



UNIVERSITAT DE
BARCELONA

Structure-based drug design of novel antiviral compounds against human herpesvirus and coronavirus

Sara Muriel Goñi

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Structure-based drug design of novel antiviral compounds against human herpesvirus and coronavirus

SARA MURIEL GOÑI

Codirected by

Prof. Miquel Coll Capella and Dr. Montserrat Fabrega



**UNIVERSITAT DE
BARCELONA**

Facultat de Farmàcia i Ciències de l'Alimentació

Biotechnology program

Barcelona, 2022



UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

BIOTECHNOLOGY PROGRAM

**STRUCTURE-BASED DRUG DESIGN OF NOVEL ANTIVIRAL COMPOUNDS AGAINST
HUMAN HERPESVIRUS AND CORONAVIRUS**

Thesis submitted by Sara Muriel Goñi, enrolled in the Biotechnology program at the University of Barcelona, for degree of Doctor.

This work was carried out with a FPI scholarship awarded by the Spanish Ministry of Economy and Competitiveness in the Structural Biology of Protein & Nucleic Acid Complexes and Molecular Machines Group at the Institute for Research in Biomedicine (IRB-Barcelona) and Molecular Biology Institute of Barcelona (IBMB-CSIC), under the supervision of Prof. Miquel Coll Capella and Dr. Montserrat Fabrega and Josefa Badia Palacín as tutor.

SARA MURIEL GOÑI
IRB BARCELONA
IBMB-CSIC

JOSEFA BADIA PALACIN
University of Barcelona

MIQUEL COLL CAPELLA
IRB BARCELONA
IBMB-CSIC

MONTSERRAT FABREGA FERRER
IRB BARCELONA
IBMB-CSIC

Index

ABSTRACT.....	7
LIST OF FIGURES.....	8
LIST OF TABLES	12
LIST OF ABBREVIATIONS AND SYMBOLS.....	13
AMINO ACIDS ABBREVIATIONS	15
PREFACE.....	16
PART I: Herpesvirus	17
CHAPTER 1: INTRODUCTION	17
1.1. Herpesviridae family.....	18
1.1.1. Classification	18
1.1.2. Structure	19
1.1.3. Nucleic acid.....	20
1.1.4. Viral life-cycle.....	21
1.1.5. Genome organization	22
1.1.6. Epidemiology and transmission.....	25
1.1.7. Pathology	26
1.1.8. Treatment.....	28
1.2. Herpesvirus DNA encapsidation	30
1.2.1. Herpesvirus terminase complex	33
1.2.1.1. Tripartite Terminase Subunit 1 (TRM1).....	34
1.2.1.2. Tripartite Terminase Subunit 2 (TRM2).....	35
1.2.1.3. Tripartite Terminase Subunit 3 (TRM3).....	35
1.2.2. DNA encapsidation as a therapeutic target.....	36
CHAPTER 2: OBJECTIVES.....	42
CHAPTER 3: MATERIALS AND METHODS	44
3.1.1. Chemicals and reagents.....	45
3.1.2. Oligonucleotides	45
3.1.3. Plasmid	45
3.1.4. Bacterial strains.....	46
3.1.5. Media	47
3.1.6. Crystallization.....	47
3.1.6.1. Crystallization plates	47
3.1.6.2. Cryo-crystallography loops.....	47
3.1.6.3. Robotic platforms.....	47
3.1.6.4. Software.....	48
3.2. METHODS.....	49

3.2.1.	Sample preparation and analysis.....	49
3.2.2.	Construct design.....	49
3.2.3.	Cloning.....	49
3.2.3.1.	Bacterial transformation.....	50
3.2.3.2.	Plasmid extraction.....	50
3.2.3.3.	Plasmid analysis.....	50
3.2.4.	Protein expression.....	50
3.2.4.1.	Large-scale protein expression.....	50
3.2.4.2.	Protein electrophoresis.....	51
3.2.5.	Protein purification.....	51
3.2.5.1.	Cell lysis.....	52
3.2.5.2.	Chromatographic techniques.....	52
3.2.5.3.	Protein concentration and quantification.....	53
3.2.6.	Protein analysis.....	53
3.2.6.1.	Western blot.....	53
3.2.6.2.	Mass spectrometry.....	54
3.2.6.3.	Dynamic light scattering.....	54
3.2.6.4.	Multi angle light scattering (MALS).....	54
3.2.6.5.	Nuclear magnetic resonance (NMR).....	55
3.2.7.	Biochemical assays.....	55
3.2.7.1.	In vitro nuclease activity assays.....	55
3.2.8.	Crystallization.....	56
3.2.8.1.	Microseeding.....	57
3.2.8.2.	Protein crystallization screening.....	57
3.2.8.3.	Protein crystallization optimization.....	57
3.2.8.4.	Crystal freezing.....	58
3.2.9.	X-ray diffraction.....	58
3.2.9.1.	Data collection.....	59
3.2.9.2.	Data processing.....	59
3.2.10.	Structure determination: molecular replacement.....	59
3.2.10.1.	Model building and refinement.....	59
3.2.10.2.	Final model validation.....	60
3.2.10.3.	Structural data visualization and analysis.....	60
3.2.11.	Model of structures with inhibitors.....	60
3.2.12.	In vitro antiviral screening.....	61
CHAPTER 4: RESULTS AND DISCUSSION.....		63
4.1.	RESULTS AND DISCUSSION.....	64
4.1.1.	Herpesvirus tripartite terminase subunit 3.....	64

4.1.2.	Construct design.....	64
4.1.3.	Cloning.....	69
4.1.4.	Protein expression.....	71
4.1.5.	Protein purification.....	72
4.1.5.1.	Herpes Simplex Virus 2 (HSV-2).....	73
4.1.5.2.	Varicella-Zoster Virus (VZV).....	75
4.1.5.3.	Epstein-Barr virus (EBV).....	77
4.1.5.4.	Human Cytomegalovirus (HCMV).....	79
4.1.5.5.	Kaposi's Sarcoma Herpesvirus (KSHV).....	82
4.1.6.	<i>In vitro</i> nuclease assays.....	85
4.1.6.1.	Ion preference.....	86
4.1.7.	Protein quality analysis.....	87
4.1.8.	Data collection and structure determination.....	94
4.1.9.	Structure refinement and validation.....	95
4.1.10.	Structural analysis TRM3-C HSV2.....	97
4.1.11.	Inhibition assays.....	100
4.1.11.1.	Herpes Simplex Virus 2 (HSV-2).....	101
4.1.11.2.	Varicella-Zoster Virus (VZV).....	102
4.1.11.3.	Epstein-Barr virus (EBV).....	103
4.1.11.4.	Human Cytomegalovirus (HCMV).....	104
4.1.11.5.	Kaposi's Sarcoma Herpesvirus (KSHV).....	105
4.1.11.6.	Summary of results.....	106
4.1.12.	Modeled structures with inhibitors.....	106
4.1.12.1.	Herpes Simplex Virus 2 (HSV-2).....	107
4.1.12.2.	Varicella-Zoster Virus (VZV).....	109
4.1.12.3.	Epstein-Barr virus (EBV).....	111
4.1.12.4.	Kaposi's Sarcoma Herpesvirus (KSHV).....	113
4.1.13.	<i>In vitro</i> antiviral screening report.....	114
4.1.14.	Design of novel terminase inhibitors.....	116
CHAPTER 5: CONCLUSIONS.....		118
BIBLIOGRAPHY.....		121
PART II: Coronavirus.....		124
CHAPTER 6: INTRODUCTION.....		124
6.1.	Coronaviridae family.....	125
6.1.1.	Classification.....	125
6.1.2.	Epidemiology, transmission and pathology.....	127
6.1.3.	Treatment.....	129
6.1.4.	Structure.....	130

6.1.5.	Genome organization	131
6.1.6.	Infection cycle.....	133
6.1.7.	Structure and function of the main protease.....	134
6.1.8.	Coronavirus proteins as drug targets: the relevance of the main protease.....	137
6.1.9.	Main protease inhibitors.....	138
6.1.10.	Antiviral compounds rupintrivir and compound 1	139
CHAPTER 7: OBJECTIVES.....		141
CHAPTER 8: MATERIALS AND METHODS		143
8.1.	MATERIALS	144
8.1.1.	Crystallization.....	144
8.1.1.1.	Crystallization plates	144
8.1.1.2.	Cryo-crystallography loops.....	144
8.1.1.3.	Robotic platforms	144
8.1.1.4.	Software.....	144
8.2.	METHODS	146
8.2.1.	Complex preparation.....	146
8.2.2.	Crystallization trials.....	146
8.2.3.	Structure determination and analysis	146
CHAPTER 9: RESULTS AND DISCUSSION		147
9.1.	RESULTS AND DISCUSSION	148
9.1.1.	Crystallization trials.....	148
9.1.2.	Data collection and structure determination	148
9.1.3.	Structure refinement and validation.....	150
9.1.4.	Structure analysis	151
9.1.5.	Comparison with SARS-CoV-2 Compound 1 and SARS-CoV-2 rupintrivir	155
9.1.6.	Design of novel M ^{pro} inhibitors	159
CHAPTER 10: CONCLUSIONS.....		162
BIBLIOGRAPHY		164

ABSTRACT

The herpesvirus diseases are increasing in importance as a public health problem throughout the world. Members of the human herpesvirus family are global in distribution and infect 60-95% of the world's population, both in developed and in developing countries. The terminase complex plays a key role in genome packaging and is a promising drug target for developing new antiviral compounds. The tripartite terminase 3 subunit C-terminal domain (TRM3-C) of five human herpesviruses (HSV-2, VZV, EBV, HCMV and KSHV) has been biochemically characterized, showing the need of Mn^{2+} cations in order to enhance the nuclease activity. The three-dimensional structure of TRM3-C HSV-2 has been determined by X-ray crystallography and showed a new position of the two Mn^{+2} ions in the active centre. Diverse inhibitors have been tested against the TRM3-C terminal domain of the five herpesviruses. Some of them were developed in our laboratory (BS14 and BS17) and the others are specific inhibitors of the HIV integrase. Docking models of TRM3-C have been obtained with the inhibitors that were able to inhibit the nuclease activity in a low micromolar range (BS14, BS17 and Bictegravir). Cell assays with BS14 showed that is moderately active against HCMV. Considering all the biochemical and structural data, new compounds have been designed to obtain better inhibitors against the TRM3-C target.

Coronaviruses are a large family of viruses that are known to cause human illness ranging from the common cold to more severe diseases such as MERS and SARS. The emergence of SARS-CoV-2 in 2019 has triggered an ongoing global pandemic of the severe pneumonia-like disease coronavirus disease 2019 (COVID-19). The main protease (M^{pro}) is one of the best characterized drug targets among coronaviruses. Compound 1 inhibits the protease of the rhinovirus, which is homolog to the SARS-CoV-1 M^{pro} . Thus, could be a promising drug against coronaviruses. The three-dimensional structure of SARS-CoV-1 M^{pro} in complex with the inhibitor compound 1 has been determined.

LIST OF FIGURES

Figure 1. Structure of herpesviruses.....	20
Figure 2. Major classes of herpesvirus genome architecture.....	21
Figure 3. Schematic representation of the lytic replication cycle of a representative herpesvirus in permissive cells.....	21
Figure 4. Organisation of conserved gene blocks in human herpesviruses representing the three subfamilies.....	25
Figure 5. From full genome to cleavage site.....	30
Figure 6. Conserved sequences at herpesvirus termini.....	31
Figure 7. Genome cleavage/packaging and the HCMV terminase complex.....	33
Figure 8. Structure of the tripartite terminase complex of HSV-1 strain 17.....	34
Figure 9. (A) Overall structure of HCMV TRM3-C.....	36
Figure 10. Rational design of BS2 and BS14 as the TRM3-C HCMV inhibitor.....	38
Figure 11. X-ray crystal structure of BS2 (6EY7) bound to TRM3-C HCMV.....	39
Figure 12. Formula of BS17.....	39
Figure 13. Formula of BS14.....	39
Figure 14. Formula of Bictegravir.....	40
Figure 15. Formula of Elvitegravir.....	40
Figure 16. Formula of Dolutegravir.....	40
Figure 17. Formula of Cabotegravir.....	40
Figure 18. Formula of 707035 inhibitor.....	41
Figure 19. Formula of MK-2048 inhibitor.....	41
Figure 20. pET-28a (+) vector map.....	46
Figure 21. Sequence alignment of TRM3-C terminal domain of HSV-2, VZV, EBV, HCMV and KSHV (Clustal Omega).....	64
Figure 22. Secondary structure prediction of TRM3-C HSV-2.....	65
Figure 23. Secondary structure prediction of TRM3-C VZV.....	66
Figure 24. Secondary structure prediction of TRM3-C EBV.....	67
Figure 25. Secondary structure prediction of TRM3-C HCMV.....	68
Figure 26. Secondary structure prediction of TRM3-C KSHV.....	69
Figure 27. One-percent agarose gel stained with SYBR-Safe revealing amplicon product of TRM3-C of EBV.....	71
Figure 28. EBV TRM3-C overexpression.....	71
Figure 29. HisTrap HP 5 mL chromatographic profile of TRM3-C HSV-2.....	73
Figure 30. SDS-PAGE gel stained with Coomassie of the HisTrap HP 5 mL of TRM3-C HSV-2. ...	73
Figure 31. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C HSV-2.....	74
Figure 32. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C HSV-2.....	74
Figure 33. HisTrap HP 5 mL chromatographic profile of TRM3-C VZV.....	75
Figure 34. SDS-PAGE gel stained with Coomassie of the HisTrap HP 5 mL of TRM3-C VZV.....	75
Figure 35. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C VZV.....	76

Figure 36. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C VZV. ..	76
Figure 37. HisTrap HP 5 mL chromatographic profile of TRM3-C EBV.	77
Figure 38. SDS-PAGE gel stained with Coomassie of the HisTrap HP 5 mL of TRM3-C EBV.....	77
Figure 39. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C EBV.....	78
Figure 40. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C EBV. ..	78
Figure 41. HisTrap HP 5 mL chromatographic profile of TRM3-C HCMV.	79
Figure 42. SDS-PAGE gel stained with Coomassie of the HisTrap HP 5 mL of TRM3-C HCMV.	79
Figure 43. Mono Q 5/50 GL chromatographic profile of TRM3-C HCMV.	80
Figure 44. SDS-PAGE gel stained with Coomassie of the Mono Q 5/50 GL of TRM3-C HCMV.....	80
Figure 45. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C HCMV.	81
Figure 46. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C HCMV.	81
Figure 47. HisTrap HP 5 mL chromatographic profile of TRM3-C KSHV.....	82
Figure 48. SDS-PAGE gel stained with Coomassie of the HisTrap HP 5 mL of TRM3-C KSHV.	82
Figure 49. Mono Q 5/50 GL chromatographic profile of TRM3-C KSHV.....	83
Figure 50. SDS-PAGE gel stained with Coomassie of the Mono Q 5/50 GL of TRM3-C KSHV.	83
Figure 51. Heparin 1mL chromatographic profile of TRM3-C KSHV.	84
Figure 52. SDS-PAGE gel stained with Coomassie of the Heparin affinity chromatography of TRM3-C KSHV.....	84
Figure 53. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C KSHV.....	85
Figure 54. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C KSHV. Since there was a lot of protein, two injections were needed and the gel shows both replicas. ..	85
Figure 55. Agarose gel (1%) of TRM3-C HSV-2, VZV, EBV, HCMV and KSHV with the products of a nuclease assay against linear DNA.	87
Figure 56. Size distribution by volume of TRM3-C EBV measurements by dynamic light scattering.	88
Figure 57. Size distribution by number of TRM3-C EBV measurements by dynamic light scattering.	88
Figure 58. MALS profile of TRM3-C EBV concentrated sample in 10/300 Superdex 75 Increase.	89
Figure 59. 1D ¹ H nuclear magnetic resonance spectra of TRM3-C EBV.	89
Figure 60. Soaking experiment of TRM3-C HCMV with BS14 inhibitor.....	92
Figure 61. Soaking experiment of TRM3-C HCMV with BS17 inhibitor..	93
Figure 62. Tiny crystals of TRM3-C HCMV with mutated active centre.	93
Figure 63. TRM3-C HSV-2 crystal grown in 24-well plate	94
Figure 64. Diffraction pattern of TRM3-C HSV-2 crystal diffracting up to 2.46 Å.	94
Figure 65. Cartoon representation of TRM3-C HSV-2 monomer protein.....	96
Figure 66. Ramachandran plot for the TRM3-C HSV-2.	97
Figure 67. Cartoon representation of chain A of TRM3-C HSV2 monomer protein.	97
Figure 68. Superposition of chain A of TRM3-C HSV2 monomer (blue) and chain D of TRM3-C HCMV (purple).....	98
Figure 69. Zoom of the residues of TRM3-C HSV-2 active site (red) interacting with Mn ⁺² (yellow).	98
Figure 70. Zoom of the residues of TRM3-C HCMV active site (red) interacting with Mn ⁺² (purple).	99
Figure 71. Zoom of the superposition of chain A of TRM3-C HSV2 monomer (blue) and chain D of TRM3-C HCMV (purple).	100

<i>Figure 72. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C HSV-2.....</i>	<i>102</i>
<i>Figure 73. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C VZV.</i>	<i>103</i>
<i>Figure 74. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C EBV.</i>	<i>104</i>
<i>Figure 75. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C HCMV.....</i>	<i>105</i>
<i>Figure 76. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C KSHV.....</i>	<i>106</i>
<i>Figure 77. A) Alphafold TRM3-C HSV-2 with Bictegravir. B) Interactions between Bictegravir and TRM3-C HSV2.....</i>	<i>107</i>
<i>Figure 78. A) Alphafold TRM3-C HSV-2 with BS17. B) Interactions between BS17 and TRM3-C HSV2.....</i>	<i>108</i>
<i>Figure 79. A) Alphafold TRM3-C HSV-2 with BS14. B) Interactions between BS14 and TRM3-C HSV2.....</i>	<i>108</i>
<i>Figure 80. A) Alphafold TRM3-C VZV with Bictegravir. B) Interactions between Bictegravir and TRM3-C VZV.....</i>	<i>109</i>
<i>Figure 81. A) Alphafold TRM3-C VZV with BS17. B) Interactions between BS17 and TRM3-C VZV.....</i>	<i>110</i>
<i>Figure 82. A) Alphafold TRM3-C VZV with BS14. B) Interactions between BS14 and TRM3-C VZV.....</i>	<i>110</i>
<i>Figure 83. A) Alphafold TRM3-C EBV with Bictegravir. B) Interactions between Bictegravir and TRM3-C EBV.....</i>	<i>111</i>
<i>Figure 84. A) Alphafold TRM3-C EBV with BS17. B) Interactions between BS17 and TRM3-C EBV.....</i>	<i>112</i>
<i>Figure 85. A) Alphafold TRM3-C EBV with BS14. B) Interactions between BS14 and TRM3-C EBV.....</i>	<i>112</i>
<i>Figure 86. A) Alphafold TRM3-C KSHV with Bictegravir. B) Interactions between Bictegravir and TRM3-C KSHV.....</i>	<i>113</i>
<i>Figure 87. A) Alphafold TRM3-C KSHV with BS17. B) Interactions between BS17 and TRM3-C KSHV.....</i>	<i>114</i>
<i>Figure 88. A) Alphafold TRM3-C KSHV with BS14. B) Interactions between BS14 and TRM3-C KSHV.....</i>	<i>114</i>
<i>Figure 89. Terminase inhibitor 1.....</i>	<i>116</i>
<i>Figure 90. Terminase inhibitor 2.....</i>	<i>117</i>
<i>Figure 91. Terminase inhibitor 3.....</i>	<i>117</i>
<i>Figure 92. Classification of the seven types of Coronavirus that infected humans.....</i>	<i>126</i>
<i>Figure 93. Global distribution of MERS-CoV and SARS-CoV, respectively.....</i>	<i>128</i>
<i>Figure 94. COVID-19 cases reported weekly by WHO Region, and global deaths, as of 21 November 2021.....</i>	<i>129</i>
<i>Figure 95. Schematic diagram of a Covid (CoV) virion.....</i>	<i>130</i>
<i>Figure 96. Diagrammatic representation of the spike trimer assembled on membranes.....</i>	<i>131</i>
<i>Figure 97. Schematic diagram representing the genomic expression of the avian coronavirus.....</i>	<i>132</i>
<i>Figure 98. Organization of the coronavirus replicase gene products.....</i>	<i>133</i>
<i>Figure 99. Replication cycle of CoVs.....</i>	<i>134</i>

<i>Figure 100. A) Organisation of the RNA genome of SARS-CoV-2 with selected genes B) Schematic representation of polyprotein cleavage sites of SARS-CoV-2.</i>	135
<i>Figure 101. Alignment of the amino acid sequences of crystallised main proteases of SARS-CoV-2 (PDB: 6Y2E), SARS-CoV (PDB: 2BX4) and MERS-CoV (PDB: 5C3N).</i>	136
<i>Figure 102. Superimposition of X-ray crystal structures of the main proteases of SARS-CoV (pink, PDB: 2BX4), MERS-CoV (cyan, PDB: 5C3N) and SARS-CoV-2 (green, PDB: 6Y2E).....</i>	136
<i>Figure 103. Polyprotein cleavage sites recognised by M^{pro} of SARS-CoV-2, SARS-CoV and MERS-CoV.</i>	137
<i>Figure 104. Inhibitors (Compound 1, 2 and 3) of the SARS-CoV-2 main protease M^{pro}.</i>	139
<i>Figure 105. Chemical structure of rupintrivir (AG7088) and its orally bioavailable analogue compound 1 (AG7404).....</i>	139
<i>Figure 106. SARS-Cov-1 M^{pro} with Compound 1 crystal grown in 24-well plate.....</i>	148
<i>Figure 107. Diffraction pattern of SARS-Cov-1 M^{pro} with Compound 1 crystal.....</i>	149
<i>Figure 108. Electron density around Compound 1.....</i>	150
<i>Figure 109. Ramachandran plot for the SARS-CoV-1 with Compound1.</i>	151
<i>Figure 110. Overall structure of SARS-Cov-1 M^{pro} with Compound 1</i>	151
<i>Figure 111. Compound 1 structure in the active site of SARS-Cov-1 M^{pro}.....</i>	152
<i>Figure 112. β-lactam ring of Compound 1 and its interactions with SARS-Cov-1 M^{pro}.....</i>	153
<i>Figure 113. P2 of Compound 1 and its interactions with SARS-Cov-1 M^{pro}.</i>	153
<i>Figure 114. P3 and P4 of Compound 1 and its hydrogen bonds with Glu166 of SARS-Cov-1 M^{pro}.</i>	153
<i>Figure 115. P3 of Compound 1 and its interactions with SARS-Cov-1 M^{pro}.</i>	154
<i>Figure 116. P4 of Compound 1 and its interactions with SARS-Cov-1 M^{pro}.</i>	154
<i>Figure 117. Electrostatic potential surface of SARS-Cov-1 M^{pro} with Compound 1.</i>	154
<i>Figure 118. Compound 1 structure in the active site of SARS-Cov-2 M^{pro}.....</i>	155
<i>Figure 119. P1', P1 and P2 of Compound 1 and its interactions with SARS-Cov-2 M^{pro}.</i>	156
<i>Figure 120. P3 and P4 of Compound 1 and its interactions with SARS-Cov-2 M^{pro}.....</i>	156
<i>Figure 121. Rupintrivir structure in the active site of SARS-Cov-2 M^{pro}.</i>	157
<i>Figure 122. P1 and P2 of rupintrivir and its interactions with SARS-Cov-2 M^{pro}.....</i>	158
<i>Figure 123. P3 and P4 of rupintrivir and its interactions with SARS-Cov-2 M^{pro}.</i>	158
<i>Figure 124. M^{pro} inhibitor 1.</i>	159
<i>Figure 125. M^{pro} inhibitor 2.</i>	160
<i>Figure 126. M^{pro} inhibitor 3.</i>	160
<i>Figure 127. M^{pro} inhibitor 4.</i>	160
<i>Figure 128. M^{pro} inhibitor 5.</i>	161
<i>Figure 129. M^{pro} inhibitor 6.</i>	161

LIST OF TABLES

Table 1. Classification and properties of the human herpesviruses	19
Table 2. Shared genes in members of the family Herpesviridae	24
Table 3. Drugs used to treat Herpesvirus infections.....	29
Table 4. Herpesvirus DNA packaging proteins.....	32
Table 5. Drugs and compounds targeting herpesviral DNA encapsidation and viral capsid assembly.....	38
Table 6. Oligonucleotides sequences.....	45
Table 7. Cell strains used in cloning and protein overexpression.....	46
Table 8. Recipe for SDS-PAGE gels.....	51
Table 9. Features of the expression vector of TRM3-C HSV-2.....	70
Table 10. Features of the expression vector of TRM3-C VZV.....	70
Table 11. Features of the expression vector of TRM3-C EBV.....	70
Table 12. Features of the expression vector of TRM3-C HCMV.....	70
Table 13. Features of the expression vector of TRM3-C KSHV.....	70
Table 14. Protein concentration per liter of initial culture.....	72
Table 15. Automated Crystallography Platform screens.....	90
Table 16. Crystallographic data of TRM3-C HCMV with BS14 inhibitor.....	91
Table 17. Refinement statistics for TRM3-C HCMV with BS14 inhibitor.....	91
Table 18. Crystallographic data of TRM3-C HCMV with BS17 inhibitor.....	92
Table 19. Refinement statistics for TRM3-C HCMV with BS17 inhibitor.....	92
Table 20. Crystallographic data and processing statistics of TRM3-C HSV-2.....	95
Table 21. Refinement statistics for TRM3-C HSV-2, as calculated with Molprobity.....	96
Table 22. Distances between the two metal cations (Mn^{+2}) and the residues of the active site of TRM3-C HSV-2.....	99
Table 23. Distances between the two metal cations (Mn^{+2}) and the residues of the active site of TRM3-C HCMV.....	100
Table 24. IC50 values of each inhibitor in the presence of the TRM3-C terminal domain of all the herpesviruses tested: HSV-2, VZV, EBV, HCMV and KSHV.....	106
Table 25. In vitro antiviral screening of BS14 against HSV-1, HSV-2, VZV, HCMV, EBV, HHV6-B and KSHV.....	115
Table 26. In vitro antiviral screening of BS17 against HSV-1, HSV-2, VZV, HCMV, EBV, HHV6-B and KSHV.....	116
Table 27. Examples of Coronavirus species.....	126
Table 28. Automated crystallography platform screens used to crystallize the M^{pro} -Compound 1 complex.....	146
Table 29. Crystallographic data and processing statistics of SARS-Cov-1 M^{pro} with Compound 1.....	149
Table 30. Refinement statistics for SARS-Cov-1 M^{pro} with Compound 1, as calculated with Molprobity.....	150
Table 31. Hydrogen bonds between SARS-CoV-1 active site and Compound1.....	152
Table 32. Hydrogen bonds between SARS-CoV-2 active site and Compound1.....	155
Table 33. Distances between SARS-CoV-2 active site and rupintrivir.....	157

LIST OF ABBREVIATIONS AND SYMBOLS

Å Angstrom
AIDS Acquired Immune Deficiency Syndrome
AU Absorbance units
Bp Base pairs
BSA Bovine serum albumin
CDV cidofovir
Chl Chloramphenicol
CoV Coronavirus
C-terminal Carboxy-terminal
COVID-19 Coronavirus Disease 2019
CSS Complexation Significance Score
DKA diketo acid
DLS Dynamic light scattering
DMSO Dimethyl sulfoxide
DNA Deoxyribonucleic acid
dsDNA Double-stranded DNA
E envelope protein
EBV Epstein-Barr virus
E. coli *Escherichia coli*
FDA Food and Drug Administration
g Gram
GCV ganciclovir
h Hour
HCMV human cytomegalovirus
HE hemagglutinin-esterase
HHV-6 human herpesvirus 6
HHV-7 human herpesvirus 7
HSV-1 herpes simplex virus 1
HSV-2 herpes simplex virus 2
His-tag Poly(6) Histidine-tag
IEX Ion exchange chromatography
IMAC immobilized metal ion affinity chromatography
IPTG Isopropyl-D-beta-galactopyranoside
IR_L internal long repeat
IR_S internal short repeat
IV intravenose
Kan Kanamycin
kDa Kilodalton
KSHV Kaposi's sarcoma associated herpesvirus
l Liter
LB Luria-Bertani broth
μ Micro (10⁻⁶)

M membrane glycoprotein
M Molar
m Milli (10^{-3})
MALS multi angle light scattering
MERS Middle East Respiratory Syndrome
M^{pro} main protease
MR Molecular replacement
min Minute
n Nano (10^{-9})
N nucleocapsid protein
NMR Nuclear magnetic resonance
N-terminal Amino-terminal
OD Optical density
PAC Automated Crystallography Platform
PAGE Polyacrilamide gel
PCR Polymerase chain reaction
PDB Protein Data Bank
PEG Polyethylene glycol
PHENIX Python-based Hierarchical ENvironment for Integrated Xtallography
ppm Parts per million
PSA Ammonium persulfate
RNA Ribonucleic acid
rpm Revolutions per minute
s Seconds
S spyke glycoprotein
SARS-CoV severe acute respiratory syndrome coronavirus
SDS Sodium dodecyl sulphate
SEC Size exclusion chromatography
TAE Tris acetate EDTA buffer
TEMED N,N,N,N,-tetramethylendiamine
TR terminal repeat
Tris (hydroxymethyl) aminomethane
TR_L terminal long repeat
TRM1 Tripartite terminase subunit 1
TRM2 Tripartite terminase subunit 2
TRM3 Tripartite terminase subunit 3
TR_s terminal short repeat
U unique
UL long unique
US short unique
UV Ultraviolet
V Volt
VZV varicella-zoster virus
[v/v] Volume per volume
WHO World Health Organization
[w/v] Weight per volume
1D Mono-dimensional
1H Hydrogen-1
3D Three-dimensional

AMINO ACIDS ABBREVIATIONS

1-LETTER CODE	3-LETTER CODE	AMINO ACID
A	Ala	alanine
R	Arg	arginine
N	Asn	asparagine
D	Asp	Aspartic acid
C	Cys	cysteine
E	Glu	glutamic acid
Q	Gln	glutamine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
L	Leu	leucine
K	Lys	lysine
M	Met	methionine
F	Phe	phenylalanine
P	Pro	proline
S	Ser	serine
T	Thr	threonine
W	Try	tryptophan
Y	Tyr	tyrosine
V	Val	valine

PREFACE

A virus is a small parasite that cannot undergo self-reproduction and once it infects a susceptible cell, the virus can direct the cell machinery to produce more viruses. There are hundreds of viruses able to infect humans, and a number of them cause diseases that range from mild to life threatening.

The nine human herpesviruses are among the most widespread infectious pathogens in the world, with 90% of the population infected with three or more different herpesviruses. While there are effective treatments for some kinds of herpesviruses, the approved drugs can be toxic or face problems of drug-resistance.

Recently, a novel coronavirus causing Coronavirus Disease 2019 (COVID-19) has led to severe respiratory damage to humans and deaths in Asia, Europe, and Americas, which allowed the World Health Organisation (WHO) to declare the pandemic state.

Due to all impacts caused by herpesviruses and coronaviruses, it is evident that the development of new antiviral agents is an unmet need.

This thesis is divided in two parts. The first part of the project was focused on structural and biochemical characterization of the tripartite terminase C-terminal subunit (TRM3-C) of herpesvirus in order to identify potent inhibitors against the TRM3-Ctarget. The aim of the second part of the project was to determine the three-dimensional structure of SARS-Cov1-M^{pro} covalently linked to compound 1 inhibitor. Each part includes introduction, objectives, materials and methods, results and discussion, conclusion and bibliography.

PART I: Herpesvirus

CHAPTER 1: INTRODUCTION

1. INTRODUCTION

1.1. Herpesviridae family

1.1.1. Classification

Herpesviruses comprise a large and diverse order of double stranded DNA viruses that infect humans and a wide range of other hosts (*Pellet and Roizman, 2007*). The former family Herpesviridae has been split into three families, which have been incorporated into the new order Herpesvirales. The revised family *Herpesviridae* retains the mammal, bird and reptile viruses, the new family *Alloherpesviridae* incorporates the fish and frog viruses, and the new family *Malacoherpesviridae* contains a bivalve virus. The *Herpesviridae* family is subdivided into three subfamilies, the *Alpha-*, *Beta-*, and *Gammaherpesvirinae*. Within subfamilies, groups of related herpesvirus species are classified into genera. The nine species of human herpesviruses are distributed across the three subfamilies and several genera (*Table 1*). Prior studies found that the *Beta-* and *Gammaherpesvirinae* are more closely related to each other than to *Alphaherpesvirinae* (*Montague and Hutchison, 2000*). In contrast to some other human viruses, the human herpesviruses have a long evolutionary history, with evidence suggesting that the primordial herpesvirus diverged into the *Alpha-*, *Beta*, and *Gammaherpesvirinae* approximately 180 million to 220 million years ago (*McGeoch et al., 1995*).

Subfamily	Genus	Species	Common name	Genome length (kb)
<i>Alphaherpesvirinae</i>				
	<i>Simplexvirus</i>	<i>Human alphaherpesvirus 1</i>	Herpes simplex 1 (HSV1)	152
	<i>Simplexvirus</i>	<i>Human alphaherpesvirus 2</i>	Herpes simplex 2 (HSV2)	155
	<i>Varicellovirus</i>	<i>Human alphaherpesvirus 3</i>	Varicella-zoster virus (VZV)	125
<i>Betaherpesvirinae</i>				
	<i>Cytomegalovirus</i>	<i>Human betaherpesvirus 5</i>	Human cytomegalovirus (HCMV)	236
	<i>Roseolovirus</i>	<i>Human betaherpesvirus 6A</i>	Human herpesvirus 6A (HHV-6A)	159
	<i>Roseolovirus</i>	<i>Human betaherpesvirus 6B</i>	Human herpesvirus 6B (HHV-6B)	162
	<i>Roseolovirus</i>	<i>Human betaherpesvirus 7</i>	Human herpesvirus 7 (HHV-7)	153
<i>Gammaherpesvirinae</i>				
	<i>Lymphocryptovirus</i>	<i>Human gammaherpesvirus</i>	Epstein-Barr Virus (EBV)	172

	4		
<i>Rhadinovirus</i>	<i>Human gammaherpesvirus 8</i>	Kaposi sarcoma- associated herpesvirus (KSHV)	138

Table 1. Classification and properties of the human herpesviruses (Zmasek et al., 2019).

The grouping of herpesviruses into sub-families serves the purpose of identifying evolutionary relatedness as well as summarizing unique properties of each member. The members of the *Alphaherpesvirinae* sub-family are characterized by an extremely short reproductive cycle (hours), prompt destruction of the host cell and the ability to replicate in a wide variety of host tissues. They characteristically establish latent infection in sensory nerve ganglia. This sub-family consists of herpes simplex virus 1 and 2 (HSV-1 and HSV-2) and varicella-zoster virus (VZV). In contrast to the *Alphaherpesvirinae*, *Betaherpesvirinae* [human cytomegalovirus (HCMV), human herpesvirus 6 (HHV-6) and human herpesvirus 7 (HHV-7)] have a restricted host range. Their reproductive life cycle is long (days), with infection progressing slowly in cell culture systems. A characteristic of these viruses is their ability to form enlarged cells. These viruses can establish latent infection in secretory glands, cells of the reticuloendothelial system, and the kidneys. Finally, the *Gammaherpesvirinae* [Epstein-Barr virus (EBV) and Kaposi's sarcoma associated herpesvirus (KSHV)] have the most limited host range. They replicate in lymphoblastoid cells *in vitro* and can cause lytic infections in certain targeted cells. Latent virus has been demonstrated in lymphoid tissue (Whitley, 1996).

1.1.2. Structure

Membership in the family *Herpesviridae* is based on the structure of the virion (Figure 1). These viruses consist of a single linear double-stranded DNA molecule in the form of a torus which is located at the central core. The DNA exists under pressure in a liquid-crystalline array that fills the entire internal volume of the capsid (Booy et al., 1991). Herpesvirus DNA varies in molecular weight from approximately 80 to 150 million Da, or 120 to 250 kb pairs, depending on the virus. This DNA core is surrounded by a capsid which consists of 162 capsomers, arranged in icosapentahedral symmetry. The capsid is approximately 100 to 110 nanometers in diameter. The tegument is structurally the most complex part of the virion, but remains poorly defined (Kelly et al., 2009). Located between the capsid and the envelope, it contains many viral (and possibly some host) proteins, as well as viral and cellular transcripts (Loret et al., 2008), not all of which are required for virion formation. Individual tegument proteins can vary markedly in abundance. Enveloped tegument structures lacking capsids can assemble in infected cells in cell culture and are released from cells along with virions. Electron tomography indicates that there are inner (capsid-associated) and outer (envelope-associated) tegument layers in virions, and that capsids may be situated non-centrally within the envelope to form an asymmetric tegument cap (Grünewald et al., 2003). Loosely surrounding the capsid and tegument is a lipid bilayer envelope derived from altered host cell membranes. The envelope consists of polyamines, lipids, and a dozen unique viral glycoproteins. These glycoproteins confer distinctive properties to each virus and provide unique antigens to which the host is capable of responding. It is spherical to pleomorphic, 150-200 nm in diameter (Whitley, 1996).

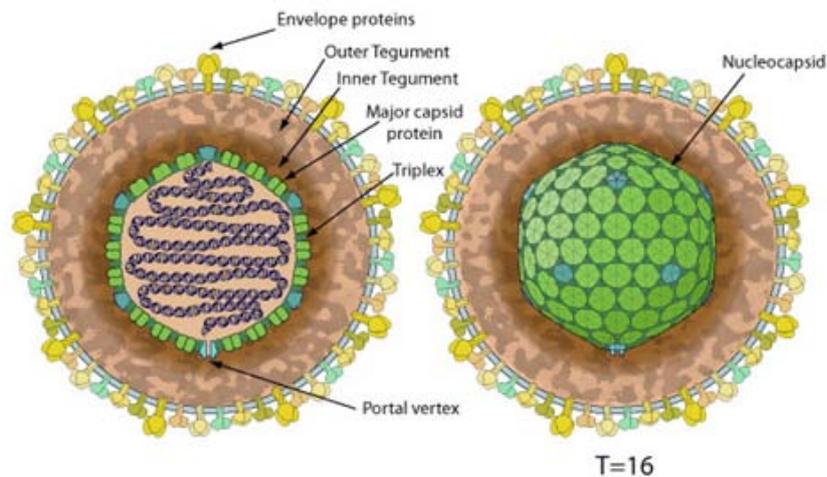


Figure 1. Structure of herpesviruses. The figure shows the different parts of the virion: the envelope proteins, the tegument, the major capsid protein, the triplex, the portal vertex and the nucleocapsid (ViralZone 2017).

1.1.3. Nucleic acid

The linear, double-stranded DNA molecule in herpesviruses virion cores ranges from 125 kbp (VZV) to 241 kbp (chimpanzee cytomegalovirus). Members of the subfamily *Alphaherpesvirinae*, especially those in the genus *Varicellovirus*, range widely in nucleotide composition, from 32% G+C to 78% G+C. The genomes examined in sufficient detail have been shown to contain a single nucleotide extension at the 3' end of each strand, and no terminal protein has been identified (Gatherer *et al.*, 2021).

The arrangement of reiterated sequences (direct or inverted repeats at the genome termini or internally) results in several classes of genome architecture, some including orientation isomers resulting from recombination between terminal and internal inverted repeats (Figure 2). Class 1 genomes consist of a unique sequence flanked by a direct repeat that may be as short as 31 bp (murine cytomegalovirus) or as long as several kbp (HHV-6A and HHV-6B). Class 2 genomes also contain a unique sequence, but in this case it is flanked by a variable number of direct repeats at each terminus (KSHV). Class 3 genomes contain different elements at each terminus that are also present internally in inverted orientation. The genome is thus divided into two unique regions (one long and one short), which are flanked by inverted repeats. Homologous recombination in replicated concatemeric DNA results in inversion of the two unique regions, and cleavage largely or entirely at one of the two junction regions results in unit length genomes that are one or the other of two isomers differing in the orientation of the short unique region (Davison, 1984). Class 4 genomes have the most complex architecture. Like class 3 genomes, they contain long and short unique regions, but these are both flanked by large inverted repeats. Homologous recombination and cleavage occur with equal probability at either of the two junction regions and result in the formation of four isomers differing in the orientations of the unique sequences, with each isomer equimolar in virion populations (HSV-1). In addition, class 4 genomes contain a short terminal direct repeat, which is also present internally in inverse orientation at the junction region.

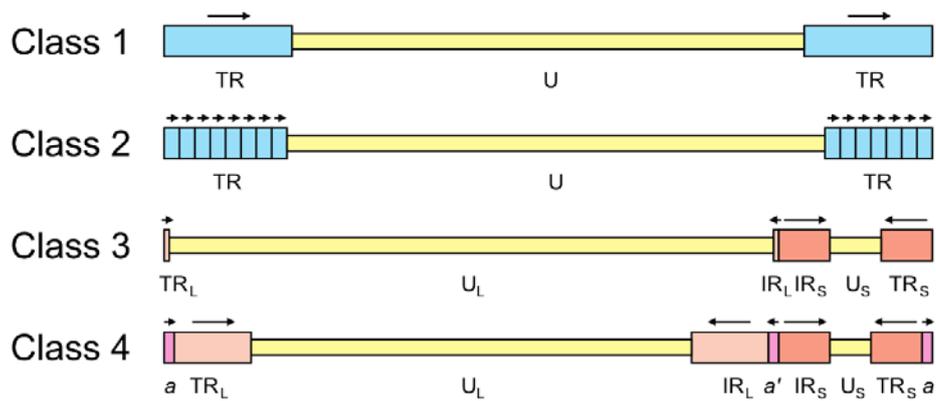


Figure 2. Major classes of herpesvirus genome architecture. In this illustration, unique and repeated sequences are shown in thinner and thicker format, respectively. The nomenclature used to describe these sequences is U (unique), U_L (long unique), U_S (short unique), TR (terminal repeat), TR_L (terminal long repeat), IR_L (internal long repeat), TR_S (terminal short repeat) and IR_S (internal short repeat). Unique sequences are shaded yellow, direct repeats are shaded blue and inverted repeats are shaded red. The orientations of repeated sequences are shown by arrows (Gatherer et al., 2021).

Particular genome structures are associated with certain taxa. Thus, class 2 genomes are common in members of the subfamily *Gammapherpesvirinae* (though not all members have this structure), and class 3 genomes are associated with members of the genus *Varicellovirus* in the subfamily *Alphaherpesvirinae*. However, distantly related viruses may have equivalent genome structures that have presumably evolved convergently (Gatherer et al., 2021).

1.1.4. Viral life-cycle

The precise details of the replication strategy during lytic infection vary from virus to virus, particularly in regard to the subfamily to which the virus belongs and the host cell type infected, but there are many general similarities (Figure 3) (Zerboni et al., 2014).

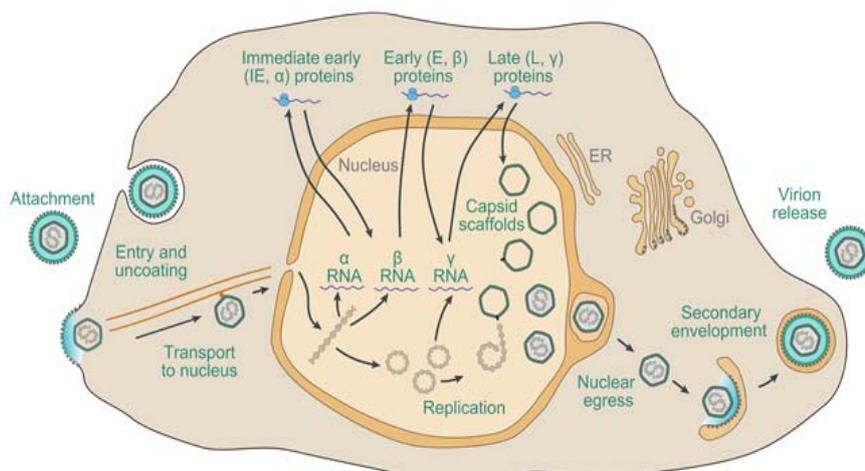


Figure 3. Schematic representation of the lytic replication cycle of a representative herpesvirus in permissive cells.

Adsorption of virions to the cell and penetration into the cytoplasm involve the interaction of multiple virion envelope proteins with multiple cell surface receptors. Entry takes place by membrane fusion either at the cell surface or following endocytosis of the attached virion, promoted by core glycoprotein complexes that include conserved glycoprotein B. The capsid is transported to the region of a nuclear pore by retrograde microtubule transport. Tegument proteins are thought to modify cellular metabolism to favour viral replication. In permissive cells, entry of the genome into the nucleus is followed by a transcriptional cascade (*Roizman and Campadelli-Fiume 2007*). Immediate early (IE or α) genes, which are largely distinct among members of different subfamilies, regulate subsequent gene expression by transcriptional and post-transcriptional mechanisms. Early (E or β) genes encode the DNA replication complex and a variety of enzymes and other proteins involved in modifying host cell metabolism. Finally, late (L or γ) genes encode virion proteins. IE genes can be transcribed in the absence of de novo protein synthesis, E gene transcription is dependent on expression of IE proteins, and L gene transcription is dependent on viral DNA synthesis. With the exception of a few small, non-translated RNAs expressed by some members of the subfamily *Gammaherpesvirinae*, transcription involves host RNA polymerase II (*Gatherer et al., 2021*).

Viral DNA synthesis occurs from one or more origins of DNA replication, probably by a rolling-circle mechanism from circularised genomes, to generate concatemers. DNA replication requires seven viral proteins: an origin-binding helicase, a single-stranded DNA-binding protein, a DNA polymerase composed of a catalytic subunit and a processivity subunit, and a helicase-primase complex comprising a primase subunit, a helicase subunit and a third subunit (*Weller and Coen 2012*). Newly synthesized viral DNA is cleaved into unit-length genomes from concatemers and packaged into preformed immature capsids within the nucleus by processes that involve several viral proteins. Capsids bud through the inner nuclear membrane into the perinuclear space in a process termed primary envelopment, and are then de-enveloped by fusion with the outer nuclear membrane and released into the cytoplasm. Assembly of tegument proteins and secondary envelopment to generate mature virions appears to involve a Golgi or post-Golgi compartment.

The alternative to lytic infection and consequent cell death is latent infection, whereby the virus enters a dormant state within the host with occasional reactivation leading to limited production of virions. Unlike lytic infection, the molecular mechanisms involved in latent infection appear to differ among members of different subfamilies and perhaps even among members of different genera within a subfamily. However, the cell lineages targeted for latency largely follow subfamily lines, with neuronal, myeloid and lymphoid cells most relevant for members of the subfamilies *Alphaherpesvirinae*, *Betaherpesvirinae* and *Gammaherpesvirinae*, respectively. Like other large eukaryotic DNA viruses, some family members are in development as vectors for gene therapy and some, particularly members of the subfamily *Betaherpesvirinae*, are being considered as platforms for self-disseminating vaccines against other pathogens (*Murphy et al., 2016*).

