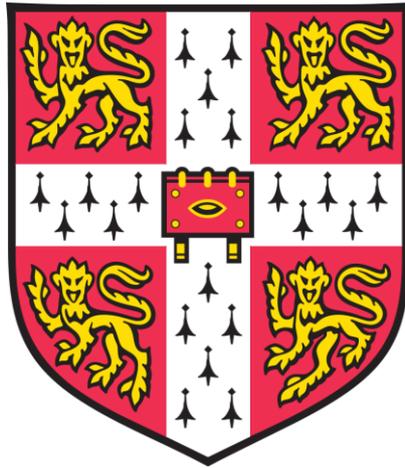


Towards optimising HIV-1 derived lentiviral vectors through structure informed genome modifications



Eirini Vamva

Department of Medicine

University of Cambridge

This dissertation is submitted for the degree of *Doctor of Philosophy*

Declaration

I, Eirini Vamva, hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others. This dissertation contains fewer than 60,000 words excluding appendices and bibliography.

1. Abstract

Eirini Vamva

Towards optimising HIV-1 derived lentiviral vectors through structure informed genome modifications

Lentiviral vectors are being successfully used as therapeutic agents in a series of clinical applications of gene therapy, genome editing and cancer immunotherapy. 3rd generation HIV-1 derived lentiviral vectors are produced from 4 independent plasmids. Here, I focus on the transfer vector that contains the therapeutic gene and the *cis*-acting elements that drive its expression including the packaging signal (*psi*). Lentiviral vector particles carry two copies of transfer vector RNA that become linked via a process known as dimerisation. I focused on improving the infectivity of vectors by targeting their dimerisation and packaging properties based on the hypothesis that WT HIV-1 regulates genome encapsidation tightly by recognising dimeric RNA. The genomic RNA (gRNA) leader region is thought to act as a switch between the monomeric conformation that is associated with translation and the dimeric conformation linked with packaging. I therefore attempted to identify the structures that are important for packaging and optimise those at the expense of structures that are used by the virus for translational control. To do so, I created mutants in the 5'UTR that target regions that play important roles in the process of dimerisation including the Dimerisation Stem Loop (DSL), the U5-AUG duplex formed by sequences located at the beginning of the U5 region and nucleotides surrounding the start of the *gag* gene, and the polyA stem loop, a region suppressed in the 5'LTR, but known to regulate polyadenylation in the 3'LTR. The introduction of these mutations aimed to create vectors whose RNA is more likely to adopt the dimeric conformation and therefore be packaged. To evaluate this, I developed a novel competitive RT-qPCR assay to measure the relative packaging efficiencies (RPE) of transfer vectors in a competitive co-transfection environment. Biochemical characterisation showed an overall negative effect of the introduced mutations on viral infectivity. Northern Blots confirmed that the propensity of mutated vector RNA to dimerise has increased in the mutants as hypothesised. Here, I report the effect of the dimerisation-stabilising mutations on infectious and physical titres of lentiviral vectors, as well as on their

packaging efficiency measured with our novel competitive qPCR assay. Our data suggest that enhancing dimerisation does not automatically lead to better packaging of vector RNA. Despite the improvement of vector RNA dimerisation efficiency and in some cases RPE, the effect of introduced mutations on the ability of the designed transfer vectors to successfully transduce 293T cells was negative, reflecting the multifunctionality of the HIV-1 leader regions and the significance of RNA flexibility. Finally, I explored by Selective 2'-Hydroxyl Acylation Analysed by Primer Extension (SHAPE) the structure of *psi* in our vector RNAs, in particular studying the influence of regions adjacent to *psi* on dimerisation and packaging. SHAPE identifies the RNA backbone flexibility, which is an indication of whether nucleotides are base-paired or not. Our single nucleotide level structural analysis revealed that the presence of *gag* sequences stabilise the *psi* element of the dimeric RNA, suggesting their role in supporting a stable RNA conformation that can be packaged and offering a potential explanation for their requirement in the transfer vector plasmid for maintenance of infectious titres. These findings will give us better insights into the biology of lentiviral vectors and enable us to design more efficient vectors for a variety of clinical applications.

Acknowledgments

My deepest thanks go to my academic supervisors Andrew Lever and Julia Kenyon who gave me the opportunity to pursue a PhD. I would like to thank them for their guidance on the direction of my research project and for helping me become an independent researcher. I would also like to thank my industrial supervisor Conrad Vink for his endless intellectual input on lentiviral vector design and for believing in the 23 year old Eirini. I would like to thank all three of them for their helpful contributions to the writing of this thesis and for teaching me to write in academic style. Additionally, I would like to thank Nina Kotsopoulou, my first ever manager for hiring me at GSK and introducing me to the world of gene therapy and HIV-1 derived lentiviral vectors and Lakis Liloglou, my first ever academic supervisor, for giving me the opportunity to move to the U.K. to work in his lab, for his unconditional support and for teaching me how to mix reagents properly. Last but not least, Harriet McClean for always being there to help, even when she didn't have to, and for being the loveliest and most caring colleague anyone could wish for.

On a personal note, I would like to thank from the bottom of my heart my forever best friend in life Spiros Papasotiriou, for being the best friend or rather brother anyone could hope for and my two best friends in Cambridge Katerina Geladaki and Luis Nobre for being my family not only when I was fun, but also when I was going through my darkest times. My time in Cambridge would have been impossible without you. Thank you for all the moments. I love you so much. Katerina for teaching me how to be beyond liberal, Luis for always being bitter with me and Spiros for showing me what unconditional giving means.

