antunes, LMAP: Lightweight Multigene Analyses in PAML, *BMC Bioinformatics*. 17 (2016) 1–11. <https://doi.org/10.1186/S12859-016-1204-5/FIGURES/5>.
- [56] J. Rozas, A. Ferrer-Mata, J.C. Sanchez-DelBarrio, S. Guirao-Rico, P. Librado, S.E. Ramos-Onsins, A. Sanchez-Gracia, DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets, *Mol Biol Evol.* 34 (2017) 3299–3302. <https://doi.org/10.1093/MOLBEV/MSX248>.

- [57] F.O. Aylward, Introduction to calculating dN/dS ratios with codeml, *Protocols.io*. (2018). <https://dx.doi.org/10.17504/protocols.io.qhwdt7e> (accessed September 21, 2023).
- [58] R. Cagliani, D. Forni, M. Biasin, M. Comabella, F.R. Guerini, S. Riva, U. Pozzoli, C. Agliardi, D. Caputo, S. Malhotra, X. Montalban, N. Bresolin, M. Clerici, M. Sironi, Ancient and Recent Selective Pressures Shaped Genetic Diversity at AIM2-Like Nucleic Acid Sensors, *Genome Biol Evol.* 6 (2014) 830–845. <https://doi.org/10.1093/GBE/EVU066>.
- [59] Z. Yang, R. Nielsen, N. Goldman, A.M.K. Pedersen, Codon-substitution models for heterogeneous selection pressure at amino acid sites., *Genetics*. 155 (2000) 431. <https://doi.org/10.1093/GENETICS/155.1.431>.
- [60] Z. Yang, PAML 4: Phylogenetic Analysis by Maximum Likelihood, *Mol Biol Evol.* 24 (2007) 1586–1591. <https://doi.org/10.1093/MOLBEV/MSM088>.
- [61] Z. Yang, W.S.W. Wong, R. Nielsen, Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection, *Mol Biol Evol.* 22 (2005) 1107–1118. <https://doi.org/10.1093/MOLBEV/MSI097>.
- [62] S.L. Kosakovsky Pond, S.D.W. Frost, Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection, *Mol Biol Evol.* 22 (2005) 1208–1222. <https://doi.org/10.1093/MOLBEV/MSI105>.
- [63] B. Murrell, S. Moola, A. Mabona, T. Weighill, D. Sheward, S.L. Kosakovsky Pond, K. Scheffler, FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection, *Mol Biol Evol.* 30 (2013) 1196–1205. <https://doi.org/10.1093/MOLBEV/MST030>.
- [64] T. Mendes, L. Silva, D. Almeida, A. Antunes, Neofunctionalization of the UCP1 mediated the non-shivering thermogenesis in the evolution of small-sized placental mammals, *Genomics*. 112 (2020) 2489–2498. <https://doi.org/10.1016/J.YGENO.2020.01.023>.
- [65] B. Murrell, S. Weaver, M.D. Smith, J.O. Wertheim, S. Murrell, A. Aylward, K. Eren, T. Pollner, D.P. Martin, D.M. Smith, K. Scheffler, S.L. Kosakovsky Pond, Gene-Wide Identification of Episodic Selection, *Mol Biol Evol.* 32 (2015) 1365–1371. <https://doi.org/10.1093/MOLBEV/MSV035>.

- [66] J.O. Wertheim, B. Murrell, M.D. Smith, S.L.K. Pond, K. Scheffler, RELAX: Detecting Relaxed Selection in a Phylogenetic Framework, *Mol Biol Evol.* 32 (2015) 820–832. <https://doi.org/10.1093/MOLBEV/MSU400>.
- [67] R. Ramani, K. Krumholz, Y.F. Huang, A. Siepel, PhastWeb: a web interface for evolutionary conservation scoring of multiple sequence alignments using phastCons and phyloP, *Bioinformatics.* 35 (2019) 2320–2322. <https://doi.org/10.1093/BIOINFORMATICS/BTY966>.
- [68] C. Duran, D. Edwards, J. Batley, Genetic maps and the use of synteny, *Methods in Molecular Biology.* 513 (2009) 41–55. https://doi.org/10.1007/978-1-59745-427-8_3/COVER.
- [69] T. Zhao, M.E. Schranz, Comparative Phylogenomic Synteny Network Analysis of Mammalian and Angiosperm Genomes, *BioRxiv.* (2018) 246736. <https://doi.org/10.1101/246736>.
- [70] N.T.T. Nguyen, P. Vincens, H.R. Crollius, A. Louis, Genomicus 2018: karyotype evolutionary trees and on-the-fly synteny computing, *Nucleic Acids Res.* 46 (2018) D816–D822. <https://doi.org/10.1093/NAR/GKX1003>.
- [71] T. Hackl, M. Ankenbrand, B. van Adrichem, gggenomes: A Grammar of Graphics for Comparative Genomics., (2023). <https://github.com/thackl/gggenomes> (accessed September 25, 2023).
- [72] M. Gouy, F. Milleret, C. Mugnier, M. Jacobzone, C. Gautier, ACNUC: a nucleic acid sequence data base and analysis system, *Nucleic Acids Res.* 12 (1984) 121–127. <https://doi.org/10.1093/NAR/12.1PART1.121>.
- [73] D.T. Hoang, O. Chernomor, A. Von Haeseler, B.Q. Minh, L.S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation, *Mol Biol Evol.* 35 (2018) 518–522. <https://doi.org/10.1093/MOLBEV/MSX281>.
- [74] J.A. Esselstyn, C.H. Oliveros, M.T. Swanson, B.C. Faircloth, Investigating Difficult Nodes in the Placental Mammal Tree with Expanded Taxon Sampling and Thousands of Ultraconserved Elements, *Genome Biol Evol.* 9 (2017) 2308–2321. <https://doi.org/10.1093/GBE/EVX168>.
- [75] R. Del Amparo, M. Arenas, Influence of substitution model selection on protein phylogenetic tree reconstruction, *Gene.* 865 (2023) 147336. <https://doi.org/10.1016/J.GENE.2023.147336>.

- [76] Y. Ishii, Topological analysis of vertebrate phylogenetic trees utilizing Horton's first law, *J Theor Biol.* 562 (2023) 111419. <https://doi.org/10.1016/J.JTBI.2023.111419>.
- [77] G.J. Szllosi, E. Tannier, V. Daubin, B. Boussau, The Inference of Gene Trees with Species Trees, *Syst Biol.* 64 (2015) e42–e62. <https://doi.org/10.1093/SYSBIO/SYU048>.
- [78] E. Noutahi, M. Semeria, M. Lafond, J. Seguin, B. Boussau, L. Guéguen, N. El-Mabrouk, E. Tannier, Efficient Gene Tree Correction Guided by Genome Evolution, *PLoS One.* 11 (2016). <https://doi.org/10.1371/JOURNAL.PONE.0159559>.
- [79] A. Veillette, Immune regulation by SLAM family receptors and SAP-related adaptors, *Nat Rev Immunol.* 6 (2006) 56–66. <https://doi.org/10.1038/NRI1761>.
- [80] P. Liston, N. Roy, K. Tamai, C. Lefebvre, S. Baird, G. Cherton-Horvat, R. Farahani, M. McLean, J.E. Ikeda, A. MacKenzie, R.G. Korneluk, Suppression of apoptosis in mammalian cells by NAIP and a related family of IAP genes, *Nature* 1996 379:6563. 379 (1996) 349–353. <https://doi.org/10.1038/379349a0>.
- [81] P.M. Petkov, J.H. Graber, G.A. Churchill, K. Dipetrillo, B.L. King, K. Paigen, Evidence of a Large-Scale Functional Organization of Mammalian Chromosomes, (n.d.). <https://doi.org/10.1371/journal.pgen.0010033>.
- [82] Mank, E. Judith, B. Vicoso, S. Berlin, B. Charlesworth, EFFECTIVE POPULATION SIZE AND THE FASTER-X EFFECT: EMPIRICAL RESULTS AND THEIR INTERPRETATION., *Evolution (N Y).* 64 (2010) 74–663. <https://www.jstor.org/stable/27784037> (accessed September 27, 2023).
- [83] Y.K. Mo, R. Lanfear, M.W. Hahn, B.Q. Minh, Updated site concordance factors minimize effects of homoplasy and taxon sampling, *Bioinformatics.* 39 (2023). <https://doi.org/10.1093/BIOINFORMATICS/BTAC741>.
- [84] M. Morra, J. Lu, F. Poy, M. Martin, J. Sayos, S. Calpe, C. Gullo, D. Howie, S. Rietdijk, A. Thompson, A.J. Coyle, C. Denny, M.B. Yaffe, P. Engel, M.J. Eck, C. Terhorst, Structural basis for the interaction of the free SH2 domain EAT-2 with SLAM receptors in hematopoietic cells, *EMBO J.* 20 (2001) 5840–5852. <https://doi.org/10.1093/EMBOJ/20.21.5840>.

- [85] T.B. Sackton, B.P. Lazzaro, T.A. Schlenke, J.D. Evans, D. Hultmark, A.G. Clark, Dynamic evolution of the innate immune system in *Drosophila*, *Nature Genetics* 2007 39:12. 39 (2007) 1461–1468. <https://doi.org/10.1038/ng.2007.60>.
- [86] A. Rhie, S. Nurk, M. Cechova, S.J. Hoyt, D.J. Taylor, N. Altemose, P.W. Hook, S. Koren, M. Rautiainen, I.A. Alexandrov, J. Allen, M. Asri, A. V. Bzikadze, N.C. Chen, C.S. Chin, M. Diekhans, P. Flicek, G. Formenti, A. Fungtammasan, C. Garcia Giron, E. Garrison, A. Gershman, J.L. Gerton, P.G.S. Grady, A. Guarracino, L. Haggerty, R. Halabian, N.F. Hansen, R. Harris, G.A. Hartley, W.T. Harvey, M. Haukness, J. Heinz, T. Hourlier, R.M. Hubley, S.E. Hunt, S. Hwang, M. Jain, R.K. Kesharwani, A.P. Lewis, H. Li, G.A. Logsdon, J.K. Lucas, W. Makalowski, C. Markovic, F.J. Martin, A.M. Mc Cartney, R.C. McCoy, J. McDaniel, B.M. McNulty, P. Medvedev, A. Mikheenko, K.M. Munson, T.D. Murphy, H.E. Olsen, N.D. Olson, L.F. Paulin, D. Porubsky, T. Potapova, F. Ryabov, S.L. Salzberg, M.E.G. Sauria, F.J. Sedlazeck, K. Shafin, V.A. Shepelev, A. Shumate, J.M. Storer, L. Surapaneni, A.M. Taravella Oill, F. Thibaud-Nissen, W. Timp, M. Tomaszewicz, M.R. Vollger, B.P. Walenz, A.C. Watwood, M.H. Weissensteiner, A.M. Wenger, M.A. Wilson, S. Zarate, Y. Zhu, J.M. Zook, E.E. Eichler, R.J. O'Neill, M.C. Schatz, K.H. Miga, K.D. Makova, A.M. Phillippy, The complete sequence of a human Y chromosome, *Nature* 2023 621:7978. 621 (2023) 344–354. <https://doi.org/10.1038/s41586-023-06457-y>.
- [87] C. Bleidorn, Third generation sequencing: Technology and its potential impact on evolutionary biodiversity research, *Syst Biodivers.* 14 (2016) 1–8. <https://doi.org/10.1080/14772000.2015.1099575>.
- [88] E.E. Schadt, S. Turner, A. Kasarskis, A window into third-generation sequencing, *Hum Mol Genet.* 19 (2010). <https://doi.org/10.1093/HMG/DDQ416>.
- [89] M.R. Vollger, G.A. Logsdon, P.A. Audano, A. Sulovari, D. Porubsky, P. Peluso, A.M. Wenger, G.T. Concepcion, Z.N. Kronenberg, K.M. Munson, C. Baker, A.D. Sanders, D.C.J. Spierings, P.M. Lansdorp, U. Surti, M.W. Hunkapiller, E.E. Eichler, Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads, *Ann Hum Genet.* 84 (2020) 125–140. <https://doi.org/10.1111/AHG.12364>.
- [90] K. Athanasopoulou, M.A. Boti, P.G. Adamopoulos, P.C. Skourou, A. Scorilas, Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics, *Life (Basel).* 12 (2021). <https://doi.org/10.3390/LIFE12010030>.

- [91] R.J. McKinlay Gardner, D.J. Amor, Sex Chromosome Translocations, Case Studies in Clinical Psychological Science: Bridging the Gap from Science to Practice. (2018) 1–7. <https://doi.org/10.1093/MED/9780199329007.003.0006>.
- [92] C. Cheng, M. Kirkpatrick, Inversions are bigger on the X chromosome, Mol Ecol. 28 (2019) 1238. <https://doi.org/10.1111/MEC.14819>.
- [93] R.J.M. Gardner, G.R. Sutherland, L.G. Shaffer, Sex Chromosome Translocations, Chromosome Abnormalities and Genetic Counseling. (2011) 112–139. <https://doi.org/10.1093/MED/9780195375336.003.0006>.

Appendix

Table S1: List of primate species chosen by family for the study and respective assemblies. Genomes taken from the NCBI genome database [30].

Primates				
Family	Species	Assembly level	Link	GenBank Assembly name
Hominidae	Human (<i>Homo sapiens</i>) GRCh38.p14	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001405.40/	GCA_000001405.29
	Chimpanzees (<i>Pan troglodytes</i>) Clint_PTRv2	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_002880755.1	GCA_002880755.3
	Chimpanzees (<i>Pan paniscus</i>) Mhudiblu_PPA_v2	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_013052645.3/	GCA_013052645.3
	Gorilla (<i>Gorilla gorilla</i>) Kamilah_GGO_v0	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_008122165.1/	GCA_008122165.1
	Orangutan (<i>Pongo abelli</i>) Susie_PABv2	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_002880775.1/	GCA_002880775.3
	Bornean orangutan (<i>Pongo pygmeus</i>) ASM2376777v1	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023767775.1/	GCA_023767775.1
Cercopithecidae	Rhesus monkey (<i>Macaca mulatta</i>) Mmul_10	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_003339765.1/	GCA_003339765.3
	Gelada (<i>Theropithecus gelada</i>) Tgel_1.0	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_003255815.1/	GCA_003255815.1
	Green monkey (<i>Chlorocebus sabaeus</i>) Chlorocebus_sabeus 1.1	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000409795.2/	GCA_000409795.2
	Mantled guereza (<i>Colobus guereza</i>) ASM2149845v1	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_021498455.1/	GCA_021498455.1
	Proboscis monkey (<i>Nasalis larvatus</i>) Charlie1.0	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_000772465.1/	GCA_000772465.1
	Golden snub-nosed monkey (<i>Rhinopithecus roxellana</i>) ASM756505v1	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_007565055.1/	GCA_007565055.1
Cebidae	White-tufted-ear marmoset (<i>Callithrix jacchus</i>) Callithrix_jacchus_cj1700_1.1	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_009663435.1/	GCA_009663435.2
	Tufted capuchin (<i>Sapajus apella</i>) GSC_monkey_1.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_009761245.1/	GCA_009761245.1
	White-fronted capuchin (<i>Cebus albifrons</i>) CebAlb_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004027755.1/	GCA_004027755.1
	Panamanian white-faced capuchin (<i>Cebus imitator</i>) Cebus_imitator-1.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_001604975.1/	GCA_001604975.1
	Bolivian squirrel monkey (<i>Saimiri boliviensis</i>) BCM_Sbol_2.1	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_016699345.2/	GCA_016699345.2
	Tamarin (<i>Saguinus imperator</i>) Saglmp_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004024885.1/	GCA_004024885.1

Continuation of Table S1

Primates				
Family	Species	Assembly level	Link	GenBank Assembly name
Cheirogaleidae)	<i>Microcebus tavaratra</i> Mtav_1.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_008750935.1/	GCA_008750935.1
	Gray mouse lemur (<i>Microcebus murinus</i>) Mmur_3.0	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000165445.2/	GCA_000165445.3
	<i>Microcebus ravelobensis</i> Mrav_1.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_008750975.1/	GCA_008750975.1
	Mittermeier's mouse lemur (<i>Microcebus mittermeieri</i>) Mmit_1.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_008750955.1/	GCA_008750955.1
	Lesser dwarf lemur (<i>Cheirogaleus medius</i>) CheMed_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004024725.1/	GCA_004024725.1
	Coquerel's mouse lemur (<i>Mirza coquereli</i>) MizCoq_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004024645.1/	GCA_004024645.1
Lemuridae	Greater bamboo lemur (<i>Prolemur simus</i>) Prosim_1.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_003258685.1/	GCA_003258685.1
	Black lemur (<i>Eulemur macaco</i>) Emacaco_refEf_BWA_oneround	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_001262655.1/	GCA_001262655.1
	Sclater's lemur (<i>Eulemur flavifrons</i>) Eflavifronsk33QCA	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_001262665.1/	GCA_001262665.1
	Brown lemur (<i>Eulemur fulvus</i>) EulFul_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004027275.1/	GCA_004027275.1
	Ring-tailed lemur (<i>Lemur catta</i>) mLemCat1.pri	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_020740605.2/	GCA_020740605.1
Hylobatidae	Siamang (<i>Symphalangus syndactylus</i>) ASM2376113v1	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023761135.1/	GCA_023761135.1
	Pileated gibbon (<i>Hylobates pileatus</i>) ASM2149846v1	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_021498465.1/	GCA_021498465.1
	Silvery gibbon (<i>Hylobates moloch</i>) HMol_V2	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_009828535.3/	GCA_009828535.3
	Southern white-cheeked gibbon (<i>Nomascus siki</i>) ASM2378306v1	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023783065.1/	GCA_023783065.1
	Northern white-cheeked gibbon (<i>Nomascus leucogenys</i>) Asia_NLE_v1	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_006542625.1/	GCA_006542625.1
	Eastern hoolock gibbon (<i>Hoolock leuconedys</i>) ASM2374817v1	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023748175.1/	GCA_023748175.1

Continuation of Table S1

Primates				
Family	Species	Assembly level	Link	GenBank Assembly name
Lorisidae	Slender loris (<i>Loris tardigradus</i>) ASM2378313v1	Contig	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023783135.1/	GCA_023783135.1
	Slow loris (<i>Nycticebus coucang</i>) mNycCou1.pri	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_027406575.1/	GCA_027406575.1
	Bengal slow loris (<i>Nycticebus bengalensis</i>) ASM2389825v1	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023898255.1/	GCA_023898255.1
Atelidae	Mantled howler monkey (<i>Alouatta palliata</i>) AloPal_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004027835.1/	GCA_004027835.1
	Black-handed spider monkey (<i>Ateles geoffroyi</i>) AteGeo_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004024785.1/	GCA_004024785.1
	Brown spider monkey (<i>Ateles hybridus</i>) ORGONE_01	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_916098195.1/	GCA_916098195.1
Daubentonidae	Aye-aye (<i>Daubentonia madagascariensis</i>) DauMad_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004027145.1/	GCA_004027145.1
Pitheciidae	White-faced saki (<i>Pithecia pithecia</i>) PitPit_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004026645.1/	GCA_004026645.1
	Bolivian titi (<i>Plecturocebus donacophilus</i>) CalDon_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004027715.1/	GCA_004027715.1
Tarsiidae	Philippine tarsier (<i>Carlito syrichta</i>) Tarsius_syrichta-2.0.1	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000164805.1/	GCA_000164805.2
	Horsfield's tarsier (<i>Cephalopachus bancanus</i>) ASM2725705v1	Contig	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_027257055.1/	GCA_027257055.1
Indridae	Coquerel's sifaka (<i>Propithecus coquereli</i>) Pcoq_1.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000956105.1/	GCA_000956105.1
	Indri indri IndInd_v1_BIUU	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004363605.1/	GCA_004363605.1
Galagidae	Small-eared galago (<i>Otolemur garnettii</i>) OtoGar3	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000181295.1/	GCA_000181295.3
	Moholi bushbaby (<i>Galago moholi</i>) ASM2378343v1	Contig	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023783435.1/	GCA_023783435.1
Aotidae	Ma's night monkey (<i>Aotus nancymae</i>) Anan_2.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000952055.2/	GCA_000952055.2

Table S2: List of mammalian species chosen by order for the study and respective assemblies. Genomes taken from the NCBI genome database [30].

Mammalian				
Order	Species	Assembly level	Link	GenBank Assembly name
Afrosoricida	Small madagascar hedgehog (<i>Echinops telfairi</i>) ASM31398v2	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000313985.2/	GCA_000313985.2
Monotremata	Platypus (<i>Ornithorhynchus anatinus</i>) mOrnAna1.pri.v4	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_004115215.2/	GCA_004115215.4
Carnivora	Dog (<i>Canis lupus familiaris</i>) ROS_Cfam_1.0	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_014441545.1/	GCA_014441545.1
Dasyuromorphia	Yellow-footed antechinus (<i>Antechinus flavipes</i>) AdamAnt_v2	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_016432865.1/	GCA_016432865.2
Macroscelidea	Long-eared elephant shrews (<i>Elephantulus edwardii</i>) EleEdw1.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000299155.1/	GCA_000299155.1
Hyracoidea	Cape rock hyrax (<i>Procavia capensis</i>) ProCapCap_v3_BIUU_UCD	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_004026925.3/	GCA_004026925.3
Didelphimorphia	Agile gracile mouse opossum (<i>Gracilinanus agilis</i>) AgileGrace	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_016433145.1/	GCA_016433145.1
Sirenia	Dugong (<i>Dugong dugon</i>) Dugong	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_905400935.1/	GCA_905400935.1
Scandentia	Large tree shrew (<i>Tupaia tana</i>) mTupTan1	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_026018925.1/	GCA_026018925.1
Eulipotyphla	White-toothed pygmy shrew (<i>Suncus etruscus</i>) mSunEtr1.pri.cur	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_024139225.1/	GCA_024139225.1
Tubulidentata	Aardvark (<i>Orycteropus afer</i>) OryAfe1.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000298275.1/	GCA_000298275.1
Chiroptera	Pallid bat (<i>Antrozous pallidus</i>) DD_ASM_mAP_20210503	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_027563665.1/	GCA_027563665.1
Diprotodontia	Woylie (<i>Bettongia penicillata ogilbyi</i>) mBetpen1.pri.20210916	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023548195.1/	GCA_023548195.1
Perissodactyla	Donkey (<i>Equus asinus</i>) ASM1607732v2	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_016077325.2/	GCA_016077325.2
Artiodactyla	Pygmy sperm whale (<i>Kogia breviceps</i>) mKogBre1 haplotype 1	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_026419965.1/	GCA_026419965.1
Pholidota	Malayan pangolin (<i>Manis javanica</i>) YNU_ManJav_2.0	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_014570535.1/	GCA_014570535.1
Cingulata	Tolypeutes matacus mTolMat1	Scaffold	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_026826555.1/	GCA_026826555.1

Continuation of Table S2

Mammalian				
Order	Species	Assembly level	Link	GenBank Assembly name
Proboscidea	Asiatic elephant (<i>Elephas maximus indicus</i>) mEleMax1 primary haplotype	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_024166365.1/	GCA_024166365.1
Rodentia	European woodmouse (<i>Apodemus sylvaticus</i>) mApoSyl1.1	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_947179515.1/	GCA_947179515.1
Microbiotheria	Monito del monte (<i>Dromiciops gliroides</i>) mDroGli1.pri	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_019393635.1/	GCA_019393635.1
Dermoptera	Philippine flying lemur (<i>Cynocephalus volans</i>) mCynVol1.pri	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_027409185.1/	GCA_027409185.1
Pilosa	Southern tamandua (<i>Tamandua tetradactyla</i>) mTamTet1.pri	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023851605.1/	GCA_023851605.1
Lagomorpha	Rabbit (<i>Oryctolagus cuniculus</i>) UM_NZW_1.0	Chromosome	https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_009806435.1/	GCA_009806435.2

Table S3: List of species by dataset for each gene. Green blocks mean the presence of the species in the dataset. Orange blocks mean that the species is the outgroup of the dataset.

Species	<i>SH2D1A</i> gene			<i>SH2D1B</i> gene			<i>XIAP</i> gene		
	GA	P	M	GA	P	M	GA	P	M
Small madagascar hedgehog (<i>Echinops telfairi</i>) ASM31398v2									
Agile gracile mouse opossum (<i>Gracilinanus agilis</i>) AgileGrace									
Platypus (<i>Ornithorhynchus anatinus</i>) mOrnAna1.pri.v4									
Dog (<i>Canis lupus familiaris</i>) ROS_Cfam_1.0									
Yellow-footed antechinus (<i>Antechinus flavipes</i>) AdamAnt_v2									
Cape rock hyrax (<i>Procapia capensis</i>) ProCapCap_v3_BIUU_UCD									
Dugong (<i>Dugong dugon</i>) Dugong									
Large tree shrew (<i>Tupaia tana</i>) mTupTan1									
White-toothed pygmy shrew (<i>Suncus etruscus</i>) mSunEtr1.pri.cur									
Aardvark (<i>Orycteropus afer afer</i>) OryAfe1.0									
Long-eared elephant shrews (<i>Elephantulus edwardii</i>) EleEdw1.0									

Continuation of Table S3

Species	SH2D1A gene			SH2D1B gene			XIAP gene		
	GA	P	M	GA	P	M	GA	P	M
Pallid bat (<i>Antrozous pallidus</i>) DD_ASM_mAP_20210503									
Woylie (<i>Bettongia penicillata ogilbyi</i>) mBetpen1.pri.20210916									
Donkey (<i>Equus asinus</i>) ASM1607732v2									
Pygmy sperm whale (<i>Kogia breviceps</i>) mKogBre1 haplotype 1									
Malayan pangolin (<i>Manis javanica</i>) YNU_ManJav_2.0									
Tolypeutes matacus mTolMat1									
Asiatic elephant (<i>Elephas maximus indicus</i>) mEleMax1 primary haplotype									
European woodmouse (<i>Apodemus sylvaticus</i>) mApoSyl1.1									
Monito del monte (<i>Dromiciops gliroides</i>) mDroGli1.pri									
Philippine flying lemur (<i>Cynocephalus volans</i>) mCynVol1.pri									
Southern tamandua (<i>Tamandua tetradactyla</i>) mTamTet1.pri									
Rabbit (<i>Oryctolagus cuniculus</i>) UM_NZW_1.0									
Rhesus monkey (<i>Macaca mulatta</i>) Mmul_10									
Brown spider monkey (<i>Ateles hybridus</i>) ORGONE_01									
Microcebus tavaratra Mtav_1.0									
Gray mouse lemur (<i>Microcebus murinus</i>) Mmur_3.0									
White-fronted capuchin (<i>Cebus albifrons</i>) CebAlb_v1_BIUU									
Gelada (<i>Theropithecus gelada</i>) Tgel_1.0									
Green monkey (<i>Chlorocebus sabaeus</i>) Chlorocebus_sabeus 1.1									
Mantled guereza (<i>Colobus guereza</i>) ASM2149845v1									
Proboscis monkey (<i>Nasalis larvatus</i>) Charlie1.0									
Golden snub-nosed monkey (<i>Rhinopithecus roxellana</i>) ASM756505v1									
White-tufted-ear marmoset (<i>Callithrix jacchus</i>) Callithrix_jacchus_cj1700_1.1									
Tufted capuchin (<i>Sapajus apella</i>) GSC_monkey_1.0									

Continuation of Table S3

Species	<i>SH2D1A</i> gene			<i>SH2D1B</i> gene			<i>XIAP</i> gene		
	GA	P	M	GA	P	M	GA	P	M
Panamanian white-faced capuchin (<i>Cebus imitator</i>)									
Cebus_imitator-1.0									
Bolivian squirrel monkey (<i>Saimiri boliviensis</i>)									
BCM_Sbol_2.1									
Tamarin (<i>Saguinus imperator</i>)									
SagImp_v1_BIUU									
<i>Microcebus ravelobensis</i>									
Mrav_1.0									
Mittermeier's mouse lemur (<i>Microcebus mittermeieri</i>)									
Mmit_1.0									
Lesser dwarf lemur (<i>Cheirogaleus medius</i>)									
CheMed_v1_BIUU									
Coquerel's mouse lemur (<i>Mirza coquereli</i>)									
MizCoq_v1_BIUU									
Greater bamboo lemur (<i>Prolemur simus</i>)									
Prosim_1.0									
Black lemur (<i>Eulemur macaco</i>)									
Emacaco_refEf_BWA_oneround									
Sclater's lemur (<i>Eulemur flavifrons</i>)									
Eflavifronsk33QCA									
Brown lemur (<i>Eulemur fulvus</i>)									
EulFul_v1_BIUU									
Ring-tailed lemur (<i>Lemur catta</i>)									
mLemCat1.pri									
Siamang (<i>Symphalangus syndactylus</i>)									
ASM2376113v1									
Pileated gibbon (<i>Hylobates pileatus</i>)									
ASM2149846v1									
Silvery gibbon (<i>Hylobates moloch</i>)									
HMol_V2									
Southern white-cheeked gibbon (<i>Nomascus siki</i>)									
ASM2378306v1									
Northern white-cheeked gibbon (<i>Nomascus leucogenys</i>)									
Asia_NLE_v1									
Eastern hoolock gibbon (<i>Hoolock leuconedys</i>)									
ASM2374817v1									
Slender loris (<i>Loris tardigradus</i>)									
ASM2378313v1									
Slow loris (<i>Nycticebus coucang</i>)									
mNycCou1.pri									
Bengal slow loris (<i>Nycticebus bengalensis</i>)									
ASM2389825v1									
Mantled howler monkey (<i>Alouatta palliata</i>)									
AloPal_v1_BIUU									
Black-handed spider monkey (<i>Ateles geoffroyi</i>)									
AteGeo_v1_BIUU									

Continuation of Table S3

Species	SH2D1A gene			SH2D1B gene			XIAP gene		
	GA	P	M	GA	P	M	GA	P	M
Aye-aye (<i>Daubentonia madagascariensis</i>) DauMad_v1_BIUU									
White-faced saki (<i>Pithecia pithecia</i>) PitPit_v1_BIUU									
Bolivian titi (<i>Plecturocebus donacophilus</i>) CalDon_v1_BIUU									
Philippine tarsier (<i>Carlito syrichta</i>) Tarsius_syrichta-2.0.1									
Horsfield's tarsier (<i>Cephalopachus bancanus</i>) ASM2725705v1									
Coquerel's sifaka (<i>Propithecus coquereli</i>) Pcoq_1.0									
Indri indri IndInd_v1_BIUU									
Small-eared galago (<i>Otolemur garnettii</i>) OtoGar3									
Moholi bushbaby (<i>Galago moholi</i>) ASM2378343v1									
Ma's night monkey (<i>Aotus nancymae</i>) Anan_2.0									
Human (<i>Homo sapiens</i>) GRCh38.p14									
Chimpanzees (<i>Pan troglodytes</i>) Clint_PTRv2									
Chimpanzees (<i>Pan paniscus</i>) Mhudiblu_PPA_v2									
Gorilla (<i>Gorilla gorilla</i>) Kamilah_GGO_v0									
Orangutan (<i>Pongo abelli</i>) Susie_PABv2									
Bornean orangutan (<i>Pongo pygmeus</i>) ASM2376777v1									
Ma's night monkey (<i>Aotus nancymae</i>) Anan_2.0									
Human (<i>Homo sapiens</i>) GRCh38.p14									
Chimpanzees (<i>Pan troglodytes</i>) Clint_PTRv2									
Chimpanzees (<i>Pan paniscus</i>) Mhudiblu_PPA_v2									
Gorilla (<i>Gorilla gorilla</i>) Kamilah_GGO_v0									
Orangutan (<i>Pongo abelli</i>) Susie_PABv2									
Bornean orangutan (<i>Pongo pygmeus</i>) ASM2376777v1									

GA – Great Apes dataset | P – Primates dataset | M – Mammalia dataset

Table S4: *Homo sapiens* protein sequences from the UniProt database used as queries for synteny analyses.

Species	Genes							
	<i>SH2D1A</i>	<i>SH2D1B</i>	<i>XIAP</i>	<i>THOC2</i>	<i>STAG2</i>	<i>TXT13D</i>	<i>UHMK1</i>	<i>C1orf226</i>
<i>Homo sapiens</i>	O60880	O14796	P98170	Q8NI27	Q8N3U4	A0A0J9YY54	Q8TAS1	A1L170

Table S5: Saturation signature values resulting from Xia's test for each gene in each dataset.

Dataset	Gene	P (inv)	Iss	Iss c	DF	P-value
Great apes	<i>SH2D1A</i>	0,002	0,008	0,754	382,0	0,000
	<i>SH2D1B</i>	0,002	0,015	0,755	394,0	0,000
	<i>XIAP</i>	0,001	0,005	0,811	1489	0,000
Dataset	Gene	P (inv)	Iss	Iss c	DF	P-value
Primates	<i>SH2D1A</i>	0,325	0,095	0,787	253,0	0,000
	<i>SH2D1B</i>	0,245	0,171	0,786	281,0	0,000
	<i>XIAP</i>	0,353	0,083	0,833	949,0	0,000
Dataset	Gene	P (inv)	Iss	Iss c	DF	P-value
Mammalia	<i>SH2D1A</i>	0,174	0,239	0,684	306,0	0,000
	<i>SH2D1B</i>	0,133	0,347	0,678	303,0	0,000
	<i>XIAP</i>	0,123	0,202	0,780	1272	0,000

P (inv) - Proposition of invariant sites; **Iss** - Index of substitution saturation; **Iss c** - Critical index of substitution saturation and **DF** - Degrees of freedom

Table S6: Statistical hypothesis testing (SHT) of the branch-site model analysis of great apes dataset, performed by the PAML program package.

Genes	Statistical hypothesis testing (SHT)																					
	k _i	Model code	np	(SC0) fg ω	(SC0) bg ω	(SC0) p	(SC1) fg ω	(SC1) bg ω	(SC1) p	(SC2a) fg ω	(SC2a) bg ω	(SC2a) p	(SC2b) fg ω	(SC2b) bg ω	(SC2b) p	Tree length	k	LnI	Model vs Model	LRT	DF	P-Value
SH2D1A	0,2	Alt	14	0,07399	0,07399	1,00000	1,00000	1,00000	0,00000	3,09638	0,07399	0,00000	3,09638	1,00000	0,00000	0,04772	3,37408	-560,068167	Alt vs Null	0,000	1	1,000
		Null	13	0,07399	0,07399	1,00000	1,00000	1,00000	0,00000	1,00000	0,07399	0,00000	1,00000	1,00000	0,00000	0,04772	3,37408	-560,068167				
	2	Alt	14	0,07399	0,07399	0,99999	1,00000	1,00000	0,00001	2,74702	0,07399	0,00001	2,74702	1,00000	0,00000	0,04772	3,37410	-560,068167	Alt vs Null	0,000	1	1,000
		Null	13	0,07399	0,07399	0,99999	1,00000	1,00000	0,00000	1,00000	0,07399	0,00001	1,00000	1,00000	0,00000	0,04772	3,37406	-560,068167				
	5	Alt	14	0,07399	0,07399	1,00000	1,00000	1,00000	0,00000	2,75002	0,07399	0,00000	2,75002	1,00000	0,00000	0,04772	3,37408	-560,068167	Alt vs Null	0,000	1	1,000
		Null	13	0,07399	0,07399	0,99999	1,00000	1,00000	0,00000	1,00000	0,07399	0,00001	1,00000	1,00000	0,00000	0,04772	3,37408	-560,068167				
SH2D1B	0,2	Alt	14	0,93103	0,93103	0,99781	1,00000	1,00000	0,00219	1,00000	0,93103	0,00000	1,00000	1,00000	0,00000	0,08344	18,6837	-604,602077	Alt vs Null	0,000	1	1,000
		Null	13	0,93100	0,93100	0,99729	1,00000	1,00000	0,00270	1,00000	0,93100	0,00001	1,00000	1,00000	0,00000	0,08344	18,6838	-604,602078				
	2	Alt	14	0,93117	0,93117	1,00000	1,00000	1,00000	0,00000	1,00000	0,93117	0,00000	1,00000	1,00000	0,00000	0,08344	18,6836	-604,602076	Alt vs Null	0,000	1	1,000
		Null	13	0,93116	0,93116	1,00000	1,00000	1,00000	0,00000	1,00000	0,93116	0,00000	1,00000	1,00000	0,00000	0,08344	18,6835	-604,602076				
	5	Alt	14	0,93116	0,93116	1,00000	1,00000	1,00000	0,00000	1,00000	0,93116	0,00000	1,00000	1,00000	0,00000	0,08344	18,6833	-604,602076	Alt vs Null	0,000	1	1,000
		Null	13	0,93116	0,93116	1,00000	1,00000	1,00000	0,00000	1,00000	0,93116	0,00000	1,00000	1,00000	0,00000	0,08344	18,6837	-604,602076				
XIAP	0,2	Alt	14	0,19002	0,19002	0,00000	1,00000	1,00000	0,00000	79,8854	0,19002	1,00000	79,8854	1,00000	0,00000	0,02696	2,25048	-2082,76278	Alt vs Null	0,465	1	0,495
		Null	13	0,18913	0,18913	0,00000	1,00000	1,00000	0,00000	1,00000	0,18913	0,99998	1,00000	1,00000	0,00002	0,02699	2,20963	-2082,99527				
	2	Alt	14	0,18999	0,18999	0,00000	1,00000	1,00000	0,00000	73,7108	0,18999	1,00000	73,7108	1,00000	0,00000	0,02696	2,24978	-2082,76306	Alt vs Null	0,464	1	0,496
		Null	13	0,18920	0,18920	0,00000	1,00000	1,00000	0,00000	1,00000	0,18920	1,00000	1,00000	1,00000	0,00000	0,02699	2,21080	-2082,99527				
	5	Alt	14	0,19000	0,19000	0,00000	1,00000	1,00000	0,00001	999,000	0,19000	1,00000	999,000	1,00000	0,00000	0,02696	2,25049	-2082,75971	Alt vs Null	0,471	1	0,493
		Null	13	0,18920	0,18920	0,00000	1,00000	1,00000	0,00000	1,00000	0,18920	1,00000	1,00000	1,00000	0,00000	0,02699	2,21079	-2082,99527				

k_i - Inicial kappa value; **k** - Kappa; **LRT** - Likelihood-ratio test; **np** - Number of parameters; **DF** - Degrees of freedom; **SC** - Site classes; **fg** - Foreground; **bg** - Background; **p** - Proportion and ω - Ratio of non-synonymous/synonymous substitutions ratio

Table S7: Statistical hypothesis testing (SHT) of the branch-site model analysis of Mammalia dataset, performed by the PAML program package.

Genes	Statistical hypothesis testing (SHT)																					
	k _i	Model code	np	(SC0) fg ω	(SC0) bg ω	(SC0) p	(SC1) fg ω	(SC1) bg ω	(SC1) p	(SC2a) fg ω	(SC2a) bg ω	(SC2a) p	(SC2b) fg ω	(SC2b) bg ω	(SC2b) p	Tree length	k	LnI	Model vs Model	LRT	DF	P-Value
SH2D1A (Apical branches)	0,2	Alt	59	0,04356	0,04356	0,84696	1,00000	1,00000	0,15304	1,00000	0,04356	0,00000	1,00000	1,00000	0,00000	3,42258	3,07180	-2226,227423	Alt vs Null	0,000	1	1,000
		Null	58	0,04356	0,04356	0,84695	1,00000	1,00000	0,15305	1,00000	0,04356	0,00000	1,00000	1,00000	0,00000	3,42258	3,07180	-2226,227423				
	2	Alt	59	0,04356	0,04356	0,84696	1,00000	1,00000	0,15304	1,00000	0,04356	0,00000	1,00000	1,00000	0,00000	3,42258	3,07180	-2226,227423	Alt vs Null	0,000	1	1,000
		Null	58	0,04356	0,04356	0,84696	1,00000	1,00000	0,15304	1,00000	0,04356	0,00000	1,00000	1,00000	0,00000	3,42258	3,07180	-2226,227423				
	5	Alt	59	0,04356	0,04356	0,84696	1,00000	1,00000	0,15304	1,00000	0,04356	0,00000	1,00000	1,00000	0,00000	3,42258	3,07180	-2226,227423	Alt vs Null	0,000	1	1,000
		Null	58	0,04356	0,04356	0,84696	1,00000	1,00000	0,15304	1,00000	0,04356	0,00000	1,00000	1,00000	0,00000	3,42258	3,07180	-2226,227423				
SH2D1A (Ancestral branch)	0,2	Alt	59	0,04297	0,04297	0,00000	1,00000	1,00000	0,00000	79,9606	0,04297	0,85043	79,9606	1,00000	0,14957	3,42298	3,07267	-2224,329092	Alt vs Null	0,489	1	0,484
		Null	58	0,04297	0,04297	0,00000	1,00000	1,00000	0,00000	1,00000	0,04297	0,85041	1,00000	1,00000	0,14959	3,42307	3,07117	-2224,573722				
	2	Alt	59	0,04297	0,04297	0,00000	1,00000	1,00000	0,00000	999,000	0,04297	0,85043	999,000	1,00000	0,14957	3,42298	3,07273	-2224,325853	Alt vs Null	0,496	1	0,481
		Null	58	0,04297	0,04297	0,00000	1,00000	1,00000	0,00000	1,00000	0,04297	0,85041	1,00000	1,00000	0,14959	3,42308	3,07117	-2224,573722				
	5	Alt	59	0,04297	0,04297	0,00000	1,00000	1,00000	0,00000	49,4283	0,04297	0,85043	49,4283	1,00000	0,14957	3,42298	3,07269	-2224,331262	Alt vs Null	0,485	1	0,484
		Null	58	0,04297	0,04297	0,00000	1,00000	1,00000	0,00000	1,00000	0,04297	0,85041	1,00000	1,00000	0,14959	3,42307	3,07117	-2224,573722				
SH2D1B (Apical branches)	0,2	Alt	57	0,08883	0,08883	0,56881	1,00000	1,00000	0,41441	1,00000	0,08883	0,00971	1,00000	1,00000	0,00707	9,02255	2,93668	-3805,843727	Alt vs Null	0,000	1	1,000
		Null	56	0,08883	0,08883	0,56881	1,00000	1,00000	0,41441	1,00000	0,08883	0,00971	1,00000	1,00000	0,00707	9,02255	2,93668	-3805,843727				
	2	Alt	57	0,08883	0,08883	0,56881	1,00000	1,00000	0,41441	1,00000	0,08883	0,00971	1,00000	1,00000	0,00707	9,02255	2,93668	-3805,843727	Alt vs Null	0,000	1	1,000
		Null	56	0,08883	0,08883	0,56881	1,00000	1,00000	0,41441	1,00000	0,08883	0,00971	1,00000	1,00000	0,00707	9,02254	2,93668	-3805,843727				
	5	Alt	57	0,08883	0,08883	0,56881	1,00000	1,00000	0,41441	1,00000	0,08883	0,00971	1,00000	1,00000	0,00707	9,02254	2,93668	-3805,843727	Alt vs Null	0,000	1	1,000
		Null	56	0,08883	0,08883	0,56881	1,00000	1,00000	0,41441	1,00000	0,08883	0,00971	1,00000	1,00000	0,00707	9,02255	2,93668	-3805,843727				
SH2D1B (Ancestral branch)	0,2	Alt	57	0,08942	0,08942	0,47408	1,00000	1,00000	0,34583	2,84640	0,08942	0,10413	2,84640	1,00000	0,75960	9,02090	2,93613	-3805,857922	Alt vs Null	0,000	1	1,000
		Null	56	0,08942	0,08942	0,47408	1,00000	1,00000	0,34583	1,00000	0,08942	0,10413	1,00000	1,00000	0,75960	9,02090	2,93614	-3805,857908				
	2	Alt	57	0,08942	0,08942	0,47408	1,00000	1,00000	0,34584	2,84669	0,08942	0,10412	2,84669	1,00000	0,75960	9,02090	2,93613	-3805,857922	Alt vs Null	0,000	1	1,000
		Null	56	0,08942	0,08942	0,47407	1,00000	1,00000	0,34583	1,00000	0,08942	0,10414	1,00000	1,00000	0,75970	9,02090	2,93613	-3805,857908				
	5	Alt	57	0,08942	0,08942	0,47409	1,00000	1,00000	0,34584	2,84661	0,08942	0,10412	2,84661	1,00000	0,75950	9,02090	2,93614	-3805,857922	Alt vs Null	0,000	1	1,000
		Null	56	0,08942	0,08942	0,47415	1,00000	1,00000	0,34588	1,00000	0,08942	0,10406	1,00000	1,00000	0,75910	9,02091	2,93613	-3805,857908				
XIAP (Apical branches)	0,2	Alt	55	0,09876	0,09876	0,82583	1,00000	1,00000	0,16649	1,00000	0,09876	0,00639	1,00000	1,00000	0,00129	4,10040	3,59918	-9267,233342	Alt vs Null	0,000	1	1,000
		Null	54	0,09876	0,09876	0,82583	1,00000	1,00000	0,16649	1,00000	0,09876	0,00639	1,00000	1,00000	0,00129	4,10040	3,59918	-9267,233342				
	2	Alt	55	0,09876	0,09876	0,82583	1,00000	1,00000	0,16649	1,00000	0,09876	0,00639	1,00000	1,00000	0,00129	4,10040	3,59918	-9267,233342	Alt vs Null	0,000	1	1,000
		Null	54	0,09876	0,09876	0,82583	1,00000	1,00000	0,16649	1,00000	0,09876	0,00639	1,00000	1,00000	0,00129	4,10040	3,59918	-9267,233342				
	5	Alt	55	0,09876	0,09876	0,82583	1,00000	1,00000	0,16649	1,00000	0,09876	0,00639	1,00000	1,00000	0,00129	4,10040	3,59918	-9267,233342	Alt vs Null	0,000	1	1,000
		Null	54	0,09876	0,09876	0,82583	1,00000	1,00000	0,16649	1,00000	0,09876	0,00639	1,00000	1,00000	0,00129	4,10040	3,59918	-9267,233342				
XIAP (Ancestral branch)	0,2	Alt	55	0,09890	0,09890	0,82865	1,00000	1,00000	0,16651	999,000	0,09890	0,00403	999,000	1,00000	0,00081	4,10878	3,60643	-9264,155738	Alt vs Null	6,128	1	1,000
		Null	54	0,09898	0,09898	0,78015	1,00000	1,00000	0,16691	1,00000	0,09898	0,05240	1,00000	1,00000	0,01054	4,09950	3,59852	-9267,219532				
	2	Alt	55	0,09890	0,09890	0,82865	1,00000	1,00000	0,16652	999,000	0,09890	0,00403	999,000	1,00000	0,00081	4,10879	3,60643	-9264,155738	Alt vs Null	6,128	1	1,000
		Null	54	0,09898	0,09898	0,78015	1,00000	1,00000	0,16691	1,00000	0,09898	0,05239	1,00000	1,00000	0,01054	4,09950	3,59852	-9267,219532				
	5	Alt	55	0,09890	0,09890	0,82865	1,00000	1,00000	0,16652	999,000	0,09890	0,00403	999,000	1,00000	0,00081	4,10879	3,60643	-9264,155738	Alt vs Null	6,128	1	1,000
		Null	54	0,09898	0,09898	0,78015	1,00000	1,00000	0,16691	1,00000	0,09898	0,05239	1,00000	1,00000	0,01054	4,09950	3,59852	-9267,219532				

κ_i - Initial kappa value; $\mathbf{k}$ - Kappa; **LRT** - Likelihood-ratio test; **np** - Number of parameters; **DF** - Degrees of freedom; **SC** - Site classes; **fg** - Foreground; **bg** - Background; **p** - Proportion and ω - Ratio of non-synonymous/synonymous substitutions ratio

Table S8: Statistical hypothesis testing (SHT) of the branch-site model analysis of primates dataset, performed by the PAML program package.