1.1.5. Genome organization

The number of human herpesvirus genes that encode functional proteins ranges from about 70 (VZV) to about 170 (HCMV). Splicing occurs in a minority of genes, the

proportion being lowest in members of the subfamily *Alphaherpesvirinae*. Ribosome profiling has identified a large number of additional open reading frames (ORFs) that may encode proteins in some family members, some of which have been detected by proteomics (Stern-Ginossar *et al.*, 2012, Whisnant *et al.*, 2020). Except where these additional ORFs correspond to shorter or longer versions of established protein-coding regions, most are generally not conserved among related viruses and whether they encode functional proteins has not been determined. In addition to proteins, family members also encode RNAs that do not encode proteins, including long noncoding RNAs and micro RNAs. High-throughput technology has enabled the sequencing of large numbers of strains of the most extensively studied viruses, including viral genomes in clinical material, and has led to a deeper understanding of the diversity and evolution of the family (Renner & Szpara 2018).

A subset of 43 genes is detectably conserved among family members, although one or two have been lost in some lineages (Table 2). These genes were presumably inherited from an ancestral herpesvirus and occur in six blocks that are arranged differently in members of the three subfamilies (Figure 4). Additional conserved genes may have been lost from members of the subfamily *Alphaherpesvirinae* but retained in members of the other subfamilies, following the early separation of this lineage (McGeoch *et al.*, 1995). The conserved genes typically encode capsid proteins, components of the DNA replication and packaging machinery, nucleotide modifying enzymes, membrane proteins and tegument proteins, and, to a lesser extent, regulatory proteins. This reinforces the view that, despite their genetic diversity, family members share common features in many aspects of their replication strategies.

<i>Alphaherpes-</i>	<i>Betaherpes-</i>	<i>Gammaherpes-</i>	Protein name
<i>virinae</i>	<i>Virinae</i>	<i>virinae</i>	
HSV-1	HCMV	KSHV	
<i>DNA replication machinery</i>			
UL30	UL54	ORF9	DNA polymerase catalytic subunit
UL42	UL44	ORF59	DNA polymerase processivity subunit
UL9	Lost	lost	DNA replication origin-binding helicase; present in members of the subfamily <i>Alphaherpesvirinae</i> and some members of the subfamily <i>Betaherpesvirinae</i>
UL5	UL105	ORF44	Helicase-primase helicase subunit
UL8	UL102	ORF40	Helicase-primase subunit
UL52	UL70	ORF56	Helicase-primase primase subunit
UL29	UL57	ORF6	Single-stranded DNA-binding protein
<i>Enzymes peripheral to DNA replication</i>			
UL23	Lost	ORF21	Thymidine kinase; present in members of the subfamilies <i>Alphaherpesvirinae</i> and <i>Gammaherpesvirinae</i>
UL39	UL45	ORF61	Ribonucleotide reductase subunit 1; not an active enzyme in members of the subfamily <i>Betaherpesvirinae</i>
UL40	Lost	ORF60	Ribonucleotide reductase subunit 2; present in members of the subfamilies <i>Alphaherpesvirinae</i> and <i>Gammaherpesvirinae</i>

UL50	UL72	ORF54	Deoxyuridine triphosphatase; not an active enzyme in members of the subfamily <i>Betaherpesvirinae</i>
UL2	UL114	ORF46	Uracil-DNA glycosylase
Processing and packaging of DNA			
UL12	UL98	ORF37	Deoxyribonuclease
UL15	UL89	ORF29	DNA packaging terminase subunit 1
UL28	UL56	ORF7	DNA packaging terminase subunit 2
UL6	UL104	ORF43	Capsid portal protein
UL25	UL77	ORF19	DNA packaging tegument protein UL25
UL32	UL52	ORF68	DNA packaging protein UL32
UL33	UL51	ORF67A	DNA packaging protein UL33
UL17	UL93	ORF32	DNA packaging tegument protein UL17
Egress of capsids from nucleus			
UL31	UL53	ORF69	Nuclear egress lamina protein
UL34	UL50	ORF67	Nuclear egress membrane protein
Capsid assembly and structure			
UL19	UL86	ORF25	Major capsid protein
UL18	UL85	ORF26	Capsid triplex subunit 2
UL38	UL46	ORF62	Capsid triplex subunit 1
UL35	UL48A	ORF65	Small capsid protein
UL26	UL80	ORF17	Capsid maturation protease
UL26.5	UL80.5	ORF17.5	Capsid scaffold protein
Tegument			
UL7	UL103	ORF42	Tegument protein UL7
UL11	UL99	ORF38	Myristylated tegument protein
UL14	UL96	ORF35	Tegument protein UL14
UL16	UL94	ORF33	Tegument protein UL16
UL36	UL48	ORF64	Large tegument protein
UL37	UL47	ORF63	Tegument protein UL37
UL51	UL71	ORF55	Tegument protein UL51
Surface and envelope			
UL27	UL55	ORF8	Envelope glycoprotein B
UL1	UL115	ORF47	Envelope glycoprotein L
UL22	UL75	ORF22	Envelope glycoprotein H
UL10	UL100	ORF39	Envelope glycoprotein M
UL49A	UL73	ORF53	Envelope glycoprotein N
Control and modulation			
UL13	UL97	ORF36	Tegument serine/threonine protein kinase
UL54	UL69	ORF57	Multifunctional expression regulator
Unknown			
UL24	UL76	ORF20	Nuclear protein UL24

Table 2. Shared genes in members of the family Herpesviridae (Gatherer et al., 2021).

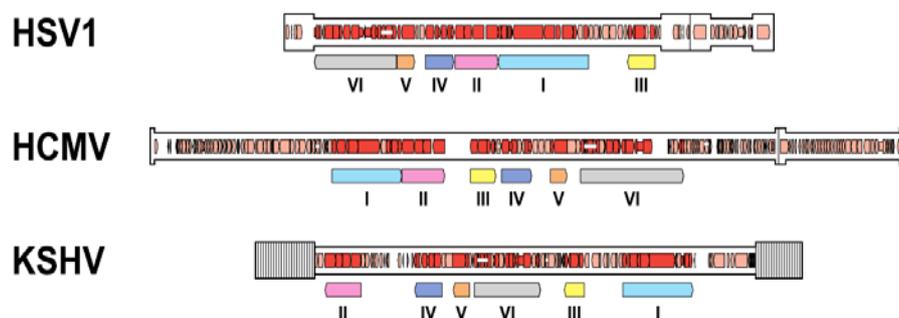


Figure 4. Organisation of conserved gene blocks in human herpesviruses representing the three subfamilies. The genomes illustrated are those of HSV-1, HCMV and KSHV. The conserved gene blocks (I-VI) are indicated by open, coloured arrows below the genomes, and are ordered relative to their arrangement in HCMV. Direct or inverted repeats are shown in a thicker format than unique regions. Functional protein-coding regions are indicated by open, coloured arrows, with bright red indicating genes shared with members of the other subfamilies and light red indicating other genes. Some protein-coding regions are designated by narrower coloured arrows merely to make their locations clearer. Introns connecting protein-coding regions are shown as narrow white bars (Gatherer et al., 2021).

1.1.6. Epidemiology and transmission

Members of different herpesvirus species have distinct epidemiological or biological characteristics and distinct genomes that represent independent replicating lineages. Replicating lineages of herpesviruses are identified primarily on the basis of information derived from genome sequences. For some well-studied genes, there are levels of sequence difference beyond which the viruses in question are presumed to have distinct epidemiological and biological properties; such viruses can be reliably recognised as members of different species on the basis of limited sequence information. There are also closely related viruses that have relatively small differences in the sequences of individual genes, but these differences extend across the respective genomes in a manner indicating that they represent independent replicating lineages

The range of vertebrate hosts in which family members have been discovered is very broad, extending from reptiles to birds and mammals, and the most extensively studied hosts harbour several different herpesviruses (for example, humans are host to nine). In general, the natural host range of individual viruses is restricted, with most having evolved in association with their host species. As a result, herpesviruses are highly adapted to their hosts, and severe infection is usually observed only in the foetus, the very young, the immunocompromised or following infection of an alternative host. Despite this general picture of coevolution, which in some instances may track host speciation, there is phylogenetic evidence that cross-species transmission has played an important part in the evolution of the family (McGeoch et al., 2000, Ehlers et al., 2008).

Host range varies considerably in experimental animal systems: some members of the subfamily *Alphaherpesvirinae* can infect a wide variety of animal species, whereas members of the subfamilies *Betaherpesvirinae* and *Gammaherpesvirinae* exhibit a very restricted experimental host range. Host range in cell culture also varies, though the same general rule holds true. The basis of host restriction is not well understood either *in vivo* or in cell culture. The presence and location of specific cell surface receptors determines tropism, and various other factors, including the interplay between innate cellular defences and immune evasion mechanisms, and the presence of host-specific transcription factors, determine the overall permissiveness of cells *in vivo* to productive infection (Adler et al., 2017). Natural transmission routes range from aerosol spread (VZV) to intimate oral (EBV) or sexual contact (HSV-2). In addition to an exogenous route of infection, the HHV6A and HHV6B genomes are transmitted in about 1% of people via the host germ line when integrated into chromosomal telomeres (Clark, 2016) via human telomere-like repeats near the

genome termini. Some family members, such as HSV1, can retain infectivity after long periods in drinking water (*Dayaram et al., 2017*), although the role of water in natural transmission is not clear. Transmission can occur through direct mucosal contact or via droplets and fomites. Most family members establish a systemic infection via a cell-associated viraemia during primary infection. Infection with some members of the subfamily *Alphaherpesvirinae* may only produce localised lesions on the skin or mucosa of the oral or genital tracts, whereas more generalised infection can occur in young, immunocompromised or non-definitive hosts (*James et al., 2014*). The ability of family members to establish life-long latent infection is considered to be key to their survival. Certain cell types that harbour latent virus have been identified, and this has suggested a general pattern in which most members of the subfamily *Alphaherpesvirinae* establish latent infection in neurons, members of the subfamily *Betaherpesvirinae* in haematopoietic cells, and members of the subfamily *Gammaherpesvirinae* in mononuclear cells, predominantly lymphocytes (*Calnek et al., 1981*).

1.1.7. Pathology

Herpes infections can be encountered either in anatomic or clinical pathology. Correlation between clinical, histologic, and laboratory findings is often necessary to synthesize a complete understanding of the disease process (*Chisholm & Lopez, 2011*).

After initial infection, all eight types of human infecting herpesviruses remain latent within specific host cells and may subsequently reactivate. Clinical syndromes due to primary infection can vary significantly from those caused by reactivation of these viruses. Herpesviruses do not survive long outside a host; thus, transmission usually requires intimate contact. In people with latent infection, the virus can reactivate without causing symptoms; in such cases, asymptomatic shedding occurs and people can transmit infection. Despite the fact that the herpesviruses are genetically and structurally similar, they cause a wide array of generally non-overlapping clinical syndromes (*Kaye, 2021*).

❖ Herpes Simplex Virus 1 and Herpes Simplex Virus 2

Although separate species, these viruses cause similar histologic and clinical findings. Classic HSV-1 and HSV-2 clinical findings are described as painful grouped vesicles on an erythematous base, usually with admixed ulcerated and crusted lesions. A systemic prodrome of fever and malaise often coincides with the primary outbreak. Hypersensitivity to HSV is also one cause of the targetoid lesions of erythema multiforme. Abnormal presentations are not uncommon, particularly in the genital area and in the immunocompromised host, including verrucous papules, solitary ulcerations, and urinary retention. HSV-2 has been recently linked with the recurrence of atypical genital symptoms such as vulvar burning and itching, skin fissures, erythema, and repeat menstrual cycles with white vaginal discharge, often mistaken by the patient for a yeast infection.

❖ **Varicella Zoster Virus**

VZV affects humans in a bimodal distribution, occurring in children (chicken pox) and adults (herpes zoster, or shingles). The virus penetrates the respiratory mucosa and rapidly disseminates throughout the body, causing a generalized eruption of vesicles on erythematous bases. Infiltration of the vesicles with inflammatory cells leads to pustule formation, which is followed by ulceration and encrustation. Multiple crops of vesicles appear throughout the course of disease, leading to lesions in various stages of evolution. Fever and lymphadenopathy frequently accompany these cutaneous signs. VZV becomes latent in the spinal cord ganglia when the host immune response is sufficient.

About half of affected individuals will develop VZV reactivation by 80 years, especially in immunocompromised individuals, with extracutaneous symptoms causing significant morbidity. Herpes zoster is heralded by a sensation of tingling and burning in a unilateral dermatomal distribution accompanied by fever, chills, headache, stomachache, and/or malaise. A maculopapular rash follows 48 to 72 hours later, with subsequent vesicles and ulcer formation similar to those of primary VZV infection. Generalized herpes zoster does occur in a minority of patients, and this is accompanied by visceral involvement in 50% of cases.

❖ **Epstein-Barr virus**

EBV causes infectious mononucleosis and is associated with several benign and malignant conditions, including Burkitt lymphoma, nasopharyngeal carcinoma, posttransplant lymphoproliferative disorders, Kikuchi histiocytic necrotizing lymphadenitis, hydroa vacciniforme, Gianotti-Crosti syndrome, and oral hairy leukoplakia. Infectious mononucleosis presents with fever, pharyngitis, lymphadenopathy, and malaise. Hepatosplenomegaly may also be present. This infection is self-limited, and most individuals are affected by young adulthood. EBV becomes latent in B lymphocytes. Skin manifestations are uncommon and, when present, are usually found in immunocompromised individuals. A maculopapular or morbilliform eruption may be seen concurrently with acute EBV symptoms, but histopathology is nonspecific.

❖ **Human Cytomegalovirus**

HCMV is acquired by late childhood in the majority of individuals. The initial infection is nonspecific and rarely causes the patient to seek medical attention. Once the primary infection is cleared, the virus becomes latent in peripheral blood leukocytes. Reactivation occurs in immunosuppressed patients and is accompanied by fever, malaise, leukopenia, and the very uncommon appearance of a macular rash. HCMV infects blood vessel endothelium when reactivated in cutaneous tissue. The biopsy will show nonspecific inflammation, dilated vessels, and possible overlying ulceration. Three distinct patterns of vascular injury have been identified in skin biopsies: leukocytoclastic vasculitis, necrotizing lymphocytic vasculitis, and pauci-inflammatory endothelial cell injury with luminal thrombosis and vascular dropout. Although the leukocytoclastic vasculitis and necrotizing lymphocytic vasculitis

patterns present with palpable purpura, the pauci-inflammatory variant is more likely to present with acute ischemic organ injury and/or progressive cutaneous sclerosis.

❖ **Kaposi Sarcoma–Associated Herpesvirus**

KSHV is likely a relatively ubiquitous virus. The route of transmission is uncertain. Many children exhibit seropositivity, indicating an innocuous form of transmission. However, sexual transmission is also likely given that seropositive rates are higher among homosexual males and patients attending sexually transmitted disease centers. In addition, transmission from latent infections in organ grafts may be possible. KSHV is the causative organism of Kaposi sarcoma, which, in the United States, occurs most frequently in association with immunosuppression, primarily human immunodeficiency virus infection. In at-risk patients, purple macules or plaques mimicking purpura, hematomas, or small angiomas appear spontaneously on any portion of the body. Some lesions may behave indolently, whereas others may grow rapidly with subsequent ulceration. KSHV has tropism for endothelial cells, similar to cytomegalovirus, and its affinity for endothelium has led some authors to propose these cells as the site of latency. Kaposi sarcoma is categorized into patch, plaque, and nodular phases. Early lesions are characterized by a proliferation of small vessels surrounding more ectatic vessels, which may be mistaken for other angiomatous tumors.

1.1.8. Treatment

Drugs that have activity against herpesviruses include acyclovir, cidofovir, famciclovir, formivirsen, foscarnet, ganciclovir, idoxuridine, penciclovir, trifluridine, valacyclovir, valganciclovir and vidarabine (Table 3).

Drug	Activity	Uses	Adverse Effects
Acyclovir	Active against (in order of potency) HSV-1, HSV-2, VZV, and EBV. Minimal activity against HCMV	Oral or IV: IV indicated when a higher serum drug level is required	Oral: Infrequent IV: Rarely, renal toxicity due to precipitation of acyclovir crystals.
Cidofovir	In vitro inhibition of a broad spectrum of viruses: HSV-1, HSV-2, VZV, HCMV, EBV, KSHV	IV: Generally used for HCMV, but use limited by renal toxicity. Intravitreal injection: For HCMV retinitis	Significant renal toxicity
Famciclovir	Antiviral spectrum similar to acyclovir	Oral: As effective as acyclovir for HSV-1, HSV-2 and VZV	Infrequent
Fomivirsen	Potent activity	Intravitreal	Increased intraocular pressure,

	against HCMV (antisense oligonucleotide inhibits HCMV protein synthesis)	injection: For patients with human immunodeficiency virus (HIV) infection and HCMV retinitis	For corticosteroid-responsive uveitis
Foscarnet	Active against EBV, KSHV, HHV-6, HSV-1, HSV-2 VZV and HCMV. Some anti-HIV activity	IV or intravitreal injection: Efficacy similar to ganciclovir for treating and delaying progression of HCMV retinitis	Renal toxicity in up to one third of patients if foscarnet is given without adequate hydration, electrolyte imbalances
Ganciclovir	In vitro activity against all herpesviruses.	IV form: Most common Intravitreal injection: For HCMV retinitis Oral: Only 6 to 9% bioavailable	Primarily, bone marrow suppression, particularly neutropenia, which sometimes requires treatment
Idoxuridine	Active against HSV-1, HSV-2, VZV, vaccinia, and CMV	Topical: high systemic toxicity; limited to HSV keratoconjunctivitis	Irritation, pain, photophobia, pruritus, inflammation or edema of the eyelids. Rarely, allergic reactions
Letermovir	Active against HCMV	IV or oral: HCMV prophylaxis in bone marrow transplant patients	Nausea, vomiting, diarrhea, peripheral edema, cough, headache, fatigue, abdominal pain
Penciclovir	Active against HSV-1, HSV-2, VZV, and EBV	Topical (cream): Used to treat recurrent herpes labialis in adults	Erythema
Trifluridine	Active against HSV-1 and HSV-2	Topical: recurrent keratitis or ulceration caused by HSV-1 and HSV-2	Ocular stinging, palpebral edema. Less commonly, punctate keratitis, allergic reactions
Valacyclovir	Antiviral spectrum similar to that of acyclovir	Oral: 3–5 times more bioavailable than acyclovir	Similar to those of acyclovir.
Valganciclovir	Similar to ganciclovir	Oral: More bioavailable than oral ganciclovir	Similar to ganciclovir
Vidarabine	For HSV infections	IV form not used anymore because of neurotoxicity	Superficial punctate keratitis with tearing, irritation, pain, and photophobia

Table 3. Drugs used to treat Herpesvirus infections. IV (intravenous) (Kaye, 2021).

1.2. Herpesvirus DNA encapsidation

Herpesviruses have a linear double-stranded DNA which is replicated by a rolling circle mechanism. Circularisation of the herpesvirus genome occurs early during infection and these circularised molecules act as templates for DNA replication. The viral DNA replicates according to an origin-dependent theta mechanism, in which circular templates are amplified. This step is followed by a rolling circle-based mode of replication that produces concatemers of the genome in head-to-tail fashion, which further act as substrates for the DNA-packaging process. A viral protein complex called terminase cleaves concatemeric DNA into unit-length genomes for DNA packaging inside the procapsid. This involves site-specific cleavage at adenine or thymine (AT)-rich core sequences within pac motifs ('cis-acting packaging signal') located in the 'a' sequence of the terminal and internal repeat segments (Figure 5) (Ligat et al, 2018).

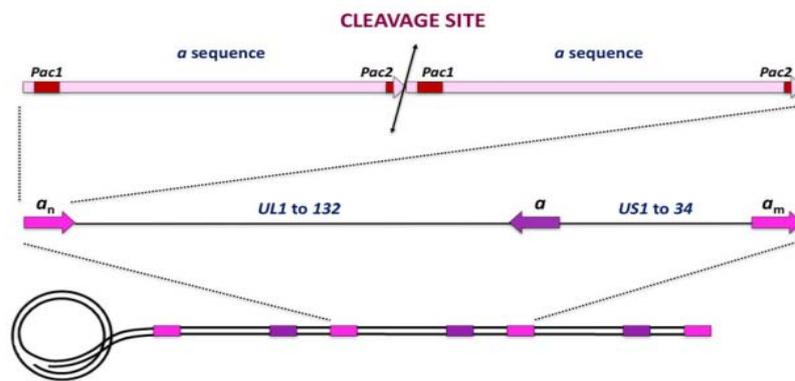


Figure 5. From full genome to cleavage site. The genome is organised as two regions. The unique long (U_L) and the unique short (U_S) segments are flanked by repeated sequences that contain the <<a>> sequence. The pac1 and pac2 sequences are present in each <<a>> sequence (Ligat et al, 2018).

The process involves specific cleavage of viral DNA at the <<a>> sequence. This is a short DNA sequence located at both termini of the human herpesvirus genomes and repeated in an inverted orientation. The <<a>> sequence contains two conserved motifs, pac1 and pac2, and both are required for cleaving the viral DNA. The pac1 motif consists of an AT-rich region of eight bases flanked by a GC-rich sequence. The pac2 motif consists of an AT-rich region of seven bases and it is associated with the nearby CGCGGCG sequence (Figure 6) (Baines & Weller, 2005).

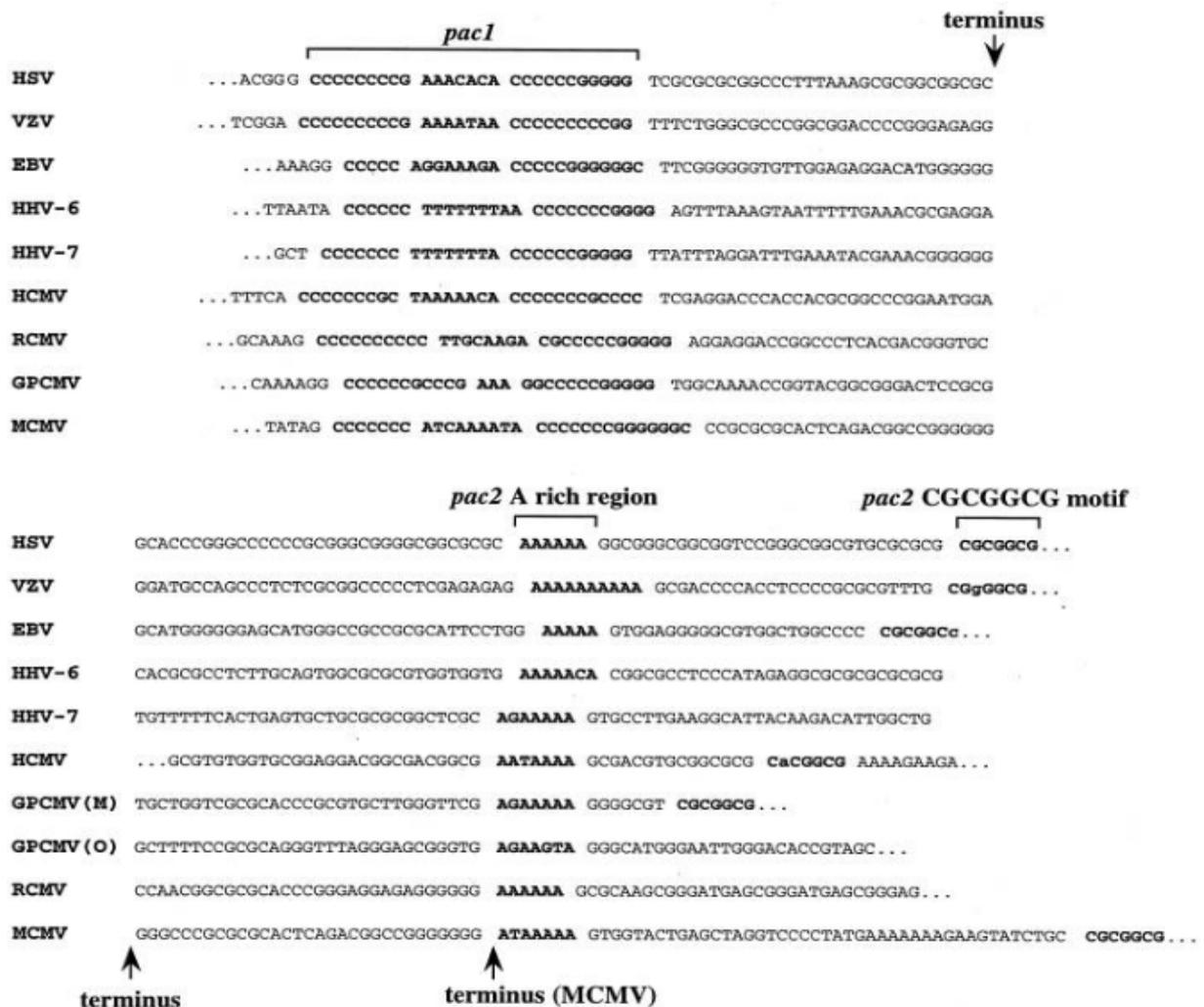


Figure 6. Conserved sequences at herpesvirus termini. Alignments of the terminal sequences of HSV-1, VZV, EBV, HCMV, HHV-6, HHV-7, RCMV, GPCM, and MCMV. Genomic termini are indicated, with the exception of the HCMV *pac2*-containing terminus, which has an additional 97 bp between the left end of the sequence and the terminus. The conserved components of *pac1* and *pac2* are shown in boldface and set off by spaces (McVoy et al., 1998).

The encapsidation mechanism is similar to that of the DNA bacteriophages. Seven highly conserved genes within the human herpesviruses have been shown to be essential in the DNA packaging process (table 4). Deletion of any of these genes entails accumulation of uncleaved DNA and empty capsids in the nucleus.

HSV-1	VZV	HCMV	EBV	KSHV	FUNCTION
UL6	ORF54	UL104	BBRF1	ORF43	Portal protein
UL28	ORF30	UL56	BALF3 TRM1	ORF7	Terminase subunit 1
TRM1	TRM1	TRM1		TRM1	
UL33	ORF25	UL51	BFRF1A TRM2	ORF67A	Terminase subunit 2
TRM2	TRM2	TRM2		TRM2	
UL15	ORF42	UL89	BGRF1/BDRF1	ORF29	Terminase subunit 3
TRM3	TRM3	TRM3	TRM3	TRM3	
UL17	ORF43	UL93	BGLF1 CVC1	ORF32	Tegument protein, role in
CVC1	CVC1	CVC1		CVC1	encapsidation proteins and capsids to replication

					compartments
UL25	ORF34	UL77	BVRF1	CVC2	Role in stabilizing DNA-filled capsid
CVC2	CVC2	CVC2		CVC2	
UL32	ORF26	UL52	BFLF1	ORF68	Role in encapsidation proteins and capsids to replication compartments

Table 4. *Herpesvirus DNA packaging proteins. Data from Vasalli & van Zeijl 2003.*

After the terminase complex is translocated into the nucleus, the process of viral DNA packaging includes six steps (Figure 7):

- ❖ Specific binding of the terminase complex to the pac site of the concatemeric DNA and binding of the DNA-terminase complex at the unique portal vertex of an empty preformed procapsid.
- ❖ Cleavage of the duplex, with a first specific cut that generates a free end at which further packaging is initiated.
- ❖ Exertion of ATPase activity to power the translocation of a unit length DNA genome into the capsid through the portal vertex.
- ❖ Cutting off of the excess DNA at the portal region, leading to C capsids (viral DNA containing capsids).
- ❖ Dissociation of the DNA/terminase complex from the filled capsid.

The DNA/terminase complex is now ready for the next packaging step.

Three types of capsids, A (empty), B (intermediate), and C (full), are found in the nucleus of infected cells and have the same shell structure. The three differ in the content of the capsid cavity: C capsids contain the viral DNA and are able to mature into infectious virus, B capsids lack DNA but are filled with the scaffolding protein and A capsids lack both DNA and protein. A and B capsids are considered to be developmental dead ends. They are distinguishable morphologically in electron micrographs, and can be separated from each other by sucrose density gradient ultracentrifugation (*Brown & Newcomb, 2011*).

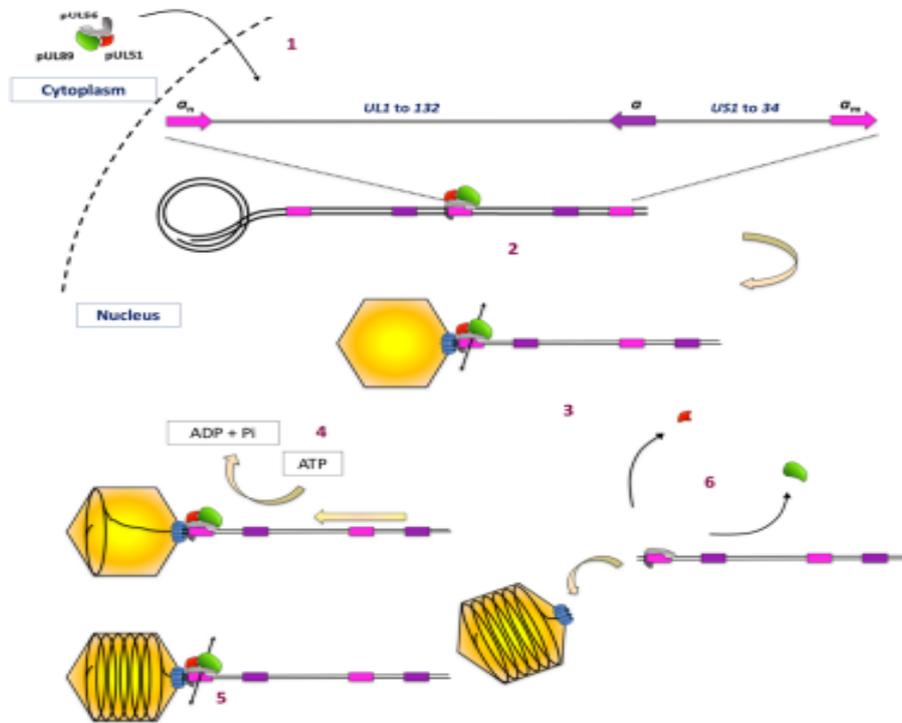


Figure 7. Genome cleavage/packaging and the HCMV terminase complex. (i) Translocation of the terminase complex into the nucleus, (ii) HCMV terminase specifically binds the *pac* site and recruits the empty capsid, (iii) cleaves the duplex, (iv) exerts its ATPase activity to power translocation of a unit-length DNA genome into the capsid and (v) completes the DNA-packaging process by cutting off excess DNA at the portal region. (vi) Finally, the DNA-terminase complex dissociates from the filled capsid and is ready for next DNA-packaging step (Ligat et al, 2018).

1.2.1. Herpesvirus terminase complex

Herpesvirus replication strategy is very similar to that of tailed bacteriophages. In bacteriophages, the terminase complex is formed by a small and a large terminase subunit assembled in various stoichiometries that dock at the portal vertex and convert ATP hydrolysis into rotation of dsDNA, which is gradually encapsidated. Herpesviruses encode two proteins equivalent to the large and small terminase subunits, but also express a smaller third terminase subunit. In HSV-1 a structure of a hexameric ring terminase complex was recently solved by cryo-electron microscopy (Yang et al. 2020). Each of the six subunits comprises three different components (Figure 8). Therefore, each subunit has a total molecular weight of 181 kDa and is formed by a heterotrimer of three terminase subunits (TRM): TRM1 (86kDa), TRM2 (14 kDa) and TRM3 (81 kDa).

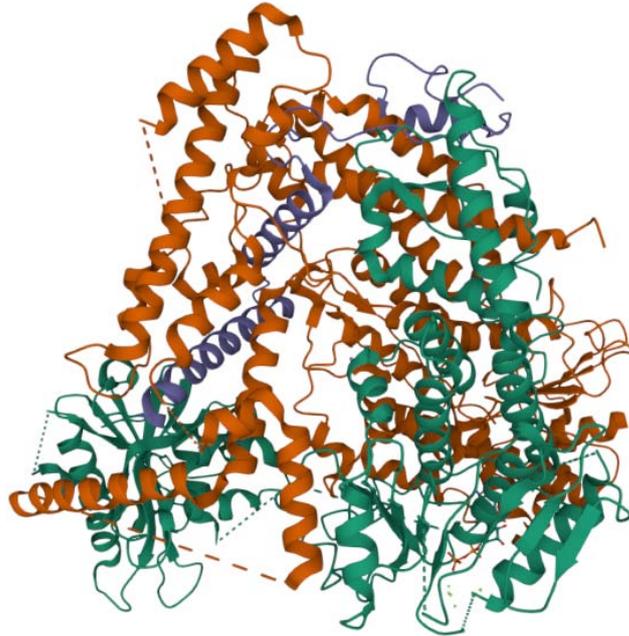


Figure 8. Structure of the tripartite terminase complex of HSV-1 strain 17 by electron microscopy at 3.80 Å (Yang et al. 2020).

This structure shows six ATPase domains that form a central channel with conserved basic-patches conducive to DNA binding and trans-acting arginine fingers, which are essential to ATP hydrolysis and sequential DNA translocation. Indeed, the structure of the heterotrimer shows a number of different interaction interfaces between the three terminase subunits.

1.2.1.1. Tripartite Terminase Subunit 1 (TRM1)

The larger subunit of the herpesvirus terminase complex, TRM1, is composed of 12 conserved regions and is located on the unique long portion of the viral genome. Three-dimensional reconstruction by electronic cryomicroscopy suggests that TRM1, when expressed alone, exists as a dimer formed by two ring-shaped structures connected to each other by a bridge to their base (Savva, Holzenburg and Bogner 2004). Numerous in vitro studies confirmed the activity of TRM1. Electrophoretic mobility shift assay demonstrated a sequence-specific interaction of TRM1 with pac motifs within 'a' sequences of the viral genome. Electron microscopy studies have shown that TRM1 binds linearized DNA and this protein-DNA complex, visible under the microscope, remains stable. However, upon addition of TRM3 the same DNA is cut into smaller fragments, indicating that these proteins mediate a concerted two-steps reaction of DNA recognition and cleavage (Scheffczik et al. 2002). Some studies indicate that TRM1 has ATP-dependent endonuclease activity that seems to be pac specific. Moreover, TRM1 could enhance the endonuclease activity driven by TRM3 (Scheffczik et al. 2002). TRM1 also interacts with the viral portal protein during DNA-packaging via its C-terminal part. The TRM1 ATPase activity is enhanced by up to 30% when it is associated with TRM3 (Hwang and Bogner 2002). TRM1 protein has two putative ATP-binding sites.

It has been reported that the C-terminal part of TRM1 interacts with TRM3 (Hwang & Bogner, 2002). The TRM1 protein is translocated into the nucleus by means of its own nuclear localization signal. An electron microscopy study revealed that TRM1 displays

a toroidal structure. Another study revealed that each ring measures about 9 nm in diameter and 2.5 nm in height with a central hole of approximately 3.5 by 2.5 nm across (*Savva et al, 2004*).

1.2.1.2. Tripartite Terminase Subunit 2 (TRM2)

TRM2 interacts with TRM1 and enhances the TRM1-TRM3 interaction. Extensive mutagenic analyses previously identified regions of TRM2 important for DNA packaging and the generation of L component termini. Several novel mutations were identified in a conserved region of TRM2 with a positive charge that do not prevent the cleavage of concatemeric DNA but preclude the release of monomeric genomes. TRM2 subunit is necessary for one of the two DNA cleavage events necessary to release unit-length genomes from concatemeric DNA.

This protein is found associated with the external surface of the viral capsid during assembly and DNA packaging, but seems absent in extracellular mature virions (*Yang et al. 2017*).

1.2.1.3. Tripartite Terminase Subunit 3 (TRM3)

TRM3 is codified on the long unit of the viral genome. Its gene consists of two exons separated by an intron. The N-terminal domain is predicted to bear the ATPase activity while the C-terminal domain would bind and cut the DNA (*Champier et al, 2007*).

In vitro translation and eukaryotic expression demonstrated that TRM3 is a protein in monomeric form. Previous studies had shown similarities in the amino acid sequence of TRM3 and the terminase subunit gp17 of phage T4. Because of the strong homology of part of TRM3 to the ATP binding motif of the bacteriophage T4 gp17 subunit, the possible role of TRM3 in DNA packaging was investigated (*Hwang and Bogner 2002*). Subsequently, in silico studies focusing on the amino acid sequence of TRM3 have highlighted the four motifs involved in the ATPase centre domains located in N-terminal part of TRM3: the adenine binding site, the Walker A box or motif I, the Walker B box or motif II and motif III (*Champier et al. 2007*). These motifs have also been identified in the terminase subunit of the bacteriophage T4 protein gp17 (*Mitchell et al. 2002*).

Despite its partial homology with the terminase subunit of T4 gp17, TRM3 did not exhibit enzymatic ATPase activity. The interaction between TRM1 and TRM3 likely takes place in the cytoplasm, after which the terminase proteins are translocated to the nucleus. Two putative NLS have been proposed to catalyse the nuclear translocation of TRM3 (*Champier et al. 2007*). Recent findings indicate that the TRM3 subunit translocates to the nucleus only in presence of TRM1 and TRM2, and otherwise remains exclusively in the cytoplasm (*Neuber et al. 2017*). It has been suggested that TRM1 is unable to exert specific cleavage by itself and that, once again, synergy with TRM3 is necessary to complete the cleavage steps of the DNA-packaging process during herpesvirus replication (*Scheffczik et al. 2002*). Structural data indicate that the TRM3 C-terminal domain (TRM3-C) of HCMV belongs to the RNase

H-like superfamily of nuclease and polynucleotidyl transferases. Indeed, it has the characteristic fold of this superfamily, and three conserved acidic residues (Asp463, Glu534 and Asp651) coordinating two Mn^{2+} cations (Figure 9) (Nadal et al. 2010).

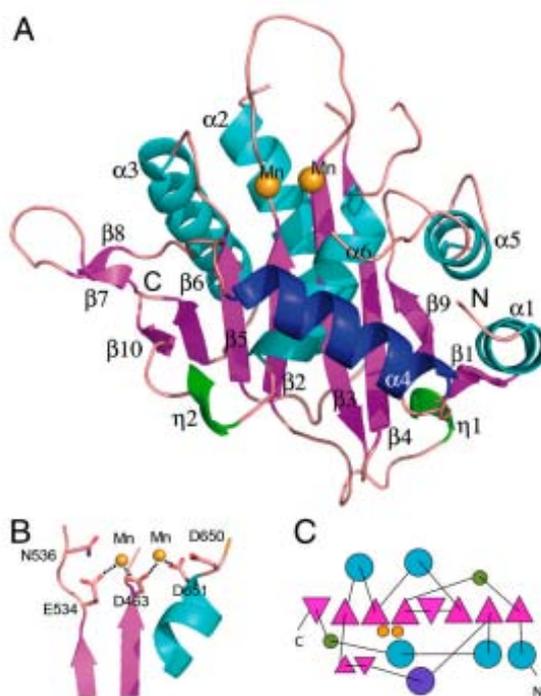


Figure 9. (A) Overall structure of HCMV TRM3-C. The metal ions are indicated by yellow spheres. UL56-interacting helix $\alpha 4$ is highlighted in blue, α helices in cyan, 3_{10} helices in green, and β strands in magenta. (B) Detailed view of the active site area. One Mn^{2+} ion is shown coordinated by Asp463 and Glu534 and a second one is coordinated by Asp463 and Glu651. (C) Topology diagram of UL89 (Nadal et al., 2010).

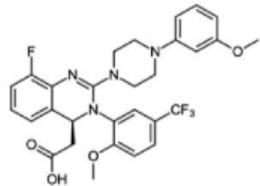
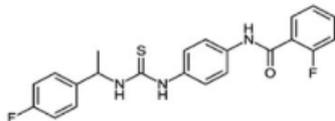
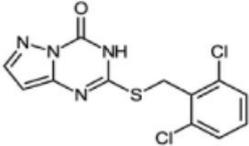
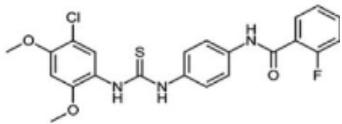
1.2.2. DNA encapsidation as a therapeutic target

Although there are effective nucleoside analogs to treat HSV, VZV, and HCMV disease, herpesvirus infections continue to contribute to significant morbidity and mortality. No successful chemotherapeutic options are available for EBV, HHV-6, 7, and KSHV. Drug resistance is a concern for HCMV, HSV, and VZV since approved drugs share common mechanisms of action. Targeting DNA encapsidation provide additional options for the development small molecule anti-herpesviral drugs. Indeed, since DNA packaging is one of the first processes during viral maturation, a terminase inhibitor would represent a blockage at an early stage. Furthermore, inhibitors of the terminase complex are attractive therapeutic targets since the cleavage of concatemeric DNA does not take place in mammalian cells. Thus, drugs targeted to the terminase proteins might be safe and highly selective (Keil et al., 2020).

Letemovir was approved for prophylaxis in stem cell transplantation validating the HCMV terminase complex as an alternative antiviral target, complementing the viral DNA polymerase, which had been the sole target for available systemic HCMV antivirals (Table 5). This development mitigates cross-resistance and drug-specific toxicities while potentially improving antiviral options by use of combination therapy. In vitro studies identified letemovir resistance mutations mapping

primarily to a region between codons 229 and 369 of the TRM1 gene and rarely to the TRM3 and TRM2 genes of the HCMV terminase complex. Preliminary data from clinical use of letermovir includes detection of TRM1 amino acid substitutions Val236Met and Cys325Trp in treated individuals, consistent with the preferential localization of mutations observed *in vitro* (Chou *et al.* 2018).

Raltegravir HIV integrase inhibitor presents also inhibition properties on HCMV TRM3-C due to the structural similarity between both proteins (Nadal *et al.*, 2010). Raltegravir (MK0518), which was approved by the Food and Drug Administration (FDA) in 2007 for the treatment of Acquired Immune Deficiency Syndrome (AIDS) turned out to be a strong inhibitor of the nuclease activity of HCMV TRM3-C. A recent structure of the prototype foamy virus integrase in complex with DNA and the inhibitor shows how raltegravir binds at the integrase active site, directly coordinating the metal ions (Hare & Cherepanov, 2010).

Drug or Compound	Virus	Mechanism of action/target	Structure
Clinically approved Letermovir	HCMV	Quinazoline derivative, TRM2	
Preclinical: <i>in vivo</i> α -methylbenzyl thiourea derivatives (Comp 1, 2 and 3)	VZV	DNA encapsidation inhibitor, portal protein	
Preclinical: <i>in vitro</i> Pyrazolo derivatives	VZV	Capsid assembly inhibitor, major capsid protein	
WAY: 150183	HSV-1, HSV-2	DNA encapsidation inhibitor, portal protein	

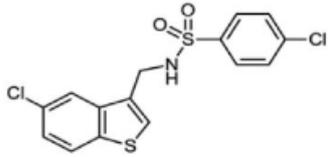
Chlorobenzothiopen derivatives (45B5)	VZV	DNA encapsidation inhibitor, portal protein	
---------------------------------------	-----	---------------------------------------------	-------------------------------------------------------------------------------------

Table 5. Drugs and compounds targeting herpesviral DNA encapsidation and viral capsid assembly (Keil et al. 2020).

The drugs licensed so far for the treatment of HCMV infection are the antisense oligonucleotide fomivirsen and four viral DNA polymerase inhibitors: ganciclovir (GCV), valganciclovir, cidofovir (CDV), and foscarnet (FOS). There are several disadvantages associated with the use of these drugs including poor bioavailability, serious side effects induced by prolonged treatment, and emergence of resistant viral strains. A promising strategy is to block the HCMV terminase machinery which mediates DNA binding, cleavage, and packaging because these processes do not occur in mammalian cells and therefore specific terminase inhibitors could have fewer undesirable side effects.

A rational design of TRM3-C HCMV inhibitors (Figure 10) by simplifying the structure of a HIV integrase inhibitor and selecting chemical features important for TRM3-C HCMV binding was done. The three coplanar oxygen atoms of the diketo acid (DKA) moiety chelate Mn^{2+} ions in the active site, explaining the requirement for metal ions for drug binding. The central oxygen, O(H), bridges both metal ions, whereas two O (carbonyl) coordinates the metal ions. On this basis, BS2 was designed, containing an α,γ -DKA motif linked to a 4-fluorophenyl ring (Figure 10).

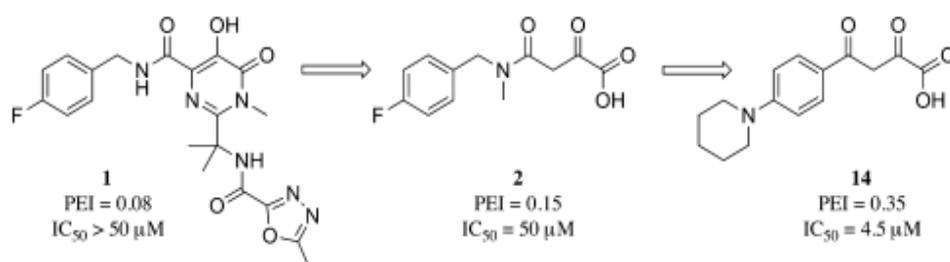


Figure 10. Rational design of BS2 and BS14 as the TRM3-C HCMV inhibitor (Bongarzone et al. 2018).

Due to the elucidation of the binding mode of BS2 in the TRM3-C HCMV binding site by X-ray crystallography (Figure 11) the α,γ -DKA series was enlarged. BS14 inhibitor 14 was able to inhibit TRM3-C HCMV in vitro in the low micromolar range, making BS14 (Figure 10) an optimal candidate for further development and virus-infected cell assay. BS17 inhibitor was also able to inhibit TRM3-C HCMV in vitro in the low micromolar range (Figure 12) (Bongarzone et al. 2018).



Figure 11. X-ray crystal structure of BS2 (6EY7) bound to TRM3-C HCMV. Mn²⁺ atoms are shown as purple spheres (Bongarzone et al. 2018).

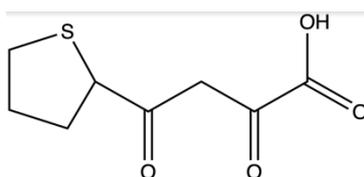


Figure 12. Formula of BS17.

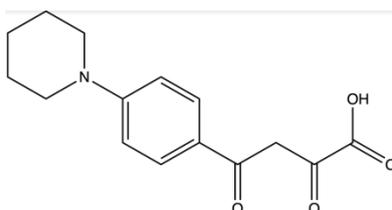


Figure 13. Formula of BS14.

The catalytic site of the HIV integrase is contained within an RNase H-like fold, and numerous drugs have been developed that bind to this site and inhibit its activity. As the *Herpesviridae* molecular motor encodes the TRM3 protein with RNase H-like fold which is essential for viral DNA cleavage and packaging, HIV integrase inhibitors might also inhibit herpesvirus replication by targeting the terminase. Some potent HIV integrase inhibitors that have been approved or are in advanced investigation are Bictegravir (Figure 14), Elvitegravir (Figure 15), Dolutegravir (Figure 16), Cabotegravir (Figure 17), 707035 (Figure 18) and MK-2048 (Figure 19) (Bongarzone et al. 2018).

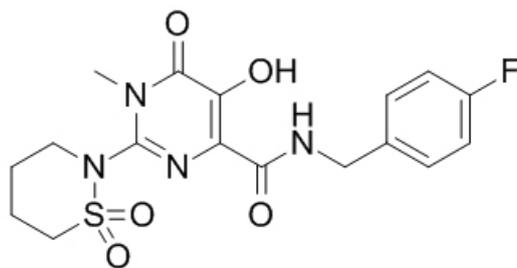


Figure 18. Formula of 707035 inhibitor.