I would like to thank everyone who has helped me run and shape the Darwin Cuban Salsa society. Everyone who has joined my society, became my friend through this society, danced with me in performances and socials, and crazy enough everyone who joined me in a trip to Cuba, especially Pablo and Fernando who became my white boys! I would have never imagined that this small Salsa society I set up in 2016 would have turned into the biggest and happiest society I could wish for by 2019. Thank you for letting me become a leader and some sort of Queen. But most importantly, thank you for helping me create a happy space for all Cambridge students. The list of people is endless but, I would like to thank the people who have selflessly helped me throughout my presidency. Alex for starting

the society with me! Also Aleix, who left early for (hopefully) a better place, and Chris Hope who became my friends in and outside the society and for literally always treating me like a Queen. I would also like to thank Joel for giving me the experience of the Salsa & Galapagos live performance and for establishing the Eirini move. Jyothi for giving me the best bed to sleep on during the last phase of my PhD and for showing me what a real sunshine is.

I would like to thank all of my rowing Darwin W2 teams for becoming my rowing family. For always pushing for each other and for teaching me how to go through an insane amount of physical pain and cold. The royal rowing Eirini of the first year could have never predicted the slaying version blood-on-bow Eirini of the last three years. Thank you for teaching me to appreciate and believe in female powerhouses and for welcoming me in my first ever team sport. In particular, I would like to thank Tatjana, as rowing would have been nothing without her; for all the motivation, dedication, hard core training and spiralling. Thank you for all the E&T moments. Eliska for the emotional rollercoasters we've been through together and for showing me how someone can be so selfless. Rowing has also given me my fabulous MRBs group including Katarina, Alicia, Jules, Dominique, T and Eli. Thank you for letting me experience what strong girls being there for each other means. You are all so fabulous in your own way. Lastly, rowing has given me my Addenbrooke's lunch crew that on top of T and Eli includes the two lovely creepy German boys, Hendrick and Toby, who alongside Moe and Amal consist of the pancake crew. Thank you for always boosting my confidence and letting me be sassy to you. You aren't very German in my head. Last but not least, Will, for being the most generous and kind coach I have ever seen in my life.

Darwin College for giving me the space to create a Cuban Salsa society, become a bartender at DarBar, allow me to develop into a savage debater at the Debate club and be an empowering part of the W2 rowing team. From college, I'd like to thank my Darwin PhD buddy Stefan for being a much better friend than I have been to him. Also, Jezebel and Torben with whom I got to hang out more when I stated writing my PhD. Thank you for sharing some of your lovely craze with me. And Yiannis for all the penne salmon he's cooked for me. From my ex GSK life, I would like to thank Albert for all the discussions on life with a romantic approach. From my Cambridge Greek friends, Orestis and Eirini, for being extremely highly intelligent and empathetic and for still looking up to me despite my endless

complaining. Ioanna, who is such an inspiring, wise and lovely human being and with whom I have had endless chats on love; sadly I still can't learn from her. Lastly, from my ex London life, I would like to thank Chronis for all his support during my emotional collapses. Thank you for picking me up and always trying to talk reason to this brain of mine. I really hope I haven't forgotten anyone who's made me feel special during these four years of my life <3

Για τους γονείς μου, δεν έχω αρκετές λέξεις για να περιγράψω πόσο ευτυχής νιώθω για την τόση αγάπη που έχω εισπράξει από εσάς. Το να κάνεις διδακτορικό είναι piece of cake μπροστά στις θυσίες που έχετε κάνει όλα τα χρόνια για εμένα. Σας ευχαριστώ που με φέρατε στη ζωή και μου δώσατε όλο το raw material για να έχω την αυτοπεποίθηση να διεκδικώ πράγματα. Μόνο αν μπορούσατε να με είχατε κάνει λιγότερο συναισθηματική και εγωίστρια. Ελπίζω να έχουμε ακόμη πολλά χρόνια να μοιραστούμε (ή κοινώς να μαλώνουμε) και να συνεχίσετε να είστε περήφανοι για εμένα! Την γιαγιά μου, που είναι ο καλύτερος άνθρωπος που έχει περάσει από την γη και που όσο μεγαλώνω θαυμάζω όλο και περισσότερο, αν είναι δυνατό.

Για το νευρικό μου σύστημα, που κάπως επιβίωσε και εξελίχθηκε μέσα από την συνολική εμπειρία. For in Cambridge I lived some of my highest and darkest moments...

And as a reminder that the hardbound copy of this thesis was prepared and submitted the last day the university was open during the COVID-19 pandemic outbreak. Quite a fitting end to this experience!

*Τοὺς Λαιστρυγόνας καὶ τοὺς Κύκλωπας,
τὸν ἄγριο Ποσειδώνα δὲν θὰ συναντήσεις,
ἂν δὲν τοὺς κουβανεῖς μὲς στὴν ψυχὴ σου,
ἂν ἡ ψυχὴ σου δὲν τοὺς στήνει ἔμπρός σου.*

Κωνσταντῖνος Π.Καβάφης-Ιθάκη

Nunca se sufre por amor, se sufre por desamor, por desencanto o por indiferencia pero nunca por amor. El amor no lastima... los que lastiman son las personas que no saben amar.