Genes	Statistical hypothesis testing (SHT)																					
	k _i	Model code	np	(SC0) fg ω	(SC0) bg ω	(SC0) p	(SC1) fg ω	(SC1) bg ω	(SC1) p	(SC2a) fg ω	(SC2a) bg ω	(SC2a) p	(SC2b) fg ω	(SC2b) bg ω	(SC2b) p	Tree length	k	LnI	Model vs Model	LRT	DF	P-Value
SH2D1A (Apical branches)	0,2	Alt	97	0,07221	0,07221	0,56496	1,00000	1,00000	0,02637	1,00000	0,07221	0,39043	1,00000	1,00000	0,01823	1,20311	2,93353	-1349,195048	Alt vs Null	-0,001	1	1,000
		Null	96	0,07217	0,07217	0,56241	1,00000	1,00000	0,02623	1,00000	0,07217	0,39304	1,00000	1,00000	0,01833	1,20308	2,92108	-1349,194778				
	2	Alt	97	0,07217	0,07217	0,56241	1,00000	1,00000	0,02623	1,00000	0,07217	0,39304	1,00000	1,00000	0,01833	1,20308	2,92107	-1349,194778	Alt vs Null	0,000	1	1,000
		Null	96	0,07217	0,07217	0,56241	1,00000	1,00000	0,02623	1,00000	0,07217	0,39304	1,00000	1,00000	0,01833	1,20308	2,92107	-1349,194778				
	5	Alt	97	0,07217	0,07217	0,56242	1,00000	1,00000	0,02623	1,00000	0,07217	0,39303	1,00000	1,00000	0,01833	1,20308	2,92107	-1349,194778	Alt vs Null	0,000	1	1,000
		Null	96	0,07217	0,07217	0,56241	1,00000	1,00000	0,02623	1,00000	0,07217	0,39303	1,00000	1,00000	0,01833	1,20308	2,92107	-1349,194778				
SH2D1A (Ancestral branch)	0,2	Alt	97	0,07688	0,07688	0,91269	1,00000	1,00000	0,03919	3,14617	0,07688	0,04614	3,14617	1,00000	0,00198	1,20226	2,91596	-1349,872734	Alt vs Null	0,000	1	1,000
		Null	96	0,07688	0,07688	0,87594	1,00000	1,00000	0,03761	1,00000	0,07688	0,08289	1,00000	1,00000	0,00356	1,20226	2,91597	-1349,872735				
	2	Alt	97	0,07688	0,07688	0,87587	1,00000	1,00000	0,03761	3,46243	0,07688	0,08296	3,46243	1,00000	0,00356	1,20226	2,91596	-1349,872733	Alt vs Null	0,000	1	1,000
		Null	96	0,07688	0,07688	0,87594	1,00000	1,00000	0,03761	1,00000	0,07688	0,08289	1,00000	1,00000	0,00356	1,20226	2,91597	-1349,872735				
	5	Alt	97	0,07689	0,07689	0,88316	1,00000	1,00000	0,03791	3,48300	0,07689	0,07567	3,48300	1,00000	0,00325	1,20226	2,91634	-1349,876615	Alt vs Null	-0,008	1	1,000
		Null	96	0,07688	0,07688	0,87756	1,00000	1,00000	0,03768	1,00000	0,07688	0,08127	1,00000	1,00000	0,00349	1,20226	2,91596	-1349,872735				
SH2D1B (Apical branches)	0,2	Alt	105	0,14665	0,14665	0,64998	1,00000	1,00000	0,35002	1,00000	0,14665	0,00000	1,00000	1,00000	0,00000	2,55591	3,35813	-2235,766408	Alt vs Null	0,000	1	1,000
		Null	104	0,14666	0,14666	0,64998	1,00000	1,00000	0,35002	1,00000	0,14666	0,00000	1,00000	1,00000	0,00000	2,55592	3,35813	-2235,766408				
	2	Alt	105	0,14666	0,14666	0,64998	1,00000	1,00000	0,35002	1,00000	0,14666	0,00000	1,00000	1,00000	0,00000	2,55592	3,35812	-2235,766410	Alt vs Null	0,000	1	1,000
		Null	104	0,14666	0,14666	0,64998	1,00000	1,00000	0,35002	1,00000	0,14666	0,00000	1,00000	1,00000	0,00000	2,55592	3,35813	-2235,766408				
	5	Alt	105	0,14666	0,14666	0,64998	1,00000	1,00000	0,35002	1,00000	0,14666	0,00000	1,00000	1,00000	0,00000	2,55592	3,35813	-2235,766408	Alt vs Null	0,000	1	1,000
		Null	104	0,14666	0,14666	0,64998	1,00000	1,00000	0,35002	1,00000	0,14666	0,00000	1,00000	1,00000	0,00000	2,55592	3,35813	-2235,766408				
SH2D1B (Ancestral branch)	0,2	Alt	105	0,14666	0,14666	0,51253	1,00000	1,00000	0,27600	3,17934	0,14666	0,13745	3,17934	1,00000	0,07402	2,55592	3,35813	-2235,766405	Alt vs Null	0,000	1	1,000
		Null	104	0,14666	0,14666	0,51245	1,00000	1,00000	0,27596	1,00000	0,14666	0,13753	1,00000	1,00000	0,07406	2,55592	3,35813	-2235,766408				
	2	Alt	105	0,14666	0,14666	0,51249	1,00000	1,00000	0,27598	3,17906	0,14666	0,13749	3,17906	1,00000	0,07404	2,55593	3,35813	-2235,766410	Alt vs Null	0,000	1	1,000
		Null	104	0,14666	0,14666	0,51248	1,00000	1,00000	0,27597	1,00000	0,14666	0,13750	1,00000	1,00000	0,07404	2,55592	3,35813	-2235,766408				
	5	Alt	105	0,14666	0,14666	0,51249	1,00000	1,00000	0,27598	3,17907	0,14666	0,13749	3,17907	1,00000	0,07404	2,55592	3,35813	-2235,766405	Alt vs Null	0,000	1	1,000
		Null	104	0,14666	0,14666	0,51249	1,00000	1,00000	0,27598	1,00000	0,14666	0,13749	1,00000	1,00000	0,07404	2,55592	3,35813	-2235,766408				
XIAP (Apical branches)	0,2	Alt	89	0,04301	0,04301	0,00000	1,00000	1,00000	0,00000	1,00000	0,04301	0,76592	1,00000	1,00000	0,23408	1,36710	3,16289	-5469,245234	Alt vs Null	0,000	1	1,000
		Null	88	0,04301	0,04301	0,00051	1,00000	1,00000	0,00015	1,00000	0,04301	0,76541	1,00000	1,00000	0,23393	1,36708	3,16328	-5469,245234				
	2	Alt	89	0,04301	0,04301	0,00000	1,00000	1,00000	0,00000	1,00000	0,04301	0,76591	1,00000	1,00000	0,23408	1,36711	3,16296	-5469,245240	Alt vs Null	0,000	1	1,000
		Null	88	0,04301	0,04301	0,00000	1,00000	1,00000	0,00000	1,00000	0,04301	0,76591	1,00000	1,00000	0,23409	1,36711	3,16301	-5469,245235				
	5	Alt	89	0,04301	0,04301	0,00000	1,00000	1,00000	0,00000	1,00000	0,04301	0,76592	1,00000	1,00000	0,23408	1,36709	3,16311	-5469,245234	Alt vs Null	0,000	1	1,000
		Null	88	0,04301	0,04301	0,00000	1,00000	1,00000	0,00000	1,00000	0,04301	0,76591	1,00000	1,00000	0,23408	1,36710	3,16301	-5469,245234				
XIAP (Ancestral branch)	0,2	Alt	89	0,040607	0,040607	0,64684	1,00000	1,00000	0,19560	2,93855	0,040607	0,12098	2,93855	1,00000	0,03658	1,36661	3,15926	-5472,053649	Alt vs Null	0,000	1	1,000
		Null	88	0,040607	0,040607	0,64679	1,00000	1,00000	0,19558	1,00000	0,040607	0,12103	1,00000	1,00000	0,03660	1,36660	3,15923	-5472,053638				
	2	Alt	89	0,040607	0,040607	0,64643	1,00000	1,00000	0,19548	2,94592	0,040607	0,12139	2,94592	1,00000	0,03671	1,36661	3,15922	-5472,053650	Alt vs Null	0,000	1	1,000
		Null	88	0,040607	0,040607	0,64680	1,00000	1,00000	0,19529	1,00000	0,040607	0,12202	1,00000	1,00000	0,03690	1,36661	3,15923	-5472,053638				
	5	Alt	89	0,040607	0,040607	0,64683	1,00000	1,00000	0,19529	2,94752	0,040607	0,12199	2,94752	1,00000	0,03689	1,36661	3,15924	-5472,053649	Alt vs Null	0,000	1	1,000
		Null	88	0,040607	0,040607	0,64615	1,00000	1,00000	0,19539	1,00000	0,040607	0,12167	1,00000	1,00000	0,03679	1,36661	3,15923	-5472,053638				

k_i - Inicial kappa value; k - Kappa; LRT - Likelihood-ratio test; np - Number of parameters; DF - Degrees of freedom; SC - Site classes; fg - Foreground; bg - Background; p - Proportion and ω - Ratio of non-synonymous/synonymous substitutions ratio

Table S9: Statistical hypothesis testing (SHT) of the site model analysis for the great apes dataset, performed by the PAML program package.

Genes	Statistical hypothesis testing (SHT)										
	k _i	Model code	np	Parameters	Tree length	k	Lnl	Model vs Model	LRT	DF	P-Value
SH2D1A	0,2	M0	9	w0 = 0,07400	0,04771	2,37295	-560,0643420	M8 vs M7	0,000	2	1,000
		M7	10	p = 7,95342 q = 99,00000	0,04771	2,37333	-560,0648980				
		M8	12	p0 = 0,99999 p= 7,95267 q = 99,00000 (p1 = 0,00001) w = 1,00000	0,04771	2,37333	-560,0649060	M8 vs M8a	0,000	1	1,000
		M8a	11	p0 = 0,99999 p= 7,98598 q = 99,00000 (p1 = 0,00001) w = 1,00000	0,04771	2,37328	-560,0648340				
	2	M0	9	w0 = 0,07400	0,04771	2,37294	-560,0643420	M8 vs M7	0,000	2	1,000
		M7	10	p = 7,95353 q = 99,00000	0,04771	2,37334	-560,0648980				
		M8	12	p0 = 0,99999 p= 7,95274 q = 99,00000 (p1 = 0,00001) w = 1,00000	0,04771	2,37337	-560,0649060	M8 vs M8a	0,000	1	1,000
		M8a	11	p0 = 0,99999 p= 7,98598 q = 99,00000 (p1 = 0,00001) w = 1,00000	0,04771	2,37331	-560,0648340				
	5	M0	9	w0 = 0,07400	0,04771	2,37296	-560,0643420	M8 vs M7	0,000	2	1,000
		M7	10	p = 7,95342 q = 99,00000	0,04771	2,37333	-560,0648980				
		M8	12	p0 = 0,99999 p= 7,95261 q = 99,00000 (p1 = 0,00001) w = 1,00000	0,04771	2,16311	-560,0649060	M8 vs M8a	0,000	1	1,000
		M8a	11	p0 = 0,99999 p= 7,98591 q = 99,00000 (p1 = 0,00001) w = 1,00000	0,04771	2,3733	-560,0648340				
SH2D1B	0,2	M0	9	w0 = 1,24063	0,07619	17,96072	-595,8931680	M8 vs M7	0,097	2	0,953
		M7	10	p = 0,99536 q = 0,00500	0,07598	17,10320	-595,9417440				
		M8	12	p0 = 0,00001 p= 0,01342 q = 1,06663 (p1 = 0,99999) w = 1,24063	0,07619	17,96083	-595,8931690	M8 vs M8a	0,097	1	0,755
		M8a	11	p0 = 0,00001 p= 3,27001 q = 0,56498 (p1 = 0,99999) w = 1,00000	0,07598	17,10320	-595,9417430				
	2	M0	9	w0 = 1,24063	0,07619	17,96083	-595,8931680	M8 vs M7	0,097	2	0,953
		M7	10	p = 0,99396 q = 0,00500	0,07598	17,10140	-595,9417440				
		M8	12	p0 = 0,00001 p= 0,01823 q = 1,06663 (p1 = 0,99999) w = 1,24064	0,07619	17,96081	-595,8931690	M8 vs M8a	0,097	1	0,755
		M8a	11	p0 = 0,00001 p= 0,00500 q = 1,97936 (p1 = 0,99999) w = 1,00000	0,07598	17,10303	-595,9417490				
	5	M0	9	w0 = 1,24062	0,07619	17,96089	-595,8931680	M8 vs M7	0,097	2	0,953
		M7	10	p = 1,00192 q = 0,00500	0,07598	17,10336	-595,9417440				
		M8	12	p0 = 0,00001 p= 0,00859 q = 1,06758 (p1 = 0,99999) w = 1,24063	0,07619	17,96083	-595,8931690	M8 vs M8a	0,097	1	0,755
		M8a	11	p0 = 0,00001 p= 0,00500 q = 1,97977 (p1 = 0,99999) w = 1,00000	0,07598	17,10327	-595,9417490				
XIAP	0,2	M0	9	w0 = 0,18102	0,02503	1,857040	-2072,908273	M8 vs M7	-0,002	2	1,000
		M7	10	p = 21,92250 q = 99,00000	0,02503	1,857440	-2072,908654				
		M8	12	p0 = 0,99999 p= 13,43934 q = 60,59679 (p1 = 0,00001) w = 15,49437	0,02504	1,857840	-2072,909816	M8 vs M8a	-0,002	1	1,000
		M8a	11	p0 = 0,99999 p= 21,94849 q = 99,00000 (p1 = 0,00001) w = 1,0000000	0,02503	1,857450	-2072,908805				
	2	M0	9	w0 = 0,18103	0,02503	1,857220	-2072,908273	M8 vs M7	0,000	2	1,000
		M7	10	p = 21,92254 q = 99,00000	0,02503	1,857420	-2072,908654				
		M8	12	p0 = 0,99999 p= 21,80197 q = 99,00000 (p1 = 0,00001) w = 1,0000000	0,02504	1,859640	-2072,908698	M8 vs M8a	0,000	1	1,000
		M8a	11	p0 = 0,99999 p= 21,94998 q = 99,00000 (p1 = 0,00001) w = 1,0000000	0,02503	1,857220	-2072,908805				
	5	M0	9	w0 = 0,18101	0,02503	1,857240	-2072,908273	M8 vs M7	-0,002	2	1,000
		M7	10	p = 21,92247 q = 99,00000	0,02503	1,857430	-2072,908654				
		M8	12	p0 = 0,99999 p= 13,39662 q = 60,40472 (p1 = 0,00001) w = 15,45704	0,02504	1,857710	-2072,909817	M8 vs M8a	-0,002	1	1,000
		M8a	11	p0 = 0,99999 p= 21,95009 q = 99,00000 (p1 = 0,00001) w = 1,0000000	0,02503	1,857480	-2072,908805				

k_i - Inicial kappa value; k - Kappa; LRT - Likelihood-ratio test; np - Number of parameters and DF - Degrees of freedom.

Table S10: Statistical hypothesis testing (SHT) of the site model analysis for the Mammalia dataset, performed by the PAML program package.

Genes	Statistical hypothesis testing (SHT)										
	k _i	Model code	np	Parameters	Tree length	k	Lnl	Model vs Model	LRT	DF	P-Value
SH2D1A	0,2	M0	39	w0 = 0,13812	2,99744	3,13510	-2023,849326	M8 vs M7	0,103	2	0,950
		M7	40	p = 0,14876 q = 0,86250	3,05124	3,09143	-1977,670095				
		M8	42	p0 = 0,95778 p= 0,17424 q = 1,37829 (p1 = 0,04222) w = 1,00000	3,04478	3,09342	-1977,618444	M8 vs M8a	0,134	1	0,714
		M8a	41	p0 = 0,94454 p= 0,16629 q = 1,35148 (p1 = 0,05546) w = 1,00000	3,05134	3,09648	-1977,685580				
	2	M0	39	w0 = 0,13812	2,99744	3,13510	-2023,849326	M8 vs M7	0,103	2	0,950
		M7	40	p = 0,14876 q = 0,86250	3,05124	3,09143	-1977,670095				
		M8	42	p0 = 0,95778 p= 0,17424 q = 1,37829 (p1 = 0,04222) w = 1,00000	3,04478	3,09342	-1977,618444	M8 vs M8a	0,134	1	0,714
		M8a	41	p0 = 0,94454 p= 0,16629 q = 1,35148 (p1 = 0,05546) w = 1,00000	3,05134	3,09648	-1977,685580				
	5	M0	39	w0 = 0,13812	2,99744	3,13510	-2023,849326	M8 vs M7	0,103	2	0,950
		M7	40	p = 0,14876 q = 0,86250	3,05124	3,09143	-1977,670095				
		M8	42	p0 = 0,95778 p= 0,17424 q = 1,37829 (p1 = 0,04222) w = 1,00000	3,04478	3,09342	-1977,618444	M8 vs M8a	0,134	1	0,714
		M8a	41	p0 = 0,94454 p= 0,16629 q = 1,35148 (p1 = 0,05546) w = 1,00000	3,05134	3,09648	-1977,685580				
SH2D1B	0,2	M0	37	w0 = 0,43989	7,53373	2,72160	-3503,505888	M8 vs M7	18,688	2	<0,001
		M7	38	p = 0,22764 q = 0,30952	8,29438	2,74200	-3351,295576				
		M8	40	p0 = 0,74437 p= 0,36979 q = 1,26008 (p1 = 0,25563) w = 1,51846	8,38802	3,01150	-3341,951569	M8 vs M8a	10,762	1	0,001
		M8a	39	p0 = 0,64796 p= 0,48516 q = 3,26767 (p1 = 0,35204) w = 1,00000	8,20383	2,79510	-3347,332525				
	2	M0	37	w0 = 0,43989	7,53373	2,72160	-3503,505888	M8 vs M7	18,688	2	<0,001
		M7	38	p = 0,22764 q = 0,30952	8,29438	2,74200	-3351,295576				
		M8	40	p0 = 0,74437 p= 0,36979 q = 1,26008 (p1 = 0,25563) w = 1,51846	8,38802	3,01150	-3341,951569	M8 vs M8a	10,762	1	0,001
		M8a	39	p0 = 0,64796 p= 0,48516 q = 3,26767 (p1 = 0,35204) w = 1,00000	8,20383	2,79510	-3347,332525				
	5	M0	37	w0 = 0,43989	7,53373	2,72160	-3503,505888	M8 vs M7	18,688	2	<0,001
		M7	38	p = 0,22764 q = 0,30952	8,29438	2,74200	-3351,295576				
		M8	40	p0 = 0,74437 p= 0,36979 q = 1,26008 (p1 = 0,25563) w = 1,51846	8,38802	3,01150	-3341,951569	M8 vs M8a	10,762	1	0,001
		M8a	39	p0 = 0,64796 p= 0,48516 q = 3,26767 (p1 = 0,35204) w = 1,00000	8,20383	2,79510	-3347,332525				

k_i - Inicial kappa value; k - Kappa; LRT - Likelihood-ratio test; np - Number of parameters and DF - Degrees of freedom.

Evolutionary genomic analyses of *SH2D1A*, *SH2D1B* and *XIAP* genes in primates: Implications on immune functionality

Table S10: Statistical hypothesis testing (SHT) of the site model analysis for the Mammalia dataset, performed by the PAML program package.

Genes	Statistical hypothesis testing (SHT)										
	k_i	Model code	np	Parameters	Tree length	k	LnI	Model vs Model	LRT	DF	P-Value
XIAP	0,2	M0	35	$w0 = 0,18784$	3,52878	3,23100	-8521,365456				
		M7	36	$p = 0,36476 \ q = 1,34372$	3,73569	3,32850	-8350,822561	M8 vs M7	18,131	2	<0,001
		M8	38	$p0 = 0,97053 \ p = 0,47198 \ q = 2,16370 \ (p1 = 0,02947) \ w = 1,65774$	3,74189	3,37528	-8341,756861				
		M8a	37	$p0 = 0,91892 \ p = 0,53786 \ q = 3,07481 \ (p1 = 0,08108) \ w = 1,00000$	3,72194	3,34479	-8343,944947	M8 vs M8a	4,376	1	0,036
	2	M0	35	$w0 = 0,18784$	3,52878	3,23100	-8521,365456				
		M7	36	$p = 0,36476 \ q = 1,34372$	3,73569	3,32850	-8350,822561	M8 vs M7	18,131	2	<0,001
		M8	38	$p0 = 0,97053 \ p = 0,47198 \ q = 2,16370 \ (p1 = 0,02947) \ w = 1,65774$	3,74189	3,37528	-8341,756861				
		M8a	37	$p0 = 0,91892 \ p = 0,53786 \ q = 3,07481 \ (p1 = 0,08108) \ w = 1,00000$	3,72194	3,34479	-8343,944947	M8 vs M8a	4,376	1	0,036
	5	M0	35	$w0 = 0,18784$	3,52878	3,23100	-8521,365456				
		M7	36	$p = 0,36476 \ q = 1,34372$	3,73569	3,32850	-8350,822561	M8 vs M7	18,131	2	<0,001
		M8	38	$p0 = 0,97053 \ p = 0,47198 \ q = 2,16370 \ (p1 = 0,02947) \ w = 1,65774$	3,74189	3,37528	-8341,756861				
		M8a	37	$p0 = 0,91892 \ p = 0,53786 \ q = 3,07481 \ (p1 = 0,08108) \ w = 1,00000$	3,72194	3,34479	-8343,944947	M8 vs M8a	4,376	1	0,036

k_i - Inicial kappa value; **k** - Kappa; **LRT** - Likelihood-ratio test; **np** - Number of parameters and **DF** - Degrees of freedom.

Table S11: Statistical hypothesis testing (SHT) of the site model analysis for the primates dataset, performed by the PAML program package.

Genes	Statistical hypothesis testing (SHT)										
	k _i	Model code	np	Parameters	Tree length	k	Lnl	Model vs Model	LRT	DF	P-Value
SH2D1A	0,2	M0	81	w0 = 0,10526	1,14115	2,99133	-1308,306562	M8 vs M7	0,000	2	1,000
		M7	82	p = 0,35069 q = 2,77193	1,14581	2,99997	-1305,478420				
		M8	84	p0 = 0,99999 p= 0,35072 q = 2,77228 (p1 = 0,00001) w = 1,00000	1,14581	2,99998	-1305,478539	M8 vs M8a	-0,077	1	1,000
		M8a	83	p0 = 0,99999 p= 0,29598 q = 2,19948 (p1 = 0,00001) w = 1,00000	1,14580	3,00009	-1305,439859				
	2	M0	81	w0 = 0,10526	1,14115	2,99131	-1308,306562	M8 vs M7	0,000	2	1,000
		M7	82	p = 0,35069 q = 2,77194	1,14581	2,99997	-1305,478420				
		M8	84	p0 = 0,99999 p= 0,35072 q = 2,77230 (p1 = 0,00001) w = 1,00000	1,14581	2,99998	-1305,478539	M8 vs M8a	-0,077	1	1,000
		M8a	83	p0 = 0,99999 p= 0,29598 q = 2,19950 (p1 = 0,00001) w = 1,00000	1,14580	3,00009	-1305,439859				
	5	M0	81	w0 = 0,10526	1,14115	2,99087	-1308,313131	M8 vs M7	0,000	2	1,000
		M7	82	p = 0,35069 q = 2,77194	1,14581	2,99997	-1305,478420				
		M8	84	p0 = 0,99999 p= 0,35071 q = 2,77228 (p1 = 0,00001) w = 1,00000	1,14581	2,99998	-1305,478539	M8 vs M8a	-0,088	1	1,000
		M8a	83	p0 = 0,99999 p= 0,29608 q = 2,20025 (p1 = 0,00001) w = 1,00000	1,14580	3,00053	-1305,434414				
SH2D1B	0,2	M0	89	w0 = 0,59101	2,44070	3,62129	-2209,528649	M8 vs M7	29,871	2	<0,001
		M7	90	p = 0,27995 q = 0,32527	2,48593	3,33905	-2178,642652				
		M8	92	p0 = 0,90907 p= 0,602810 q = 0,920180 (p1 = 0,09093) w = 3,20810	2,57261	3,71567	-2163,707142	M8 vs M8a	23,382	1	<0,001
		M8a	91	p0 = 0,65198 p= 17,03462 q = 99,98921 (p1 = 0,34802) w = 1,00000	2,48340	3,29291	-2175,398116				
	2	M0	89	w0 = 0,59101	2,44070	3,62129	-2209,528649	M8 vs M7	29,928	2	<0,001
		M7	90	p = 0,27995 q = 0,32527	2,48593	3,33905	-2178,642652				
		M8	92	p0 = 0,90985 p= 0,603060 q = 0,92205 (p1 = 0,09215) w = 3,19517	2,57261	3,71355	-2163,678411	M8 vs M8a	23,439	1	<0,001
		M8a	91	p0 = 0,65199 p= 17,03462 q = 99,0000 (p1 = 0,34801) w = 1,00000	2,48340	3,29290	-2175,398115				
	2	M0	89	w0 = 0,59101	2,44070	3,62129	-2209,528649	M8 vs M7	29,870	2	<0,001
		M7	90	p = 0,27988 q = 0,32514	2,48593	3,33984	-2178,613405				
		M8	92	p0 = 0,90985 p= 0,603060 q = 0,92205 (p1 = 0,09215) w = 3,19517	2,57261	3,71355	-2163,678411	M8 vs M8a	23,439	1	<0,001
		M8a	91	p0 = 0,65200 p= 17,03548 q = 99,0000 (p1 = 0,34800) w = 1,00000	2,48338	3,29290	-2175,398115				
XIAP	0,2	M0	73	w0 = 0,23376	1,04567	3,37430	-5059,233427	M8 vs M7	0,346	2	0,839
		M7	74	p = 0,23721 q = 0,73327	1,06479	3,40946	-5020,115671				
		M8	76	p0 = 0,88653 p= 0,37418 q = 2,05750 (p1 = 0,11347) w = 1,00000	1,06586	3,41780	-5019,942757	M8 vs M8a	0,000	1	1,000
		M8a	75	p0 = 0,87871 p= 0,36110 q = 2,00721 (p1 = 0,12129) w = 1,00000	1,06598	3,41820	-5019,942603				
	2	M0	73	w0 = 0,23376	1,04567	3,37430	-5059,233427	M8 vs M7	0,344	2	0,839
		M7	74	p = 0,23721 q = 0,73328	1,06478	3,40946	-5020,115671				
		M8	76	p0 = 0,88794 p= 0,36653 q = 1,98706 (p1 = 0,11206) w = 1,00000	1,06567	3,41700	-5019,943565	M8 vs M8a	-0,002	1	1,000
		M8a	75	p0 = 0,87871 p= 0,36110 q = 2,00719 (p1 = 0,12129) w = 1,00000	1,06598	3,41820	-5019,942603				
	5	M0	73	w0 = 0,23376	1,04567	3,37430	-5059,233427	M8 vs M7	0,344	2	0,839
		M7	74	p = 0,23721 q = 0,73327	1,06479	3,40946	-5020,115671				
		M8	76	p0 = 0,88777 p= 0,36728 q = 1,99338 (p1 = 0,11223) w = 1,00000	1,06571	3,41690	-5019,943466	M8 vs M8a	-0,002	1	1,000
		M8a	75	p0 = 0,87871 p= 0,36109 q = 2,00718 (p1 = 0,12129) w = 1,00000	1,06598	3,41820	-5019,942603				

k_i - Inicial kappa value; **k** - Kappa; **LRT** - Likelihood-ratio test; **np** - Number of parameters and **DF** - Degrees of freedom.

Table S12: Positive sites determined by Bayes Empirical Bayes (BEB) of site model analysis for mammalia dataset, performed by the PAML program package[61] .

Genes	Bayes Empirical Bayes (BEB) analysis		
	Amino acid	Pr($\omega > 1$)	E(ωX) $\pm$ SE
SH2D1B	11 Asn	0,788	1,345 $\pm$ 0,309
	13 Gln	0,963	1,476 $\pm$ 0,129
	17 Thr	0,776	1,335 $\pm$ 0,317
	24 Lys	0,828	1,376 $\pm$ 0,280
	36 Val	0,863	1,403 $\pm$ 0,251
	49 Lys	0,957	1,472 $\pm$ 0,137
	57 Lys	0,978	1,487 $\pm$ 0,098
	60 His	0,994	1,498 $\pm$ 0,056
	62 Tyr	0,711	1,278 $\pm$ 0,360
	64 Ser	0,932	1,455 $\pm$ 0,172
	67 Ser	0,513	1,086 $\pm$ 0,447
	68 Glu	0,971	1,482 $\pm$ 0,114
	70 Arg	0,642	1,222 $\pm$ 0,386
	71 Gly	0,996	1,499 $\pm$ 0,050
	72 Lys	0,946	1,464 $\pm$ 0,156
	74 Glu	0,554	1,135 $\pm$ 0,427
	75 Leu	0,870	1,409 $\pm$ 0,241
	95 Ile	0,575	1,144 $\pm$ 0,434
	96 His	0,763	1,325 $\pm$ 0,323
	98 Arg	0,856	1,392 $\pm$ 0,273
	99 Phe	0,875	1,413 $\pm$ 0,235
	102 Val	0,863	1,405 $\pm$ 0,244
	103 Lys	0,703	1,274 $\pm$ 0,360
	104 Asn	0,985	1,492 $\pm$ 0,082
	106 Val	0,937	1,458 $\pm$ 0,169
	107 Tyr	0,929	1,452 $\pm$ 0,178
	108 Gln	0,796	1,345 $\pm$ 0,317
	109 Ser	0,802	1,351 $\pm$ 0,311
	113 -	0,988	1,493 $\pm$ 0,078
	114 -	0,709	1,274 $\pm$ 0,365
	115 -	0,888	1,422 $\pm$ 0,226
	116 -	0,929	1,452 $\pm$ 0,180
	117 -	0,855	1,393 $\pm$ 0,270

Continuation of **Table S12**

Genes	Bayes Empirical Bayes (BEB) analysis		
	Amino acid	Pr($\omega > 1$)	E(ωX) $\pm$ SE
XIAP	62 Lys	0,888	0,888 $\pm$ 1,419
	110 Ala	0,656	0,656 $\pm$ 1,222
	115 Tyr	0,525	0,525 $\pm$ 1,100
	117 Pro	0,931	0,931 $\pm$ 1,454
	120 Cys	0,630	0,630 $\pm$ 1,214
	121 Ser	0,840	0,840 $\pm$ 1,380
	255 Ile	0,972	0,972 $\pm$ 1,483
	344 Lys	0,675	0,675 $\pm$ 1,240
	351 -	0,521	0,521 $\pm$ 1,074
	352 -	0,929	0,929 $\pm$ 1,452
	353 -	0,673	0,673 $\pm$ 1,246
	374 Tyr	0,945	0,945 $\pm$ 1,464
	409 Thr	0,541	0,541 $\pm$ 1,131
	415 Ser	0,826	0,826 $\pm$ 1,370

ω - Non-synonymous/synonymous substitutions ratio; **Pr** - Probability; **E($\omega|X$)** - Posterior mean of ω and **SE** - Standart error.

Table S13: Positive sites determined by Bayes Empirical Bayes (BEB) of site model analysis for primates dataset, performed by the PAML program package[61] .

Gene	Bayes Empirical Bayes (BEB) analysis		
	Amino acid	Pr($\omega > 1$)	E(ωX) $\pm$ SE
SH2D1B	9 Pro	0,716	2,440 $\pm$ 1,185
	13 Arg	0,997	3,223 $\pm$ 0,751
	49 Thr	0,999	3,227 $\pm$ 0,745
	64 Lys	0,682	2,324 $\pm$ 1,147
	71 Val	0,868	2,863 $\pm$ 1,032
	72 Pro	0,572	2,043 $\pm$ 1,190
	96 Gly	0,999	3,228 $\pm$ 0,744
	104 Thr	0,568	2,021 $\pm$ 1,151
	108 Leu	0,780	2,610 $\pm$ 1,124
	112 Glu	0,733	2,459 $\pm$ 1,119
	113 Met	0,940	3,064 $\pm$ 0,899
	119 Ser	0,895	2,932 $\pm$ 0,978

ω - Non-synonymous/synonymous substitutions ratio; **Pr** - Probability; **E($\omega|X$)** - Posterior mean of ω and **SE** - Standart error.

Table S14: FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the *SH2D1A* gene in Mammalia data set.

<i>SH2D1A</i> gene							
Site	FEL		FUBAR		SLAC		MEME
	Class	P-Value	Pr[$\alpha > \beta$]	Pr[$\alpha < \beta$]	Pr[dN/dS > 1]	Pr[dN/dS < 1]	
2	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
3	Purifying	0,000	1,000	0,000	0,999	0,005	0,670
6	Purifying	0,000	0,999	0,001	1,000	0,004	0,670
7	Purifying	0,000	0,998	0,001	1,000	0,005	0,670
8	Purifying	0,000	1,000	0,000	1,000	0,002	0,670
13	Purifying	0,000	1,000	0,000	1,000	0,003	0,670
14	Purifying	0,000	1,000	0,000	1,000	0,002	0,670
15	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
16	Purifying	0,005	0,984	0,011	1,000	0,037	0,670
19	Purifying	0,009	0,973	0,018	1,000	0,043	0,670
21	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
23	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
26	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
28	Purifying	0,008	0,968	0,022	1,000	0,040	0,670
30	Purifying	0,000	0,999	0,001	1,000	0,011	0,670
32	Purifying	0,001	0,997	0,002	1,000	0,015	0,670
34	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
36	Purifying	0,014	0,955	0,013	0,994	0,033	0,670
38	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
39	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
43	Purifying	0,001	0,996	0,003	1,000	0,018	0,670
44	Purifying	0,001	0,992	0,005	1,000	0,014	0,670
45	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
53	Purifying	0,004	0,987	0,009	1,000	0,037	0,670
54	Purifying	0,000	0,997	0,002	1,000	0,003	0,670
55	Purifying	0,004	0,993	0,004	1,000	0,021	0,670
66	Purifying	0,019	0,962	0,025	0,996	0,045	0,670
71	Purifying	0,006	0,982	0,012	1,000	0,037	0,670
72	Purifying	0,004	0,989	0,007	1,000	0,037	0,670
75	Purifying	0,005	0,984	0,011	1,000	0,029	0,670
77	Purifying	0,009	0,972	0,019	1,000	0,039	0,670
93	Purifying	0,002	0,990	0,007	1,000	0,012	0,670
98	Purifying	0,001	0,999	0,001	1,000	0,014	0,670
99	Purifying	0,002	0,991	0,006	1,000	0,017	0,670
102	Purifying	0,002	0,994	0,004	1,000	0,012	0,670

Pr[$\alpha < \beta$] - Posterior probability of positive selection; **Pr[$\alpha > \beta$]** - Posterior probability of negative selection; **α** - Mean posterior synonymous substitution rate; **β** - Mean posterior non-synonymous substitution rate; **Pr[dN/dS > 1]** - The binomial probability that synonymous substitution in no greater than the observed value, with a probability of success (PS); **Pr[dN/dS < 1]** - The binomial probability that synonymous substitution in no less than the observed value, with a probability of success (PS); **dS** - synonymous substitution rate; **dN** - non-synonymous substitution rate.

Table S15: FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the *XIAP* gene in Mammalia dataset.

<i>XIAP</i> gene							
Site	FEL		FUBAR		SLAC		MEME
	Class	P-Value	Pr[$\alpha > \beta$]	Pr[$\alpha < \beta$]	Pr[dN/dS > 1]	Pr[dN/dS < 1]	
19	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
32	Purifying	0,000	1,000	0,000	1,000	0,004	0,670
36	Purifying	0,000	1,000	0,000	1,000	0,003	0,670

Continuation of Table S15

<i>XIAP</i> gene							
Site	FEL		FUBAR		SLAC		MEME
	Class	P-Value	Pr[$\alpha > \beta$]	Pr[$\alpha < \beta$]	Pr[dN/dS > 1]	Pr[dN/dS < 1]	P-Value
40	Purifying	0,000	1,000	0,000	1,000	0,003	0,670
41	Purifying	0,001	0,998	0,001	1,000	0,037	0,670
43	Purifying	0,006	0,996	0,001	0,998	0,028	0,670
44	Purifying	0,008	0,993	0,003	0,993	0,047	0,670
46	Purifying	0,000	1,000	0,000	1,000	0,004	0,670
47	Purifying	0,003	0,997	0,001	1,000	0,029	0,670
48	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
50	Purifying	0,000	1,000	0,000	1,000	0,004	0,670
51	Purifying	0,000	0,999	0,000	1,000	0,012	0,670
53	Purifying	0,008	0,994	0,003	0,998	0,019	0,670
54	Purifying	0,000	1,000	0,000	1,000	0,004	0,670
56	Purifying	0,002	0,996	0,002	1,000	0,037	0,670
59	Purifying	0,000	1,000	0,000	1,000	0,003	0,670
60	Purifying	0,000	1,000	0,000	0,999	0,004	0,670
61	Purifying	0,001	0,999	0,000	1,000	0,004	0,670
65	Purifying	0,000	1,000	0,000	1,000	0,003	0,670
69	Purifying	0,001	0,998	0,001	0,999	0,018	0,310
74	Purifying	0,000	1,000	0,000	1,000	0,003	0,670
76	Purifying	0,002	0,997	0,002	1,000	0,041	0,670
79	Purifying	0,001	0,999	0,001	1,000	0,012	0,670
84	Purifying	0,002	0,998	0,001	1,000	0,011	0,670
87	Purifying	0,020	0,975	0,017	1,000	0,037	0,670
94	Purifying	0,001	0,997	0,002	0,999	0,010	0,670
95	Purifying	0,002	0,996	0,002	1,000	0,012	0,670
110	Diversifying	0,014	0,004	0,977	0,112	0,981	0,020
132	Purifying	0,000	1,000	0,000	1,000	0,008	0,670
136	Purifying	0,004	0,994	0,003	0,996	0,045	0,650
140	Purifying	0,000	0,999	0,000	1,000	0,004	0,670
141	Purifying	0,003	0,996	0,002	1,000	0,037	0,670
145	Purifying	0,005	0,994	0,003	0,993	0,048	0,670
148	Purifying	0,016	0,985	0,009	0,996	0,042	0,670
152	Purifying	0,000	1,000	0,000	1,000	0,003	0,670
156	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
158	Purifying	0,003	0,994	0,004	1,000	0,011	0,670
162	Purifying	0,017	0,967	0,021	0,986	0,094	0,300
166	Purifying	0,007	0,990	0,006	0,996	0,045	0,630
167	Purifying	0,002	0,998	0,002	1,000	0,027	0,670
168	Purifying	0,001	0,999	0,000	1,000	0,022	0,670
173	Purifying	0,001	0,998	0,001	1,000	0,003	0,670
180	Purifying	0,000	1,000	0,000	0,995	0,048	0,670
182	Purifying	0,003	0,997	0,002	1,000	0,037	0,670
186	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
189	Purifying	0,004	0,995	0,003	1,000	0,041	0,670
194	Purifying	0,001	0,998	0,001	1,000	0,037	0,670

Pr[$\alpha < \beta$] - Posterior probability of positive selection; **Pr[$\alpha > \beta$]** - Posterior probability of negative selection; **α** - Mean posterior synonymous substitution rate; **β** - Mean posterior non-synonymous substitution rate; **Pr[dN/dS > 1]** - The binomial probability that synonymous substitution in no greater than the observed value, with a probability of success (PS); **Pr[dN/dS < 1]** - The binomial probability that synonymous substitution in no less than the observed value, with a probability of success (PS); **dS** - synonymous substitution rate; **dN** - non-synonymous substitution rate.

Continuation of Table S15

XIAP gene							
Site	FEL		FUBAR		SLAC		MEME
	Class	P-Value	Pr[$\alpha > \beta$]	Pr[$\alpha < \beta$]	Pr[dN/dS > 1]	Pr[dN/dS < 1]	P-Value
197	Purifying	0,000	0,998	0,001	1,000	0,011	0,670
199	Purifying	0,003	0,997	0,002	1,000	0,012	0,670
205	Purifying	0,002	0,997	0,002	1,000	0,037	0,670
206	Purifying	0,004	0,994	0,004	1,000	0,041	0,670
208	Purifying	0,009	0,990	0,006	1,000	0,045	0,670
210	Purifying	0,001	0,998	0,001	1,000	0,011	0,670
213	Purifying	0,000	0,999	0,000	1,000	0,012	0,670
216	Purifying	0,003	0,995	0,003	0,995	0,048	0,400
217	Purifying	0,001	0,999	0,000	1,000	0,003	0,670
219	Purifying	0,011	0,991	0,006	1,000	0,049	0,670
220	Purifying	0,001	0,998	0,001	1,000	0,013	0,670
223	Purifying	0,000	1,000	0,000	1,000	0,006	0,670
224	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
225	Purifying	0,001	0,998	0,001	1,000	0,011	0,670
226	Purifying	0,000	1,000	0,000	1,000	0,012	0,670
228	Purifying	0,003	0,994	0,004	1,000	0,014	0,670
231	Purifying	0,000	1,000	0,000	1,000	0,004	0,670
235	Purifying	0,001	0,994	0,003	0,997	0,026	0,670
249	Purifying	0,005	0,994	0,004	1,000	0,030	0,670
275	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
299	Purifying	0,001	0,998	0,001	1,000	0,012	0,670
307	Purifying	0,007	0,992	0,005	1,000	0,038	0,670
308	Purifying	0,003	0,997	0,002	1,000	0,031	0,670
313	Purifying	0,009	0,988	0,008	1,000	0,037	0,670
316	Purifying	0,001	0,998	0,001	1,000	0,003	0,670
319	Purifying	0,001	0,998	0,001	1,000	0,015	0,670
320	Purifying	0,000	1,000	0,000	1,000	0,002	0,670
328	Purifying	0,001	0,998	0,001	1,000	0,014	0,670
330	Purifying	0,002	0,994	0,002	0,995	0,038	0,670
331	Purifying	0,010	0,988	0,007	1,000	0,044	0,670
335	Purifying	0,005	0,994	0,004	1,000	0,019	0,670
336	Purifying	0,000	1,000	0,000	1,000	0,002	0,670
350	Purifying	0,015	0,982	0,004	0,995	0,030	0,670
368	Purifying	0,006	0,994	0,004	1,000	0,037	0,670
371	Purifying	0,000	0,999	0,000	1,000	0,012	0,670
375	Purifying	0,000	1,000	0,000	1,000	0,005	0,670
386	Purifying	0,009	0,979	0,001	0,998	0,014	0,670
393	Purifying	0,004	0,996	0,003	1,000	0,013	0,670
400	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
402	Purifying	0,006	0,996	0,002	1,000	0,028	0,670
435	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
443	Purifying	0,001	0,998	0,001	1,000	0,003	0,670
446	Purifying	0,001	0,998	0,001	1,000	0,014	0,670

Pr[$\alpha < \beta$] - Posterior probability of positive selection; Pr[$\alpha > \beta$] - Posterior probability of negative selection; α - Mean posterior synonymous substitution rate; β - Mean posterior non-synonymous substitution rate; Pr[dN/dS > 1] - The binomial probability that synonymous substitution in no greater than the observed value, with a probability of success (PS); Pr[dN/dS < 1] - The binomial probability that synonymous substitution in no less than the observed value, with a probability of success (PS); dS - synonymous substitution rate; dN - non-synonymous substitution rate.

Continuation of Table S15

XIAP gene							
Site	FEL		FUBAR		SLAC		MEME
	Class	P-Value	Pr[$\alpha > \beta$]	Pr[$\alpha < \beta$]	Pr[dN/dS > 1]	Pr[dN/dS < 1]	
452	Purifying	0,004	0,993	0,004	0,996	0,045	0,550
455	Purifying	0,001	0,998	0,001	1,000	0,011	0,670
465	Purifying	0,008	0,988	0,008	0,997	0,045	0,540
466	Purifying	0,001	0,997	0,000	0,992	0,033	0,670
475	Purifying	0,009	0,990	0,006	1,000	0,037	0,670
479	Purifying	0,002	0,997	0,002	1,000	0,037	0,670
482	Purifying	0,026	0,971	0,019	0,989	0,104	0,670
485	Purifying	0,000	0,999	0,000	1,000	0,007	0,670

Pr[$\alpha < \beta$] - Posterior probability of positive selection; **Pr[$\alpha > \beta$]** - Posterior probability of negative selection; **α** - Mean posterior synonymous substitution rate; **β** - Mean posterior non-synonymous substitution rate; **Pr[dN/dS > 1]** - The binomial probability that synonymous substitution in no greater than the observed value, with a probability of success (PS); **Pr[dN/dS < 1]** - The binomial probability that synonymous substitution in no less than the observed value, with a probability of success (PS); **dS** - synonymous substitution rate; **dN** - non-synonymous substitution rate.

Table S16: FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the *SH2D1B* gene in Mammalia data set.

SH2D1B gene							
Site	FEL		FUBAR		SLAC		MEME
	Class	P-Value	Pr[$\alpha > \beta$]	Pr[$\alpha < \beta$]	Pr[dN/dS > 1]	Pr[dN/dS < 1]	
2	Purifying	0,002	0,994	0,004	1,000	0,009	0,670
3	Purifying	0,013	0,968	0,021	1,000	0,033	0,670
5	Purifying	0,001	0,997	0,001	1,000	0,003	0,670
7	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
8	Purifying	0,002	0,996	0,002	1,000	0,012	0,670
16	Purifying	0,003	0,992	0,005	1,000	0,014	0,670
19	Purifying	0,005	0,983	0,011	1,000	0,016	0,670
25	Purifying	0,001	0,997	0,001	1,000	0,001	0,670
29	Purifying	0,009	0,976	0,016	1,000	0,027	0,670
30	Purifying	0,017	0,962	0,020	0,997	0,026	0,670
31	Purifying	0,001	0,996	0,002	1,000	0,009	0,670
33	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
35	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
37	Purifying	0,003	0,993	0,004	0,999	0,018	0,670
41	Purifying	0,001	0,998	0,001	1,000	0,007	0,670
42	Purifying	0,013	0,979	0,014	1,000	0,049	0,670
52	Purifying	0,000	0,999	0,001	1,000	0,002	0,670
53	Purifying	0,001	0,994	0,004	1,000	0,009	0,670
61	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
63	Purifying	0,015	0,959	0,009	0,996	0,025	0,080
76	Purifying	0,000	0,997	0,002	1,000	0,002	0,670
79	Purifying	0,004	0,986	0,008	1,000	0,008	0,670
82	Purifying	0,002	0,994	0,001	0,999	0,007	0,670
89	Purifying	0,001	0,997	0,002	1,000	0,012	0,670
92	Purifying	0,001	0,996	0,002	1,000	0,013	0,670
97	Purifying	0,008	0,986	0,009	1,000	0,039	0,670
100	Purifying	0,007	0,984	0,010	0,999	0,018	0,670
101	Purifying	0,005	0,985	0,009	0,998	0,016	0,670

Pr[$\alpha < \beta$] - Posterior probability of positive selection; **Pr[$\alpha > \beta$]** - Posterior probability of negative selection; **α** - Mean posterior synonymous substitution rate; **β** - Mean posterior non-synonymous substitution rate; **Pr[dN/dS > 1]** - The binomial probability that synonymous substitution in no greater than the observed value, with a probability of success (PS); **Pr[dN/dS < 1]** - The binomial probability that synonymous substitution in no less than the observed value, with a probability of success (PS); **dS** - synonymous substitution rate; **dN** - non-synonymous substitution rate.

Table S17: FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the *XIAP* gene in primates dataset.

<i>XIAP</i> gene							
Site	FEL		FUBAR		SLAC		MEME
	Class	P-Value	Pr[$\alpha > \beta$]	Pr[$\alpha < \beta$]	Pr[dN/dS > 1]	Pr[dN/dS < 1]	
60	Purifying	0,000	1,000	0,000	1,000	0,000	0,670
94	Purifying	0,004	0,989	0,007	1,000	0,011	0,670
155	Purifying	0,000	1,000	0,000	1,000	0,004	0,670
167	Purifying	0,005	0,998	0,000	1,000	0,037	0,670
185	Purifying	0,001	0,999	0,000	1,000	0,012	0,670
220	Purifying	0,017	0,955	0,032	1,000	0,049	0,670
274	Purifying	0,011	0,989	0,007	1,000	0,037	0,670
345	Purifying	0,016	0,980	0,013	1,000	0,037	0,670
347	Purifying	0,006	0,995	0,003	1,000	0,049	0,670
349	Purifying	0,005	0,992	0,005	1,000	0,017	0,670
350	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
360	Purifying	0,007	0,997	0,001	0,996	0,046	0,670
363	Purifying	0,002	0,996	0,002	1,000	0,037	0,670
371	Purifying	0,024	0,953	0,034	1,000	0,049	0,670
383	Purifying	0,003	0,995	0,003	1,000	0,041	0,670

Pr[$\alpha < \beta$] - Posterior probability of positive selection; **Pr[$\alpha > \beta$]** - Posterior probability of negative selection; **α** - Mean posterior synonymous substitution rate; **β** - Mean posterior non-synonymous substitution rate; **Pr[dN/dS > 1]** - The binomial probability that synonymous substitution in no greater than the observed value, with a probability of success (PS); **Pr[dN/dS < 1]** - The binomial probability that synonymous substitution in no less than the observed value, with a probability of success (PS); **dS** - synonymous substitution rate; **dN** - non-synonymous substitution rate.

Table S18: FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the *SH2D1B* gene in primates dataset.

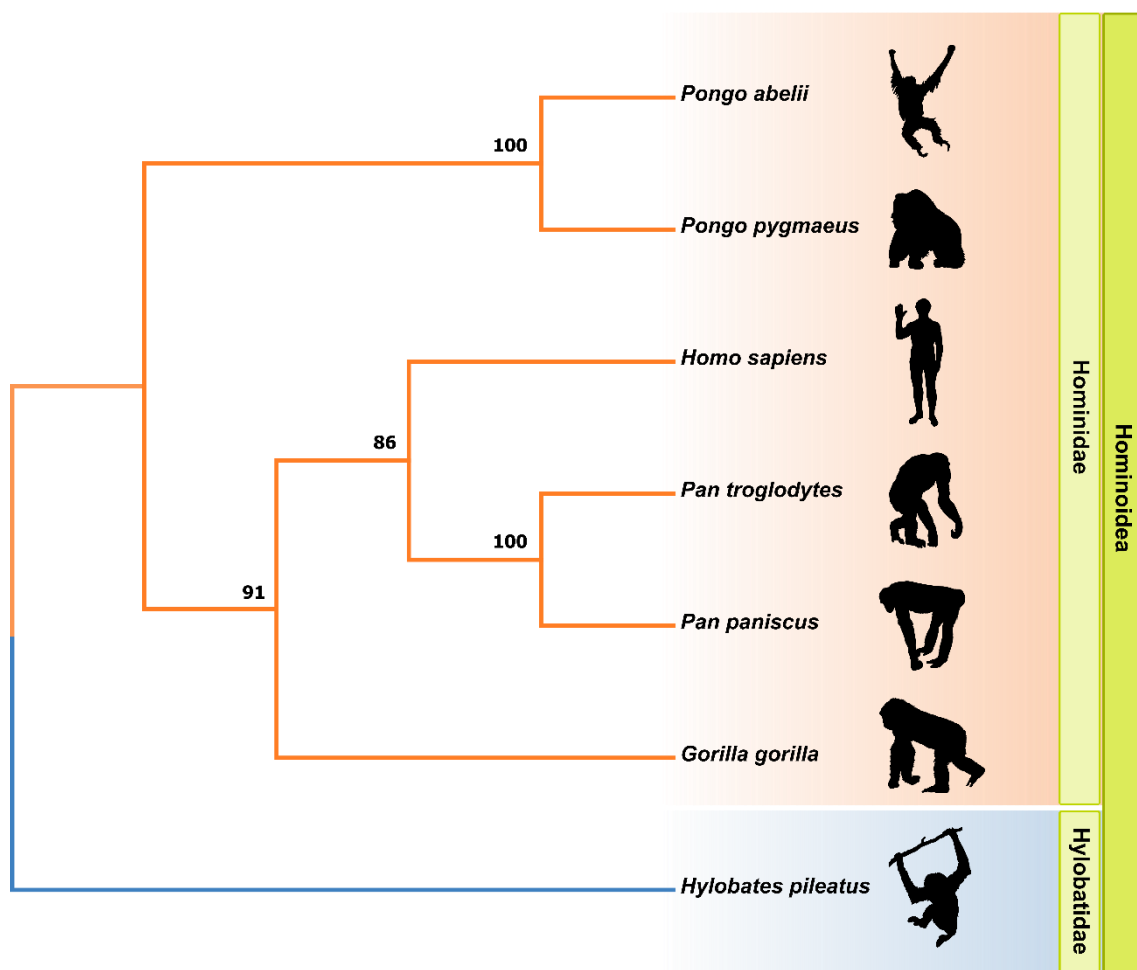
<i>SH2D1B</i> gene							
Site	FEL		FUBAR		SLAC		MEME
	Class	P-Value	Pr[$\alpha > \beta$]	Pr[$\alpha < \beta$]	Pr[dN/dS > 1]	Pr[dN/dS < 1]	
7	Purifying	0,001	0,993	0,004	1,000	0,007	0,670
13	Diversifying	0,035	0,001	0,988	0,095	0,978	0,020
14	Purifying	0,000	1,000	0,000	1,000	0,002	0,670
33	Purifying	0,000	1,000	0,000	1,000	0,001	0,670
35	Purifying	0,001	0,998	0,001	1,000	0,005	0,670
49	Diversifying	0,007	0,000	0,997	0,044	1,000	0,010
60	Purifying	0,013	0,987	0,002	0,996	0,024	0,670
61	Purifying	0,007	0,963	0,025	1,000	0,037	0,670
71	Diversifying	0,031	0,011	0,971	0,088	1,000	0,030
100	Purifying	0,005	0,968	0,022	1,000	0,037	0,670
101	Purifying	0,037	0,961	0,018	0,997	0,026	0,670
122	Purifying	0,026	0,958	0,026	0,997	0,041	0,670

Pr[$\alpha < \beta$] - Posterior probability of positive selection; **Pr[$\alpha > \beta$]** - Posterior probability of negative selection; **α** - Mean posterior synonymous substitution rate; **β** - Mean posterior non-synonymous substitution rate; **Pr[dN/dS > 1]** - The binomial probability that synonymous substitution in no greater than the observed value, with a probability of success (PS); **Pr[dN/dS < 1]** - The binomial probability that synonymous substitution in no less than the observed value, with a probability of success (PS); **dS** - synonymous substitution rate; **dN** - non-synonymous substitution rate.