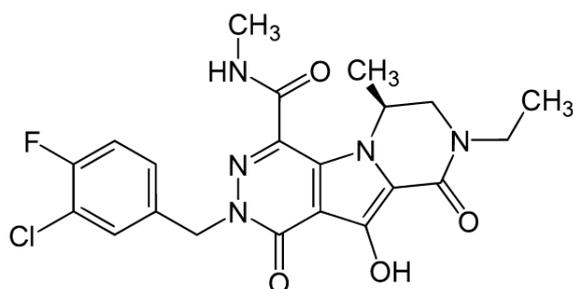


Figure 19. Formula of MK-2048 inhibitor.

Structures of the integrase in complex with DNA and some of these drugs showed that the compounds displaced the viral DNA end from the active site, disarming the viral nucleoprotein complex. Moreover, the compounds are bound to the two magnesium atoms in the active site and interact both with the protein and the DNA (*Hare et al. 2010, Cook et al. 2020*)

CHAPTER 2: OBJECTIVES

2. OBJECTIVES

The global aim of this part of the thesis was to structurally and functionally characterize the *Herpesviridae* family TRM3-C terminal domain. Moreover, in order to block the infections generated by this family of viruses, we intended to identify specific drugs against the TRM3 C-terminal domain that could inhibit the viral DNA packaging process. Finally, the design of novel inhibitors using a structure-based approach was aimed.

The following objectives were established:

- I. Prepare soluble samples of the TRM3-C domains of HSV-2, VZV, EBV, HCMV and KSHV at a high level of purity that allows further functional and structural studies.
- II. Characterize the nuclease activity of the domains in the presence of different metal ions.
- III. Obtain high-quality crystals and determine the three-dimensional structure of the TRM3-C terminal domains of HSV-2, VZV, EBV and KSHV.
- IV. Perform nuclease inhibition assays against TRM3-C of HSV-2, VZV, EBV, HCMV and KSHV with different commercial drugs (MK-2048, 707035, Cabotegravir, Dolutegravir, Elvitegravir and Bictegravir) and optimized drugs developed in the laboratory (BS14 and BS17).
- V. Determine the IC₅₀ of each inhibitor with the different protein domains that were tested.
- VI. Use the structures that could be solved as well as the HCMV model previously determined to solve the structure of protein-compound complexes.
- VII. Considering all the biochemical and structural data, design new compounds to obtain better inhibitors against the TRM3-C target.

CHAPTER 3: MATERIALS AND METHODS

3.1. MATERIALS

3.1.1. Chemicals and reagents

General laboratory chemicals were purchased from BioRad, Clontech, Fermentas, Invitrogen, Merck, New England Biolabs, Roche, Sigma and Condalab.

Buffers were prepared as aqueous solutions in distilled water. Solutions were sterilized by filtration (Millipore, 0.22 μ M) and degassed.

Sequence optimized genes were purchased from GeneArt (ThermoFisher) and the corresponding DNA primers were obtained in lyophilized form from Merck.

Inhibitors for soaking experiments and biochemical assays were purchased from Merck, Selleckchem and AdooQ Bioscience.

3.1.2. Oligonucleotides

Synthetic oligonucleotide primers for use in polymerase chain reaction amplification and sequencing of TRM3-C terminal domain of EBV are described in the following table (Table 6).

Protein	Usage	Primers	Sequence (5' \rightarrow 3')
EBV	Amplification	Forward	CTTTAGCATATGAGCCAGGGTAGCC
		Reverse	CTCGCGGGATCCTTAATGAACTTTCTTGAACGT
EBV	Sequencing	T7 promoter	TAATACGACTCACTATAGGG
		T7 terminator	TACGACGACTCACTATACCC

Table 6. Oligonucleotides sequences.

Synthetic genes of TRM3-C terminal domain of HSV-2, VZV and KSHV with the optimized codon were bought. The plasmid of TRM3-C terminal domain of HCMV was already in the lab.

3.1.3. Plasmid

Linearized pET28a vector (Figure 20) was obtained from our lab. pET-28a (+) vector (5369bp) carries an N-terminal His-Tag®/thrombin/T7-Tag® configuration plus an optional C-terminal His-Tag sequence. The sequence is numbered by the pBR322 convention, so the T7 expression region is reverse on the circular map. pET28a has bacterial resistance to Kanamycin.

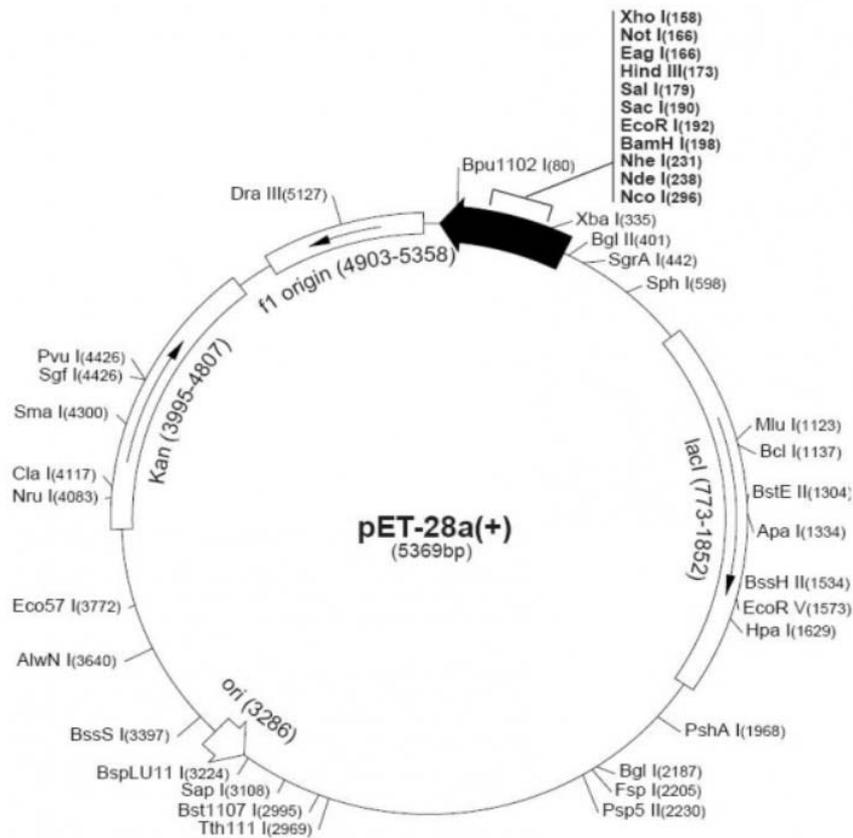


Figure 20. pET-28a (+) vector map.

3.1.4. Bacterial strains

The bacterial cells from all studies are *Escherichia coli* strains. DH5 α -T1R strain was used as a cloning host whereas Rosetta 2 (DE3) cells were used for protein expression (Table 7).

Strain	Genotype	Description
DH5α-T1R	F- ϕ 80lacZ Δ M15 Δ (lacZYAargF)U169 recA1 endA1 hsdR17(rk - , mk +) phoA supE44 thi-1 gyrA96 relA1 λ tonA	Subcloning into plasmid vectors
Rosetta 2 (DE3)	F-ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (ChIR)	Suitable for protein expression

Table 7. Cell strains used in cloning and protein overexpression.

3.1.5. Media

Luria-Bertani (LB) medium (Sambrook et al., 1989): 1% (w/v) tryptone (Condalab), 1% (w/v) NaCl (Merck), 0.5% (w/v) yeast extract (Condalab). Sterilised by autoclaving.

LB agar plates: Luria-Bertani (LB) agar (1.5% w/v). Agar sterilized by autoclaving and poured into the plates after adding the corresponding antibiotic when the LB agar temperature drops to 50°C.

3.1.6. Crystallization

The main materials used for the crystallization of EBV TRM3-C were the following:

3.1.6.1. Crystallization plates

- ❖ CrystalQuick 96 well plates (Hampton Research): Plates for sitting drop vapour diffusion crystallization. They are suited for high throughput crystallization and are fully compatible with robotic equipment.
- ❖ Cryschem plates (Hampton Research): Plates for crystallization by sitting drop and hanging drop with 24 wells.

3.1.6.2. Cryo-crystallography loops

- ❖ Mounted CryoLoop (Molecular Dimensions): It is a nylon loop used to mount, freeze, and secure the crystal during cryo-crystallographic procedures and X-ray data collection.
- ❖ CrystalCap HP (Molecular Dimensions): It is a complete crystal vial for cryocrystallography. It has a magnetic base and cap. It also has a bar code on cap.

3.1.6.3. Robotic platforms

- ❖ Cartesian (Genomic Solutions): Liquid handling robot specially designed to perform crystallization assays.
- ❖ Crystal farm (Bruker AXS): Instrument for protein crystallization storage and automated imaging.
- ❖ Phoenix crystallization robot (Art Robbins Instruments): protein crystallography dispenser, featuring accuracy, speed and precision. Instrument for sitting drop, hanging drop and microbatch reactions.

3.1.6.4. Software

- ❖ XDS: Software for processing crystal X-ray diffraction data (*Kabsch, 1988*).
- ❖ CCP4i: It is a graphical user interface that allows running programs from the CCP4 suite (*Potterton et al, 2003*).
- ❖ Refmac: Software for macromolecular refinement (*Murshudov et al, 1997*).
- ❖ Coot: Software for molecular graphics applications (*Emsley & Cowtan, 2004*).
- ❖ Phenix: Software package for macromolecular structure determination using crystallographic (X-ray, neutron and electron) and electron cryo-microscopy data (*Adams et al., 2002*).

3.2. METHODS

3.2.1. Sample preparation and analysis

Protein engineering is a powerful tool for improving protein physiochemical properties leading to proteins that are more stable and soluble to have a greater propensity to crystallize.

3.2.2. Construct design

The first step was the design of protein constructs containing the TRM3-C terminal domain of HSV-2, VZV, EBV and KSHV with a high chance to be expressed in a soluble manner. Protein alignment with TRM3 from HCMV was done using the Clustal Omega server in order to design equivalent constructs to TRM3-C HCMV. It is important to take into account that TRM3 codifying genes consist of 2 exons separated by an intron. The designed constructs were entirely codified in the second exon.

Prediction of secondary structure elements, identification of protein domains and disordered regions from aminoacid sequence using PSIPRED (Jones, 1999), TMpred (Hofmann & Stoffel, 1993) and GlobPlot (Linding et al., 2003) were also taken into account to design the TRM3-C domain constructs.

3.2.3. Cloning

Expression vectors corresponding to the HCMV and KSHV TRM3-C constructs were already available in the lab. In the case of HSV-2, VZV and HCMV mutant synthesis and subcloning of equivalent TRM3-C constructs was ordered (GeneArt, ThermoFisher).

Genomic DNA of EBV was ordered from National Collection of Pathogenic Viruses. The EBV TRM3-C domain was inserted into the pET-28a (+) vector (5369bp), which contains N-terminal His-Tag®/thrombin/T7-Tag® configuration plus an optional C-terminal His-Tag sequence.

The primers were designed with 15 bp extensions (5') complementary to the ends of the linearized vector and between 18-25 bases specific of the target gene in the 3' end. Amplification of the gene was performed with PfuUltraII Fusion polymerase (Agilent technologies, Inc.) according to manufacturer's instructions and the PCR product was verified on an agarose gel.

Agarose was suspended in 1x TAE at 1% (w/v) concentration and boiled. DNA samples were mixed with 6x DNA Loading Dye (ThermoFisher) and loaded into the agarose gel wells. Constant voltage was applied and then the gel was submerged with 1x TAE supplemented with SYBR-Safe (ThermoFisher) for 30 min. 1 kb/100 pb Plus DNA Ladder (Invitrogen) was used for sizing the DNA. The resulted gel was visualized on a UV transilluminator (Syngene).

The digestion of both, insert and vector, was carried out with *BamHI* and *NdeI* (Fermentas) restriction enzymes. The protocol of Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) was followed to purify the PCR product. With the

digested vector pET28a (+), an electrophoresis gel was ran and the band of biggest size was cut and purified using the QIAquick Gel Extraction Kit 250 (QIAGEN). DNA concentration and purity were checked with a Nanodrop 1000 spectrometer according to manufacturer's instructions (ThermoFisher). The ligation reaction was set up and incubated for 10 hours at 16°C. This temperature is optimal for the T4 DNA Ligase of *E. coli* that was used (Fermentas).

3.2.3.1. Bacterial transformation

Heat shock transformation of *E. coli* DH5- α competent cells prepared for α -complementarity and cDNA cloning was performed. After 30 minutes of incubation with 10 μ l of the ligation reaction on ice, heat shock of 45 seconds at 42°C was applied using a water bath. Immediately after, the mixture was placed back on ice during 2 minutes to reduce damage of cells. 1 mL of LB was added and the sample was incubated for 1 hour at 37°C with 220 rpm of agitation.

As the DNA came from a ligation and to maximize the probability of success in the transformation, cells were concentrated before being plated on LB-agar supplemented with kanamycin (Kan), for resistance selection, and kept overnight at 37°C. The culture plates were examined the next day for colony formation.

3.2.3.2. Plasmid extraction

Formed colonies were pre-cultured in 4 ml of LB medium with 50 μ g/ml of Kan and incubated overnight at 37°C with 220 rpm of agitation. 1200 μ L of each culture were stocked with 15% glycerol at -80°C. The rest of the liquid culture was used for isolating the plasmidic DNA following manufacturer's instructions of the QIAprep Miniprep Kit (QIAGEN). Briefly, the bacterial culture was harvested and lysed by high alkaline conditions. The plasmid DNA was eluted with nuclease free water.

3.2.3.3. Plasmid analysis

Clones were checked by PCR (Taq polymerase) and agarose gel electrophoresis. The absence of possible mutations was analyzed by DNA sequencing (Macrogen) using the T7 promoter and T7 terminator.

3.2.4. Protein expression

E. coli is the most popular host for the production of recombinant proteins since it produces large quantities of protein quickly and it can grow on economic media under well defined conditions and has a very short doubling time.

3.2.4.1. Large-scale protein expression

The Miniprep kit (QIAGEN) was used for extracting the expression plasmid from the DH5- α cells, which was subsequently transformed into Rosetta 2 (DE3) host cells. Colonies were picked up on the plate and transferred into 4 ml of LB medium

supplemented with of antibiotic (Chloramphenicol (Chl): 34 µg/ml; Kan: 50 µg/ml) for resistance selection of the cell line. The precultures were incubated overnight at 37°C with 220 rpm of agitation.

The IPTG induction protocol was performed. The optical density of each pre-culture was measured so as to determine the volume of inoculum that was needed for each 500 ml culture. The following calculation was used for obtaining an initial O.D. of 0.05:

$$V_i \times C_i = V_f \times C_f$$

$$V_i \times O.D. = 500 \text{ ml} \times 0.05$$

Inoculated 500 ml cultures were incubated at 37°C with 220 rpm of agitation until acquiring an optical density of 0.6. At this point, they were induced with 1 mM of IPTG and incubated overnight with 220 rpm of agitation at 16°C in order to improve protein folding. The samples extracted before and after induction (T₀ and T₁) were analyzed by SDS-PAGE electrophoresis

3.2.4.2. Protein electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed in order to verify protein overexpression. Proteins are separated by electrophoresis using a discontinuous polyacrylamide gel as a support medium and sodium dodecyl sulfate (SDS) to denature them (Table 8). All protein electrophoresis were conducted on a BioRad system.

System	Separating gel	Stacking gel	Running buffer
TrisGlycine (15%)	30% (w/v) acrylamide, 1.5 M Tris pH 8.8, 10% PSA, 10% SDS, 0.05% (v/v) TEMED	30% (w/v) acrylamide, 1.5 M Tris pH 6.8, 10% PSA, 10% SDS, 0.125% (v/v) TEMED	0.25 M Tris 0.2M Glycine 0.1% (w/v) SDS

Table 8. Recipe for SDS-PAGE gels.

The samples were dissolved in 5x Loading buffer containing bromophenol blue in 25 mM Tris-HCl (pH 6.8), 5% (w/v) SDS, 10% (v/v) glycerol and 5% (v/v) β-mercaptoethanol. Then, samples were boiled at 94 °C for 5 minutes. The SDS-PAGE gel was ran at 200V until the dye front reached the bottom of the gel. Afterwards, gels were stained with One Step Blue Protein Gel Stain (Biotium) and destained with water. PageRuler Prestained Protein Ladder (ThermoFisher) was used as size standards (10 to 180 kDa).

3.2.5. Protein purification

In X-ray crystallography, the purity of the protein is key for successful, reproducible crystallisation. Ideally a highly soluble and monodispersed protein should be available before starting crystallisation trials. If the protein sample is not pure, only

poorly diffracting crystals will be obtained, resulting in a low-resolution protein structure. Poor protein purity may also lead to the failure of protein crystals to form at all. The most common methods for preparative purification of proteins involve chromatography.

3.2.5.1. Cell lysis

After expression time, cells were collected by centrifuging the cultures during 20 minutes at 5000 g and 4 °C. Cell pellets were harvested and resuspended in the appropriate buffer (ProtParam values for the isoelectric point of the designed constructs were taken into account) and lysed to extract the proteins using a Cell disruptor (Constant Systems Ltd) from the Protein Expression Facility (IRB) under pressure of 25 kpsi. The soluble fraction was separated from the insoluble fraction by ultracentrifugation.

3.2.5.2. Chromatographic techniques

Chromatography is an important biophysical technique that enables the separation, identification, and purification of proteins for qualitative and quantitative analysis. A wide range of chromatographic procedures make use of differences in size and shape, binding capacity with the stationary phase, total charge, and other properties to purify proteins. This method is based on the principle that components of a mixture are separated when the mixture added to a mobile phase is moved through a stationary phase, resulting in some components of the mixture being attached to the stationary phase. At the same time, the rest is passed along with the mobile phase. Proteins in this work have been purified using two or three different chromatographies:

❖ Affinity chromatography

HisTrap HP is a nickel-charged immobilized metal ion affinity chromatography (IMAC) column (GE Healthcare) for high resolution his-tagged protein purification. The column is packed with Ni Sepharose High Performance affinity resin, which consists of highly cross-linked 6% agarose beads coupled to a chelating group. The chelating group is precharged with nickel, which selectively retains proteins with exposed histidine groups, allowing them to be purified from cellular contaminants or cell-free systems. Its performance characteristics include:

- Negligible leakage of the Ni²⁺ ion.
- Compatibility with a very wide range of reducing agents, detergents and other additives.
- Very high protein binding capacities.

❖ Ion exchange chromatography (IEX)

Ion exchange chromatography can separate molecules that have only slight differences in charge. Separation is based on the reversible interaction between a charged molecule and an oppositely charged chromatography medium. Conditions are selected to ensure that the molecules of interest bind to the medium as they are loaded onto the column, being then altered so that the bound substances are eluted

differentially. Mono Q is a strong anion exchanger pre-packed with MonoBeads in a Tricorn column that binds proteins with an isoelectric point below pH 7. It is an excellent choice for small scale polishing in purification of proteins when high purity is required and provides high resolution separations with greater loading capacity. The sample was filtered (0.22 μ M pore size) before loading it in a Mono Q (GE Healthcare) column.

❖ Size exclusion chromatography (SEC)

Size exclusion chromatography, also known as gel filtration chromatography, is a chromatographic method that separates molecules based on their particle size (actually particle's Stokes radius) as they pass through a SEC resin packed in a column (GE Healthcare). The resin consists of a porous matrix of spherical beads with specific size distribution that lack reactivity and adsorptive properties. Therefore, molecules are eluted in order of decreasing molecular weight and isocratically, so there is no need to use different buffers during the separation. The sample was centrifuged (16000g for 20 minutes) and filtered before loading it in a gel filtration column.

3.2.5.3. Protein concentration and quantification

Purified protein was concentrated using Amicon centrifugal filters (Millipore) of appropriate molecular weights cut-off (10K for 29 kDa protein) according to manufacturer's instruction. The Bradford protein quantification assay was performed for measuring the protein concentration by an absorbance spectrophotometer. The standard curve was prepared by plotting the absorbance at 595 nm of samples containing known concentrations (0, 1, 2, 4, 6 μ g/ml) of the standard protein bovine serum albumin (BSA). The protein sample absorbance at 595 nm was measured and its concentration was then interpolated onto the Standard curve formula.

3.2.6. Protein analysis

Several techniques for protein identification and characterization were employed.

3.2.6.1. Western blot

Western blotting, also called protein immunoblotting, enables to detect target proteins (as low as 1 ng in concentration) based on molecular weight due to high-resolution capacity of gel electrophoresis and strong sensitivity and specificity of the immunoassay. It was used to verify the production of the protein of interest after its expression. An electrophoresis gel was run and the gel was placed next to a nitrocellulose membrane. Electrical current (50 V overnight at 4°C) induced protein migration from the gel to the membrane where they adhered by using transfer buffer (25 mM Tris, 190 mM glycine, 20% methanol). The membrane was then blocked for 1 hour with 10% powdered milk in TBST (20 mM Tris pH 7.5, 150 mM NaCl, 0.1% Tween 20). Penta-his mouse primary antibody (dilution 1:1000) (Qiagen) diluted in TBST buffer with 5% of powdered milk was used for protein detection. The membrane was washed for 10 minutes with TBST buffer 3 times. Straight after, the blot was incubated for 45 minutes with anti-mouse IgG secondary antibody (dilution

1:1000) (Sigma-Aldrich) diluted in TBST buffer with 5% of powdered milk. The membrane was washed again, following the same procedure explained above. BenchMark His-tagged Protein Standard (Invitrogen) was used as positive control and for molecular weight sizing of the his-tagged fusion protein. Finally, membranes were imaged by means of the Odyssey®CLx Infrared Imaging System (Li-Cor Biosciences).

3.2.6.2. Mass spectrometry

Peptide mass fingerprinting (PMF) is a high throughput protein identification technique in which the mass of a protein can be determined. PMF is always performed with Matrix-assisted laser/desorption ionization time of flight (MALDI-TOF) mass spectrometry. The protein is cleaved with a proteolytic enzyme (usually trypsin) to yield the constituent small peptides. The accurate mass of these peptides is determined by MS analysis. This gives the peak list of peptides of the protein. This peak list is compared with the theoretical peptide peak list obtained from the *in silico* digestion of the database proteins and the best match is identified by computer software. The main advantage of this method is that it does not depend on protein sequencing for protein identification. PMF was carried out in the Proteomics and Genomics Service (Centro de Investigaciones Biológicas – CSIC).

3.2.6.3. Dynamic light scattering

The Dynamic Light Scattering (DLS) technique measures motion optically by recording the scattered light signal at a fixed angle. The particles are illuminated with a monochromatic, coherent light source (laser) and the light scattered by the particles is recorded. The recorded scattered light contains information on the diffusion speed and thus enables to detect aggregates in macromolecular solutions and determine the size of proteins based on the Brownian motion of disperse particles, which states that smaller particles move faster, while larger ones move slower in a liquid. DLS analyses were performed using Zetasizer Nano ZS from the Automated Crystallography Platform of Barcelona Science Park.

3.2.6.4. Multi angle light scattering (MALS)

A multi angle light scattering (MALS) detector is a form of static light scattering detector which allows the absolute molecular weight (M_w) and potentially the radius of gyration (R_g) of a sample to be measured.

The most common method of using a MALS is to connect to a HPLC size exclusion chromatography (SEC). As part of this system it is essential that a concentration detector, Refractive index or UV, is also connected.

The SEC-MALS detector measures the light scattered by a sample at many angles and for a Debye plot to be created. The Debye plot models the angular dependence of a samples scattering and is used to determine the M_w and R_g at every data slice within the chromatogram. Importantly the fit order and model chosen for this analysis needs to be carefully considered as it will have a significant impact on the results. Isotropic scatterers, smaller than 10-15nm in radius, will scatter light evenly in all directions

meaning only the M_w can be measured. Anisotropic scattering samples, more than 10-15 nm in radius, will scatter more light in the forward direction allowing both M_w and R_g to be measured. Using the R_g a conformation plot can be generated (plot of R_g against M_w) which allows any structural differences between samples to be measured.

ASTRA software is used to determine the molar masses and root mean square radii of a sample whose light scattering properties have been measured. The most important application of the ASTRA program is its ability to convert measurements of a fractionated sample, passing through appropriate instrumentation, into an accurate representation of the molar masses and sizes present in the sample. MALS analyses were performed at the Automated Crystallography Platform of Barcelona Science Park.

3.2.6.5. Nuclear magnetic resonance (NMR)

Nuclear magnetic resonance spectroscopy is a widely used and powerful method that takes advantage of the magnetic properties of certain nuclei. A 1D proton NMR experiment enables to determine whether a protein is well-folded and therefore if it is likely to be functional. The most noticeable difference in NMR spectra obtained for the same protein molecule in folded and unfolded states lies in the chemical shift dispersion observed. In folded states, ring current and other dipolar interactions lead to increased chemical shift dispersion. These interactions are averaged out by conformational fluctuations in unfolded or partly folded states and chemical shifts are close to random coil values. Another parameter of critical evaluation is the dispersion rate and the width of the signals in the amide region. Particularly, the appearance of intensities at chemical shifts near 8.3 ppm is representative of disordered proteins as this is contributed majorly by the backbone amides randomly structured. On the other hand, signal dispersion beyond 8.5 ppm (8.5–11 ppm) is indicative of globular folded protein structure. Because of the different chemical environment and, thus, the varying shielding effects, the resonances of the single protons will be distributed over a wide range of frequencies. Furthermore, in the aliphatic region of the spectrum, between 1.0 and -1.0 ppm, a large signal dispersion versus a steep flank of the dominant peaks at approximately 1 ppm separates a structured protein from an unfolded one. 1D NMR experiments were performed using 500MHz NMR spectrometer of the NMR Unit of the Centros Científicos y Tecnológicos de la Universidad de Barcelona (UB).

3.2.7. Biochemical assays

3.2.7.1. In vitro nuclease activity assays

Nuclease assays were performed with the purified protein. The protocol was the following:

1. Prepare the following reaction in an eppendorf tube:

- ❖ 15 µl of EBV TRM3-C at 0.1 mg/ml. The final protein concentration is 2 µM
- ❖ 2 µl DNA (100 ng/µl). Vector pUC18: It is a high copy number *E. coli* plasmid of 2686 bp. The reaction was done with linear DNA (digested with Hind III).
- ❖ 7,5 µl activity buffer (30 mM Tris pH 9, 50 mM NaCl)
- ❖ 0,5 µl ions solution at 150 mM (CaCl₂, MgCl₂ or MnCl₂). The final concentration is 3 mM

2. Incubate the reaction mix 1 hour at 37°C.
3. Stop the reaction by adding EDTA to 30 mM final concentration.
4. Analyse the samples by agarose gel with SYBR-Safe staining.

The structure of the tripartite terminase complex of HSV-1 shows the presence of Mg⁺² in the active centre of the protein (*Yang et al. 2020*).

3.2.1.1. Inhibition assays

Different inhibitors were tested at different concentrations in order to determine which was able to inhibit the nuclease activity of the EBV TRM3-C in a lower micromolar range. BS14 and BS17 are specific terminase inhibitors developed by Salvatore Bongarzone at the lab while Bictegravir, Elvitegravir, Dolutegravir, Cabotegravir, 707035 and MK-2048 are commercial HIV-integrase inhibitors. A stock solution of these inhibitors at 5mM was prepared with 50 % DMSO. This solution was diluted with the reaction buffer (30 mM TrisHCl pH 8, 50 mM NaCl) to 0.5 mM and 0.05 mM respectively.

The GelAnalyzer 19.1 program was used to analyze the agarose gels. The obtained information was used to calculate the IC50 of each inhibitor.

3.2.8. Crystallization

Crystallography is the major contributor to three-dimensional protein structure information. However, the production of soluble purified protein and diffraction-quality crystals are clearly the major roadblocks preventing the realization of high-throughput structure determination. In the first step of crystallization screening, the protein is exposed to a wide variety of reagents at different concentrations in order to find 'hits' or 'leads' that point to conditions that may be conducive to crystallization. Once a 'hit' is identified, parameters such as precipitant concentration, pH, and temperature are used to optimize the production of crystals suitable for analysis by X-ray diffraction.

Crystallization requires bringing the macromolecule to supersaturation. The sample should therefore be concentrated to the highest possible concentration without causing aggregation or precipitation of the macromolecule (usually 2-50 mg/mL). The crystallization process proceeds in two rather distinct but inseparable steps: nucleation (labile zone) and growth (metastable zone). Nucleation is the most difficult problem to address theoretically and experimentally because it represents a first-order phase transition by which molecules pass from a wholly disordered state

to an ordered one. The growth of macromolecular crystals is a better characterized process than nucleation. Protein crystals grow principally by the classical mechanisms of dislocation growth and growth by two-dimensional nucleation, along with two other less common mechanisms known as normal growth and three-dimensional nucleation.

The physical process of causing a change in concentration can be carried out by two main techniques: vapor diffusion and batch crystallization. In this project vapor diffusion was used. In vapor diffusion, a drop containing a mixture of precipitant and protein solutions is sealed in a chamber with the reservoir solution. Water vapor then diffuses out of the drop until the osmolarity of the drop and the reservoir solution are equal. The dehydration of the drop causes a slow concentration of both protein and precipitant until equilibrium is achieved. With the correctly chosen conditions, crystallization of the protein will occur over the course of days or weeks.

3.2.8.1. Microseeding

This technique takes advantage of the fact that the formation of a crystal is a two step process divided into nuclei formation and crystal growth. The initial step, nuclei formation, is more likely to occur if the protein solution is highly supersaturated. In contrast, the growth of crystals, an ordered process, is maintained in the metastable zone of the phase diagram.

Seeding methods separate the two events of nucleation and crystal growth. In microseeding this separation is accomplished by transferring a seed, a submicroscopic crystal, from one condition, where the level of supersaturation is high, to a similar condition at a lower level of supersaturation. In order to have lower levels of supersaturation either the protein or the precipitant concentration is lowered in a crystallization setup.

3.2.8.2. Protein crystallization screening

As it is essentially impossible to predict crystallization conditions for proteins, the process usually relies on extensive screening of hundreds to thousands of conditions, including different precipitants, salts, buffers, additives and etc. 96-well 2-Drop MRC plates were prepared at the Automated Crystallography Platform (PAC) using Freedom EVO robot (TECAN). Crystallization drops were dispensed using both Cartesian Honeybee-X8 (Cartesian Technologies) and Phoenix Rigaku Edition (Art Robbins Instruments) for high-throughput screening featuring accuracy, speed, and precision. Crystallization plates were incubated at 4°C or 20°C and plate inspection was done manually using SMZ1000 (Nikon) optical microscopes or automatically by a Crystal Farm 400 (Bruker Corporation) robot.

3.2.8.3. Protein crystallization optimization

Initial crystals frequently possess something less than the desired optimal characteristics for diffraction. The crystals may be too small or too large, have unsuitable morphology or yield poor X-ray diffraction intensities. It is therefore

necessary to improve upon these initial crystallization conditions in order to obtain crystals of sufficient quality for X-ray data collection.

The optimization process entails sequential, incremental changes in the chemical parameters that influence crystallization, such as pH, ionic strength and reagent concentration, as well as physical parameters such as temperature and sample volume. 24-well plates (Crystalgen) were set up manually using two different experimental setups for vapor diffusion. In the hanging drop experiment the crystallization drop is hanging from a cover slide, which closes the reservoir, whereas in the sitting drop experiment the crystallization drop is placed on a shelf above the reservoir. The crystallization drops were regularly inspected to determine whether better crystals were obtained with the optimized conditions.

3.2.8.4. Crystal freezing

In order to minimise radiation damage and crystal desiccation during data collection, the crystals were cryo-cooled with liquid nitrogen during storage and with nitrogen gas during data collection. The radiation damage is due to the formation of free radicals which can spread within the crystal altering the protein and producing changes in the diffraction pattern and the unit cell dimensions. The diffusion of free radicals is much lower in frozen crystals. A critical point during cryo-cooling is to avoid the crystallization of solvent molecules, mainly water, which could interfere in the data collection. For this purpose, we added a cryo-protectant to the reservoir solutions.

Before being frozen in liquid nitrogen, crystals were soaked in the reservoir solution containing increasing amounts of cryoprotectant. Determining the optimal cryoprotectant is a process of trial and error. We tested the most common ones (glycerol, ethylene glycol and PEG400) at different concentrations (10%, 20%, 30%). Protein crystals were fished using 20 micron nylon cryoloops (Molecular Dimensions) of different size (0,05-0,1 mm; 0,1-0,2 mm; 0,2-0,3 mm).

3.2.9. X-ray diffraction

X-ray crystallography is a powerful non-destructive technique for determining the three dimensional molecular structure of a crystal that uses the principles of X-ray diffraction to analyze the sample in many different directions so that the 3D structure can be built.

Crystals are exposed to an X-ray beam and the resulting diffraction patterns are processed, initially to yield information about the crystal packing symmetry and the size of the repeating unit that forms the crystal. This information is obtained from the pattern of the diffraction spots. The intensities of the spots can be used to determine the structure factors from which a map of the electron density can be calculated, if phases can be determined. Several methods can be used to improve the quality of this map until it is clear enough to allow the building of the molecular structure using the protein sequence. The resulting structure is then refined to fit the map more accurately and to adopt a thermodynamically favoured conformation.

3.2.9.1. Data collection

Diffraction data was collected at the BL13-XALOX beamline at ALBA synchrotron, a 3rd-generation synchrotron facility. The data collection detector was a photon-counting Pilatus 6M (DECTRIS Ltd). For collecting diffraction data from single small crystals ($\leq 15 \mu\text{m}$ in one dimension) ID30A-3 (or MASSIF-3) was used, which is a mini-focus (beam size $18 \mu\text{m} \times 14 \mu\text{m}$) highly intense (2.0×10^{13} photons s^{-1}), fixed-energy (12.81 keV) beamline for macromolecular crystallography (MX) experiments at the European Synchrotron Radiation Facility (ESRF).

3.2.9.2. Data processing

The collected diffraction data were indexed, integrated and scaled using the program XDS (Kabsch, 2010). XDS needs only one input file, XDS.INP, which contains all relevant information about the data collection, from beam parameters to detector parameters and crystal parameters as well as the data collection geometry.

3.2.10. Structure determination: molecular replacement

In order to determine the 3D structure from protein crystals, amplitudes and phases of the diffracted X-rays are required. Amplitudes can be deduced from the measured intensities of the diffracted X-rays but phase information cannot be directly measured in the native diffraction data. This is the 'phase problem' of X-ray crystallography. Without phase information it is impossible to reconstruct the electron density in the unit cell. An alternative approach to the phase problem may be used when the molecule under study is similar to another molecule whose structure is already known. In this case the molecular replacement (MR) method (Rossmann, 1972) allows phases to be obtained from the known structure. Solved structures of HSV-1 (PDB: 6M5T) were used to estimate the phases for our target structure.

The molecular replacement method (MR) is focussed on two 3-D searches to place the search model in the new unit cell: the Rotation Function (RF), which determines the best orientation of the model covering the rotational asymmetric unit for the space group and the Translation Function (TF), which determines its absolute position in the unit cell. Multiple components can be placed sequentially to solve the structure of a crystal containing multiple copies in the asymmetric unit.

The program used for phasing the structures was Phaser from PHENIX (Python-based Hierarchical ENvironment for Integrated Xtallography).

3.2.10.1. Model building and refinement

After the initial phasing and building, the model of a protein is generally far from perfect. To improve the phases and also the interpretation of the electron density map refinement methods are a very important step. Refining is achieved through statistical adjustment of the atomic coordinates to fit the diffraction data better. As a measure of the fitting the work R-factor is used, that measures how far the calculated amplitudes differ from the observed amplitudes. The work R-factor can get trapped in

local minima giving the false impression of having a good model. For this reason the free R-factor is used to validate the refinement process using a random small set of data not included in the refinement. Two methods are widely used in refinement: maximum likelihood and simulated annealing. Both methods use restraints to how an atomic model has to look like in respect to bond distances, angles and torsions and temperature factors (B-factors). In maximum likelihood the phases are adjusted to minimise the R-factor. In simulated annealing the structure is heated to add randomness and slowly cooled and refined. Several rounds of refinement using Refmac (*Murshudov et al., 2011*) were performed in combination with manual model correction using Coot (*Emsley & Cowtan, 2004*).

3.2.10.2. Final model validation

MolProbity, a general-purpose web server offering quality validation for 3D structures of proteins, nucleic acids and complexes was used. It provides detailed all-atom contact analysis of any steric problems within the molecules as well as updated dihedral-angle diagnostics, and it can calculate and display the H-bond and van der Waals contacts in the interfaces between components. An integral step in the process is the addition and full optimization of all hydrogen atoms, both polar and nonpolar. MolProbity results are reported in multiple forms: as overall numeric scores, as lists or charts of local problems, as downloadable PDB and graphics files, and most notably as informative, manipulable 3D kinemage graphics shown online in the KiNG viewer.

3.2.10.3. Structural data visualization and analysis

Structure visualization was performed using Coot (*Emsley & Cowtan, 2004*), UCSF-Chimera (*Pettersen et al., 2004*) and PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

3.2.11. Model of structures with inhibitors

In order to analyze the interactions between the inhibitors and each protein (TRM3-C of HSV-2, VZV, EBV, HCMV and KSHV), the interacting models were built based on the structure solved of TRM3-C HCMV in complex with BS2 (*Bongarzone et al., 2017*).

The protein structures used to build such models were for TRM3-C HCMV (PDB: 3N4P) and for the rest of the proteins (TRM3-C HSV-2, VZV, EBV and KSHV) AlphaFold models.

AlphaFold is an AI system developed by DeepMind that directly predicts the 3D coordinates of all heavy atoms for a given protein using the primary amino acid sequence and aligned sequences of homologues as inputs. AlphaFold greatly improves the accuracy of structure prediction by incorporating novel neural network architectures and training procedures based on the evolutionary, physical and geometric constraints of protein structures.

The network comprises two main stages. First, the trunk of the network processes the inputs through repeated layers of a novel neural network block, the Evoformer, that

produces an $N_{\text{seq}} \times N_{\text{res}}$ array (N_{seq} , number of sequences; N_{res} , number of residues) that represents a processed MSA and an $N_{\text{res}} \times N_{\text{res}}$ array that represents residue pairs. The trunk of the network is followed by the structure module that introduces an explicit 3D structure in the form of a rotation and translation for each residue of the protein (global rigid body frames). These representations are initialized in a trivial state with all rotations set to the identity and all positions set to the origin, but rapidly develop and refine a highly accurate protein structure with precise atomic details (Jumper *et al.* 2021).

3.2.12. In vitro antiviral screening

Cell culture and virus strains. Human foreskin fibroblast (HFF) cells prepared from human foreskin tissue were obtained from the University of Alabama at Birmingham tissue procurement facility with approval from its IRB. Akata cells were kindly provided by John Sixbey (Louisiana State University, Baton Rouge, LA). BCBL-1 cells were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH. Molt-3 cells were obtained from Scott Schmid at the Centers for Disease Control and Prevention, Atlanta, GA. The E-377 strain of HSV-1 was a gift of Jack Hill (Burroughs Wellcome). The HCMV strain AD169 and HSV-2 strain G were obtained from the American Type Culture Collection (ATCC, Manassas, VA). VZV, strain Ellen, was obtained from the ATCC. Akata cells latently infected with EBV were obtained from John Sixbey. The Z29 strain of HHV-6B was a gift of Scott Schmid at the Centers for Disease Control and Prevention, Atlanta GA. HHV-8 was obtained as latently infected BCBL-1 cells through the NIH AIDS Research and Reference Reagent Program.

Antiviral Assays. Each experiment that evaluates the antiviral activity of the compounds includes both positive and negative control compounds to ensure the performance of each assay. Concurrent assessment of cytotoxicity is also performed for each study in the same cell line and with the same compound exposure.

❖ CPE assays for HSV-1, HSV-2, VZV and HCMV.

Assays were performed in monolayers. Cells were seeded in 384 well plates and incubated for 24h to allow the formation of confluent monolayers. Dilutions of test drug were prepared directly in the plates and the monolayers infected at a predetermined MOI based on virus used. After incubation, cytopathology was determined by the addition of CellTiter-Glo (CTG) reagent. Concentrations of test compound sufficient to reduce CPE by 50% (EC_{50}) or decrease cell viability by 50% (CC_{50}) were interpolated using standard methods in Microsoft Excel. Acyclovir was used as negative control. Ganciclovir was used as negative control.

❖ Assays for EBV, HHV-6B, and HHV-8.

For EBV assays, Akata cells were induced to undergo a lytic infection with 50 $\mu\text{g}/\text{ml}$ of a goat anti-human IgG antibody. Experimental compounds were diluted within plates; the cells were added and incubated for 72 h. For HHV-6 assays, compounds were serially diluted plates then uninfected Molt-3 cells were added to each well. Infection was initiated by adding HHV-6B infected Molt-3 cells, at a ratio of approximately 1 infected cell for every 10 uninfected cells. Assay plates were incubated for seven days at 37°C. Assays for HHV-8 were performed in BCBL-1 cells. Similar plates were

initiated without virus induction/addition and used for measuring cytotoxicity by the addition of CTG.

For all assays, the replication of the virus was assessed by the quantification of viral DNA. Compound concentrations sufficient to reduce genome copy number by 50% were calculated from experimental data as well as compound cytotoxicity.

CHAPTER 4: RESULTS AND DISCUSSION

4.1. RESULTS AND DISCUSSION

4.1.1. Herpesvirus tripartite terminase subunit 3

Herpesvirus genome packaging molecular motor is a complex composed of three subunits: a large subunit, a small subunit, and a third subunit, which are collectively known as terminase. TRM3 is highly conserved in herpesvirus. It mainly includes two domains: the C-terminal nuclease domain, which cuts the viral concatemeric DNA into a monomeric genome, and the N-terminal ATPase domain, which hydrolyzes ATP to provide energy for the genome cutting and transfer activities.

In this work we report the cloning, expression, purification and functional activity of the TRM3-C terminal nuclease domain of several herpesvirus: HSV-2, VZV, EBV, HCMV and KSHV. We also characterize structurally the TRM3 C-terminal nuclease domain of HSV-2.

4.1.2. Construct design

Protein engineering is a key step for improving protein physiochemical properties leading to proteins that are more stable, soluble, and have a higher propensity to crystallize. The main objective in construct design is to produce large, homogenous quantities of soluble proteins with a high likelihood to crystallize. With a high-throughput method, ESPRIT (*Mas & Hart 2017*), it was possible to obtain a soluble form of TRM3 HCMV (*Nadal et al. 2010*). This construct corresponds to the TRM3 C-terminal domain of HCMV and it was 91 residues shorter than the predicted domain. This fragment was called TRM3-C HCMV. The same soluble fragment was cloned for HSV-2, VZV, EBV and KSHV. These fragments were called TRM3-C HSV-2, TRM3-C VZV, TRM3-C EBV and TRM3-C KSHV. The sequence alignment of all these fragments is shown in the next figure (Figure 21).

```

EBV      SQGSLSRVTRDDAINQLELCRVDTLNPRVAGRLASSLYVYVDPAYTNNTSASGTGIAAVT 60
KSHV     SNATLYRVVGDAAALQFDMCRVDTTAQQVQKCLGKQLFVYIDPAYTNNTSASGTGVAIV 60
HCMV     KISQNTVLITDQSRREEFDILRYSTLNTNAYDYFGKTLVYVLDPAFTTNRKASGTGVAIV 60
HSV-2    ETGDDRPLVTKSAGERFLLYRPSTT--TNSGLMAPELYVYVDPAFNTANRASGTGIAIV 58
VZV      GVGFDRLPFTKTARERFLYRPSTV--ANCAILSSVLYVYVDPAFNTANRASGTGVAIV 58
          . . . . : : * .* . . : * : * : * * : * * : * * : * * : * *
EBV      HDRADPNRVIVLGLHEHFFLKDLTGDAALQIATCVVALVYSIVTLHPH-LEEVKVAVEGNS 119
KSHV     TSQTPTRSILIGMEHFFLRDLTGAAAYEIASCACTMIKAIKAVLHPT-IERVNAAVEGNS 119
HCMV     AYRH---QFLIYGLHEHFFLRDLSESSEVAIAECAAHMIIISVLSLHPY-LDELRIAVEGNT 116
HSV-2    RYRD---DFIIFALEHFFLRALTGSAADIARCVAHSLAQVLAHPGAFRSVRVAVEGNS 115
VZV      RYKS---DWIIFGLEHFFLRALGTSSSEIGRCVTQCLGHILALHPNTFTWVHVSIEGNS 115
          : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
EBV      SQDSAVAIASIIIGESCP-----LPCAFVHTKDKTSSLQWPMYLLTNEKSKAFE 167
KSHV     SQDSGVAIATVINEICP-----LPIHFLHYTDKSSALQWPIYMLGGEKSSAFE 167
HCMV     NQAAAVRIAQLRQSVQ-----SSTLRVLFYH-TPDQNHIEQPFYLMGRDKALAVE 167
HSV-2    SQDSAVAIATHVHTEMHRI LASAGANGPPELLFYHCEPPGGAVLYPFFLLNKQKTPAFE 175
VZV      SQDSAVAI SLAI AQQF -AVLEKGNVLSAPVLLFYHSIPPGCSVAYPFFLLQKQKTPAVD 174
          .* : .* * : : * * : * * : * : * : * : * : * : * : * : * : * : *
EBV      RLIYAVNTASLSASQVTVSNTIQLSFDPVLYLISQIRAIKPIPLR---DGYTYTGK-- 221
KSHV     TFIYALNSGTLASQTVSNTIKISFDPVTYLVEQVRAIKCVPLR---DGGQSYSAK-- 221
HCMV     QFISRFNSGYIKASQELVSYTIKLSHDIPEYLLEIQNLHRVTLA---EGTTARYSAK-R 223
HSV-2    YFIKRFNSGGVMASQELVSVTVRLQDTPVEYLSQNLNLIETVSP---NTDVRMYSYGK-- 230
VZV      YFVKRFNSGNIASQELVSLTVKLGDPVEYLKQLDNLTEVTKGGMNLDTKTYTGKGT 234
          : : .* : . : * * * * : * : * : * : * : * : * : * : * : * : *
EBV      QRNLSDDVLVALVMAHFLATTQKH--TFKKVH-- 251
KSHV     QKHMSDDLVAVVMAHFMATDDRH--MYKPISPQ- 253
HCMV     QNRISDDLIAVIMATYLCDDIHAIR--FRVSV-- 253
HSV-2    RINGAADDLHVAVIMAIYLAAPTGIFFAFPIITRTS 265
VZV      TGTHSDDLHVALIMSVYIGSSCIPDSVFMPPIK--- 266
          : * : : * : * : * : * : * : * : * : * : * : * : * : * : * : *

```

Figure 21. Sequence alignment of TRM3-C terminal domain of HSV-2, VZV, EBV, HCMV and KSHV (Clustal Omega).