Για τους γονείς μου

Ευχαριστώ για το DNA αλλά πιο πολύ για την αγάπη

Para Aleix

Te quiero mucho mi querido amigo. I hope you left us for a better world...

Στην γιαγιά μου

Που δεν πρόλαβα ποτέ να αποχαιρετήσω <3

Table of Contents

| | |
|--|----|
| Declaration..... | 2 |
| 1. Abstract..... | 3 |
| Acknowledgments | 5 |
| List of Figures..... | 13 |
| List of tables | 16 |
| Abbreviations | 17 |
| 2. Introduction..... | 20 |
| 2.1. HIV-1 biology..... | 20 |
| 2.1.1. The HIV-1 genome..... | 20 |
| 2.1.2. Overview of the gRNA journey..... | 21 |
| 2.1.3. Gag/gRNA interaction and packaging..... | 22 |
| 2.1.4. Importance of RNA structure | 24 |
| 2.1.5. Techniques used to study RNA structure | 25 |
| 2.1.6. Structural packaging signals in the HIV-1 leader | 28 |
| 2.1.7. Structure of the HIV-1 RNA genome..... | 29 |
| 2.1.8. HIV-1 dimerisation | 30 |
| 2.1.9. HIV-1 RNA packaging..... | 32 |
| 2.1.10. The relationship between RNA dimerisation and packaging..... | 33 |
| 2.1.11. Packaging vs Translation: Maintaining the Balance..... | 34 |
| 2.1.12. Established methods for RNA packaging measurement | 37 |
| 2.2. Lentiviral vectors for gene therapy | 38 |
| 2.2.1. History of gene therapy | 38 |
| 2.2.2. Viral vectors as gene delivery tools | 41 |
| 2.2.3. Evolution of lentiviral vector design..... | 43 |
| 2.2.4. Properties of lentiviral vectors..... | 46 |
| 2.2.5. Clinical use of lentiviral vectors..... | 48 |
| 2.2.6. SIN (Self Inactivating) LVs | 49 |
| 2.2.7. Transfer Vector design..... | 51 |
| 2.2.8. Provisional lentiviral vector design modifications..... | 52 |
| 2.2.9. Transduction efficiency-What are the challenges?..... | 54 |
| 2.3. Aims of the project | 56 |
| 3. Materials and Methods | 57 |
| 3.1. Lentivirus production by transient transfection | 57 |
| 3.2. Vector titration with flow cytometry | 57 |

| | |
|--|-----|
| 3.3. Design and cloning of mutants | 58 |
| 3.3.1. Mutagenesis primer design..... | 58 |
| 3.3.3. Transfer of mutant inserts into pCCL..... | 61 |
| 3.4. Intracellular RNA purification | 61 |
| 3.5. Extracellular RNA purification | 61 |
| 3.6. RT-qPCR assays..... | 62 |
| 3.6.1. WPRE RT-qPCR..... | 62 |
| 3.6.2. Competitive RT-qPCR BglG and MSL assay..... | 63 |
| 3.7. Production of a DIG-labelled RNA probe for Northern Blotting..... | 64 |
| 3.8. Northern Blotting | 65 |
| 3.9. In-Gel SHAPE | 65 |
| 3.9.1. <i>In Vitro</i> production of RNA | 65 |
| 3.9.2. In-gel chemical acylation of RNAs | 66 |
| 3.9.3. Reverse Transcription and Structural analysis | 67 |
| 3.10. Western Blotting of viral supernatants | 68 |
| 3.11. Cell lysate preparation and Western Blotting..... | 68 |
| 3.12. ELISA p24 assay..... | 69 |
| 3.13. ELLA p24 assay..... | 70 |
| 4. Rational design and biochemical characterisation of transfer vector mutants | 71 |
| 4.1. Verification of integrity of 3 rd generation lentiviral plasmids | 71 |
| 4.2 Transfection optimisation | 73 |
| 4.2.1. Calculation of DNA mass..... | 74 |
| 4.2.2. Calculation of transfection and transduction efficiencies | 75 |
| 4.3. Design and cloning of U5-AUG mutants | 77 |
| 4.4. Cloning of S1 (Abbink <i>et al.</i> , 2003) and LU5-AUG (Lu <i>et al.</i> , 2011) constructs..... | 82 |
| 4.5. Rational design and cloning of dimerisation mutants | 84 |
| 4.6. Measurement of viral titres via Flow cytometry..... | 86 |
| 4.7. Measurement of virion production..... | 90 |
| 4.8. Effect of mutations in 5'UTR on Gag processing and budding | 93 |
| 4.9. Chapter conclusions..... | 96 |
| 5. RNA dimerisation propensity..... | 98 |
| 5.1. Probe generation | 98 |
| 5.2. Non-denaturing (TBE) Northern Blots | 99 |
| 5.3. <i>In Vitro</i> Dimerisation RNA Studies | 104 |
| 5.3.1. <i>In vitro</i> synthesis of RNA for dimerisation studies..... | 104 |
| 5.3.2. Optimisation of <i>in vitro</i> RNA dimerisation..... | 106 |