Table S19: FEL, FUBAR, SLAC and MEME analyses from HyPhy software package for the *SH2D1A* gene in primates dataset.

<i>SH2D1A</i> gene							
Site	FEL		FUBAR		SLAC		MEME
	Class	P-Value	Pr[$\alpha > \beta$]	Pr[$\alpha < \beta$]	Pr[dN/dS > 1]	Pr[dN/dS < 1]	P-Value
3	Purifying	0,000	1,000	0,000	1,000	0,007	0,670
15	Purifying	0,001	1,000	0,000	1,000	0,004	0,670
16	Purifying	0,013	0,989	0,006	1,000	0,037	0,670
21	Purifying	0,001	1,000	0,000	1,000	0,005	0,670
26	Purifying	0,003	0,995	0,003	1,000	0,019	0,670
36	Purifying	0,003	0,992	0,005	1,000	0,012	0,670
38	Purifying	0,002	0,997	0,002	1,000	0,037	0,670
116	Purifying	0,023	0,995	0,001	0,995	0,048	0,670

Pr[$\alpha < \beta$] - Posterior probability of positive selection; **Pr[$\alpha > \beta$]** - Posterior probability of negative selection; **α** - Mean posterior synonymous substitution rate; **β** - Mean posterior non-synonymous substitution rate; **Pr[dN/dS > 1]** - The binomial probability that synonymous substitution is no greater than the observed value, with a probability of success (PS); **Pr[dN/dS < 1]** - The binomial probability that synonymous substitution is no less than the observed value, with a probability of success (PS); **dS** - synonymous substitution rate; **dN** - non-synonymous substitution rate.

**Figure S1:** Cladogram representation of the phylogenetic tree based on the concatenated *XIAP*, *SH2D1A* and *SH2D1B* genes of great apes dataset, created using IQ-TREE 2.0.7 with the Maximum Likelihood method. Bootstrap values above 70% are marked in bold.

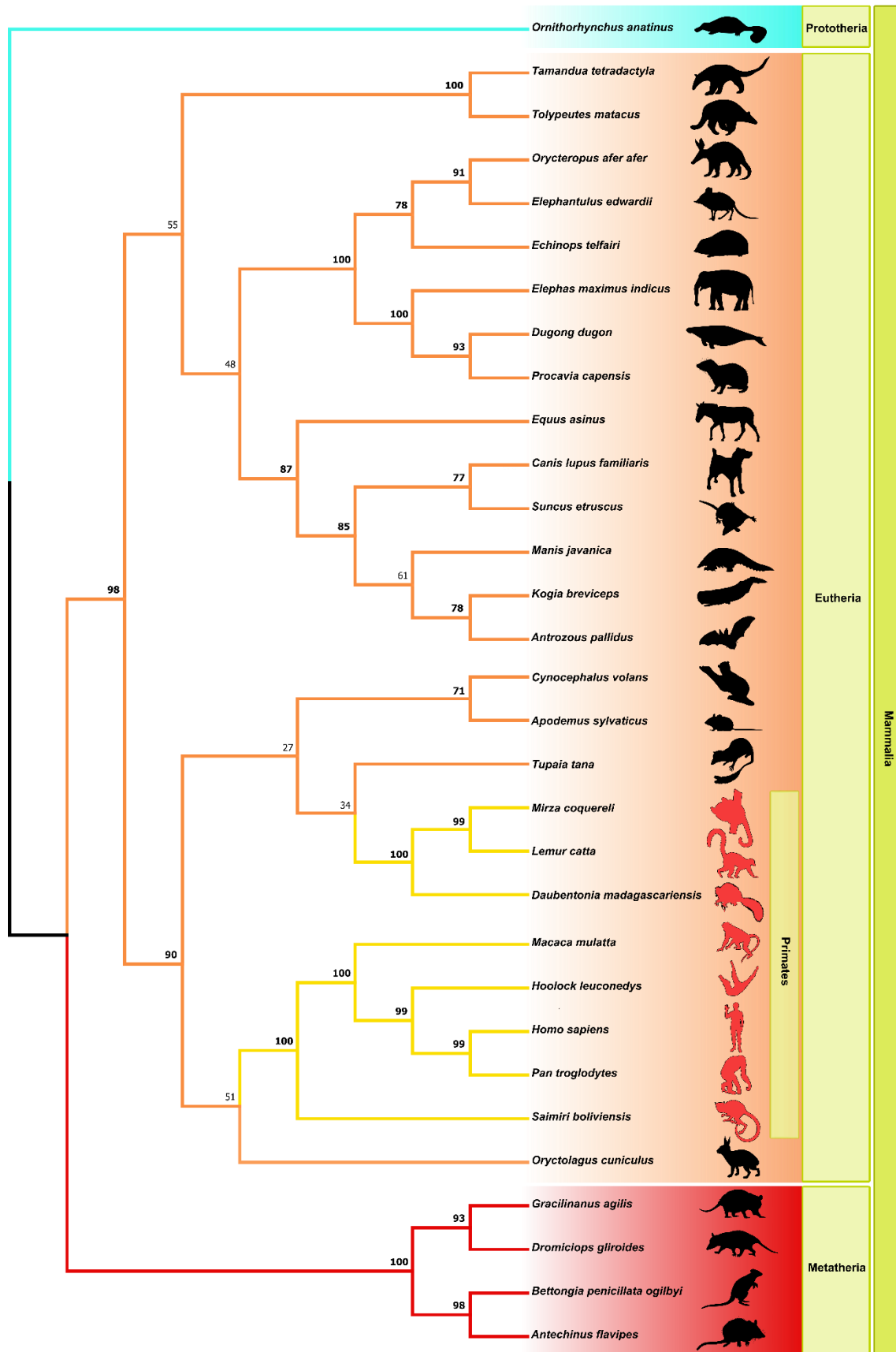


Figure S2: Cladogram representation of the phylogenetic tree based on the concatenated *XIAP*, *SH2D1A* and *SH2D1B* genes of Mammalia dataset, created using IQ-TREE 2.0.7 with the Maximum Likelihood method. Bootstrap values above 70% are marked in bold.

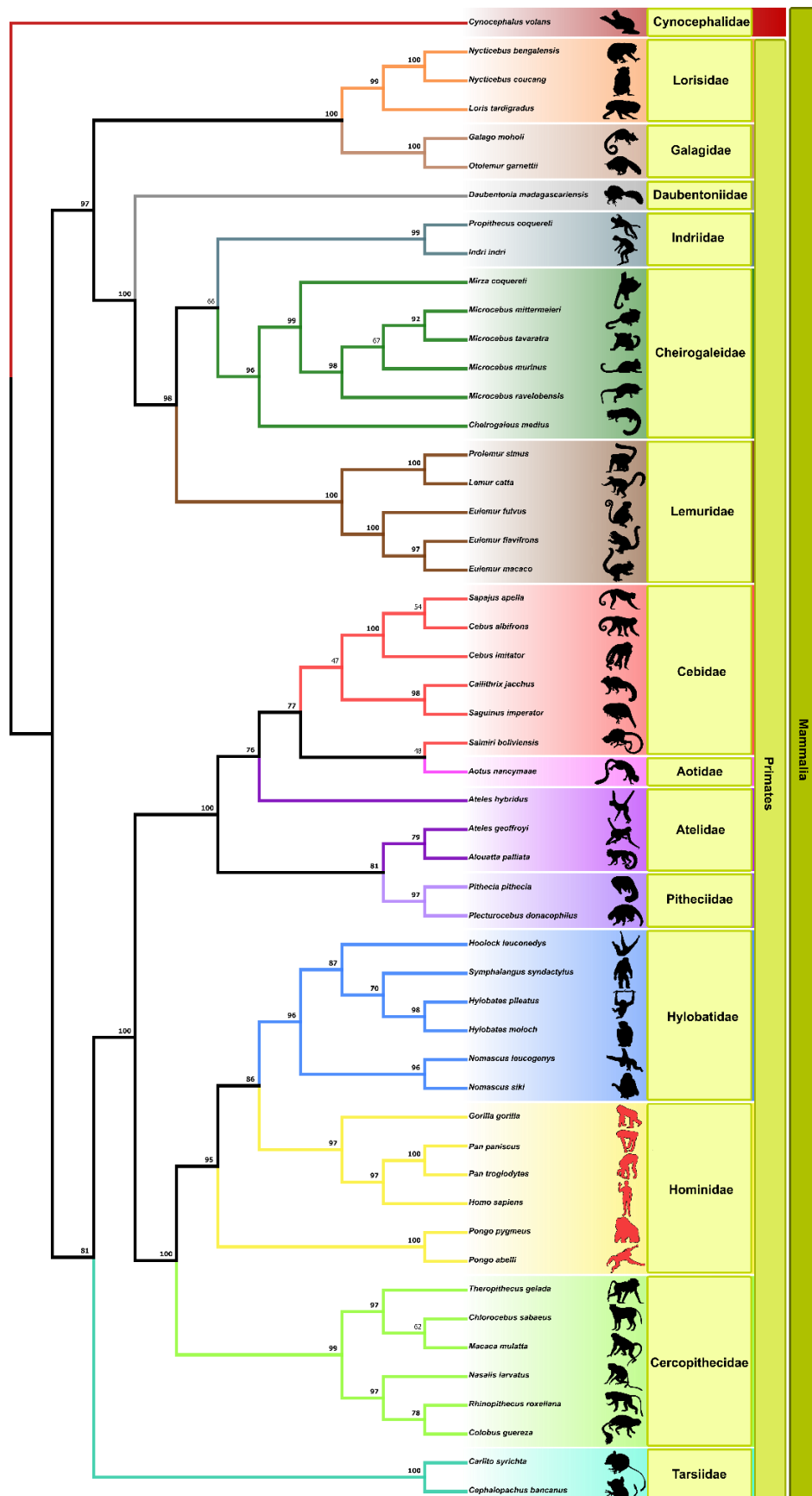


Figure S3: Cladogram representation of the phylogenetic tree based on the concatenated XIAP, SH2D1A and SH2D1B genes of primates dataset, created using IQ-TREE 2.0.7 with the Maximum Likelihood method. Bootstrap values above 70% are marked in bold.

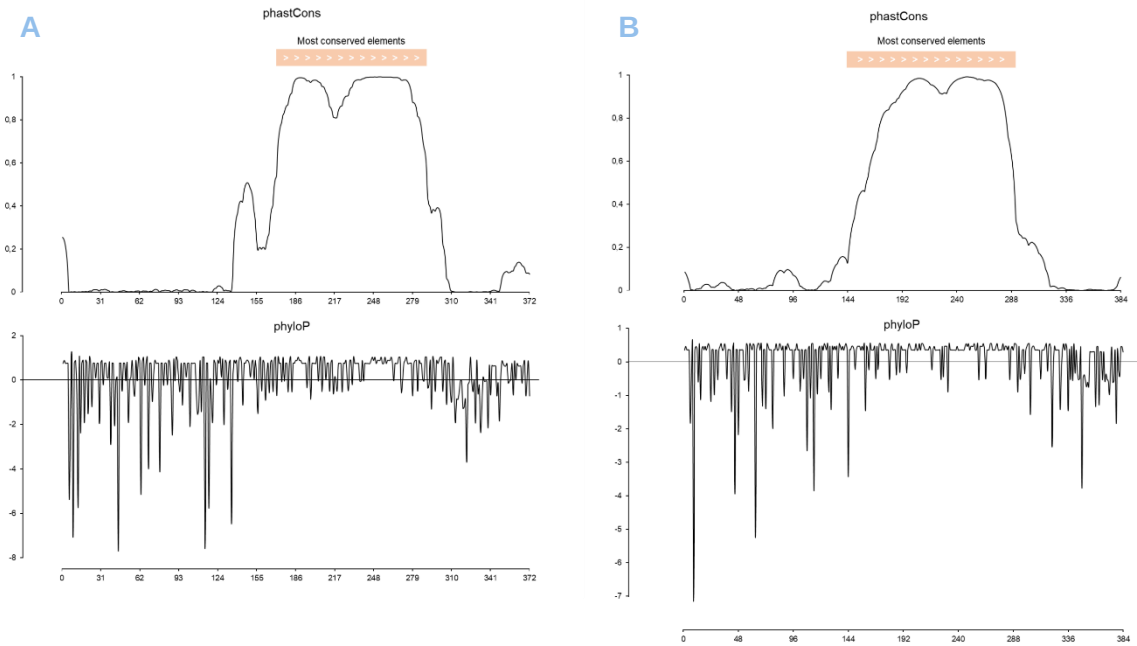


Figure S4: Graphs showing the PhyloP and PhastCons scores for the *SH2D1A* gene in the primate (A) and Mammalia (B) datasets. In PhyloP, negative sites are predicted to be rapidly evolving and positive sites are predicted to be conserved. In PhastCons, the score varies between 0 and 1, from least to most conserved, respectively.

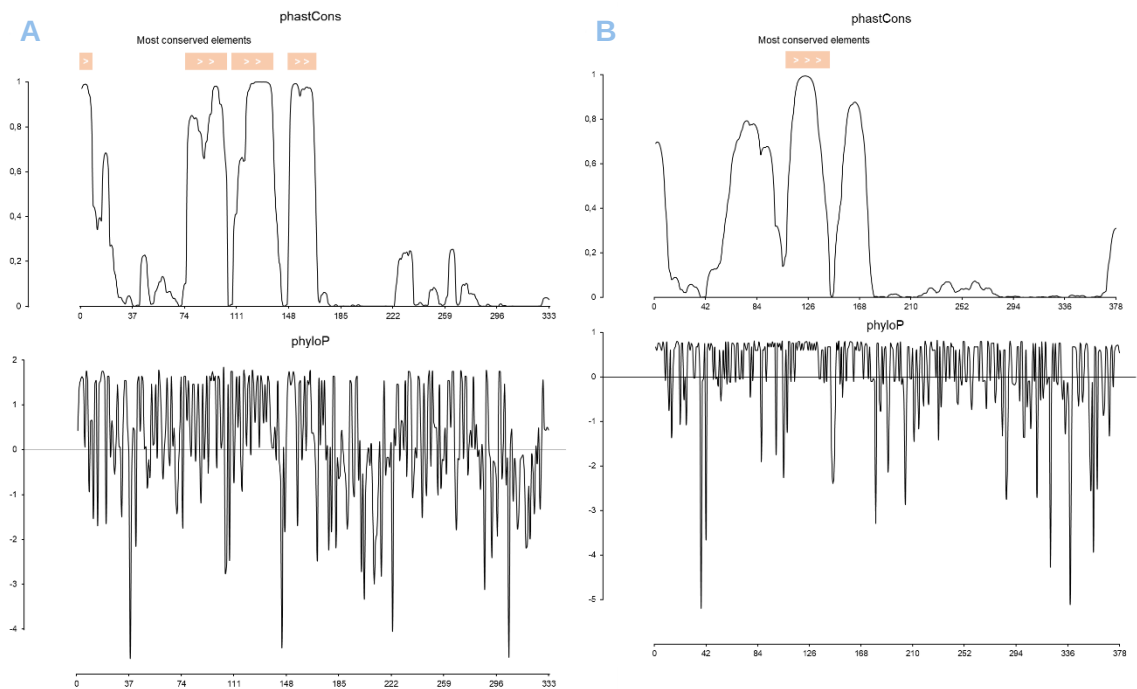


Figure S5: Graphs showing the PhyloP and PhastCons scores for the *SH2D1B* gene in the primate (A) and Mammalia (B) datasets. In PhyloP, negative sites are predicted to be rapidly evolving and positive sites are predicted to be conserved. In PhastCons, the score varies between 0 and 1, from least to most conserved, respectively.

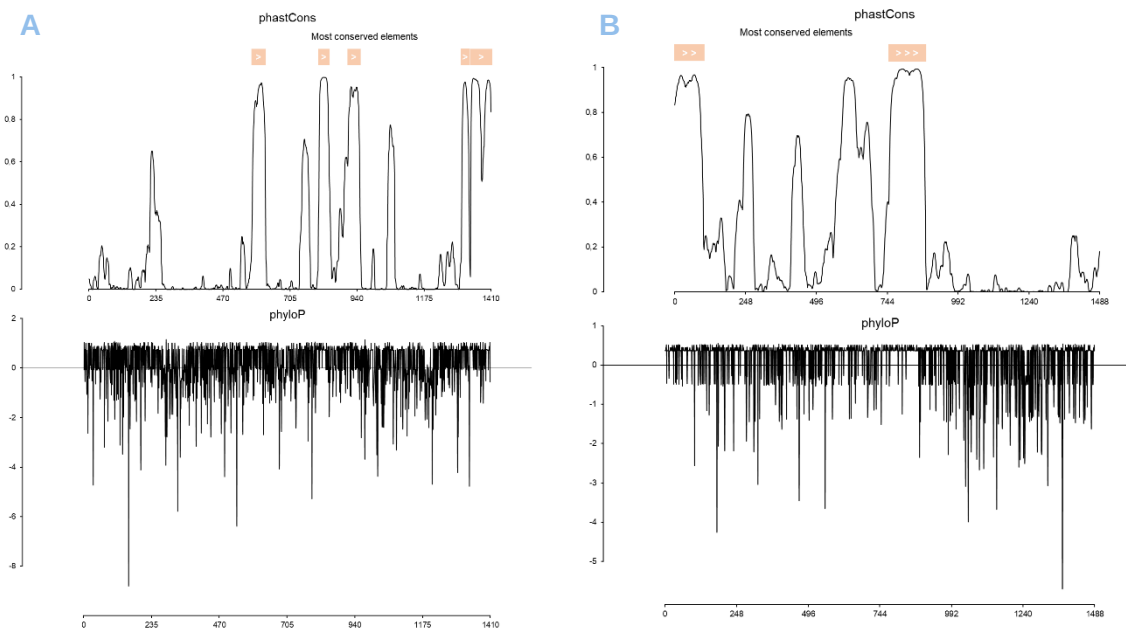
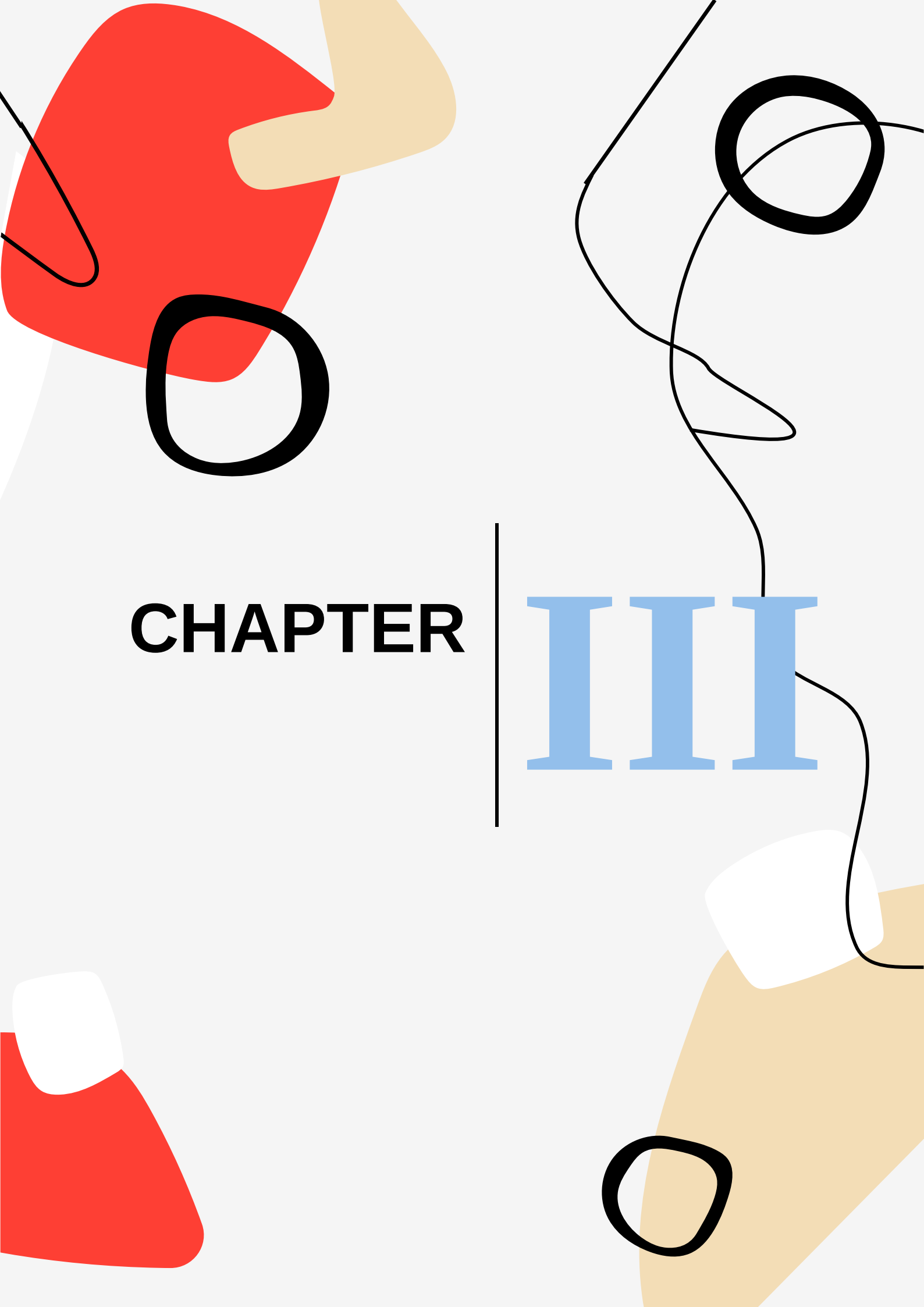
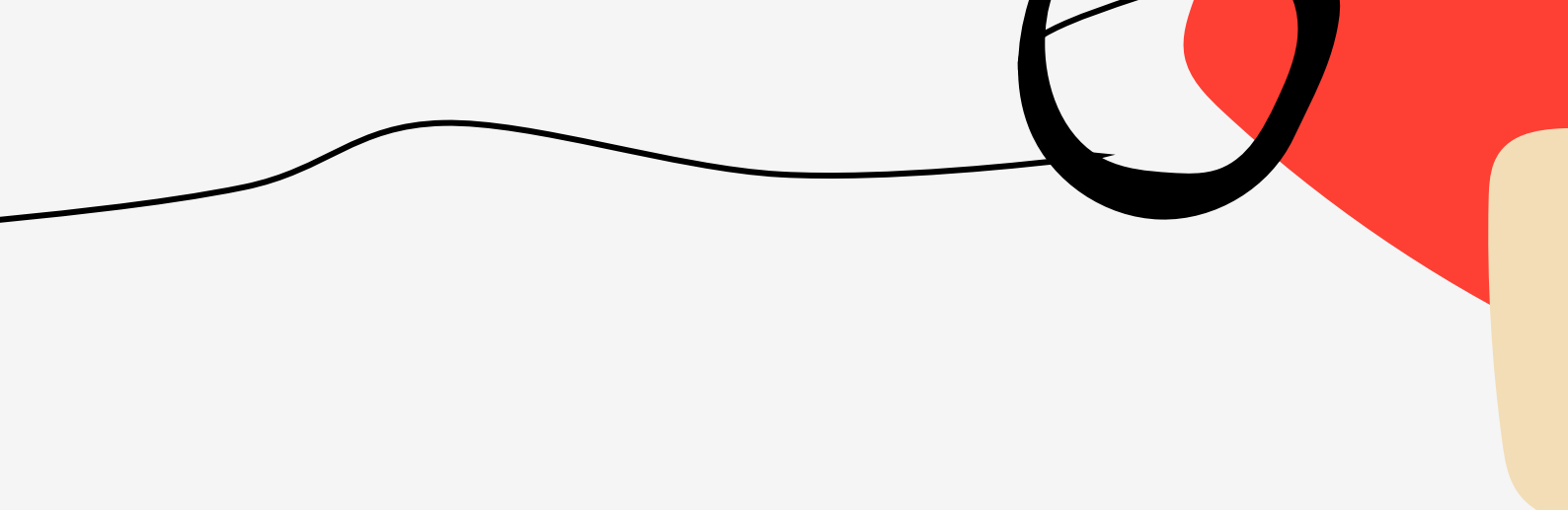


Figure S6: Graphs showing the PhyloP and PhastCons scores for the *XIAP* gene in the primate (A) and Mammalia (B) datasets. In PhyloP, negative sites are predicted to be rapidly evolving and positive sites are predicted to be conserved. In PhastCons, the score varies between 0 and 1, from least to most conserved, respectively.



CHAPTER

III

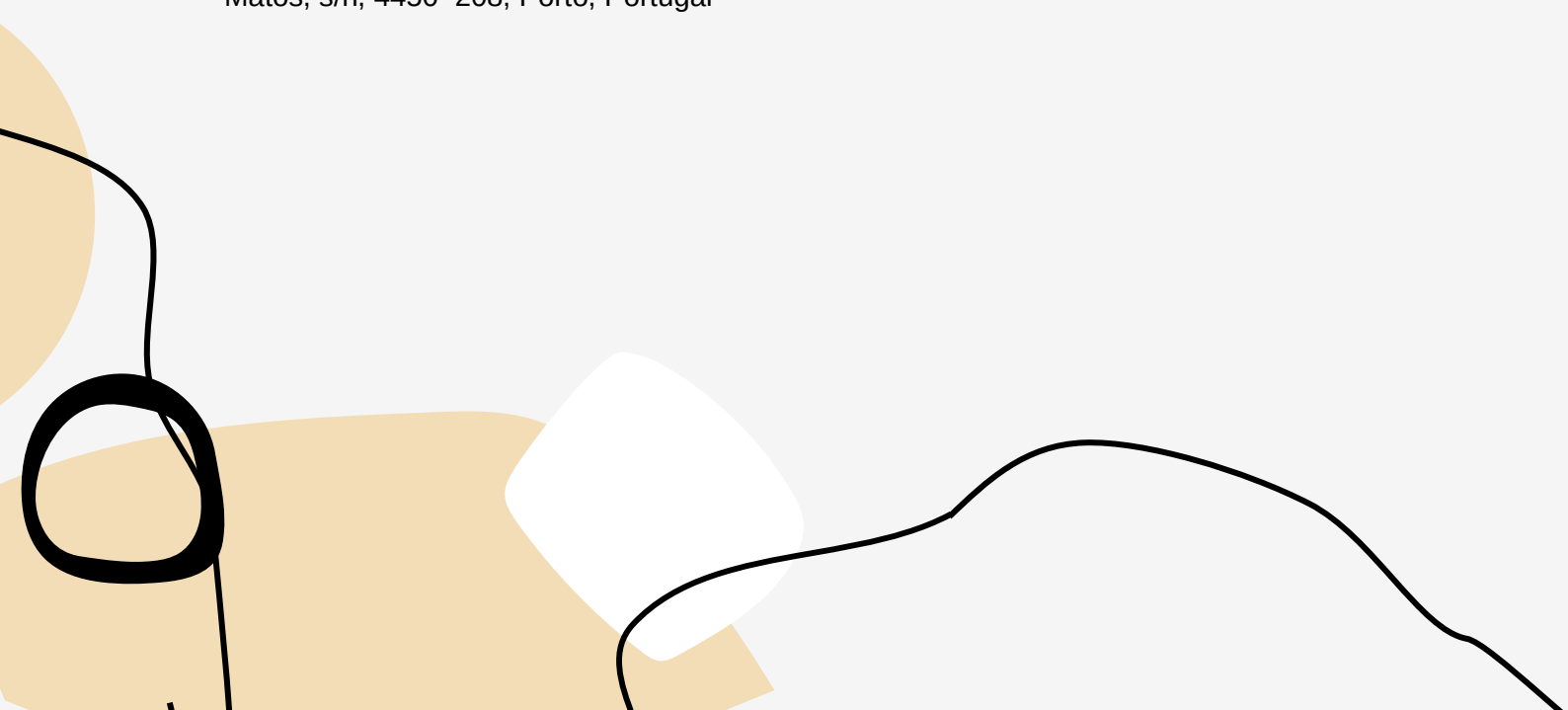


***In silico* analysis of non-synonymous SNPs and their implications on X-linked lymphoproliferative syndrome type 1 (XLP1)**

Nuno Sousa^{1,2}, Luana Ramos^{1,2}, Agostinho Antunes^{1,2}

¹ Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre, 4169-007, Porto, Portugal

² CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, 4450-208, Porto, Portugal



3.1. Abstract

In genomics and personalized medicine, researchers rely on automated methods to analyze single nucleotide polymorphisms (SNPs), particularly nonsynonymous SNPs (nsSNPs). These nsSNPs play dual roles as genetic markers and disease-related variants, often affecting protein stability and interactions. The last decade has witnessed significant progress in understanding the genetic basis of individual reactions to medications, crucial for personalized medicine. In response, *in silico* prediction tools have been developed to classify harmful nsSNPs based on sequence data and physicochemical properties. Focusing on the *SH2D1A* gene associated with X-linked lymphoproliferative disease, we employed molecular dynamics simulations to comprehensively investigate the effects of deleterious nsSNPs, offering insights for disease diagnosis and personalized treatments. Missense SNPs linked to the *SH2D1A* gene were retrieved from ClinVar and HGMD using the human SH2D1A protein sequence as basis for analyses. We resorted to PredictSNP1.0 to assess the impact of the nsSNPs using nine prediction tools and to I-Mutant 2.0 and MUpro to explore stability changes. Further characterization involved conservation, physicochemical changes, and structural and functional feature analyses. The native 3D structure of SH2D1A was obtained from the Protein Data Bank, and missing residues were added and validated. *In silico* mutagenesis was performed for selected SNPs. Molecular dynamics simulations were carried out, and the data analyzed for various structural parameters. In total, we analyzed 60 missense SNPs and identified 23 highly deleterious and unstable SNPs associated with the *SH2D1A* gene. Structural analysis revealed that these mutations typically affect conserved amino acids, with changes in physicochemical properties impacting stability. Molecular dynamics simulations showed varying effects on protein stability, flexibility, and structure. Complex relationships between secondary structures, hydrogen bonds, and solvent exposure influenced the mutations' impact. Further experimental validation is essential and future research should explore potential drug targets for the treatment of these XLP-related mutations. Personalized medicine may benefit from tailored interventions based on genetic profiles.

Keywords: X-linked Lymphoproliferative Syndrome, SH2D1A, Molecular Simulation Dynamics, nonsynonymous SNPs

3.2. Introduction

In the field of genomics and personalised medicine, researchers face a major challenge in the extensive analysis of data obtained by whole genome sequencing [1,2]. To handle this complexity, the use of automated methods for annotation and prioritisation of variants has become essential. In this context, computational analysis of common single nucleotide polymorphisms (SNPs) has emerged as a powerful and easily accessible tool to drive the progress of personalised genomics [3–5].

SNPs, particularly nonsynonymous SNPs (nsSNPs), play a dual role as genetic markers in genetic mapping and as potential functional polymorphic variants involved in complex diseases and drug responses [6]. Many nsSNPs associated with human diseases exert their effects through changes in structural stability or by directly disrupting sites of interaction with ligands [7].

In the last decade, research aimed to comprehensively understand the genetic basis of variability in response to medicines in different individuals has experienced a remarkable upsurge, which is fundamental to the development of personalised medicine [8].

There is an urgent need for efficient methods to classify nsSNPs as harmful or disease-causing. Hence, sophisticated and affordable *in silico* prediction tools have been developed. These tools evaluate the harmful potential of a nsSNP based on sequence information, structural attributes and the physicochemical properties of amino acids [9]. Each algorithm has its own advantages and limitations, but they all contribute to the overall goal of classifying nsSNPs as pathogenic or neutral [10].

Molecular dynamics (MD) simulation is a computational technique used to analyse the physical behaviour of atoms and molecules over time [11]. It is widely employed in the field of structural biology to investigate the dynamics and thermodynamics of biomolecules, including proteins and nucleic acids. So, comprehending the three-dimensional (3D) structure of a gene product is crucial to assess the impact of mutations on disease development. Using the 3D structure of proteins, MD simulations can predict the impact of nsSNPs on the dynamics of protein structure [12]. Currently, many macromolecular structures are available in the Protein Data Bank (PDB), which serve as valuable resources for mapping mutations, creating 3D models using homology modelling techniques and running MD simulations [13].

These tools are particularly important to explore the genetic basis of poorly studied diseases, including the X-linked lymphoproliferative (XLP) syndrome. Mutations in the

SH2D1A gene (Figure 1) lead to the development of XLP syndrome type 1 (XLP1), which mainly affects men. In this rare immune disease, the immune response to the Epstein-Barr virus (EBV) is impaired [14].

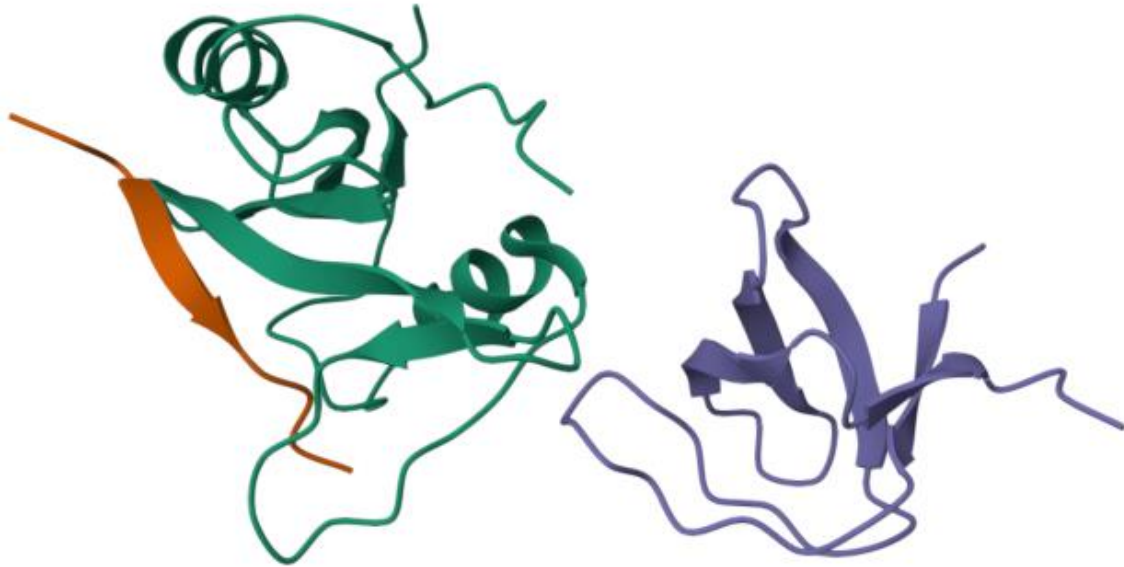


Figure 1: Crystal structure of SH2D1A/FYN/SLAM in ternary complex (ID: 1M27). In orange the SLAM protein, in green the SH2D1A protein and in purple the FYN protein. Extracted from the Protein Data Bank (PDB) [15].

SH2D1A mutations lead to reduced or absent invariant natural killer T cells (iNKT), impaired cell death mechanisms, dysregulated cytokine production and an increased risk of lymphoma, especially B-cell lymphoma such as Burkitt's lymphoma [16]. The severity and timing of symptoms vary depending on the type and location of the *SH2D1A* mutation, highlighting the need for individualised approaches for diagnosis and treatment [17] (see Chapter I). XLP1 highlights the complicated relationship between genetics and immunity and, consequently, the importance of understanding genetic factors in immune diseases for personalised medical care.

The aim of this study is to comprehensively investigate the functional consequences of deleterious nsSNPs in the *SH2D1A* gene. By using an integrated *in silico* framework that includes prediction methods and MD simulations, we intend to find out how these mutations affect the structure and function of the SH2D1A protein. Ultimately, our results may provide valuable insights for disease diagnosis and the development of personalised therapeutic strategies for *SH2D1A*-related diseases.

3.3. Material and Methods

Retrieval of SNPs and SH2D1A protein sequence

To comprehensively investigate the effects of amino acid alterations within the SH2D1A protein underlying the phenotypes associated with XLP, we first performed a detailed identification and extraction of missense SNPs, specific to the *SH2D1A* gene, involved in the disease. This crucial step was taken resorting to two widely recognised genetic variant databases: the ClinVar database, managed by the National Center for Biotechnology Information (NCBI) [18], and the Human Gene Mutation Database (HGMD) [19]. These databases are valuable scientific resources due to their expertise in classifying genetic variants and are considered highly reliable sources of genetic information. After extracting SNPs from these databases, we associated each of the extracted SNPs with identifiers from the NCBI dbSNP (Single Nucleotide Polymorphism) database [20]. This ensures accurate tracking and correlation of the genetic variations associated with the *SH2D1A* gene with the corresponding entries in this database.

In addition, we obtained the human protein sequence corresponding to the longest transcriptional variant of the SH2D1A protein from the UniProt database [21] (UniProt ID: O60880). This served as an essential basis for our subsequent investigations into the functional consequences and potential impact of the identified missense SNPs within the *SH2D1A* gene.

Missense SNPs analysis

Prediction of most deleterious SNPs

To assess changes in protein function resulting from missense SNPs, we used PredictSNP1.0 web server [22], a powerful predictive tool from the field of computational biology. PredictSNP1.0 acts as a unified classifier that harmoniously integrates information from nine of the most popular prediction tools - PolyPhen-1, PolyPhen-2, SIFT, MAPP, SNAP, PredictSNP, PANTHER, PhD-SNP and nsSNPAnalyzer. This comprehensive approach allows for a versatile assessment of the impact of genetic variants on protein function, since each of these tools offers unique approaches to predicting the effects of genetic mutations on proteins [23].

PolyPhen-1 and PolyPhen-2 use machine learning algorithms to predict the effects of mutations on protein structure and function and classify mutations as "benign" or "harmful" [24,25]. SIFT evaluates mutations based on the conservation of amino acids

at similar positions in related proteins and predicts whether a mutation is "tolerated" or "intolerant" [26]. MAPP considers various aspects of protein structure and biochemistry to predict the effects of mutations [27]. SNAP evaluates mutations based on sequence information, protein structure and function [28]. PredictSNP, on the other hand, uses machine learning to predict the functional impact of mutations by integrating various genetic and structural features [22]. PANTHER classifies mutations based on protein family evolution and predicts whether a mutation is "likely benign" or "likely harmful" [29,30]. PhD-SNP uses a set of decision rules based on protein structure and function to classify mutations as "neutral" or "non-neutral" [31]. Finally, the nsSNPAnalyzer evaluates mutations based on sequence and structural analyses and identifies mutations that may affect protein function [32].

By integrating these different approaches, PredictSNP1.0 provides a more comprehensive and reliable assessment of the effects of genetic mutations, enabling a deeper understanding of the functional implications of genetic variants [33]. SNPs showing deleterious effects in at least six prediction tools were accepted for protein stability analyses [34].

Predicting alterations of stability

The SNPs selected in the previous steps were analysed by I-Mutant 2.0 [35] and MUpro [36], two automated web servers based on neural networks that play an essential role in investigating changes in protein stability due to nsSNP mutations. Both tools employ machine learning principles with the Support Vector Machine (SVM) [37].

In the case of I-Mutant 2.0, it can predict variations in free energy ($\Delta\Delta G$) and classify them into two distinct categories: an increase in protein stability ($\Delta\Delta G > 0$ kcal/mol) or a decrease in stability ($\Delta\Delta G < 0$ kcal/mol) [35]. The results provided by MUpro are expressed in terms of confidence values ranging from -1 to 1, in which a confidence value below 0 indicates a reduction in stability, while a confidence value above 0 indicates an increase in stability [36]. Mutations demonstrating instability effects in at least three out of the four results obtained through the methods will be chosen for subsequent analysis.

Characterization of the selected SNPs

Once the most deleterious and instability-causing polymorphisms had been uncovered, their comprehensive characterisation was carried out, which comprises the assessment of various aspects - the degree of conservation of the amino acids involved, changes in

the physicochemical properties of the protein and the identification of structural and functional pathogenic features. Three different bioinformatics tools were used to facilitate this multi-faceted analysis: the ConSurf server [38], the Have Your Protein Explained (HOPE) [39] server and MutPred2 [40].

The ConSurf server measures the evolutionary conservation of amino acids in proteins based on phylogenetic relationships between similar sequences. The conservation value, which ranges from 1 to 9, indicates the degree of conservation of amino acids during evolution, with 9 representing maximum conservation and 1 representing low conservation [38,41].

The Have Your Protein Explained (HOPE) server is an advanced tool that provides a complete analysis of the molecular structures, mutations and molecular properties of human proteins. This state-of-the-art automated resource sheds light on mutation-induced changes in structural and functional stability [39]. It integrates reliable data from various sources, including UniProt, WHAT IF [42] and Distributed Annotation System (DAS) [43,44].

MutPred2 is a web server that plays an essential role in uncovering structural and functional knowledge about protein variants. MutPred2 is specifically designed to categorise amino acid substitutions in humans as pathogenic or benign and extends its predictive capabilities to assess more than 100 different protein properties. This comprehensive analysis enables the inference of potentially pathogenic molecular mechanisms and contributes to the understanding of the molecular basis of disease [40].

With these tools, we obtain a comprehensive understanding of the complex interrelationships between genetic variations, protein structure and function, thus gaining new insights into the molecular basis of diseases and potential therapeutic approaches.

Loop modelling, validation and mutagenesis

To assess the effects of mutations on the molecular dynamics of the protein, we first extracted the native three-dimensional structure of the human SH2D1A protein from the Protein Data Bank [15] (PDB ID: 1KA6). Protein structures obtained by X-ray diffraction and nuclear magnetic resonance (NMR) have advantages over predicted protein structures because they are experimentally validated and represent the real structure of the protein [45,46]. However, structures obtained by these direct methods are often not complete due to technical complications or structural flexibility [47]. Therefore, the missing residues of the protein were added to the native structure with PyMol v2.5 [48]

and validated with the ModLoop web server [49]. Based on the complete structure of the native protein, an *in silico* mutagenesis was performed for the selected SNPs using the Swiss PDB Viewer v4.1 [50].

Molecular dynamics simulation

We used the GROMACS 2023.2 package [51] to study the dynamics and equilibrium of native and mutant SH2D1A proteins. Molecular dynamics simulation (MDS) generated time trajectories revealing conformational changes in the proteins after mutation. We used the AMBER99SB-ILDN force field with the TIP3P water model [52]. After solvation, we neutralised the system with Na⁺ and Cl⁻ ions using GENION. After neutralisation, we optimised the energy to avoid improper conformations and sterics by using the "steepest descent" algorithm to ensure that the three-dimensional structures were free of obstacles [53]. We then equilibrated the system in two regimes: NVT (constant number of particles, volume and temperature) and NPT (constant number of particles, pressure and temperature), both with 50,000 steps. We then ran the MDS for 25 nanoseconds at a temperature of 300 K, generating trajectories every 0.2 picoseconds.

The resulting trajectories were used to analyse important GROMACS parameters such as the RMSD (Root Mean Square Deviation) to assess structural stability over time, the RMSF (Root Mean Square Fluctuation) to assess the flexibility of individual residues, the Rg (Radius of Gyration) to measure protein compaction, the number of hydrogen bonds to assess protein stability, the solvent accessible surface area (SASA) to analyse the area of the protein available for interaction with solvents [54–56] and the DSSP algorithm to analyse the secondary structure elements of the proteins [57].

These trajectories were also used to generate diagrams through XMGrace v5.1.19 [58], providing detailed information about the structural differences that occurred during the simulation and a detailed look at the dynamics of these proteins. These analyses provided crucial insights into the structural responses of proteins to mutations and the changes in their dynamics over time.

The data extracted from the MDS analyses were subjected to statistical analysis using Excel 2016. This analysis encompassed the calculation of mean values, standard deviations, differences between the means of mutant and wild-type proteins, and *p*-values for the disparities of the mutants through the two-tailed T-test.

3.4. Results

Missense SNPs collection

In order to extract only the SNPs that are linked to the appearance of the XLP phenotype and result in a change on the encoded amino acid (missense), a total of 60 SNPs were extracted from the ClinVar and HGMD databases (Table S1).

Analyses and selection of the most deleterious SNPs

The different algorithms used to analyse the harmful potential of the missense SNPs on the PredictSNP1.0 webserver yielded different results (see Table S2). Of the 60 SNPs originally identified, PredictSNP classified 46 as harmful, while MAPP detected 53 deleterious SNPs. PhD-SNP identified 42 SNPs as deleterious, PolyPhen-1 labelled 41 as deleterious and PolyPhen-2 identified 45 with deleterious potential. Similarly, SIFT classified 48 SNPs as deleterious, while SNAP pointed to 37. It is noteworthy that both PANTHER and nsSNP Analyzer did not produce any results.

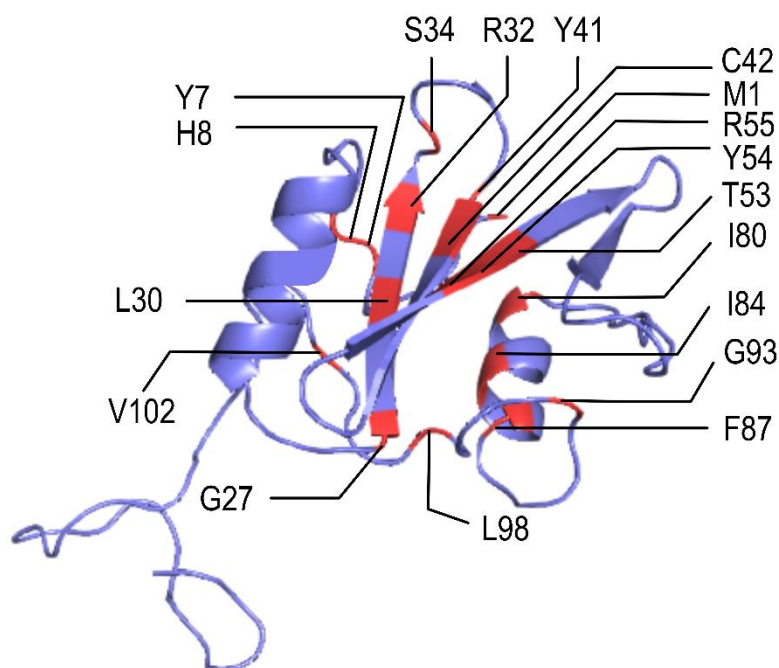


Figure 2: Protein structure of SH2D1A modelled with the sites affected by the SNPs highlighted in red.

To refine the selection, a comparative selective approach centred on SNPs with deleterious signals was used in at least six of the aforementioned algorithms, resulting in the identification of 40 SNPs with harmful effects. These SNPs were subjected to stability analyses using the I-mutant2.0 and MUpro web servers (see Table S3). Of these 40 SNPs, 23 showed structural instability and were selected for further analyses.

Prior to performing MDS, the 23 SNPs selected from previous analyses were characterised (Figure 2). This characterisation of their conservation level, physicochemical properties and potential pathogenic mechanisms using the ConSurf, HOPE and MutPred2 web servers are presented in Table S4. Remarkably, the substitutions involved amino acids that are typically buried within the protein and exhibit a high degree of conservation. Approximately 60% of the selected mutations resulted in the placement of smaller amino acids compared to their wild-type counterparts. In addition, about 48% of the mutant amino acids were less hydrophilic than the wild type, while 24% had increased hydrophobicity.

Molecular Dynamics of native and mutant SH2D1A protein

Despite the absence of certain amino acids in the protein structure, they could be effectively positioned and validated by using PyMol and ModLoop software. In addition, in silico mutagenesis was performed successfully for all SNPs using the Swiss PDB Viewer. To gain a deeper understanding of how the mutant SH2D1A proteins behave at the structural level, from the output trajectory files of the 25-ns molecular dynamics simulation, we created plots to analyse various parameters, including Rg (gyration radius), RMSF (root mean square fluctuation), SASA (solvent-accessible surface area), RMSD (root mean square deviation) and the total number of intramolecular hydrogen bonds. This analysis allowed us to compare the parameter results of the MD simulations for the wild-type protein (WT) with those of the 23 previously selected mutants given a detailed understanding at the atomic level of how the amino acid changes affect the stability and functionality of these proteins.

The deviation of the atomic positions (RMSD) between the initial and the often-simulated protein was calculated for the protein backbone (Figure 3). In general, we observed statistically significant differences in the variances between the native WT and mutant proteins, except for mutants H8D and R32T, which had *p*-values of 0,527 and 0,571, respectively (Table S5). Notably, the Y54N mutant structure exhibited the highest average structural deviation for backbones ($\bar{x}$ = 0,312), contrasting with the lower deviation values for WT ($\bar{x}$ = 0,576) and the other mutants (Figure S1, S2, S3 and S4).

To evaluate the dynamic behaviour of all 23 structures, we performed RMSF calculations in carbon alpha atoms for both the native and mutant proteins (Figure 4). The C α -RMSF values were then plotted based on the residues within the structures (Figure S5, S6, S7 and S8). Except for mutants Y41N, C42Y, V102G, R32T, Y7C, G27D, R55Q, I84T, the

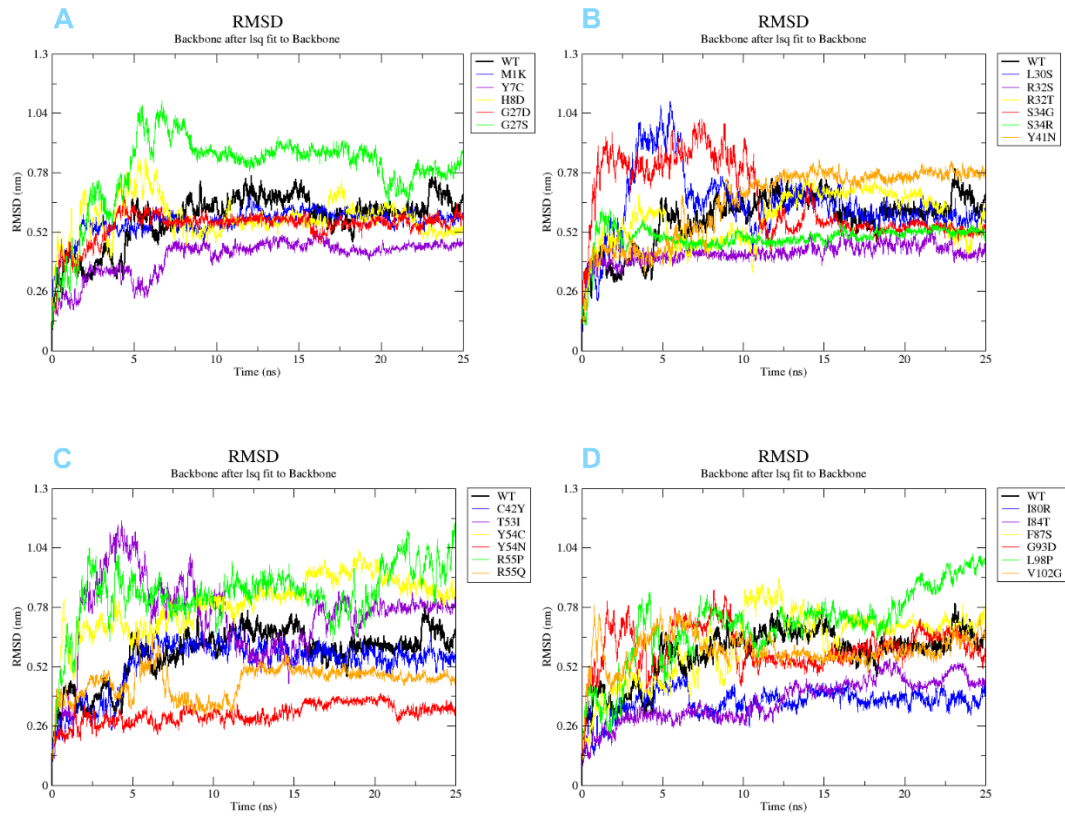


Figure 3: Analysis of the evolution of backbone RMSD of all mutants and native (WT) SH2D1A protein at 25 ns.