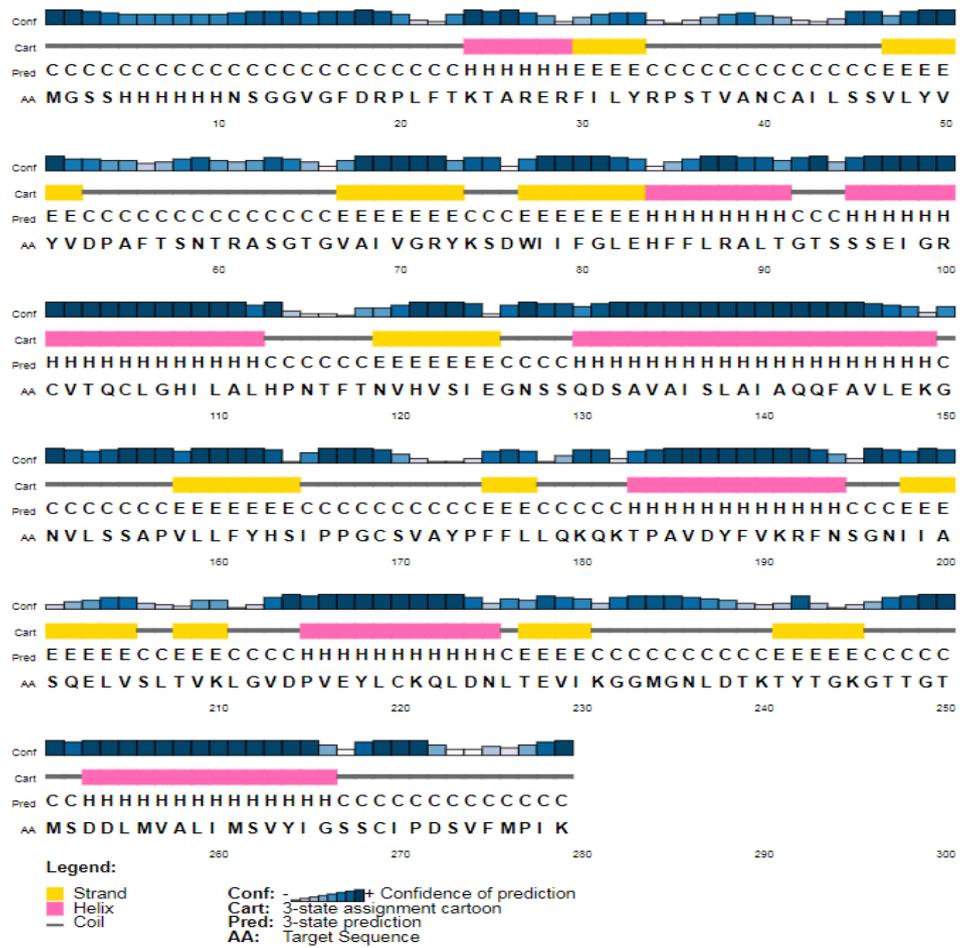


Figure 23. Secondary structure prediction of TRM3-C VZV.

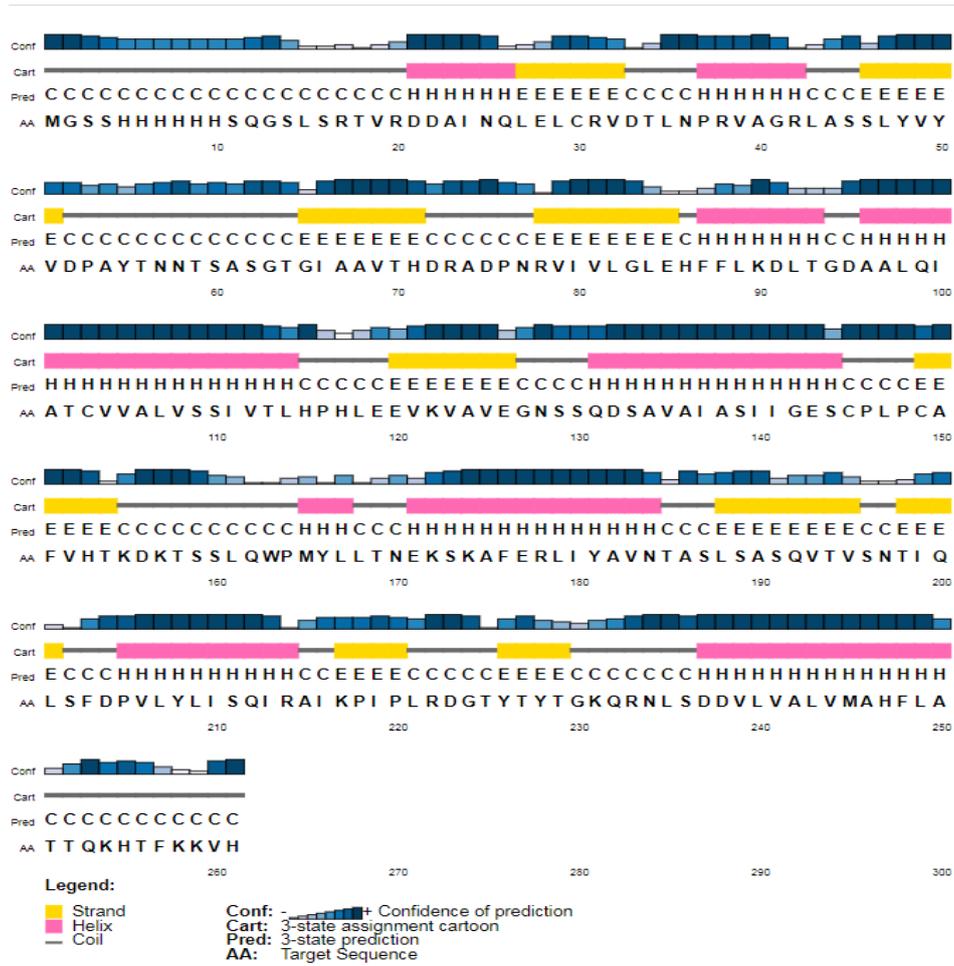


Figure 24. Secondary structure prediction of TRM3-C EBV.

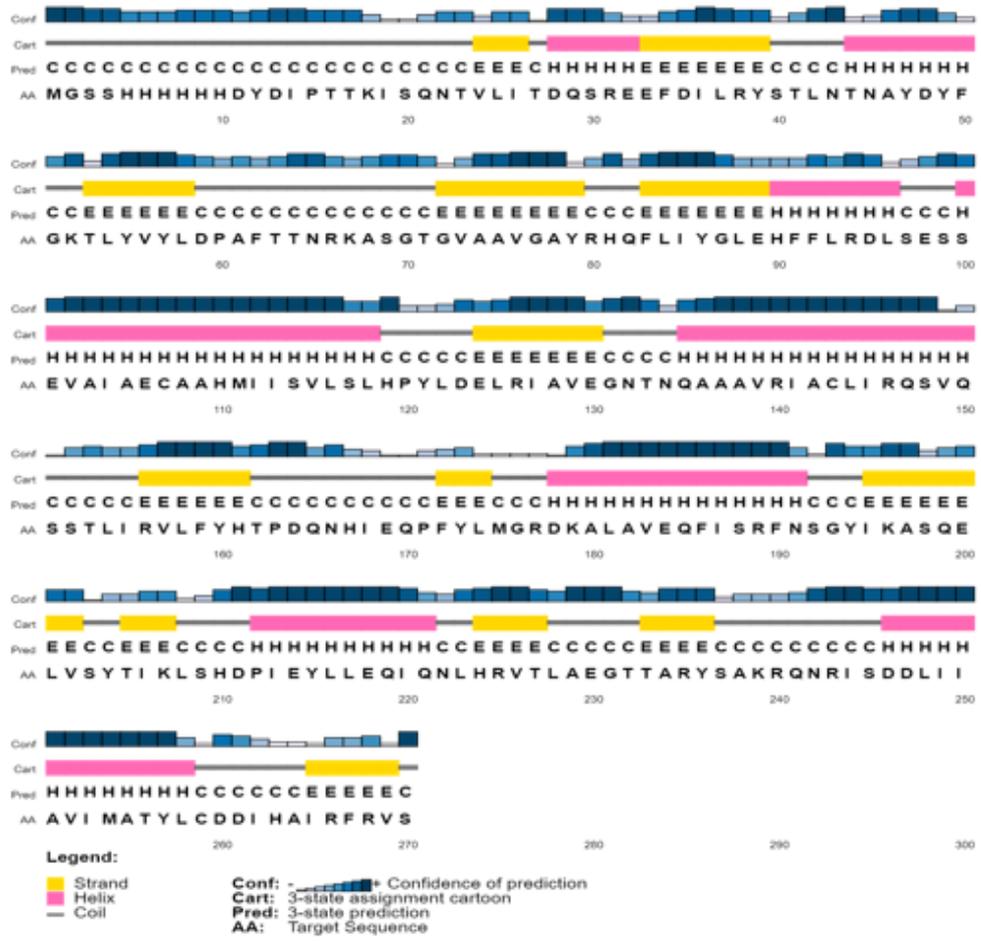


Figure 25. Secondary structure prediction of TRM3-C HCMV.

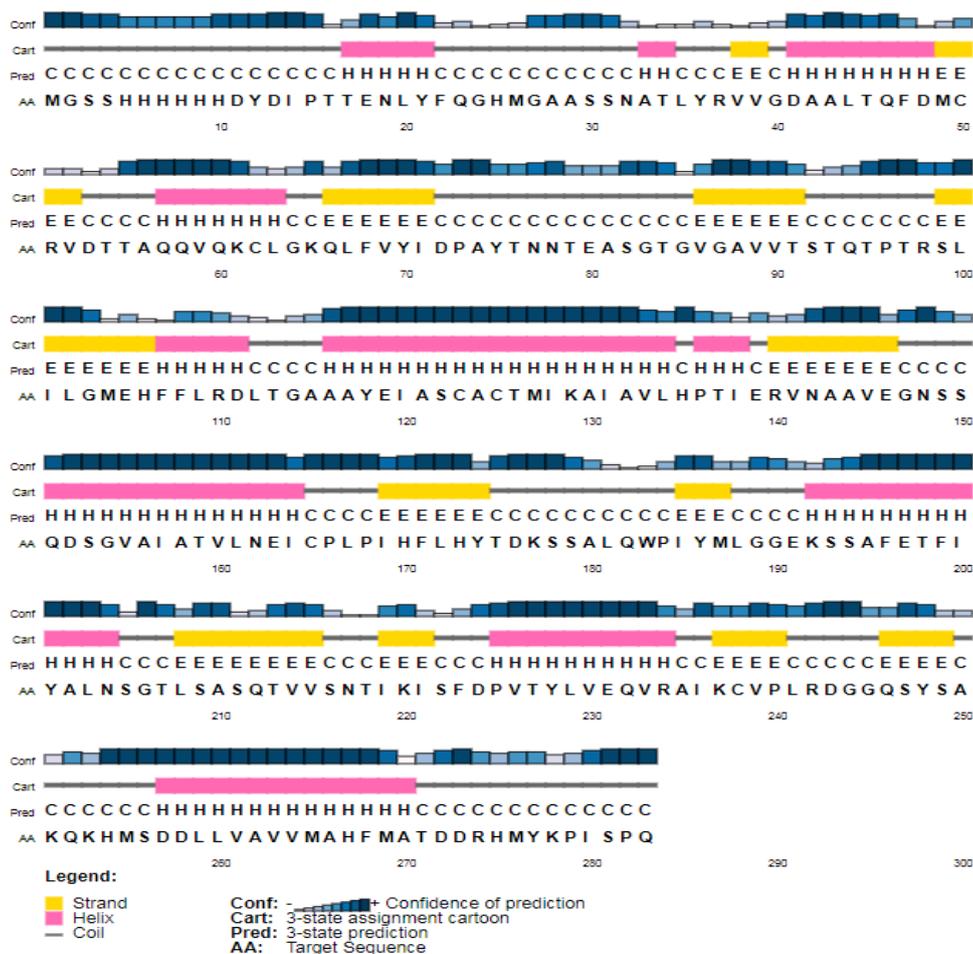


Figure 26. Secondary structure prediction of TRM3-C KSHV.

4.1.3. Cloning

Once the desired protein fragment has been determined, the choice of the cloning method to insert the gene into an expression vector is important. The TRM3-C of HSV-2 (Table 9), VZV (Table 10) and EBV (Table 11) of approximately 250 bp was cloned into the pET-28a (+) expression vector. The expression vector pET-28a (+) contains an N-terminal His-Tag®/thrombin/T7-Tag® configuration plus an optional C-terminal His-Tag sequence. The affinity tag facilitates the detection and purification of the recombinant protein.

Vector	pET-28a (+)
Tag Protein	MGSSHHHHHHMQEIIGG (N-terminal on insert)
Antibiotic resistance	Kanamycin
Promoter	T7 promoter
Terminator	T7 terminator

Size (w/o insert)	5369bp
--------------------------	--------

Table 9. Features of the expression vector of TRM3-C HSV-2.

Vector	pET-28a (+)
Tag Protein	MGSSHHHHHHNSGGVGF (N-terminal on insert)
Antibiotic resistance	Kanamycin
Promoter	T7 promoter
Terminator	T7 terminator
Size (w/o insert)	5369bp

Table 10. Features of the expression vector of TRM3-C VZV.

Vector	pET-28a (+)
Tag Protein	MGSSHHHHHHSQGSLSR (N-terminal on insert)
Antibiotic resistance	Kanamycin
Promoter	T7 promoter
Terminator	T7 terminator
Size (w/o insert)	5369bp

Table 11. Features of the expression vector of TRM3-C EBV.

Vector	pET-28a (+)
Tag Protein	MGSSHHHHHHHDYDIPTT (N-terminal on insert)
Antibiotic resistance	Kanamycin
Promoter	T7 promoter
Terminator	T7 terminator
Size (w/o insert)	5369bp

Table 12. Features of the expression vector of TRM3-C HCMV.

Vector	pET-28a (+)
Tag Protein	MGSSHHHHHHHDYDIPTT (N-terminal on insert)
Antibiotic resistance	Kanamycin
Promoter	T7 promoter
Terminator	T7 terminator
Size (w/o insert)	5369bp

Table 13. Features of the expression vector of TRM3-C KSHV.

Previously, amplification of the TRM3-C of EBV was performed by PCR with specifically designed primers using PfuUltraII Fusion polymerase and the final construct was checked by PCR amplification with the same primers (Figure 27) and sequencing.

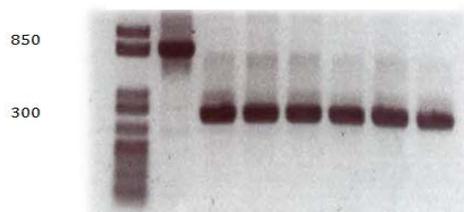


Figure 27. One-percent agarose gel stained with SYBR-Safe revealing amplicon product of TRM3-C of EBV.

4.1.4. Protein expression

Different expression systems are regularly used to generate recombinant protein suitable for crystallization purposes, including bacteria, insect cells, yeast and mammalian cells. Economic considerations as well as ease of use system directed us towards *E. coli* as cell host.

After small-scale expression assays with different *E. coli* strains and several induction temperatures and times the optimal overexpression condition was identified.

Large-scale protein expression was initiated by transforming the Rosetta 2 (DE3) *E. coli* strain with the expression plasmids (pET-28a (+) TRM3-C EBV/ pET-28a (+) TRM3-C HSV-2/ pET-28a (+) TRM3-C VZV). Plates with antibiotic selection (Kan-Chl) were incubated overnight at 37°C. Single colonies were resuspended in 4 ml liquid culture with antibiotic to produce a starter culture. The preinoculum was incubated at 37°C with shaking overnight. 500 ml cultures were inoculated and incubated at 37°C until OD600 reached 0.6-0.8. Afterwards, a final concentration of 1 mM IPTG was added and protein expression was performed overnight at 16°C. Cell pellets obtained by centrifugation were stored at -20°C. Samples were taken before induction and at the end of the incubation (Figure 28). The C-terminal nuclease domain of TRM3 in HSV-2, VZV, EBV, HCMV and KSHV was successfully overexpressed, leading us to design a protein purification workflow.

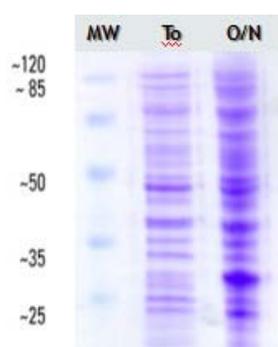


Figure 28. EBV TRM3-C overexpression.

4.1.5. Protein purification

Purified proteins are required for structural studies. Often more than one purification step is necessary to reach the desired purity taking into account the yield.

Cell pellets from 500 ml culture of *E. coli* BL21 (DE3) overexpressing TRM3-C protein of HSV-2, VZV, EBV, HCMV and KSHV were resuspended in buffer A (20 mM Tris [pH 8], 300 mM NaCl, 20 mM imidazole) supplemented with 50 µg/ml DNase I and a protease inhibitor cocktail (Roche). Cells were lysed using cell disruptor and the soluble fraction was filtered before column loading.

Affinity chromatography is the common first purification step when the target protein is affinity tagged. A second purification step (polishing step) with 10/300 Superdex 75 to remove remaining impurities or aggregates was followed. In some particular cases, ion exchange chromatographies were required between affinity chromatography and SEC steps to improve the quality of the samples. Fractions corresponding to the protein peaks were pooled and concentrated. Protein concentration was determined using a Nanodrop instrument measuring absorbance at 280 nm. Protein yield obtained per liter of initial culture with each protein are indicated in the following table (Table 14).

Protein	Concentration (ug/ml)
TRM3-C HSV-2	18,1 mg/ml
TRM3-C VZV	1,2 mg/ml
TRM3-C EBV	16,7 mg/ml
TRM3-C HCMV	8,3 mg/ml
TRM3-C KSHV	7.5 mg/ml

Table 14. Protein concentration per liter of initial culture.

4.1.5.1. Herpes Simplex Virus 2 (HSV-2)

TRM3-C protein of HSV-2 was eluted with a lineal gradient of imidazole, using buffer B (20 mM Tris [pH 9], 300 mM NaCl, 500 mM imidazole) in a HisTrap 5 ml HP affinity column (GE Healthcare) (Figure 29, Figure 30).

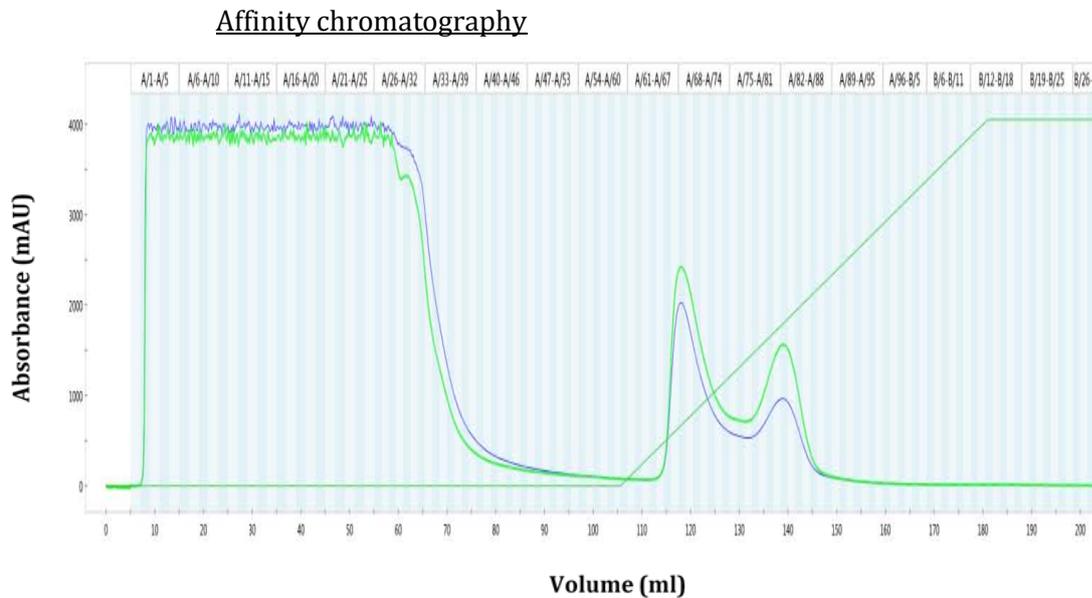


Figure 29. HisTrap HP 5 mL chromatographic profile of TRM3-C HSV-2. The first peak corresponds to contaminant proteins and the second one corresponds mostly to TRM3-C HSV-2.

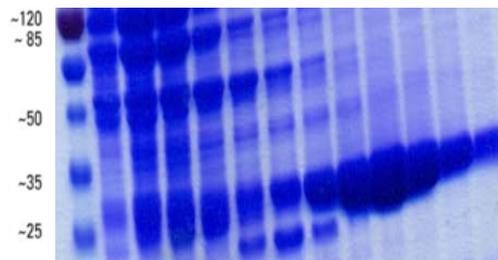


Figure 30. SDS-PAGE gel stained with Coomassie of the HisTrap HP 5 mL of TRM3-C HSV-2. The first seven lanes correspond to contaminant proteins and the last five lanes to TRM3-C HSV-2.

A second purification step with 10/300 Superdex 75 to remove remaining impurities or aggregates was followed using buffer C (50 mM Tris [pH 8], 300 mM NaCl) (Figure 31). The SDS-PAGE gel shows a large quantity of pure protein (Figure 32).

Size Exclusion Chromatography (SEC)

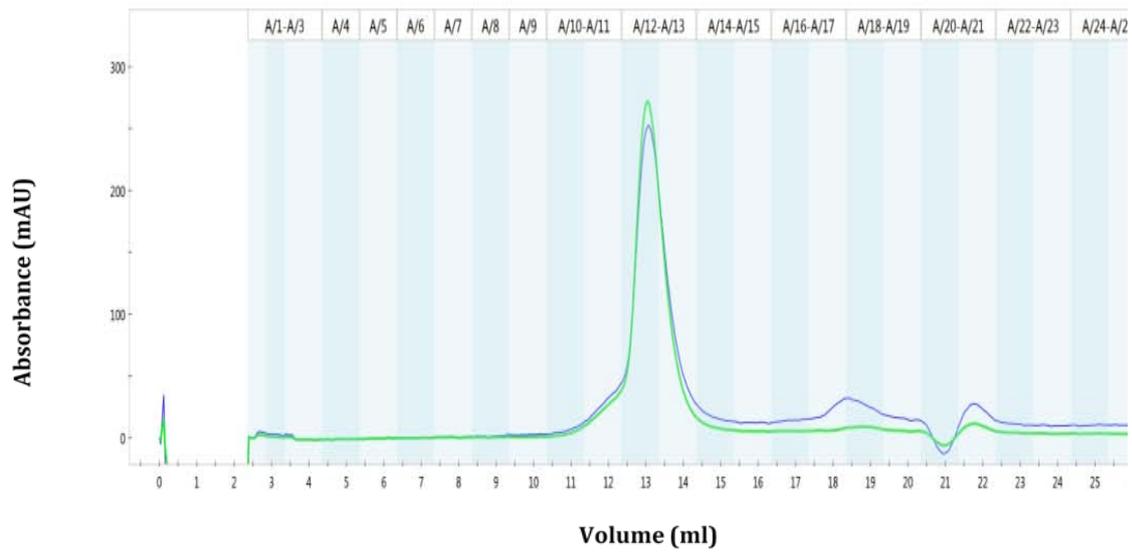


Figure 31. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C HSV-2 monomer.

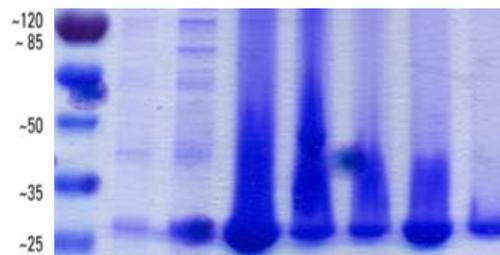


Figure 32. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C HSV-2. The first two lanes correspond to aggregated protein while the last five correspond to TRM3-C HSV-2.

4.1.5.2. Varicella-Zoster Virus (VZV)

The same protocol described in 4.1.5.1 was used to purify the TRM3-C domain of VZV. Chromatograms and SDS-PAGE of the affinity chromatographies (Figure 33, Figure 34) are shown.

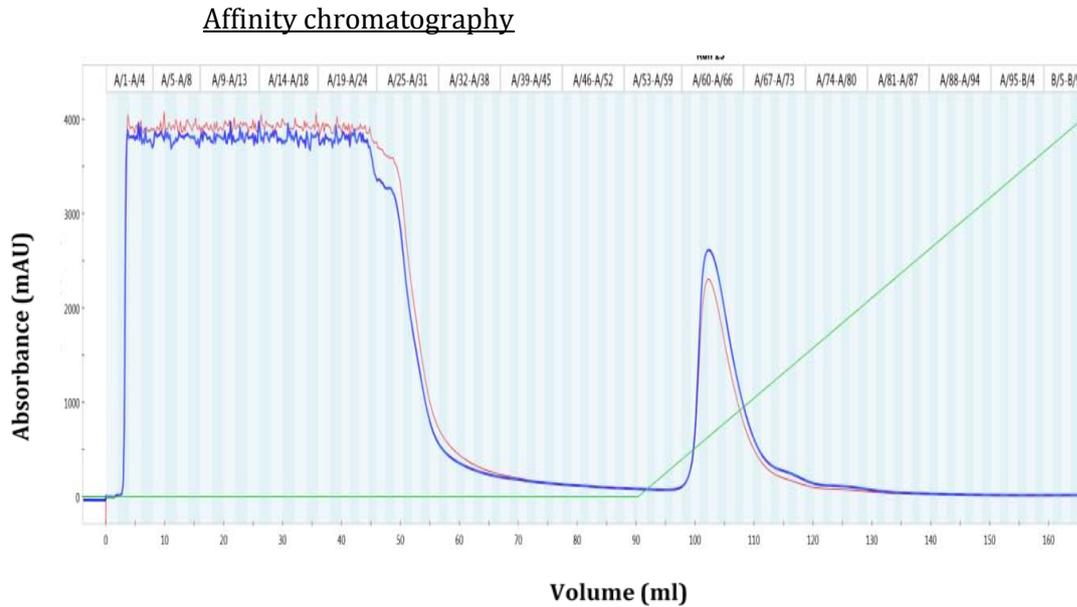


Figure 33. HisTrap HP 5 mL chromatographic profile of TRM3-C VZV. The main peak corresponds to contaminant proteins and the small one mostly to TRM3-C VZV.

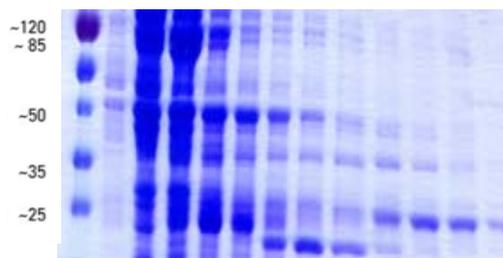


Figure 34. SDS-PAGE gel stained with Coomassie of the HisTrap HP 5 mL of TRM3-C VZV. The first eight lanes correspond to contaminant proteins and the last four lanes to TRM3-C VZV.

SEC results (Figure 35, Figure 36) indicate that the sample was not completely pure. Indeed, the obtained protein yield was very low (Table 14). For this reason, this sample was only used for performing biochemical assays.

Size Exclusion Chromatography (SEC)

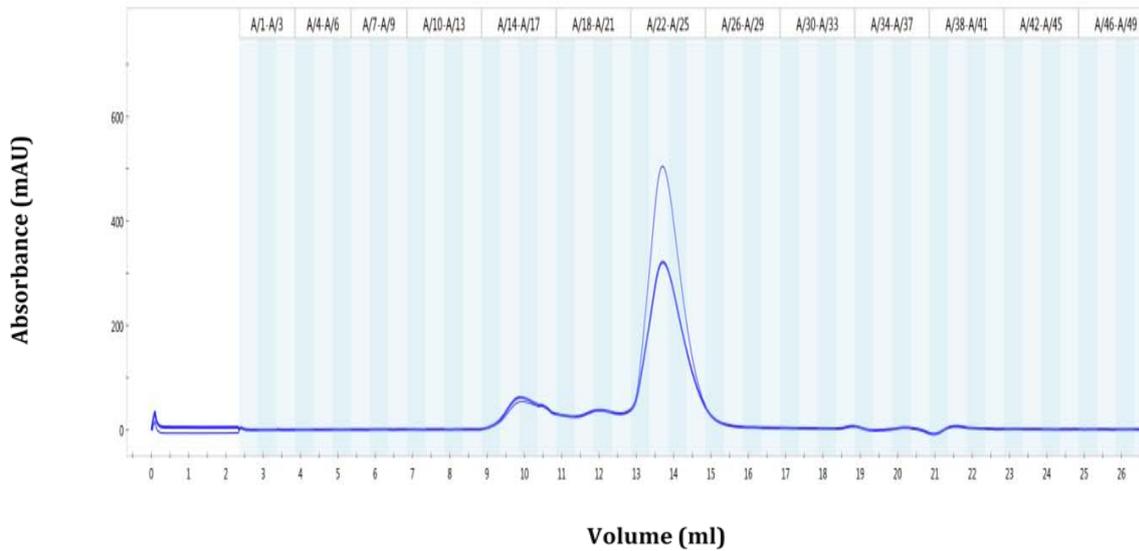


Figure 35. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C VZV. The first peak corresponds to aggregated protein, the second one to the dimer form of TRM3-C VZV and the main peak mostly to TRM3-C VZV monomer.

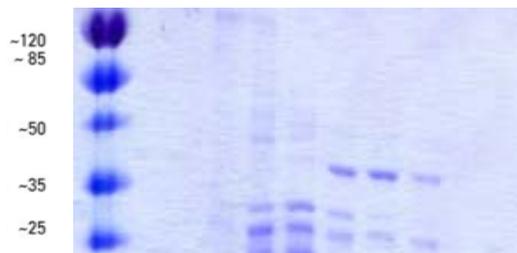


Figure 36. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C VZV. The first two lanes correspond to aggregated protein while the last three correspond to TRM3-C VZV.

4.1.5.3. Epstein-Barr virus (EBV)

The same protocol described in 4.1.5.1 was used to purify the TRM3-C domain of EBV. Chromatograms and SDS-PAGE of the affinity chromatographies (Figure 37, Figure 38) are shown.

Affinity chromatography

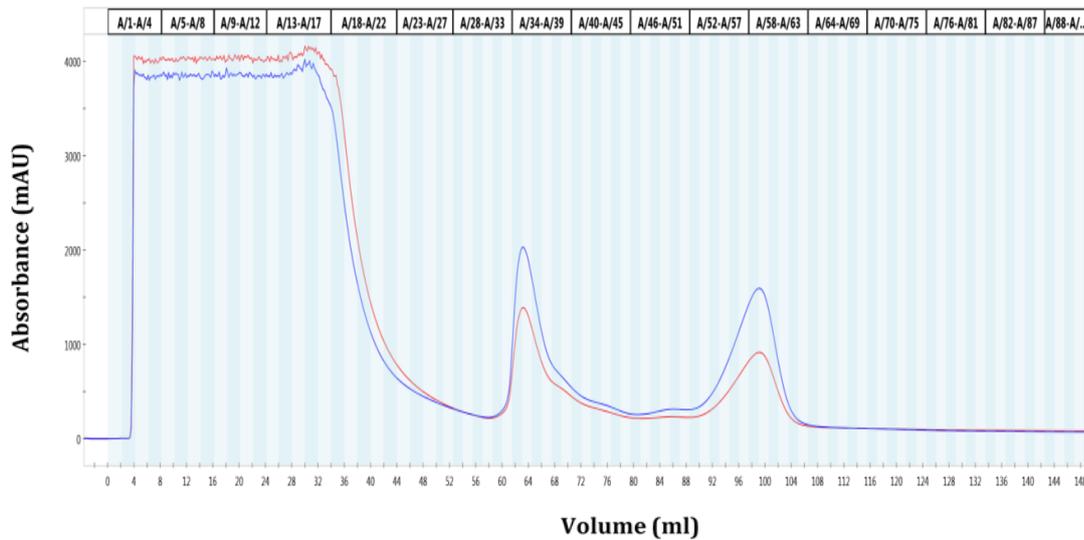


Figure 37. HisTrap HP 5 mL chromatographic profile of TRM3-C EBV. The first peak corresponds to contaminant proteins and the second one mostly to TRM3-C EBV.

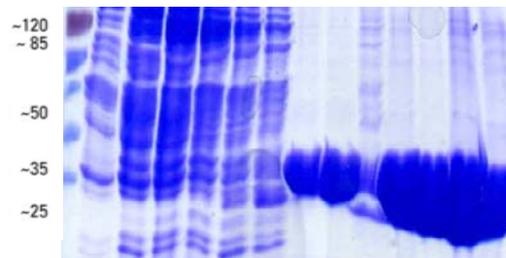


Figure 38. SDS-PAGE gel stained with Coomassie of the HisTrap HP 5 mL of TRM3-C EBV. The first six lanes correspond to contaminant proteins and the last seven lanes to TRM3-C EBV.

SEC results (Figure 39, Figure 40) indicated that the sample was completely pure. The obtained protein yield was high enough (Table 14) to perform crystallization assays.

Size Exclusion Chromatography (SEC)

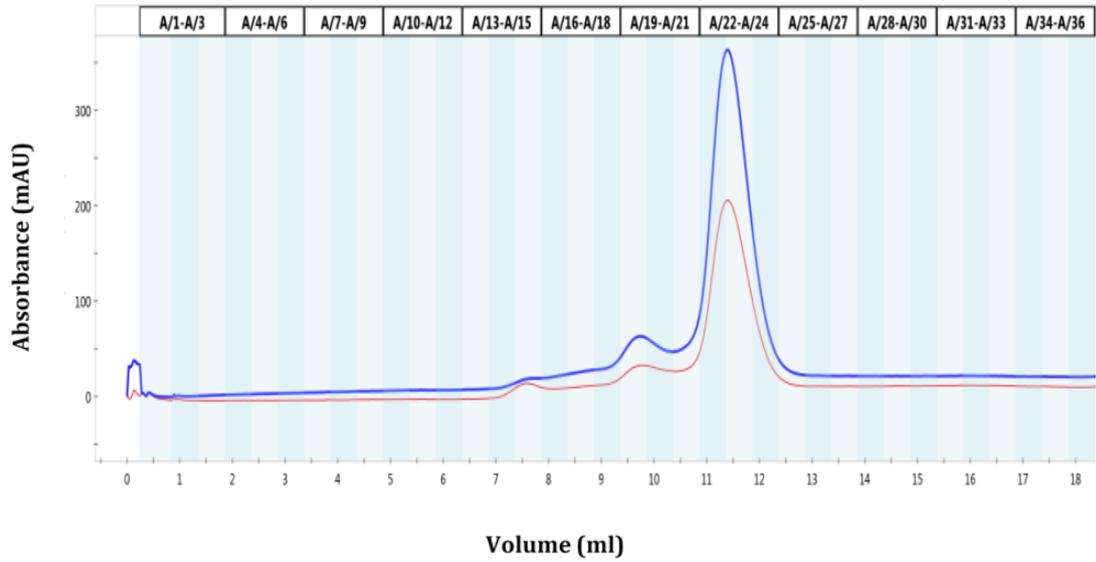


Figure 39. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C EBV. The first peak corresponds to the dimer form of TRM3-C EBV and the main peak to the monomeric form of TRM3-C EBV.

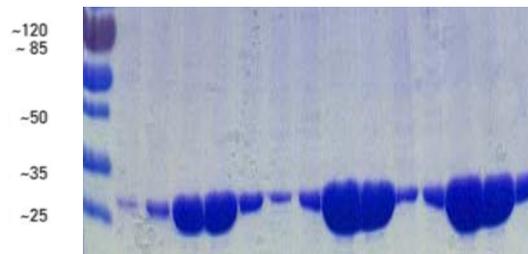


Figure 40. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C EBV. Since there was a lot of protein, three injections were needed and the gel shows the three replicas. The first two lanes of each replica correspond to the dimer form of TRM3-C EBV while the last two correspond to TRM3-C EBV monomer.

4.1.5.4. Human Cytomegalovirus (HCMV)

TRM3-C protein of HCMV was eluted with a lineal gradient of imidazole (50 mM Hepes [pH 8], 300 mM NaCl, 500 mM imidazole) in a HisTrap 5 ml HP affinity column (GE Healthcare). Chromatograms and SDS-PAGE of the affinity chromatographies (Figure 41, Figure 42) are shown.

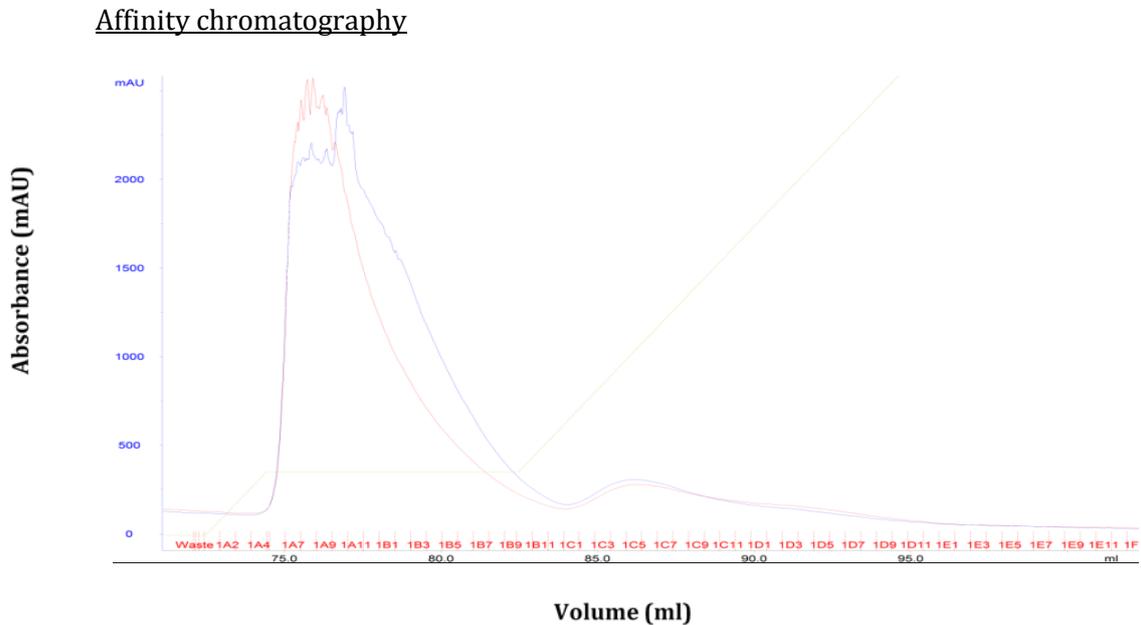


Figure 41. HisTrap HP 5 mL chromatographic profile of TRM3-C HCMV. The first part of the peak corresponds to contaminant proteins and the second one mostly to TRM3-C HCMV.

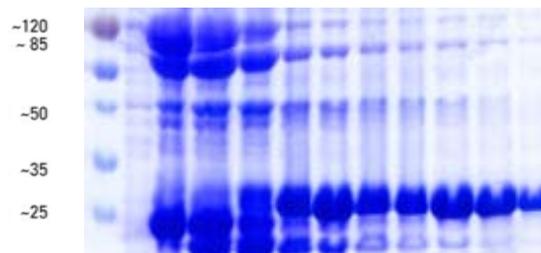


Figure 42. SDS-PAGE gel stained with Coomassie of the HisTrap HP 5 mL of TRM3-C HCMV. The first six lanes correspond to contaminant proteins and the last five lanes to TRM3-C HCMV.

An ion exchange chromatography (Mono Q 5/50) with a lineal gradient of NaCl (from 50 mM NaCl to 1 M NaCl) was performed. Chromatograms and SDS-PAGE of the ion exchange chromatography (Figure 43, Figure 44) are shown.

Ion exchange chromatography

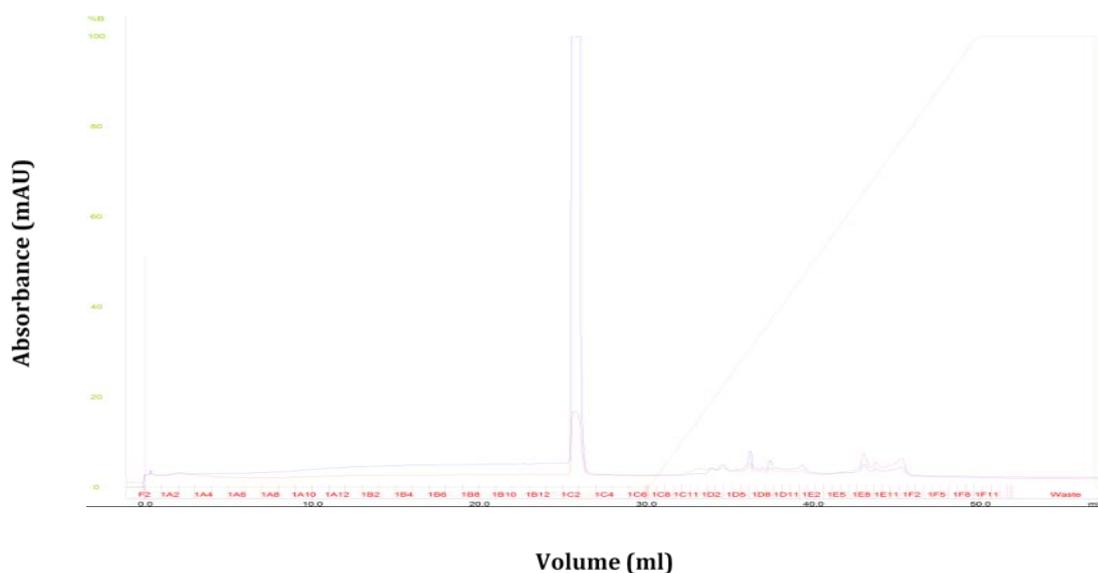


Figure 43. Mono Q 5/50 GL chromatographic profile of TRM3-C HCMV. The first part of the chromatogram corresponds to TRM3-C HCMV since it does not bind to the column and elutes in the flow-through. The peaks correspond to contaminant proteins that elute during the NaCl gradient.



Figure 44. SDS-PAGE gel stained with Coomassie of the Mono Q 5/50 GL of TRM3-C HCMV. All the lanes correspond to the flow-through fractions were the TRM3-C HCMV elutes.

A third purification step with 10/300 Superdex 75 to remove remaining impurities was performed (50 mM Hepes [pH 8], 300 mM NaCl). Chromatograms and SDS-PAGE of the SEC are shown (Figure 45, Figure 46).

Size Exclusion Chromatography (SEC)

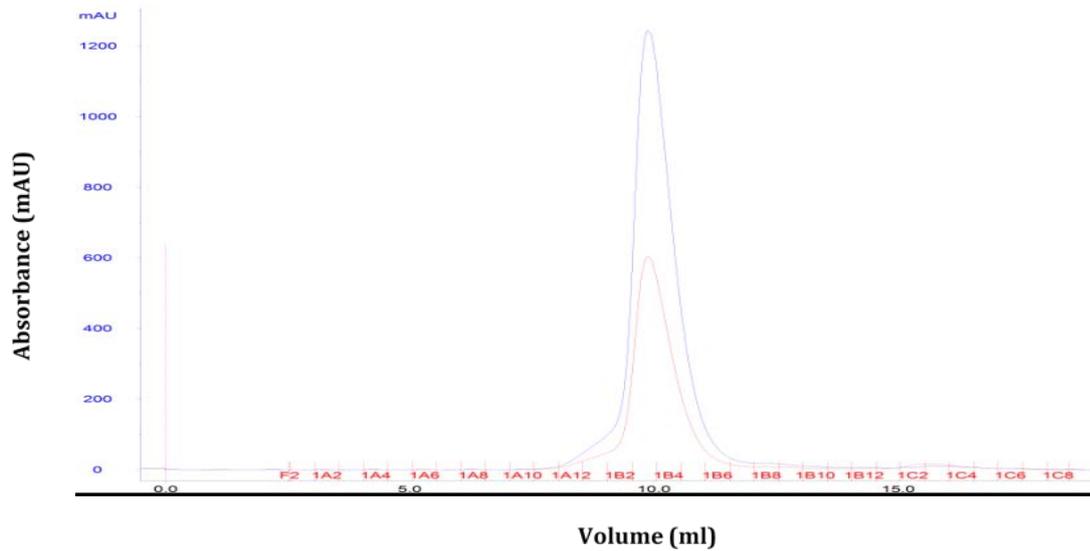


Figure 45. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C HCMV. The peak corresponds to the monomeric form of TRM3-C HCMV.



Figure 46. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C HCMV. The first three lanes correspond to the beginning of the peak where the amount of protein is lower and the last four lanes to the highest point of the peak where TRM3-C HCMV is most concentrated.

An ion exchange chromatography (Mono Q 5/50) with a lineal gradient of NaCl (from 50 mM NaCl to 1 M NaCl) was performed. Chromatograms and SDS-PAGE of the affinity chromatographies (Figure 49, Figure 50) are shown.

Ion exchange chromatography

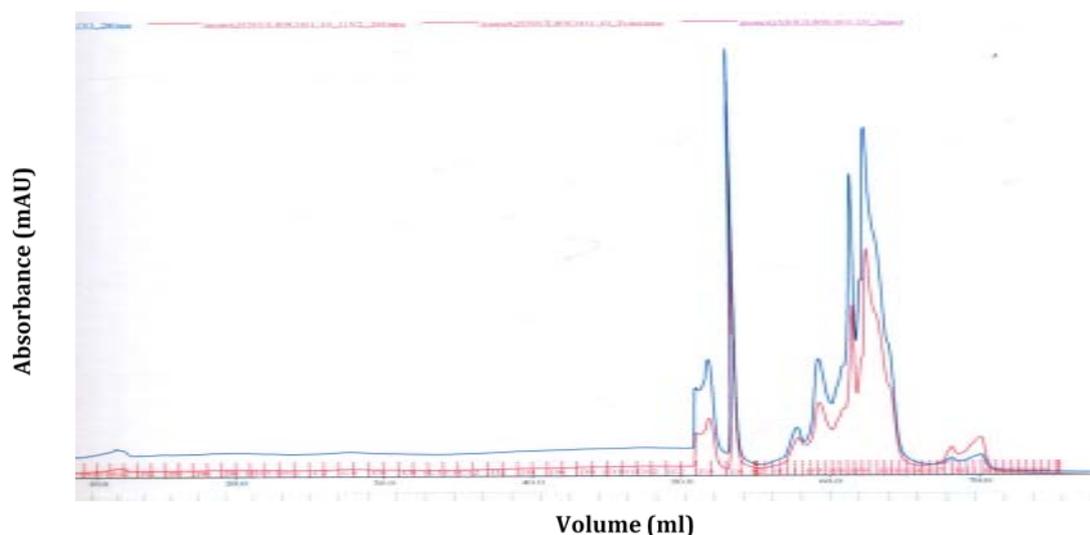


Figure 49. Mono Q 5/50 GL chromatographic profile of TRM3-C KSHV. The first part of the chromatogram corresponds to TRM3-C KSHV since it does not bind to the column and elutes in the flow-through. The peaks correspond to contaminant proteins that elute during the NaCl gradient.

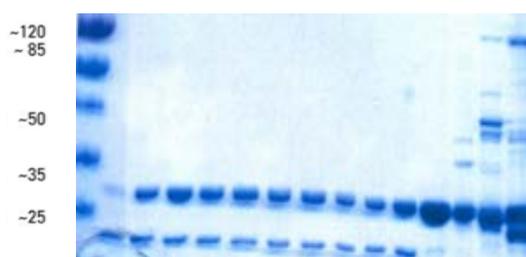


Figure 50. SDS-PAGE gel stained with Coomassie of the Mono Q 5/50 GL of TRM3-C KSHV. The first eleven lanes correspond to the flow-through fractions where the TRM3-C KSHV elutes together with a smaller contaminant protein. The last 3 lanes correspond to the peaks where most of the contaminant proteins elute.

A Heparin affinity chromatography was performed with a linear gradient of NaCl (from 250 mM NaCl to 1 M NaCl). Chromatograms and SDS-PAGE of the affinity chromatographies (Figure 51, Figure 52) are shown.