| | |
|--|-----|
| 5.3.3. <i>In vitro</i> dimerisation of WT and mutant RNA..... | 110 |
| 5.4. Chapter conclusions..... | 112 |
| 6. Measurement of relative packaging efficiency (RPE) by RT-qPCR | 113 |
| 6.1. Measurement of the packaging efficiency of rationally designed U5-AUG mutants with a WPRES-targeting RT-qPCR assay | 113 |
| 6.2. Design of a novel competitive qPCR assay to measure RPE..... | 116 |
| 6.3. Cloning of transfer vectors containing the unique BglG and MS2 sequences | 119 |
| 6.4. Optimisation of a competitive qPCR assay to measure RPE | 120 |
| 6.5. Optimisation of a qPCR protocol for genetic material extraction and data analysis to measure RPE | 130 |
| 6.6. Use of a novel qPCR protocol to measure relative packaging efficiency of lentiviral vectors in a competitive environment | 139 |
| 6.7. Chapter conclusions..... | 144 |
| 7. RNA structural analysis with in-gel SHAPE..... | 146 |
| 7.1. In-gel SHAPE methodology..... | 147 |
| 7.2. Structural analysis of the monomeric WT HIV-1 RNA | 150 |
| 7.3. Structural analysis of the dimeric HIV-1 leader | 154 |
| 7.4. Structural analysis of the J9+Gag mutant..... | 162 |
| 7.5. Structural analysis of the S1 mutant | 165 |
| 7.6. Chapter conclusions..... | 169 |
| 8. Discussion..... | 171 |
| 8.1. Conclusions | 171 |
| 8.2. Future directions | 178 |
| 8.2.1. <i>In vivo</i> SELEX for the production and selection of best replicating variants..... | 178 |
| 8.2.2. Improvement of gRNA stability..... | 180 |
| 8.2.3. Investigation of the effect of secondary gRNA structure on other parts of the HIV-1 life cycle..... | 181 |
| 9. Bibliography..... | 182 |

List of Figures

| | |
|---|-----|
| Figure 1. HIV-1 genome organisation | 20 |
| Figure 2. Overview of the assembly pathway of an HIV-1 virion ²⁸⁶ | 22 |
| Figure 3. Common structural motifs of RNA ³⁴ | 25 |
| Figure 4. Secondary structural conformation of the one half of the dimeric HIV-1 leader (hemidimer)..... | 28 |
| Figure 5. Conformational changes of the HIV-1 leader featuring a) a monomeric RNA structure, b) a dimeric RNA structure and c) a dimer interaction. Adapted from source: ⁹ | 35 |
| Figure 6. Representation of vectors used in clinical trials. Adapted from source: ¹⁴⁹ | 42 |
| Figure 7. Schematic presentation of the evolution of lentiviral vector design..... | 45 |
| Figure 8. Schematic representation of non-SIN and SIN vector. | 50 |
| Figure 9. Components of a typical 3rd generation lentiviral transfer vector..... | 52 |
| Figure 10. Plasmid maps of constructs comprising the 3rd generation lentiviral vector system. | 72 |
| Figure 11. Agarose gel electrophoresis of 3rd generation lentiviral plasmids..... | 73 |
| Figure 12. Measurement of Transfection efficiency with fluorescent microscopy. | 75 |
| Figure 13. Measurement of transduction efficiency..... | 76 |
| Figure 14. Design of mutants with either strengthened or impaired U5-AUG base pairing. | 78 |
| Figure 15. mFold predicted secondary structures of WT and mutants. | 78 |
| Figure 16. Intermediate steps of the cloning process..... | 79 |
| Figure 17. Agarose gel analysis of intermediates during the cloning process. | 81 |
| Figure 18. Design and in vitro dimerisation propensity of S1 and LU5AUG mutants..... | 82 |
| Figure 19. Stages of cloning for S1 and LU5AUG transfer vector mutants..... | 83 |
| Figure 20. Depiction of dimerisation mutant design targeting the polyA, U5-AUG and SL-1 structural motifs. Adapted from source: ⁸ | 84 |
| Figure 21. Agarose gel analysis of intermediates during the cloning process of the dimerisation mutants..... | 86 |
| Figure 22. Comparison of infectious titres between WT and U5s viral vectors..... | 87 |
| Figure 23. Combined comparison of infectious titres between WT and U5s viral vectors. | 88 |
| Figure 24. Comparison of infectious titres between WT, U5s viral vectors and the S1 and LU5AUG U5-AUG mutants. | 89 |
| Figure 25. Comparison of infectious titres between WT and the dimerisation vector mutants designed: PolyAs, J9+Gag, PolyAsw and SIVS1. | 90 |
| Figure 26. Comparison of virion production between the various mutants..... | 91 |
| Figure 27. Measurement of virion production for the U5-AUG and dimerisation mutants..... | 92 |
| Figure 28. Investigation of Gag processing of first group of transfer vector mutants..... | 93 |
| Figure 29. Gag processing study of U5-AUG mutants. | 94 |
| Figure 30. Effects of 5'UTR mutations on Gag budding. | 95 |
| Figure 31. Production of an RNA probe that targets eGFP and WPRE sequences..... | 99 |
| Figure 32. Visualisation of virion RNA of WT and mutant samples with Northern Blotting. | 100 |
| Figure 33. Visualisation of virion RNA of WT and mutant samples with Northern Blotting..... | 101 |
| Figure 34. Visualisation of virion RNA of WT and U5-AUG mutants. | 102 |
| Figure 35. Visualisation of virion RNA of WT and S1 with Northern Blotting..... | 103 |
| Figure 36. Production of in vitro transcribed 690 nt RNA fragments of interest..... | 105 |
| Figure 37. In vitro dimerisation studies of WT and U5s RNAs visualised with the use of different types of gels..... | 107 |