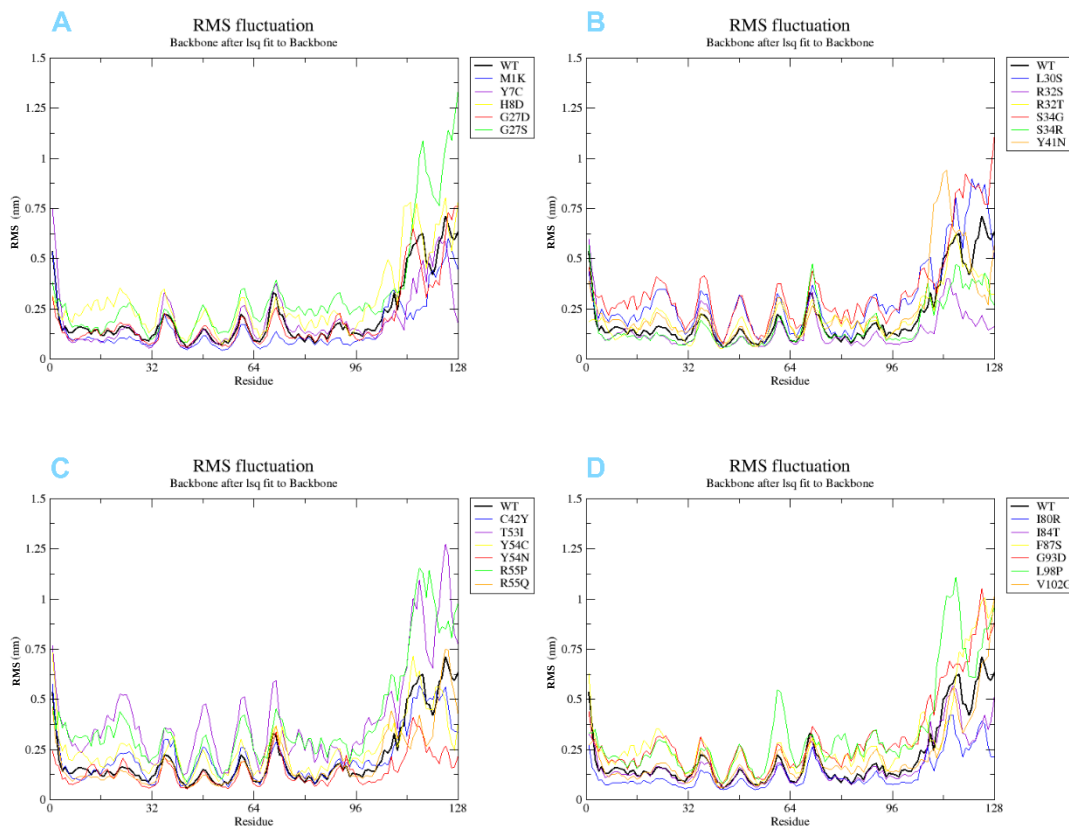


Figure 4: Analysis RMS fluctuation of carbon alpha on all mutants and native (WT) SH2D1A protein.

average C α -RMSF values for the remaining mutants were significantly different from those of the native SH2D1A protein (Table S6).

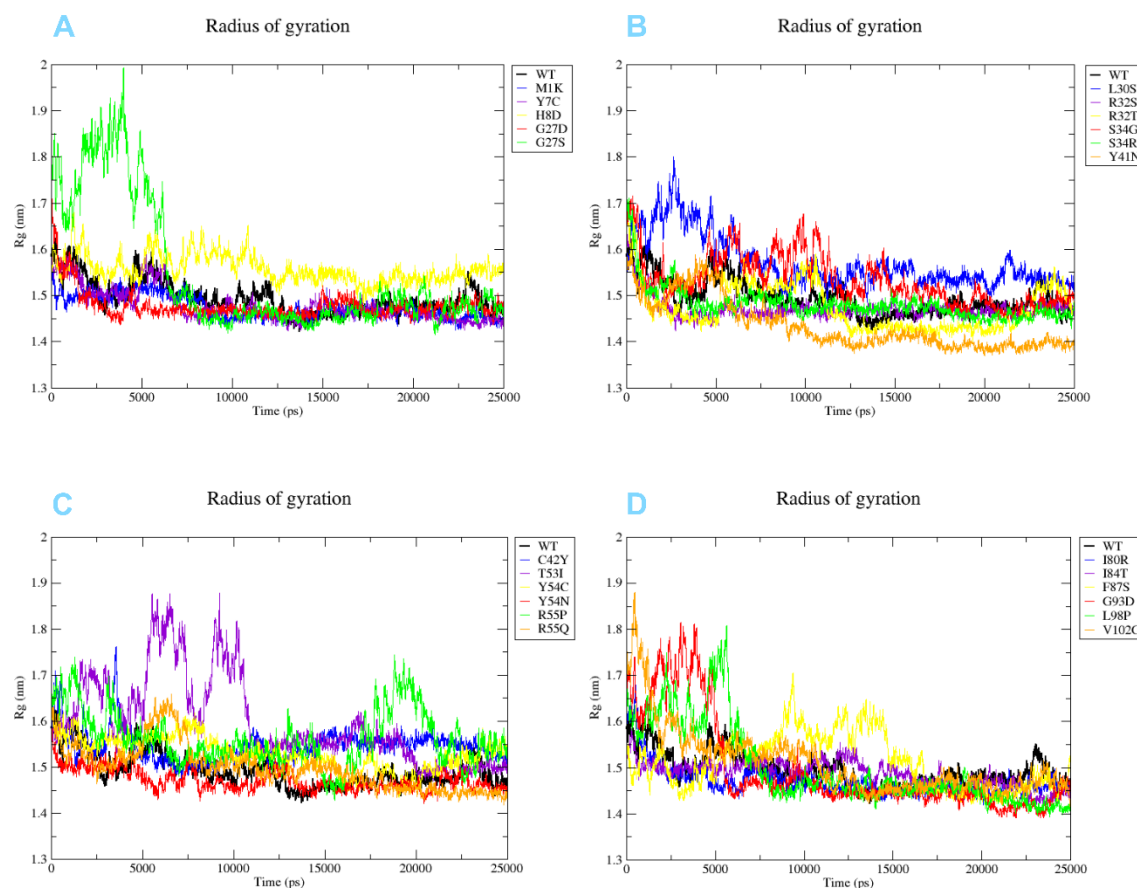


Figure 5: Analysis of radius gyration (Rg) on the backbone of all mutants and native (WT) SH2D1A protein.

Analysis of the gyration radius (Rg), which provides information about the dimensions and conformational geometry of the protein (Figure 5), showed that all mutants had average Rg values that were significantly different from the wild-type structure (Table S7), apart from the G93D mutant (Figure S9, S10, S11 and S12).

To understand the flexibility of both native and mutant SH2D1A proteins, we calculated the intramolecular hydrogen bonds (Figure 6). In general, all mutations had significant differences in the average number of hydrogen bonds compared to the wild-type protein, with most mutants having more bonds (Table S8). The exceptions were mutant proteins 34G, G27S, H8D, C42Y, which had the same average number of hydrogen bonds, and mutant R55P, which had fewer bonds (Figure S13, S14, S15 and S16).

Solvent accessible surface area analysis (SASA) was performed to evaluate the available surface for binding interactions (Figure 7). The average SASA value for the native protein was 85 nm² (Figure S17, S18, S19 and S20). Most mutants had higher

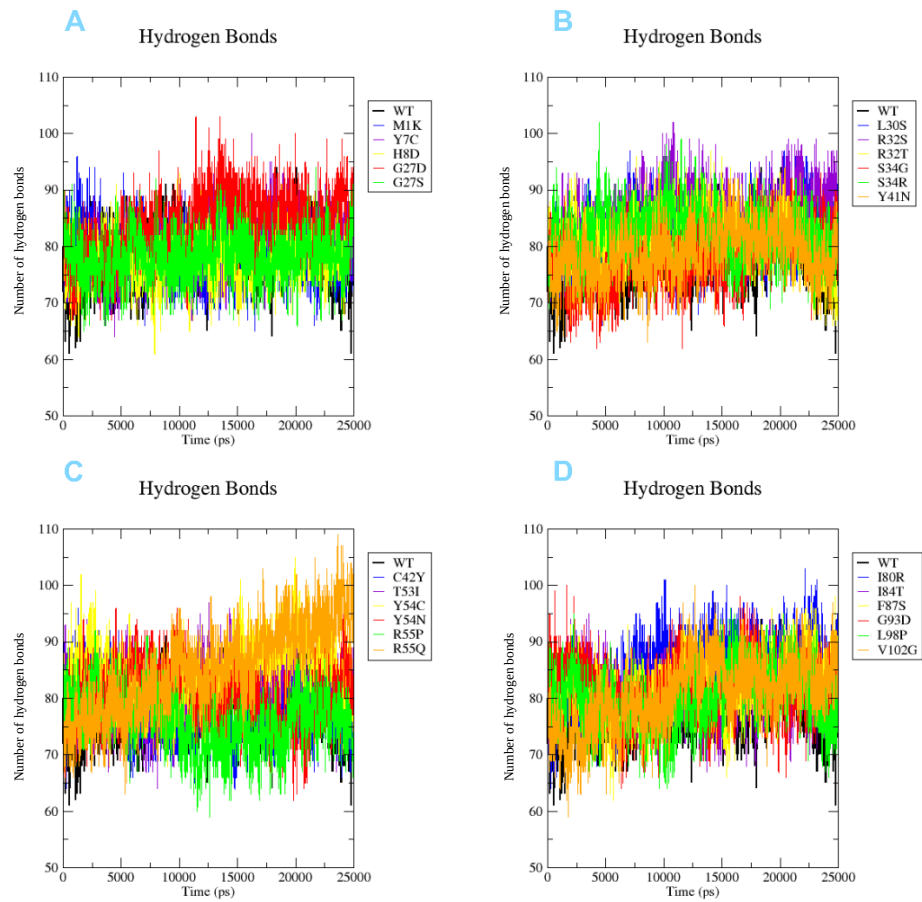


Figure 6: Analysis of intramolecular hydrogen bonds on all mutants and native (WT) SH2D1A protein.

III

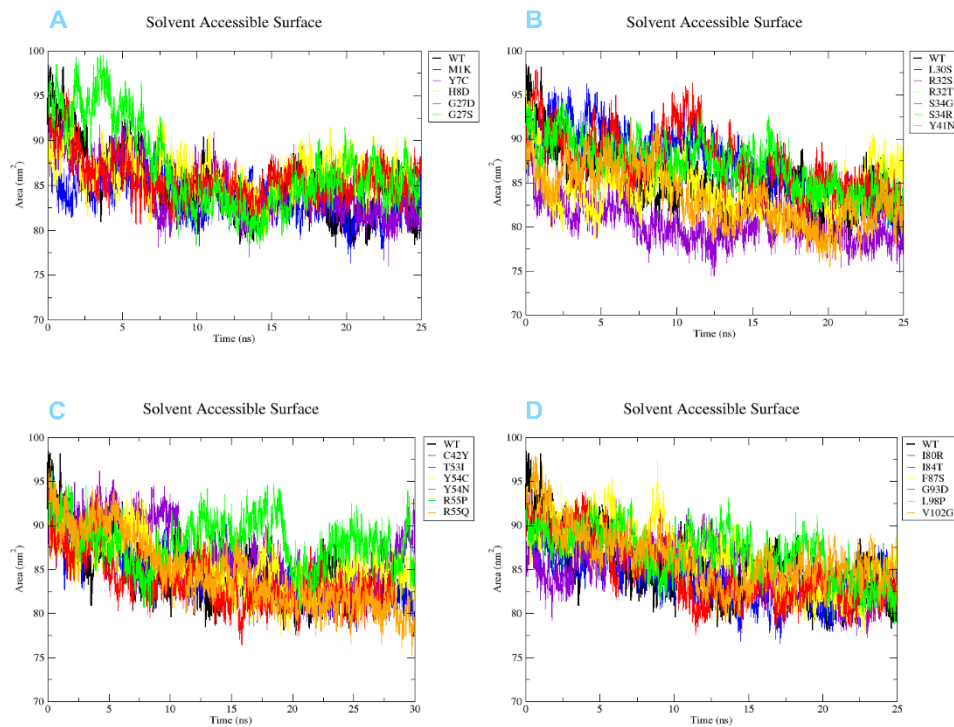


Figure 7: Analysis of solvent accessible surface area (SASA) on all mutants and native (WT) SH2D1A protein.

average SASA values. However, some mutants, such as C42Y, I84T, G93D, Y7C, R32T, Y54N, I80R, M1K, Y41N and R32S, had lower SASA values, indicating compaction of the protein due to gain of binding (Table S9).

The statistical significance of the mean values of each parameter obtained from the MDS of each of the mutant proteins and the native protein are shown in tables S5, S6, S7, S8 and S9.

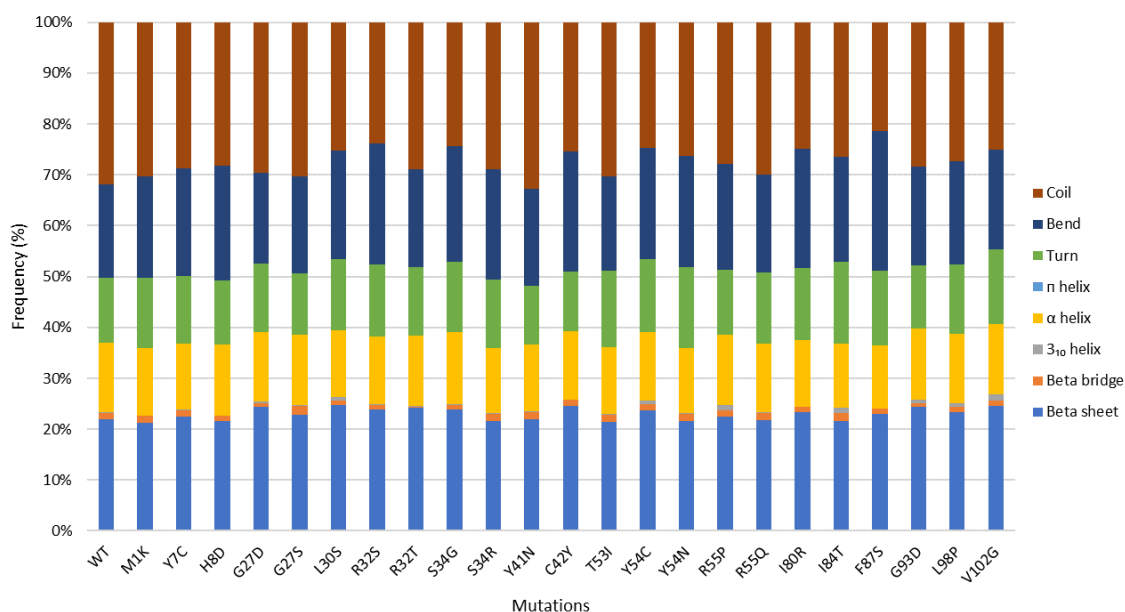


Figure 8: Graphical summary of the average frequency of protein secondary structures in the native protein and by mutant.

The type of secondary structures was evaluated through the simulation time (25ns) by the Define Secondary Structure of Proteins (DSSP) algorithm. When comparing the mutant proteins with the wild-type protein, we can see that in general there is a decrease in coil residues and an increase in residues in non-regular secondary structures (turns and bends). More than half of the mutant proteins also show an increase in secondary structures (helices and sheets) (Figure 8).

3.5. Discussion

This research primarily addressed mutations within the *SH2D1A* gene associated with XLP to assess their impact on the structural stability and protein functionality of SH2D1A mutants. In the general population, SNPs are the predominant form of DNA variation and play a critical role in unravelling complex human diseases characterised by genetic phenotypic variations. The primary goal of studying deleterious SNPs is to uncover their

significant impact on the function and structure of DNA. However, comprehensive assessment of the deleterious effects of all nsSNPs from the vast pool of available nsSNPs through experimental research is challenging and time-consuming [34,59–70]. Therefore, various bioinformatics tools and databases come into play to distinguish between harmful nsSNPs and benign nsSNPs from the vast collection of reported nsSNPs. Previous research has highlighted the critical role of various bioinformatics algorithms in identifying harmful nonsynonymous SNPs associated with several complex diseases [71,72]. These computational tools are indispensable resources in the field of protein stability analyses, offering precise and detailed insights into the consequences of genetic variations on the structural integrity and function of proteins [73].

Although mutations within the SH2D1A protein have been frequently documented in various physiological conditions, the exact mechanisms underlying the onset of various diseases remain obscure [14]. Therefore, an investigation of the effects of nsSNPs in the SH2D1A protein has the potential to shed light on the deleterious consequences. This endeavour involved deciphering the nature and structural location of the SNPs to understand their impact on protein functionality and stability to elucidate the severity of the associated disease [59,74]. Various computational tools were used to classify non-synonymous polymorphisms as either neutral or disease related. A substantial number of SH2D1A mutations have been documented, with almost half of these mutations falling into the missense/nonsense category.

A bioinformatics framework was used to prioritise nsSNPs within the SH2D1A gene. This prioritisation process included a structural analysis using PredictSNP1.0, I-Mutant 2 and MUpro, which identified 23 nsSNPs that were classified as "high-risk" (Tables S2 and S3). This harmfulness of the mutations can be understood quite generally due to the high degree of conservation of the amino acids involved, according to the ConSurf web server [38]. Put simply, a high level of conservation in a region indicates that the region performs a vital function and is essential to the structure and function of the protein [75,76]. This would be expected given that the 23 mutations affect the only functional domain of the protein (SH2 domain), which is crucial for the interprotein interactions of the SH2D1A protein required for the initiation of signalling cascades. Therefore, any mutation in this region could have a significantly detrimental effect by disrupting a critical property of the protein [77–81].

Mutations can also lead to changes in the physicochemical properties of a protein, affecting its stability [82]. These proteins properties include size, charge distribution, hydrophobic interactions and structural integrity [83]. Therefore, the evaluation of

nsSNPs based on the properties of the amino acids they affect is a crucial aspect of understanding protein stability, functionality and folding [84,85]. For all mutations in the SH2D1A protein, the prediction tool HOPE has found that they have a negative impact on the behaviour of the protein due to their altered physicochemical properties (Tables S4).

This collective method using different algorithms that evaluate multiple mutation-related traits, significantly improves the accuracy of predicting the likelihood that reported SNPs are associated with disease, surpassing the accuracy achieved by individual methods [64–68,86–90].

Biomolecules are not static and are in constant motion [91]. With MDS, we can observe how amino acid exchanges caused by nsSNPs affect the dynamic behaviour of proteins. This dynamic perspective helps us understand how genetic variations can lead to changes in the molecular flexibility, dynamics and adaptability of proteins, which are crucial factors in disease mechanisms [92–94]. Hence, MDS analyses were performed to confirm and understand these deleterious and instable effects that showed significant phenotypic effects of disease-associated nsSNPs on native protein conformation and their interaction dynamics.

In the M1K, Y7C, R32S, Y54N, I80R and I84T mutants (Figure S21, S22, S23, S24, S26, S27, S32, S39, S43 and S44), we observed a remarkable trend towards greater structural stability compared to the native SH2D1A protein (Figure S21 and S25), proven by the reduction of RMSD values [95]. This greater stability is supported too by the reduction in RMSF levels, indicating less flexibility of the protein [96]. This, in turn, could limit the protein's ability to make conformational changes and adapt to different functional states, as more stable conformations are generally naturally less flexible [97–99]. In addition, we observed a significant decrease in Rg and SASA values in the mutations, indicating greater structural compaction [100]. These results add to the idea that these mutations have a stabilising effect on the protein. The decrease in accessible surface area indicates a decrease in the ability to interact with other molecules [101], which could explain the inability of SH2D1A mutants to bind effectively to SLAMF receptor proteins [102]. The increase in structural stability can be explained by the increase in the number of intramolecular hydrogen bonds in the mutated proteins [103], observed in the respective results. These bonds play a crucial role in stabilising regular secondary structures, such as helices and beta-sheets, as well as irregular structures, such as those found in turns [104]. These observations, based on the values obtained in the molecular dynamics analyses, are consistent with the data from the DSSP. In the DSSP, we find a

decrease in residues in coil conformations, which normally contribute to a more open structure, while there is an increase in residues in regular and non-regular secondary structures [105]. This change in secondary structure generally contributes to a more compact protein structure [106].

The G27D, S34R and R55Q (Figure S21, S22, S23, S29, S43 and S44) mutants also show lower average RMSD and RMSF values than the wild-type protein, again supporting and indicating greater stability and less structural flexibility. In addition, the average number of hydrogen bonds is higher than that of the wild-type protein, supported by the change from coil residues to regular and non-regular secondary structures. However, the average SASA values are higher than the wild-type protein, indicating greater unzipping and a consequent increase in the contact surface with the solvent. These values seem to contradict the average Rg values obtained, where there is a decrease indicating greater compaction of the protein (except for R55Q where Rg increases). However, if the change that leads to increased surface exposure is the result of the collapse of a flexible region of the protein, this flexible region that collapses was previously shielding certain residues from the solvent, this could justify the decrease in Rg [107]. This idea can be further supported by the fact that there is a decrease in flexibility due to a decrease in the RMSF value in the G27D and S34R mutants, in contrast to the R55Q, where the reduction is not significant (Table S6).

The H8D and C42Y (Figure S21, S23, S28 and S37) mutants show greater stability than the wild-type protein (as indicated by the lower RMSD), but with greater flexibility (as indicated by the higher RMSF). In addition, the mutant proteins appear to be less compact (as indicated by the higher Rg), but with less solvent-accessible surface area (as indicated by the lower SASA, except for C42Y). The equal number of hydrogen bonds suggests that the mutations did not significantly disrupt the protein's hydrogen bond network. Finally, the increase in secondary structures may indicate that the mutation led to a more structured local conformation [108]. These results may seem contradictory at first glance, but they can be explained by considering the complexity of the protein structure. It is possible that the mutation introduced flexibility into certain regions of the protein, helping with global conformation by relieving certain conformational tensions caused, increasing stability. Similarly, the increase in Rg and decrease in SASA may be due to changes in the protein's conformation that expose less of the protein's surface to the solvent, but also result in a less compact and structure. This could be the case if the mutation leads to a conformation that is more extended overall (increase in Rg) but has internal regions that are less accessible to the solvent (decrease in SASA). It might

happen if the mutation leads to the formation of a new pocket or cavity in the protein or if it causes a restructuring of the protein domains in such a way that some areas of the surface become less accessible [107].

In the L30S, T53I, Y54C, F87S, L98P and V102G (Figure S22, S23, S24, S31, S38, S40, S45, S47 and S48) mutants, we observed a notable reduction in the structural stability of the mutant proteins, as indicated by the RMSD values. This decrease in stability is supported by the increase in RMSF values, suggesting an increase in structural flexibility. This higher flexibility can be explained by the increase in turns and bends in the protein structure. In addition, the Rg and SASA values are greater than those observed in the wild-type protein, indicating an expansion in the protein structure [109]. However, it is important to note the increase in the number of intramolecular hydrogen bonds in the mutants. This observation suggests that the mutations may have introduced more opportunities for hydrogen bond formation which reflect the shift from coil residues to secondary structures. Similarly to these mutants, the G27S, S34G and R55P mutants (Figure S21, S22, S23, S30, S34 and S42) also show higher average RMSD, RMSF, Rg and SASA values than the wild-type protein, but the average number of hydrogen bonds are equal to the native protein, (except the R55P mutant, which shows a decrease), also accompanied by a alteration of coil residues to secondary structures. The new apparent contradiction between the rise/maintenance of hydrogen bonds with the increase in RMSD can be explained by the possibility of the mutations introduce/change hydrogen bonds that do not contribute significantly to the protein's stability. Protein stability depends not only on the quantity of hydrogen bonds, but also on their nature. The exposure of hydrogen bonds to solvent plays a vital role on the quality of this bonds [107,110]. A hydrogen bond that is in an environment protected from water will be more stable than one that is exposed to solvent [111]. If many of these hydrogen bonds are more exposed to solvent (as indicated by the increase in SASA), they may not contribute significantly to the stability of the protein. Although secondary structures can contribute to the protein's stability, other changes caused by the mutation can upset this balance and lead to a less stable overall structure [112]. For example, if a mutation introduces an amino acid residue with a large side chain or a different electrical charge into a region of the protein that is normally hydrophobic and therefore generally has without charge, can disrupt hydrophobic interactions and lead to a less stable structure, even if more secondary structures are formed [107,113,114]. Other factors, such as hydrophobic interactions, disulfide bridges and ionic interactions [115–119], also play a crucial role in the stability of the protein and in the formation of the secondary structures, explaining

the increase in secondary structures in mutants without an increase in hydrogen bonds [105,120].

In the case of the R32T, Y41N and G93D mutants (Figure S22, S24, S33, S36 and S46), we observed a parametric pattern identical to the L30S, T53I, Y54C, F87S, L98P and V102G mutants, but with lower average of Rg and SASA values than the wild-type protein (except for the G93D mutant, where average SASA values increase). Despite the increase in hydrogen bonds and secondary structures (except for the Y41N mutant, where the number of secondary structures decrease), the changes in charge, the reduction in hydrophobicity and the decrease in residue size (resulting in a potential vacant space in the protein structure) in the G93D, Y41N and R32T mutants respectively, could have a more significant impact on structural stability [121–123]. This emphasizes the substantial influence of these factors over others in determining the overall stability of the protein.

In summary, unlike the M1K, Y7C, R32S, Y54N, I80R and I84T mutants, the analysis of the H8D, G27D, G27S, L30S, R32T, S34G, S34R, Y41N, C42Y, T53I, Y54C, R55P, R55Q, F87S, G93D, L98P and V102G mutations highlights the complexity of protein stability. Stability is not only determined by the number of secondary structures and hydrogen bonds, but also by quality, secondary interactions and solvent exposure. Understanding these factors is essential for assessing the impact of mutations on protein structure and function.

3.6. Conclusion and Future Perspectives

This study has thoroughly investigated the impact of mutations within the *SH2D1A* gene associated with X-linked lymphoproliferative syndrome (XLP) on the structural stability and functionality of SH2D1A mutants. Using advanced bioinformatics tools and molecular dynamics simulations (MDS), we have acquired valuable insights into the intricate relationship between genetic variations and protein behavior. The research underscores the pivotal role of computational algorithms in discerning detrimental non-synonymous SNPs (nsSNPs) and their effects on protein stability and function.

Our study has unveiled a considerable number of mutations, each manifesting unique patterns of conformational alterations within the SH2D1A mutants. These findings contribute significantly to our comprehension of how nsSNPs can disrupt protein function and potentially contribute to disease onset.

While computational tools are indispensable for predicting the effects of mutations, it is imperative to emphasize the need for experimental validation. Future investigations

should encompass *in vitro* and *in vivo* experiments to conclusively confirm the impact of these mutations on protein structure and function.

Following this validation process, we propose a complementary approach. By employing computational techniques, we can identify transient pockets within the protein structure using the trajectories obtained from MDS. The identification of pockets that remain consistent during the simulation and exhibit potential druggability can be leveraged for virtual screening. This screening aims to identify molecules (molecular chaperones) capable of restoring the mutant protein to its native conformation and thus reinstating its normal functions.

In this manner, our research provides a robust foundation for the development of targeted therapies for XLP and related diseases by unraveling the specific ways in which mutations affect protein structure and function. Furthermore, the insights gleaned from this study have the potential to revolutionize personalized medicine by tailoring medical interventions and treatment strategies to the specific genetic profiles of individual patients.

References

- [1] P. Suwinski, C.K. Ong, M.H.T. Ling, Y.M. Poh, A.M. Khan, H.S. Ong, Advancing personalized medicine through the application of whole exome sequencing and big data analytics, *Front Genet.* 10 (2019) 422886. <https://doi.org/10.3389/FGENE.2019.00049/BIBTEX>.
- [2] B. Rabbani, M. Tekin, N. Mahdieh, The promise of whole-exome sequencing in medical genetics, *Journal of Human Genetics* 2014 59:1. 59 (2013) 5–15. <https://doi.org/10.1038/jhg.2013.114>.
- [3] S.L. Nickerson, R. Marquis-Nicholson, K. Claxton, F. Ashton, I.U.S. Leong, D.O. Prosser, J.M. Love, A.M. George, G. Taylor, C. Wilson, R.J.M. Gardner, D.R. Love, SNP Analysis and Whole Exome Sequencing: Their Application in the Analysis of a Consanguineous Pedigree Segregating Ataxia, *Microarrays*. 4 (2015) 490. <https://doi.org/10.3390/MICROARRAYS4040490>.
- [4] R.J. Pengelly, J. Gibson, G. Andreoletti, A. Collins, C.J. Mattocks, S. Ennis, A SNP profiling panel for sample tracking in whole-exome sequencing studies, *Genome Med.* 5 (2013) 1–7. <https://doi.org/10.1186/GM492/FIGURES/3>.

- [5] J.T.L. Mah, E.S.H. Low, E. Lee, *In silico* SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery, *Drug Discov Today*. 16 (2011) 800–809. <https://doi.org/10.1016/J.DRUDIS.2011.07.005>.
- [6] C.M. Yates, M.J.E. Sternberg, The Effects of Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) on Protein–Protein Interactions, *J Mol Biol*. 425 (2013) 3949–3963. <https://doi.org/10.1016/J.JMB.2013.07.012>.
- [7] F. Robert, J. Pelletier, Exploring the Impact of Single-Nucleotide Polymorphisms on Translation, *Front Genet*. 9 (2018) 507. <https://doi.org/10.3389/FGENE.2018.00507>.
- [8] C.P.I. Schärfe, R. Tremmel, M. Schwab, O. Kohlbacher, D.S. Marks, Genetic variation in human drug-related genes, *Genome Med*. 9 (2017) 1–15. <https://doi.org/10.1186/S13073-017-0502-5/FIGURES/4>.
- [9] J. Wu, R. Jiang, Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases, *The Scientific World Journal*. 2013 (2013). <https://doi.org/10.1155/2013/675851>.
- [10] M. Zhang, C. Huang, Z. Wang, H. Lv, X. Li, *In silico* analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in the human GJA3 gene associated with congenital cataract, *BMC Mol Cell Biol*. 21 (2020) 1–13. <https://doi.org/10.1186/S12860-020-00252-7/FIGURES/6>.
- [11] J.D. Durrant, J.A. McCammon, Molecular dynamics simulations and drug discovery, *BMC Biol*. 9 (2011) 1–9. <https://doi.org/10.1186/1741-7007-9-71/FIGURES/4>.
- [12] M.S. Badar, S. Shamsi, J. Ahmed, Md.A. Alam, Molecular Dynamics Simulations: Concept, Methods, and Applications, (2022) 131–151. https://doi.org/10.1007/978-3-030-94651-7_7.
- [13] G. Glusman, P.W. Rose, A. Prlić, J. Dougherty, J.M. Duarte, A.S. Hoffman, G.J. Barton, E. Bendixen, T. Bergquist, C. Bock, E. Brunk, M. Buljan, S.K. Burley, B. Cai, H. Carter, J.J. Gao, A. Godzik, M. Heuer, M. Hicks, T. Hrabe, R. Karchin, J.K. Leman, L. Lane, D.L. Masica, S.D. Mooney, J. Moulton, G.S. Omenn, F. Pearl, V. Pejaver, S.M. Reynolds, A. Rokem, T. Schwede, S. Song, H. Tilgner, Y. Valasatava, Y. Zhang, E.W. Deutsch, Mapping genetic variations to three-dimensional protein structures to enhance variant interpretation: A proposed

- framework, *Genome Med.* 9 (2017) 1–10. <https://doi.org/10.1186/S13073-017-0509-Y/FIGURES/1>.
- [14] K. Zhang, E. Wakefield, R. Marsh, Lymphoproliferative Disease, X-Linked, GeneReviews®. (2016) 1–7. <https://www.ncbi.nlm.nih.gov/books/NBK1406/> (accessed April 7, 2023).
- [15] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, *Nucleic Acids Res.* 28 (2000) 235–242. <https://doi.org/10.1093/NAR/28.1.235>.
- [16] N. Panchal, C. Booth, J.L. Cannons, P.L. Schwartzberg, X-linked lymphoproliferative disease type 1: A clinical and molecular perspective, *Front Immunol.* 9 (2018). <https://doi.org/10.3389/FIMMU.2018.00666>.
- [17] M. Morra, D. Howie, M.S. Grande, J. Sayos, N. Wang, C. Wu, P. Engel, C. Terhorst, X-Linked Lymphoproliferative Disease: A Progressive Immunodeficiency, <https://doi.org/10.1146/Annurev.Immunol.19.1.657>. 19 (2003) 657–682. <https://doi.org/10.1146/ANNUREV.IMMUNOL.19.1.657>.
- [18] E.W. Sayers, E.E. Bolton, J.R. Brister, K. Canese, J. Chan, D.C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B.W. Trawick, K.D. Pruitt, S.T. Sherry, Database resources of the national center for biotechnology information, *Nucleic Acids Res.* 50 (2022) D20–D26. <https://doi.org/10.1093/NAR/GKAB1112>.
- [19] P.D. Stenson, M. Mort, E. V. Ball, K. Shaw, A.D. Phillips, D.N. Cooper, The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine, *Hum Genet.* 133 (2014) 1–9. <https://doi.org/10.1007/S00439-013-1358-4>.
- [20] S.T. Sherry, M.H. Ward, M. Kholodov, J. Baker, L. Phan, E.M. Smigielski, K. Sirotkin, dbSNP: the NCBI database of genetic variation, *Nucleic Acids Res.* 29 (2001) 308–311. <https://doi.org/10.1093/NAR/29.1.308>.
- [21] A. Bateman, M.J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-AJee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L.G. Castro, L. Figueira, P. Garmiri, G. Georgiou, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi,

- A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimò, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M.C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Junco, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nospikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A.L. Veuthey, C.H. Wu, C.N. Arighi, L. Arminski, C. Chen, Y. Chen, J.S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D.A. Natale, K. Ross, C.R. Vinayaka, Q. Wang, Y. Wang, L.S. Yeh, J. Zhang, UniProt: the universal protein knowledgebase, *Nucleic Acids Res.* 45 (2017) D158–D169. <https://doi.org/10.1093/NAR/GKW1099>.
- [22] J. Bendl, J. Stourac, O. Salanda, A. Pavelka, E.D. Wieben, J. Zendulka, J. Brezovsky, J. Damborsky, PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations, *PLoS Comput Biol.* 10 (2014). <https://doi.org/10.1371/JOURNAL.PCBI.1003440>.
- [23] S. Alam, M. Sayem, M. Bello, S.I. Mou, N.A. Faruqi, F. Hossain, Md.K. Hasan, Computational analysis uncovers the deleterious SNPs along with the mutational spectrum of p53 gene and its differential expression pattern in pan-cancer, *Bulletin of the National Research Centre* 2022 46:1. 46 (2022) 1–17. <https://doi.org/10.1186/S42269-022-00859-0>.
- [24] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, S.R. Sunyaev, A method and server for predicting damaging missense mutations, *Nat Methods.* 7 (2010) 248–249. <https://doi.org/10.1038/NMETH0410-248>.
- [25] V. Ramensky, P. Bork, S. Sunyaev, Human non-synonymous SNPs: server and survey, *Nucleic Acids Res.* 30 (2002) 3894–3900. <https://doi.org/10.1093/NAR/GKF493>.
- [26] P.C. Ng, S. Henikoff, Predicting deleterious amino acid substitutions, *Genome Res.* 11 (2001) 863–874. <https://doi.org/10.1101/GR.176601>.

- [27] E.A. Stone, A. Sidow, Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity, Krishnan and Westhead. (2001). <https://doi.org/10.1101/gr.3804205>.
- [28] A.D. Johnson, R.E. Handsaker, S.L. Pulit, M.M. Nizzari, C.J. O'Donnell, P.I.W. De Bakker, SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap, *Bioinformatics*. 24 (2008) 2938–2939. <https://doi.org/10.1093/BIOINFORMATICS/BTN564>.
- [29] H. Mi, A. Muruganujan, P.D. Thomas, PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees, *Nucleic Acids Res.* 41 (2013). <https://doi.org/10.1093/NAR/GKS1118>.
- [30] P.D. Thomas, D. Ebert, A. Muruganujan, T. Mushayahama, L.P. Albou, H. Mi, PANTHER: Making genome-scale phylogenetics accessible to all, *Protein Science*. 31 (2022) 8–22. <https://doi.org/10.1002/PRO.4218>.
- [31] E. Capriotti, R. Calabrese, R. Casadio, Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information, *Bioinformatics*. 22 (2006) 2729–2734. <https://doi.org/10.1093/BIOINFORMATICS/BTL423>.
- [32] L. Bao, M. Zhou, Y. Cui, nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms, *Nucleic Acids Res.* 33 (2005). <https://doi.org/10.1093/NAR/GKI372>.
- [33] L. Elkhatabi, I. Morjane, H. Charoute, S. Amghar, H. Bouafi, Z. Elkarhat, R. Saile, H. Rouba, A. Barakat, In silico analysis of coding/noncoding SNPs of human RETN gene and characterization of their impact on resistin stability and structure, *J Diabetes Res.* 2019 (2019). <https://doi.org/10.1155/2019/4951627>.
- [34] D. Thirumal Kumar, E. Judith, J. Priyadharshini Christy, R. Siva, I.A. Tayubi, C. Chakraborty, C. George Priya Doss, H. Zayed, Computational and modeling approaches to understand the impact of the Fabry's disease causing mutation (D92Y) on the interaction with pharmacological chaperone 1-deoxygalactonojirimycin (DGJ), *Adv Protein Chem Struct Biol.* 114 (2019) 341–407. <https://doi.org/10.1016/BS.APCSB.2018.10.009>.
- [35] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res.* 33 (2005) W306. <https://doi.org/10.1093/NAR/GKI375>.

- [36] J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, *Proteins*. 62 (2006) 1125–1132. <https://doi.org/10.1002/PROT.20810>.
- [37] S.M. Khan, A.R.M. Faisal, T.A. Nila, N.N. Binti, M.I. Hosen, H.U. Shekhar, A computational in silico approach to predict high-risk coding and non-coding SNPs of human PLCG1 gene, *PLoS One*. 16 (2021) e0260054. <https://doi.org/10.1371/JOURNAL.PONE.0260054>.
- [38] F. Glaser, T. Pupko, I. Paz, R.E. Bell, D. Bechor-Shental, E. Martz, N. Ben-Tal, ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information, *Bioinformatics*. 19 (2003) 163–164. <https://doi.org/10.1093/BIOINFORMATICS/19.1.163>.
- [39] H. Venselaar, T.A.H. te Beek, R.K.P. Kuipers, M.L. Hekkelman, G. Vriend, Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, *BMC Bioinformatics*. 11 (2010) 1–10. <https://doi.org/10.1186/1471-2105-11-548/FIGURES/5>.
- [40] V. Pejaver, J. Urresti, J. Lugo-Martinez, K.A. Pagel, G.N. Lin, H.J. Nam, M. Mort, D.N. Cooper, J. Sebat, L.M. Iakoucheva, S.D. Mooney, P. Radivojac, Inferring the molecular and phenotypic impact of amino acid variants with MutPred2, *Nature Communications* 2020 11:1. 11 (2020) 1–13. <https://doi.org/10.1038/s41467-020-19669-x>.
- [41] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, *Nucleic Acids Res*. 44 (2016) W344. <https://doi.org/10.1093/NAR/GKW408>.
- [42] M.L. Hekkelman, T.A.H. te Beek, S.R. Pettifer, D. Thorne, T.K. Attwood, G. Vriend, WIWS: a protein structure bioinformatics Web service collection, *Nucleic Acids Res*. 38 (2010) W719–W723. <https://doi.org/10.1093/NAR/GKQ453>.
- [43] A. Prlić, T.A. Down, E. Kulesha, R.D. Finn, A. Kähäri, T.J.P. Hubbard, Integrating sequence and structural biology with DAS, *BMC Bioinformatics*. 8 (2007) 1–9. <https://doi.org/10.1186/1471-2105-8-333/FIGURES/4>.
- [44] N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, *J Mol Biol*. 294 (1999) 1351–1362. <https://doi.org/10.1006/JMBI.1999.3310>.

- [45] M. Billeter, Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals, *Q Rev Biophys.* 25 (1992) 325–377. <https://doi.org/10.1017/S0033583500004261>.
- [46] A.W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A.W.R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D.T. Jones, D. Silver, K. Kavukcuoglu, D. Hassabis, Improved protein structure prediction using potentials from deep learning, *Nature* 2020 577:7792. 577 (2020) 706–710. <https://doi.org/10.1038/s41586-019-1923-7>.
- [47] H. Mouhib, B. Stringer, H. van Ingen, J. Gavaldá-García, K. Waury, S. Abeln, K.A. Feenstra, Structure determination, *An Introduction to High-Pressure Science and Technology.* (2021) 189–218. <https://doi.org/10.1017/cbo9780511565113.005>.
- [48] Schrödinger LLC, The PyMOL Molecular Graphics System, Version~1.8, 2015.
- [49] A. Fiser, A. Sali, ModLoop: automated modeling of loops in protein structures, *Bioinformatics.* 19 (2003) 2500–2501. <https://doi.org/10.1093/BIOINFORMATICS/BTG362>.
- [50] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis.* 18 (1997) 2714–2723. <https://doi.org/10.1002/ELPS.1150181505>.
- [51] S. Páll, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, B. Hess, E. Lindahl, Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS, *Journal of Chemical Physics.* 153 (2020) 134110. <https://doi.org/10.1063/5.0018516/199476>.
- [52] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins.* 78 (2010) 1950. <https://doi.org/10.1002/PROT.22711>.
- [53] S.S. Petrova, A.D. Solov'Ev, The Origin of the Method of Steepest Descent, *Historia Mathematica.* 24 (1997) 361–375. <https://doi.org/10.1006/HMAT.1996.2146>.
- [54] A. Kumar, V. Rajendran, R. Sethumadhavan, R. Purohit, In silico prediction of a disease-associated STIL mutant and its affect on the recruitment of centromere protein J (CENPJ), *FEBS Open Bio.* 2 (2012) 285–293. <https://doi.org/10.1016/J.FOB.2012.09.003>.

- [55] C.G. Priya Doss, C. Chakraborty, L. Chen, H. Zhu, Integrating *in silico* prediction methods, molecular docking, and molecular dynamics simulation to predict the impact of ALK missense mutations in structural perspective, *Biomed Res Int.* 2014 (2014). <https://doi.org/10.1155/2014/895831>.
- [56] S.S. Sohrab, M.A. Kamal, Screening, Docking, and Molecular Dynamics Study of Natural Compounds as an Anti-HER2 for the Management of Breast Cancer, *Life.* 12 (2022). <https://doi.org/10.3390/LIFE12111729>.
- [57] K. Sekihara, Y. Kawabata, S. Ushio, S. Sumiya, S. Kawabata, Y. Adachi, S.S. Nagarajan, Dual signal subspace projection (DSSP): a novel algorithm for removing large interference in biomagnetic measurements, *J Neural Eng.* 13 (2016). <https://doi.org/10.1088/1741-2560/13/3/036007>.
- [58] P.J. Turner, XMGRACE, Version 5.1. 19, Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR. 2 (2005).
- [59] S. Pandey, N. Maurya, H. Awasthi, P. Katara, S. Singh, B. Gautam, D.B. Singh, Comprehensive analysis of non-synonymous SNPs related to Parkinson's disease and molecular dynamics simulation of PRKN mutants, *Results Chem.* 5 (2023) 100817. <https://doi.org/10.1016/J.RECHEM.2023.100817>.
- [60] D. Thirumal Kumar, C. George Priya Doss, P. Sneha, I.A. Tayubi, R. Siva, C. Chakraborty, R. Magesh, Influence of V54M mutation in giant muscle protein titin: a computational screening and molecular dynamics approach, *J Biomol Struct Dyn.* 35 (2017) 917–928. <https://doi.org/10.1080/07391102.2016.1166456>.
- [61] D. Thirumal Kumar, C. George Priya Doss, P. Sneha, I.A. Tayubi, R. Siva, C. Chakraborty, R. Magesh, Influence of V54M mutation in giant muscle protein titin: a computational screening and molecular dynamics approach, *J Biomol Struct Dyn.* 35 (2017) 917–928. <https://doi.org/10.1080/07391102.2016.1166456>.
- [62] D. Thirumal Kumar, L. Jerushah Emerald, C. George Priya Doss, P. Sneha, R. Siva, W. Charles Emmanuel Jebaraj, H. Zayed, Computational approach to unravel the impact of missense mutations of proteins (D2HGDH and IDH2) causing D-2-hydroxyglutaric aciduria 2, *Metab Brain Dis.* 33 (2018) 1699–1710. <https://doi.org/10.1007/S11011-018-0278-3/FIGURES/4>.
- [63] D. Thirumal Kumar, H.G. Eldous, Z.A. Mahgoub, C. George Priya Doss, H. Zayed, Computational modelling approaches as a potential platform to understand the

- molecular genetics association between Parkinson's and Gaucher diseases, *Metab Brain Dis.* 33 (2018) 1835–1847. <https://doi.org/10.1007/S11011-018-0286-3/FIGURES/1>.
- [64] P. Sneha, D. Kumar Thirumal, H. Tanwar, R. Siva, C. George Priya Doss, H. Zayed, Structural Analysis of G1691S Variant in the Human Filamin B Gene Responsible for Larsen Syndrome: A Comparative Computational Approach, *J Cell Biochem.* 118 (2017) 1900–1910. <https://doi.org/10.1002/JCB.25920>.
- [65] P. Sneha, K.D. Thirumal, P.D.C. George, R. Siva, H. Zayed, Determining the role of missense mutations in the POU domain of HNF1A that reduce the DNA-binding affinity: A computational approach, *PLoS One.* 12 (2017). <https://doi.org/10.1371/JOURNAL.PONE.0174953>.
- [66] P. Sneha, E.A. Ebrahimi, S.A. Ghazala, D. Thirumal Kumar, R. Siva, G. Priya Doss C, H. Zayed, Structural analysis of missense mutations in galactokinase 1 (GALK1) leading to galactosemia type-2, *J Cell Biochem.* 119 (2018) 7585–7598. <https://doi.org/10.1002/JCB.27097>.
- [67] P. Sneha, T.U. Zenith, U.S. Abu Habib, J. Evangeline, D. Thirumal Kumar, C. George Priya Doss, R. Siva, H. Zayed, Impact of missense mutations in survival motor neuron protein (SMN1) leading to Spinal Muscular Atrophy (SMA): A computational approach, *Metab Brain Dis.* 33 (2018) 1823–1834. <https://doi.org/10.1007/S11011-018-0285-4/TABLES/5>.
- [68] C. George Priya Doss, C. Chakraborty, V. Narayan, D.T. Kumar, Computational Approaches and Resources in Single Amino Acid Substitutions Analysis Toward Clinical Research, *Adv Protein Chem Struct Biol.* 94 (2014) 365–423. <https://doi.org/10.1016/B978-0-12-800168-4.00010-X>.
- [69] A.K. Agrahari, M. Muskan, C. George Priya Doss, R. Siva, H. Zayed, Computational insights of K1444N substitution in GAP-related domain of NF1 gene associated with neurofibromatosis type 1 disease: a molecular modeling and dynamics approach, *Metabolic Brain Disease* 2018 33:5. 33 (2018) 1443–1457. <https://doi.org/10.1007/S11011-018-0251-1>.
- [70] A.K. Agrahari, P. Sneha, C. George Priya Doss, R. Siva, H. Zayed, A profound computational study to prioritize the disease-causing mutations in PRPS1 gene, *Metab Brain Dis.* 33 (2018) 589–600. <https://doi.org/10.1007/S11011-017-0121-2/FIGURES/7>.

- [71] K. Shinwari, Y. Wu, H.M. Rehman, N. Xiao, M. Bolkov, I. Tuzankina, V. Cheresnev, In-silico assessment of high-risk non-synonymous SNPs in ADAMTS3 gene associated with Hennekam syndrome and their impact on protein stability and function, BMC Bioinformatics. 24 (2023) 1–33. <https://doi.org/10.1186/S12859-023-05361-6/FIGURES/14>.
- [72] D.F. Burke, C.L. Worth, E.-M. Priego, T. Cheng, L.J. Smink, J.A. Todd, T.L. Blundell, Genome bioinformatic analysis of nonsynonymous SNPs, BMC Bioinformatics 2007 8:1. 8 (2007) 1–15. <https://doi.org/10.1186/1471-2105-8-301>.
- [73] K.M. Pawłowski, J. Mucha, K. Majchrzak, T. Motyl, M. Król, Expression and role of PGP, BCRP, MRP1 and MRP3 in multidrug resistance of canine mammary cancer cells, BMC Vet Res. 9 (2013). <https://doi.org/10.1186/1746-6148-9-119>.
- [74] H.C. Lu, J. Herrera Braga, F. Fraternali, PinSnps: structural and functional analysis of SNPs in the context of protein interaction networks, Bioinformatics. 32 (2016) 2534–2536. <https://doi.org/10.1093/BIOINFORMATICS/BTW153>.
- [75] V.M. Prabantu, N. Naveenkumar, N. Srinivasan, Influence of Disease-Causing Mutations on Protein Structural Networks, Front Mol Biosci. 7 (2021) 620554. <https://doi.org/10.3389/FMOLB.2020.620554/BIBTEX>.
- [76] B.J. Livesey, J.A. Marsh, The properties of human disease mutations at protein interfaces, PLoS Comput Biol. 18 (2022) e1009858. <https://doi.org/10.1371/JOURNAL.PCBI.1009858>.
- [77] R. Planas, M. Felber, S. Vavassori, J. Pachlopnik Schmid, The hyperinflammatory spectrum: from defects in cytotoxicity to cytokine control, Front Immunol. 14 (2023) 1163316. <https://doi.org/10.3389/FIMMU.2023.1163316/BIBTEX>.
- [78] M. Morra, M. Simarro-Grande, M. Martin, A.S.I. Chen, A. Lanyi, O. Silander, S. Calpe, J. Davis, T. Pawson, M.J. Eck, J. Sumegi, P. Engel, S.C. Li, C. Terhorst, Characterization of SH2D1A missense mutations identified in X-linked lymphoproliferative disease patients, J Biol Chem. 276 (2001) 36809–36816. <https://doi.org/10.1074/JBC.M101305200>.
- [79] L. Torralba-Raga, B. Tesi, S.C.C. Chiang, H. Schlums, M. Nordenskjöld, A.C. Horne, J.I. Henter, M. Meeths, M. Abdelhaleem, S. Weitzman, Y. Bryceson, Diagnostic challenges for a novel SH2D1A mutation associated with X-linked lymphoproliferative disease, Pediatr Blood Cancer. 67 (2020). <https://doi.org/10.1002/PBC.28184>.