Adsorption chromatography

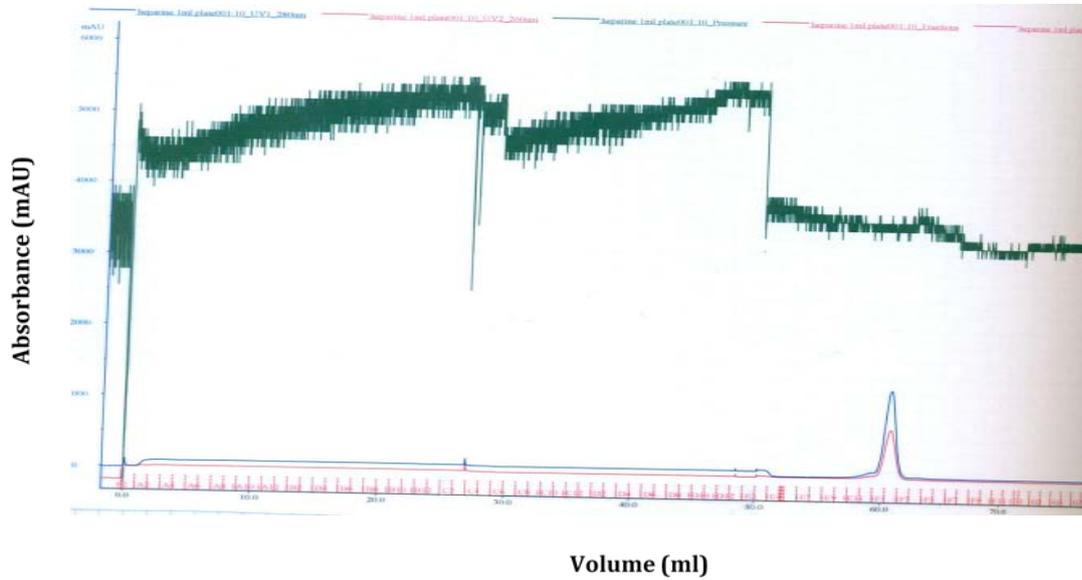


Figure 51. Heparin 1mL chromatographic profile of TRM3-C KSHV. The first part of the chromatogram corresponds to TRM3-C KSHV since it does not bind to the column and elutes in the flow-through. The peak corresponds to contaminant proteins that elute during the NaCl gradient.



Figure 52. SDS-PAGE gel stained with Coomassie of the Heparin affinity chromatography of TRM3-C KSHV. All the lanes correspond to flow-through fractions where TRM3-C KSHV elutes.

A fourth purification step with 10/300 Superdex 75 was performed (50 mM Hepes [pH 8], 300 mM NaCl). Chromatograms and SDS-PAGE of the SEC are shown (Figure 53, Figure 54).

Size Exclusion Chromatography (SEC)

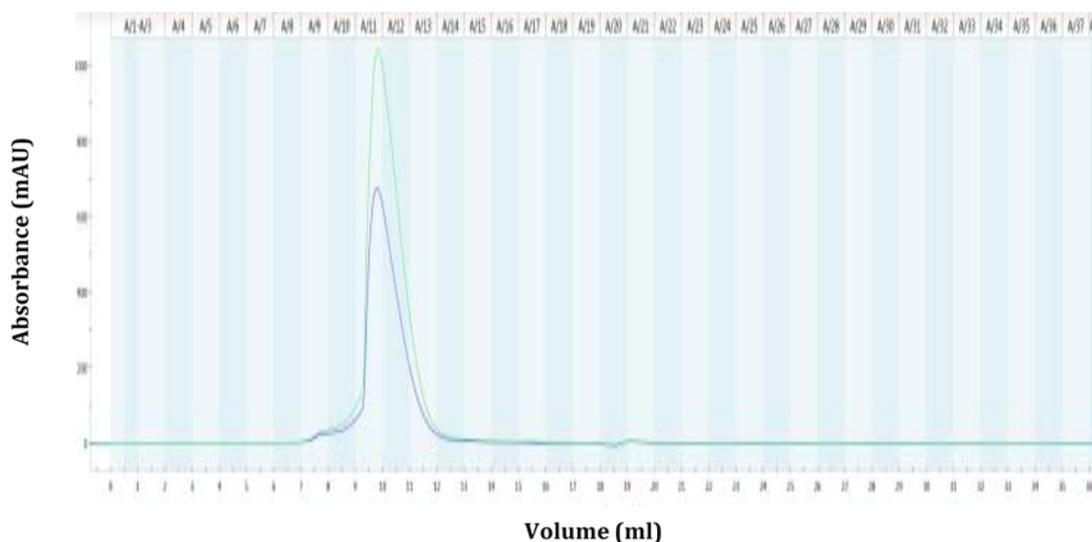


Figure 53. 10/300 Superdex 75 size-exclusion chromatographic profile of TRM3-C KSHV. The peak corresponds to the monomeric form of TRM3-C KSHV.

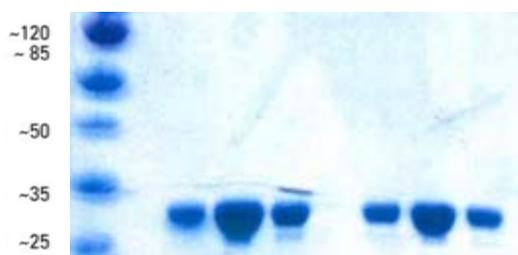


Figure 54. SDS-PAGE gel stained with Coomassie of the 10/300 Superdex 75 of TRM3-C KSHV. Since there was a lot of protein, two injections were needed and the gel shows both replicas. The three lanes of each replica correspond to TRM3-C KSHV monomeric form.

Not only high purity of the sample is crucial for crystallization, but also homogeneity of the sample is needed.

4.1.6. *In vitro* nuclease assays

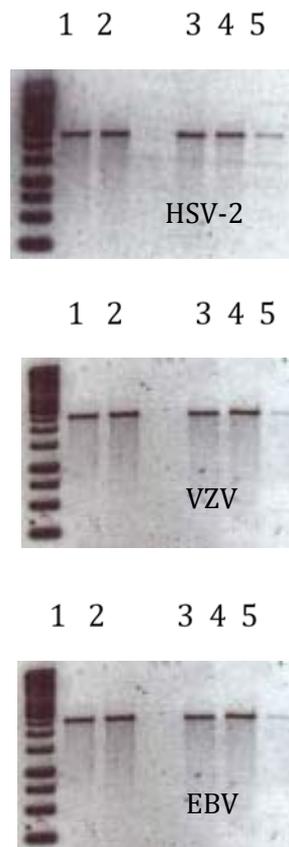
It has already been demonstrated by other laboratories that the full length TRM3-C HCMV protein has an endonuclease activity. A previous study reported that the full length protein was able to convert supercoiled plasmid DNA to a linear DNA

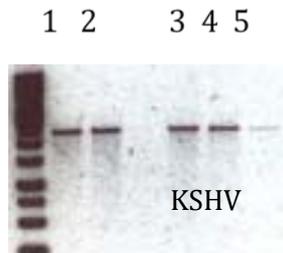
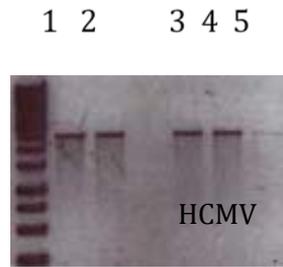
(Scheffczik *et al*, 2002). This experiment was done with the plasmid pUC18 as substrate (digested with Hind III) in order to quantify the amount of DNA that is cut and obtain an IC50 value for each inhibitor.

4.1.6.1. Ion preference

An ion preference experiment was performed in order to verify that TRM3-C HSV-2, TRM3-C VZV, TRM3-C HCMV, TRM3-C EBV and TRM3-C KSHV encode the nuclease activity. With the aim of finding the appropriate conditions for the nuclease reaction, different reactions with the presence of different ions (Mg^{2+} , Ca^{2+} and Mn^{2+}) were tested. The results were that the nuclease activity was strongly activated by Mn^{2+} ions with all the proteins that were tested. In the presence of Mn^{2+} , the DNA was completely degraded. The reaction done in the same conditions but in presence of Mg^{2+} or Ca^{2+} instead of Mn^{2+} shown that TRM3-C protein of HSV-2, EBV, EBV, HCMV and KSHV was less efficient, not being able to degrade the DNA (Figure 55).

In spite of the high sequence similarity between HSV-1 and HSV-2 (98%), the nuclease activity of HSV-1 was strongly activated by Mg^{2+} (Yang *et al*. 2020).



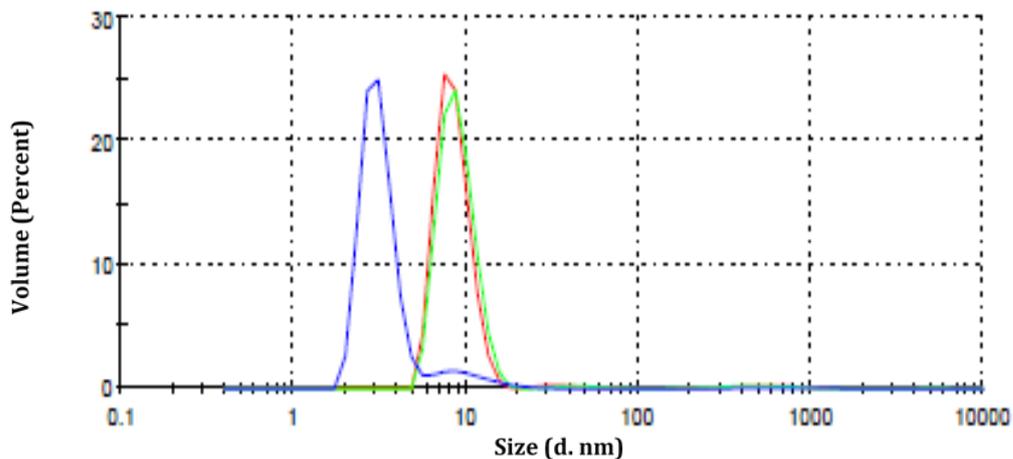


- 1: pUC18 digested with HindIII
- 2: Nuclease reaction in absence of divalent cations
- 3: Nuclease reaction in presence of Ca^{2+}
- 4: Nuclease reaction in presence of Mg^{2+}
- 5: Nuclease reaction in presence of Mn^{2+}

Figure 55. Agarose gel (1%) of TRM3-C HSV-2, VZV, EBV, HCMV and KSHV with the products of a nuclease assay against linear DNA.

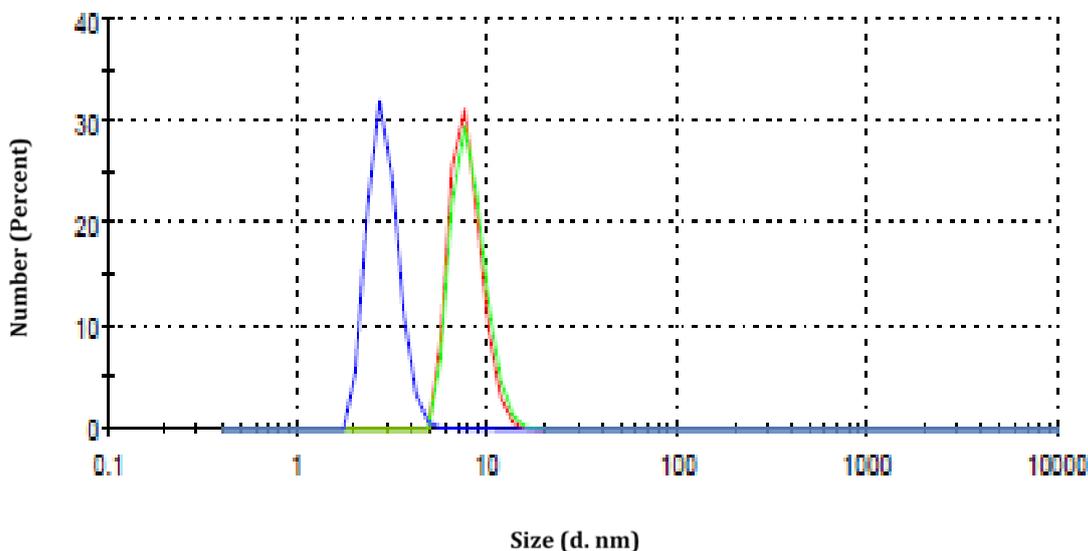
4.1.7. Protein quality analysis

One of the most important factors for achieving successful crystallization is the quality of the protein of interest. Biophysical properties were checked for quality assessment of the protein preparation.



_Record 1: TRM3-C EBV 1
 _Record 2: TRM3-C EBV 2
 _Record 3: TRM3-C EBV 3

Figure 56. Size distribution by volume of TRM3-C EBV measurements by dynamic light scattering.



_Record 1: TRM3-C EBV 1
 _Record 2: TRM3-C EBV 2
 _Record 3: TRM3-C EBV 3

Figure 57. Size distribution by number of TRM3-C EBV measurements by dynamic light scattering.

Dynamic light scattering is particularly well suited for evaluating homogeneity of protein solution. The volume and number distribution of nearly identical peaks from multiple measurements indicated that the final sample was highly pure although it contained both the dimer and monomer forms of TRM3-C EBV (Figure 56, Figure 57). These results were confirmed by Multi-angle light scattering (MALS). MALS profile of TRM3-C EBV showed a flat and constant M_w trace indicating homogeneous monodisperse species across the peak (Figure 58). According to ASTRA7 calculations they belong to monomer and dimer forms of the protein. ASTRA7 also indicated the fraction of each from the amount of protein calculated from dRI signal. The monomer amount is the 75% and the dimer fraction is 25%.

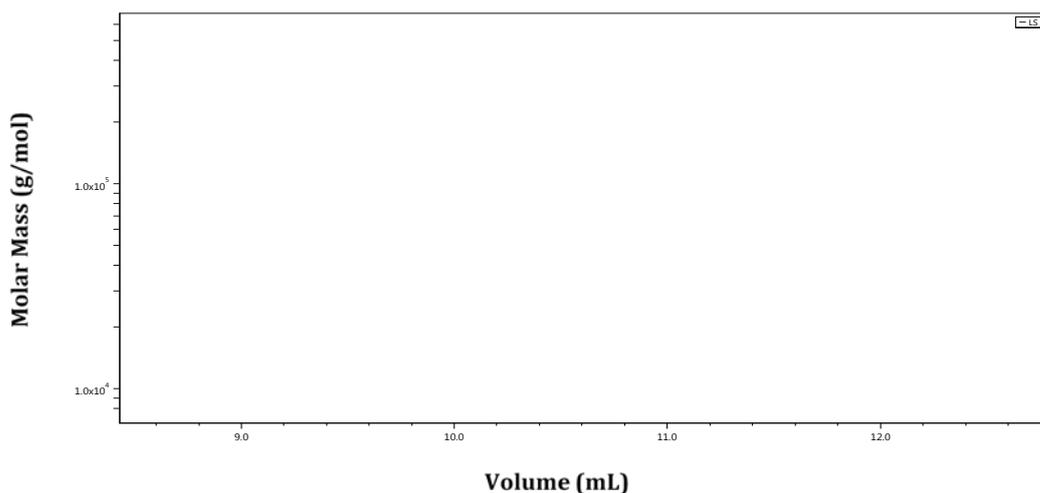


Figure 58. MALS profile of TRM3-C EBV concentrated sample in 10/300 Superdex 75 Increase.

Nuclear magnetic resonance spectroscopy can provide information on the global fold of proteins and verify the folded state. From the 1D ^1H NMR spectra, we determined that TRM3-C EBV protein was well-folded. The dispersion of the NMR signals in the regions of the methyl protons (-0.5 to 1.5 ppm) and amide protons (6-10 ppm) provides the main indicators of well-folded globular protein (Figure 59).

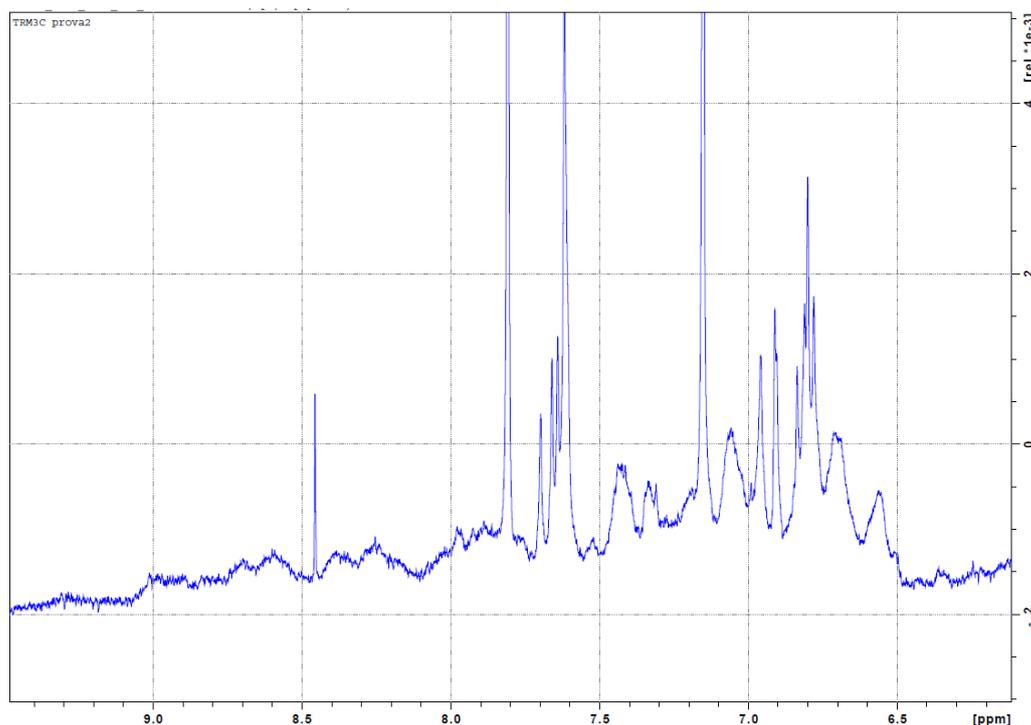


Figure 59. 1D ^1H nuclear magnetic resonance spectra of TRM3-C EBV.

4.1.1. Crystallization trials

Several crystallization screening plates (Table 15) were set up for TRM3-C HSV-2,

TRM3-C EBV and TRM3-C KSHV.

Code	Original screen	Nº of conditions
PAC 1	Crystal Screen I	48
	Crystal Screen II	48
PAC 2	Wizard Screen I	48
	Wizard Screen II	48
PAC 3	Index	96
PAC 4	Salt RX	96
PAC 5	A/S Ion Screen	48
	Ammonium sulfate	24
	Quick phosphate	24
PAC 6	PEG 6000	24
	PEG 6000/LiCl	24
	PEG 400	24
	PEG 4000/LiCl	24
PAC 7	Crystal Screen Lite	48
	Crystal Screen Cryo	48
PAC 8	PEG Ion Screen	48
	Grid MPD	24
	Grid NaCl	24
PAC 9	Matrix	48
	Complex Screen	48
PAC 11	MemFac	48
	MemPlus	48
PAC 12	MemStart	48
	MemSys	48
PAC 13	MemGold	96
PAC 14	Wizard Cryo I	48
	Wizard Cryo II	48
PAC 20	Citrate Screen	24
	Mc Pherson Screen	24
	Clear Strategy I	24
	Clear Strategy II	24
PAC 21	PACT premier	96
PAC 22	Pi- PEG Screen	96
PAC 23	Pi- Minimal Screen	96
PAC plus	JCSG plus	96
PAC TOP 96	TOP 96	96

Table 15. Automated Crystallography Platform screens.

No crystals were obtained with TRM3-C KSHV. TRM3-C HCMV was crystallized in the condition published by Nadal et al. 2012 (150mM Calcium Acetate Hydrate, 0.1 MES pH 6, 9 % w/v PEG 800). Crystal soaking experiments were performed by adding BS14 and BS17 inhibitors at different concentrations and incubation times. Crystals incubated during 24h and 48h with 5mM of BS14 (Figure 60) and BS17 (Figure 61)

were the most promising ones. Crystals with BS14 were yellow, indicating that the inhibitor could be incorporated. Nevertheless, the majority of the crystals did not diffract. Several data sets were collected with BS14 (Table 16, Table 17) and BS17 (Table 18, Table 19) but after processing them we noticed that the inhibitors were not incorporated into the active site of the protein.

Parameters	Values
Wavelength (Å)	0.97895 Å
Resolution Range	82.087 - 2.998 (3.11 - 2.99)
Space group	P2 ₁ 2 ₁ 2 ₁
Unit cell dimensions	a= 81.442 Å b= 87.207 Å c= 184.174 Å α= 90° β= 90° γ= 90°
Total reflections	80637 (4076)
Unique reflections	19641 (1007)
Multiplicity	4.1 (4.0)
Completeness (%)	96.1 (99.1)
Mean I/σ (I)	11.1 (1.2)
Rmerge(%)	7.9 (115.4)
Rmeas (%)	9.1 (132.7)
Rpim (%)	4.4 (64.0)
CC1/2 (%)	99.8 (62.0)

Table 16. Crystallographic data of TRM3-C HCMV with BS14 inhibitor.

Parameters	Values
Resolution	2,998 Å
Space group	P2 ₁ 2 ₁ 2 ₁
R work	0.3073
R free	0,2498
RMS (Angles)	1.22
RMS (Bonds)	0.013

Table 17. Refinement statistics for TRM3-C HCMV with BS14 inhibitor.

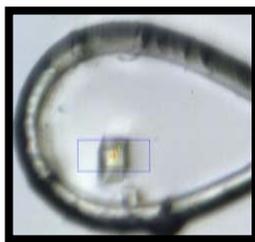


Figure 60. Soaking experiment of TRM3-C HCMV with BS14 inhibitor. Crystals grew in 24-well plates and were individually moved to new drops for the addition of the inhibitor.

Parameters	Values
Wavelength (Å)	0.97895 Å
Resolution Range	79.033 - 2.974 (3.06 - 2.97)
Space group	P2 ₁ 2 ₁ 2 ₁
Unit cell dimensions	a= 80.961 Å b= 86.987 Å c= 183.151 Å α= 90° β= 90° γ= 90°
Total reflections	77428 (5118)
Unique reflections	18323 (1245)
Multiplicity	4.8 (4.2)
Completeness (%)	97.3 (99.6)
Mean I/σ (I)	10.8 (1.7)
Rmerge(%)	8.4 (123.7)
Rmeas (%)	9.7 (143.8)
Rpim (%)	5.2 (71.8)
CC1/2 (%)	98.9 (73.2)

Table 18. Crystallographic data of TRM3-C HCMV with BS17 inhibitor.

Parameters	Values
Resolution	2,974 Å
Space group	P2 ₁ 2 ₁ 2 ₁
R work	0.3118
R free	0,2566
RMS (Angles)	1.26
RMS (Bonds)	0.010

Table 19. Refinement statistics for TRM3-C HCMV with BS17 inhibitor.

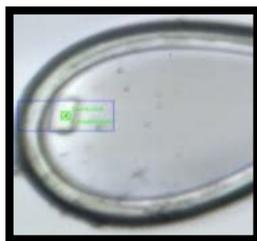


Figure 61. Soaking experiment of TRM3-C HCMV with BS17 inhibitor. Crystals grew in 24-well plates and were individually moved to new drops for the addition of the inhibitor.

Only very tiny crystals were obtained with TRM3-C EBV in one condition (1 M Ammonium Sulfate, 0.1 M HEPES pH 7, 0.5 % w/v PEG 800). Variation of crystallization conditions including protein concentration, pH, incubation temperature, addition of manganese etc. did not result in the growth of larger crystals. Crossed-microseeding (place a seed or solution of seeds of TRM3-C HCMV in a drop which is saturated to the Metastable Zone to grow larger single crystals) was also tried but we were not able to grow larger crystals. TRM3-C HCMV and TRM3-C EBV were superposed and the residues of the active centre were mutated in TRM3-C HCMV in order to obtain information about TRM3-C EBV active center.

- ❖ **F466Y**
- ❖ **T537S**
- ❖ **M579L**
- ❖ **G580T**

Once again, only tiny crystals that did not diffract were obtained (Figure 62).

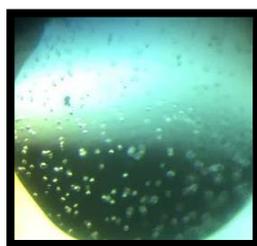


Figure 62. Tiny crystals of TRM3-C HCMV with mutated active centre.

Unfortunately, although protein analysis indicated a good quality of the protein sample it was not possible to determine the TRM3-C EBV structure due to the lack of larger crystals. We discarded by monodimensional NMR the presence of disordered regions as the cause of crystal growth inhibition. Our hypothesis is the monomer/dimer equilibrium to be the matter. Although chromatographies and dynamic light scattering seem to indicate that TRM3-C EBV behaves predominantly as a monomer, probably it is not fully stable. We suggest that the total protein concentration includes two oligomeric states, monomers and dimers. This heterogeneity could be the reason why we could not obtain better protein crystals.

Regarding TRM3-C HSV-2, several crystals of similar shape were obtained in 96-well plates that were optimized into 24-well optimization plates (Figure 63). The crystallization condition contained 0.2 M Lithium Sulfate Anhydre, 0.1 M Phosphate Citrate pH 4.2, 20 % w/v PEG 1000.

4.1.8. Data collection and structure determination

Optimal strategies for data collection are influenced by several factors. Some are general and others depend on the characteristics of a particular crystal or detector. Frozen crystals were diffracted at beamline BL13 – XALOC at ALBA synchrotron.

The best data collected diffracted up to 2.46 Å resolution (Figure 64), and it was indexed, integrated, scaled and reduced to unique reflections with XDS (Kabsch, 2010).



Figure 63. TRM3-C HSV-2 crystal grown in 24-well plate of the crystallization condition: 0.2 M Lithium Sulfate Anhydre, 0.1 M Phosphate Citrate pH 4.2, 20 % w/v PEG 1000.

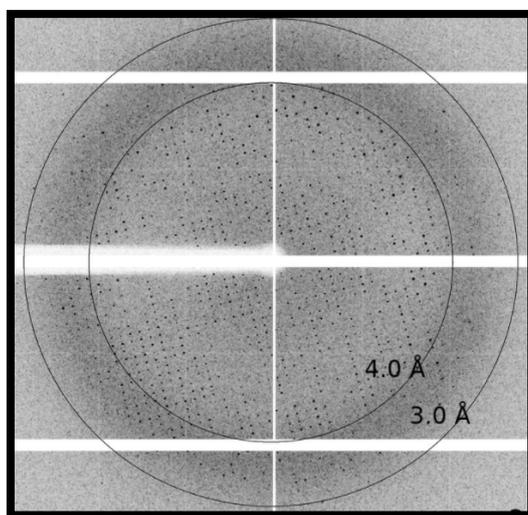


Figure 64. Diffraction pattern of TRM3-C HSV-2 crystal diffracting up to 2.46 Å.

The structure of TRM3-C HSV-2 was determined by molecular replacement using Phaser (McCoy *et al.*, 2007). The TRM3-C HSV-1 crystallographic structure solved at 2.46 Å (PDB: 4IOX) by Selvarajan Sigamani *et al.* (2013) was a suitable model to solve the phases by the molecular replacement method (Rossmann & Blow, 1962). Since the analysis of the data set revealed that the crystals suffered from both extreme

anisotropy and strong tNCS, extremely weak reflections with the largest combined effects of anisotropy and tNCS were eliminated. The asymmetric unit of these crystals contained four TRM3-C HSV-2 monomers. The Table 20 shows the processing statistics.

Parameters	Values
Wavelength (Å)	0.979178 Å
Resolution Range	56.235 - 2.46 (2.502 - 2.46)
Space group	P2 ₁
Unit cell dimensions	a= 75.111 Å b= 84.828 Å c= 89.715 Å $\alpha= 90^\circ$ $\beta= 90^\circ$ $\gamma= 90^\circ$
Total reflections	154534 (7817)
Unique reflections	21401 (1054)
Multiplicity	7.2 (7.4)
Completeness (%)	99.9 (100.0)
Mean I/σ (I)	7.2 (0.9)
Rmerge(%)	15.8 (238.0)
Rmeas (%)	17.1 (256.3)
Rpim (%)	6.3 (93.5)
CC1/2 (%)	99.8 (59.9)

Table 20. Crystallographic data and processing statistics of TRM3-C HSV-2. Overall and last shell (in parentheses) values are shown.

4.1.9. Structure refinement and validation

Once the atomic model has been obtained, it needs to be optimized to best fit the experimental data while also preserving good agreement with prior chemical knowledge.

Refmac (Murshudov et al., 2011) for 50 cycles of jelly-body refinement was run, using the pruned data and restraining the NCS copies. The map from the refinement ignoring twinning already showed some missing loops as well as parts of the structure that didn't have density in two of the four monomers that were in the asymmetric unit. Such missing fragments were rebuilt and refined again restraining the NCS copies. The final statistics are shown (Table 21).

Parameters	Values
Resolution (Å)	2.46
R-work/R-free	0.31637/0.3681
r.m.s.d. Bond lengths (Å)	0.0143
r.m.s.d Bond angles (°)	1.67
Ramachandran outliers (%)	0.85%
Ramachandran favored (%)	95.73%
Clashcore	14.27
Molprobit score	2.53

Table 21. Refinement statistics for TRM3-C HSV-2, as calculated with Molprobit

Rfree value remained quite high in spite of all the refinement cycles performed. This could reflect the fact that the crystal was not completely ordered, since the density of two of the four chains is not well defined. While monomers A and C have good density, monomers B and D do not. Each of those pair of chains are positioned one above the other (Figure 65). The fact that the Rfree value does not improve after many cycles of refinement usually happens when the crystal has tNCS and high anisotropy. In following sections the structure of monomer A will be detailed.

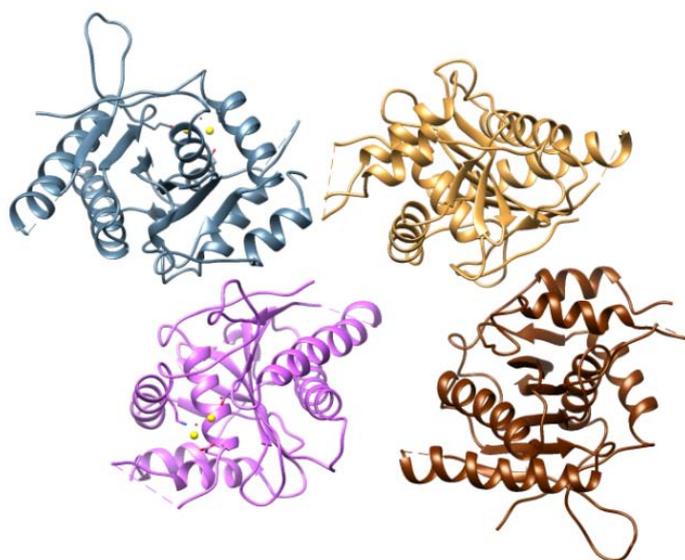


Figure 65. Cartoon representation of TRM3-C HSV-2 monomer protein. There are 4 molecules in the asymmetric unit. Monomer A (blue), monomer B (wheat), monomer C (pink), monomer D (brown).

All residues were located in allowed regions in the Ramachandran plot (Figure 66) excepting seven. These residues are located in loops, where density is poor due to flexibility.

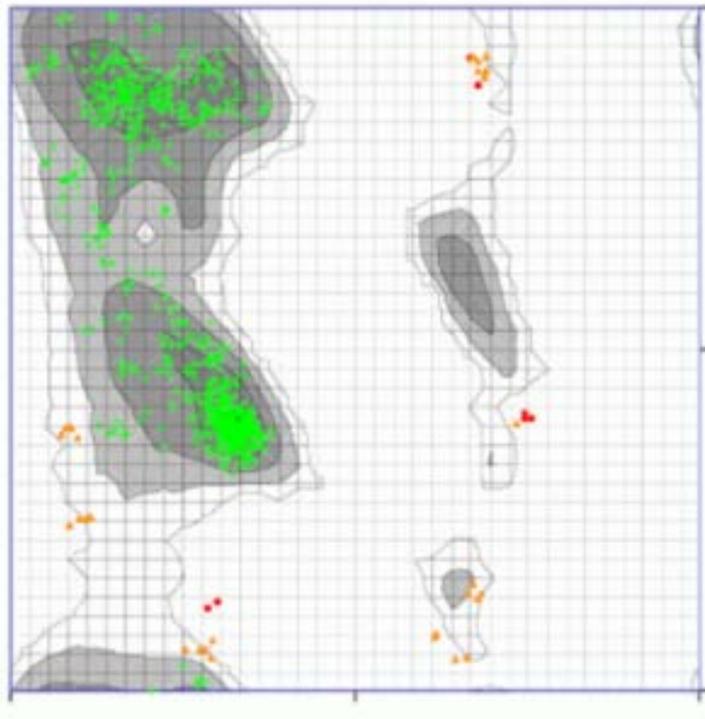


Figure 66. Ramachandran plot for the TRM3-C HSV-2. In green favoured regions (95.73%). In yellow allowed regions (3.42%). In red high energy backbone conformations (0.85%).

4.1.10. Structural analysis TRM3-C HSV2

TRM3-C HSV-2 contains a central eight-stranded β -sheet, with parallel and antiparallel strands, which is flanked by helices α on both sides (Figure 67). At one side, hydrophobic interactions pack $\alpha 2$ and $\alpha 3$ against the sheet. The strand order in the central sheet is 1, 9, 4, 3, 2, 5, 6, and 10 with topology $\beta 4_{-1, -1}, \beta 3_x, \beta 1_x, -5_x, \beta 6$ (Nadal *et al.* 2012). It shows loops of different sizes. At one end of the β sheet, long loops surround a cleft that typically harbors the active site in proteins sharing this fold.

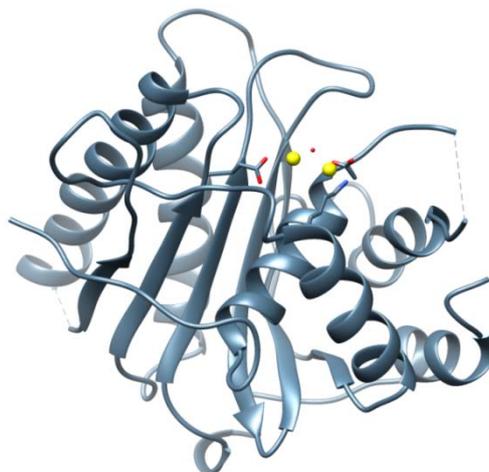


Figure 67. Cartoon representation of chain A of TRM3-C HSV2 monomer protein. In yellow Mn^{+2} . In red residues that interact with Mn^{+2} .

Overall fold of TRM3-C HSV-2 is the same as previously determined TRM3-C HCMV (Nadal et al. 2012). It has the characteristic fold of the RNase H-like superfamily of nucleases and polynucleotidyl transferases.

Superposition of both structures (Figure 68), TRM3-C HSV-2 and TRM3-C HCMV, match in alpha and beta strands but there are loops present in one structure that are not in the other.

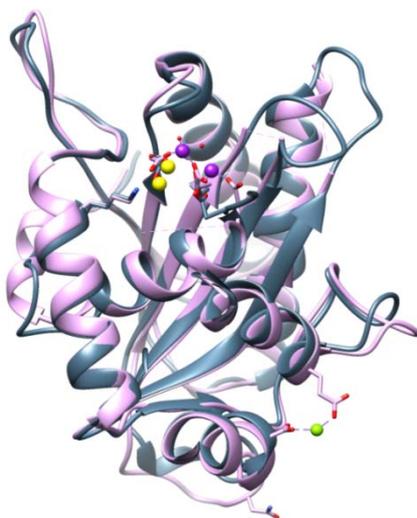


Figure 68. Superposition of chain A of TRM3-C HSV2 monomer (blue) and chain D of TRM3-C HCMV (purple). Mn^{+2} of TRM3-C HSV-2 in yellow and Mn^{+2} of TRM3-C HCMV in purple.

The active site cleft is found at one end of the β -sheet where two parallel β strands (β 2- β 5) separate. The electron density map shows the presence of two Mn^{+2} . Although Mn^{2+} was not included in the crystallization drops, in vitro assays showed that such cation strongly enhances the nuclease activity of TRM3. Our hypothesis is that such atoms come from the bacterial expression host and are maintained during the whole purification process due to its strong binding to the protein. Four aminoacids fully conserved amongst human infecting herpesvirus interact with the Mn^{+2} .

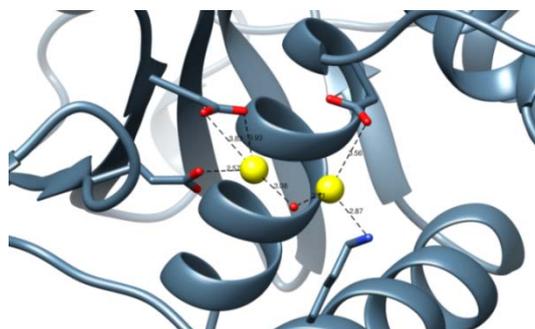


Figure 69. Zoom of the residues of TRM3-C HSV-2 active site (red) interacting with Mn^{+2} (yellow).

The active site is located at one end of the central β -sheet in a cleft formed by conserved residues. Asp461, Glu533, Lys639 and Asp663 coordinate two metal cations. Asp461 is located at the C-terminal end of β 2 whereas Glu533 is present at the end of β 5. Lys639 is present at the N-terminal of α 4. Asp663 is found at the

beginning of α_6 , the last α -helix in the structure, which lies diagonally to the two β -strands on one of the faces of the central β -sheet (Figure 69). These four amino acids are fully conserved and confer a strong electronegative character to the active site. There is also a H_2O molecule that interacts with both Mn^{+2} ions.

ATOM 1	ATOM2	DISTANCE
$Mn^{+2}688$	Glu533	2.57 Å
$Mn^{+2}687$	Asp663	3.56 Å
$Mn^{+2}687$	Lys639	2.87 Å
$Mn^{+2}688$	Asp461	3.83 Å
$Mn^{+2}687$	H_2O689	2.71 Å
$Mn^{+2}688$	H_2O689	3.08 Å

Table 22. Distances between the two metal cations (Mn^{+2}) and the residues of the active site of TRM3-C HSV-2.

Glu533 and Asp461 interact with $Mn^{+2}688$ while Asp663 and Lys639 interact with $Mn^{+2}687$. The H_2O689 molecule interacts with both cations, $Mn^{+2}687$ and $Mn^{+2}688$ (Table 22).

Previously solved TRM3-C HSV-1 structure showed no metal ions. However, TRM3-C HCMV structure Mn^{+2} soaked structure showed two Mn^{+2} cations in the active site cleft.

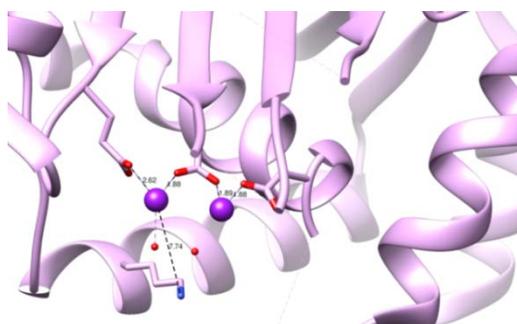


Figure 70. Zoom of the residues of TRM3-C HCMV active site (red) interacting with Mn^{+2} (purple).

Asp463, Glu534, and Asp651 coordinate two metal cations. Asp463 is located at the C-terminal end of β_2 whereas Glu534 is present at the end of β_5 . Asp651 is found at the beginning of α_6 , the last α -helix in the structure, which lies diagonally to the two β -strands on one of the faces of the central β -sheet (Figure 70). The Lys639 of TRM-3 HSV2 that corresponds to the Lys640 of TRM3-C HCMV does not interact with the Mn^{+2} because is at a distance of 7.74 Å from the active site. These three acidic amino acids are fully conserved and confer a strong electronegative character to the active site.

ATOM 1	ATOM2	DISTANCE
$Mn^{+2}675$	Asp651	1.88 Å
$Mn^{+2}675$	Asp463	1.89 Å

Mn ⁺² 676	Asp463	1.88 Å
Mn ⁺² 676	Glu534	2.62 Å

Table 23. Distances between the two metal cations (Mn⁺²) and the residues of the active site of TRM3-C HCMV.

Asp463 interacts with both Mn⁺² cations, Mn⁺²675 and Mn⁺²676. Asp651 interacts with Mn⁺²675 and Glu534 with Mn⁺²676 (Table 23).

Both structures were superposed in order to describe in more detail the position of the Mn⁺² and the side chain of the residues.

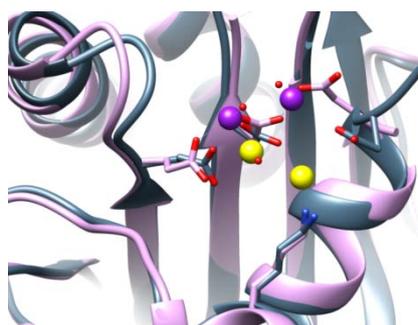


Figure 71. Zoom of the superposition of chain A of TRM3-C HSV2 monomer (blue) and chain D of TRM3-C HCMV (purple). Mn⁺² of TRM3-C HSV-2 in yellow and Mn⁺² of TRM3-C HCMV in purple.

The Asp663 of HSV-2 and the Asp651 of HCMV are maintained in both structures, as well as Glu533 of HSV-2 and Glu534 of HCMV. In the case of HSV-2 the Asp461 interacts with one Mn⁺² and the Asp463 of HCMV interacts with both of them. This is probably because the loop of HSV-2 is longer than the one of HCMV. The fact that the conserved Lys639 interacts with the Mn⁺² only in the case of HSV-2 could be due to the position of the Mn⁺² cations, which is not identical (Figure 71).

Interestingly Lys 639 is located in α -4 chain and interacts with TRM1 (Nadal *et al.* 2012) and in the structure of Yang *et al.* (2020) is positioned in the interface of the three terminase subunits.

4.1.11. Inhibition assays

The fact that the viral DNA encapsidation machinery has no counterpart in the mammalian cell implies that the proteins involved in this process represent promising selective targets for antiviral therapy.

Several studies have reported that inhibitors of DNA packaging in herpesviruses specifically target TRM3 and TRM1 although the binding regions to the protein have not been elucidated (Visalli & van Zeijl, 2003).

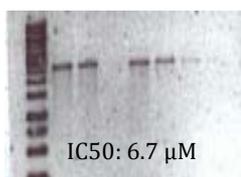
The structurally-related HIV integrase has been extensively studied as a drug target and there are several inhibitors of this enzyme in study. One of them, Raltegravir, was approved by the FDA in October 2007 (Summa *et al.* 2008). BS14 and BS17 inhibitors were designed in our laboratory by Salvatore Bongarzone and are optimizations of Raltegravir.

We have tested the effect of BS14 inhibitor, BS17 inhibitor, Bictegravir, Elvitegravir, Dolutegravir, MK-2048 inhibitor, 707035 inhibitor and Cabotegravir on the TRM3-C nuclease activity of HSV-2, VZV, EBV, HCMV and KSHV.

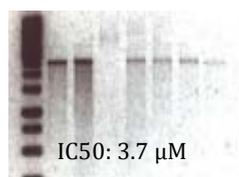
In our experiments, the concentration of TRM3-C HSV-2, TRM3-C VZV, TRM3-C EBV, TRM3-C HCMV and TRM3-C KSHV was always 2 μM . First, the optimal concentration range for determining the IC₅₀ of each compound was determined. The concentration of BS14 and BS17 increased from 2.5 to 15 μM (2.5, 5, 7.5, 10 μM). The concentration of Bictegravir, Elvitegravir and Dolutegravir increased from 10 to 40 μM (10, 20, 30, 40 μM). The concentration of Cabotegravir and 707035 inhibitor increased from 20 to 50 μM (20, 30, 40, 50 μM). The concentration of MK-2048 increased from 50 to 80 μM (50, 60, 70, 80 μM).

4.1.11.1. Herpes Simplex Virus 2 (HSV-2)

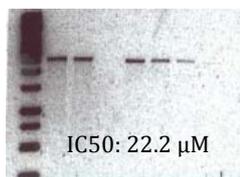
The best inhibitor of the nuclease activity of the TRM3-C of HSV-2 is BS17, with an IC₅₀ of 3.7 μM . The worst inhibitor is MK-2048 with an IC₅₀ of 67.4 μM (Figure 72).



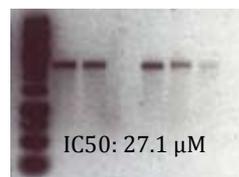
BS14 inhibitor



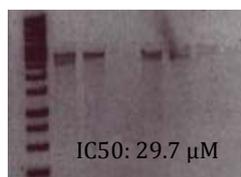
BS17 inhibitor



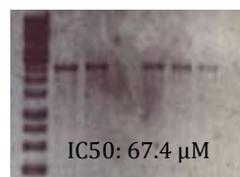
Bictegravir



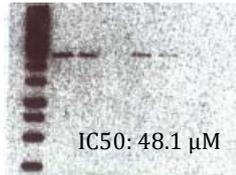
Elvitegravir



Dolutegravir



MK-2048 inhibitor



707035 inhibitor

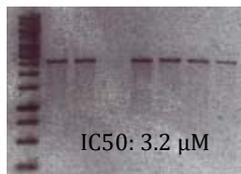


Cabotegravir

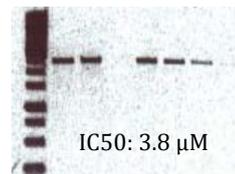
Figure 72. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C HSV-2. The first two lanes are controls. Control 1 corresponds to lineal DNA and control 2 to lineal DNA and the respective protein. The third is an empty lane. The next four lanes correspond to decreasing concentrations of the inhibitor according to the previously mentioned values.

4.1.11.2. Varicella-Zoster Virus (VZV)

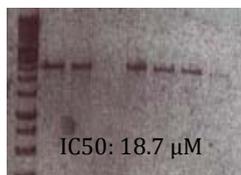
The best inhibitor of the nuclease activity of the TRM3-C of VZV is BS14, with an IC₅₀ of 3.2 μM. The worst inhibitor is MK-2048 with an IC₅₀ of 69.8 μM (Figure 73).



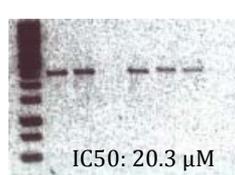
BS14 inhibitor



BS17 inhibitor



Bictegravir



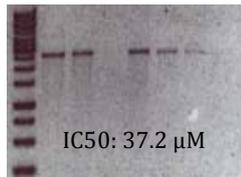
Elvitegravir



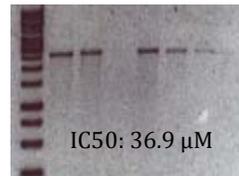
Dolutegravir



MK-2048 inhibitor



707035 inhibitor

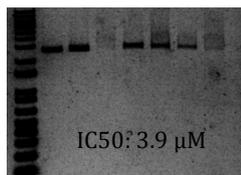


Cabotegravir

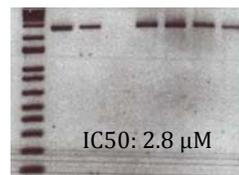
Figure 73. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C VZV. The first two lanes are controls. Control 1 corresponds to lineal DNA and control 2 to lineal DNA and the respective protein. The third is an empty lane. The next four lanes correspond to decreasing concentrations of the inhibitor according to the previously mentioned values.

4.1.11.3. Epstein-Barr virus (EBV)

The best inhibitor of the nuclease activity of the TRM3-C of EBV is BS17, with an IC₅₀ of 2.8 μM. The worst inhibitor is MK-2048 with an IC₅₀ of 64.2 μM (Figure 74).