| | |
|--|-----|
| Figure 38. Optimisation of the refolding buffer conditions to enhance in vitro RNA dimerisation efficiency..... | 108 |
| Figure 40. Side by side comparison of WT RNA of two different lengths in a 4% polyacrylamide TBM gel used for in-gel SHAPE experiments..... | 109 |
| Figure 39. In vitro dimerisation of U5s RNA incubated in the optimised Db visualised in a 4% polyacrylamide TBM gel..... | 109 |
| Figure 41. In vitro dimerisation assay assessing the dimerisation efficiency of WT and S1 transcribed RNA fragments produced with the optimised protocol developed in 5.3.b. | 110 |
| Figure 42. Side by side comparison of in vitro RNA dimerisation efficiencies of WT to a range of U5-AUG and dimerisation mutants in 4% polyacrylamide non denaturing TBM gels..... | 111 |
| Figure 43. qPCR measuring relative packaging efficiency of mutant vectors against WT..... | 114 |
| Figure 44. Relative packaging efficiency comparison between WT and U5s..... | 115 |
| Figure 45. Relative packaging efficiency of WT, NC4s and UC4d with the use of a WPRE-targeting RT-qPCR assay. | 116 |
| Figure 46. Design of a multiplex qPCR assay consisting of a common primer set and two TaqMan probes targeting the unique sequences inserted within the transfer vector. | 118 |
| Figure 47. Cloning of transfer vectors containing the unique BglG and MS2 sequences. | 120 |
| Figure 48. Primer optimisation process for the competitive MS2/BSL RT-qPCR assay. | 122 |
| Figure 49. Optimisation of BglG and MS2 probes' concentration..... | 123 |
| Figure 50. Investigation of the efficiency and linearity of the duplex BglG-MS2 qPCR assay..... | 124 |
| Figure 51. Investigation of new BglG assay linearity and efficiency. | 125 |
| Figure 52. Effect of primer concentration on the multiplex assay. | 126 |
| Figure 53. Effect of primer concentration on the multiplex assay. | 127 |
| Figure 54. Effect of constant and variable plasmid concentrations on MS2 and BglG assays. | 128 |
| Figure 55. Comparison of monoplex and duplex reactions for MS2 and BglG assays..... | 129 |
| Figure 56. Representation of qPCR workflow to measure relative packaging efficiency..... | 131 |
| Figure 57. Measurement of WTMS2/WTBglG RPE. | 132 |
| Figure 58. Measurement of WTMS2/ Δ P1BglG RPE..... | 133 |
| Figure 59. Effect of MS2 and BglG sequence introduction on viral vector transduction efficiency. | 133 |
| Figure 60. Measurement of RPE of U5-AUG and dimerisation mutants of interest..... | 134 |
| Figure 61. Assessment of WTMS2/WTBglG RPE with the use of a modified protocol. | 135 |
| Figure 62. Measurement of WTMS2/U5sBglG and U5sMS2/WTBglG RPEs..... | 136 |
| Figure 63. Measurement of WTMS2/WTBglG RPE with the use of an optimised protocol. | 137 |
| Figure 64. Measurement of WTMS2/U5sBglG and U5sMS2/WTBglG RPEs with the use of an optimised protocol..... | 138 |
| Figure 65. Relative packaging efficiency measurement of U5s mutant. | 140 |
| Figure 66. Relative packaging efficiency comparison between U5s and Δ P1 mutants. | 141 |
| Figure 67. Relative packaging efficiency measurement of J9+Gag mutant. | 142 |
| Figure 68. Relative packaging efficiency measurement of S1 mutant. | 143 |
| Figure 69. Relative packaging efficiency measurement of U5s, Δ P1, J9+Gag and S1 mutants. | 145 |
| Figure 70. Comparison of transfer vector and HIV-1 mRNAs. | 146 |
| Figure 71. Example of in-gel SHAPE data analysis with the SHAPEfinder software. | 148 |
| Figure 72. Structure Illustration of a 356 nt long monomeric RNA fragment corresponding to the HIV-1 leader..... | 151 |
| Figure 73. Structure Illustration of a 690 nt long monomeric RNA fragment including sequences of the HIV-1 leader and partial <i>gag</i> of the 3rd generation HIV-1 derived lentiviral transfer vector. . | 153 |
| Figure 74. Structure Illustration of a dimeric WT HIV-1 leader RNA fragment..... | 155 |

| | |
|--|------------|
| Figure 75. Structure Illustration of a 690 nt long dimeric RNA fragment including sequences of the HIV-1 leader and partial gag of the 3rd generation HIV-1 derived lentiviral transfer vector. | 156 |
| Figure 76. Visual representation of gag region sequence variability. | 159 |
| Figure 77. Visualisation of modelled HIV-1 leader-gag interactions with the RNAstructure software. | 161 |
| Figure 78. Structure Illustration of a 690 nt long dimeric RNA fragment including sequences of the J9+Gag HIV-1 leader and partial gag of the 3rd generation HIV-1 derived lentiviral transfer vector with XRNA. | 163 |
| Figure 79. Structure Illustration of a 690 nt long monomeric RNA fragment including sequences of the introduced S1 mutations in the HIV-1 leader and partial gag of the 3rd generation HIV-1 derived lentiviral transfer vector with XRNA. | 165 |
| Figure 80. Structure Illustration of a 690 nt long dimeric RNA fragment including sequences of the introduced S1 mutations in the HIV-1 leader and partial gag of the 3rd generation HIV-1 derived lentiviral transfer vector with XRNA. | 167 |
| Figure 81. In vivo SELEX approach for the creation of HIV-1 derived lentiviral vector libraries to select the best replicating variants. Adapted from source²⁷⁶ | 179 |