- [80] N. Sen, I. Anishchenko, N. Bordin, I. Sillitoe, S. Velankar, D. Baker, C. Orengo, Characterizing and explaining the impact of disease-associated mutations in proteins without known structures or structural homologs, *Brief Bioinform.* 23 (2022). <https://doi.org/10.1093/BIB/BBAC187>.
- [81] A. Diop, D. Santorelli, F. Malagrino, C. Nardella, V. Pennacchietti, L. Pagano, L. Marcocci, P. Pietrangeli, S. Gianni, A. Toto, SH2 Domains: Folding, Binding and Therapeutical Approaches, *International Journal of Molecular Sciences* 2022, Vol. 23, Page 15944. 23 (2022) 15944. <https://doi.org/10.3390/IJMS232415944>.
- [82] I.M. Hernández, Y. Dehouck, U. Bastolla, J.R. López-Blanco, P. Chacón, Predicting protein stability changes upon mutation using a simple orientational potential, *Bioinformatics.* 39 (2023). <https://doi.org/10.1093/BIOINFORMATICS/BTAD011>.
- [83] T.G. Kucukkal, M. Petukh, L. Li, E. Alexov, Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins, *Curr Opin Struct Biol.* 32 (2015) 18–24. <https://doi.org/10.1016/J.SBI.2015.01.003>.
- [84] S.C. Das, M.A. Rahman, S. Das Gupta, In-silico analysis unravels the structural and functional consequences of non-synonymous SNPs in the human IL-10 gene, *Egyptian Journal of Medical Human Genetics.* 23 (2022) 1–14. <https://doi.org/10.1186/S43042-022-00223-X/FIGURES/6>.
- [85] H. Venkata Subbiah, P. Ramesh Babu, U. Subbiah, In silico analysis of non-synonymous single nucleotide polymorphisms of human DEFB1 gene, *Egyptian Journal of Medical Human Genetics.* 21 (2020) 1–9. <https://doi.org/10.1186/S43042-020-00110-3/TABLES/5>.
- [86] O.K. Zaki, C. George Priya Doss, S.A. Ali, G.G. Murad, S.A. Elashi, M.S.A. Ebnou, D. Thirumal Kumar, O. Khalifa, R. Gamal, H.S.A. El Abd, B.N. Nasr, H. Zayed, Genotype–phenotype correlation in patients with isovaleric acidemia: comparative structural modelling and computational analysis of novel variants, *Hum Mol Genet.* 26 (2017) 3105–3115. <https://doi.org/10.1093/HMG/DDX195>.
- [87] R. Rajasekaran, C. Sudandiradoss, C.G.P. Doss, R. Sethumadhavan, Identification and in silico analysis of functional SNPs of the BRCA1 gene, *Genomics.* 90 (2007) 447–452. <https://doi.org/10.1016/J.YGENO.2007.07.004>.

- [88] C.G.P. Doss, R. Sethumadhavan, Investigation on the role of nsSNPs in HNPCC genes - A bioinformatics approach, *J Biomed Sci.* 16 (2009) 1–14. <https://doi.org/10.1186/1423-0127-16-42/FIGURES/3>.
- [89] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2, *Curr Protoc Hum Genet.* 76 (2013) 7.20.1-7.20.41. <https://doi.org/10.1002/0471142905.HG0720S76>.
- [90] C. George Priya Doss, H. Zayed, Comparative computational assessment of the pathogenicity of mutations in the Aspartoacylase enzyme, *Metab Brain Dis.* 32 (2017) 2105–2118. <https://doi.org/10.1007/S11011-017-0090-5/TABLES/4>.
- [91] Y. Gu, The Dynamic Nature of Biomolecules: Structure, Function, and Interactions, *Current Synthetic and Systems Biology.* 11 (2023) 1–2. <https://doi.org/10.35248/2332-0737.23.11.029>.
- [92] A. Kumar, R. Purohit, Use of Long Term Molecular Dynamics Simulation in Predicting Cancer Associated SNPs, *PLoS Comput Biol.* 10 (2014) e1003318. <https://doi.org/10.1371/JOURNAL.PCBI.1003318>.
- [93] M.J. Ul Hasnain, M. Shoaib, S. Qadri, B. Afzal, T. Anwar, S.H. Abbas, A. Sarwar, H.M.T. Malik, M.T. Pervez, Computational analysis of functional single nucleotide polymorphisms associated with SLC26A4 gene, *PLoS One.* 15 (2020) e0225368. <https://doi.org/10.1371/JOURNAL.PONE.0225368>.
- [94] M. Jia, B. Yang, Z. Li, H. Shen, X. Song, W. Gu, Computational Analysis of Functional Single Nucleotide Polymorphisms Associated with the CYP11B2 Gene, *PLoS One.* 9 (2014) e104311. <https://doi.org/10.1371/JOURNAL.PONE.0104311>.
- [95] I. Aier, P.K. Varadwaj, U. Raj, Structural insights into conformational stability of both wild-type and mutant EZH2 receptor, *Scientific Reports* 2016 6:1. 6 (2016) 1–10. <https://doi.org/10.1038/srep34984>.
- [96] A. Bornot, C. Etchebest, A.G. De Brevern, Predicting protein flexibility through the prediction of local structures, *Proteins.* 79 (2011) 839. <https://doi.org/10.1002/PROT.22922>.
- [97] R. Zrieq, I. Ahmad, M. Snoussi, E. Noumi, M. Iriti, F.D. Algahtani, H. Patel, M. Saeed, M. Tasleem, S. Sulaiman, K. Aouadi, A. Kadri, Tomatidine and Patchouli Alcohol as Inhibitors of SARS-CoV-2 Enzymes (3CLpro, PLpro and NSP15) by Molecular Docking and Molecular Dynamics Simulations, *International Journal of*

- Molecular Sciences 2021, Vol. 22, Page 10693. 22 (2021) 10693. <https://doi.org/10.3390/IJMS221910693>.
- [98] F.O. Balogun, K. Singh, A. Rampadarath, A. Akoonjee, K. Naidoo, S. Sabiu, Cheminformatics identification of modulators of key carbohydrate-metabolizing enzymes from *C. cujete* for type-2 diabetes mellitus intervention, *J Diabetes Metab Disord.* (2023) 1–19. <https://doi.org/10.1007/S40200-023-01249-7/TABLES/5>.
- [99] S. Ghahremanian, M.M. Rashidi, K. Raeisi, D. Toghraie, Molecular dynamics simulation approach for discovering potential inhibitors against SARS-CoV-2: A structural review, *J Mol Liq.* 354 (2022) 118901. <https://doi.org/10.1016/J.MOLLIQ.2022.118901>.
- [100] K. Haji-Allahverdipoor, M.J. Javaran, S.R. Monfared, M.B. Khadem-Erfan, B. Nikkhoo, Z.B. Rad, H. Eslami, S. Nasser, Insights Into The Effects of Amino Acid Substitutions on The Stability of Retheplase Structure: A Molecular Dynamics Simulation Study, *Iran J Biotechnol.* 21 (2023) e3175. <https://doi.org/10.30498/IJB.2022.308798.3175>.
- [101] K. Konstantinidis, I. Karakasiliotis, K. Anagnostopoulos, G.C. Boulougouris, On the estimation of the molecular inaccessible volume and the molecular accessible surface of a ligand in protein–ligand systems, *Mol Syst Des Eng.* 6 (2021) 946–963. <https://doi.org/10.1039/D1ME00053E>.
- [102] W. Li, M.H. Sofi, S. Rietdijk, N. Wang, C. Terhorst, C.H. Chang, The SLAM-Associated Protein (SAP)/Fyn/PKC θ Pathway is Required for Thymocyte-mediated CD4 T Cell Development, *Immunity.* 27 (2007) 763. <https://doi.org/10.1016/J.IMMUNI.2007.10.008>.
- [103] J.T.S. Coimbra, R. Feghali, R.P. Ribeiro, M.J. Ramos, P.A. Fernandes, The importance of intramolecular hydrogen bonds on the translocation of the small drug piracetam through a lipid bilayer, *RSC Adv.* 11 (2020) 899. <https://doi.org/10.1039/D0RA09995C>.
- [104] I. Rehman, M. Farooq, S. Botelho, Biochemistry, Secondary Protein Structure, StatPearls. (2022). <https://www.ncbi.nlm.nih.gov/books/NBK470235/> (accessed November 7, 2023).

- [105] J. Reeb, B. Rost, Secondary Structure Prediction, Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics. 1–3 (2019) 488–496. <https://doi.org/10.1016/B978-0-12-809633-8.20267-7>.
- [106] Q. Sun, X. He, Y. Fu, The “Beacon” Structural Model of Protein Folding: Application for Trp-Cage in Water, Molecules. 28 (2023) 5164. <https://doi.org/10.3390/MOLECULES28135164/S1>.
- [107] F.C.L. Almeida, K. Sanches, R. Pinheiro-Aguiar, V.S. Almeida, I.P. Caruso, Protein Surface Interactions—Theoretical and Experimental Studies, Front Mol Biosci. 8 (2021) 706002. <https://doi.org/10.3389/FMOLB.2021.706002/BIBTEX>.
- [108] J. Baussand, A.C. Camproux, Deciphering the shape and deformation of secondary structures through local conformation analysis, BMC Struct Biol. 11 (2011) 1–17. <https://doi.org/10.1186/1472-6807-11-9/FIGURES/7>.
- [109] C. Savojardo, M. Manfredi, P.L. Martelli, R. Casadio, Solvent Accessibility of Residues Undergoing Pathogenic Variations in Humans: From Protein Structures to Protein Sequences, Front Mol Biosci. 7 (2021) 626363. <https://doi.org/10.3389/FMOLB.2020.626363/BIBTEX>.
- [110] M. Guncheva, Role of ionic liquids on stabilization of therapeutic proteins and model proteins, Protein Journal. 41 (2022) 369–380. <https://doi.org/10.1007/S10930-022-10058-5/SCHEMES/1>.
- [111] V. Bianco, S. Iskov, G. Franzese, Effect of hydrogen bonds on protein stability, (2010). <https://arxiv.org/abs/1011.3785v1> (accessed November 7, 2023).
- [112] K.A. Dill, Dominant Forces in Protein Folding, Biochemistry. 29 (1990) 7133–7155. https://doi.org/10.1021/BI00483A001/ASSET/BI00483A001.FP.PNG_V03.
- [113] S. Khan, M. Vihinen, Spectrum of disease-causing mutations in protein secondary structures, BMC Struct Biol. 7 (2007) 1–18. <https://doi.org/10.1186/1472-6807-7-56/FIGURES/6>.
- [114] V.M. Prabantu, N. Naveenkumar, N. Srinivasan, Influence of Disease-Causing Mutations on Protein Structural Networks, Front Mol Biosci. 7 (2021) 620554. <https://doi.org/10.3389/FMOLB.2020.620554/BIBTEX>.
- [115] S. Khan, M. Vihinen, Spectrum of disease-causing mutations in protein secondary structures, BMC Struct Biol. 7 (2007) 1–18. <https://doi.org/10.1186/1472-6807-7-56/FIGURES/6>.

- [116] J. Cramer, X. Jiang, W. Schönemann, M. Silbermann, P. Zihlmann, S. Siegrist, B. Fiege, R.P. Jakob, S. Rabbani, T. Maier, B. Ernst, Enhancing the enthalpic contribution of hydrogen bonds by solvent shielding, *RSC Chem Biol.* 1 (2020) 281–287. <https://doi.org/10.1039/D0CB00108B>.
- [117] M.J. Feige, I. Braakman, L.M. Hendershot, Disulfide Bonds in Protein Folding and Stability, (2018) 1–33. <https://doi.org/10.1039/9781788013253-00001>.
- [118] Q.W. Ding, M. Lin, Protein Disulfide Bonds Detected by Tagging with High Molecular Weight Maleimide Derivative, *Mol Biol.* 55 (2021) 449–457. <https://doi.org/10.1134/S0026893321020187/FIGURES/5>.
- [119] D. Andreu, F. Albericio, N.A. Solé, M.C. Munson, M. Ferrer, G. Barany, Formation of disulfide bonds in synthetic peptides and proteins., *Methods Mol Biol.* 35 (1994) 91–169. <https://doi.org/10.1385/0-89603-273-6:91/TABLES/1>.
- [120] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers.* 22 (1983) 2577–2637. <https://doi.org/10.1002/BIP.360221211>.
- [121] A. Misiura, C. Dutta, W. Leung, J. Zepeda O, T. Terlier, C.F. Landes, The competing influence of surface roughness, hydrophobicity, and electrostatics on protein dynamics on a self-assembled monolayer, *Journal of Chemical Physics.* 156 (2022) 94707. https://doi.org/10.1063/5.0078797/16539158/094707_1_ACCEPTED_MANUSCRIPT.PDF.
- [122] H. Lan, H. Liu, Y. Ye, Z. Yin, The Role of Surface Properties on Protein Aggregation Behavior in Aqueous Solution of Different pH Values, *AAPS PharmSciTech.* 21 (2020) 1–13. <https://doi.org/10.1208/S12249-020-01663-7/FIGURES/10>.
- [123] Q. Wu, C. Cao, S. Wei, H. He, K. Chen, L. Su, Q. Liu, S. Li, Y. Lai, J. Li, Decreasing hydrophobicity or shielding hydrophobic areas of CH₂ attenuates low pH-induced IgG4 aggregation, *Front Bioeng Biotechnol.* 11 (2023) 1257665. <https://doi.org/10.3389/FBIOE.2023.1257665/BIBTEX>.

Appendix

Table S1: Missense SNPs of the *SH2D1A* gene. Extracted from the ClinVar [18] and HGMD [19] databases.

SNP ID	ClinVar ID	HGMD ID	Position	Residue Change	Impacted codons
rs111033629	VCV000010908	CM000425	1	M [Met] → I [Ile]	atG → atT
---	---	CM1413494	1	M [Met] → K [Lys]	aTg → aAt
rs2147519353	VCV001455661	CM056698	1	M [Met] → T [Thr]	aTg → aCg
---	VCV002442798	---	1	M [Met] → V [Val]	Atg → Gtg
rs1556619319	VCV000547771	---	2	D [Asp] → G [Gly]	gAc → gGc
rs148554414	VCV001169457	CM112387	3	A [Ala] → S [Ser]	Gca → Tca
rs1569527111	VCV000633415	CM002851	7	Y [Tyr] → C [Cys]	tAt → tGt
---	---	CM012656	8	H [His] → D [Asp]	Cat → Gat
---	VCV002501803	CM064262	8	H [His] → P [Pro]	cAt → cCt
---	---	CM195261	12	S [Ser] → I [Ile]	aGc → aTc
---	---	CM051984	16	G [Gly] → D [Asp]	gGc → gAc
---	---	CM056699	22	A [Ala] → P [Pro]	Gcc → Ccc
---	---	CM1111179	26	D [Asp] → V [Val]	gAt → gTt
---	---	CM171571	27	E [Glu] → D [Asp]	Ggc → Agc
---	---	CM012657	27	G [Gly] → S [Ser]	gGc → gAc
---	---	CM002852	28	S [Ser] → R [Arg]	agC → agG
---	---	CM142694	29	Y [Tyr] → C [Cys]	tAt → tGt
---	---	CM1010329	29	Y [Tyr] → H [His]	Tat → Cat
---	---	CM1414324	30	L [Leu] → S [Ser]	tTg → tCg
---	---	CM002853	31	L [Leu] → P [Pro]	cTg → cCg
---	---	CM1510942	32	R [Arg] → S [Ser]	aGg → aCg
rs111033624	VCV000010901	CM981808	32	R [Arg] → T [Thr]	agG → agC
---	---	CM012658	33	D [Asp] → Y [Tyr]	Gac → Tac
---	---	CM023984	34	S [Ser] → G [Gly]	Agc → Ggc
rs2059993746	VCV002138714	CM1010330	34	S [Ser] → R [Arg]	agC → agA
rs1556619338	VCV000536456	CM014598	39	G [Gly] → V [Val]	gGc → gTc
rs199639961	VCV000952469	---	40	V [Val] → M [Met]	Gtg → Atg
---	VCV001722010	---	41	Y [Tyr] → N [Asn]	Tac → Aac
---	---	CM110891	42	C [Cys] → Y [Tyr]	tGc → tAc
rs2059993925	VCV000944992	CM1010331	44	C [Cys] → Y [Tyr]	tGt → tAt
---	VCV001718192	---	---	Y [Tyr] → C [Cys]	tAt → tGt
---	---	CM108046	48	H [His] → P [Pro]	cAc → cCc
rs2147531318	VCV001410736	CM012659	49	G [Gly] → V [Val]	gGt → gTt
---	VCV002444355	CM994330	53	T [Thr] → I [Ile]	aCa → aTa
rs2147531326	VCV001686182	---	54	Y [Tyr] → N [Asn]	Tac → Aac
---	---	CM002855	54	Y [Tyr] → C [Cys]	tAc → tGc
rs111033629	VCV000871639	CM110892	55	R [Arg] → Q [Gln]	cGa → cAa

Continuation of Table S1

SNP ID	ClinVar ID	HGMD ID	Position	Residue Change	Impacted codons
rs111033630	VCV000010910	CM014961	55	R [Arg] → L [Leu]	cGa → cTa
---	---	CM058456	55	R [Arg] → P [Pro]	cGa → cCa
---	---	CM034449	57	S [Ser] → P [Pro]	Tcc → Ccc
---	---	CM175598	64	W [Trp] → C [Cys]	tgG → tgT
rs770028051	VCV002139610	---	66	A [Ala] → P [Pro]	Gct → Cct
---	---	CM169895	67	E [Glu] → D [Asp]	gaG → gaC
---	---	CM110890	67	E [Glu] → G [Gly]	gAg → gGg
rs111033627	VCV000010905	CM981811	68	T [Thr] → I [Ile]	aCa → aTa
rs954824608	VCV000963742	---	73	H [His] → Q [Gln]	caT → caA
---	---	CM1010333	80	I [Ile] → R [Arg]	aTa → aGa
---	VCV001805004	CM066231	84	I [Ile] → T [Thr]	aTt → aCt
---	---	CM002857	87	F [Phe] → S [Ser]	tTt → tCt
---	---	CM025343	88	Q [Gln] → P [Pro]	cAg → cCg
rs954824608	VCV001397679	---	91	D [Asp] → E [Glu]	gaT → gaG
---	---	CM034450	93	G [Gly] → D [Asp]	gGc → gAc
---	---	CM137748	98	L [Leu] → P [Pro]	cTg → cCg
---	---	CM002858	99	Q [Gln] → P [Pro]	cAg → cCg
rs111033626	VCV000010904	CM981812	101	P [Pro] → L [Leu]	cCa → cTa
---	---	CM981812	102	V [Val] → G [Gly]	gTt → gGt
rs757900199	VCV001425284	---	113	G [Gly] → V [Val]	gGt → gTt
rs2147534671	VCV001488910	---	117	I [Ile] → M [Met]	atA → atG
rs2060068693	VCV000858571	---	121	P [Pro] → L [Leu]	cCt → cTt
rs2060068721	VCV000913201	---	124	C [Cys] → S [Ser]	tGc → tCc

Table S2: Results of the analysis of missense SNPs by the PredictSNP1.0 web server [22]. Mutations marked in orange correspond to SNPs with at least six tests showing deleterious results.

Mutation	PredictSNP	MAPP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
M1I	N	D	N	N	N	D	N
M1K	D	D	D	D	D	D	N
M1T	D	D	N	D	D	D	N
M1V	N	D	N	N	D	D	N
D2G	D	D	N	D	D	N	N
A3S	N	N	N	N	N	N	N
Y7C	D	D	D	D	D	D	D
H8D	D	D	D	D	D	D	D
H8P	D	D	D	D	D	D	D
S12I	D	D	D	D	D	D	D
G16D	D	D	D	D	D	N	N

D - Deleterious; N - Neutral

Continuation of Table S2

Mutation	PredictSNP	MAPP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
A22P	N	D	N	N	N	N	N
D26V	D	D	D	D	D	D	D
G27D	D	D	D	D	D	D	D
G27S	D	D	D	D	D	D	D
S28R	D	D	D	D	D	D	N
Y29C	D	D	D	D	D	D	D
Y29H	D	D	D	D	D	D	D
L30S	D	D	D	D	D	D	D
L31P	D	D	D	D	D	D	D
R32S	D	D	D	D	D	D	D
R32T	D	D	D	D	D	D	D
D33Y	D	D	D	D	D	D	D
S34G	D	D	D	D	D	D	D
S34R	D	D	D	D	D	D	D
G39V	D	D	D	D	D	D	D
V40M	N	D	N	N	D	N	N
Y41N	D	D	D	D	D	D	D
C42Y	D	D	D	D	D	D	D
C44Y	D	D	D	D	D	D	D
Y47C	N	N	D	N	N	N	N
H48P	D	D	D	N	N	D	N
G49V	D	D	D	D	D	D	N
T53I	D	D	D	D	D	D	D
Y54N	D	D	D	D	D	D	D
Y54C	D	D	D	D	D	D	D
R55Q	D	D	D	N	D	D	D
R55L	D	D	D	D	D	D	D
R55P	D	D	D	D	D	D	D
S57P	D	D	D	D	D	N	N
W64C	D	D	N	D	D	D	D
A66P	D	D	N	N	N	D	D
E67D	N	D	N	N	N	N	N
E67G	D	D	N	N	N	D	D
T68I	D	D	D	N	D	D	N
H73Q	N	N	N	N	N	N	N
I80R	D	D	D	D	D	D	D
I84T	D	D	D	D	D	D	D
F87S	D	D	N	D	D	D	D
Q88P	D	D	D	D	D	D	D
D91E	N	D	N	N	N	D	N

D - Deleterious; N – Neutral

Continuation of Table S2

Mutation	PredictSNP	MAPP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
G93D	D	D	D	D	D	D	D
L98P	D	D	D	D	D	D	D
Q99P	N	D	D	N	N	N	N
P101L	D	D	D	D	D	D	D
V102G	D	D	D	D	D	D	D
G113V	N	N	N	N	N	N	N
I117M	N	N	N	N	N	N	N
P121L	N	---	N	N	N	D	N
C124S	N	---	N	N	N	D	N

D - Deleterious; N – Neutral

Table S3: Results of the structural stability analysis of proteins with missense SNPs using the I-Mutant2.0 [34] and MUpro [35] web servers. Mutations marked in orange correspond to SNPs with at least three tests showing deleterious results.

Mutation	I-Mutant2.0 DDG value (kcal/mol)	MUpro DDG value	MUpro SVM value	MUpro NN value (kcal/mol)
M1K	-0,75 (Decrease stability)	-2,64 (Decrease stability)	-1,00 (Decrease stability)	-0,95 (Decrease stability)
Y7C	1,18 (Increase stability)	-0,89 (Decrease stability)	-1,00 (Decrease stability)	-0,82 (Decrease stability)
H8D	-0,85 (Decrease stability)	-1,16 (Decrease stability)	-0,40 (Decrease stability)	-0,80 (Decrease stability)
H8P	-0,32 (Decrease stability)	-0,96 (Decrease stability)	0,18 (Increase stability)	0,53 (Increase stability)
S12I	0,68 (Increase stability)	0,53 (Increase stability)	-0,01 (Decrease stability)	0,60 (Increase stability)
D26V	-0,10 (Decrease stability)	-0,17 (Decrease stability)	0,95 (Increase stability)	0,90 (Increase stability)
G27D	-0,75 (Decrease stability)	-0,17 (Decrease stability)	0,01 (Increase stability)	-0,75 (Decrease stability)
G27S	-0,70 (Decrease stability)	-0,52 (Decrease stability)	-0,84 (Decrease stability)	-0,71 (Decrease stability)
S28R	-0,88 (Decrease stability)	-1,05 (Decrease stability)	0,30 (Increase stability)	0,89 (Increase stability)
Y29C	0,41 (Increase stability)	-0,64 (Decrease stability)	-0,01 (Decrease stability)	0,68 (Increase stability)
Y29H	-0,44 (Decrease stability)	-0,81 (Decrease stability)	0,19 (Increase stability)	0,62 (Increase stability)
L30S	-3,05 (Decrease stability)	-1,95 (Decrease stability)	-1,00 (Decrease stability)	-1,00 (Decrease stability)
L31P	-1,09 (Decrease stability)	-2,65 (Decrease stability)	0,53 (Increase stability)	0,98 (Increase stability)
R32S	-3,47 (Decrease stability)	-1,31 (Decrease stability)	-1,00 (Decrease stability)	0,98 (Increase stability)
R32T	-2,25 (Decrease stability)	-1,68 (Decrease stability)	-1,00 (Decrease stability)	-0,95 (Decrease stability)
D33Y	-1,18 (Decrease stability)	-0,86 (Decrease stability)	0,31 (Increase stability)	0,63 (Increase stability)
S34G	-2,91 (Decrease stability)	-1,36 (Decrease stability)	-0,93 (Decrease stability)	-0,75 (Decrease stability)
S34R	-1,47 (Decrease stability)	-0,64 (Decrease stability)	0,35 (Increase stability)	-0,56 (Decrease stability)
G39V	0,39 (Increase stability)	0,06 (Increase stability)	0,55 (Increase stability)	0,54 (Increase stability)
Y41N	-1,84 (Decrease stability)	-0,40 (Decrease stability)	-0,82 (Decrease stability)	-0,54 (Decrease stability)
C42Y	-0,35 (Decrease stability)	-0,67 (Decrease stability)	-0,04 (Decrease stability)	-0,85 (Decrease stability)
C44Y	0,14 (Increase stability)	-1,14 (Decrease stability)	0,09 (Increase stability)	0,70 (Increase stability)
G49V	1,04 (Increase stability)	-1,01 (Decrease stability)	0,20 (Increase stability)	0,63 (Increase stability)
T53I	0,21 (Increase stability)	-0,12 (Decrease stability)	-0,71 (Decrease stability)	-0,70 (Decrease stability)
Y54N	-1,64 (Decrease stability)	-2,01 (Decrease stability)	-0,60 (Decrease stability)	-0,78 (Decrease stability)
Y54C	0,01 (Increase stability)	-1,80 (Decrease stability)	-0,26 (Decrease stability)	-0,73 (Decrease stability)
R55Q	-1,00 (Decrease stability)	-0,49 (Decrease stability)	-0,72 (Decrease stability)	-0,98 (Decrease stability)
R55L	-0,53 (Decrease stability)	0,19 (Increase stability)	-0,06 (Decrease stability)	0,73 (Increase stability)
R55P	-0,42 (Decrease stability)	-0,77 (Decrease stability)	-0,36 (Decrease stability)	-0,87 (Decrease stability)
W64C	-0,90 (Decrease stability)	-0,42 (Decrease stability)	0,35 (Increase stability)	0,68 (Increase stability)
I80R	-1,51 (Decrease stability)	-1,66 (Decrease stability)	-1,00 (Decrease stability)	-1,00 (Decrease stability)
I84T	-2,24 (Decrease stability)	-1,87 (Decrease stability)	-1,00 (Decrease stability)	-0,95 (Decrease stability)
F87S	-1,24 (Decrease stability)	-2,23 (Decrease stability)	-1,00 (Decrease stability)	-1,00 (Decrease stability)
Q88P	-2,44 (Decrease stability)	-1,92 (Decrease stability)	0,11 (Increase stability)	0,70 (Increase stability)
G93D	-0,66 (Decrease stability)	-0,71 (Decrease stability)	-0,81 (Decrease stability)	-0,81 (Decrease stability)
L98P	-0,93 (Decrease stability)	-1,10 (Decrease stability)	-0,24 (Decrease stability)	-0,81 (Decrease stability)
P101L	-0,38 (Decrease stability)	-0,03 (Decrease stability)	0,68 (Increase stability)	0,96 (Increase stability)
V102G	-4,37 (Decrease stability)	-2,03 (Decrease stability)	-1,00 (Decrease stability)	-0,98 (Decrease stability)

DDG - Prediction of the value of energy change; SVM - Support vector machines; NN - Neural network

Table S4: Degree of conservation, molecular pathological mechanism and physicochemical characteristics of selected missense SNPs. Results obtained through ConSurf, MutPred2 and HOPE web servers.

Position	ConSurf Web Server		Variant	MutPred2 Web Server	InterPro domain	Hope Web Server
	Conservation score	Prediction		Molecular mechanisms of pathogenicity (P < 0.05)		Characteristic
M1	9	Highly conserved and exposed (f)	M1K	Altered Metal binding Gain of Ubiquitylation at M1 Loss of N-terminal acetylation at M1	IPR017289	The mutant residue is larger and less hydrophobic than the wild-type residue, which can cause repulsion between neighbouring residues. It has a positive charge, unlike the wild residue, which is neutral.
Y7	8	Conserved and exposed (f)	Y7C	Altered Metal binding Altered Ordered interface Gain of Helix		The mutant residue is smaller and more hydrophobic than the wild-type residue, which can cause repulsion between neighbouring residues. The mutation will leave a space in the core of the protein and there will be folding disorders resulting from the loss of hydrogen bonds.
H8	9	Highly conserved and buried (s)	H8D	Altered Metal binding Altered Ordered interface Altered DNA binding Gain of Sulfation at Y7	IPR000980	The mutated residue is smaller than the wild residue and can cause repulsion between neighbouring residues. The mutation is negatively charged, while the wild-type residue is neutral. May cause loss of external interactions.
G27	9	Highly conserved and buried (s)	G27D	Altered Metal binding	IPR036860	The mutant residue is larger and less hydrophobic than the wild-type residue. Since the wild-type residue is buried in the core of the protein, the mutant will not fit, leading to folding and conformation problems.
				Gain of Relative solvent accessibility	IPR017289	
				Gain of Allosteric site at R32	IPR035876	
			G27S	Altered Ordered interface		The mutant residue is larger than the wild-type residue. Since the wild-type residue is buried in the core of the protein, the mutant will not fit, leading to folding and conformation problems.
				Gain of Catalytic site at R32		
				Gain of Sulfation at Y29		
				Altered Metal binding		
				Gain of Relative solvent accessibility		
				Gain of Allosteric site at R32		
				Altered Ordered interface		
				Gain of Catalytic site at R32		

P - p-value

Continuation of Table S4

Position	ConSurf Web Server		Variant	MutPred2 Web Server	Hope Web Server	
	Conservation score	Prediction		Molecular mechanisms of pathogenicity (P < 0.05)	InterPro domain	Characteristic
L30	9	Highly conserved and buried (s)	L30S	Altered Metal binding		The mutant residue is smaller and less hydrophobic than the wild-type residue. The mutation will cause loss of hydrophobic interactions and will leave a space in the core of the protein.
				Gain of Relative solvent accessibility		
				Gain of Allosteric site at R32		
				Altered Ordered interface		
				Altered Stability		
				Gain of Intrinsic disorder		
				Loss of Loop		
R32	9	Highly conserved and buried (s)	R32S	Gain of Catalytic site at R32	IPR000980	The mutant residue is smaller and more hydrophobic than the wild-type residue, leading to the modification of hydrogen bonds, causing a possible loss of external interactions.The mutation will leave a space in the core of the protein and there will be folding disorders resulting from the loss of hydrogen bonds.
				Altered Metal binding		
				Loss of Allosteric site at R32		
				Altered Ordered interface		
			R32T	Altered Stability	IPR017289 IPR035876	The mutant residue is smaller and more hydrophobic than the wild-type residue, leading to the modification of hydrogen bonds, causing disturbs in the folding and a possible loss of external interactions. The mutant residue is neutrally charged, unlike the wild-type residue.The mutation will leave a space in the core of the protein.
				Loss of Catalytic site at R32		
				Gain of Sulfation at Y29		
				Loss of Allosteric site at R32		
				Altered Ordered interface		
				Altered Stability		
S34	9	Highly conserved and buried (s)	S34G	Loss of Catalytic site at R32		The mutant residue is smaller than the wild-type residue, leading to modification of the hydrogen bonds and creating a space in the protein core.
				Altered Ordered interface		
				Loss of Allosteric site at R32		
			S34R	Altered Ordered interface		The mutant residue is larger and less hydrophobic than the wild-type residue and may not fit, causing a loss of hydrophobic bonds in the protein core. The mutant residue is positively charged, unlike the neutrally charged wild-type residue, which leads to folding problems.
				Loss of Loop		
				Gain of Catalytic site at R32		
Y41	6	Highly average and exposed	Y41N	Loss of Strand		The mutant residue is smaller and less hydrophobic than the wild-type residue, causing loss of hydrophobic bonds and leaving a space in the core of the protein.
				Altered Ordered interface		

P – p-value

Continuation of [Table S4](#)

Position	ConSurf Web Server		Variant	MutPred2 Web Server	InterPro domain	Hope Web Server
	Conservation score	Prediction		Molecular mechanisms of pathogenicity (P < 0.05)		Characteristic
C42	9	Highly conserved and buried (s)	C42Y	Altered Ordered interface		The mutant residue is larger and less hydrophobic than the wild-type residue and may not fit, causing loss of hydrophobic bonds in the protein core and a possible loss of external interactions.
T53	9	Highly conserved and buried (s)	T53I	Altered Transmembrane protein Altered Ordered interface		The mutated residue is larger and more hydrophobic than the wild-type residue, causing a possible loss of external interactions. The residue is located on the surface of the protein, mutation of this residue can disrupt interactions with other molecules or other parts of the protein.
Y54	9	Highly conserved and buried (s)	Y54N	Altered Metal binding Altered Ordered interface	IPR000980 IPR036860	The mutant residue is smaller and less hydrophobic than the wild-type residue, affecting the formation of hydrogen bonds and causing a possible loss of external interactions.
			Y54C	Altered Stability Altered Transmembrane protein	IPR017289 IPR035876	The mutant residue is smaller and more hydrophobic than the wild-type residue, affecting the formation of hydrogen bonds and causing a possible loss of external interactions.
R55	9	Highly conserved and exposed (f)	R55Q	Altered Metal binding Altered Ordered interface		The mutant residue is smaller than the wild-type residue, affecting the formation of hydrogen bonds. It has a positive charge, while the wild-type residue has a neutral charge. Possible loss of external interactions.
			R55P	Loss of Strand Altered Transmembrane protein		The mutant residue is smaller and more hydrophobic than the wild-type residue, affecting the formation of hydrogen bonds. It has a positive charge, while the wild-type residue has a neutral charge. Possible loss of external interactions.

P - p-value

Continuation of Table S4

Position	ConSurf Web Server		Variant	MutPred2 Web Server	InterPro domain	Hope Web Server
	Conservation score	Prediction		Molecular mechanisms of pathogenicity (P < 0.05)		Characteristic
I80	6	Highly average and buried	I80R	Altered Ordered interface	IPR000980 IPR036860 IPR017289 IPR035876	The mutant residue is larger and less hydrophobic than the wild-type residue, which means that it doesn't fit and there is a loss of hydrophobic bonds in the core of the protein. The mutant residue is positively charged, unlike the wild type which is neutrally charged. Thus, the mutation leads to folding disorders.
				Loss of Loop		
				Loss of Acetylation at K81		
				Altered Disordered interface		
I84	8	Conserved and buried	I84T	Altered Stability		The mutant residue is smaller and less hydrophobic than the wild-type residue, leading to a gap and a loss of hydrophobic bonds in the protein core.
				Altered Ordered interface		
				Loss of Acetylation at K89		
				Gain of Allosteric site at F87		
F87	5	Average and buried	F87S	Gain of N-linked glycosylation at N82		The mutant residue is smaller and less hydrophobic than the wild-type residue, leading to a gap and a loss of hydrophobic bonds in the protein core.
				Altered Stability		
				Loss of Acetylation at K89		
				Gain of Relative solvent accessibility		
G93	9	Highly conserved and exposed (F)	G93D	Gain of B-factor		The mutant residue is larger and less hydrophobic than the wild-type residue, leading to an incorrect conformation of the protein and disturbance with other molecules. The charge of the mutant residue is negative, unlike the wild type, which is neutral, leading to repulsion from neighbouring residues.
				Altered Disordered interface		
				Altered Metal binding		
				Loss of Acetylation at K89		
L98	9	Highly conserved and buried (S)	L98P	Altered Ordered interface	IPR000980 IPR036860 IPR017289 IPR035876	The mutant residue is smaller than the wild-type residue, leading to an empty space in the protein core.
				Loss of Strand		
				Gain of Intrinsic disorder		
				Altered Stability		
V102	8	Highly conserved and buried	V102G	Gain of B-factor		The mutant residue is larger and less hydrophobic than the wild-type residue, leading to the possible loss of hydrophobic bonds and the creation of a space in the protein core.
				Gain of Intrinsic disorder		
				Altered Stability		
				Gain of B-factor		

P - p-value

Table S5: Average RMSD values of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, *p*-values are provided for the entire simulation duration and at specific time intervals, offering insights into structural changes and significance. *p*-values above 0,05 marked in italic.

Proteins	WT	M1K	Y7C	H8D	G27D	G27S	L30S	R32S	R32T	S34G	S34R	Y41N	C42Y	T53I	Y54C	Y54N	R55P	R55Q	I80R	I84T	F87S	G93D	L98P	V102G
$\bar{x}$	0,683	0,602	0,493	0,000	0,565	0,864	0,581	0,432	0,617	0,532	0,520	0,803	0,567	0,795	0,840	0,327	1,123	0,459	0,412	0,443	0,735	0,564	0,967	0,667
σ	0,115	0,058	0,080	0,087	0,069	0,165	0,148	0,044	0,112	0,157	0,055	0,154	0,111	0,177	0,126	0,045	0,131	0,070	0,058	0,093	0,152	0,094	0,166	0,071
$\Delta\bar{x}$	---	-0,081	-0,190	-0,683	-0,118	0,181	-0,102	-0,251	-0,066	-0,151	-0,163	0,120	-0,116	0,112	0,157	-0,356	0,440	-0,224	-0,271	-0,240	0,052	-0,119	0,284	-0,016
P-Value (Total)	---	0,000	0,000	0,527	0,000	0,000	0,000	0,000	0,575	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (5 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,117	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,258	0,000	0,000	0,000
P-Value (10 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (15 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (20 ns)	---	0,315	0,000	0,373	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (25 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,847	0,000	0,000

$\bar{x}$ - Average of the RMSD values; σ - Standard deviations and $\Delta\bar{x}$ - Difference between the average of the RMSD values between the mutant and the wild type.

Table S6: Average RMSF values of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, *p*-values are provided for the entire structure and at specific amino acid (aa) intervals, offering insights into structural changes and significance. *p*-values above 0,05 marked in italic.

Proteins	WT	M1K	Y7C	H8D	G27D	G27S	L30S	R32S	R32T	S34G	S34R	Y41N	C42Y	T53I	Y54C	Y54N	R55P	R55Q	I80R	I84T	F87S	G93D	L98P	V102G
$\bar{x}$	0,207	0,149	0,192	0,286	0,191	0,318	0,298	0,140	0,209	0,340	0,171	0,246	0,212	0,405	0,253	0,139	0,388	0,187	0,130	0,174	0,294	0,308	0,333	0,209
σ	0,159	0,123	0,136	0,177	0,157	0,257	0,191	0,090	0,117	0,205	0,114	0,182	0,128	0,241	0,133	0,073	0,245	0,138	0,084	0,114	0,209	0,205	0,232	0,156
$\Delta\bar{x}$	---	-0,058	-0,015	0,079	-0,016	0,111	0,091	-0,067	0,002	0,133	-0,036	0,039	0,005	0,198	0,046	-0,068	0,181	-0,020	-0,077	-0,033	0,087	0,101	0,126	0,002
P-Value (Total)	---	0,001	0,420	0,000	0,441	0,000	0,000	0,000	0,906	0,000	0,039	0,067	0,766	0,000	0,012	0,000	0,000	0,293	0,000	0,066	0,000	0,000	0,000	0,913
P-Value (1-32 aa)	---	0,133	0,816	0,000	0,270	0,022	0,000	0,570	0,710	0,000	0,281	0,049	0,367	0,000	0,001	0,072	0,000	0,216	0,000	0,453	0,000	0,000	0,000	0,706
P-Value (33-64 aa)	---	0,031	0,239	0,000	0,645	0,000	0,000	0,603	0,196	0,000	0,117	0,041	0,012	0,000	0,000	0,108	0,000	0,304	0,002	0,319	0,002	0,003	0,000	0,580
P-Value (65-96 aa)	---	0,000	0,071	0,001	0,463	0,000	0,000	0,044	0,023	0,000	0,385	0,047	0,500	0,000	0,001	0,407	0,000	0,454	0,021	0,487	0,000	0,000	0,000	0,019
P-Value (97-128 aa)	---	0,013	0,025	0,015	0,607	0,004	0,006	0,000	0,230	0,001	0,004	0,249	0,238	0,000	0,842	0,000	0,000	0,232	0,000	0,030	0,034	0,001	0,001	0,612

$\bar{x}$ - Average of the RMSF values; σ - Standard deviations and $\Delta\bar{x}$ - Difference between the average of the RMSF values between the mutant and the wild type.

Table S7: Average Rg values of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, *p*-values are provided for the entire simulation duration and at specific time intervals, offering insights into structural changes and significance. *p*-values above 0,05 marked in italic.

Proteins	WT	M1K	Y7C	H8D	G27D	G27S	L30S	R32S	R32T	S34G	S34R	Y41N	C42Y	T53I	Y54C	Y54N	R55P	R55Q	I80R	I84T	F87S	G93D	L98P	V102G
$\bar{x}$	1,493	1,474	1,478	1,558	1,478	1,545	1,571	1,471	1,474	1,534	1,481	1,435	1,546	1,599	1,530	1,472	1,563	1,499	1,471	1,487	1,511	1,497	1,499	1,511
σ	0,037	0,025	0,036	0,030	0,031	0,138	0,056	0,023	0,051	0,054	0,035	0,054	0,035	0,094	0,036	0,024	0,058	0,049	0,034	0,028	0,059	0,103	0,087	0,074
$\Delta\bar{x}$	---	-0,019	-0,015	0,065	-0,015	0,052	0,078	-0,022	-0,019	0,041	-0,012	-0,058	0,053	0,106	0,037	-0,021	0,070	0,006	-0,022	-0,006	0,018	0,004	0,006	0,018
P-Value (Total)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,116	0,002	0,000	0,000
P-Value (5 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (10 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,021	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (15 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,026	0,000	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,833
P-Value (20 ns)	---	0,667	0,762	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,949	0,000	0,000	0,000	0,000	0,000	0,000	0,003	0,000	0,057	0,940	0,000	0,000	0,000
P-Value (25 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,030	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

$\bar{x}$ - Average of the Rg values; σ - Standard deviations and $\Delta\bar{x}$ - Difference between the average of the Rg values between the mutant and the wild type.

Table S8: Average hydrogen bonds number of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, *p*-values are provided for the entire simulation duration and at specific time intervals, offering insights into structural changes and significance. *p*-values above 0,05 marked in italic.

Proteins	WT	M1K	Y7C	H8D	G27D	G27S	L30S	R32S	R32T	S34G	S34R	Y41N	C42Y	T53I	Y54C	Y54N	R55P	R55Q	I80R	I84T	F87S	G93D	L98P	V102G
$\bar{x}$	78	79	80	78	84	78	82	85	81	78	82	79	78	82	84	80	76	85	85	81	82	82	80	81
σ	5	4	5	4	6	4	5	5	5	5	5	5	5	5	5	5	5	8	6	4	5	5	5	6
$\Delta\bar{x}$	---	1	2	0	6	0	4	7	3	0	4	1	0	4	6	2	-2	7	7	3	4	4	2	3
P-Value (Total)	---	0,009	0,000	0,001	0,000	0,029	0,000	0,000	0,000	0,044	0,000	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (5 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,110	0,000	0,059	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (10 ns)	---	0,807	0,729	0,000	0,000	0,006	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,824	0,000	0,000	0,001	0,000	0,000	0,079	0,041	0,005	0,000	0,000
P-Value (15 ns)	---	0,000	0,970	0,000	0,000	0,057	0,102	0,000	0,000	0,000	0,000	0,211	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,033	0,000	0,000	0,000	0,000
P-Value (20 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,473	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (25 ns)	---	0,030	0,000	0,829	0,000	0,451	0,000	0,000	0,381	0,000	0,000	0,622	0,000	0,000	0,000	0,013	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

$\bar{x}$ - Average of the hydrogen bonds number; σ - Standard deviations and $\Delta\bar{x}$ - Difference between the average of the hydrogen bonds number between the mutant and the wild type.

Table S9: Average SASA value of both mutant and native proteins, along with their respective standard deviations, highlight the deviation differences compared to the native protein. Additionally, *p*-values are provided for the entire simulation duration and at specific time intervals, offering insights into structural changes and significance. *p*-values above 0,05 marked in italic.

Proteins	WT	M1K	Y7C	H8D	G27D	G27S	L30S	R32S	R32T	S34G	S34R	Y41N	C42Y	T53I	Y54C	Y54N	R55P	R55Q	I80R	I84T	F87S	G93D	L98P	V102G
$\bar{x}$	85,111	83,940	84,480	86,176	85,715	86,975	87,170	80,444	84,263	87,496	87,012	83,340	84,931	87,262	85,510	84,032	88,169	85,418	84,030	84,869	86,476	84,797	86,834	86,248
σ	3,422	2,236	2,900	2,259	2,425	4,765	3,824	2,506	2,593	3,435	2,846	3,192	2,772	3,643	2,847	3,012	2,819	4,051	2,756	2,654	3,714	3,501	2,763	3,166
$\Delta\bar{x}$	---	-1,171	-0,631	1,065	0,604	1,864	2,059	-4,667	-0,848	2,385	1,901	-1,771	-0,180	2,151	0,399	-1,079	3,058	0,307	-1,081	-0,242	1,365	-0,314	1,723	1,137
P-Value (Total)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,041	0,000	0,000	0,000	0,000	0,004	0,000	0,005	0,000	0,001	0,000	0,000
P-Value (5 ns)	---	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,012	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,336	0,000
P-Value (10 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,280	0,000	0,000	0,000	0,192	0,000	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (15 ns)	---	0,838	0,692	0,000	0,000	0,000	0,000	0,000	0,019	0,000	0,000	0,000	0,000	0,000	0,000	0,859	0,000	0,000	0,023	0,000	0,000	0,000	0,000	0,952
P-Value (20 ns)	---	0,000	0,747	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,046	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
P-Value (25 ns)	---	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,162	0,000	0,026	0,000	0,000	0,979	0,000	0,907	0,000	0,034	0,000

$\bar{x}$ - Average of the SASA values; σ - Standard deviations and $\Delta\bar{x}$ - Difference between the average of the SASA values between the mutant and the wild type.

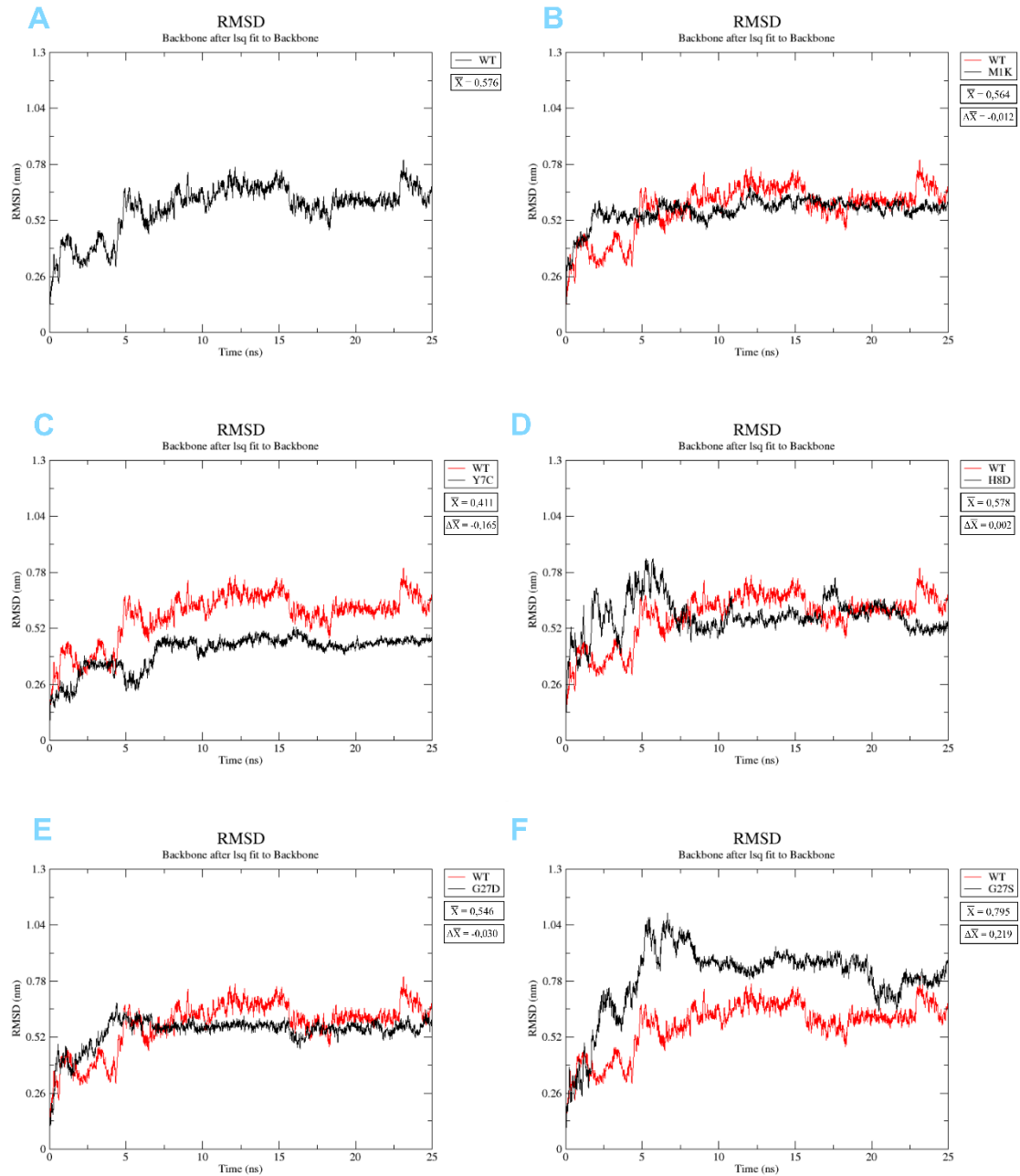


Figure S1: Analysis of the evolution of backbone RMSD of native and mutants SH2D1A protein at 25 ns. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{x}$) Average of the RMSD values and ($\Delta\bar{x}$) difference between the average of the RMSD values between the mutant and the wild type.