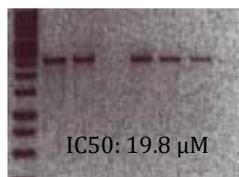
BS14 inhibitor



BS17 inhibitor



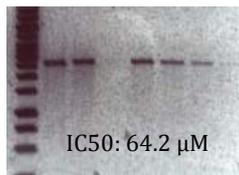
Bictegravir



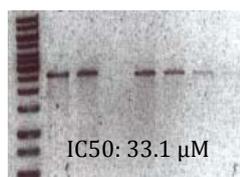
Elvitegravir



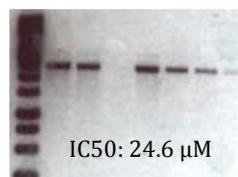
Dolutegravir



MK-2048 inhibitor



707035 inhibitor

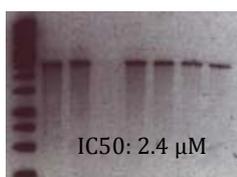


Cabotegravir

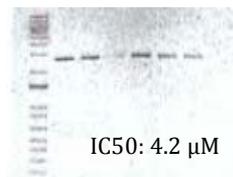
Figure 74. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C EBV. The first two lanes are controls. Control 1 corresponds to lineal DNA and control 2 to lineal DNA and the respective protein. The third is an empty lane. The next four lanes correspond to decreasing concentrations of the inhibitor according to the previously mentioned values.

4.1.11.4. Human Cytomegalovirus (HCMV)

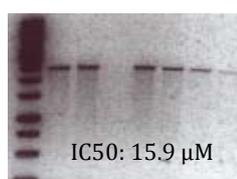
The best inhibitor of the nuclease activity of the TRM3-C of HCMV is BS14, with an IC₅₀ of 2.4 μM. The worst inhibitor is MK-2048 with an IC₅₀ of 60.7 μM (Figure 75).



BS14 inhibitor



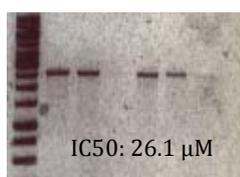
BS17 inhibitor



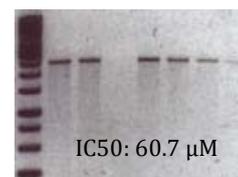
Bictegravir



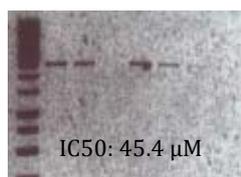
Elvitegravir



Dolutegravir



MK-2048 inhibitor



707035 inhibitor

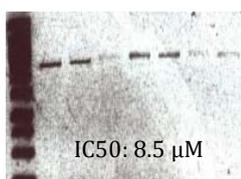


Cabotegravir

Figure 75. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C HCMV. The first two lanes are controls. Control 1 corresponds to lineal DNA and control 2 to lineal DNA and the respective protein. The third is an empty lane. The next four lanes correspond to decreasing concentrations of the inhibitor according to the previously mentioned values.

4.1.11.5. Kaposi's Sarcoma Herpesvirus (KSHV)

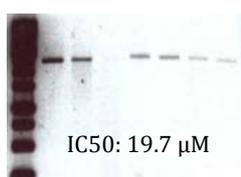
The best inhibitor of the nuclease activity of the TRM3-C of KSHV is BS17, with an IC₅₀ of 3.1 μM. The worst inhibitor is MK-2048 with an IC₅₀ of 61.3 μM (Figure 76).



BS14 inhibitor



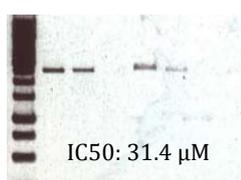
BS17 inhibitor



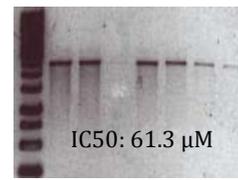
Bictegravir



Elvitegravir



Dolutegravir



MK-2048 inhibitor



Figure 76. Inhibition assays with BS14, BS17, Bictegravir, Elvitegravir, Dolutegravir, MK-2048, 707035 and Cabotegravir against TRM3-C KSHV. The first two lanes are controls. Control 1 corresponds to lineal DNA and control 2 to lineal DNA and the respective protein. The third is an empty lane. The next four lanes correspond to decreasing concentrations of the inhibitor according to the previously mentioned values.

4.1.11.6. Summary of results

We could see that the nuclease activity was inhibited proportionally to the concentration of inhibitor in all cases (Figure 72 to Figure 76).

In summary, BS17 inhibitor was able to inhibit the nuclease activity of TRM3-C HSV-2, TRM3-C EBV and TRM3-C KSHV in a lower micromolar range, while BS14 was able to inhibit the nuclease function of TRM3-C VZV and TRM3-C HCMV in a lower micromolar range. MK-2048 was the worst commercial inhibitor tested, showing small inhibition of the activity at 60 μM with all the proteins. Bictegravir was the best commercial inhibitor tested, showing inhibition of the nuclease activity of the terminase proteins up to 15 μM (Table 24).

	HSV-2	VZV	EBV	HCMV	KSHV
BS14	6.7 μM	3.2 μM	3.9 μM	2.4 μM	8.5 μM
BS17	3.7 μM	3.8 μM	2.8 μM	4.2 μM	3.1 μM
Bictegravir	22.2 μM	18.7 μM	15.4 μM	15.9 μM	19.7 μM
Elvitegravir	27.1 μM	20.3 μM	19.8 μM	23.1 μM	27.4 μM
Dolutegravir	29.7 μM	28.6 μM	27.2 μM	26.1 μM	31.4 μM
Cabotegravir	37.3 μM	36.9 μM	24.6 μM	32.6 μM	28.9 μM
707035	48.1 μM	37.2 μM	33.1 μM	45.4 μM	35.8 μM
MK-2048	67.4 μM	69.8 μM	64.2 μM	60.7 μM	61.3 μM

Table 24. IC50 values of each inhibitor in the presence of the TRM3-C terminal domain of all the herpesviruses tested: HSV-2, VZV, EBV, HCMV and KSHV.

4.1.12. Modeled structures with inhibitors

In order to analyze the interactions between the best inhibitors (BS14, BS17 and Bictegravir) and each protein (TRM3-C of HSV-2, VZV, EBV, HCMV and KSHV), we built the interacting models based on the structure solved of TRM3-C HCMV in complex with BS2 (Bongarzone *et al.*, 2017). Such structure was used as a guide regarding inhibitor orientation and distance respect to the Mn^{2+} positions.

The protein structures used to build such models were for HCMV (PDB: 3N4P) and for

the rest of the cases AlphaFold models. Although an experimental structure of HSV-2 was obtained, the different position of the Mn^{2+} atoms in the active site did not allow the modeling of the inhibitor in an equivalent manner to those previously observed experimentally in HCMV.

After the inhibitors were manually fitted in their positions some minimizations were required in the case of Bictegravir to avoid clashes with specific side chains (Lys639).

4.1.12.1. Herpes Simplex Virus 2 (HSV-2)

The interacting model of TRM3-C HSV-2 with Bictegravir shows the three oxygen atoms positioned toward the Mn^{2+} . Indeed, these oxygen atoms form hydrogen bonds with the Asp453, Glu542, Asp658 and Lys636 (Figure 77).

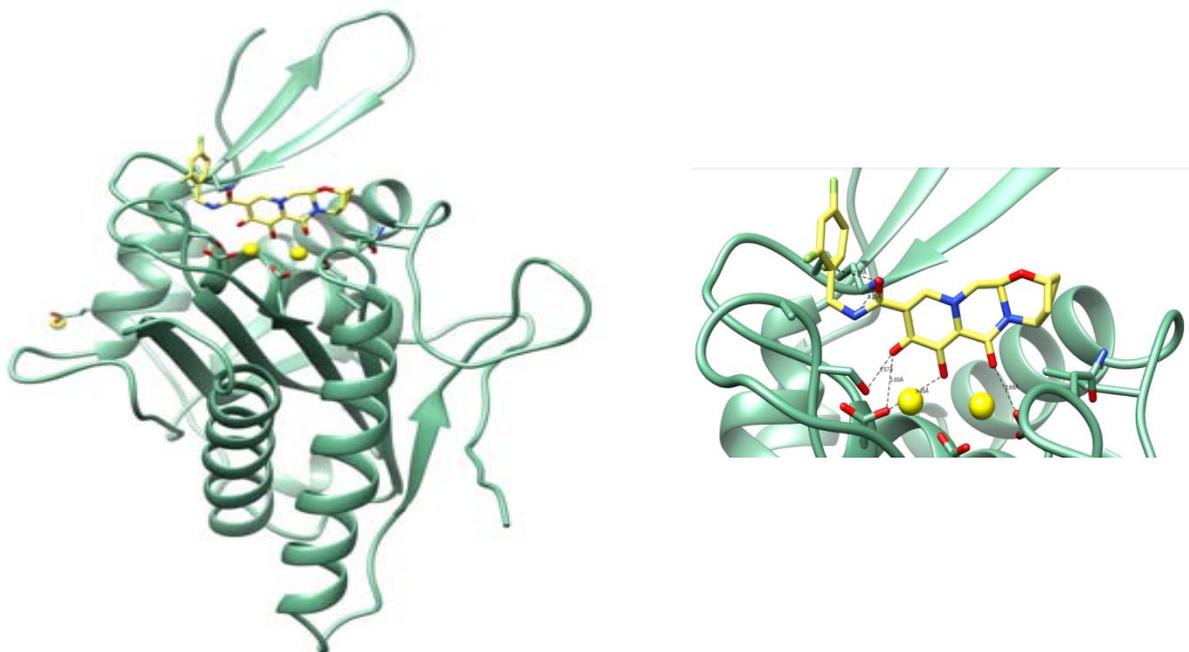


Figure 77. A) AlphaFold TRM3-C HSV-2 with Bictegravir. B) Interactions between Bictegravir and TRM3-C HSV2.

The interacting model of TRM3-C HSV-2 with BS17 and BS14 shows both inhibitors solvent exposed. The three coplanar oxygen atoms of the diketo acid (DKA) moiety chelate Mn^{2+} ions in the active site. The central oxygen, O (H), bridges both metal ions, whereas two O (carbonyl) coordinates the metal ions. Additionally, hydrogen bonds are formed between the three oxygen atoms and Asp453, Glu542, Asp658 (Figure 78, figure 79).

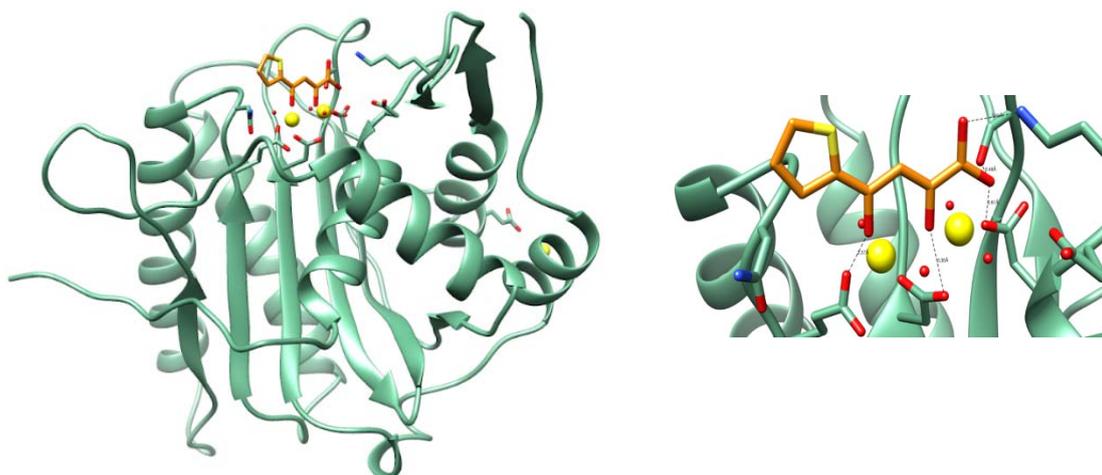


Figure 78. A) AlphaFold TRM3-C HSV-2 with BS17. B) Interactions between BS17 and TRM3-C HSV2.

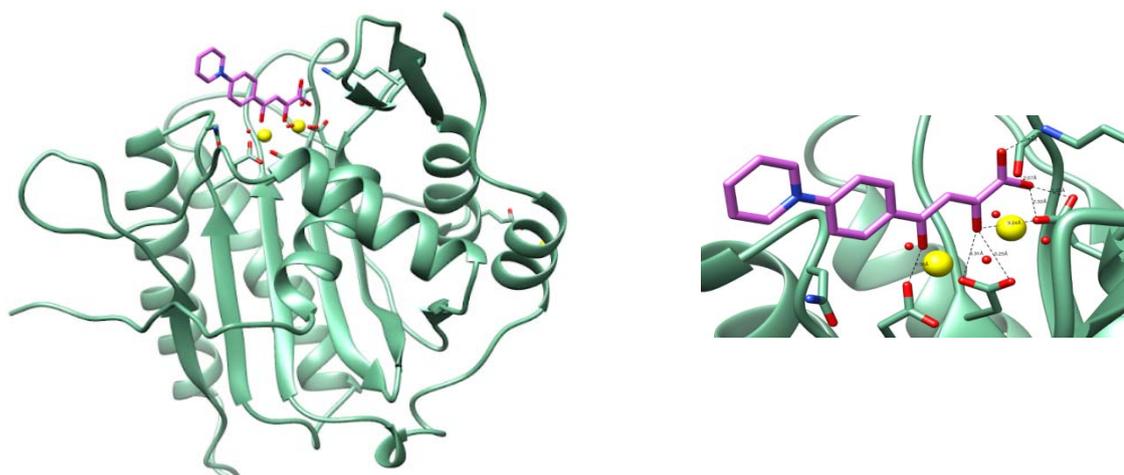


Figure 79. A) AlphaFold TRM3-C HSV-2 with BS14. B) Interactions between BS14 and TRM3-C HSV2.

4.1.12.2. Varicella-Zoster Virus (VZV)

The interacting model of TRM3-C VZV with Bictegravir shows the three oxygen atoms positioned toward the Mn^{2+} . Indeed, two of these oxygen atoms form hydrogen bonds with the Asp450 and Asp651. The nitrogen positioned near the trifluorobenzene is also forming a hydrogen bond (Figure 80).

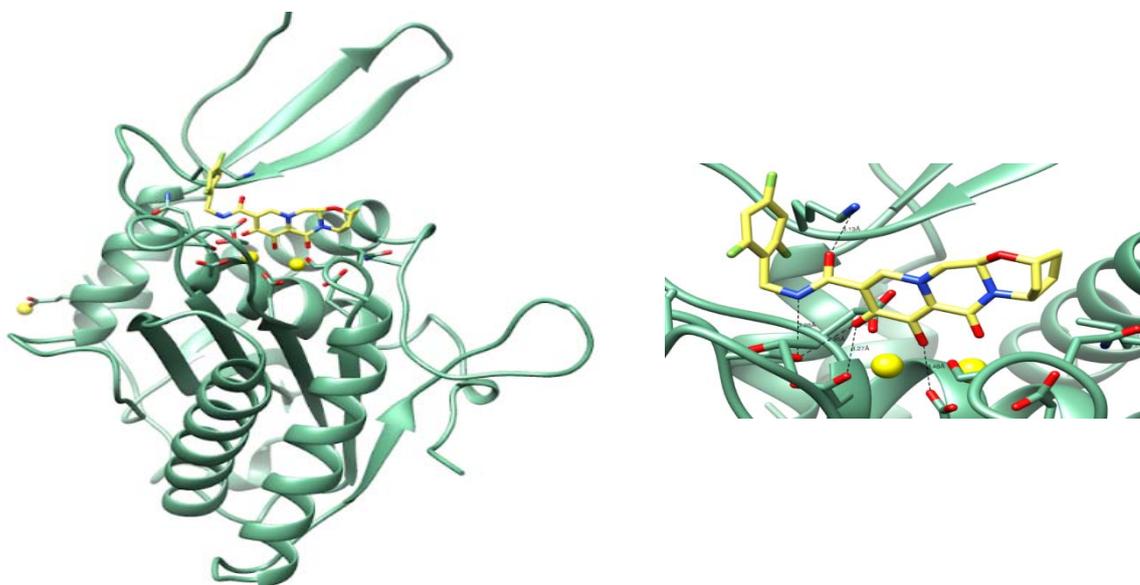


Figure 80. A) AlphaFold TRM3-C VZV with Bictegravir. B) Interactions between Bictegravir and TRM3-C VZV.

The interacting model of TRM3-C VZV with BS17 and BS14 shows both inhibitors solvent exposed. The three coplanar oxygen atoms of the diketo acid (DKA) moiety chelate Mn^{2+} ions in the active site. The central oxygen, O (H), and one O (carbonyl) form hydrogen bonds with Asp453 and Asp658 (Figure 81, Figure 82).

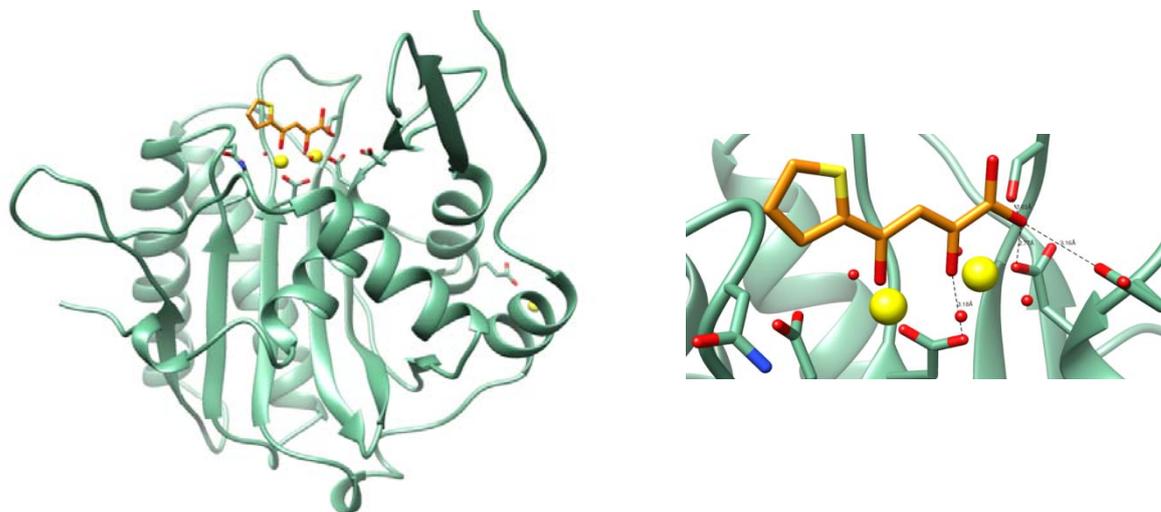


Figure 81. A) AlphaFold TRM3-C VZV with BS17. B) Interactions between BS17 and TRM3-C VZV.

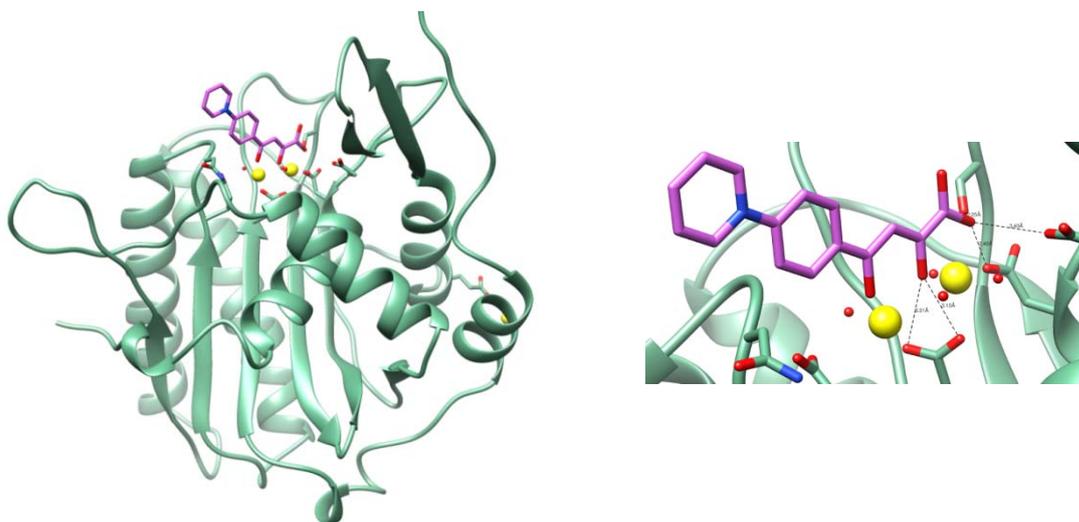


Figure 82. A) AlphaFold TRM3-C VZV with BS14. B) Interactions between BS14 and TRM3-C VZV.

4.1.12.3. Epstein-Barr virus (EBV)

The interacting model of TRM3-C EBV with Bictegravir shows the three oxygen atoms positioned toward the Mn^{2+} . Indeed, two of these oxygen atoms form hydrogen bonds with the Glu539 and Asp649. The nitrogen positioned near the trifluorobenzene is also forming a hydrogen bond with Glu539 (Figure 83).

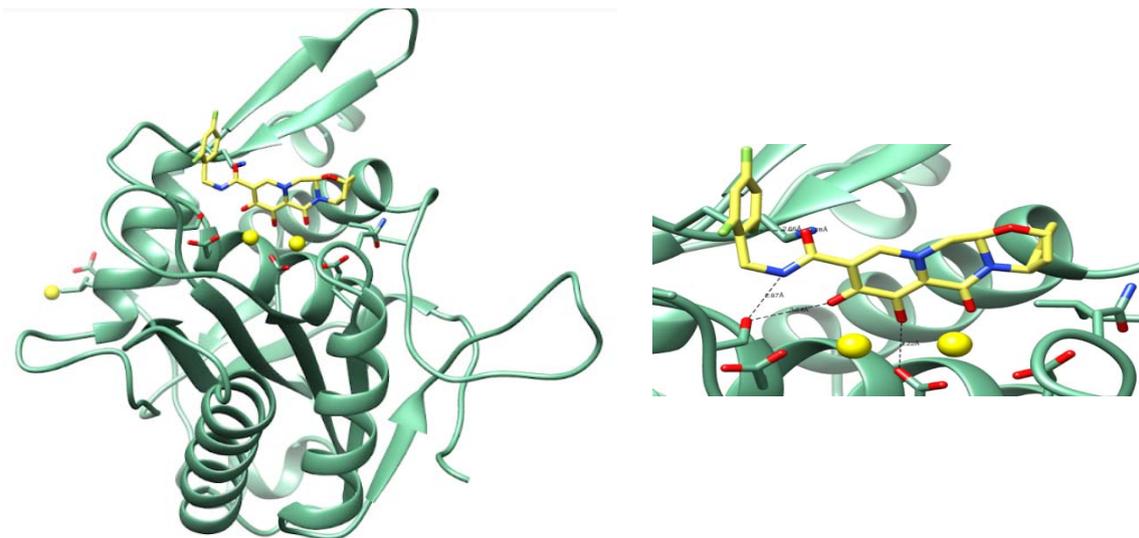


Figure 83. A) AlphaFold TRM3-C EBV with Bictegravir. B) Interactions between Bictegravir and TRM3-C EBV.

The interacting model of TRM3-C VZV with BS17 and BS14 shows both inhibitors solvent exposed. The three coplanar oxygen atoms of the diketo acid (DKA) moiety chelate Mn^{2+} ions in the active site. The central oxygen, O (H), and one O (carbonyl) form hydrogen bonds with Asp451 and Asp655 (Figure 84, Figure 85).

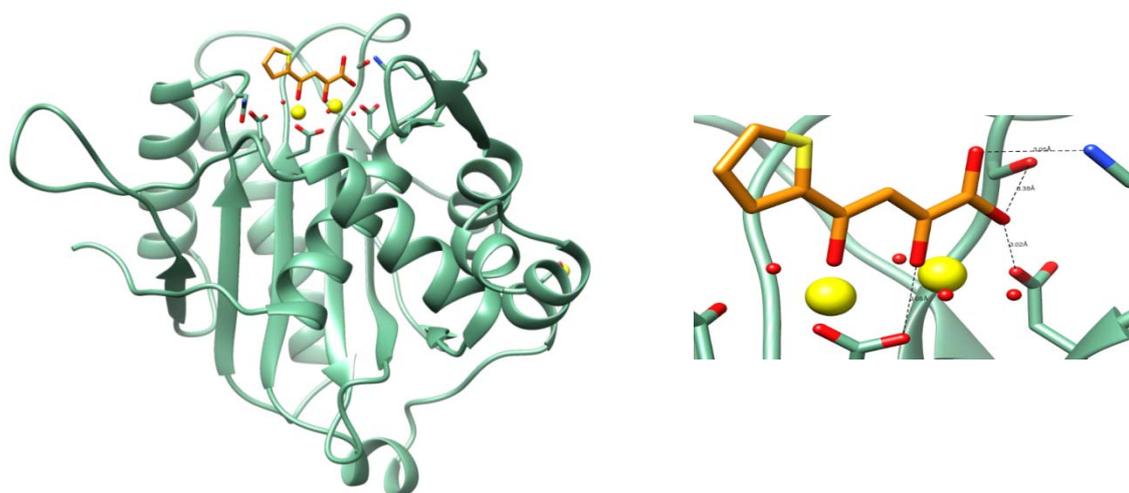


Figure 84. A) Alphafold TRM3-C EBV with BS17. B) Interactions between BS17 and TRM3-C EBV.

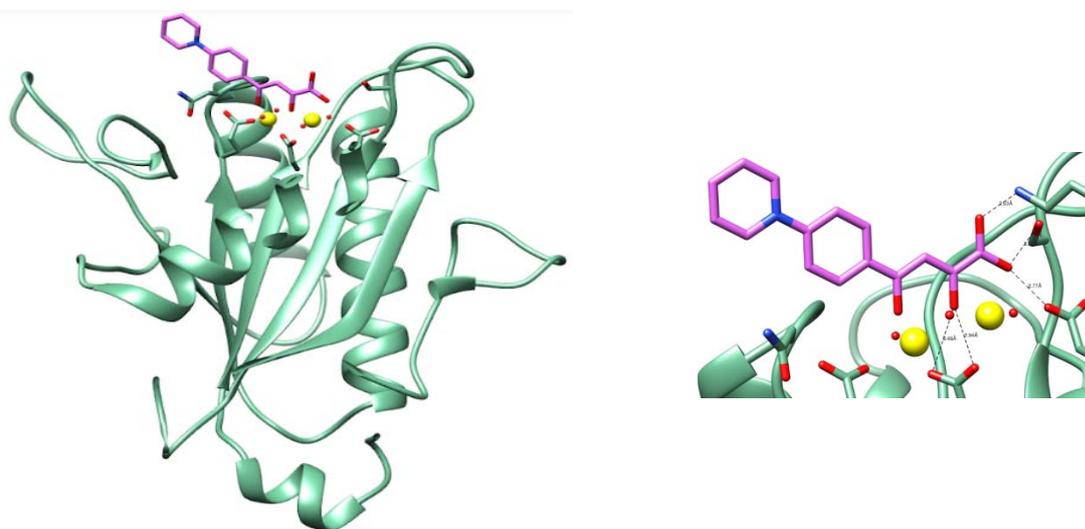


Figure 85. A) Alphafold TRM3-C EBV with BS14. B) Interactions between BS14 and TRM3-C EBV.

4.1.12.4. Kaposi's Sarcoma Herpesvirus (KSHV)

The interacting model of TRM3-C KSHV with Bictegrovir shows the three oxygen atoms positioned toward the Mn^{2+} . Indeed, these oxygen atoms form hydrogen bonds with the Asp449, Glu540, Asp657 and Lys633. The nitrogen positioned near the trifluorobenzene is also forming a hydrogen bond with Glu536 (Figure 86).

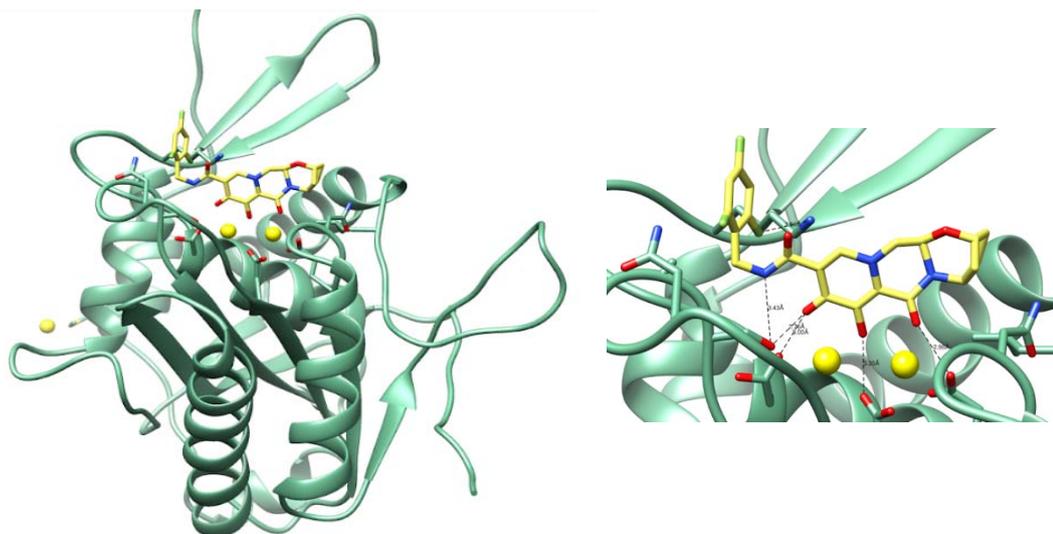


Figure 86. A) AlphaFold TRM3-C KSHV with Bictegrovir. B) Interactions between Bictegrovir and TRM3-C KSHV.

The interacting model of TRM3-C KSHV with BS17 and BS14 shows both inhibitors solvent exposed. The three coplanar oxygen atoms of the diketo acid (DKA) moiety chelate Mn^{2+} ions in the active site. The central oxygen, O (H), bridges both metal ions, whereas two O (carbonyl) coordinates the metal ions. Additionally, hydrogen bonds are formed between the three oxygen atoms and Asp449, Glu547, Asp652 (Figure 87, figure 88).

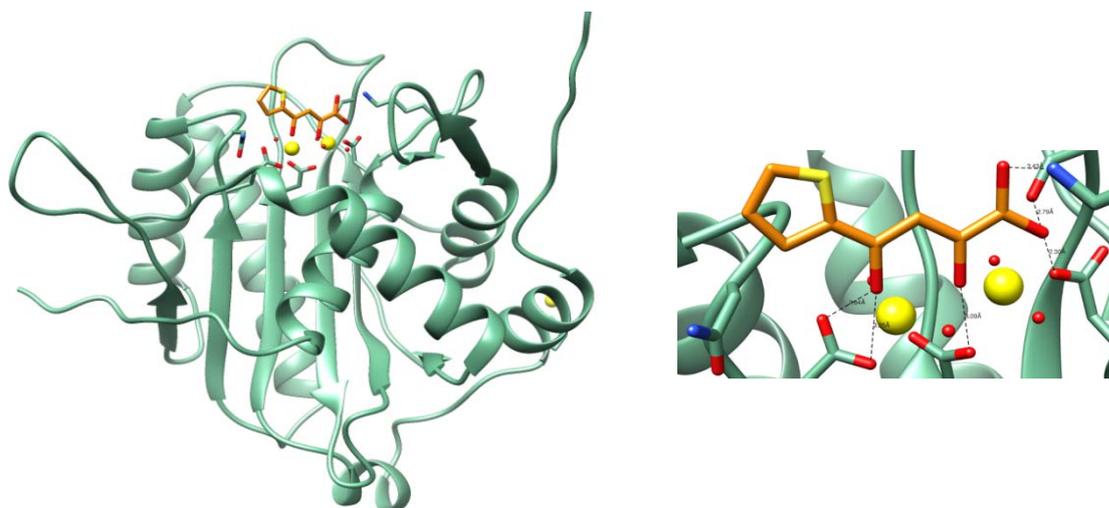


Figure 87. A) AlphaFold TRM3-C KSHV with BS17. B) Interactions between BS17 and TRM3-C KSHV.

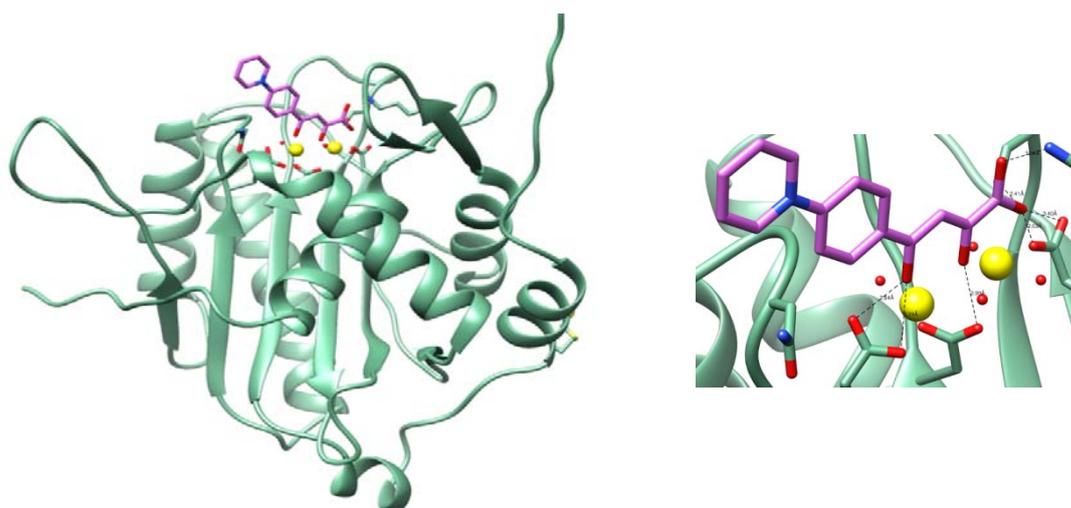


Figure 88. A) AlphaFold TRM3-C KSHV with BS14. B) Interactions between BS14 and TRM3-C KSHV.

4.1.13. In vitro antiviral screening report

The optimized inhibitors developed in the laboratory (BS14 and BS17) were sent to the NIH in order to perform an in vitro antiviral screening. The effectiveness of BS14 (Table 25) and BS17 (Table 26) was tested in HSV-1, HSV-2, HCMV, VZV, EBV, HHV6-B and KSHV.

Virus (strain)	Cell line	Drug assay	EC ₅₀	EC ₉₀	CC ₅₀	SI ₅₀	SI ₉₀
HSV-1 (E-377)	HFF	CellTiter-Glo (Cytopathic effect/Toxicity)	>30.00	>30.00	137.41	<5	<5

HSV-2 (G)	HFF	CellTiter-Glo (Cytopathic effect/Toxicity)	>30.00	>30.00	147.57	<5	<5
HCMV (AD169)	HFF	CellTiter-Glo (Cytopathic effect/Toxicity)	4.74	>30.00	127.28	27	<4
VZV (Ellen)	HFF	CellTiter-Glo (Cytopathic effect/Toxicity)	>30.00	>30.00	130.11	<4	<4
EBV (Akata)	Akata	Quantitative polymerase chain reaction (DNA)/ CellTiter-Glo (Toxicity)	93.09	>100.00	>100.00	>1	1
HHV6-B (Z29)	MOLT-3	Quantitative polymerase chain reaction (DNA)/ CellTiter-Glo (Toxicity)	70.70	90.30	>100.00	>1	>1
KSHV (BCBL-1)	BCBL-1	Quantitative polymerase chain reaction (DNA)/ CellTiter-Glo (Toxicity)	49.74	89.86	>100.00	>2	>1

Table 25. *In vitro* antiviral screening of BS14 against HSV-1, HSV-2, VZV, HCMV, EBV, HHV6-B and KSHV.

BS14 appeared to be moderately active against HCMV. As seen in Table 25, the SI₅₀ is 27, what means that could be an effective and safe drug during in vivo treatment.

Virus (strain)	Cell line	Drug assay	EC₅₀	EC₉₀	CC₅₀	SI₅₀	SI₉₀
HSV-1 (E-377)	HFF	CellTiter-Glo (Cytopathic effect/Toxicity)	>150.00	>150.00	>150.00	1	1
HSV-2 (G)	HFF	CellTiter-Glo (Cytopathic effect/Toxicity)	>150.00	>150.00	>150.00	1	1
HCMV (AD169)	HFF	CellTiter-Glo (Cytopathic effect/Toxicity)	>150.00	>150.00	>150.00	1	1
VZV (Ellen)	HFF	CellTiter-Glo (Cytopathic effect/Toxicity)	>150.00	>150.00	>150.00	1	1
EBV (Akata)	Akata	Quantitative polymerase chain reaction (DNA)/ CellTiter-Glo (Toxicity)	>100.00	>100.00	>100.00	1	1
HHV6-B (Z29)	MOLT-3	Quantitative polymerase	>100.00	>100.00	>100.00	1	>1

		chain reaction (DNA)/ CellTiter-Glo (Toxicity)					
KSHV (BCBL-1)	BCBL-1	Quantitative polymerase chain reaction (DNA)/ CellTiter-Glo (Toxicity)	>100.00	>100.00	>100.00	1	1

Table 26. *In vitro* antiviral screening of BS17 against HSV-1, HSV-2, VZV, HCMV, EBV, HHV6-B and KSHV.

4.1.14. Design of novel terminase inhibitors

Considering all the biochemical and structural data, new compounds were designed to obtain better inhibitors against the TRM3-C target. Since the DNA is not present in the solved structure of the HSV-2 TRM3-C terminal domain or in any previously solved structure of its homologs, the design of new inhibitors remains a great challenge. Indeed, the active site is wide and poorly defined. For this reason we decided to enlarge the best inhibitors we had (BS14 and BS17) in order to obtain better ones.

Our design was partially based in the structure of the best commercial inhibitor tested, Bictegravir, but keeping the three coplanar oxygen atoms of the diketo acid (DKA) present in BS14 and BS17, which were able to inhibit the nuclease activity of the TRM3-C terminal domain in a low micromolar range. These oxygens chelate Mn^{2+} ions in the active site, explaining the requirement for metal ions for drug binding. The central O bridges both metal ions, whereas each of the two adjacent O coordinates one metal ions. All three inhibitors proposed keep the three oxygens present in the α,γ -DKA motif linked to the same ring as in BS14, according to its better performance respect to BS17 in the results of the cellular tests.

Terminase inhibitor 1 is the smallest of the proposed inhibitors and the only change with respect to BS14 is that the part of the inhibitor containing the three chelating oxygens is now cycled in two additional rings, in a likewise manner to that present in Bictegravir equivalent region (Figure 89).

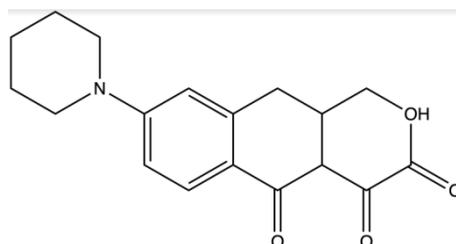


Figure 89. Terminase inhibitor 1.

All the changes introduced in terminase inhibitor 1 are maintained in terminase inhibitor 2. An additional moiety present in bicitegravir which might interact with the DNA was added (Figure 90).

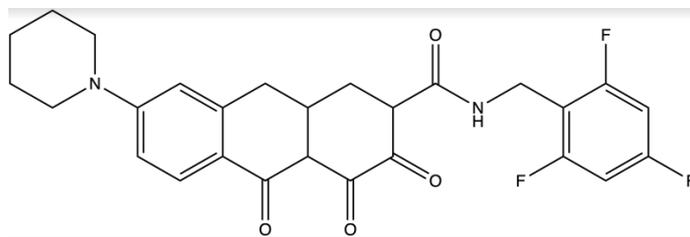


Figure 90. Terminase inhibitor 2.

Finally, terminase inhibitor 3 has two additional rings with respect to terminase inhibitor 1, with the same aim of establishing interactions with the DNA substrate (Figure 91).

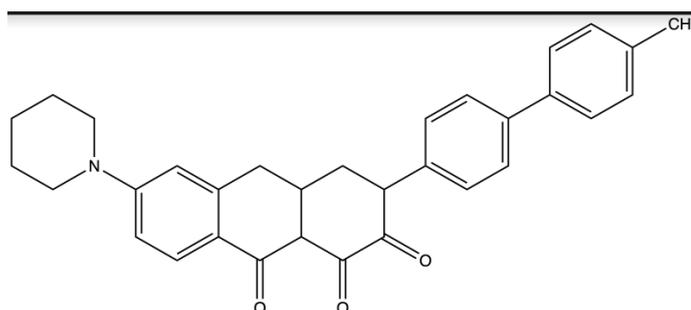


Figure 91. Terminase inhibitor 3.

CHAPTER 5: CONCLUSIONS

1. TRM3-C soluble domain equivalent to those previously identified from HCMV were cloned for HSV-2, VZV and EBV. Such domains, together with the previously cloned ones from HCMV and KSHV were successfully expressed in *E. coli* Rosetta 2 (DE3) cells.
2. TRM3-C HSV-2, TRM3-C HCMV and TRM3-C EBV were purified in quantities suitable for crystallization.
3. TRM3-C HCMV was crystallized and soakings with BS14 and BS17 were performed. Unfortunately, the collected data sets showed that the inhibitors were not incorporated.
4. TRM3-C EBV was crystallized but only tiny crystals grew. Microseeding experiments were performed but we were not able to grow larger crystals. TRM3-C HCMV and TRM3-C EBV were superposed and the residues of the active centre were mutated in TRM3-C HCMV in order to obtain information about TRM3-C EBV active centre. Tiny crystals that didn't diffract were obtained.
5. TRM3-C HSV-2 was crystallized in the space group $P2_1$. The crystals diffracted at 2.46 Å at the ALBA synchrotron with overall good quality statistics.
6. Although the dataset presented anisotropy and tNCS pathology it was possible to solve the structure of TRM3-C HSV-2 by molecular replacement using Phaser. The TRM3-C HSV-1 crystallographic structure solved by Selvarajan Sigamani et al. (2013) was a suitable model to solve the phases by the molecular replacement method.
7. The asymmetric unit of TRM3-C HSV-2 crystals consists of 4 monomers.
8. The TRM3-C HSV-2 three-dimensional structure was solved in complex with Mn^{2+} . Although Mn^{2+} was not included in the crystallization drops, such atoms come from the bacterial expression host and are maintained during the whole purification process due to its strong binding to the protein.
9. The position of the Mn^{2+} atoms in the active site of TRM3-C HSV-2 is different to the position previously observed experimentally in HCMV.
10. In vitro assays showed that Mn^{2+} cation strongly enhances the nuclease activity of TRM3-C terminal domain of HSV-2, VZV, EBV, HCMV and KSHV.
11. BS14 is the best inhibitor of the TRM3-C terminal domain of VZV (IC₅₀: 3.2 μM) and HCMV (IC₅₀: 2.4 μM).
12. BS17 is the best inhibitor of the TRM3-C terminal domain of HSV-2 (IC₅₀: 3.7 μM), EBV (IC₅₀: 2.8 μM) and KSHV (IC₅₀: 3.1 μM).

13. In vitro antiviral screening showed that BS14 was moderately active against HCMV, meaning that could be an effective and safe drug during in vivo treatment.
14. Considering all the biochemical and structural data, new compounds were designed to obtain better inhibitors against the TRM3-C target.

BIBLIOGRAPHY

Adler, B., Sattler, C. & Adler, H. Herpesviruses and Their Host Cells: A Successful Liaison. *Trends in Microbiology* (2017).

Baines J.D. & Weller S.K. Cleavage and Packaging of Herpes Simplex Virus 1 DNA. Viral Genome Packaging Machines: Genetics, Structure, and Mechanism. *Molecular Biology Intelligence Unit* (2005).

Bongarzone, S., Nadal, M., Kaczmarek, Z., Machón, C., Álvarez, M., Albericio, F., & Coll, M. Structure-Driven Discovery of α,γ -Diketoacid Inhibitors Against UL89 Herpesvirus Terminase. *ACS omega* (2018).

Booy, F.P., Newcomb, W.W., Trus, B.L., Brown, J.C., Baker, T.S. & Steven, A.C. Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus. *Cell* (1991).

Brown, J. C., & Newcomb, W. W. Herpesvirus capsid assembly: insights from structural analysis. *Current opinion in virology* (2011).

Champier, G., Hantz, S., Couvreur, A., Stuppfler, S., Mazon, M.C., Bouaziz, S., Denis, F. & Alain, S. New functional domains of human cytomegalovirus pUL89 predicted by sequence analysis and three-dimensional modelling of the catalytic site DEXDc. *Antiviral therapy* (2007).

Chisholm, C. & Lopez, L. Cutaneous infections caused by Herpesviridae: a review. *Archives of Pathology & Laboratory Medicine* (2011).

Clark, D.A. Clinical and laboratory features of human herpesvirus 6 chromosomal integration. *Clinical Microbiology and Infection* (2016).

Cook, N.J., Li, W., Berta, D., Badaoui, M., Ballandras-Colas, A., Nans, A., Kotecha, A., Rosta, E., Engelman, A.N. & Cherepanov, P. Structural basis of second-generation HIV integrase inhibitor action and viral resistance. *Science*. (2020)

Davison, J. Mechanism of control of DNA replication and incompatibility in ColE1-type plasmids. *Gene* (1984).

Dayaram, A., Franz, M. & Schattschneider, A. Long term stability and infectivity of herpesviruses in water. *Scientific Reports* (2017).

Ehlers, A., Worm, B. & Reusch, T. Importance of genetic diversity in eelgrass *Zostera marina* for its resilience to global warming. *Marine Ecology Progress Series* (2008).

Gatherer, D., Depledge, D. P., Hartley, C. A., Szpara, M. L., Vaz, P. K., Benkő, M., Brandt, C. R., Bryant, N. A., Dastjerdi, A., Doszpoly, A., Gompels, U. A., Inoue, N., Jarosinski, K. W., Kaul, R., Lacoste, V., Norberg, P., Origi, F. C., Orton, R. J., Pellett, P. E., Schmid, D. S., Spatz, S. J., Stewart, J. P., Trimpert, J., Waltzek, T. B. & Davison, A. J. ICTV Virus

Taxonomy Profile: *Herpesviridae*. *Journal of General Virology* (2021).

Grünewald, K., Desai, P., Winkler, D.C., Heymann, J.B., Belnap, D.M., Baumeister, W. & Steven, A.C. Three-dimensional structure of herpes simplex virus from cryo-electron tomography. *Science* (2003).

Hare, S., Gupta, S.S., Valkov, E., Engelman, A. & Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* (2010).

Jumper, J., Evans, R. & Pritzel, A. Highly accurate protein structure prediction with AlphaFold. *Nature* (2021).

Kaye, K. M. Overview of Herpesvirus Infections. *Harvard Medical School* (2021).

Keil, T., Liu, D., Lloyd, M., Coombs, W., Moffat, J. & Visalli, R. DNA Encapsidation and Capsid Assembly Are Underexploited Antiviral Targets for the Treatment of Herpesviruses. *Frontiers in Microbiology* (2020).

Kelly, B., Fraefel, C., Cunningham, A. & Diefenbach, R. roles of the tegument proteins of herpes simplex virus type 1. *Virus Research* (2009).

Ligat, G., Cazal, R., Hantz, S. & Alain, S. The human cytomegalovirus terminase complex as an antiviral target: a close-up view. *FEMS Microbiology Reviews* (2018).

Loret, S., Guay, G. & Lippé, R. Comprehensive characterization of extracellular herpes simplex virus type 1 virions. *Journal of Virology* (2008).