List of tables

| | |
|---|-----|
| Table 1. Variations of SHAPE technique for structural analysis of RNA | 26 |
| Table 2. Overview of methods to measure RNA packaging¹²⁹ | 38 |
| Table 3. Ongoing clinical trials using lentiviral vectors.Source:¹⁹⁴ | 49 |
| Table 4. List of primers used for SDM | 59 |
| Table 5. Calculation of plasmid copy number for transfections | 74 |
| Table 6. Calculation of plasmid DNA mass for transfections | 75 |
| Table 7. Sequence conservation analysis of the regions of polyA (72-95nt) and gag (1340-1360 nt). | 157 |
| Table 8. Sequence conservation of the regions of U5 (103-122 nt) and gag (1553-1172 nt). | 159 |
| Table 9. Sequence conservation of the regions of PBS (142-147 nt) and gag (1127-1132 nt) | 160 |
| Table 10. List of rationally designed transfer vector mutants | 172 |

Abbreviations

| | |
|------------------|--|
| AAV..... | Adeno-associated virus serotype |
| ADA..... | Adenosine deaminase |
| AIDS..... | Acquired immunodeficiency syndrome |
| ALD..... | Adrenoleukodystrophy |
| ALL..... | Acute lymphoblastic leukemia |
| BMH..... | Branched multiple hairpin |
| CA..... | Capsid |
| CAR..... | Chimeric Antigen Receptor |
| cHS4..... | Chicken HS4 Insulator Element |
| CLIP..... | Crosslinking-immunoprecipitation |
| CP..... | Copy number |
| cPPT..... | central polypurine tract |
| cRNA..... | cellular RNA |
| cryo-EM..... | cryo-electron microscopy |
| CTE..... | Constitutive transport element |
| dATP..... | deoxyadenosine triphosphate |
| Db..... | Dimerisation buffer |
| dCTP..... | deoxycytidine triphosphate |
| dGTP..... | deoxyguanosine triphosphate |
| DIS..... | Dimerisation Initiation Site |
| DLS..... | Dimer linkage sequences |
| dPCR..... | digital droplet PCR |
| DSL..... | Dimerisation Stem Loop |
| DTT..... | Dithiothreitol |
| dUTP..... | deoxyuridine triphosphate |
| EIAV..... | Equine infectious anaemia virus |
| ELISA..... | Enzyme-linked immunosorbent assay |
| EMA..... | European Medicines Agency |
| <i>env</i> | envelope gene |
| GNRs..... | Glass Nano Reactors |
| GORS..... | Global genome-wide RNA structure |
| gRNA..... | genomic RNA |
| HBS..... | Hepes Buffered Solution |
| HIV..... | Human immunodeficiency virus |
| HLA..... | Human leukocyte antigen |
| hnRNPs..... | heterogeneous nuclear ribonucleoproteins |
| HR..... | High Range |
| HSPCs..... | Hematopoietic Stem and Progenitor cells |
| IN..... | Integrase |
| INS..... | Instability sequences |
| LDI..... | Long-distance interaction |
| LDLR..... | low-density lipoprotein receptor family |
| lncRNAs..... | long noncoding RNAs |
| LPL..... | Lipoprotein lipase |

| | |
|------------------|--|
| LR..... | Low Range |
| LTR..... | Long Terminal Repeat |
| LVs..... | Lentiviral vectors |
| MA..... | Matrix |
| MCS..... | Multiple cloning site |
| MGB..... | Minor groove binder |
| MIQE | Minimum Information for Publication of Quantitative Real-Time PCR Experiments guidelines |
| MLV..... | Murine leukemia virus |
| MoMLV..... | Moloney murine leukaemia virus |
| mRNA..... | messenger RNA |
| MTOC..... | Microtubule organization center |
| NC..... | Negative control or Nucleocapsid (context dependent) |
| NEAT1..... | Nuclear paraspeckle assembly transcript 1 |
| NFQ..... | Nonfluorescent quencher |
| NICE..... | National Institute for Health and Care Excellence |
| NMR..... | Nuclear magnetic resonance spectroscopy |
| ORF..... | Open Reading Frame |
| OTC..... | <u>O</u> rnithine transcarbamylase |
| PBS..... | Primer binding site or Phosphate-Buffered Saline (context dependent) |
| PIC..... | Preintegration complex |
| PR..... | Protease |
| <i>psi</i> | Packaging signal |
| Rb..... | Refolding buffer |
| RCL..... | Replication competent lentivirus |
| RF..... | Restriction factor |
| RPA..... | Ribonuclease protection assay |
| RPE..... | Relative packaging efficiency |
| RRE..... | Rev Response Element |
| RT..... | Reverse Transcriptase |
| RTC..... | Reverse transcription complex |
| SCID..... | Severe combined immunodeficiency |
| SD..... | Splice donor |
| SDM..... | Site directed mutagenesis |
| SHAPE..... | Selective 2'-Hydroxyl Acylation Analysed by Primer Extension |
| SELEX..... | Systematic evolution of ligands by exponential enrichment |
| SIN..... | Self inactivating |
| SL1-4..... | Stem Loop 1-4 |
| SU..... | glycoprotein gp120 |
| TBE..... | Tris-borate-EDTA |
| TBS..... | Tris-Buffered Saline |
| TM..... | Transmembrane glycoprotein gp41 |
| TRiP..... | Transgene Repression In vector Production system |
| TSS..... | Transcription start site |
| TU..... | transduction units |
| UCOEs..... | Ubiquitously acting chromatin opening elements |