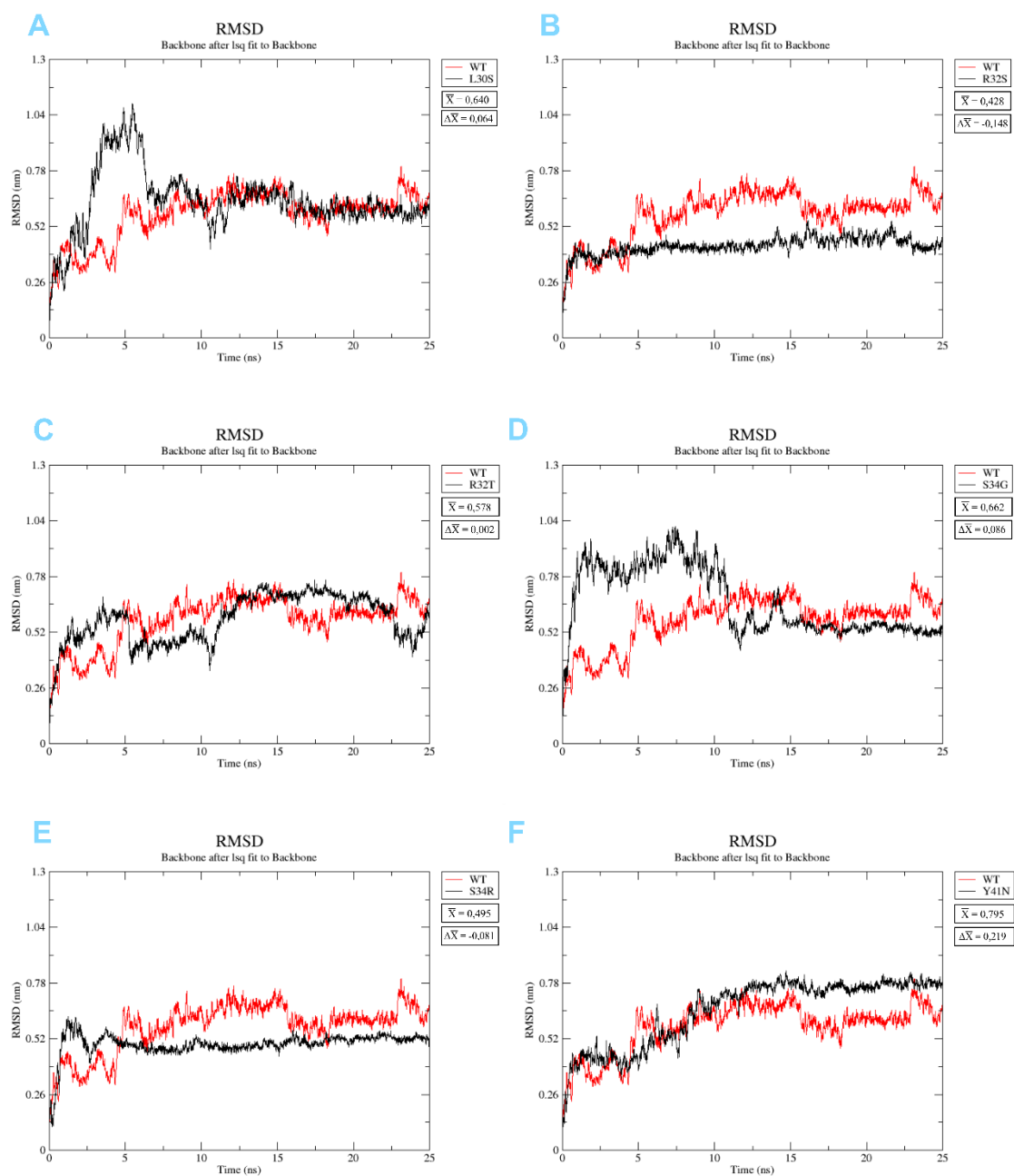


Figure S2: Analysis of the evolution of backbone RMSD of mutants SH2D1A protein at 25 ns. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{x}$) Average of the RMSD values and ($\Delta\bar{x}$) difference between the average of the RMSD values between the mutant and the wild type.

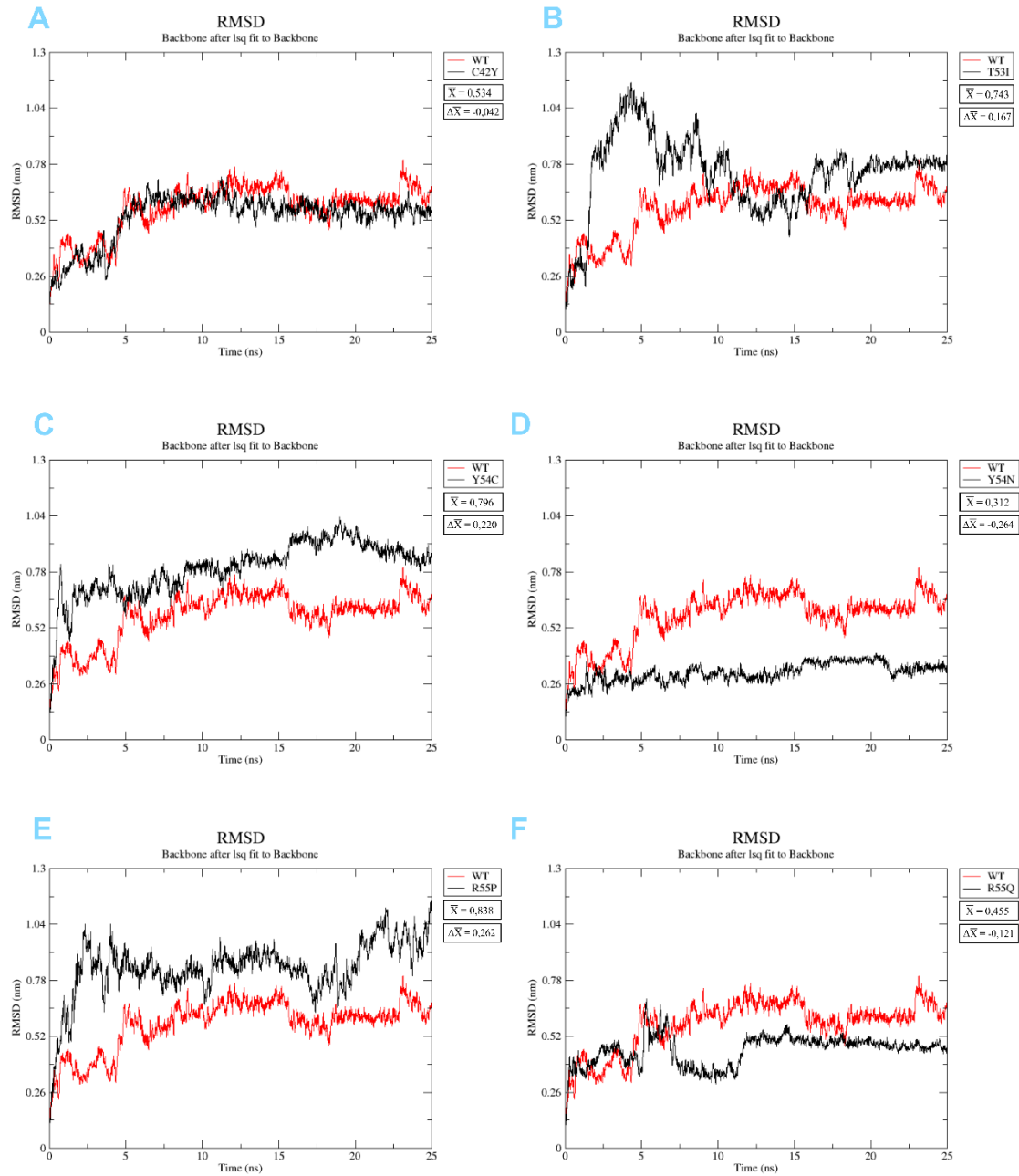


Figure S3: Analysis of the evolution of backbone RMSD of mutants SH2D1A protein at 25 ns. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{x}$) Average of the RMSD values and ($\Delta\bar{x}$) difference between the average of the RMSD values between the mutant and the wild type.

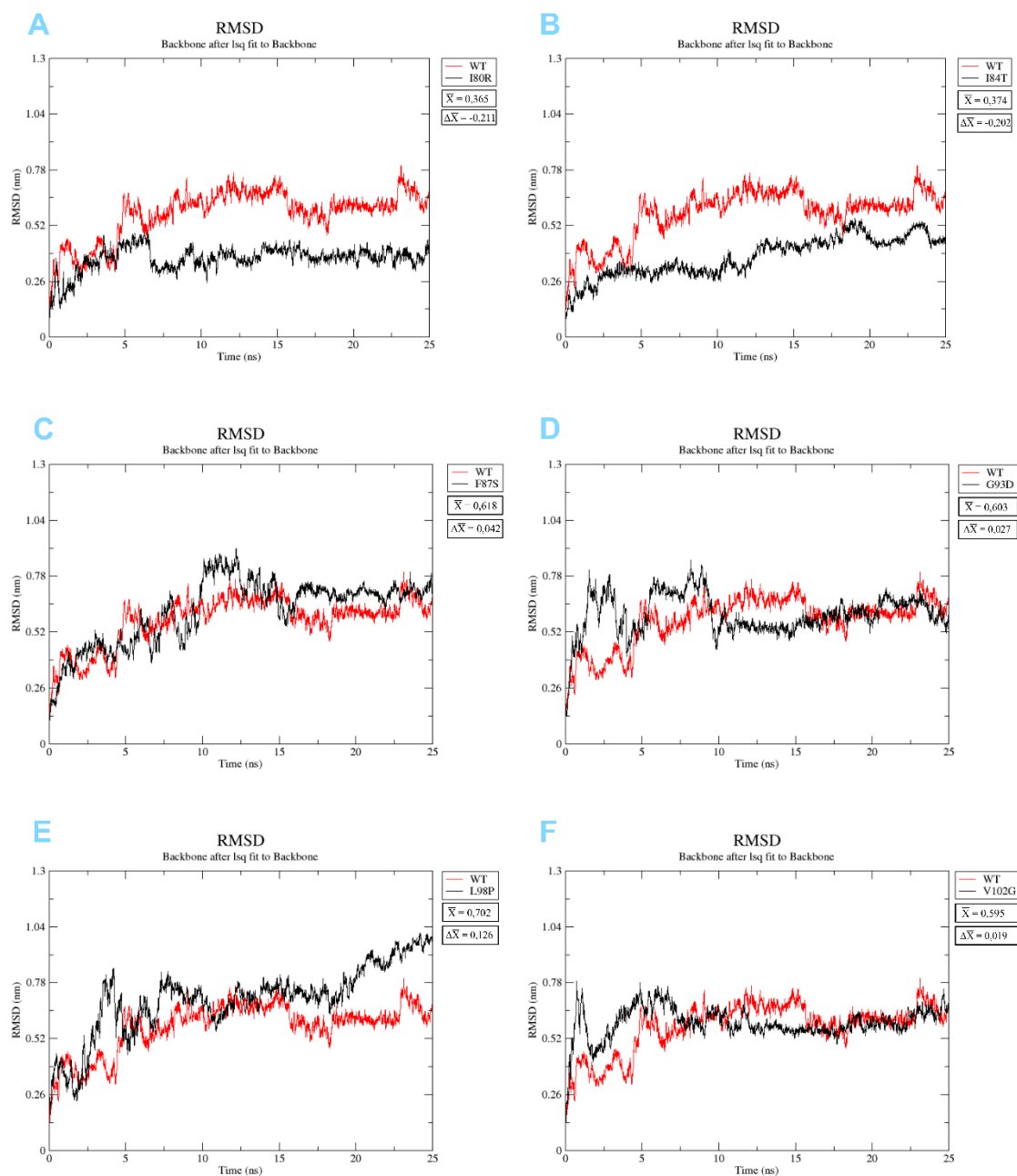


Figure S4: Analysis of the evolution of backbone RMSD of mutants SH2D1A protein at 25 ns. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G. ($\bar{x}$) Average of the RMSD values and ($\Delta\bar{x}$) difference between the average of the RMSD values between the mutant and the wild type.

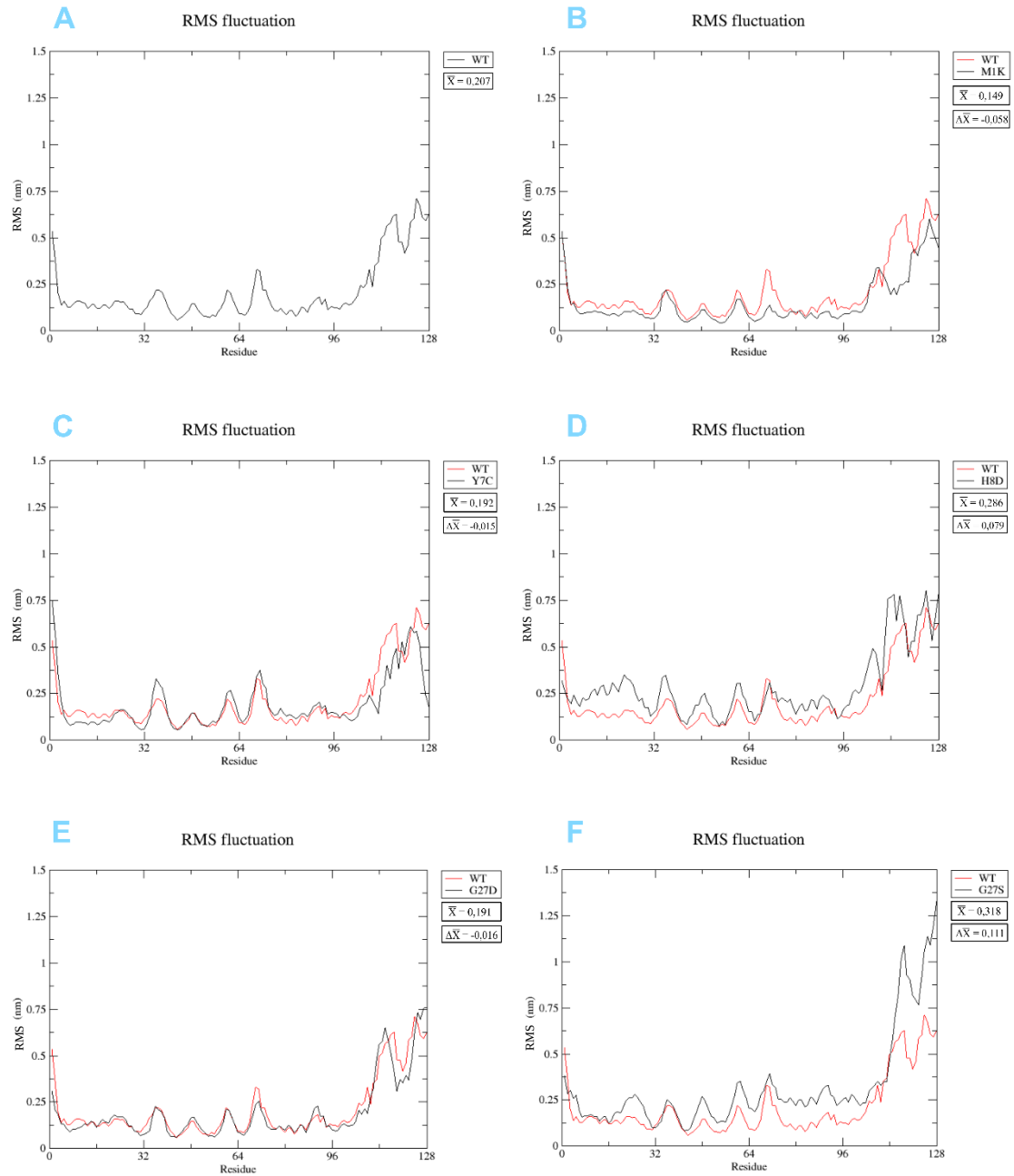


Figure S5: Analysis RMS fluctuation of carbon alpha on native and mutants SH2D1A protein. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{X}$) Average of the RMSF values and ($\Delta\bar{X}$) difference between the average of the RMSF values between the mutant and the wild type.

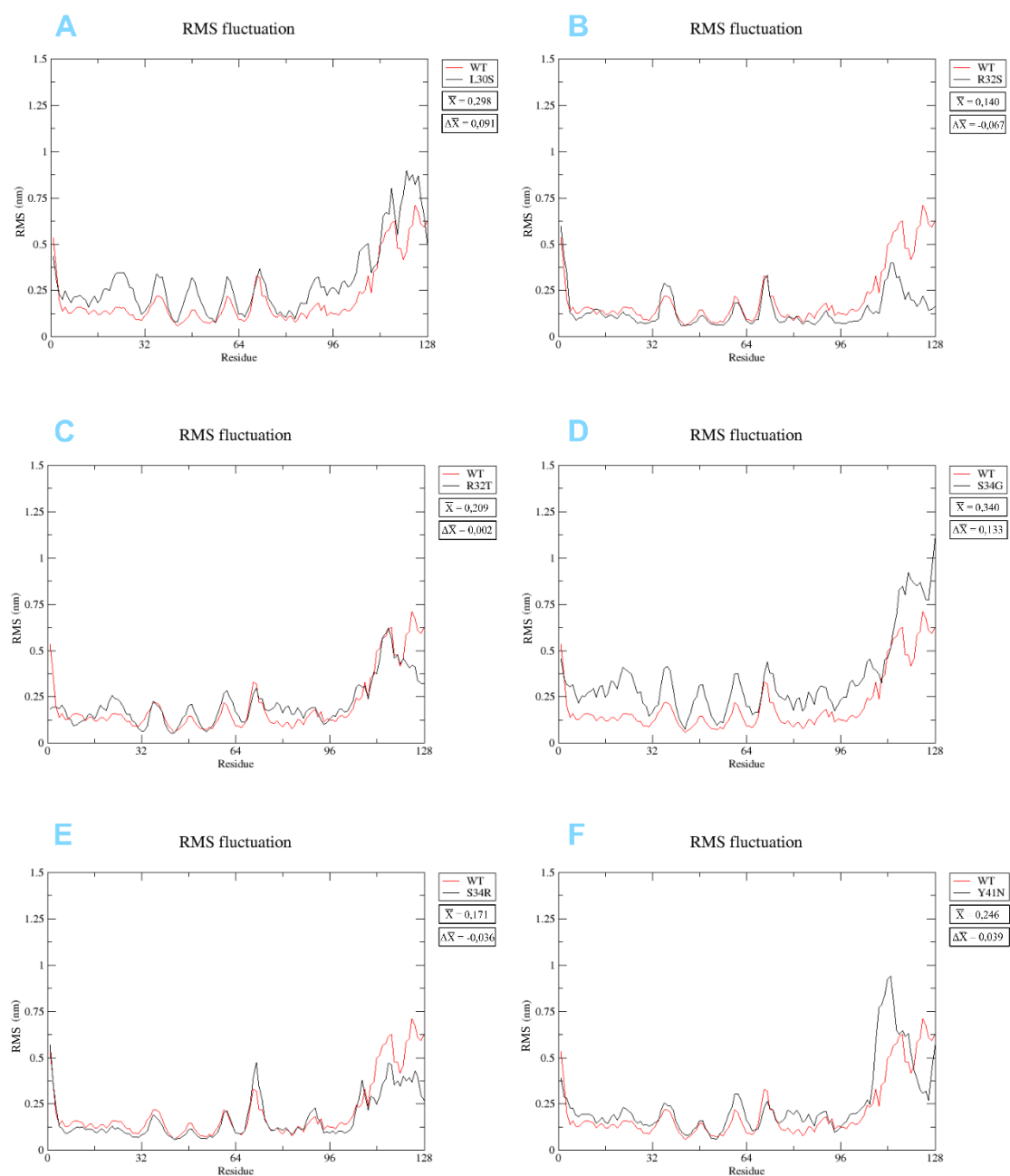


Figure S6: Analysis RMS fluctuation of carbon alpha on mutants SH2D1A protein. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{x}$) Average of the RMSF values and ($\Delta\bar{x}$) difference between the average of the RMSF values between the mutant and the wild type.

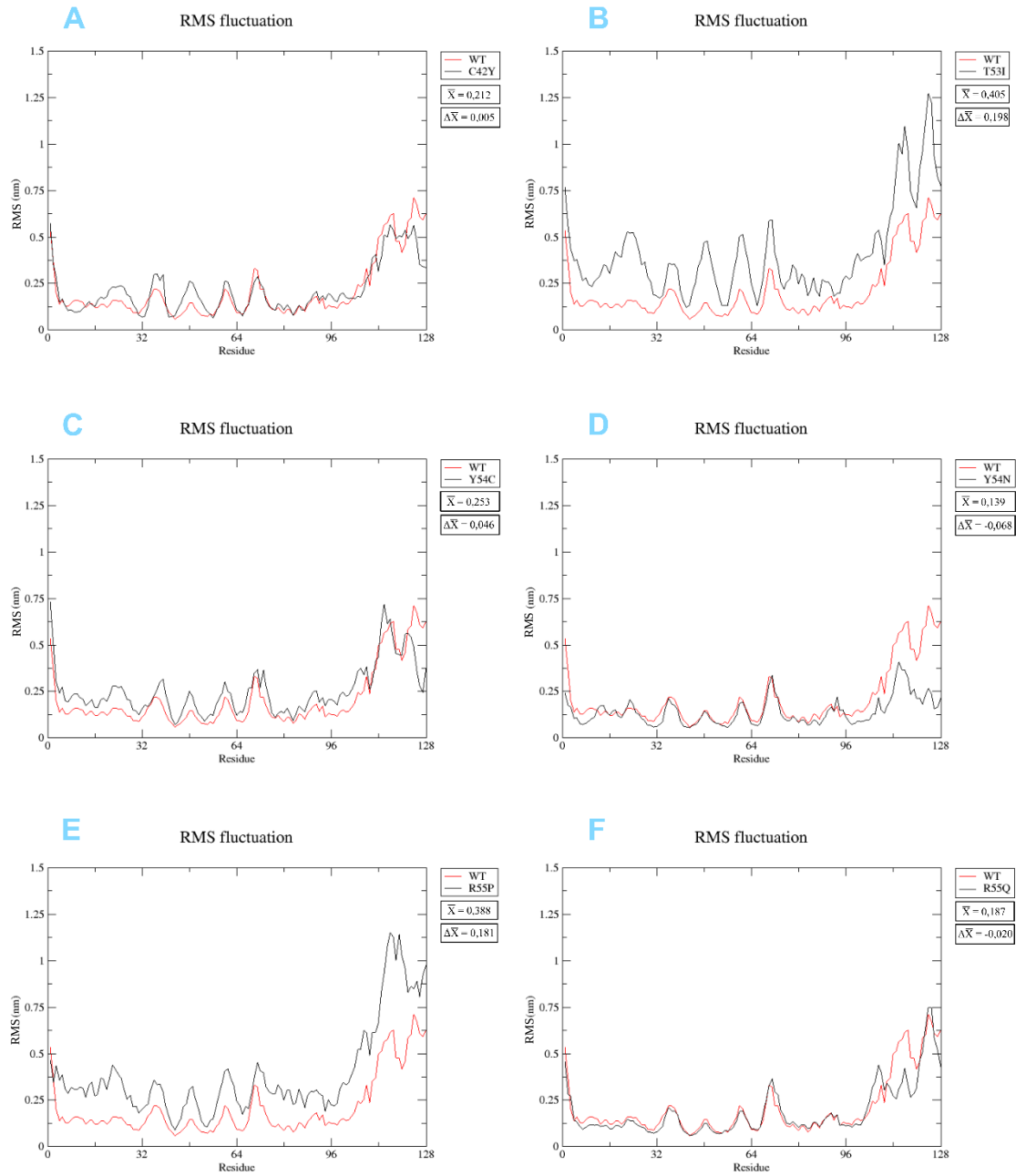


Figure S7: Analysis RMS fluctuation of carbon alpha on mutants SH2D1A protein. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{X}$) Average of the RMSF values and ($\Delta\bar{X}$) difference between the average of the RMSF values between the mutant and the wild type.

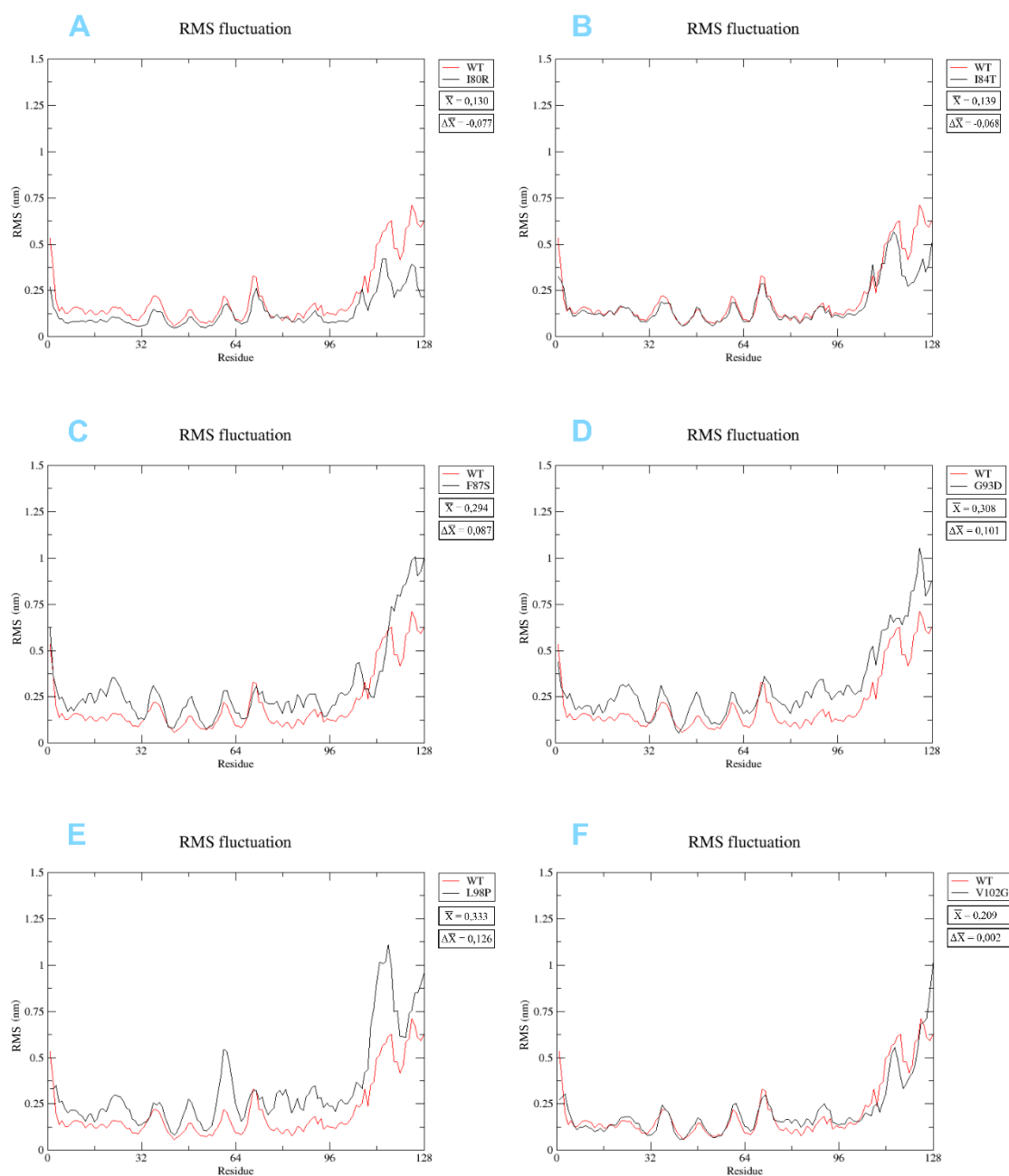


Figure S8: Analysis RMS fluctuation of carbon alpha on mutants SH2D1A protein. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G. ($\bar{x}$) Average of the RMSF values and ($\Delta\bar{x}$) difference between the average of the RMSF values between the mutant and the wild type.

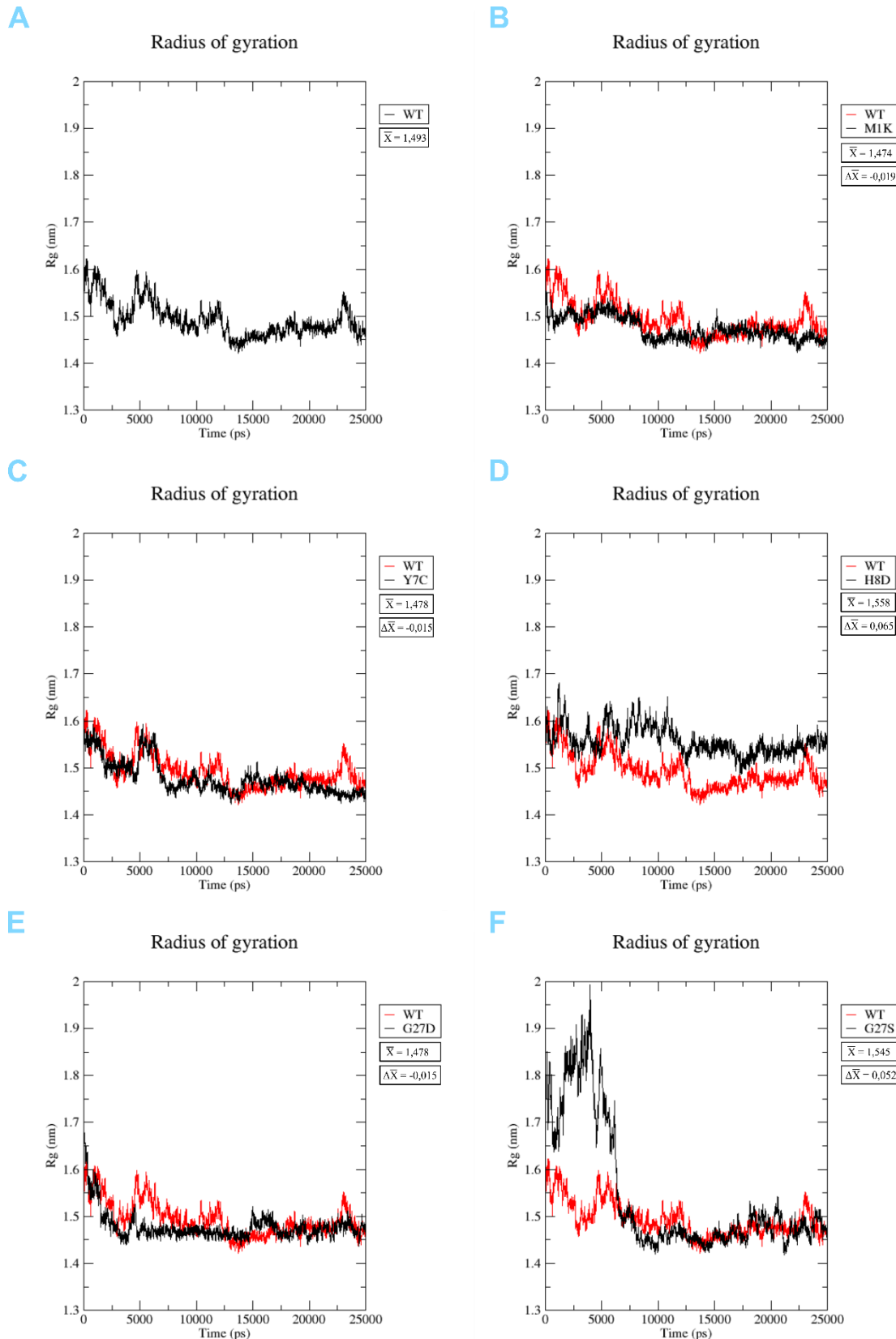


Figure S9: Analysis of radius gyration (Rg) on the backbone of native and mutants SH2D1A protein. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{X}$) Average of the Rg values and ($\Delta\bar{X}$) difference between the average of the Rg values between the mutant and the wild type.

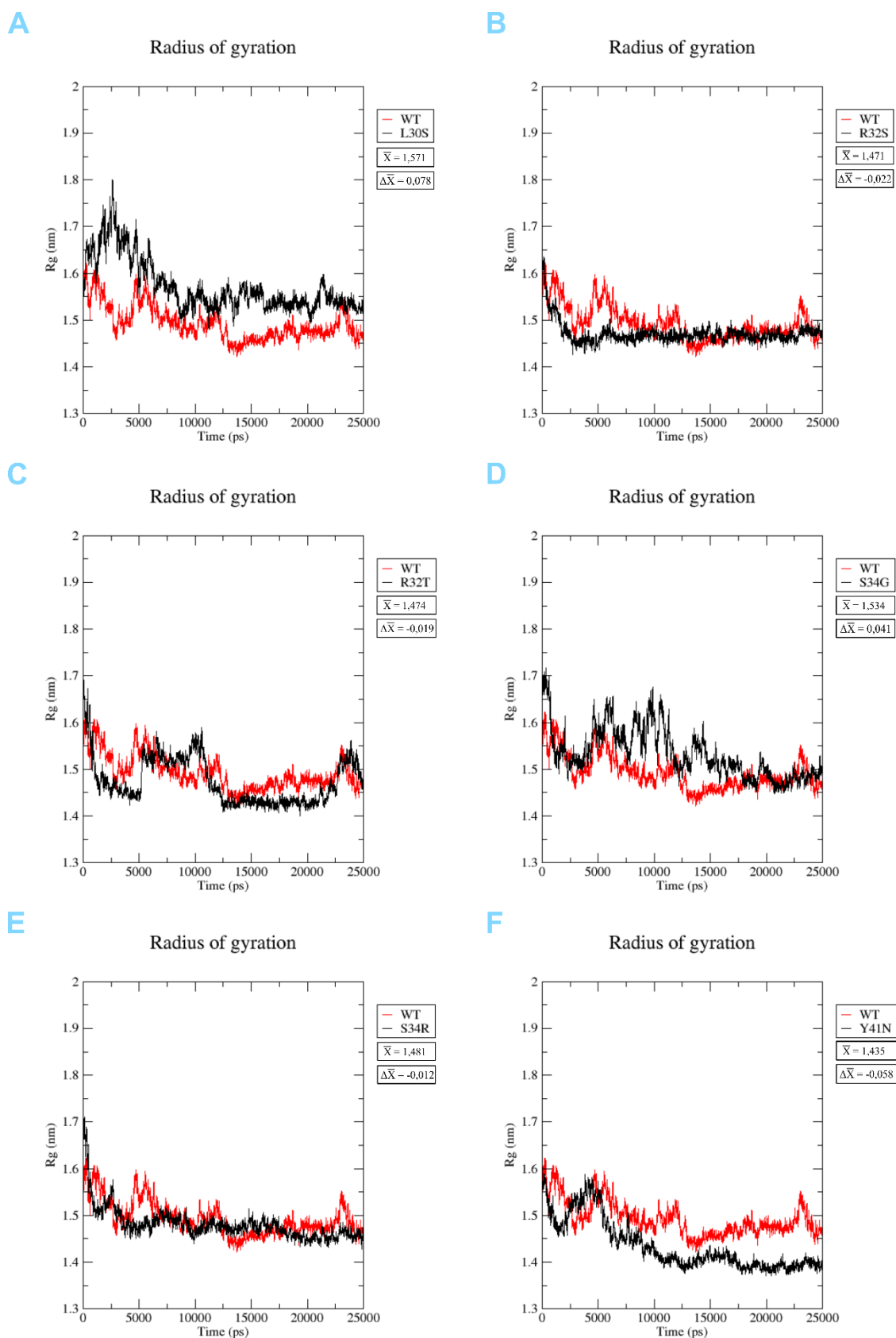


Figure S10: Analysis of radius gyration (Rg) on the backbone of mutants SH2D1A protein. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{x}$) Average of the Rg values and ($\Delta\bar{x}$) difference between the average of the Rg values between the mutant and the wild type.

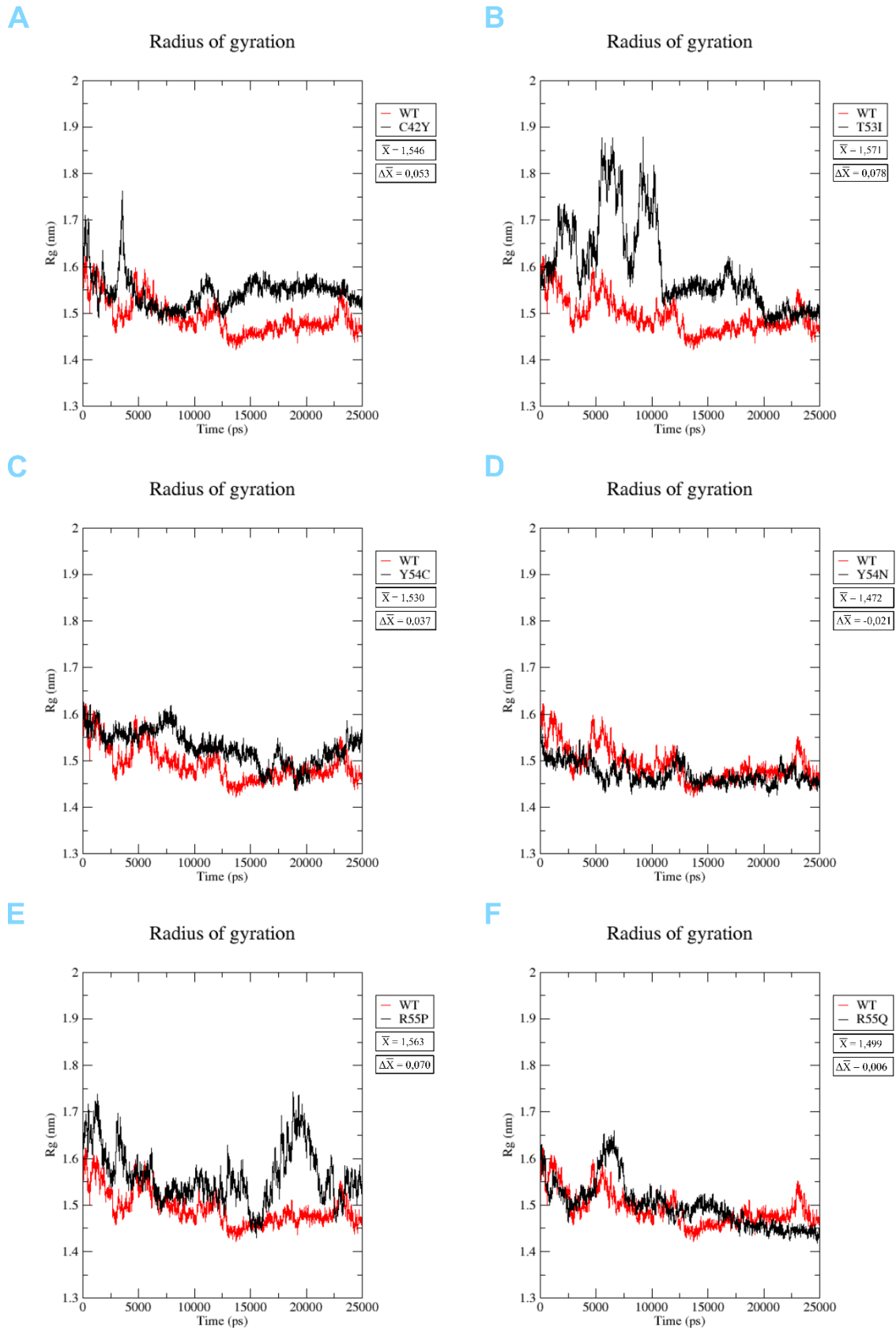


Figure S11: Analysis of radius gyration (Rg) on the backbone of mutants SH2D1A protein. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{x}$) Average of the Rg values and ($\Delta\bar{x}$) difference between the average of the Rg values between the mutant and the wild type.

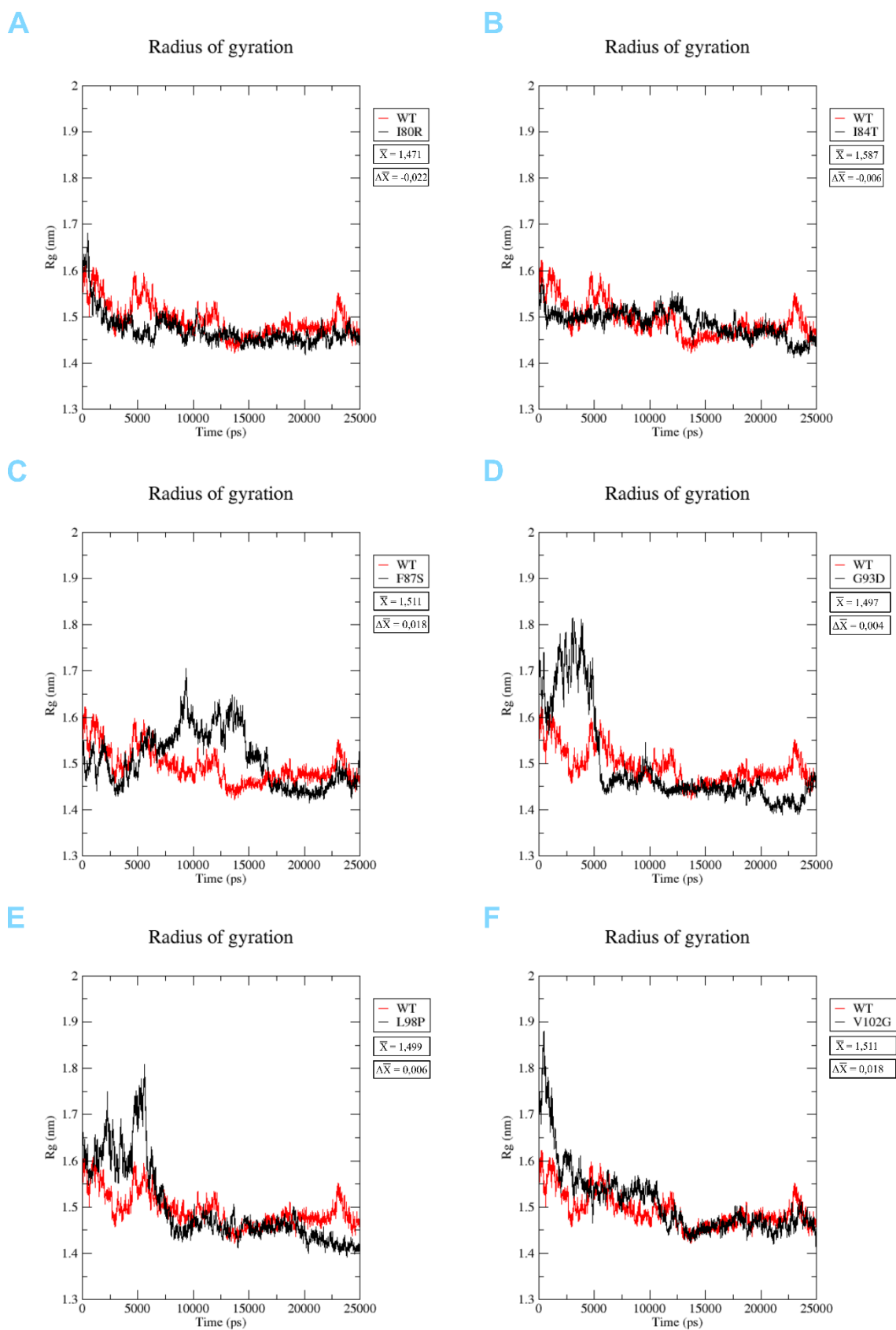


Figure S12: Analysis of radius gyration (Rg) on the backbone of mutants SH2D1A protein. **(A)** mutant I80R; **(B)** mutant I84T; **(C)** mutant F87S; **(D)** mutant G93D; **(E)** mutant L98P; **(F)** mutant V102G. ($\bar{x}$) Average of the Rg values and ($\Delta\bar{x}$) difference between the average of the Rg values between the mutant and the wild type.

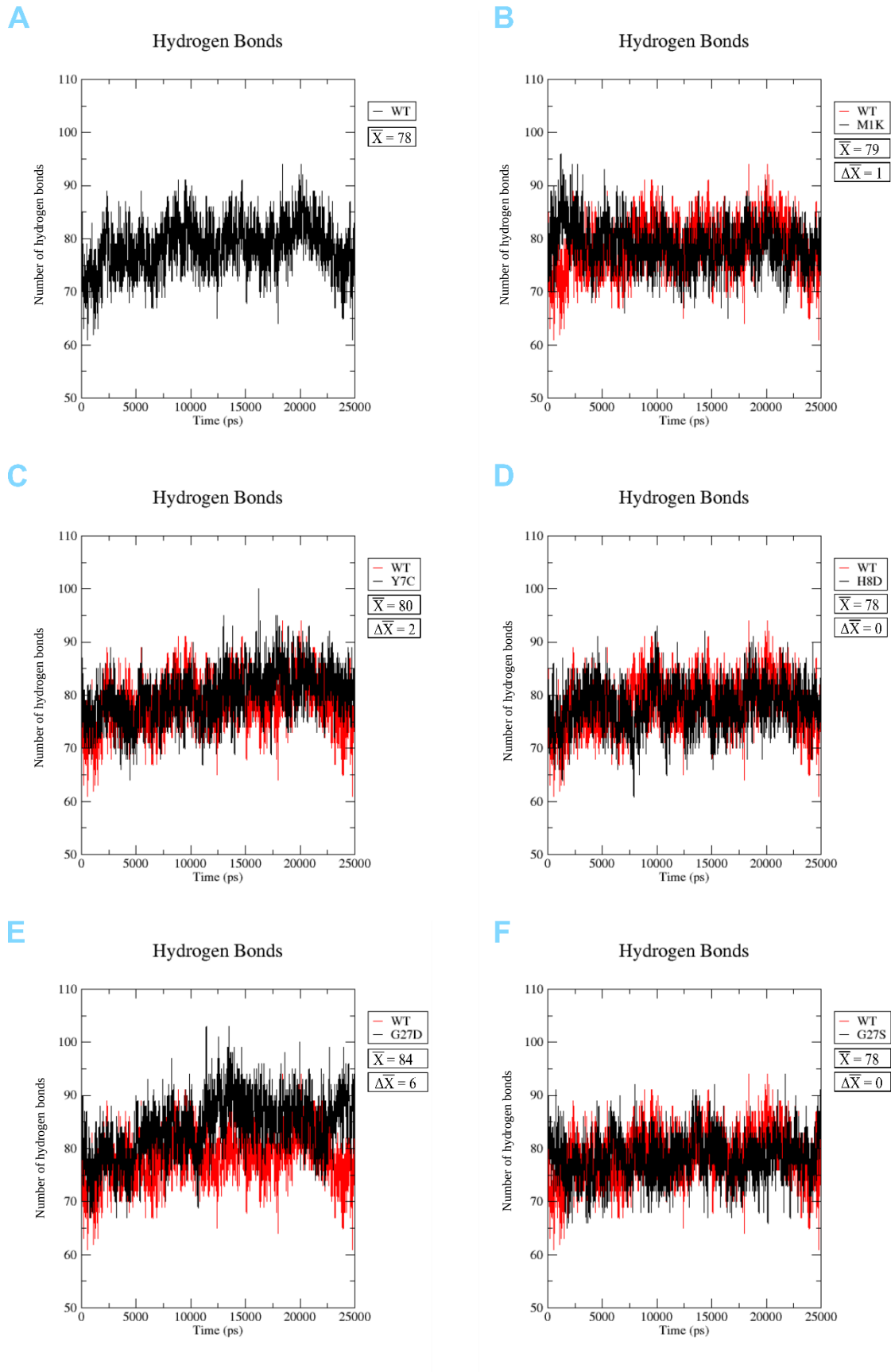


Figure S13: Analysis of intramolecular hydrogen bonds on the native and mutants SH2D1A protein. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{X}$) Average of bonds number and ($\Delta\bar{X}$) difference between the average of bond numbers between the mutant and the wild type.

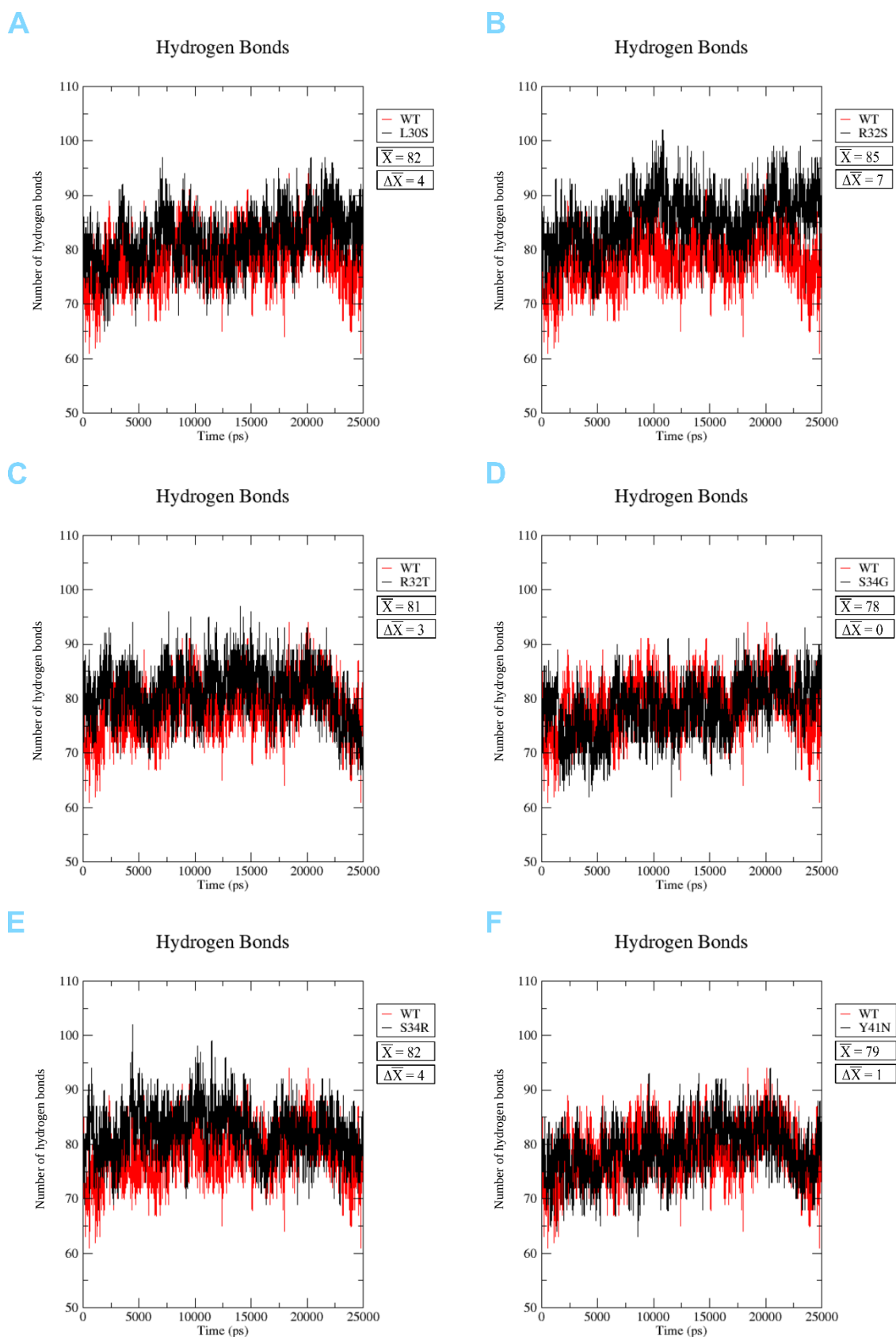


Figure S14: Analysis of intramolecular hydrogen bonds on the mutants SH2D1A protein. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{X}$) Average of bonds number and ($\Delta\bar{X}$) difference between the average of bond numbers between the mutant and the wild type.