McGeoch, D. J., Cook, S., Dolan, A., Jamieson, F.E. & Telford, E.A. Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *Journal of Molecular Biology* (1995).

McVoy, M. A., Nixon, D. E., Adler, S. P., & Mocarski, E. S. Sequences within the herpesvirus-conserved pac1 and pac2 motifs are required for cleavage and packaging of the murine cytomegalovirus genome. *Journal of virology* (1998).

Montague, M. G. & Hutchison, C. A. Gene content phylogeny of herpesviruses. *Proceedings of the National Academy of Sciences of the United States of America, PNAS* (2000).

Murphy, C. M., Xu, Y., Li, F., Nio, K., Reszka-Blanco, N., Li, X., Wu, Y., Yu, Y., Xiong, Y. & Su, Y. Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication. *Cell Reports* (2016).

Nadal, M., Mas, P.J., Blanco, A.G., Arnan, C., Solà, M., Hart, D.J & Coll, M. Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain. *Proceedings of the National Academy of Sciences* (2010).

Pellet, P. & Roizman, B. *Herpesviridae: A Brief Introduction. Fields Virology*, 5th Edition, 2480-2499 (2007).

Renner, D.W. & Szpara, M.L. Impacts of Genome-Wide Analyses on Our Understanding of Human Herpesvirus Diversity and Evolution. *Journal of Virology* (2018).

Roizman, B. & Campadelli-Fiume, G. Alphaherpes viral genes and their functions. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge University Press* (2007).

Savva, C.G., Holzenburg, A. & Bogner, E. Insights into the structure of human cytomegalovirus large terminase subunit pUL56. *FEBS Letters* (2004)

Scheffczik, H., Savva, C., Holzenburg, A., Kolesnikova, L. & Bogner, E. The terminase subunits pUL56 and pUL89 of human cytomegalovirus are DNA-metabolizing proteins with toroidal structure. *Nucleic Acids Research*. (2002)

Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V.T., Hein, M.Y., Huang, S.X., Ma, M., Shen, B., Qian, S.B., Hengel, H., Mann, M., Ingolia, N.T. & Weissman, J.S. Decoding human cytomegalovirus. *Science* (2012).

Weller, S.K. & Coen, D.M. Herpes simplex viruses: mechanisms of DNA replication. *Cold Spring Harbor Perspectives in Biology* (2012).

Whisnant, A.W., Jürges, C.S. & Hennig, T. Integrative functional genomics decodes herpes simplex virus 1. *Nature Communications* (2020).

Whitley, R.J. Herpesviruses. *Medical Microbiology*, 4th edition, Chapter 68 (1996).

Yang, Y., Yang, P., Wang, N. Architecture of the herpesvirus genome-packaging complex and implications for DNA translocation. *Protein Cell* (2020).

Zerboni, L., Sen, N., Oliver, S. L., & Arvin, A. M. Molecular mechanisms of varicella zoster virus pathogenesis. *Nature reviews* (2014).

Zmasek, C. M., Knipe, D. M., Pellett, P. E., & Scheuermann, R. H. Classification of human Herpesviridae proteins using Domain-architecture Aware Inference of Orthologs (DAIO). *Virology*, 529, 29–42 (2019).

PART II: Coronavirus

CHAPTER 6: INTRODUCTION

6. INTRODUCTION

6.1. Coronaviridae family

6.1.1. Classification

The genus *Coronavirus* together with the genus *Torovirus* form the family *Coronaviridae*. These two genera are similar morphologically. The *Coronaviridae*, *Arteriviridae*, and *Roniviridae* have been classified as members of the Order *Nidovirales*. They mutate and also recombine frequently. Members of this order have a similar genome organization and produce a nested set of subgenomic mRNAs. Traditionally, coronaviruses (CoV) have been classified into three groups (Table 27). Initially, this was on the basis of serological relationships which subsequently have been supported by gene sequencing.

Group 1	Group 2	Group 3
Subgroup 1a	Subgroup 2a	
Transmissible gastroenteritis virus	Murine hepatitis virus	Infectious bronchitis virus (IBV)
Feline coronavirus	Bovine coronavirus	Turkey coronavirus
Canine coronavirus	Porcine haemagglutinating encephalomyelitis virus	Pheasant coronavirus
Ferret coronavirus	Equine coronavirus	Duck coronavirus
	Canine respiratory coronavirus	Goose coronavirus
	Human coronavirus HKU1	Pigeon coronavirus
	Human coronavirus OC43	
	Human enteric coronavirus	
	Rat coronavirus	
	Puffinosis coronavirus	
Subgroup 1b	Subgroup 2b	
Human coronavirus 229E	Severe acute respiratory syndrome (SARS) coronavirus	
Porcine epidemic		

diarrhoea virus	Bat-CoV-HKU3-1
Bat coronavirus-61	
Bat coronavirus-HKU2	
Human coronavirus NL63	

Table 27. Examples of Coronavirus species.

Up to now, seven different types of Coronavirus that can infect humans have been identified. They are classified in two distinct genera, *Alphacoronavirus* and *Betacoronavirus*. The *Betacoronavirus* genera is divided at its turn in three different lineages. Lineage A includes OCH3 and HKU1 viruses, SARS and SARS-2 are classified in lineage B and Middle East respiratory syndrome (MERS) coronavirus is the unique member of lineage C (Figure 92).

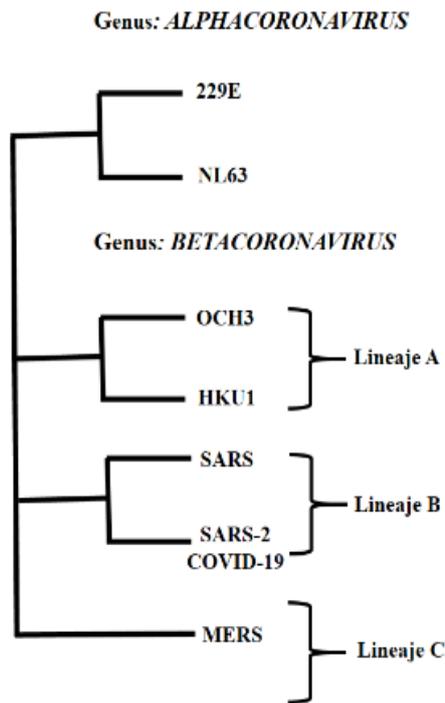


Figure 92. Classification of the seven types of Coronavirus that infected humans.

6.1.2. Epidemiology, transmission and pathology

Some coronaviruses like 229E, OC43, HKU1 and NL63 are common viruses which cause various respiratory diseases, including pneumonia (*Zeng et al. 2018*). They are distributed globally and tend to be transmitted predominantly during the winter season in temperate-climate countries, while NL63 showed a spring–summer peak of activity from a study in Hong Kong.

According to the WHO the emergence of viral diseases represents a serious public health risk. In the past two decades, several epidemics caused by viruses have been described which have had a significant impact on global health: the severe acute respiratory syndrome coronavirus (SARS-CoV) from 2002 to 2003, H1N1 influenza in 2009, the Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 and the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from 2019 and up to now.

❖ SARS-CoV (Betacoronavirus, lineage B)

Patients infected with SARS-CoV initially present with fever, myalgia, headache, malaise, and chills, followed by a nonproductive cough, dyspnea, and respiratory distress generally 5 to 7 days later, which may result in death. Other notable features in some cases include infection of the gastrointestinal tract, liver, kidney, and brain. Diffuse alveolar damage, epithelial cell proliferation, and an increase in macrophages is seen in SARS-CoV infection of the lung. Lymphopenia, hemophagocytosis in the lung, in addition to white-pulp atrophy of the spleen observed in SARS patients, are similar to fatal H5N1 influenza virus infections. Diarrhea is observed in approximately 30–40% of SARS infections.

An outbreak of disease caused by SARS-CoV, originating from Guangdong Province in southern China during November 2002, eventually spread to other countries in Asia, in addition to North America and Europe (37 countries/regions in total) over 9 months (Figure 93). 8,273 cases were reported with 775 deaths, implying a case fatality rate of 9%. The elderly were more susceptible to SARS disease, with a mortality rate of over 50%. The majority of cases and deaths occurred in mainland China and in Hong Kong.

❖ MERS-CoV (Betacoronavirus, lineage C)

MERS-CoV was first isolated from the lungs of a 60-year-old patient who had died from a severe respiratory illness in Jeddah (Saudi Arabia) in 2012. Clinical manifestations of MERS-CoV infection range from asymptomatic to severe pneumonia with acute respiratory distress, septic shock, and renal failure resulting in death. A typical disease course begins with fever, cough, chills, sore throat, myalgia, and arthralgia, followed by dyspnea and rapid progression to pneumonia. Approximately

one-third of patients present with gastrointestinal symptoms, such as diarrhea and vomiting. Acute renal impairment was the most striking feature of disease caused by MERS-CoV, which is thus far unique for human CoV infections. Seventy-five percent of patients with MERS disease also had at least one other comorbidity, and patients who died were more likely to have a pre-existing/underlying condition. Countries around the Arabian Peninsula are known to be endemic for MERS-CoV, and Saudi Arabia has reported the most cases, but since its discovery in 2012, cases have been occasionally exported to other countries through travel, sometimes causing clusters of secondary outbreaks (Figure 93). A total of 2562 confirmed MERS cases with 150 case clusters were reported with a case fatality rate of 32.7%.

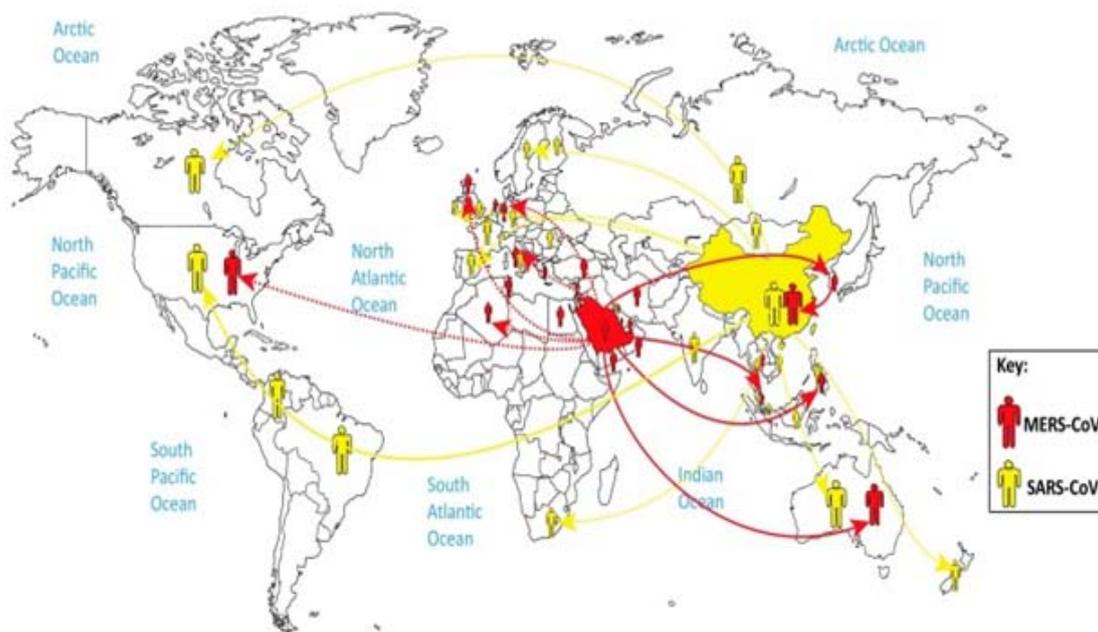


Figure 93. Global distribution of MERS-CoV and SARS-CoV, respectively.

❖ SARS-CoV-2 (Betacoronavirus, lineage B)

Since being declared a global pandemic by the WHO, SARS-CoV-2, the virus responsible for Coronavirus Disease 2019 (COVID-19) has spread to 223 countries with more than 178 million confirmed cases, and more than 3.8 million deaths reported globally as of November 2021. The U.S. experienced the highest number of SARS-CoV-2 infections and COVID-19 related deaths followed by Brazil and India. In fact, COVID-19 was the third leading cause of death in the U.S. in 2020 after heart disease and cancer, with approximately 375,000 deaths reported (Figure 94). As of 22 June 2021, the Alpha (B.1.1.7) variant has spread to 170 countries, the Beta (B.1.351) variant has been reported in 119 countries, the Gamma (P.1) variant has been

detected in 71 countries and the Delta variant (B.1.617.2) has spread to 85 countries around the world based on the weekly epidemiological update by the WHO. The new variant Omicron is the fastest-spreading virus known to humankind that has traversed the planet in a single month. The WHO's current estimate of the global case fatality rate for COVID-19 is 2.2%. However, the case fatality rate is affected by factors that include age, underlying preexisting conditions, and severity of illness and significantly varies between countries.

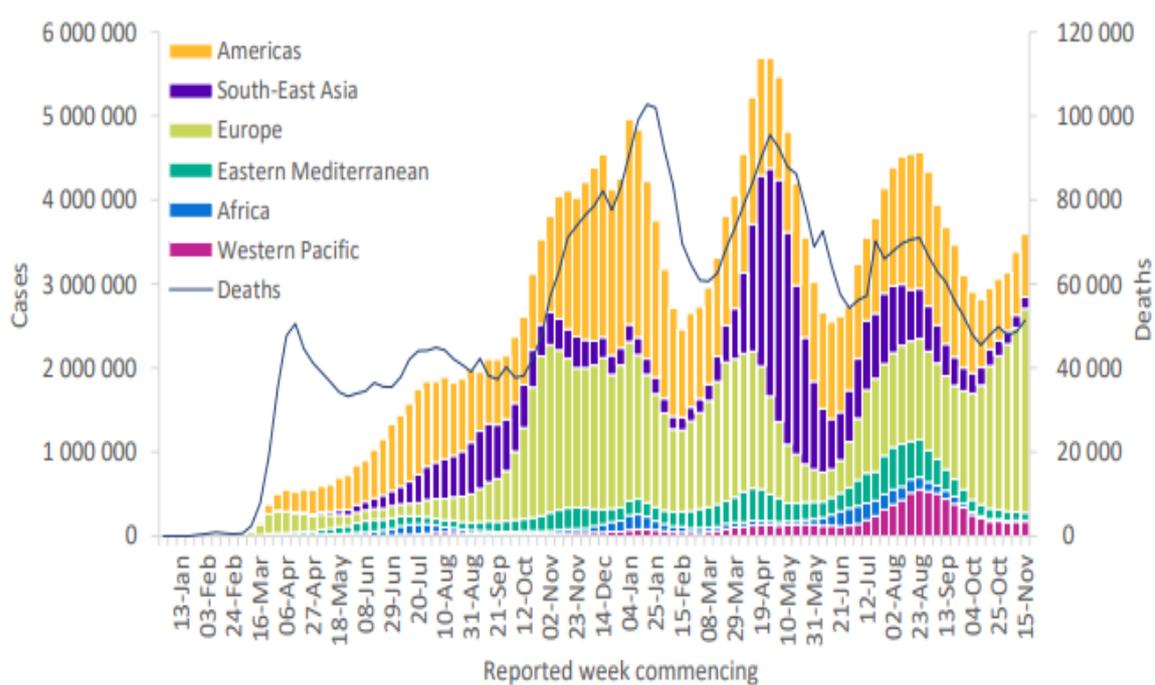


Figure 94. COVID-19 cases reported weekly by WHO Region, and global deaths, as of 21 November 2021.

6.1.3. Treatment

Initially, the understanding of COVID-19 and its therapeutic management was limited, creating an urgency to mitigate this new viral illness with experimental therapies and drug repurposing. Since then, due to the intense efforts of clinical researchers globally, significant progress has been made, which has led to a better understanding of not only COVID-19 and its management but also has resulted in the development of novel therapeutics and vaccine development at an unprecedented speed.

Currently, a variety of therapeutic options are available that include antiviral drugs (e.g., remdesivir), anti-SARS-CoV-2 monoclonal antibodies (e.g., bamlanivimab/etesevimab, casirivimab/imdevimab), anti-inflammatory drugs (e.g., dexamethasone) and immunomodulators agents (e.g., baricitinib, tocilizumab) (*Han et al., 2021*).

The clinical utility of these treatments is specific and is based on the severity of illness or certain risk factors. The clinical course of the COVID-19 illness occurs in 2 phases, an early phase when SARS-CoV-2 replication is greatest before or soon after the onset of symptoms. Antiviral medications and antibody-based treatments are likely to be more effective during this stage of viral replication. The later phase of the illness is driven by a hyperinflammatory state induced by the release of cytokines and the coagulation system's activation that causes a prothrombotic state. Anti-inflammatory drugs such as corticosteroids, immunomodulating therapies, or a combination of these therapies may help combat this hyperinflammatory state than antiviral therapies (Drożdżal *et al.*, 2021).

Three COVID-19 vaccines are authorized or approved for use in the United States to prevent COVID-19. Pfizer-BioNTech or Moderna (COVID-19 mRNA vaccines) and Johnson & Johnson's Janssen COVID-19 vaccine (Viral Vector COVID-19 Vaccines).

6.1.4. Structure

All coronaviruses have four structural proteins in common (Figure 95): a large surface spike glycoprotein (S), a small envelope protein (present in very small amounts in virions) (E), an integral membrane glycoprotein (M) and a phosphorylated nucleocapsid protein (N). Group 2a viruses have an additional structural glycoprotein, the hemagglutinin-esterase (HE) protein. This is not essential for replication *in vitro* but may affect tropism *in vivo*.

Virions are 100-120 nm in diameter, although they can be up to twice that size, and the ring of spike glycoprotein is approximately 20 nm deep. When present, the hemagglutinin-esterase protein forms a layer 5–10 nm deep. In some species, the spike glycoprotein is cleaved into two subunits, the N-terminal S1 fragment being slightly smaller than the C-terminal S2 sequence (Figure 96). The spike glycoprotein is anchored in the envelope by a transmembrane region near the C-terminus of S2. It is highly glycosylated and exists as a trimer. The bulbous outer part of the mature spike glycoprotein is formed largely by S1 while the stalk is formed largely by S2, having a coiled-coil structure. S1 is the most variable part of the spike glycoprotein and is the major inducer of protective immune responses.

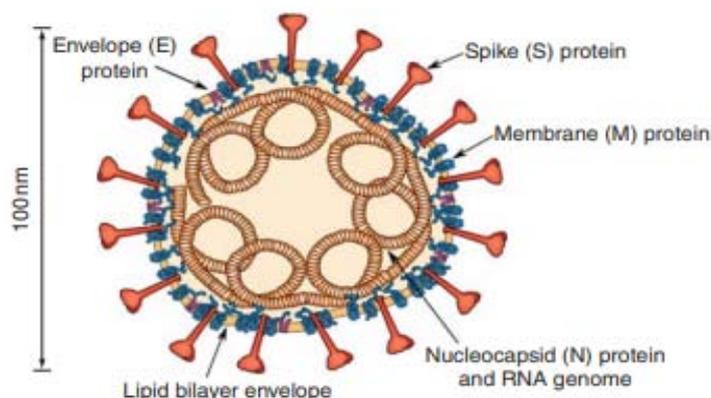


Figure 95. Schematic diagram of a Covid (CoV) virion with the minimal set of four structural proteins required for efficient assembly of the infectious virus particles: S, spike glycoprotein; M, membrane

glycoprotein; E, envelope protein; and N, nucleocapsid phosphoprotein which encapsidates the genome (Mousavizadeh & Ghasemi, 2021).

The M glycoprotein is the most abundant protein in virions. In most cases, only a small part at the N-terminus protrudes at the surface of the virus. There are three membrane-spanning segments and the C-terminal half of it is within the lumen of the virus. The E protein is anchored in the membrane by a sequence near its N-terminus.

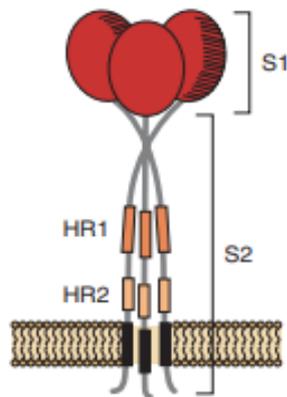


Figure 96. Diagrammatic representation of the spike trimer assembled on membranes, with the S1 receptor binding domain (RBD), S2 fusion domain and two heptad repeat (HR) regions, HR1 and HR2 indicated.

6.1.5. Genome organization

Coronaviruses have the largest known RNA genomes, which comprise 28–32 kb of positive sense, single-stranded RNA. The overall genome organization is being 5' UTR–polymerase gene–structural protein genes–3' UTR, where the UTRs are untranslated regions (Figure 97). The first 60–90 nucleotides at the 5' end form a leader sequence. The structural protein genes are in the same order in all coronaviruses: (HE)–S–E–M–N. Interspersed among these genes are one or more genes (depending on the species; for instance SARS-CoV has four) that encode small proteins of unknown function. Some of these genes encode two or three proteins. In some cases, translation of the third and second open reading frame (ORF), respectively, is effected by the preceding ORFs acting as internal ribosome entry sites. The proteins encoded by these small ORFs are mostly not required for replication in vitro and some of them might function as antagonists of innate immune responses.

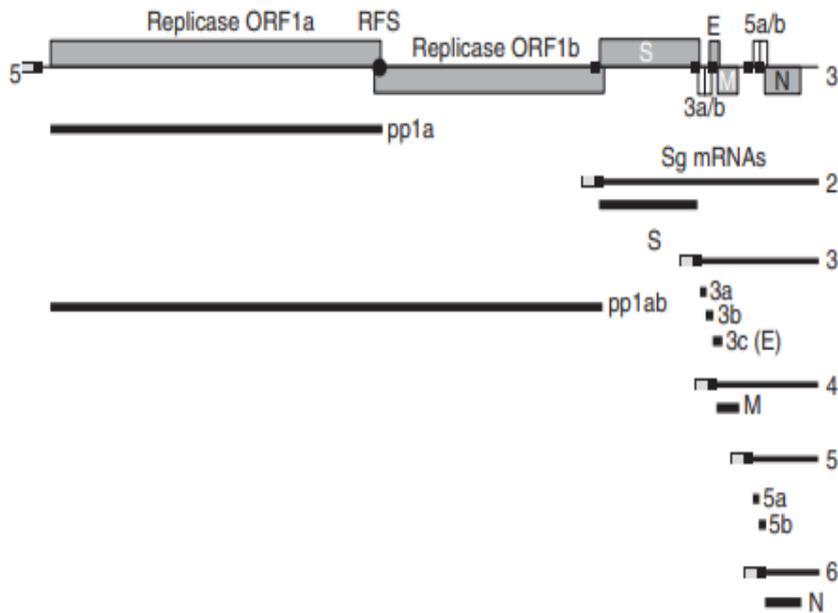


Figure 97. Schematic diagram representing the genomic expression of the avian coronavirus IBV. The upper part of the diagram shows the IBV genomic RNA. The black boxes represent the transcription regulatory sequences (TRSs). The leader sequence, represented by a gray box, is at the 5' end of the genomic RNA and at the 5' ends of the subgenomic (sg) mRNAs. The genomic RNA is translated to produce two polyproteins, pp1a and pp1ab. The structural proteins, S, E, M, and E, and the accessory proteins, 3a, 3b, 5a, and 5b, produced from IBV genes 3 and 5, respectively, are translated from the sg mRNAs. The proteins produced by the sg mRNAs are represented by lines below the corresponding sg mRNA. The ribosome frameshift (RFS) region, denoted as a black circle on the genomic RNA, directs the -1 frameshift event for the synthesis of pp1ab. Translation of the genomic RNA results in the production of pp1a. The 5' and 3' UTR sequences are represented as single lines downstream of the leader and N gene sequences, respectively.

Following entry into a cell and the release of the virus ribonucleoprotein into the cytoplasm, ribosomes translate gene 1, which is approximately 20 kb, into two polyproteins (pp1a and pp1ab). These are cleaved by gene 1-encoded proteases, to generate 15 or 16 proteins (Figure 98). Translation of ORF 1b involves ribosomal frameshifting, which has two elements, a slippery site followed by an RNA pseudoknot. At the slippery site, the ribosome slips one nucleotide backward and then moves forward, this time in a -1 frame compared with translation ORF 1a, resulting in the synthesis polyprotein 1ab. Proteins, including the RNA-dependent RNA polymerase, associate to form the replicase complex, which is membrane associated. Coronavirus subgenomic mRNAs (sg mRNAs) are generated by a discontinuous process. At the beginning of each gene is a common sequence called a transcription regulatory sequence (TRS). It is believed that when the polymerase producing the nascent negative sense RNA reaches a TRS, RNA synthesis is attenuated, followed by continuation at the 5' end of genomic RNA. This results in the addition of a negative copy of the leader sequence to the negative-sense RNA, resulting in a negative-sense copy of an sg mRNA. Progress of the polymerase is not always halted at a TRS. Rather, it sometimes continues, producing a nested set of negativesense sg mRNAs. These are the templates for the generation of the positive-

sense sg mRNAs (Figure 97). The leader sequence is found at the very 5' end of the genomic RNA and at the 5' ends of each sg mRNA.

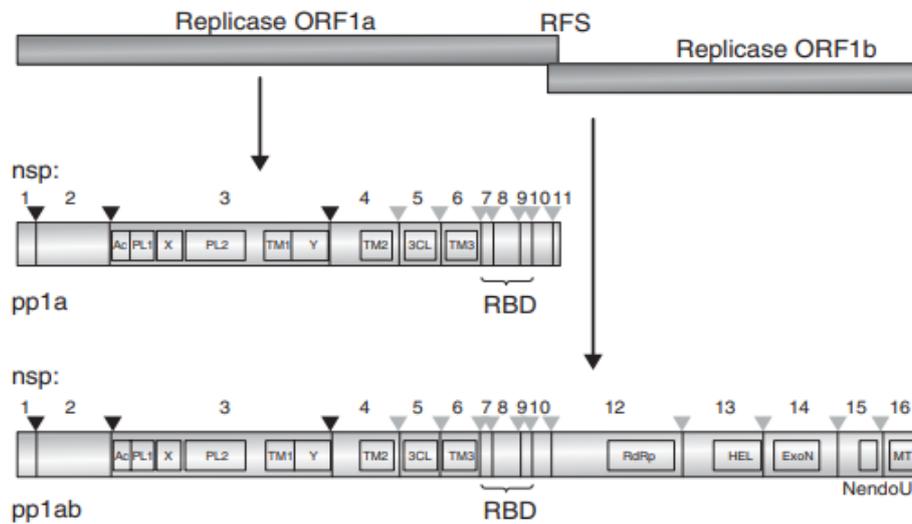


Figure 98. Organization of the coronavirus replicase gene products. Translation of the coronavirus replicase ORF 1a and ORF 1b sequences results in pp1a and pp1ab. The two polyproteins are proteolytically cleaved into 10 (pp1a; nsp1–11) and 16 (pp1ab; nsp1–16) products by the papain-like proteinases (PL1pro and PL2pro) and the 3C-like protease (3CLpro) proteinase. The PLpro proteinases cleave at the sites indicated with a black triangle and the 3CLpro proteinase cleaves at the sites indicated with a gray triangle. Ribosome Binding Site (RBD).

6.1.6. Infection cycle

The S1 part of the S protein mediates the attachment to cells. It is a determinant of host species specificity and, in some cases, pathogenicity, by determining susceptible cell range within a host. The S2 part triggers fusion of the virus envelope with cell membranes, which can occur at neutral or slightly acidic pH, depending on species or even strain. After the entrance of the virus in the cell the virus glycoproteins (S, M and HE, when present) are synthesized at the endoplasmic reticulum (Figure 99). Following infection of a susceptible cell, the coronavirus genomic RNA is released from the virion into the cytoplasm and immediately recognized as an mRNA for the translation of the replicase pp1a and pp1ab proteins. These proteins are cleaved by ORF1a-encoded proteases, after which they become part of replicase complexes for the synthesis of either complete negative-sense copies of the genomic RNA or negative-sense copies of the sg mRNAs. The negative-sense RNAs are used as templates for the synthesis of genomic RNA and sg mRNAs (Figure 97). Following synthesis of the sg mRNAs, the structural proteins are produced for the assembly and encapsidation of the de novo-synthesized genomic RNA, resulting in the release of new infectious coronavirus virions.

Early and late in infection, formation of virus particles can occur in the endoplasmic reticulum–Golgi intermediate compartment (ERGIC) and endoplasmic reticulum, but most assembly occurs in the Golgi membranes. The M protein is not transported to the plasma membrane and its location at internal membranes determines the sites of

virus particle formation. The release of new virions starts 3–4 h after the initial infection.

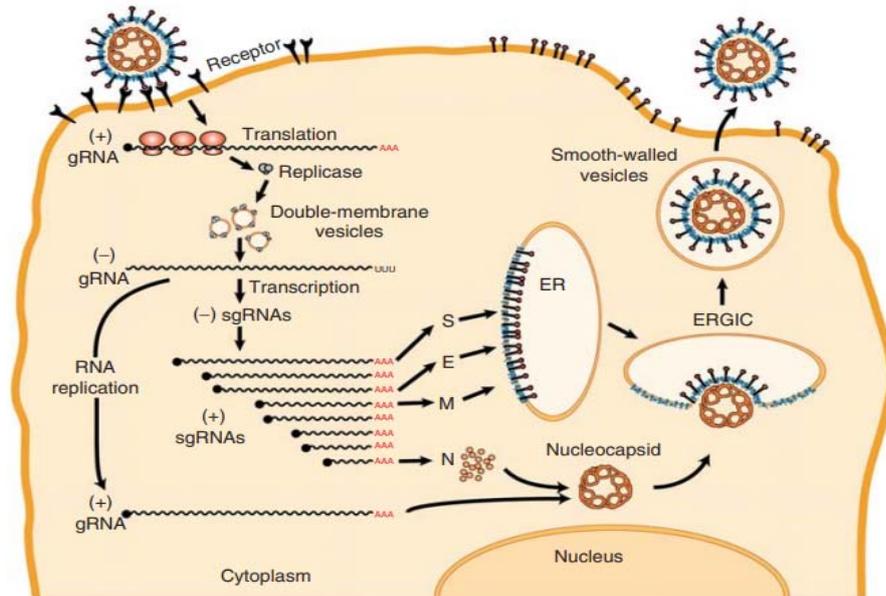


Figure 99. Replication cycle of CoVs. The spike glycoprotein on the virus particle interacts with host cell receptors to mediate fusion of the virus and host cell membranes and release of the positive-strand RNA genome into the cytoplasm. Reprinted from Masters PS (2006) *The molecular biology of coronaviruses. Advances in Virus Research* 66: 193–292.

As indicated above, the synthesis of the sg mRNAs is the result of a discontinuous process in which the synthesis of a negative-sense copy of an sg mRNA is completed by the addition of the negative-sense leader sequence by a recombination mechanism. If a cell is infected with two related coronaviruses, the polymerase may swap between two RNA templates, in a similar way to addition of the leader sequence. This ‘copy-choice’ mechanism of genetic recombination results in a chimeric RNA. Such RNAs may give rise to new viruses with modified genomes with a capacity to infect a different cell and, in some cases, new host species.

6.1.7. Structure and function of the main protease

Proteolytic cleavage of the two coronavirus polyproteins generates the various viral proteins needed to form a replication complex, required for transcription and replication of the viral genome and subgenomic mRNAs. The key viral enzymes responsible are the papain-like (PLP, nsp3) and main proteases (Mpro, nsp5) (Meyer *et al.* 2021).

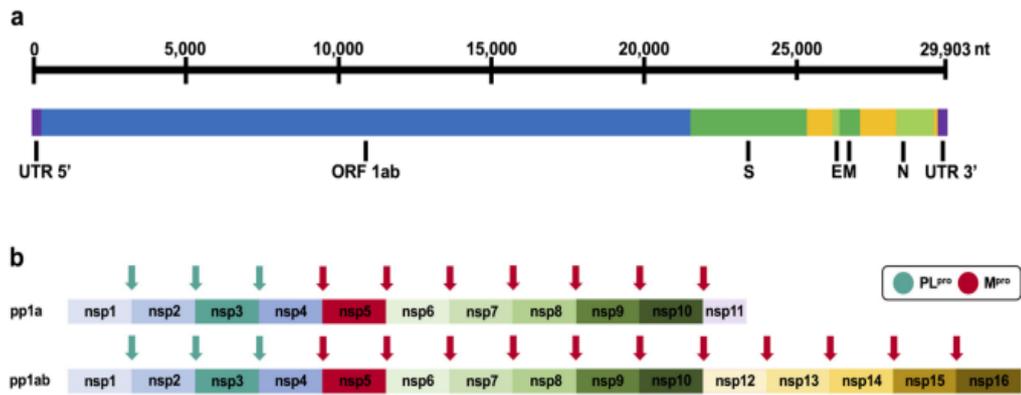


Figure 100. A) Organisation of the RNA genome of SARS-CoV-2 with selected genes (Wuhan-Hu-1 isolate MN908947). B) Schematic representation of polyprotein cleavage sites of SARS-CoV-2. The papain-like protease PL^{pro} cleaves at 3 distinct sites. The main protease M^{pro} (3CL^{pro}) cleaves at 11 distinct sites (Ullrich & Nitsche, 2020).

The SARS-CoV-2 M^{pro} proteolytically cleaves the overlapping pp1a and pp1ab polyproteins to functional proteins (Figure 100), which is a critical step during viral replication. Replication-essential enzymes such as RdRp or nsp13 cannot fully function without prior proteolytic release, positioning M^{pro} as a key enzyme in the viral replication cycle. Thus, M^{pro} is one of the most attractive viral targets for antiviral drug discovery against SARS-CoV-2.

The M^{pro} is a homodimer of 33.797 kDa, containing two protomers each, comprising three domains (Domains I, II, and III). Domains I and II are made up of six antiparallel β -barrels. An antiparallel globular cluster of five α helices forms domain III, which is connected to domain II *via* a long loop region. In the cleft between domains I and II, there is a Cys-His catalytic dyad which is thought to have a vital role in proteolytic activity. The substrate-binding site is located in the cleft between domains I and II and the protomers are located between domains II and III with roles in the formation of the substrate-binding site. The substrate-binding cleft is comprised of four subsites namely; S1', S1, S2, and S4 (Mengist *et al.*, 2021).

Coronavirus main proteases perform a key role during the infection cycle and are conserved among species. Amino acid sequence alignments reveal ~99% identity with the bat coronavirus RaTG13 M^{pro} and ~96% with the previous SARS-CoV M^{pro}. In contrast, sequence identity with MERS-CoV M^{pro} is only ~50% (Figure 101). Superimposition of the X-ray crystal structures of the main proteases of SARS-CoV-2, SARS-CoV and MERS-CoV indicates a high degree of structural similarity and conservation of the active site (Figure 102).

SARS-CoV-2	1	SGFRKMAFPPSGKVEGCMVQVTCGTTTLNGLWLDLVVYCPRHVICT*	SEDMLNPNYEDLLIR
SARS-CoV	1	SGFRKMAFPPSGKVEGCMVQVTCGTTTLNGLWLDLVVYCPRHVICTAEDMLNPNYEDLLIR	
MERS-CoV	1	SGLVKMSHPSGDEVEACMVQVTCGSMTLNGLWLDLVVYCPRHVICTAEDMLNPNYEDLLIS	
SARS-CoV-2	61	KSNHNFVLQAG---NVQLRVIGHSMQNCVLLKLVDTANPKTPKYKVFRIQPGQTFSVLAC	
SARS-CoV	61	KSNHSFVLQAG---NVQLRVIGHSMQNCVLLKLVDTANPKTPKYKVFRIQPGQTFSVLAC	
MERS-CoV	61	MLNHSFSLVQKHIGAPALRLVGHAGGTLKLVVAVANSTEAAYTETTKPGAAFVSLAC	
SARS-CoV-2	118	YNGSPSGVYQCAMPNHTIKGSFLNGSCGTVGFNIDYDCVSFCYMHMELPTGVHAGTDL*	
SARS-CoV	118	YNGSPSGVYQCAMPNHTIKGSFLNGSCGTVGFNIDYDCVSFCYMHMELPTGVHAGTDL	
MERS-CoV	121	YNGRPTGTITVVMRPNHTIKGSFLGSCGTVGFNIDYDCVSFCYMHMELPTGVHAGTDL	
SARS-CoV-2	178	EGNFYGFVDRQTAQAAGDPTTITVNVLAWLYAAVINGDRWFLNRFETTLNDFNLVAMKY	
SARS-CoV	178	EGNFYGFVDRQTAQAAGDPTTITVNVLAWLYAAVINGDRWFLNRFETTLNDFNLVAMKY	
MERS-CoV	181	EGTMYGAFHDKVHGVQLTETKYCSVNVVAVLYAALINGCAWFKPNRTFVSVFNEWALAN	
SARS-CoV-2	238	NYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSTLEDEFTPFDV	
SARS-CoV	238	NYEPLTQDHVDILGPLSAQTGIAVLDMCAALKELLQNGMNGRTILGSTLEDEFTPFDV	
MERS-CoV	241	QTEFVGTQSDML---AVRFGVAEQLLYALQQL-YTEFQCNILGSTMLEDEFTPFDV	
SARS-CoV-2	297	VRQCSGVTFQ	
SARS-CoV	297	VRQCSGVTFQ	
MERS-CoV	297	NMGLMGVVMQ	

Figure 101. Alignment of the amino acid sequences of crystallised main proteases of SARS-CoV-2 (PDB: 6Y2E), SARS-CoV (PDB: 2BX4) and MERS-CoV (PDB: 5C3N). Domains I, II and III comprise residues 8–101, 102–184 and 201–306, respectively. The catalytic dyads are indicated by asterisks. The alignment was generated using T-Coffee and shaded with Boxshade (Ullrich & Nitsche, 2020).

Mpro is a cysteine protease with a catalytic dyad (cysteine and histidine) in its active centre (Figure 102). While other cysteine and serine proteases contain a third catalytic residue, a buried water molecule occupies this place in the active site of Mpro. The proteolytic process is believed to follow a multi-step mechanism. After the cysteine side chain proton is abstracted by the histidine's imidazole, the resulting thiolate nucleophile attacks the amide bond of the substrate. The N-terminal peptide product is released by proton abstraction from histidine before the thioester is hydrolysed to release the C-terminal product and restore the catalytic dyad.

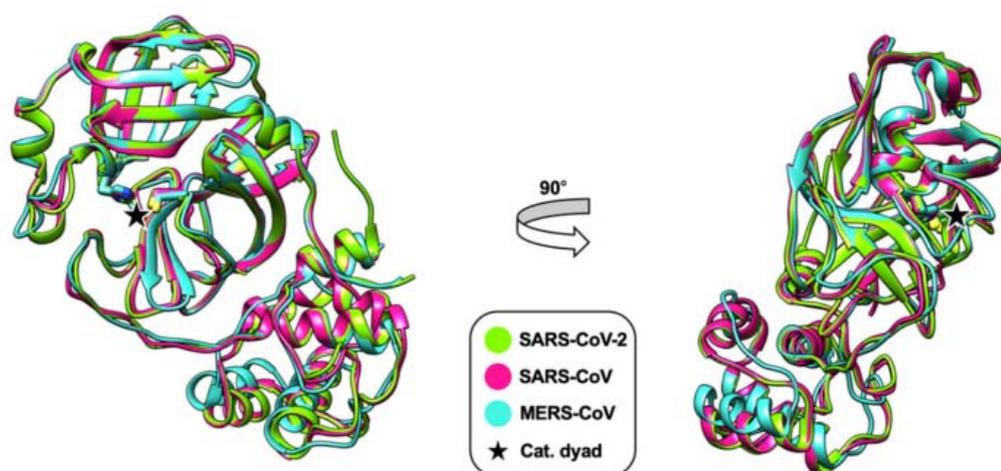


Figure 102. Superimposition of X-ray crystal structures of the main proteases of SARS-CoV (pink, PDB: 2BX4), MERS-CoV (cyan, PDB: 5C3N) and SARS-CoV-2 (green, PDB: 6Y2E). Only the monomers are shown. Residues of the catalytic dyad are indicated (His41/ Cys145 for SARS-CoV and SARS-CoV-2 and His41/ Cys148 for MERS-CoV). This figure was generated with UCSF Chimera (Ullrich & Nitsche, 2020).

According to the nomenclature introduced by Schechter and Berger, Mpro mainly recognises substrate residues ranging from P4 to P1. Prime site recognition beyond P1' is not conserved (Figure 103). Specificity is mostly determined by P1, P2 and P1', which show the highest degree of conservation amongst the cleavage sites. Glutamine in P1 is highly conserved in all polyprotein cleavage sites of SARS-CoV, MERS-CoV and SARS-CoV-2 (Figure 103). In P2 more hydrophobic amino acids are tolerated with a clear preference for leucine. P1 tolerates small residues like serine or alanine. Analysis of all polyprotein cleavage sites processed by Mpro for SARS-CoV, MERS-CoV and SARS-CoV-2 illustrates very similar substrate recognition profiles amongst these viruses (Figure 103).

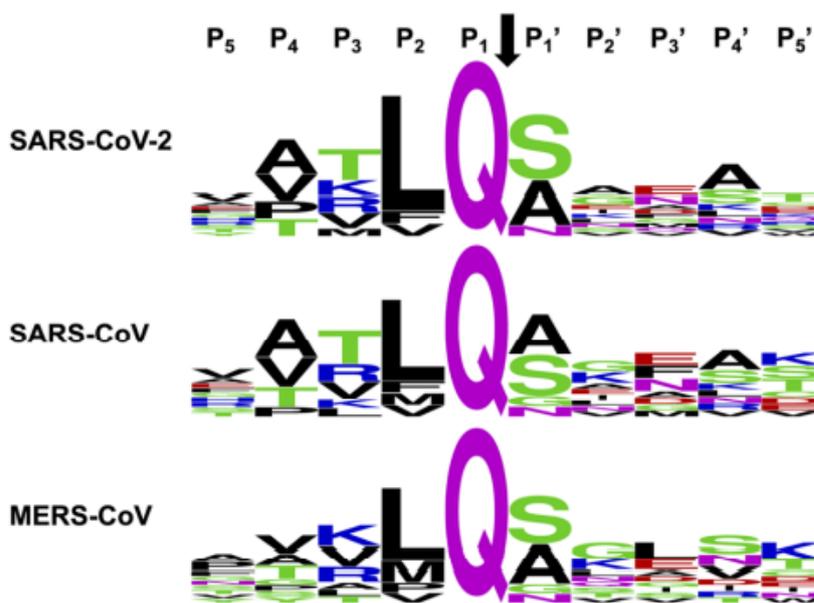


Figure 103. Polyprotein cleavage sites recognised by Mpro of SARS-CoV-2, SARS-CoV and MERS-CoV. Peptide sequences cover residues P5 to P5' according to the nomenclature of Schechter and Berger. Data were generated from pp1ab polyprotein sequences reported in the UniProt database with the accession codes P0DTD1 (SARS-CoV-2), P0C6X7 (SARS-CoV) and K9N7C7 (MERS-CoV) (Ullrich & Nitsche, 2020).

6.1.8. Coronavirus proteins as drug targets: the relevance of the main protease

The current COVID-19 pandemic has triggered global efforts for the rapid identification of vaccines and specific antiviral treatments. Amongst the coronaviral targets that have been studied in the past, the main protease (Mpro, 3CLpro, nsp5) received major attention, particularly following the first SARS-CoV outbreak in the early 2000s.

Alternative coronaviral targets include the S protein, RNA-dependent RNA-polymerase (RdRp, nsp12), NTPase/helicase (nsp13) and papain-like protease (PLpro, part of nsp3). The papain-like protease also recognises the C-terminal sequence of ubiquitin. Therefore, substrate-derived inhibitors of PLpro would be expected to also inhibit hostcell deubiquitinases, making drug-discovery campaigns against PLpro challenging. In stark contrast, the main protease Mpro exclusively cleaves polypeptide sequences after a glutamine residue, positioning the main

protease as an ideal drug target because no human host-cell proteases are known with this substrate specificity.

Viral proteases are well validated drug targets that have led to various approved drugs, for example, against chronic infections with human immunodeficiency virus (HIV) or hepatitis C virus (HCV), which employ aspartyl and serine proteases, respectively.

This might prove valuable for the development of pan-coronaviral drugs and has already been employed for the development of SARS-CoV-2 Mpro inhibitors that were based on previous compounds targeting the SARS-CoV or MERS-CoV homologs.

6.1.9. Main protease inhibitors

Although SARS-CoV-2 emerged only very recently, several inhibitors have already been identified and successfully co-crystallised with Mpro. They are often derived from previous campaigns which targeted the main proteases of SARS-CoV or MERS-CoV and contain cysteine-reactive warheads.

Particularly important is the pronounced preference for glutamine in P1, strongly informing inhibitor design. Since no human host-cell proteases with similar specificity are reported, reduced offtarget effects are assumed for peptidomimetic inhibitors.

Peptidomimetics and small molecules have been reported with affinities in the micromolar to nanomolar range. They often depend on warhead-based design strategies, employing different reactive groups to attack covalently the catalytic cysteine residue. Warheads utilised include Michael acceptors, aldehydes, epoxy ketones and other ketones.

The first reported inhibitors were covalently binding peptidomimetics addressing the major substrate-recognition motif from P1' to P3. They all comprise an α -ketoamide functionality that forms a hemithioacetal with Cys145. Compound 1 has previously been investigated as a broadspectrum corona- and enteroviral protease inhibitor. Like many other Mpro inhibitors, the P1 side chains of 1–3 employ a γ -lactam as a glutamine mimetic. P2 comprises hydrophobic cyclohexyl (1, 2) or smaller cyclopropyl (3) groups as leucine mimetics and P1' contains cyclopropyl (2) or benzyl (1, 3) residues. Compounds 1 and 3 displayed sub-micromolar Mpro inhibition (Figure 104). Compound 3 is similarly active against the SARS-CoV and MERS-CoV main proteases and inhibits SARS-CoV-2 replication in human Calu3 lung cells.

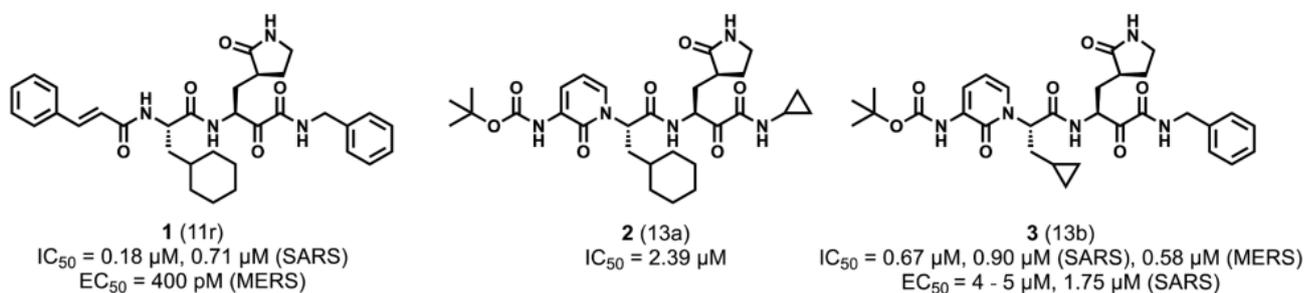


Figure 104. Inhibitors (compound 1, 2 and 3) of the SARS-CoV-2 main protease *Mpro*. IC_{50} indicates enzymatic inhibition. EC_{50} indicates antiviral activity in cells (Ullrich & Nitsche, 2020).