VLPs.....Virus like particles
VSVg.....Vesicular Stomatitis Virus glycoprotein
WAS.....Wiskott-Aldrich syndrome
WPRE.....Woodchuck hepatitis virus post transcriptional regulatory element

external gp120 glycoprotein (SU) and the transmembrane glycoprotein gp41 (TM)^{3 4}. The accessory proteins Vpu, Vpr, Vif and Nef are important for the virulence of the pathogen, whereas the regulatory proteins Tat and Rev are responsible for transcription initiation and export of full length and singly spliced RNAs to the cytoplasm respectively. These various products are produced from their respective RNAs spliced from the gRNA⁵.

2.1.2. Overview of the gRNA journey

Upon transcription of the integrated proviral DNA, the full length 9.7 kb unspliced transcript, which contains the Rev Response Element (RRE), is transported from the nucleus to the cytoplasm in a Rev-dependent manner⁶. The gRNA is an active player in the production of an infectious virion. The transport from the transcriptional site to the virion involves the adoption of several RNA structures that display the relevant *cis*-acting elements for each process the virus needs to perform^{7 8}. It has a dual role: as a template for translation of *gag* and *gag/pol*, and as the gRNA for the newly assembled virions. A conformational switch in the *cis*-acting elements of the 5'UTRs of the gRNA controls the balance between the monomeric RNA, characterised by a structural conformation that contains an exposed *gag* start codon (AUG) for translation, and the dimerisation competent RNA, which is preferentially packaged, from the same template⁹. The presence of *cis*-acting elements in the 5'UTR of the full-length transcript promotes the dimerisation and packaging of the gRNA dimer in the newly formed virions, and the specific binding of the NC domain of the Gag polyprotein on the gRNA promotes its trafficking from the cytoplasm to the membrane site of budding¹⁰. Furthermore, the viral RNA molecules act as a scaffold for their own capture by the NC in the context of Gag, and in a later stage, act as a nucleation site for the assembly of newly formed viral particles^{11 12}. The final, late stage contribution of the gRNA for virion production is at the budding site, where the interaction between gRNA, Gag and the ESCRT machinery are necessary for the completion of the budding process. Although, the exact mechanism is unclear, mutations in the 5'UTR *cis*-acting elements of the gRNA affect maturation, highlighting the importance of the interaction between gRNA and Gag at this stage^{13 14}. In the next sections I will be looking at these different steps of the gRNA journey in further detail.

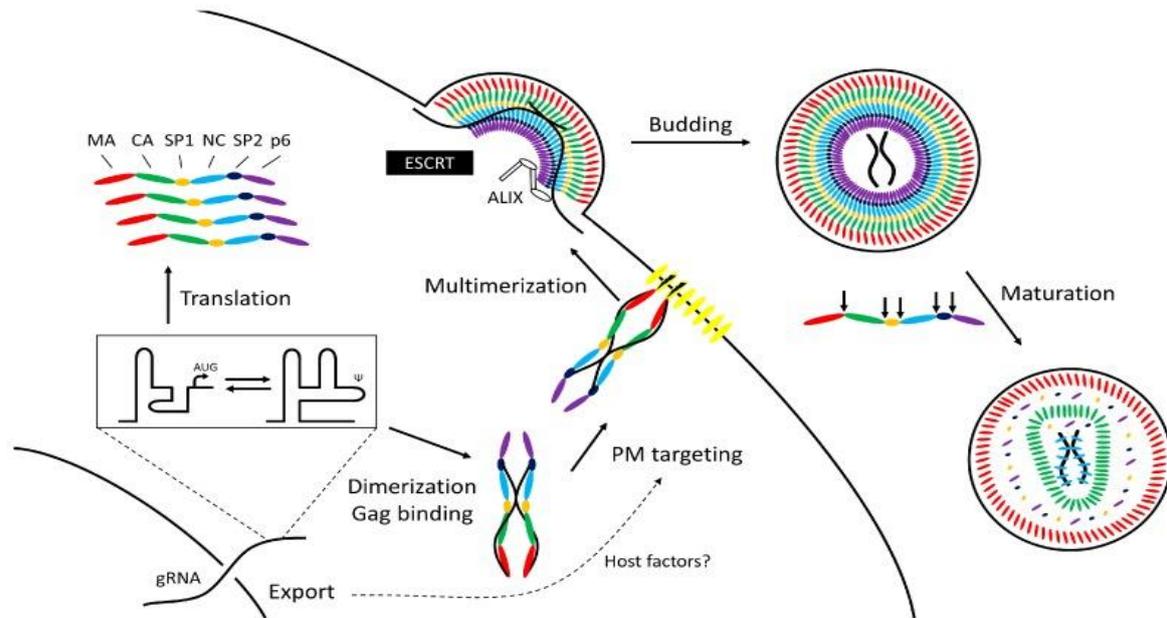


Figure 2. Overview of the assembly pathway of an HIV-1 virion²⁸⁶.