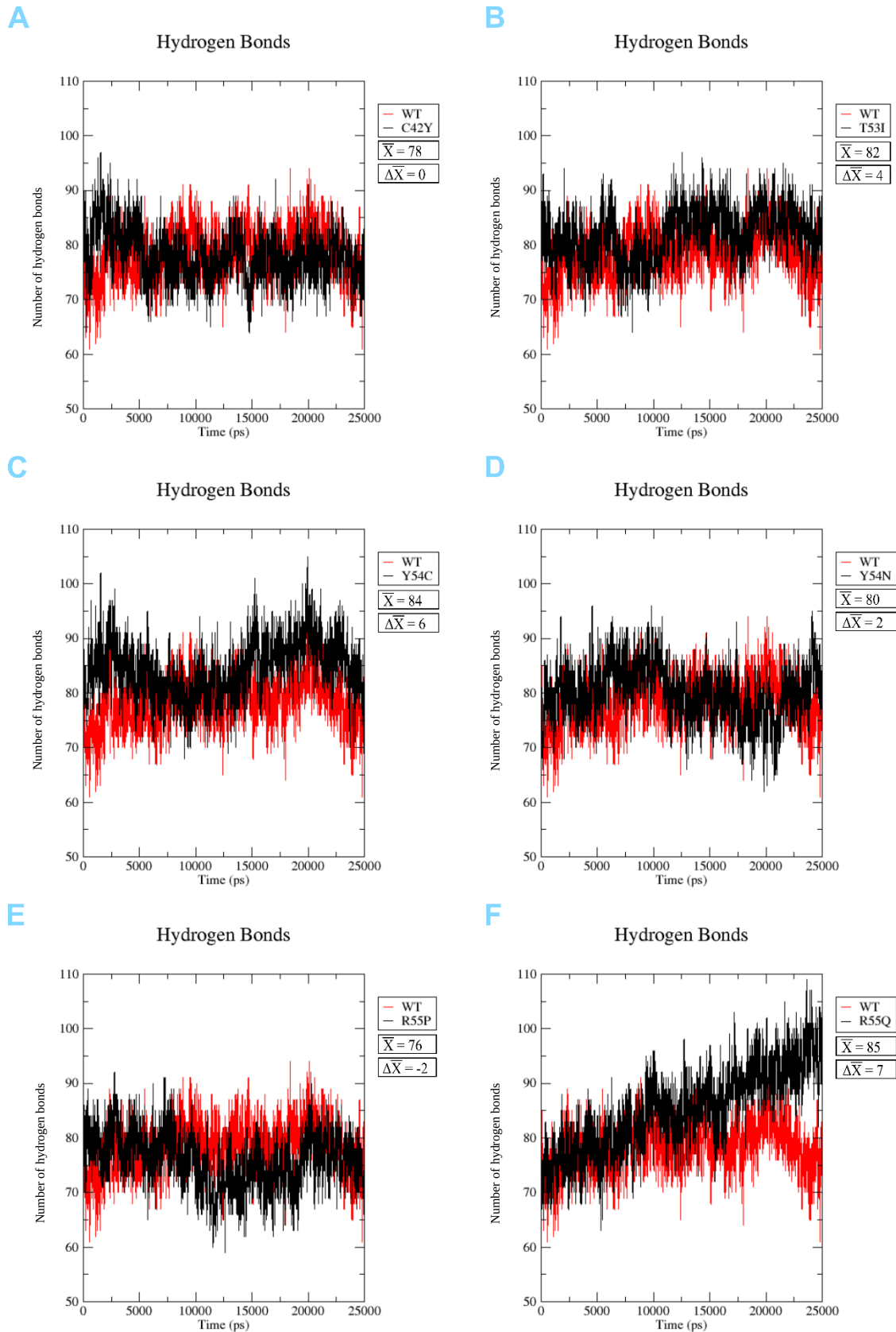


Figure S15: Analysis of intramolecular hydrogen bonds on the mutants SH2D1A protein. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{X}$) Average of bonds number and ($\Delta\bar{X}$) difference between the average of bond numbers between the mutant and the wild type.

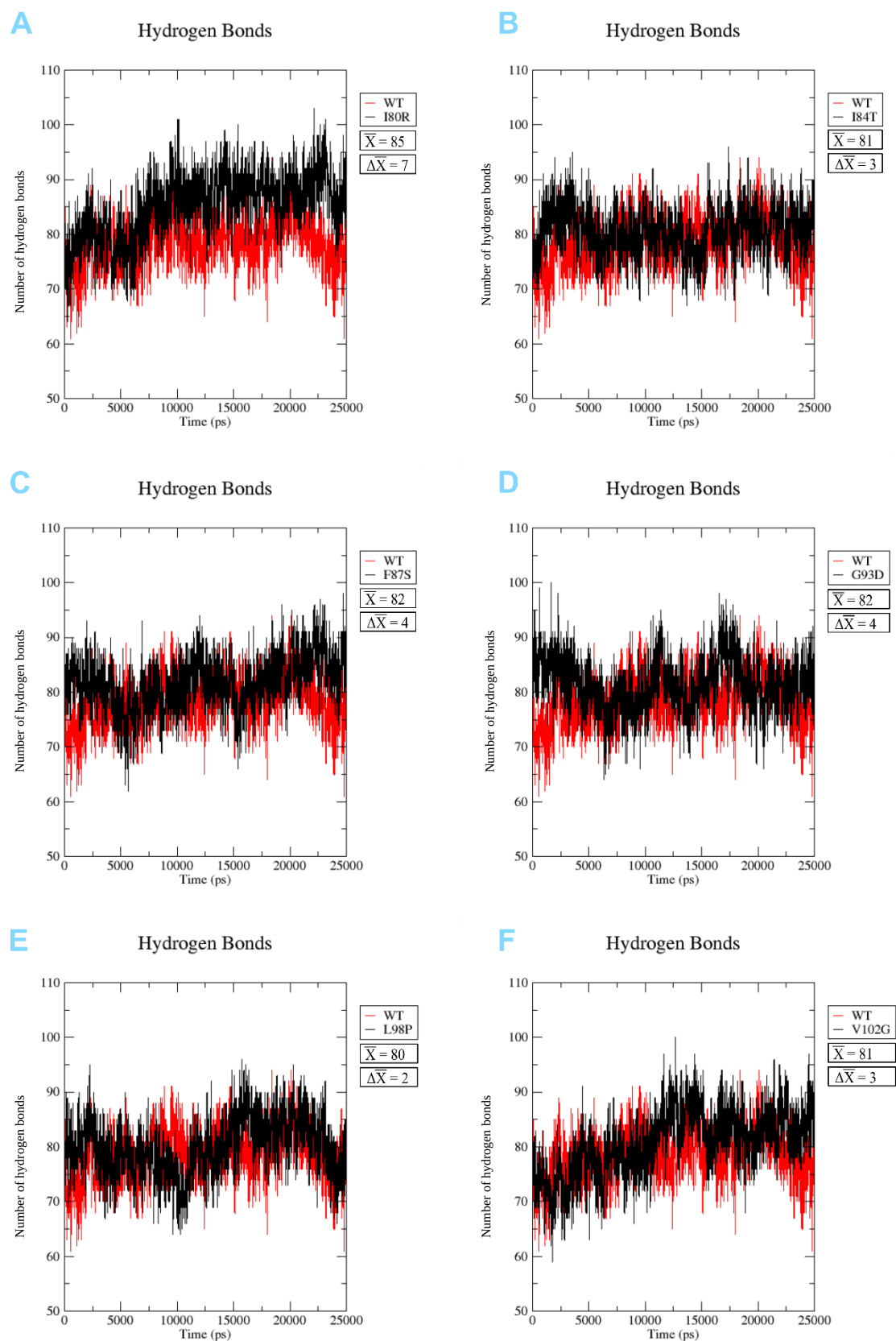


Figure S16: Analysis of intramolecular hydrogen bonds on the mutants SH2D1A protein. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G. ($\bar{X}$) Average of bonds number and ($\Delta\bar{X}$) difference between the average of bond numbers between the mutant and the wild type.

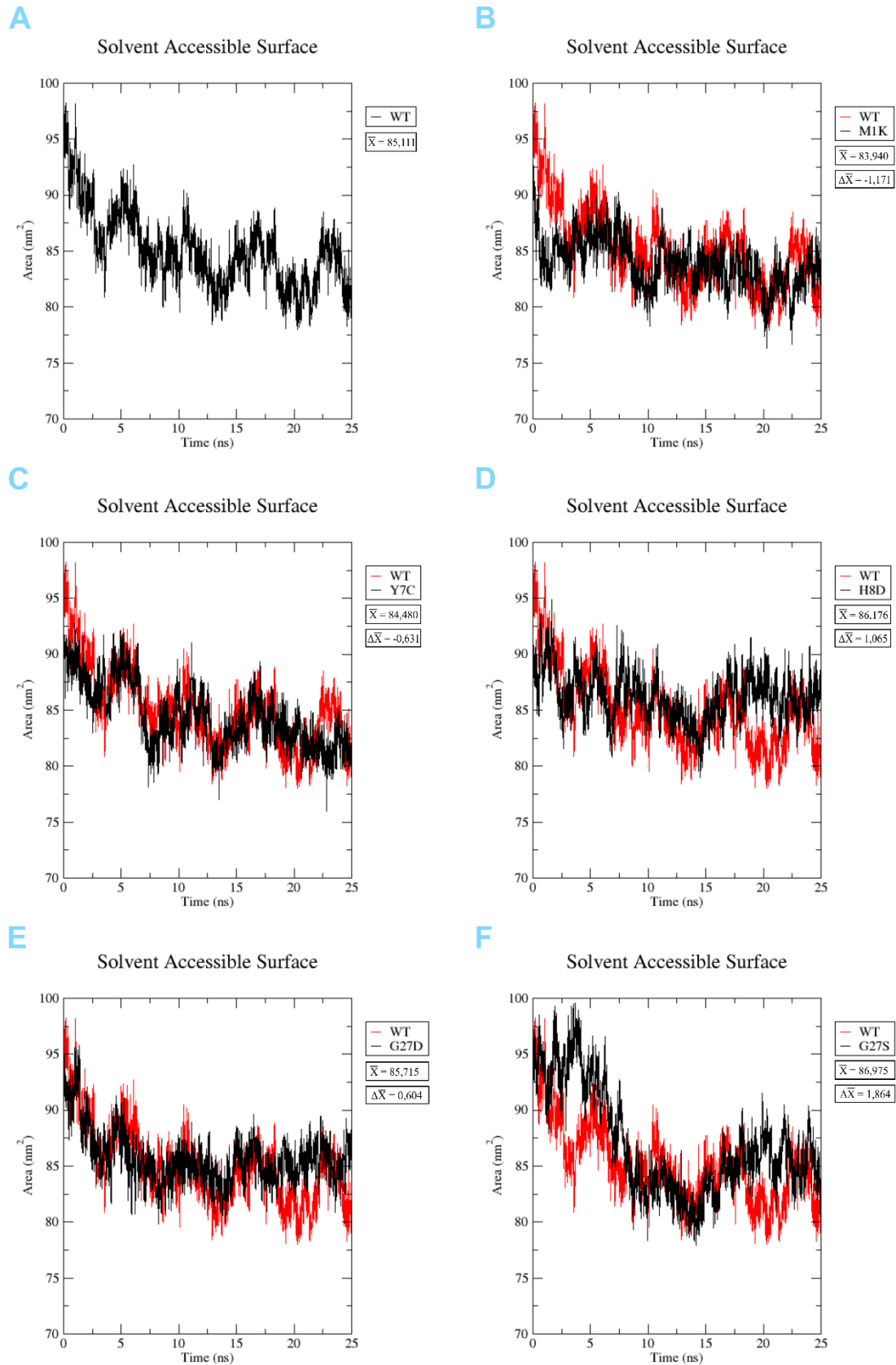


Figure S17: Analysis of solvent accessible surface area (SASA) on the native and mutants SH2D1A protein. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S. ($\bar{X}$) Average of SASA value and ($\Delta\bar{X}$) difference between the average of SASA value between the mutant and the wild type.

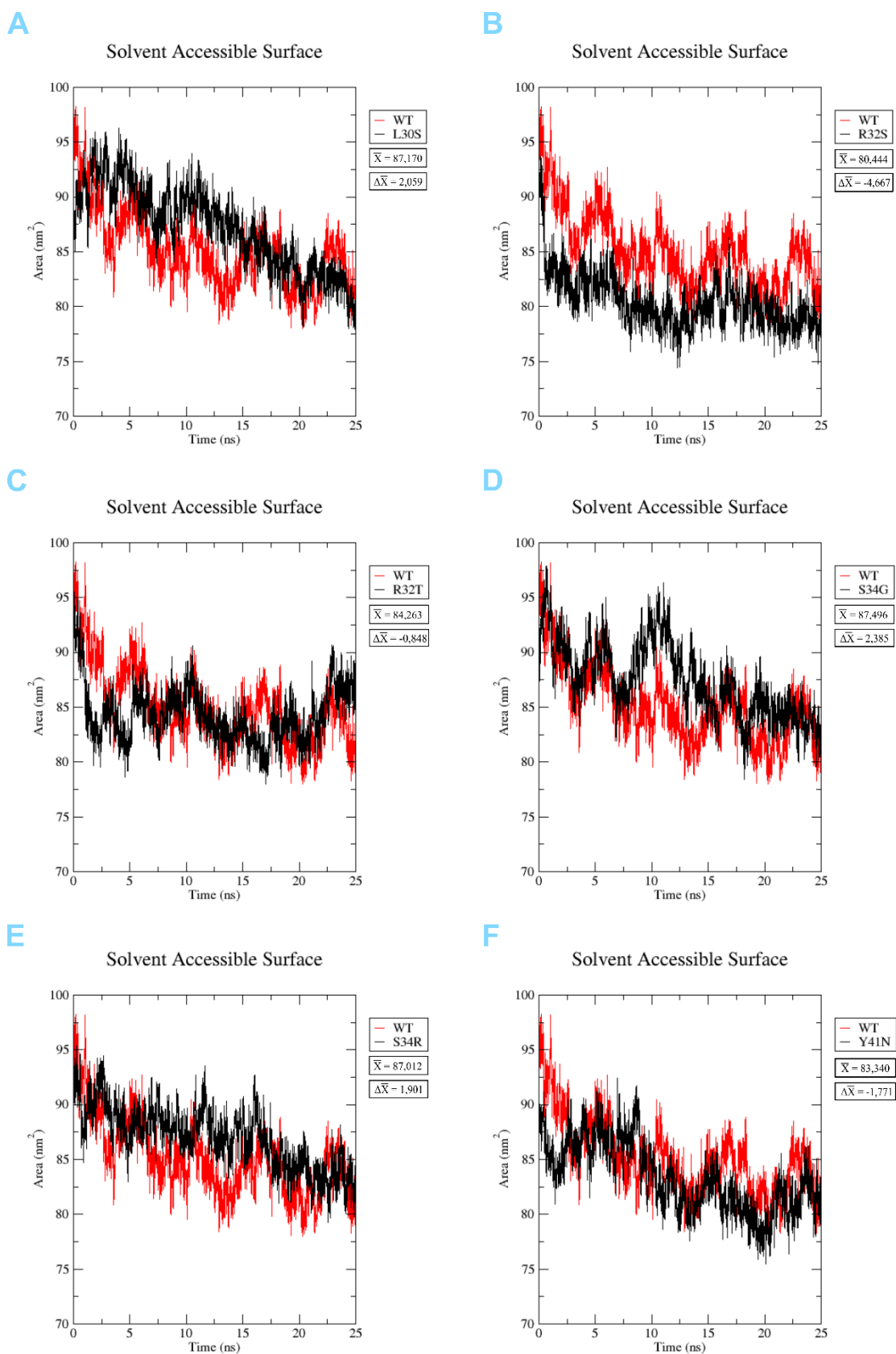


Figure S18: Analysis of solvent accessible surface area (SASA) on the mutants SH2D1A protein. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N. ($\bar{X}$) Average of SASA value and ($\Delta\bar{X}$) difference between the average of SASA value between the mutant and the wild type.

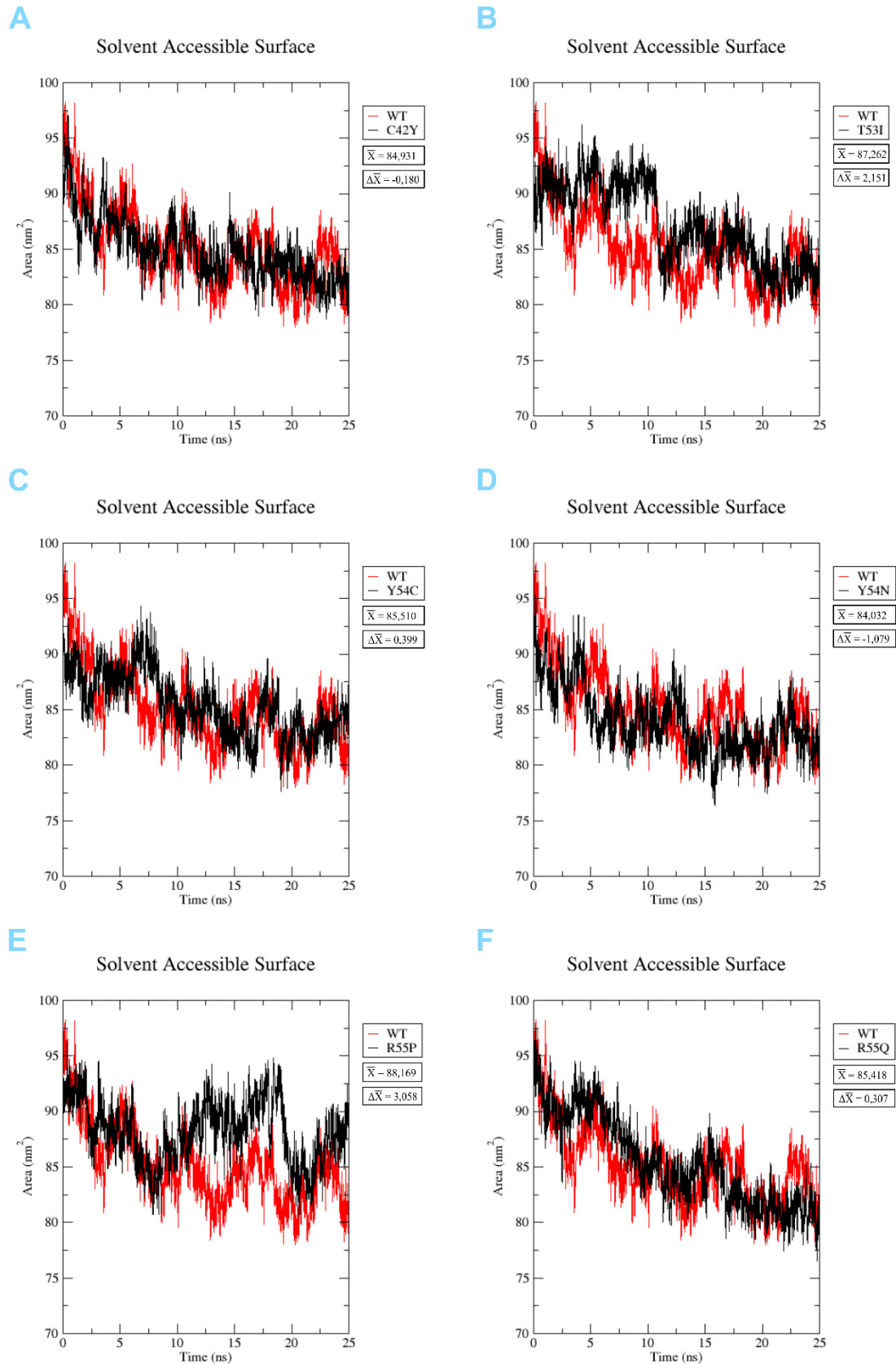


Figure S19: Analysis of solvent accessible surface area (SASA) on the mutants SH2D1A protein. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q. ($\bar{x}$) Average of SASA value and ($\Delta\bar{x}$) difference between the average of SASA value between the mutant and the wild type.

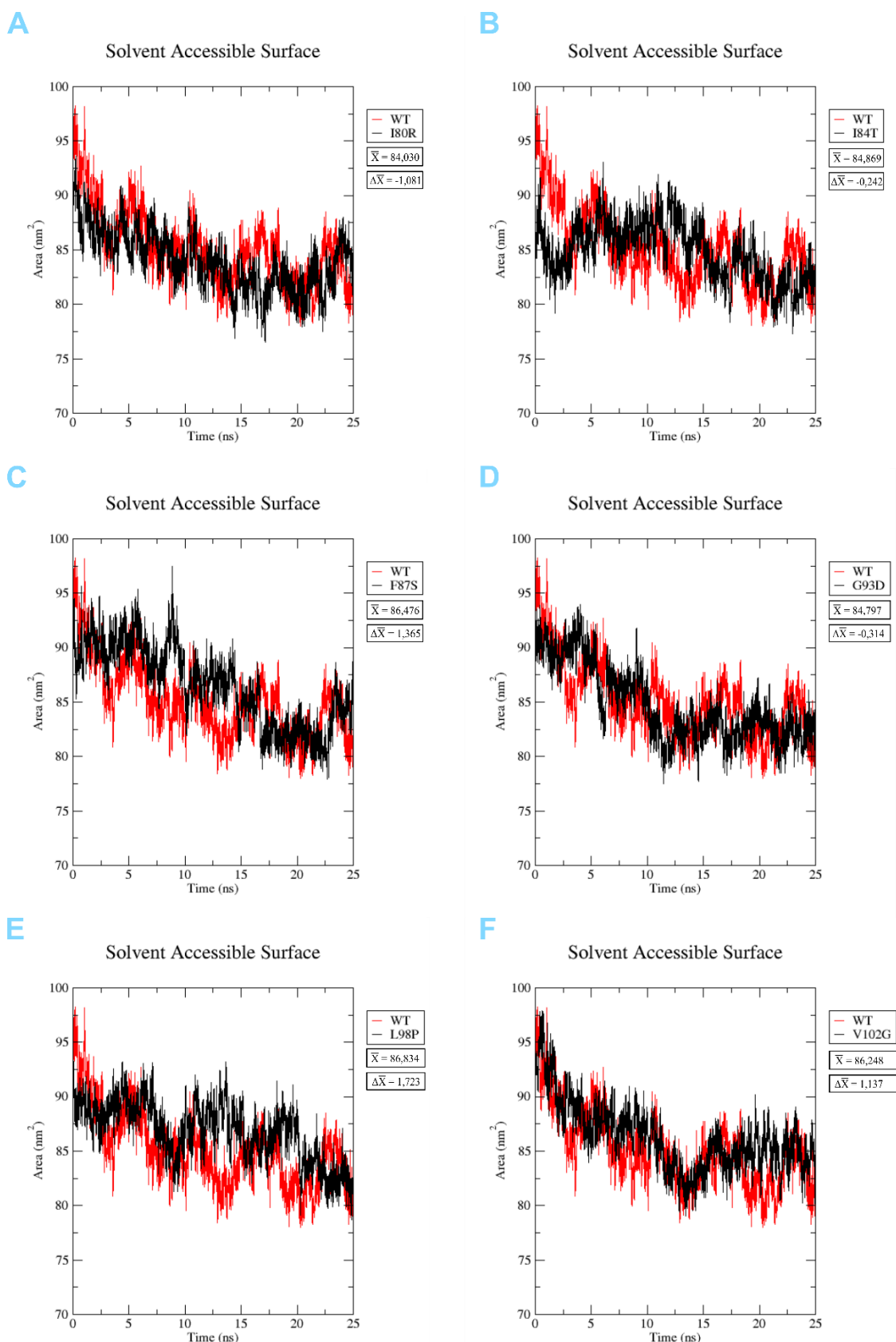


Figure S20: Analysis of solvent accessible surface area (SASA) on the mutants SH2D1A protein. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G. ($\bar{X}$) Average of SASA value and ($\Delta\bar{X}$) difference between the average of SASA value between the mutant and the wild type.

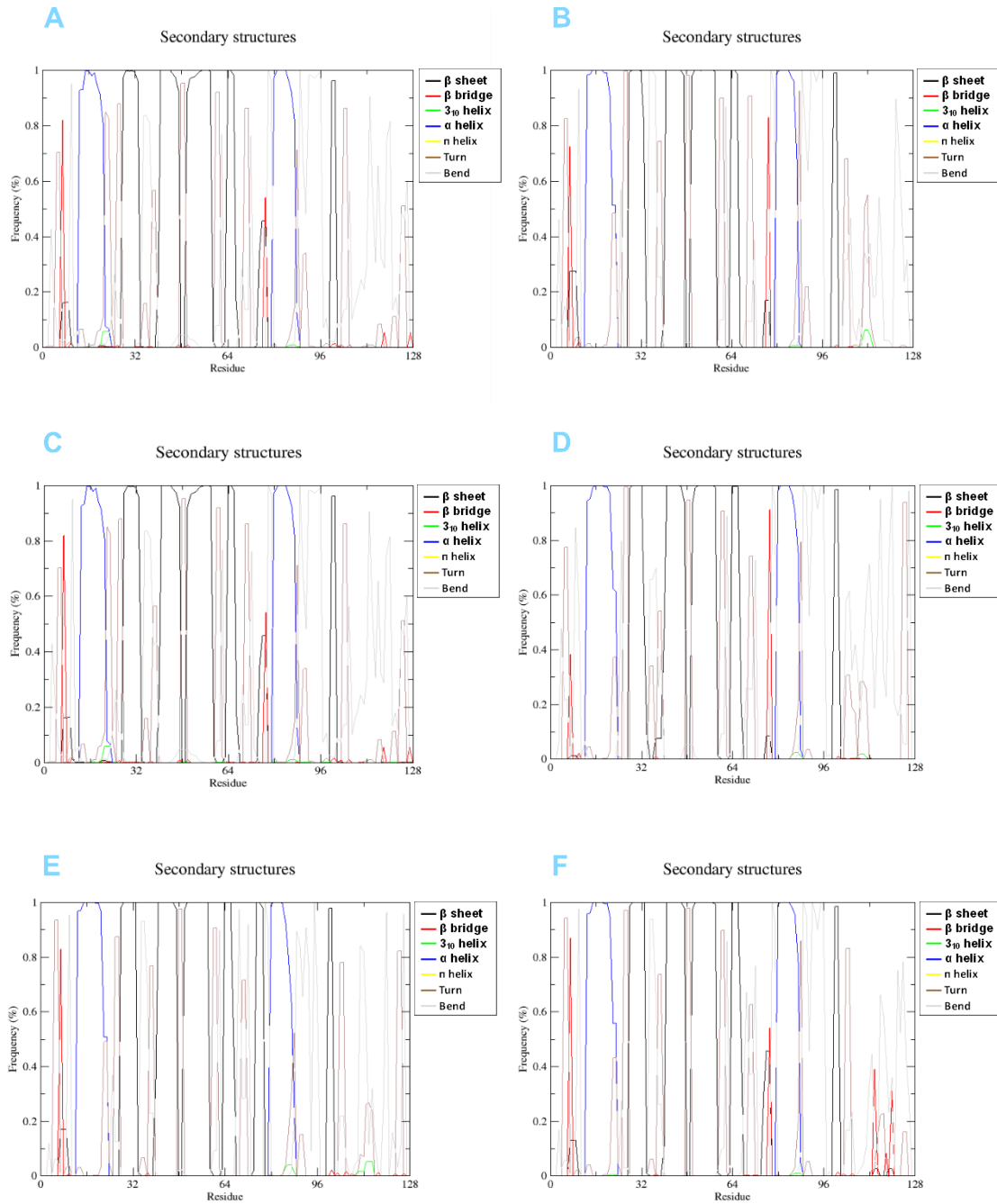


Figure S21: Frequency of secondary structures of wild-type and mutants SH2D1A proteins during simulation, by residue. (A) native protein (WT); (B) mutant M1K; (C) mutant Y7C; (D) mutant H8D; (E) mutant G27D; (F) mutant G27S.

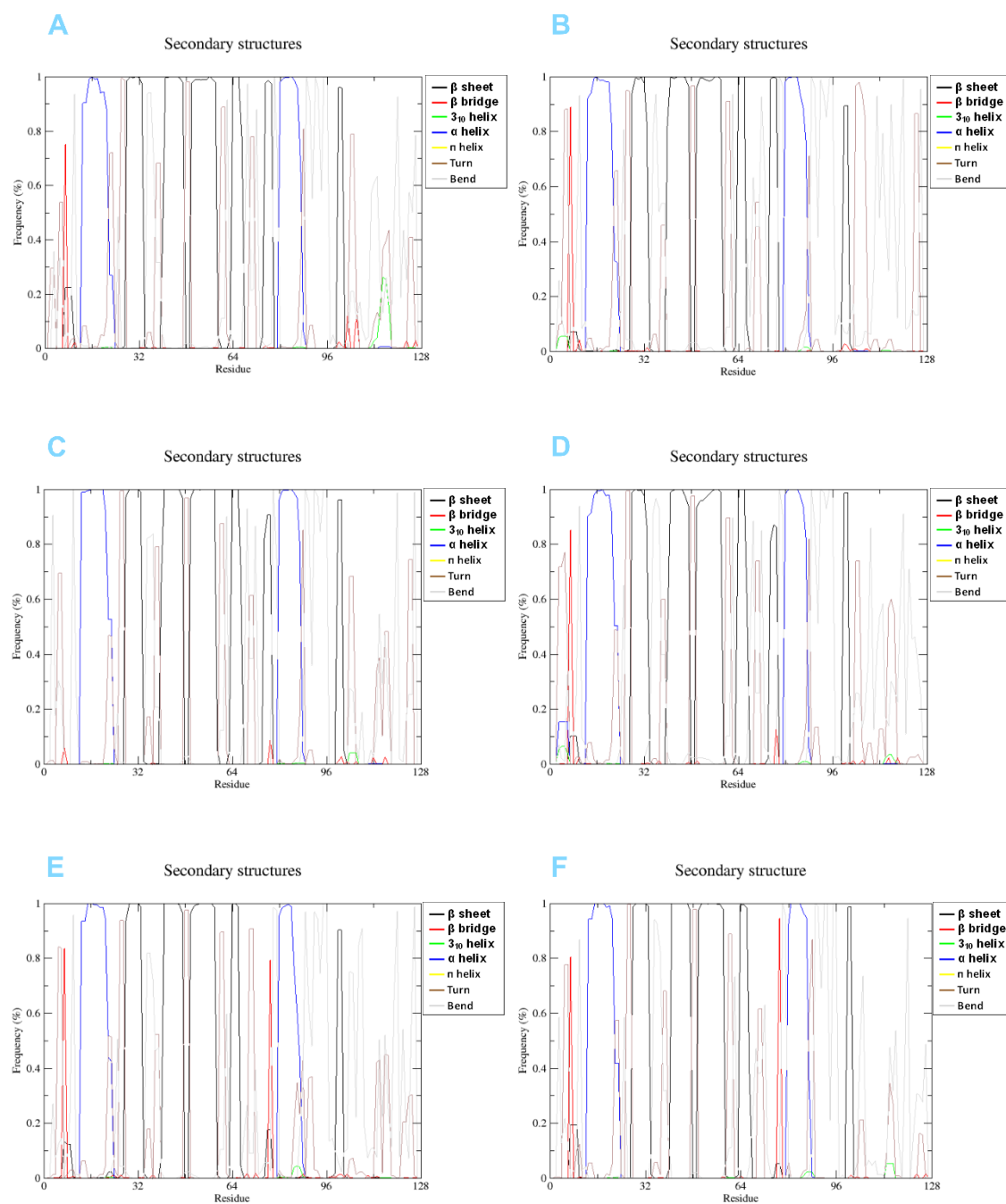


Figure S22: Frequency of secondary structures of wild-type and mutants SH2D1A proteins during simulation, by residue. (A) mutant L30S; (B) mutant R32S; (C) mutant R32T; (D) mutant S34G; (E) mutant S34R; (F) mutant Y41N.

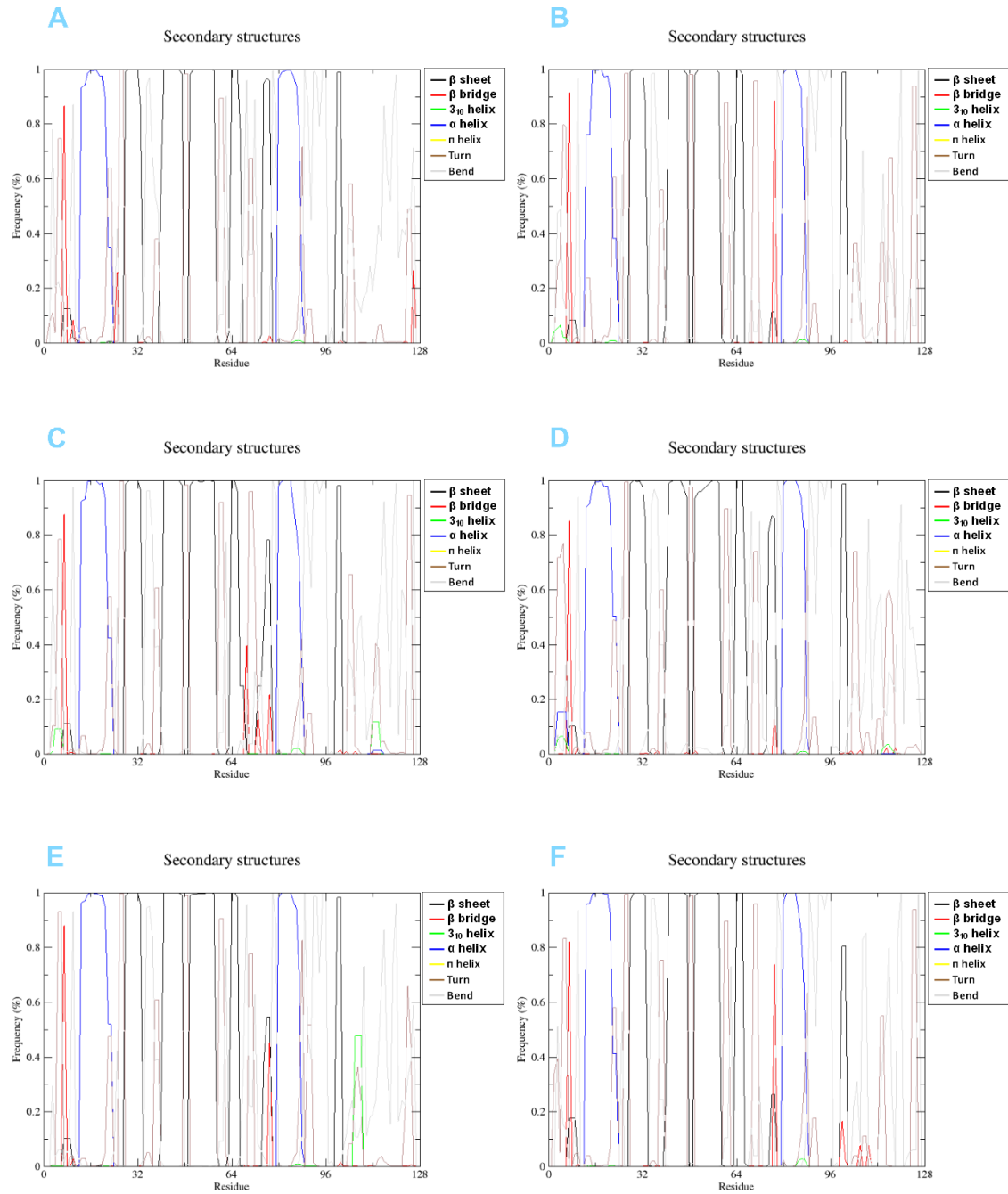


Figure S23: Frequency of secondary structures of wild-type and mutants SH2D1A proteins during simulation, by residue. (A) mutant C42Y; (B) mutant T53I; (C) mutant Y54C; (D) mutant Y54N; (E) mutant R55P; (F) mutant R55Q.

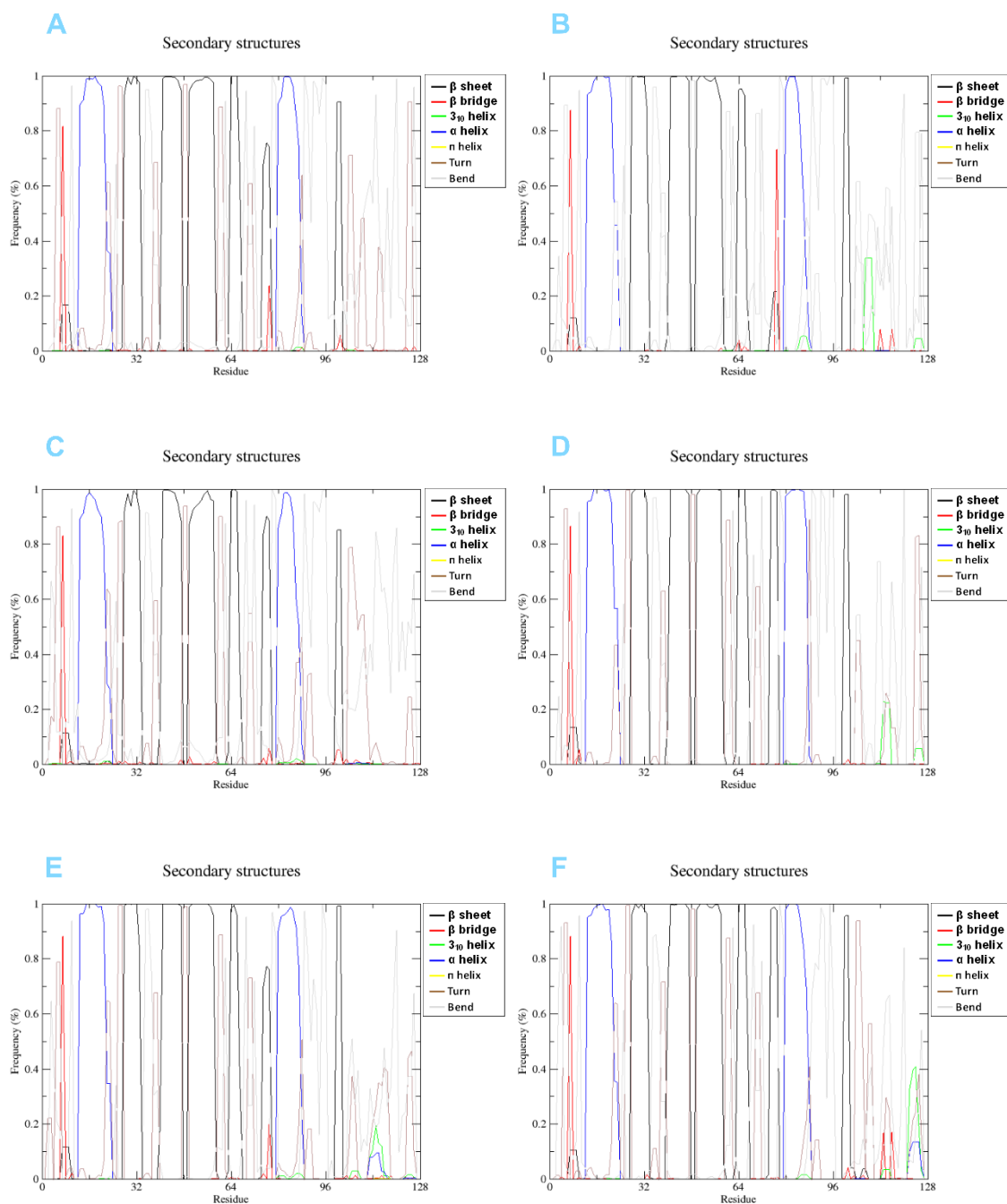


Figure S24: Frequency of secondary structures of wild-type and mutants SH2D1A proteins during simulation, by residue. (A) mutant I80R; (B) mutant I84T; (C) mutant F87S; (D) mutant G93D; (E) mutant L98P; (F) mutant V102G.

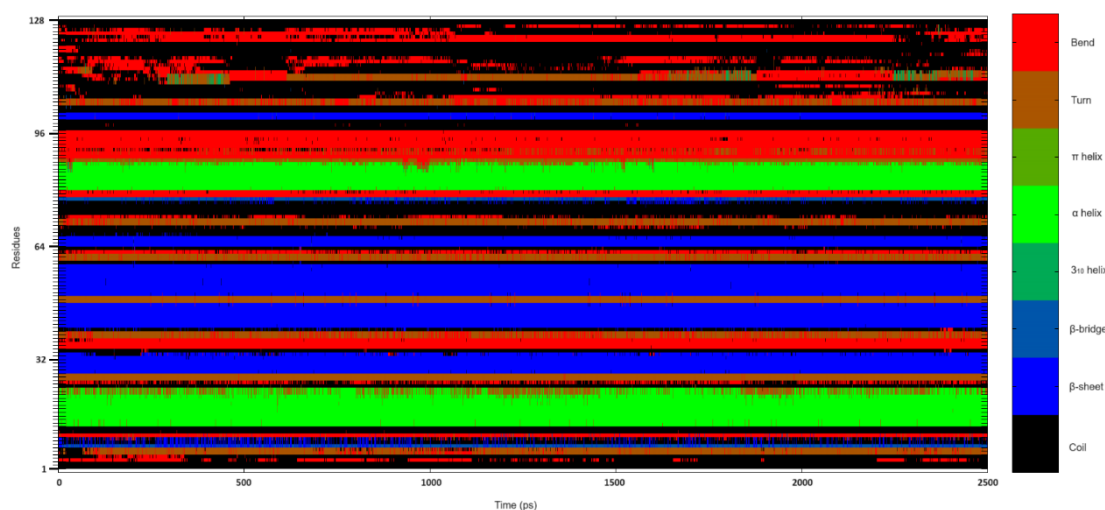


Figure S25: Secondary structures of the native protein (WT) per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

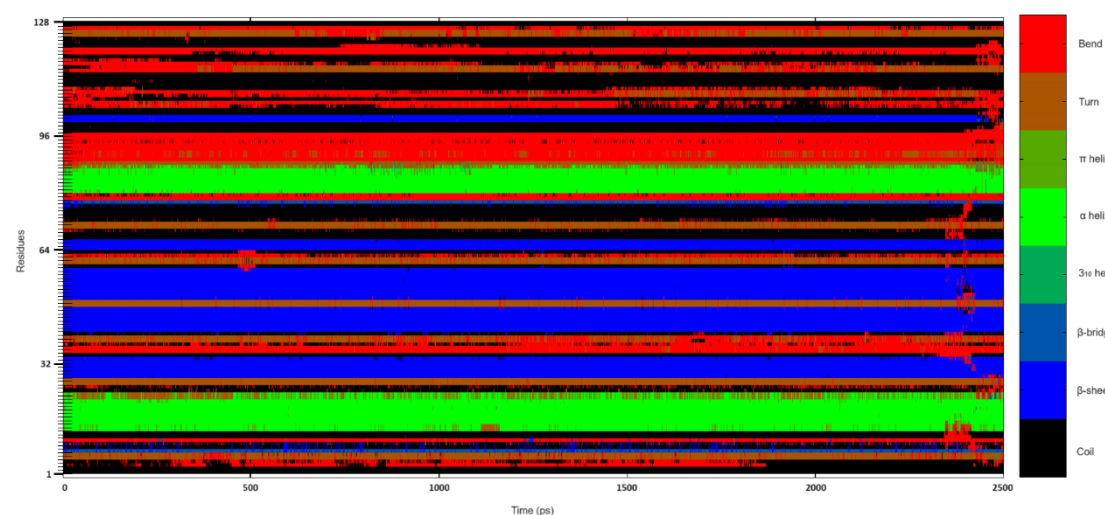


Figure S26: Secondary structures of the M1K mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

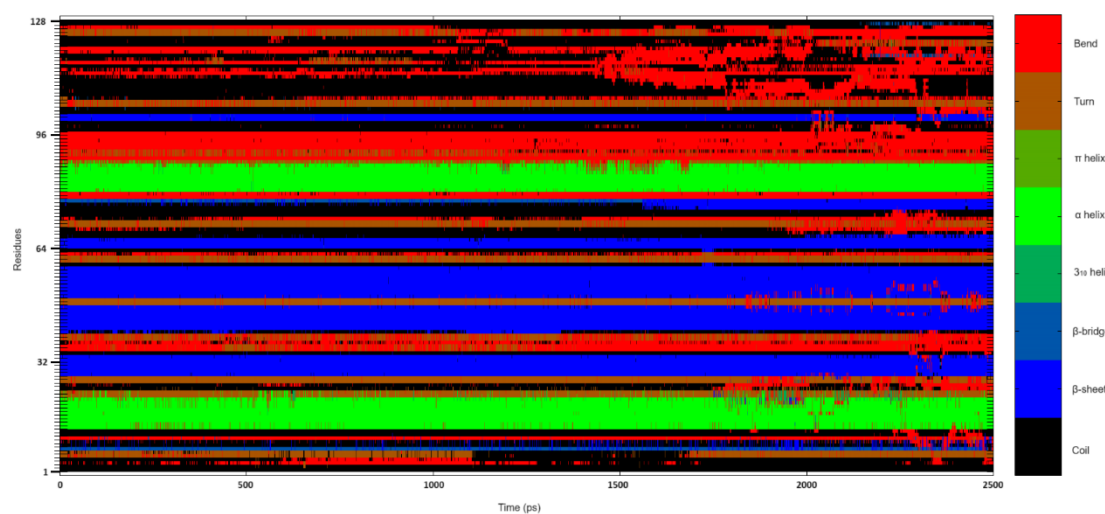


Figure S27: Secondary structures of the Y7C mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

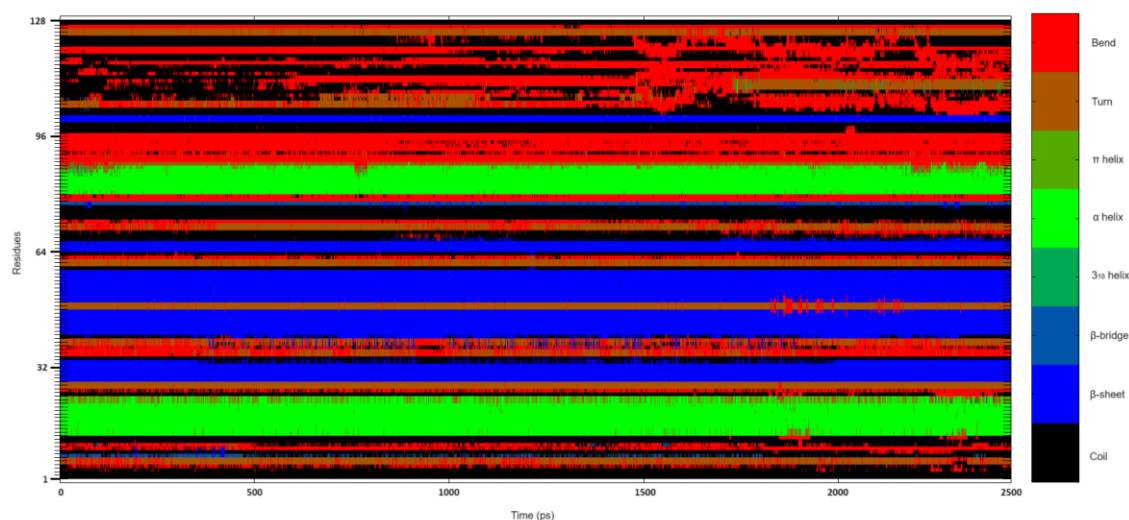


Figure S28: Secondary structures of the H8D mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

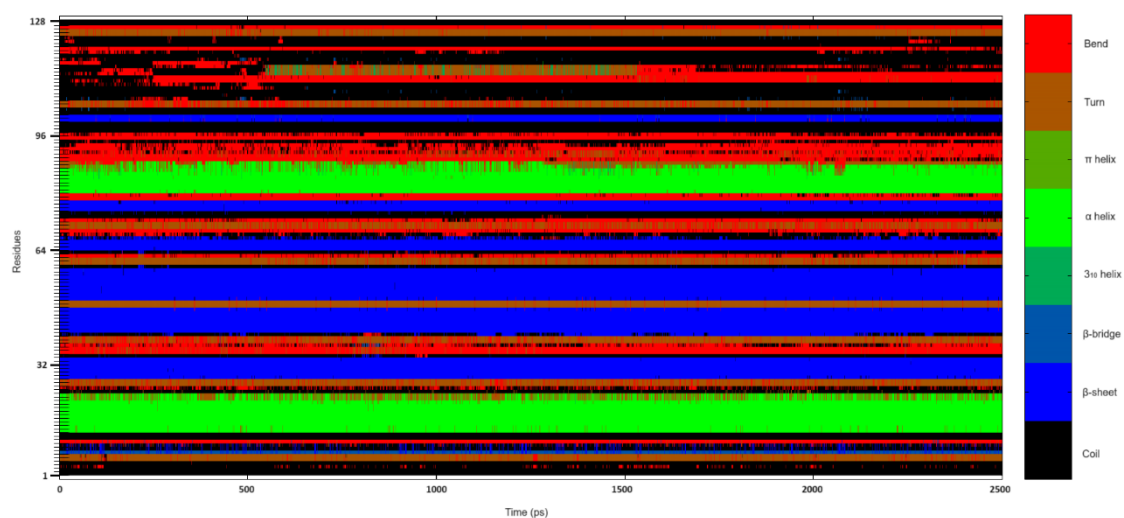


Figure S29: Secondary structures of the G27D mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

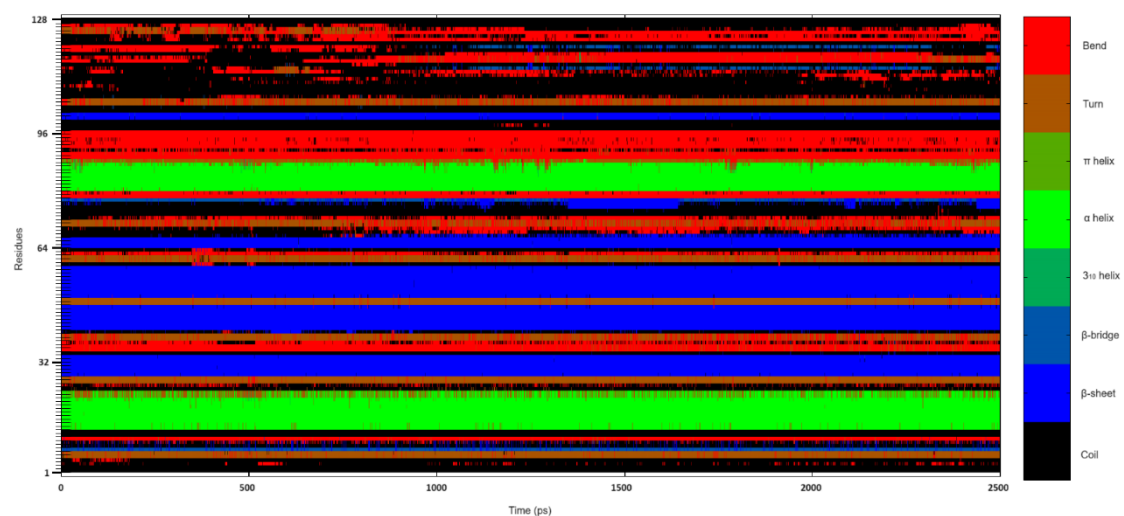


Figure S30: Secondary structures of the G27S mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