6.1.10. Antiviral compounds rupintrivir and compound 1

Rupintrivir (AG7088) and its orally bioavailable analogue compound 1 (AG7404) are compounds against the common cold that progressed to phase-II/I clinical trials. Both compounds (Figure 105) are peptidomimetic inhibitors and imitate the P4 to P1 peptide substrate, with an α,β -unsaturated ester at P1' as a Michael acceptor to form an irreversible covalent bond with the active-site Cys residue of the 3C protease ($3C^{pro}$) from an enterovirus B, EV-93, pathogen.

The structure of the first $3C^{pro}$ from an enterovirus B was solved in complex with rupintrivir and compound 1. Structures showed that rupintrivir and compound 1 present a similar binding mode, with an electrophilic carbon bound covalently to the active-site cysteine residue. Also, these compounds form a stable tetrahedral adduct that results in the irreversible inactivation of the protease.

Rupintrivir was shown to have low toxicity and potent antiviral activity against all human rhinovirus (HRV) serotypes tested, with a mean 50% effective concentration (EC_{50}) of 23 nM, and also against four related enteroviruses (EVs). Activity of both compounds against $3C^{pro}$ was also successfully characterized both in vitro and in infected cells. (Costenaro *et al.* 2011).

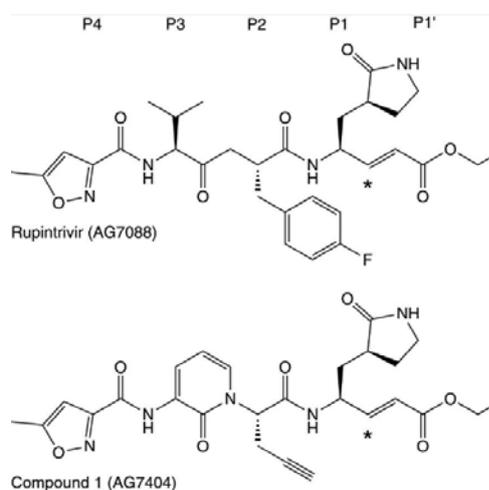


Figure 105. Chemical structure of rupintrivir (AG7088) and its orally bioavailable analogue compound 1 (AG7404). Asterisks indicate β -carbons that make irreversible covalent bonds with active site Cys. (Costenaro *et al.* 2011).

Enterovirus and rhinovirus present a similar mechanism as the one present in coronavirus, as they synthesize a polyprotein which is processed by an enzyme, in this case the 3C^{pro}, with a Cys-His-Glu catalytic triad.

CHAPTER 7: OBJECTIVES

7. OBJECTIVES

The global aim of this part of the thesis was to determine the structure of the SARS-CoV-1 main protease (M^{pro}) with compound 1 and design improved antiviral compounds against CoV M^{pro} by analysis of other structures solved in our laboratory and available in the literature.

The following specific objectives were established:

- I. Obtain high-quality crystals of the SARS-CoV-1 M^{pro} - compound 1 complex.
- II. Determine the three-dimensional structure of SARS-Cov-1 M^{pro} with compound 1.
- III. Characterize the binding mode of the inhibitor to the protein.
- IV. Compare it with the previously solved structures of SARS-CoV-2 M^{pro} with rupintrivir and compound 1 and other complexes available in the literature.
- V. Design new compounds that may have an improved antiviral activity against CoV M^{pro} .

CHAPTER 8: MATERIALS AND METHODS

8.1. MATERIALS

SARS-CoV-1 M^{pro} protein was previously expressed and purified in our lab. In this section the materials needed for its crystallization and structure determination are detailed.

8.1.1. Crystallization

The main materials used for the crystallization of SARS-Cov-1 M^{pro} with compound 1 were the following:

8.1.1.1. Crystallization plates

- ❖ MRC 2 Well Crystallization Plate (Hampton Research): Plates for sitting drop vapour diffusion crystallization. They are suited for high throughput crystallization and are fully compatible with robotic equipment.
- ❖ Cryschem plates (Hampton Research): Plates for crystallization by sitting drop with 24 wells.
- ❖ VDX™ Plate with sealant (Hampton Research): Plates for crystallization by hanging drop with 24 wells.

8.1.1.2. Cryo-crystallography loops

- ❖ Mounted CryoLoop (Hampton Research): It is a nylon loop used to mount, freeze, and secure the crystal during cryo-crystallographic procedures and X-ray data collection.
- ❖ Magnetic CryoVials and CryoCaps (Molecular Dimensions): It is a complete crystal vial for cryocrystallography. It has a magnetic base and cap. It also has a bar code on the base.

8.1.1.3. Robotic platforms

- ❖ Phoenix crystallization robot (Art Robbins Instruments): Protein crystallography dispenser, featuring accuracy, speed and precision. Instrument for sitting drop, hanging drop and microbatch reactions.
- ❖ Crystal farm (Bruker AXS): Instrument for protein crystallization storage and automated imaging.

8.1.1.4. Software

- ❖ XDS: Software for processing crystal X-ray diffraction data (*Kabsch, 1988*)

- ❖ CCP4i: Graphical user interface to run programs from the CCP4 suite (*Potterton et al, 2003*).
- ❖ Refmac: Software for macromolecular refinement (*Murshudov et al, 1997*).
- ❖ Coot: Software for molecular graphics applications and real space refinement (*Emsley & Cowtan, 2004*).

8.2. METHODS

General crystallization and crystallography methods were already described in part I. Here only specific details applied to the project will be detailed.

8.2.1. Complex preparation

Compound 1 powder was weighted and diluted to a 100 mM stock in DMSO. Then, diluted compound 1 was added to a 4 mg/ml sample of M^{pro} 4 mg/mL at a 30-fold molar excess. The sample was incubated overnight at 4°C to allow complex formation. After incubation some precipitation was observed in the tubes, which could be removed by discarding the pellet after centrifugation for 15 minutes at 16,000g and 4°C.

8.2.2. Crystallization trials

A number of crystallization screening plates (Table 28) were set up for SARS-Cov1 M^{pro} with compound 1.

Code	Original screen	Nº of conditions
PAC 3	Index	96
PAC plus	JCSG plus	96
PAC PEG	Clear Strategy	96
BCS	BCS	96
ProPlex	ProPlex	96

Table 28. Automated crystallography platform screens used to crystallize the M^{pro}-compound 1 complex.

8.2.3. Structure determination and analysis

Phenix.refine (Liebschner *et al.* 2019) and Coot (Emsley & Cowtan, 2004) were used for refinement.

Jligand was used to generate the restraints of the covalent bound between the protein and the compound (Lebedev *et al.*, 2012).

CHAPTER 9: RESULTS AND DISCUSSION

9.1. RESULTS AND DISCUSSION

The severe acute respiratory syndrome coronavirus main protease (SARS-CoV-1 M^{pro}) is a key protease of coronavirus that mediates viral replication and transcription. SARS-CoV-1 M^{pro} has been emerged as an attractive target for SARS-CoV-1 drug design and development. M^{pro} holds a pivotal role during the onset of the infection and its function is intimately related with the beginning of viral replication. The interruption of its catalytic activity could represent a relevant strategy for the development of anti-coronavirus drugs.

In this work we report the crystallization and structure determination of SARS-Cov-1 M^{pro} with the inhibitor compound 1.

9.1.1. Crystallization trials

Microcrystals were obtained in a condition that contained 0,2 M LiSO₄ monohydrate, 0,1 M Hepes pH 7.5 and 25% PEG 3350. These crystals were optimized into 24 well plates. Many tiny plate-shaped crystals appeared in the condition 0,2 M LiSO₄ monohydrate, 0,1 M Hepes pH 8 and 27% PEG 3350.

9.1.2. Data collection and structure determination

Crystals from the 24 well optimization plate were frozen with 20% glycerol. Frozen crystals were diffracted at beamline BL13 – XALOC at ALBA synchrotron.

The best data collected diffracted up to 2.528 Å resolution (Figure 106, Figure 107), and it was indexed, integrated, scaled and reduced to unique reflections with XDS (*Kabsch, 2010*). The Table 29 shows the processing statistics. The STARANISO output MTZ file was used for structure refinement. It contains all measurements to the highest observed resolution limit and amplitudes are derived via the French & Wilson method, using the correct anisotropic prior distribution of the expected intensity. Using such file, the values of R_{free} and R_{work} improved considerably.

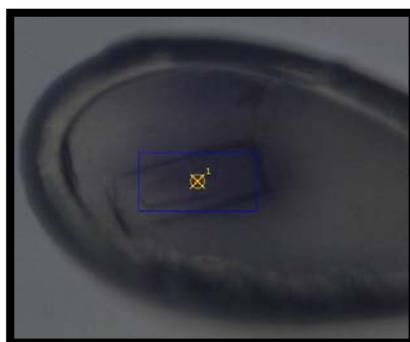


Figure 106. SARS-Cov-1 M^{pro} with compound 1 crystal grown in 24-well plate of the crystallization condition: 0,2 M LiSO₄ monohydrate, 0,1 M Hepes pH 8 and 27% PEG 3350.

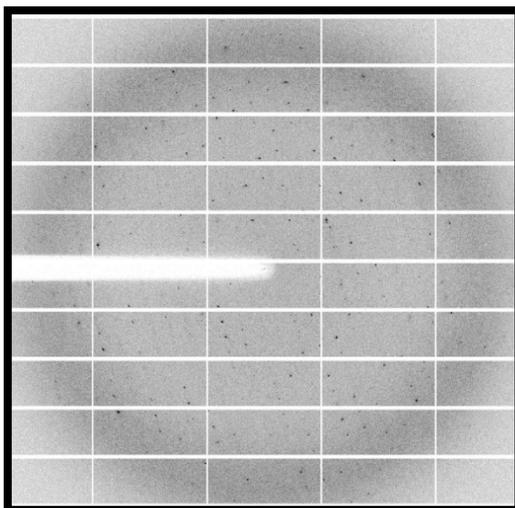


Figure 107. Diffraction pattern of SARS-Cov-1 *M^{pro}* with compound 1 crystal diffracting up to 2.528 Å.

Parameters	Values
Wavelength (Å)	0.97918 Å
Resolution Range	47.875 - 2.528 (2.693 - 2.528)
Space group	P 2 ₁ 2 ₁ 2
Unit cell dimensions	a= 106.854 Å b= 45.220 Å c= 53.550 Å α= 90° β= 90° γ= 90°
Total reflections	32205 (1985)
Unique reflections	7634 (382)
Multiplicity	4.2 (5.2)
Completeness (%)	91.9 (46.3)
Mean I/σ (I)	10.1 (1.1)
Rmerge(%)	9.3 (149.5)
Rmeas (%)	10.6 (166.4)
Rpim (%)	5.1 (72.0)
CC1/2 (%)	99.8 (43.1)

Table 29. Crystallographic data and processing statistics of SARS-Cov-1 *M^{pro}* with compound 1. Overall and last shell (in parentheses) values are shown.

The structure of SARS-Cov-1 Mpro with the inhibitor compound 1 was determined by molecular replacement using Phaser (*McCoy et al., 2007*). The SARS-CoV-1 Mpro crystallographic structure (PDB: 1UK4) solved at 2.50 Å by Yang et al. (2003) was a suitable model to solve the phases. The asymmetric unit of these crystals contained one SARS-CoV-1 M^{pro} monomer with compound 1. The LLG value is 1060 and the TFZ value 28.0.

The electron density map showed an extra density in the active site that corresponded to the compound 1 inhibitor. Compound 1 is covalently bound to the Cys145 of SARS-CoV-1 M^{pro}. The inhibitor was added and manually adjusted in Coot.

9.1.3. Structure refinement and validation

Several rounds of refinement using phenix.refine were performed in combination with manual model correction using Coot. Jligand was used to generate a CIF file containing the restraints for the covalent bond between the reactive C of compound 1 and the S of the Cys of SARS-CoV-1 M^{pro}. The final statistics are shown (Table 30).

Parameters	Values
Resolution (Å)	2.53
R-work/R-free	0.1970/0.2534
r.m.s.d. Bond lengths (Å)	0.018
r.m.s.d Bond angles (°)	2.050
Ramachandran outliers (%)	0.38%
Ramachandran favored (%)	94.65%
Clashcore	21.73
Molprobit score	3.00

Table 30. Refinement statistics for SARS-Cov-1 Mpro with compound 1, as calculated with Molprobit.

Compound 1 perfectly fitted the extra density that appeared in the active site of SARS-CoV-1 M^{pro} (Figure 108).

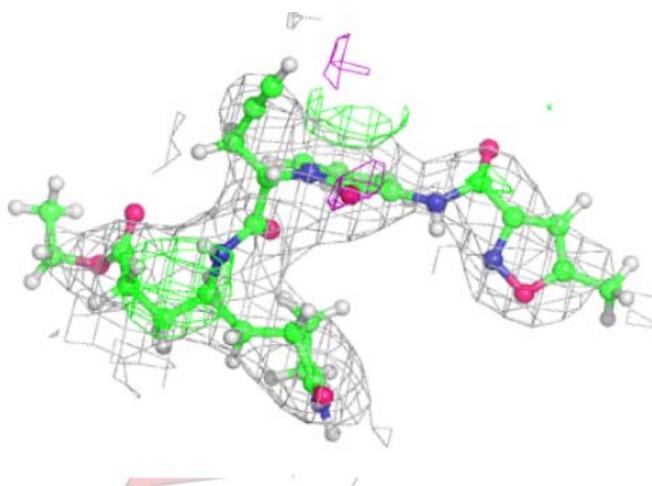


Figure 108. Electron density around compound 1.

All residues were located in allowed regions in the Ramachandran plot (Figure 109) excepting one. This residue is located in a loop, where density is worse due to flexibility.

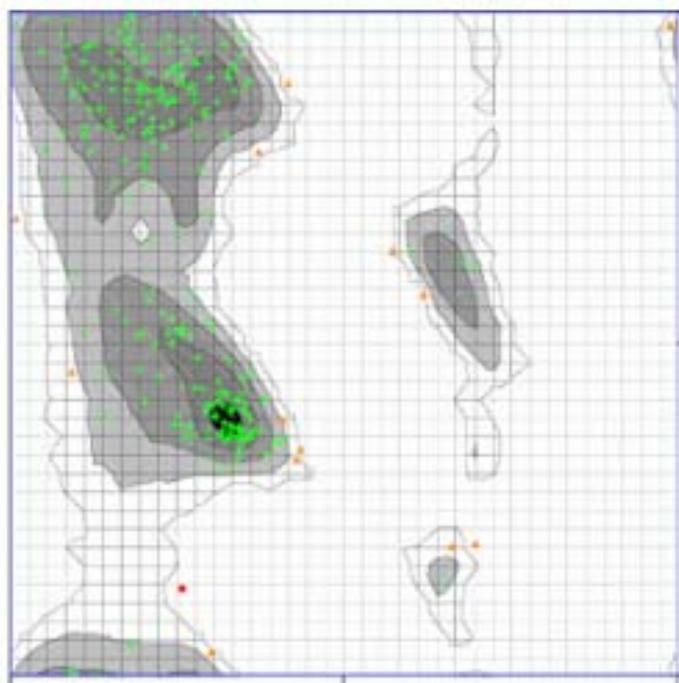


Figure 109. Ramachandran plot for the SARS-CoV-1 with compound1. In green favoured regions (94.65%). In yellow allowed regions (4.96%). In red high energy backbone conformations (0.38%).

9.1.4. Structure analysis

The structure of SARS-Cov-1 M^{pro} together with compound 1 inhibitor is described in detail this section (Figure 110). Although the asymmetric unit of these crystals contained one SARS-CoV-1 M^{pro} monomer with compound 1 when crystallographic symmetry is applied the homodimer is reconstituted.

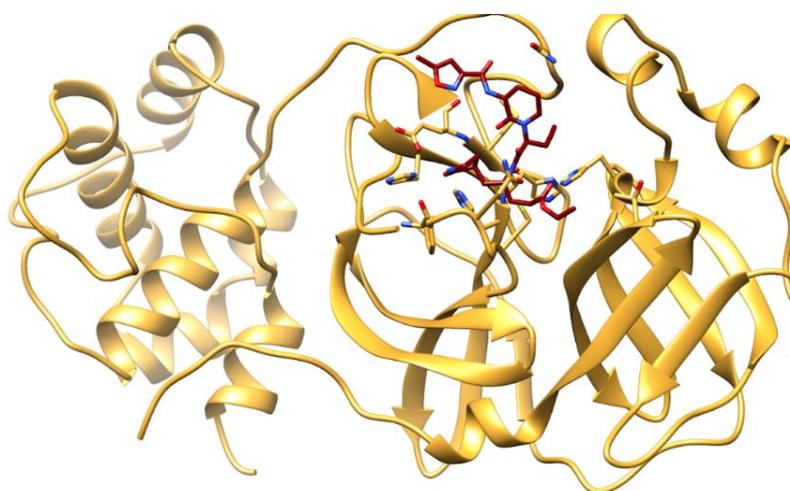


Figure 110. Overall structure of SARS-Cov-1 M^{pro} with compound 1 .

Compound 1 is placed in the cleft between domains I and II, where the Cys-His catalytic dyad is found. Compound 1 forms a covalent bond with the Cys145 which is part of the catalytic dyad. The inhibitor electrophonic β -carbon is bound to the Cys after Michael addition and forms a stable tetrahedral adduct that irreversibly inactivates the protein. The peptidomimetic backbone adopts a partially extended conformation along the subsites of the active site cleft (Figure 111).

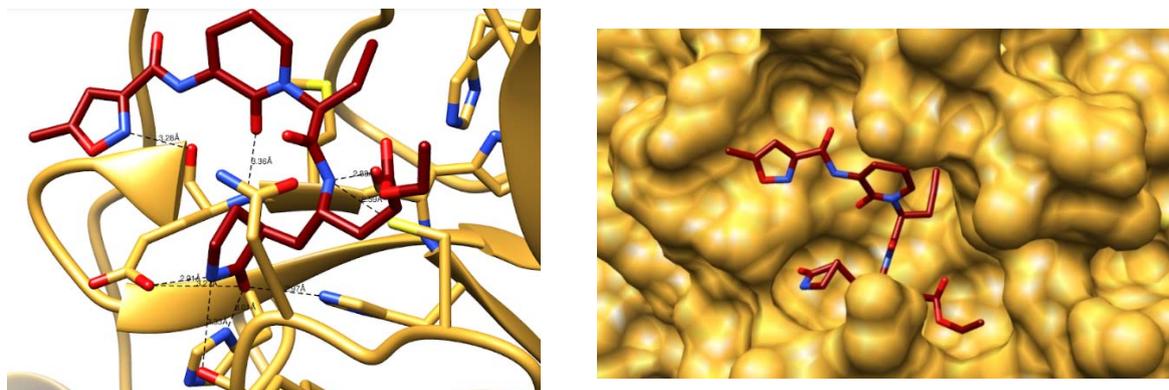


Figure 111. Compound 1 structure in the active site of SARS-Cov-1 M^{pro} . Left, protein shown as ribbon and right, protein shown as surface.

Interactions between the protein and the compound were analysed for each of the inhibitor parts (Table 31). P1 β -lactam ring makes hydrogen bonds with the protein mimicking the recognition of the P1 Gln highly conserved in substrates of the coronavirus M^{pro} . N35 interacts with Glu166 and Phe140, while O37 interacts with Glu166, His172 and His163 (Figure 112). The N8 located between P1 and P2 interacts with the O of the His164 and is at 2.59 Å from the S of the Cys145.

Inhibitor moiety	Atom1	Atom2	Distance
P1	Compound 1 N35	Glu166 OE1	2.91 Å
	Compound 1 N35	Phe140 O	3.33 Å
	Compound 1 O37	Glu166 OE1	3.27 Å
	Compound 1 O37	His172 ND1	3.01 Å
	Compound 1 O37	His163 NE2	2.97 Å
P1-P2	Compound 1 N8	His164 O	2.88 Å
	Compound 1 N8	Cys145 SG	2.59 Å
P3	Compound 1 O26	Glu166 N	3.36 Å
P4	Compound 1 N17	Glu166 O	3.28 Å

Table 31. Hydrogen bonds between SARS-CoV-1 active site and compound1

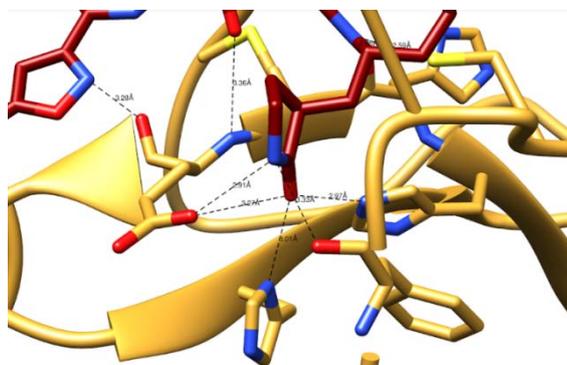


Figure 112. β -lactam ring of compound 1 and its interactions with SARS-Cov-1 M^{pro}.

P2 2-propenyl stacks against His41 from the catalytic dyad. Moreover, residues Met165 and Asp187 further restraint its conformation (Figure 113).

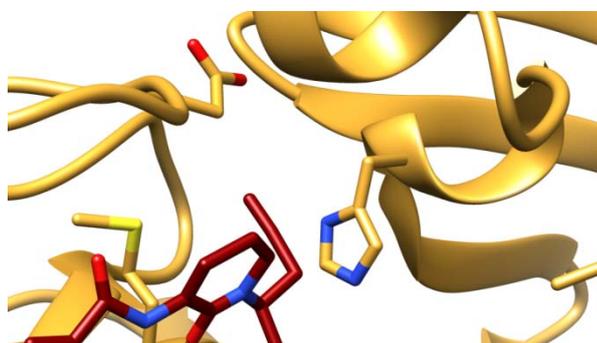


Figure 113. P2 of compound 1 and its interactions with SARS-Cov-1 M^{pro}.

Both P3 and P4 interact with Glu166, while P3 makes one main chain hydrogen bond with the N of the residue, the N of the P4 ring interacts with the carboxyl group (Figure 114).

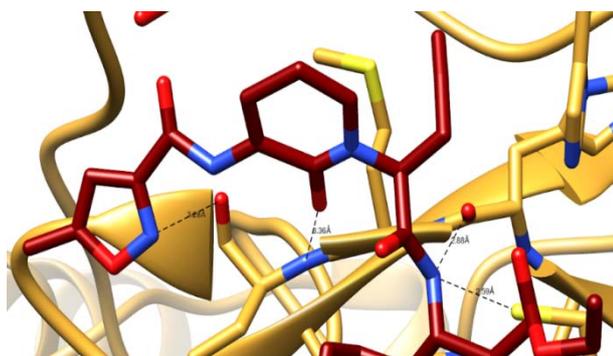


Figure 114. P3 and P4 of compound 1 and its hydrogen bonds with Glu166 of SARS-Cov-1 M^{pro}.

The heteroring of P3 is mainly solvent exposed and interacts with Gln189 on one side at 4.88 Å (Figure 115).

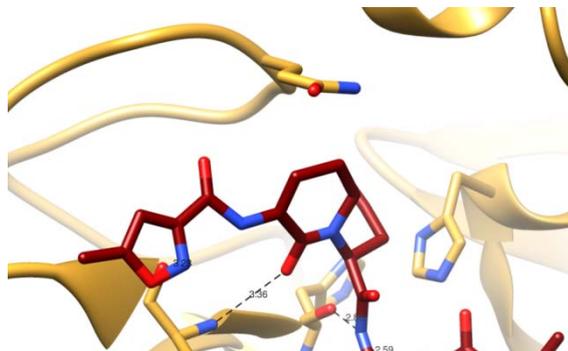


Figure 115. P3 of compound 1 and its interactions with SARS-Cov-1 Mpro.

P4 lies in a groove formed by β -strands placed close to residues Glu166, Leu167 and Pro168 (Figure 116).

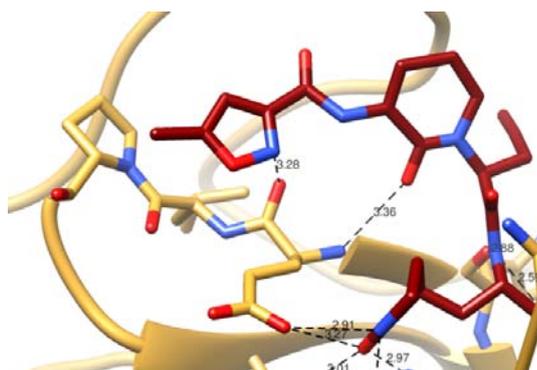


Figure 116. P4 of compound 1 and its interactions with SARS-Cov-1 Mpro.

The electrostatic potential surface shows that the β -lactam ring is in a negatively charged pocket while P3 and P4 are close to slightly negatively charged regions (Figure 117).

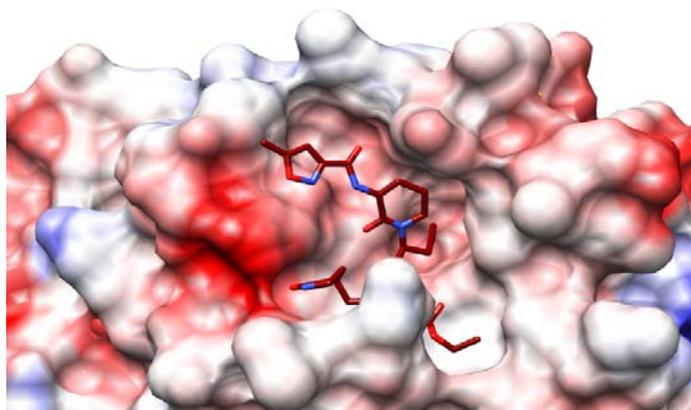


Figure 117. Electrostatic potential surface of SARS-Cov-1 Mpro with compound 1. Positively charged regions are colored in blue and negatively charged regions are colored in red.

9.1.5. Comparison with SARS-CoV-2 Compound 1 and SARS-CoV-2 rupintrivir

Recently, the crystallographic structures of SARS-CoV-2 compound 1 and SARS-CoV-2 rupintrivir have been determined in our laboratory. The SARS-CoV-2 compound 1 structure was solved at 2.26 Å and the SARS-CoV-2 rupintrivir structure at 2.26 Å (PDB: 7P35).

SARS-CoV-2 Compound1

In order to check whether the binding modes of compound 1 to SARS-CoV-1 M^{pro} and SARS-CoV-2 M^{pro} are equivalent, interactions between compound 1 and SARS-CoV-2 M^{pro} were also analyzed (Figure 118, Table 32).

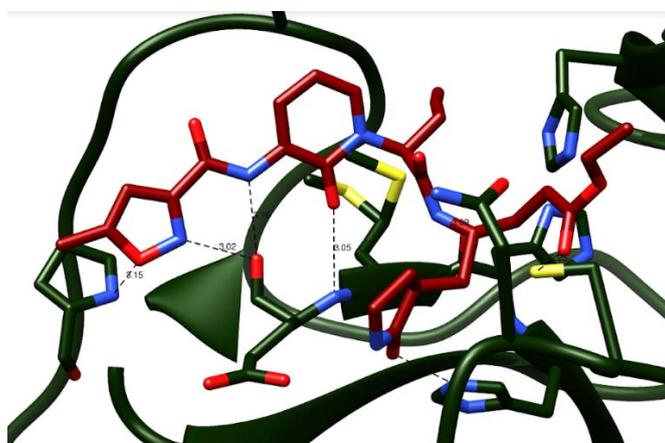


Figure 118. Compound 1 structure in the active site of SARS-Cov-2 M^{pro}.

Inhibitor moiety	Atom1	Atom2	Distance
P1'	Compound 1 O38	Cys145 N	3.14 Å
	Compound 1 O38	Gly143 O	3.24 Å
	Compound 1 O38	Gly143 N	3.40 Å
	Compound 1 O38	Cys145 SG	3.25 Å
P1	Compound 1 N35	Glu166 OE1	2.97 Å
	Compound 1 O37	His163 NE2	2.91 Å
	Compound 1 N35	Phe140 O	3.06 Å
P1-P2	Compound1 N8	His164 O	3.09 Å
	Compound1 N8	Cys145 SG	3.24 Å
P3	Compound 1 O26	Glu166 N	3.05 Å
	Compound 1 O26	Glu166 O	3.37 Å
P3-P4	Compound 1 N14	Glu166 O	3.16 Å
	Compound 1 O22	Thr190 O	3.42 Å
P4	Compound 1 N17	Glu166 O	3.02 Å
	Compound 1 O18	Pro168 N	3.15 Å

Table 32. Hydrogen bonds between SARS-CoV-2 active site and compound1.

The P1' carbonyl oxygen of the ethyl ester is positioned in SARS-CoV-2 structure above the oxyanion hold formed by the amide groups of Gly143, Ser144 and Cys145. This conformation is slightly different to the one observed in SARS-CoV-1. Furthermore, the O38 is placed at 3.25 Å from the S of the Cys145. This S in SARS-CoV-1 was closer to N8. N35 and O37 from the B-lactam interact also with SARS-CoV-2 although they perform less hydrogen bonds (Figure 119).

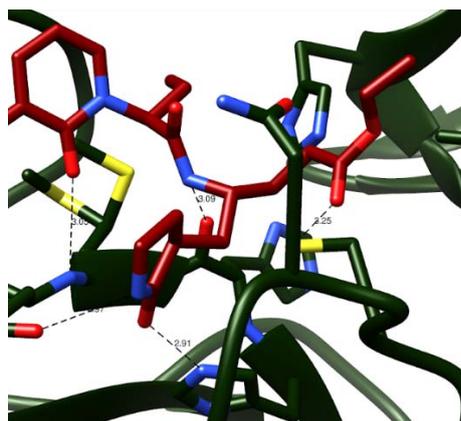


Figure 119. P1', P1 and P2 of compound 1 and its interactions with SARS-Cov-2 M^{pro}.

The rest of the interactions are maintained in both proteins except for an extra hydrogen bond in P4 (Figure 120).

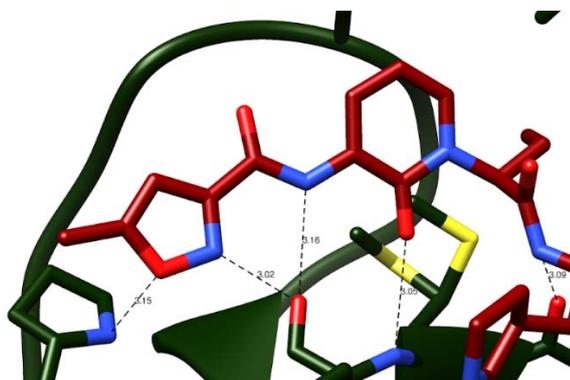


Figure 120. P3 and P4 of compound 1 and its interactions with SARS-Cov-2 M^{pro}.

In summary, we conclude that both proteins bind compound 1 in a similar manner.

SARS-CoV-2 rupintrivir

In order to compare the binding modes of compound 1 and rupintrivir to the M^{pro}, interactions between rupintrivir and SARS-CoV-2 M^{pro} were also analysed. (Figure 121, Table 33). Remarkably, the electron density corresponding to rupintrivir compound was less defined in some zones, specially P2 and P4.

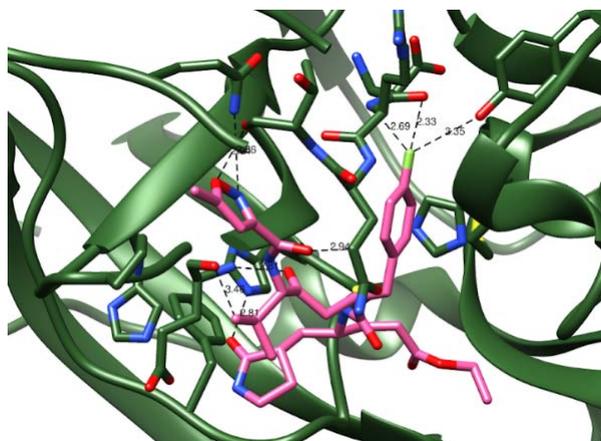


Figure 121. Rupintrivir structure in the active site of SARS-Cov-2 M^{pro}.

Inhibitor moiety	Atom1	Atom2	Distance
P1'	Rupintrivir O23	Gly143 N	2.60 Å
	Rupintrivir O23	Asn142 ND2	2.66 Å
P1	Rupintrivir O18	His163 NE2	2.81 Å
	Rupintrivir O18	Glu166 OE2	3.48 Å
	Rupintrivir N17	Glu166 OE2	2.93 Å
	Rupintrivir N17	Phe140 O	3.09 Å
P1-P2	Rupintrivir N12	Cys145 SG	2.91 Å
P2	Rupintrivir F1	Tyr54 OH	3.35 Å
	Rupintrivir F1	Asp187 O	2.33 Å
	Rupintrivir F1	Arg188 N	2.69 Å
P2-P3	Rupintrivir O3	Glu166 N	2.81 Å
	Rupintrivir O3	Glu166 O	3.09 Å
P3-P4	Rupintrivir N58	Glu166 O	2.68 Å
	Rupintrivir O60	Gln189 OE1	3.17 Å
P4	Rupintrivir O4	Thr190 O	2.94 Å
	Rupintrivir N5	Thr190 O	3.22 Å
	Rupintrivir N5	Glu192 NE2	3.38 Å

Table 33. Distances between SARS-CoV-2 active site and rupintrivir.

The position of the P1' and P1 of rupintrivir are equivalent in compound 1, although the specific hydrogen bonds are not exactly the same. Even though the F in P2 seems to make several interactions, the electron density of this moiety is hardly visible (Figure 122).

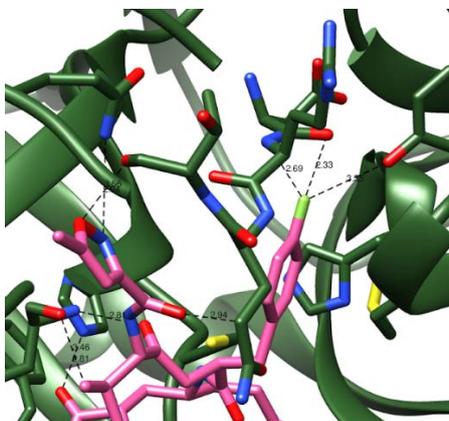


Figure 122. P1 and P2 of rupintrivir and its interactions with SARS-Cov-2 M^{pro}.

Glu166 is forming many hydrogen bonds with N and O atoms around P3 in both cases. The N and O of the P4 ring form hydrogen bonds in both cases but are not exactly the same (Figure 123). However, we must consider the lack of density in the P4 of rupintrivir, which suggests that probably this part might be orientated differently. Therefore, interactions between compound 1 and the protein seem much stronger in this region.

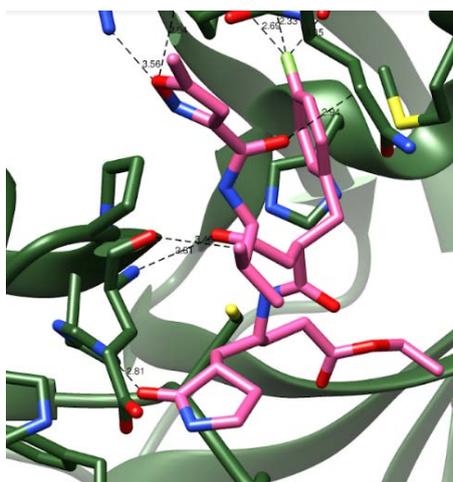


Figure 123. P3 and P4 of rupintrivir and its interactions with SARS-Cov-2 M^{pro}.

9.1.6. Design of novel M^{pro} inhibitors

The analysis of the solved structure SARS-CoV-1 compound 1 and its comparison with the previously solved structures SARS-CoV-2 Compound 1 and SARS-CoV-2 rupintrivir served as a guide to design novel M^{pro} inhibitors. We also based our design in two compounds developed by Pfizer, PAXLOVID (PF-07321332) (Macchiagodena *et al.*, 2022) and PF00835231 (Halford, 2020). Both are M^{pro} inhibitors that have demonstrated potent *in vitro* antiviral activity against SARS-CoV-2 and other coronaviruses.

The electron density of the P1' carbonyl oxygen of the ethyl ester is not as well defined as in the 3C protease of Enterovirus. In the case of SARS-CoV-1 it is even not well defined in the electron density map. For this reason, making it smaller could be an improvement. M^{pro} inhibitor 1 (Figure 124) and M^{pro} inhibitor 4 (Figure 127) have a reduced P1', while M^{pro} inhibitor 2 (Figure 125) shows the imine nitrogen present in the P1' of PF07321332 from Pfizer. M^{pro} inhibitor 5 (Figure 128) has also the imine nitrogen but is quite smaller. M^{pro} inhibitor 3 and 6 (Figure 126, Figure 129) present the same P1' as PF00835231.

The β -lactam from P1 has been maintained in all the designed inhibitors since the electron density fitted this ring perfectly. The electron density from the P2 moiety was not well defined neither in rupintrivir nor in compound 1, but specially in the first case. Since there is enough space, we think that keeping an hydrophobic group but making it slightly bigger than in compound 1 could be a good choice.

Taking into account that the P3 moiety was better defined in compound 1 than in rupintrivir, we decided to close the ring and maintain it as in compound 1 in all the designed inhibitors. The P4 moiety is similar to the one of compound 1 except for the methyl group for M^{pro} inhibitors 1, 2 and 3 (Figure 124, Figure 125, Figure 126). In the case of M^{pro} inhibitors 4, 5 and 6 (Figure 127, Figure 128, Figure 129) P3 and P4 rings are together, making the compound smaller.

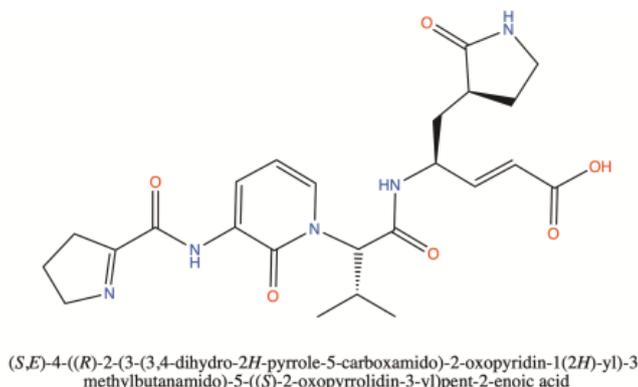
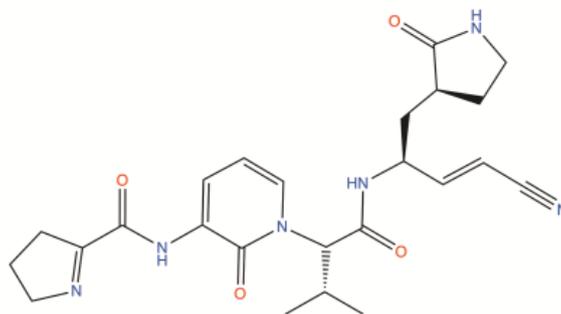
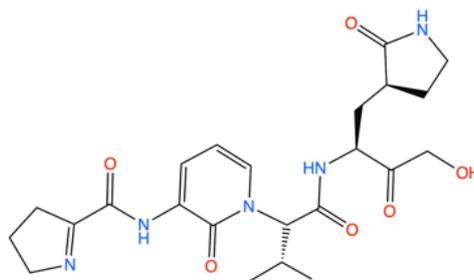


Figure 124. M^{pro} inhibitor 1.



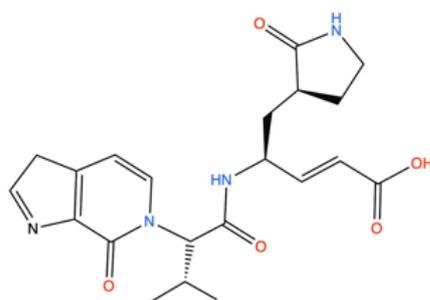
N-(1-((*R*)-1-(((*S,E*)-4-cyano-1-((*S*)-2-oxopyrrolidin-3-yl)but-3-en-2-yl)amino)-3-methyl-1-oxobutan-2-yl)-2-oxo-1,2-dihydropyridin-3-yl)-3,4-dihydro-2*H*-pyrrole-5-carboxamide

Figure 125. *M^{pro}* inhibitor 2.



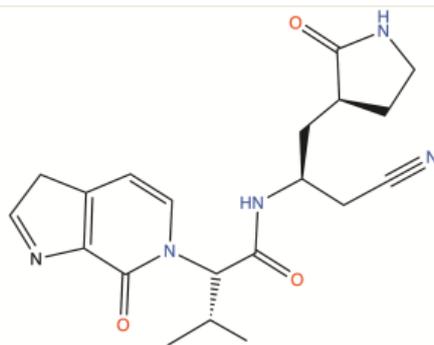
N-(1-((*R*)-1-(((*S*)-4-hydroxy-3-oxo-1-((*S*)-2-oxopyrrolidin-3-yl)butan-2-yl)amino)-3-methyl-1-oxobutan-2-yl)-2-oxo-1,2-dihydropyridin-3-yl)-3,4-dihydro-2*H*-pyrrole-5-carboxamide

Figure 126. *M^{pro}* inhibitor 3.



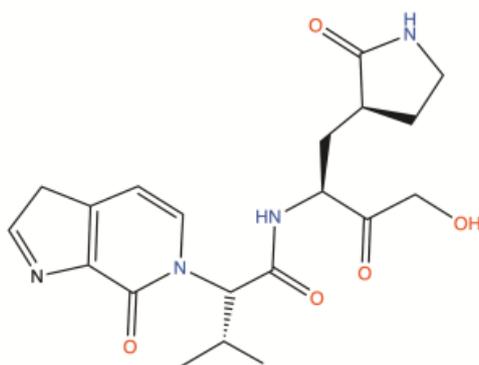
(*S,E*)-4-((*R*)-3-methyl-2-(7-oxo-3,7-dihydro-6*H*-pyrrolo[2,3-*c*]pyridin-6-yl)butanamido)-5-((*S*)-2-oxopyrrolidin-3-yl)pent-2-enoic acid

Figure 127. *M^{pro}* inhibitor 4.



(R)-*N*-((*S*)-1-cyano-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)-3-methyl-2-(7-oxo-3,7-dihydro-6*H*-pyrrolo[2,3-*c*]pyridin-6-yl)butanamide

Figure 128. *M^{pro}* inhibitor 5.



(R)-*N*-((*S*)-4-hydroxy-3-oxo-1-((*S*)-2-oxopyrrolidin-3-yl)butan-2-yl)-3-methyl-2-(7-oxo-3,7-dihydro-6*H*-pyrrolo[2,3-*c*]pyridin-6-yl)butanamide

Figure 129. *M^{pro}* inhibitor 6.

CHAPTER 10: CONCLUSIONS

1. SARS-CoV-1 M^{pro} compound 1 complex was crystallized in the space group P2₁2₁2. The crystals diffracted at 2.53 Å at the ALBA synchrotron with overall good quality statistics.
2. It was possible to solve the structure of SARS-CoV-1 M^{pro} compound 1 complex by molecular replacement using Phaser. The SARS-CoV-1 crystallographic structure solved by Yang et al. (2003) was a suitable model to solve the phases by the molecular replacement method.
3. The asymmetric unit of SARS-CoV-1 M^{pro} compound 1 complex crystals consists of 1 monomer that has a compound 1 molecule covalently attached to Cys145.
4. SARS-CoV-1 and SARS-CoV-2 M^{pro} bind compound 1 in an equivalent manner, also similar to the binding mode of rupintrivir to SARS-CoV-2 M^{pro}.
5. The P1' carbonyl oxygen of the ethyl ester is positioned in SARS-CoV-2 structures with rupintrivir and compound 1 above the oxyanion hole formed by the amides of residues Gly143, Ser144 and Cys145, being this conformation slightly different to the one observed in SARS-CoV-1.
6. P1 β-lactam ring of compound 1 makes hydrogen bonds with the SARS-CoV-1 M^{pro} mimicking the recognition of the P1 Gln highly conserved in substrates of the coronavirus M^{pro}. Same atoms from the β-lactam of rupintrivir and compound 1 interact also with SARS-CoV-2 although they perform less hydrogen bonds.
7. P2 2-propinyl from compound 1 stacks against His41 from the catalytic dyad. Although electron density is not strong in this part of the inhibitor in compound 1 structures, the equivalent moiety in SARS-CoV-2 M^{pro} structure with rupintrivir is even less well defined.
8. The heteroring of compound 1 P3 is mainly solvent exposed. Glu166 is forming many hydrogen bonds with N and O atoms around P3 of both compound 1 and rupintrivir in all the structures analysed.
9. Compound 1 P4 lies in a groove formed by β-strands. Considering the lack of density in the P4 of rupintrivir, interactions between compound 1 and the protein seem much stronger in this region.
10. Considering all the structural data and the literature, new compounds were designed to obtain better inhibitors against the SARS-CoV-1 and SARS-CoV-2 M^{pro} targets.

BIBLIOGRAPHY

Costenaro, L., Kaczmarek, Z., Arnan, C., Janowski, R., Coutard, B., Solà, M., Gorbalenya, A. E., Norder, H., Canard, B., & Coll, M. Structural basis for antiviral inhibition of the main protease, 3C, from human enterovirus 93. *Journal of virology* (2011).

Drożdżal, S., Rosik, J., Lechowicz, K., Machaj, F., Szostak, B., Przybyciński, J., Lorzadeh, S., Kotfis, K., Ghavami, S. & Łos, M.J. An update on drugs with therapeutic potential for SARS-CoV-2 (COVID-19) treatment. *Drug Resistance Updates* (2021).

Halford, B. Pfizer's novel COVID-19 antiviral heads to clinical trials. *Chemical and Engineering News* (2020).

Han, F., Liu, Y., Mo, M., Chen, J., Wang, C., Yang, Y. & Wu, J. Current treatment strategies for COVID-19 (Review). *Molecular Medicine Reports* (2021).

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Poon, B.K., Prisant, M.G., Read, R.J., Richardson, J.S., Richardson, D.C., Sammito, M.D., Sobolev, O.W., Stockwell, D.H., Terwilliger, T.C., Urzhumtsev, A.G., Videau, L.L., Williams, C.J. & Adams, P.D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica* (2019).

Macchiagodena, M., Pagliai, M. & Procacci, P. Characterization of the non-covalent interaction between the PF-07321332 inhibitor and the SARS-CoV-2 main protease. *Journal of Molecular Graphics and Modelling* (2022).

Mengist, H.M., Dilnessa, T. & Jin T. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Frontiers in Chemistry* (2021).

Meyer, B., Chiaravalli, J., Gellenoncourt, S. et al. Characterising proteolysis during SARS-CoV-2 infection identifies viral cleavage sites and cellular targets with therapeutic potential. *Nature Communications* (2021).

Mousavizadeh, L. & Ghasemi S. Genotype and phenotype of COVID-19: Their roles in pathogenesis. *Journal of Microbiology, Immunology and Infection* (2021).

Ullrich, S & Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorganic & Medicinal Chemistry Letters*. (2020).

Wallace, A.C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering* (1995).

Zeng, Z. Q., Chen, D. H., Tan, W. P., Qiu, S. Y., Xu, D., Liang, H. X., Chen, M. X., Li, X., Lin, Z. S., Liu, W. K., & Zhou, R. Epidemiology and clinical characteristics of human coronaviruses OC43, 229E, NL63, and HKU1: a study of hospitalized children with

acute respiratory tract infection in Guangzhou, China. *European journal of clinical microbiology & infectious diseases* : official publication of the European Society of Clinical Microbiology (2018).

Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A., & Murshudov, G. N. Jligand: a graphical tool for the CCP4 template-restraint library. *Acta crystallographica* (2012).

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G.F., Anand, K., Bartlam, M., Hilgenfeld, R., Rao, Z. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proceedings of the National Academy of Sciences* (2003).