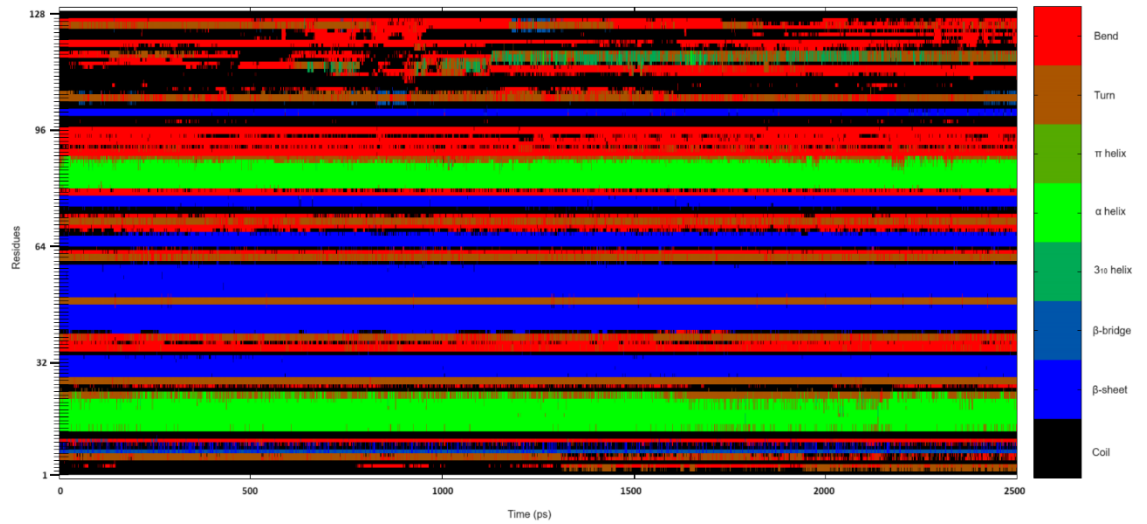


Figure S31: Secondary structures of the L30S mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

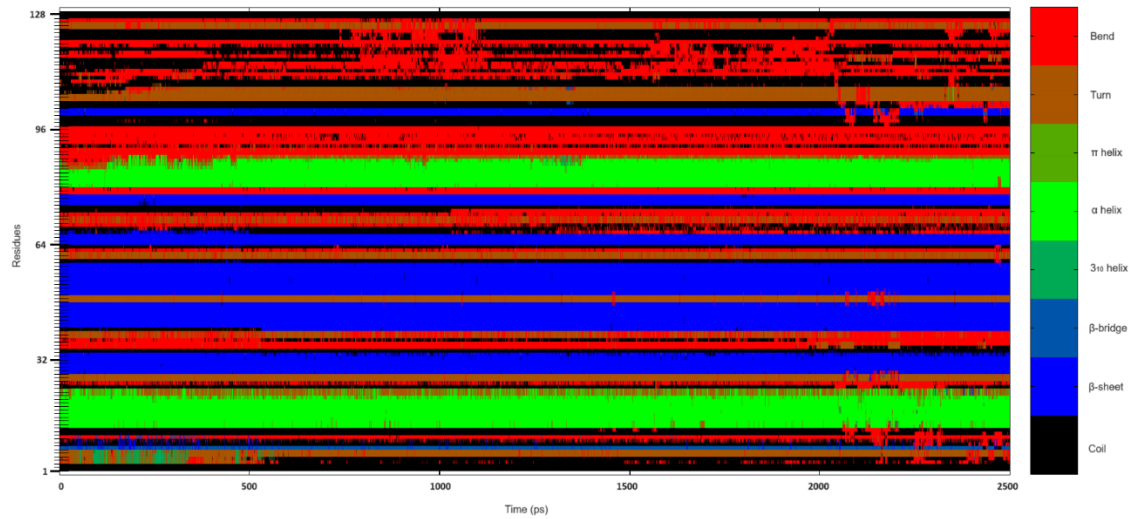


Figure S32: Secondary structures of the R32S mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

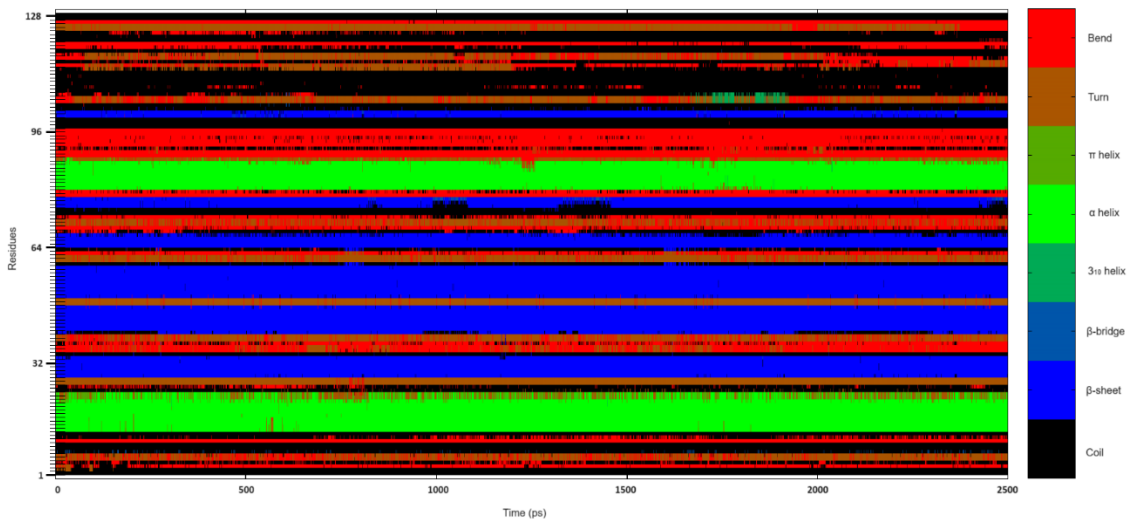


Figure S33: Secondary structures of the R32T mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

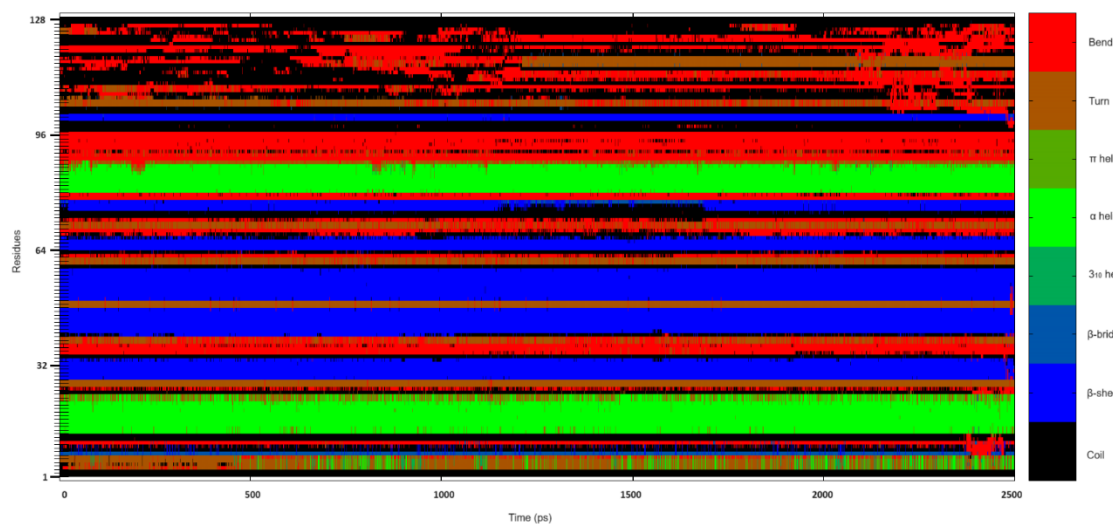


Figure S34: Secondary structures of the S34G mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

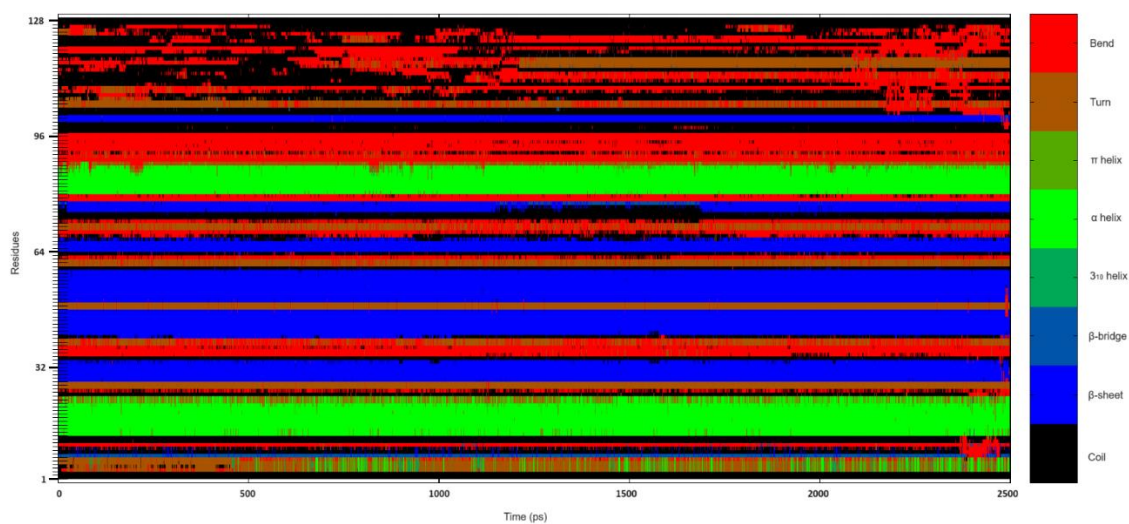


Figure S35: Secondary structures of the S34R mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

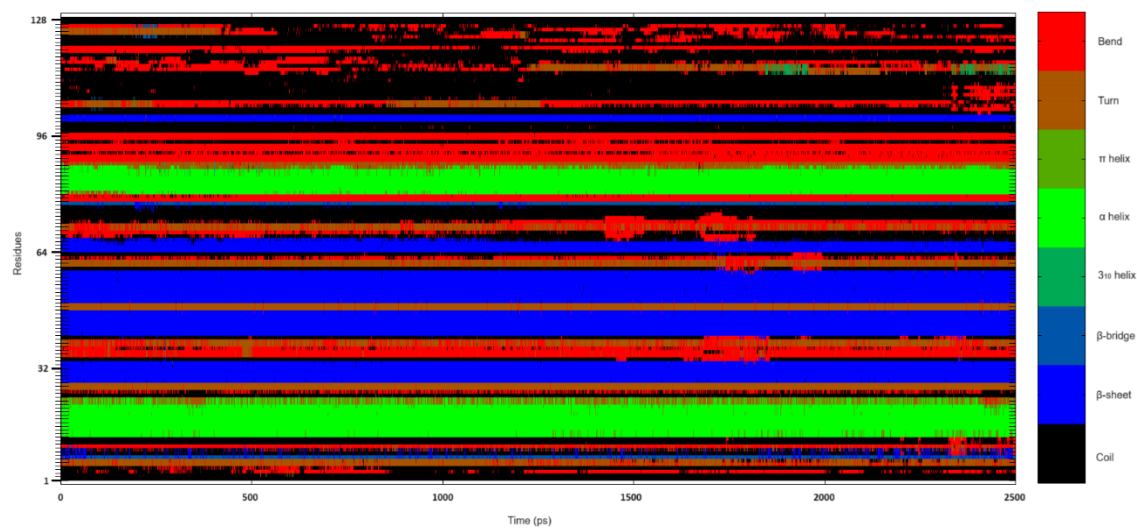


Figure S36: Secondary structures of the Y41N mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

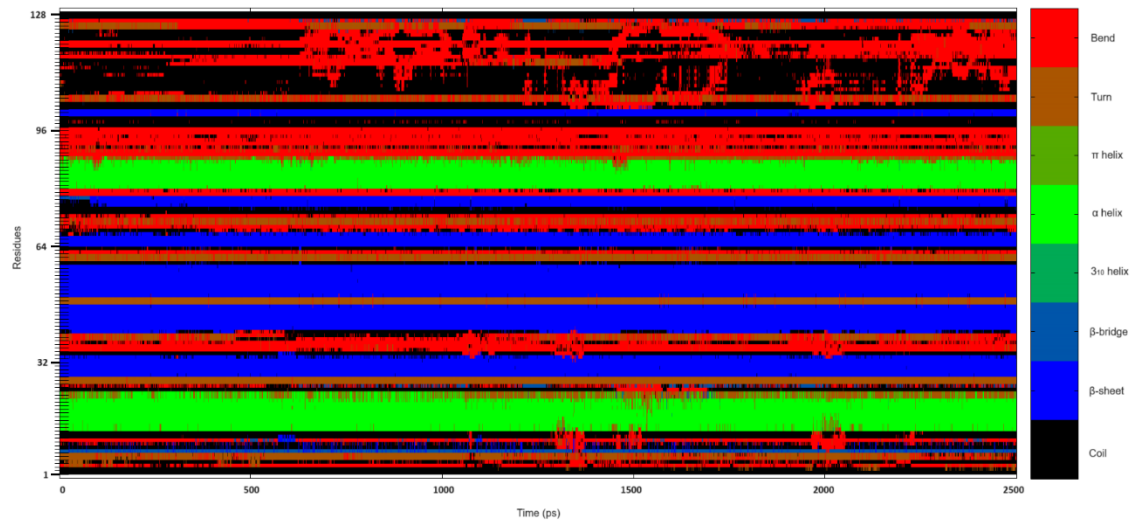


Figure S37: Secondary structures of the C42Y mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

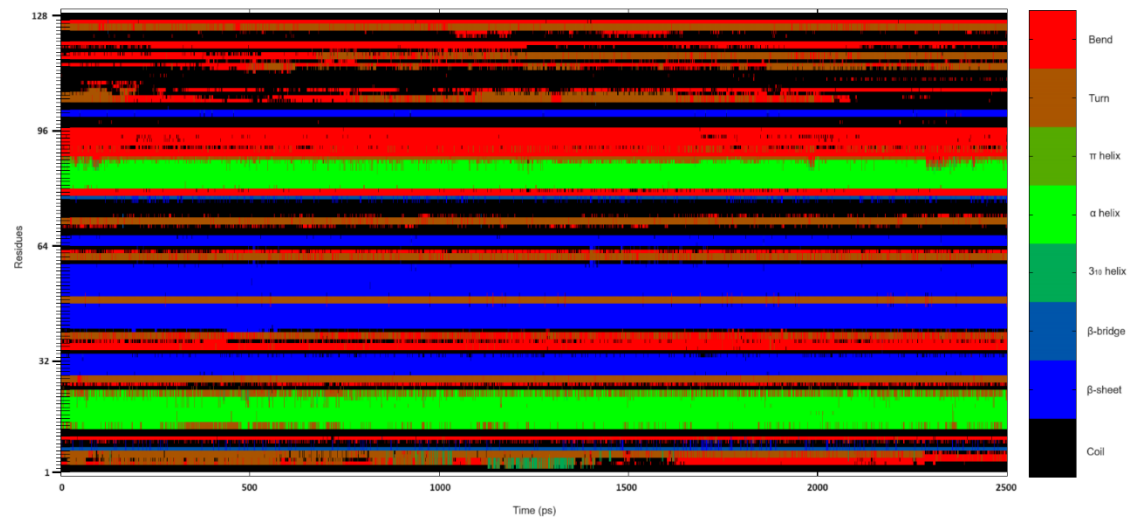


Figure S38: Secondary structures of the T53I mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

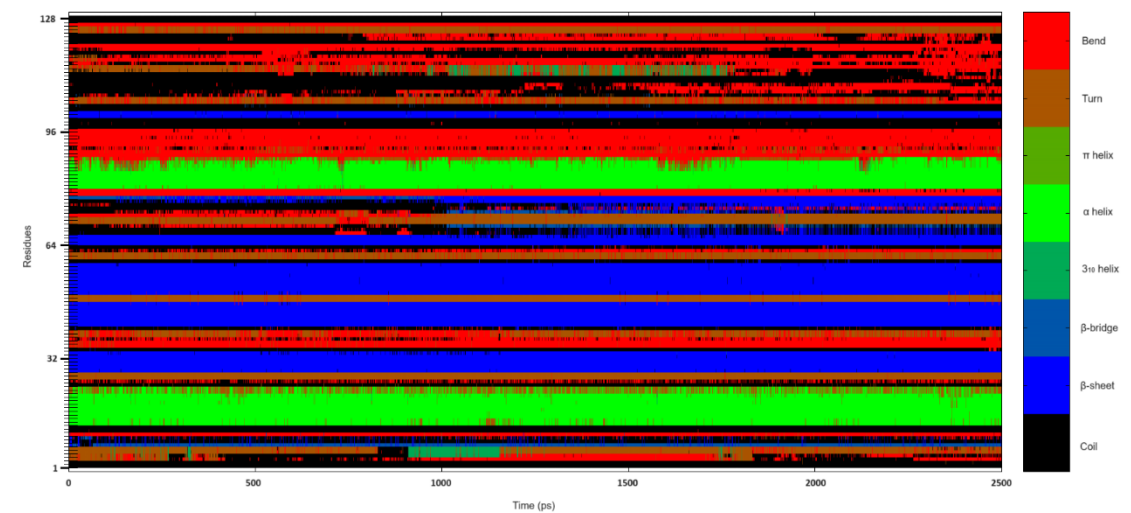


Figure S39: Secondary structures of the Y54N mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

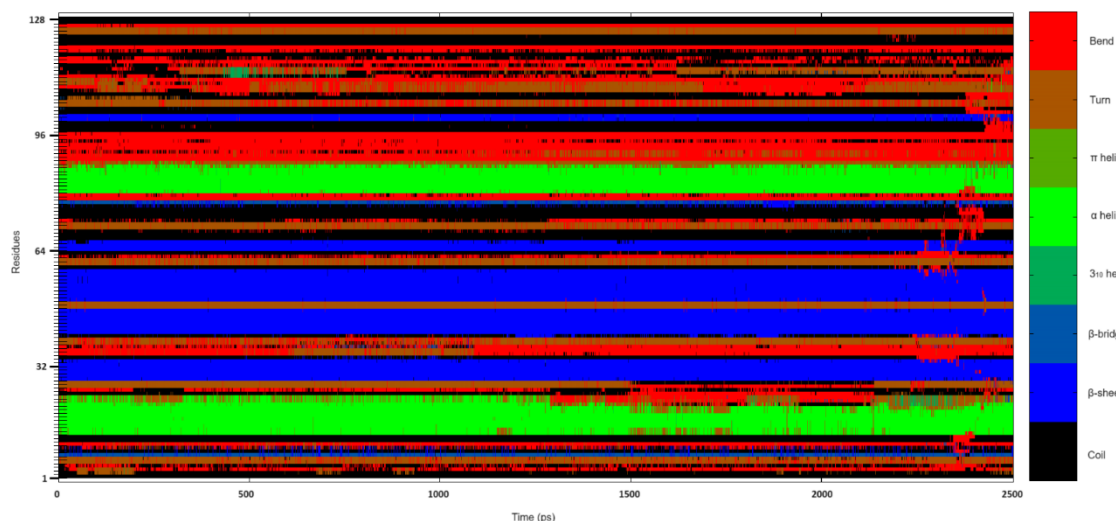


Figure S40: Secondary structures of the Y54C mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

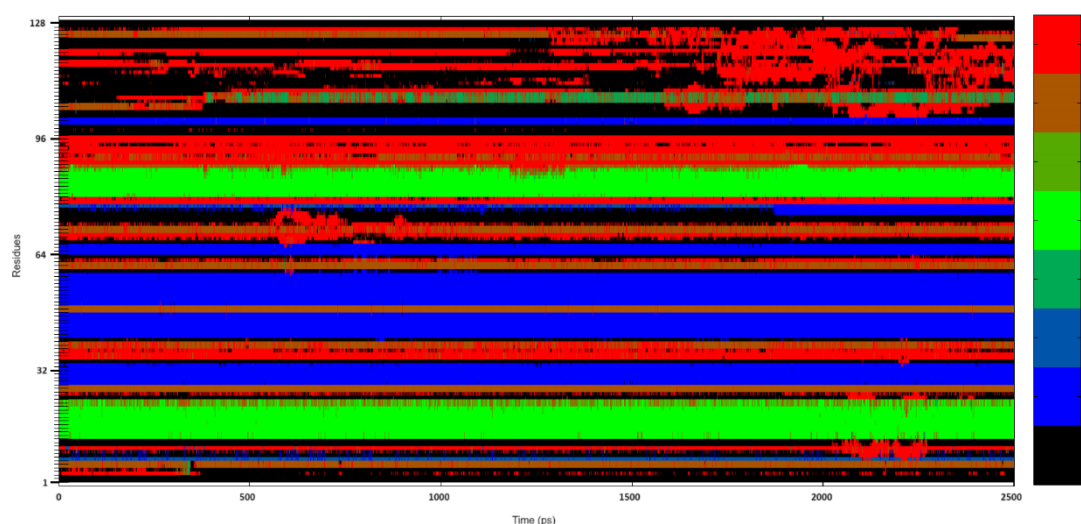


Figure S41: Secondary structures of the R55P mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

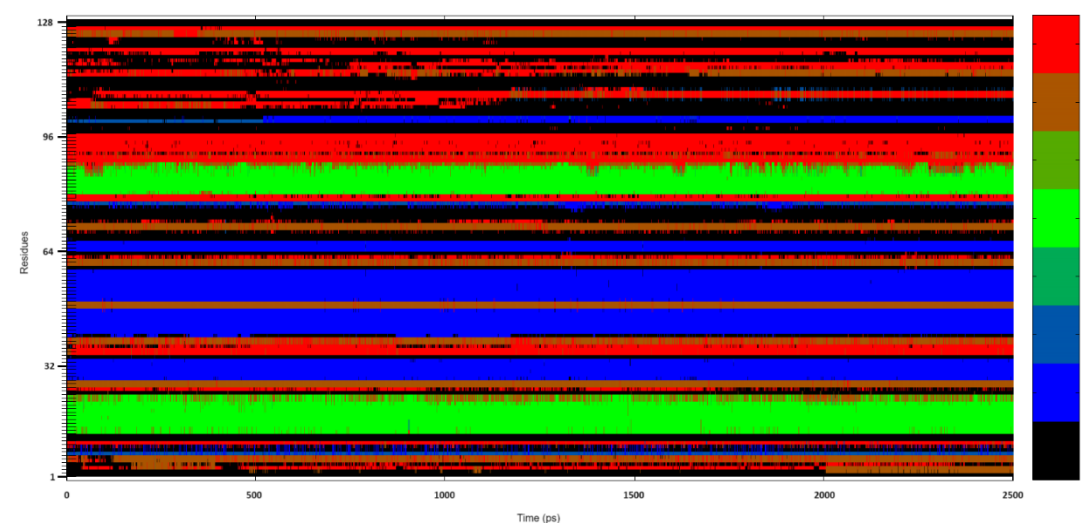


Figure S42: Secondary structures of the R55Q mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

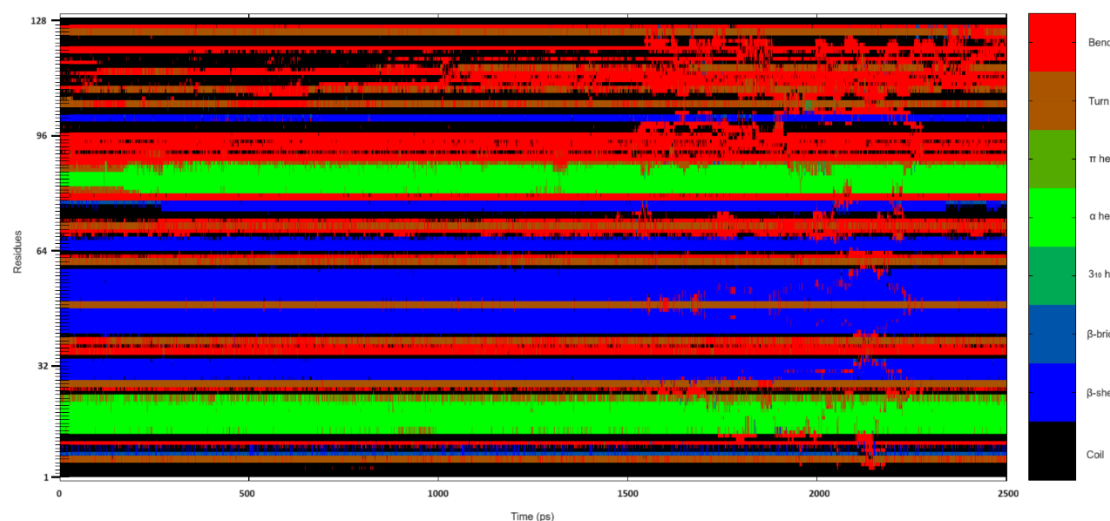


Figure S43: Secondary structures of the I80R mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

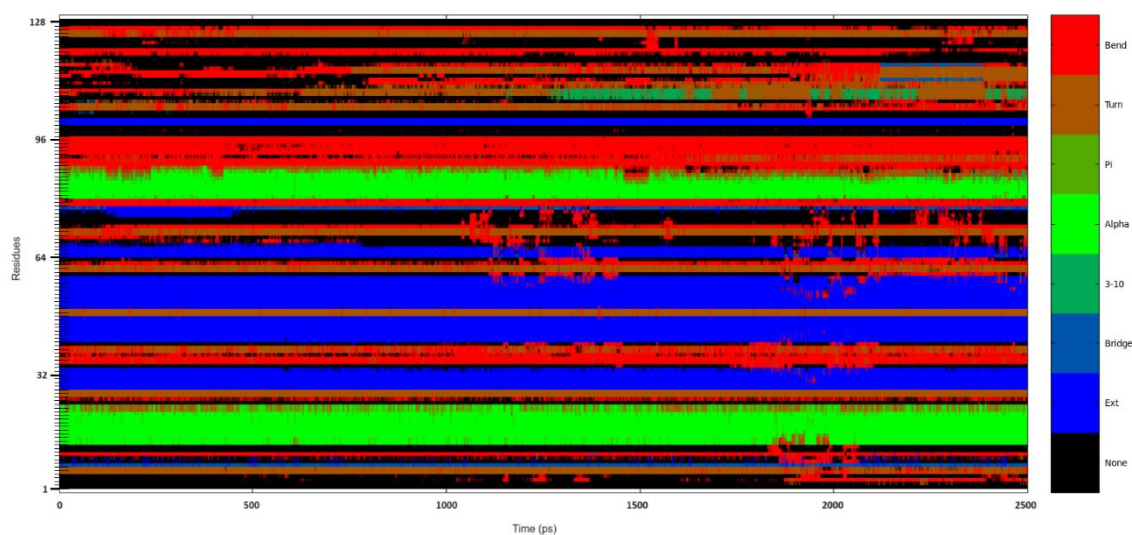


Figure S44: Secondary structures of the I84T mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

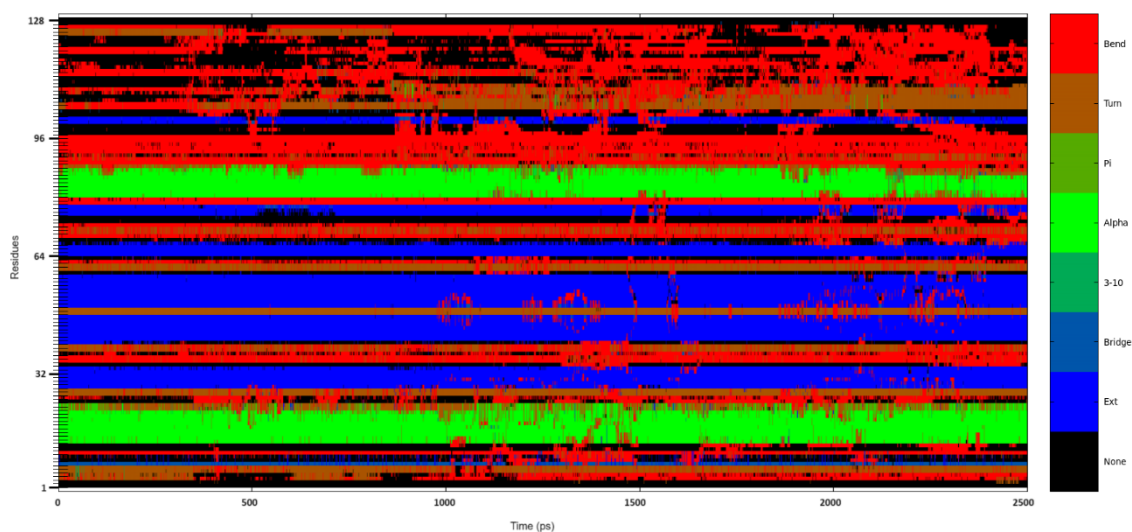


Figure S45: Secondary structures of the F87S mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

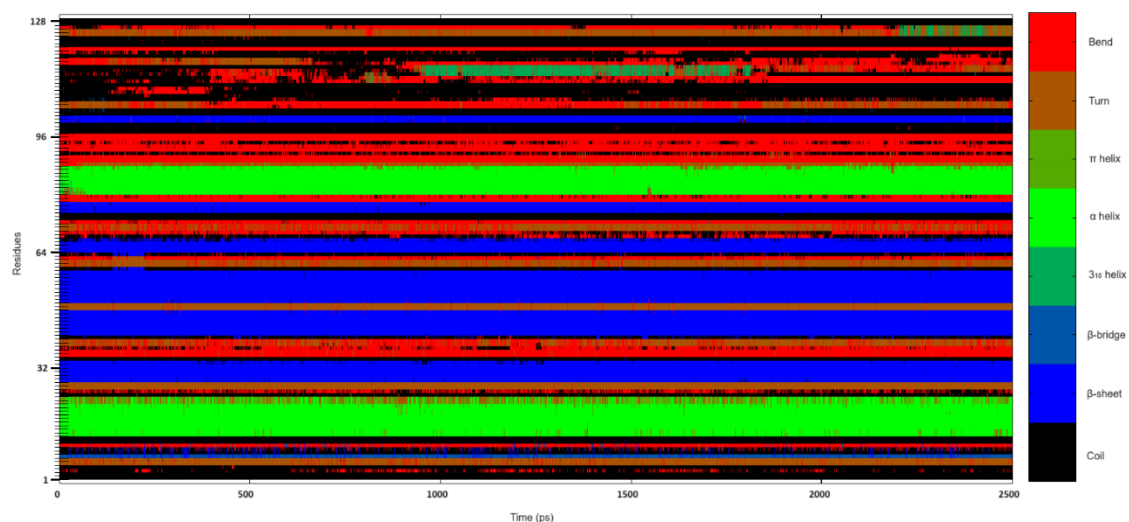


Figure S46: Secondary structures of the G93D mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

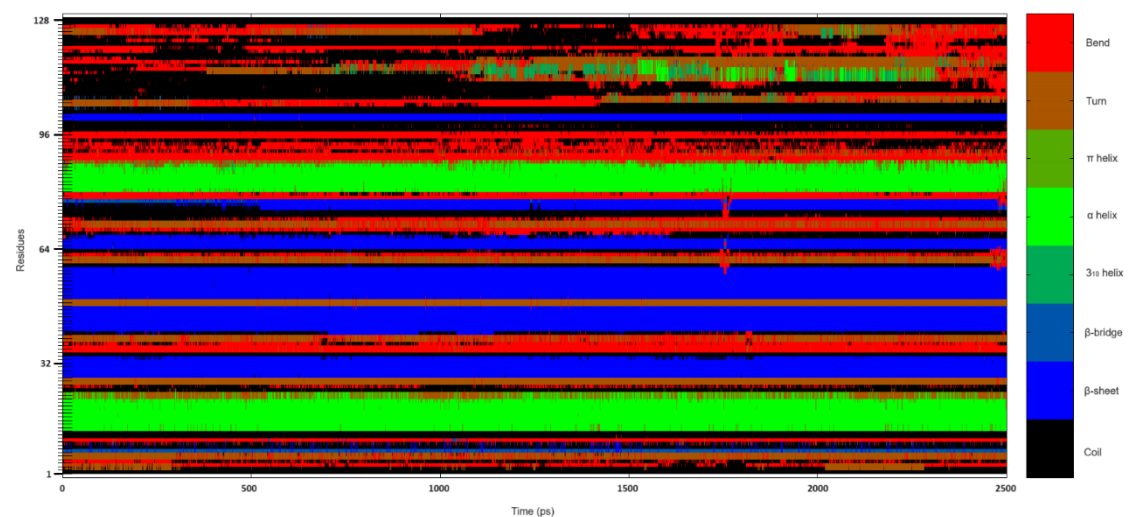


Figure S47: Secondary structures of the I98P mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.

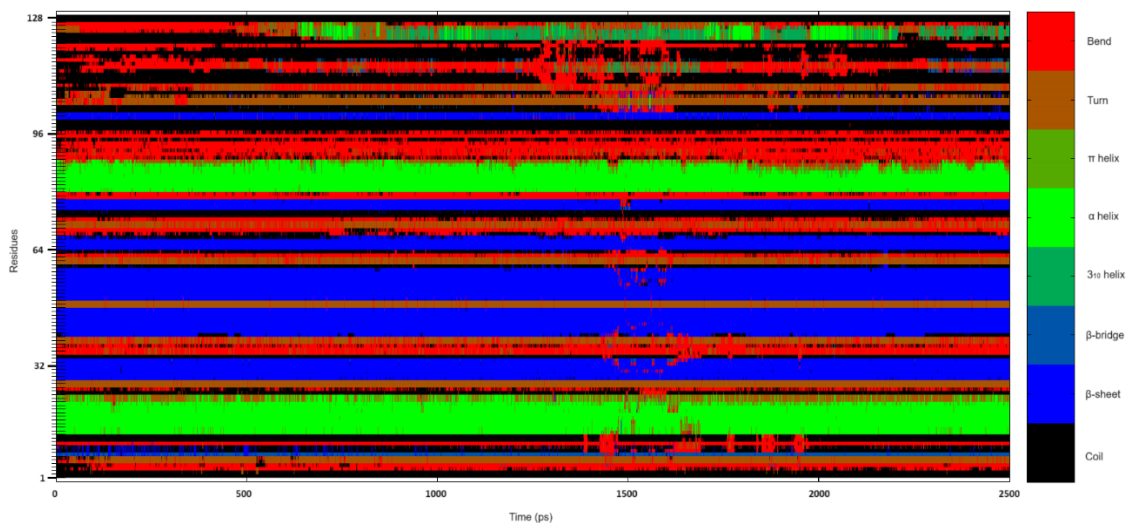


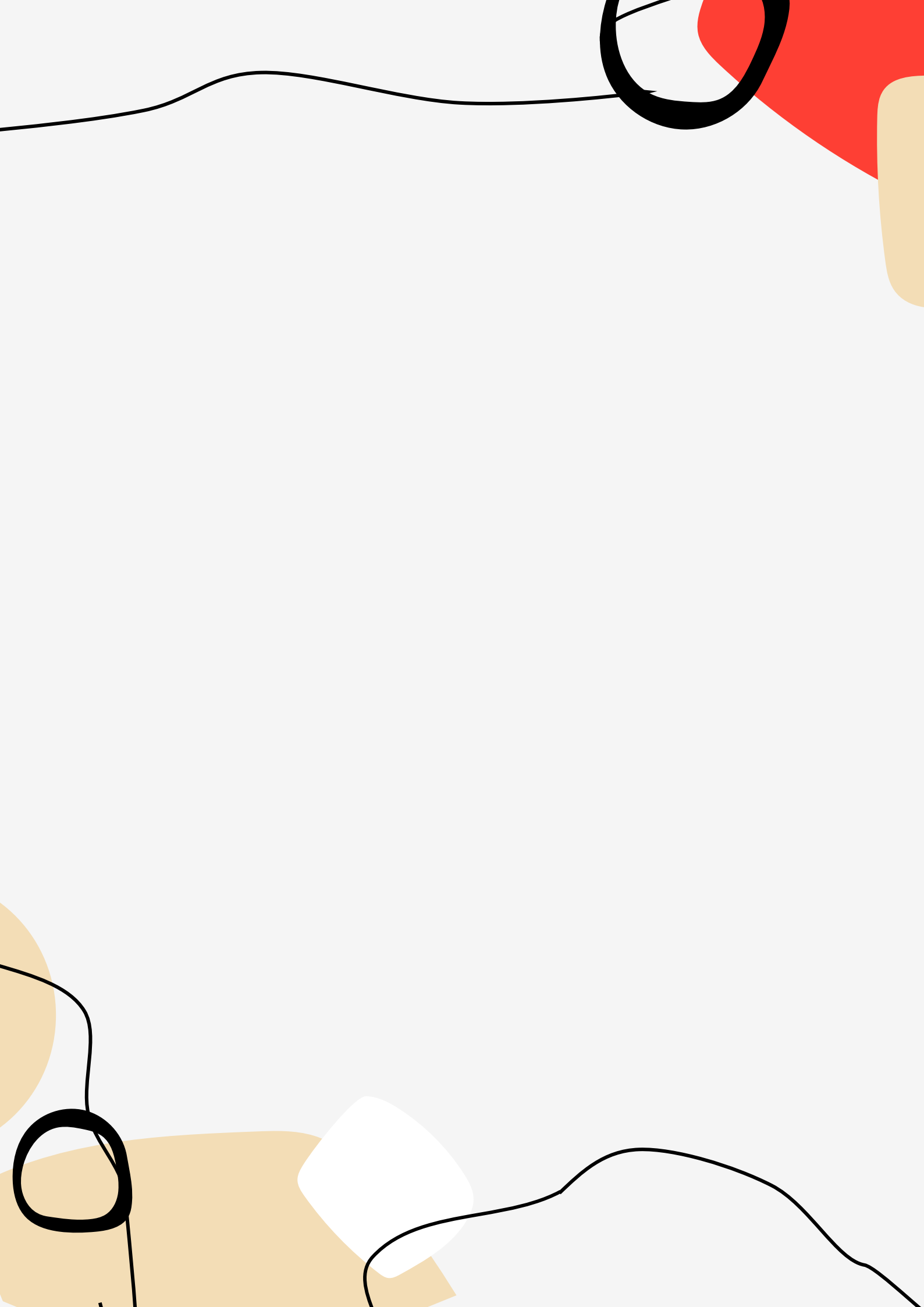
Figure S48: Secondary structures of the V102G mutant protein per residue during the simulation period. Obtained using the GROMACS DSSP algorithm.



CHAPTER

General discussion,
conclusion and future
perspectives

IV



4.1. General Discussion

XLP, or X-linked lymphoproliferative disease, is a rare and complex genetic disorder characterized by an incomplete understanding of its pathophysiological and molecular mechanisms [1]. Mutations in the *SH2D1A* and *XIAP* genes are identified as the primary causes of this condition [2–10]. These genes play a crucial role in the immune response by aiding in the activation of T lymphocytes and NK cells [7,11,12]. Furthermore, the intricate nature of XLP necessitates a deeper exploration of the molecular intricacies underlying the disease. In-depth analyses of the *SH2D1A*, *SH2D1B*, and *XIAP* genes across various primates and other representative mammalian species have provided insights into the unique evolutionary pressures acting on these genetic components, unveiling strong negative to neutral selection signals. Understanding the selective forces shaping these genes is paramount, as it may unveil critical clues about their functional significance in immune regulation.

The identification of specific disease-associated SNPs in the *SH2D1A* and *XIAP* genes, as opposed to the relative scarcity in *SH2D1B*, highlights the potential functional distinctions among these genes [13]. This disparity in disease-causing SNPs occurrence may signify varying degrees of susceptibility to genetic alterations and, consequently, shed light on the specific roles each gene plays in maintaining immune homeostasis.

Moreover, the study delves into the structural impact of specific amino acid substitutions in the *SH2D1A* gene, revealing noteworthy alterations in the protein structure. The substitutions M1K, Y7C, H8D, G27D, G27S, L30S, R32S, R32T, S34G, S34R, Y41N, C42Y, T53I, Y54C, Y54N, R55P, R55Q, I80R, I84T, F87S, G93D, L98P and V102G, disrupt the native structural organization, in accordance with the signals of strong negative selection acting on the *SH2D1A* gene. These results not only emphasize the significance of the *SH2D1A* gene in immune function but also provide a foundation for further investigations into potential therapeutic interventions targeting these specific structural elements.

It is important to emphasise the discrepancy between the I-mutant2.0 and MUpro results, indicating a decrease in the stability of the mutant proteins, and the MDS results, which suggest a possible increase in stability. Different computational methods for assessing protein stability can provide different perspectives, since they use different theoretical approaches and models [14–16]. I-mutant2.0 and MUpro generally estimate the thermodynamic effects of mutations based on physicochemical properties, such as

Gibbs free energy [17,18]. If both indicated a decrease in stability, this suggests that, from the point of view of these predictions, the mutations may be jeopardising the structural stability of the protein. However, MDS is a dynamic approach that simulates the behaviour of the protein over time, allowing conformational fluctuations and molecular dynamics to be observed [19,20]. If the parameter values obtained in the simulation indicate a possible increase in the stability of the mutants compared to the native structure, this may suggest that, in the simulation, the mutants are in fact exploring more stable conformations or showing less structural variation than the native protein.

These discrepancies can be attributed to limitations inherent in the methods used [16]. Prediction models based solely on free energies may not fully capture the dynamic and conformational effects that occur during real molecular dynamics. In addition, it is crucial to consider the quality of the initial structural models and the accuracy of the parameters used in the simulations [19]. Integration of multiple computational approaches and, if possible, experimental validation, can provide a more comprehensive understanding of the effects of mutations on the structural and dynamic properties of proteins [21,22].

4.2. General Conclusion and Future Perspectives

The collective findings provide a comprehensive understanding of XLP, shedding light on potential therapeutic strategies. Genes such as *SH2D1A* and *XIAP* play crucial roles in immune responses, showing conservation across species and undergoing purifying selection, suggesting their ancestral and critical functions.

While the main focus is on the *SH2D1A* gene, the elusive structure of the human XIAP protein poses a challenge for comprehensive molecular dynamics analyses. Future research should prioritise the complete clarification of the XIAP protein structure, as this is key to understanding the impact of SNPs on its configuration. This knowledge gap highlights the continued importance of efforts in molecular biology and structural bioinformatics to decipher the complexities of XIAP and broaden our understanding of its role in XLP.

Beyond molecular insights, the severity of XLP as a life-threatening condition demands a multifaceted approach to treatment. The current reliance on bone marrow transplantation highlights the urgency of therapeutic alternatives. The emergence of chemical substances capable of slowing down or stabilising the pathology represents a promising frontier in XLP research, complementing existing treatment modalities and

offering new strategies for managing the progression of the disease. In the pursuit of therapeutic strategies XLP, leveraging virtual screening to identify compounds interacting with specific binding sites on SH2D1A mutant proteins is crucial. Exploring transient pockets within the protein structure, often overlooked in static analyses, reveals dynamic regions susceptible to conformational changes. Molecular chaperones targeting these pockets can assist in proper protein folding, potentially restoring structural stability. Advanced molecular dynamics simulations and structural bioinformatics dynamically map these pockets, offering accurate insights under physiological conditions. Virtual screening of identified transient pockets for potential molecular chaperone-binding compounds presents a promising strategy to remodel SH2D1A mutant protein integrity. Integrating artificial intelligence and machine learning accelerates drug candidate identification by predicting binding affinities and interactions, advancing targeted therapeutic interventions at the molecular level for XLP.

The XLP remains largely unexplored. Nonetheless, the results presented in this study significantly improve our understanding of the genetic, evolutionary and structural dimensions of the genes associated with XLP. These findings are not only promising for potential therapeutic interventions but also emphasize the importance of careful consideration when interpreting computational predictions.

References

- [1] Y. Wang, Y. Wang, W. Lu, L. Tao, Y. Xiao, Y. Zhou, X. He, Y. Zhang, L. Li, Potential pathogenic mechanism of type 1 X-linked lymphoproliferative syndrome caused by a mutation of SH2D1A gene in an infant: A case report, *Medicine*. 101 (2022) E30951. <https://doi.org/10.1097/MD.00000000000030951>.
- [2] K. Zhang, E. Wakefield, R. Marsh, Lymphoproliferative Disease, X-Linked, *GeneReviews®*. (2016) 1–7. <https://www.ncbi.nlm.nih.gov/books/NBK1406/> (accessed April 7, 2023).
- [3] J.K. Hamilton, L.A. Paquin, J.L. Sullivan, H.S. Maurer, F.G. Cruz, A.J. Provisor, C. Philip Steuber, E. Hawkins, D. Yawn, J.A. Cornet, K. Clausen, G.Z. Finkelstein, B. Landing, M. Grunnet, D.T. Putilo, X-linked lymphoproliferative syndrome registry report, *J Pediatr*. 96 (1980) 669–673. [https://doi.org/10.1016/S0022-3476\(80\)80735-9](https://doi.org/10.1016/S0022-3476(80)80735-9).
- [4] M. Morra, D. Howie, M.S. Grande, J. Sayos, N. Wang, C. Wu, P. Engel, C. Terhorst, X-Linked Lymphoproliferative Disease: A Progressive

- Immunodeficiency, <https://doi.org/10.1146/Annurev.Immunol.19.1.657>. 19 (2003) 657–682. <https://doi.org/10.1146/ANNUREV.IMMUNOL.19.1.657>.
- [5] A. Majeranowski, Z. Biologii Komórki I Immunologii, W. Biotechnologii, N.A. Dułak, J. Juwa, R. Trzciński, N. Irga-Jaworska, Hematological manifestations of X-linked lymphoproliferative disease, *Hematology in Clinical Practice*. 13 (2022) 1–6. <https://doi.org/10.5603/HCP.A2022.0002>.
- [6] K.E. Nichols, R.A. Marsh, The X-Linked Lymphoproliferative Syndromes Chapter Outline, *Stiehm's Immune Deficiencies*. (2014) 475–496. <https://doi.org/10.1016/B978-0-12-405546-9.00022-4>.
- [7] S. Rigaud, M.C. Fondanèche, N. Lambert, B. Pasquier, V. Mateo, P. Soulas, L. Galicier, F. Le Deist, F. Rieux-Laucat, P. Revy, A. Fischer, G. De Saint Basile, S. Latour, XIAP deficiency in humans causes an X-linked lymphoproliferative syndrome, *Nature* 2006 444:7115. 444 (2006) 110–114. <https://doi.org/10.1038/nature05257>.
- [8] N. Panchal, C. Booth, J.L. Cannons, P.L. Schwartzberg, X-linked lymphoproliferative disease type 1: A clinical and molecular perspective, *Front Immunol*. 9 (2018). <https://doi.org/10.3389/FIMMU.2018.00666>.
- [9] M. Morra, M. Simarro-Grande, M. Martin, A.S.I. Chen, A. Lanyi, O. Silander, S. Calpe, J. Davis, T. Pawson, M.J. Eck, J. Sumegi, P. Engel, S.C. Li, C. Terhorst, Characterization of SH2D1A missense mutations identified in X-linked lymphoproliferative disease patients, *J Biol Chem*. 276 (2001) 36809–36816. <https://doi.org/10.1074/JBC.M101305200>.
- [10] L. Torralba-Raga, B. Tesi, S.C.C. Chiang, H. Schlums, M. Nordenskjöld, A.C. Horne, J.I. Henter, M. Meeths, M. Abdelhaleem, S. Weitzman, Y. Bryceson, Diagnostic challenges for a novel SH2D1A mutation associated with X-linked lymphoproliferative disease, *Pediatr Blood Cancer*. 67 (2020). <https://doi.org/10.1002/PBC.28184>.
- [11] P. Engel, M.J. Eck, C. Terhorst, The SAP and SLAM families in immune responses and X-linked lymphoproliferative disease, *Nat Rev Immunol*. 3 (2003) 813–821. <https://doi.org/10.1038/NRI1202>.
- [12] J.L. Sullivan, K.S. Byron, F.E. Brewster, D.T. Purtilo, Deficient Natural Killer Cell Activity in X-Linked Lymphoproliferative Syndrome, *Science* (1979). 210 (1980) 543–545. <https://doi.org/10.1126/SCIENCE.6158759>.

- [13] P.D. Stenson, M. Mort, E. V. Ball, K. Shaw, A.D. Phillips, D.N. Cooper, The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine, *Hum Genet.* 133 (2014) 1–9. <https://doi.org/10.1007/S00439-013-1358-4>.
- [14] V. Potapov, M. Cohen, G. Schreiber, Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details, *Protein Engineering, Design and Selection.* 22 (2009) 553–560. <https://doi.org/10.1093/PROTEIN/GZP030>.
- [15] N. Tokuriki, D.S. Tawfik, Stability effects of mutations and protein evolvability, *Curr Opin Struct Biol.* 19 (2009) 596–604. <https://doi.org/10.1016/J.SBI.2009.08.003>.
- [16] R. Guerois, J.E. Nielsen, L. Serrano, Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations, *J Mol Biol.* 320 (2002) 369–387. [https://doi.org/10.1016/S0022-2836\(02\)00442-4](https://doi.org/10.1016/S0022-2836(02)00442-4).
- [17] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res.* 33 (2005) W306. <https://doi.org/10.1093/NAR/GKI375>.
- [18] J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, *Proteins.* 62 (2006) 1125–1132. <https://doi.org/10.1002/PROT.20810>.
- [19] M.S. Badar, S. Shamsi, J. Ahmed, Md.A. Alam, Molecular Dynamics Simulations: Concept, Methods, and Applications, (2022) 131–151. https://doi.org/10.1007/978-3-030-94651-7_7.
- [20] B.J. Alder, T.E. Wainwright, Studies in Molecular Dynamics. I. General Method, *J Chem Phys.* 31 (1959) 459–466. <https://doi.org/10.1063/1.1730376>.
- [21] R. Cárdenas, J. Martínez-Seoane, C. Amero, Combining Experimental Data and Computational Methods for the Non-Computer Specialist, *Molecules* 2020, Vol. 25, Page 4783. 25 (2020) 4783. <https://doi.org/10.3390/MOLECULES25204783>.
- [22] Experimental validation, anyone?, *Nature Computational Science* 2023 3:5. 3 (2023) 361–361. <https://doi.org/10.1038/s43588-023-00462-x>.