2.1.3. Gag/gRNA interaction and packaging

As mentioned above, the Gag/gRNA interaction plays an important role in coordinating and facilitating the transport of the gRNA molecule from the export site to the budding site. However, where and when the interaction between these two determinants occurs in the cell remains controversial. Live cell co-imaging showed accumulation of HIV-1 gRNA at the nuclear envelope but detection of Gag protein in the cytoplasm and not at the nuclear rim. This observation suggests initiation of the Gag/gRNA interaction in the cytoplasm¹⁵. A different study reported the pericentriolar microtubule organization center (MTOC) as the primary site of HIV-1 Gag/gRNA interaction¹⁶, while an imaging study detected higher order Gag-RNA complexes at the plasma membrane and smaller oligomeric complexes in the cytoplasm¹⁷. Use of total internal reflection fluorescence microscopy demonstrated that stabilisation of the RNA dimer on the plasma membrane requires Gag protein¹⁰. Crosslinking-immunoprecipitation (CLIP) sequencing experiments suggest a two-step genome packaging process that starts with interactions between monomeric Gag and the ψ element, followed by interactions between multimeric Gag and the A-rich viral mRNA¹⁸. Even though it still remains unclear where exactly the initial HIV-1 Gag/gRNA takes place,

the aforementioned studies suggest formation of ribonucleoprotein particles prior to plasma membrane localisation.

The presence of sequences important for efficient gRNA packaging were first described when deletion of 19 bp between the SD in the 5'LTR and the *gag* ORF resulted in reduced HIV-1 replication and gRNA encapsidation in human T-lymphocytes¹⁹. Although the packaging defect was profound, it was not complete, suggesting some redundancy in the system. gRNAs get successfully packaged in a specific way despite the fact that they only represent approximately 1% of the cellular 5'-capped, polyadenylated mRNAs²⁰. The specificity of this interaction is mainly driven by the presence of two highly conserved zinc finger motifs that lie in the NC domain of Gag²¹. The HIV-1 NC is highly basic and contains two zinc fingers, with the zinc-chelating cysteine and histidine residues arranged as C-X2-C-X4-H-X4-C. These fingers and basic residues are very important for the interaction between NC and gRNA sequences. This electrostatic interaction is considered almost independent of the nucleotide sequence of the gRNA and mainly dependent of the zinc finger nucleic acid-binding specificity^{22 23}. The NC domain of Gag selects the gRNA during assembly while the cleaved NC coats the dimeric RNA within the mature viral particle. The ability of NC to catalyse the rearrangement of nucleic acids into the most thermodynamically stable structures makes it an important DNA synthesis cofactor²⁴. More specifically, the NC domain, in the context of Gag, catalyses the unwinding of tRNA^{Lys,3} and its annealing to the PBS²⁵.

On the gRNA level the selection is governed by the ψ (*psi*) packaging element that lies in the 5'UTR and is located downstream of the splice donor, such that the full length transcripts can preferentially get encapsidated^{26 27}. Certain mutations in the NC domain gave a similar defective phenotype to deletions of the RNA packaging signal, highlighting the mutual importance of these two key players in RNA packaging²⁸. Because of the high affinity of Gag to the ψ element on gRNA, it has been hypothesised that ψ -containing RNAs have a thermodynamic advantage over the cellular RNAs since their binding to Gag and subsequent encapsidation is energetically more favourable²⁹.

2.1.4. Importance of RNA structure

The central dogma of molecular biology states that "the coded genetic information hard-wired into DNA is transcribed into individual transportable cassettes, composed of messenger RNA (mRNA); each mRNA cassette contains the program for synthesis of a particular protein (or small number of proteins)³⁰."

To understand the processes that govern the present-day living cells deeper, one needs to consider their evolution. Today, we understand the requirement of complex machineries that convert genetic information, stored in DNA, to function, performed by proteins through an RNA intermediate. However, it is still unclear how this machinery arose. One of the popular hypotheses, originally suggested by Alex Rich in 1962 and known as the *RNA world*, suggests that RNA was once able to both store genetic information (as it still does for some viruses, including HIV-1) and catalyze chemical reactions³¹. RNA still catalyzes several fundamental reactions in modern-day cells and plays a central role in gene regulation³². RNA performs key biological functions via folding into diverse structures directed by its nucleotide sequence.

RNA acts both as a coding and noncoding molecule and as shown in Figure.3 can adopt complicated secondary structures that are in part responsible for its diverse functionality. Comparisons of many RNA structures have revealed certain frequently appearing conserved motifs shown in Figure 3³³. RNA structure plays important regulatory roles not just for specialized classes of RNAs such as riboswitches, small RNAs, and enzymatic RNAs ribozymes but essentially for all classes of RNA, including mRNAs and long noncoding RNAs (lncRNAs). Under the light of this recent realisation, RNA structure can be now seen as another layer of the genetic code whose understanding may enable us to study the evolution and function of biological systems³⁴. Advancement of high-throughput sequencing methods has improved our understanding of transcriptome modifications and their ability to impact RNA secondary structure to alter processes like splicing, translation and RNA stability³⁵. Notably, even the addition of small groups can affect the ability of bases to pair, to stack against neighbouring bases, to adopt one conformation over another, to favour one folded structure over another and to interact with proteins³⁵.

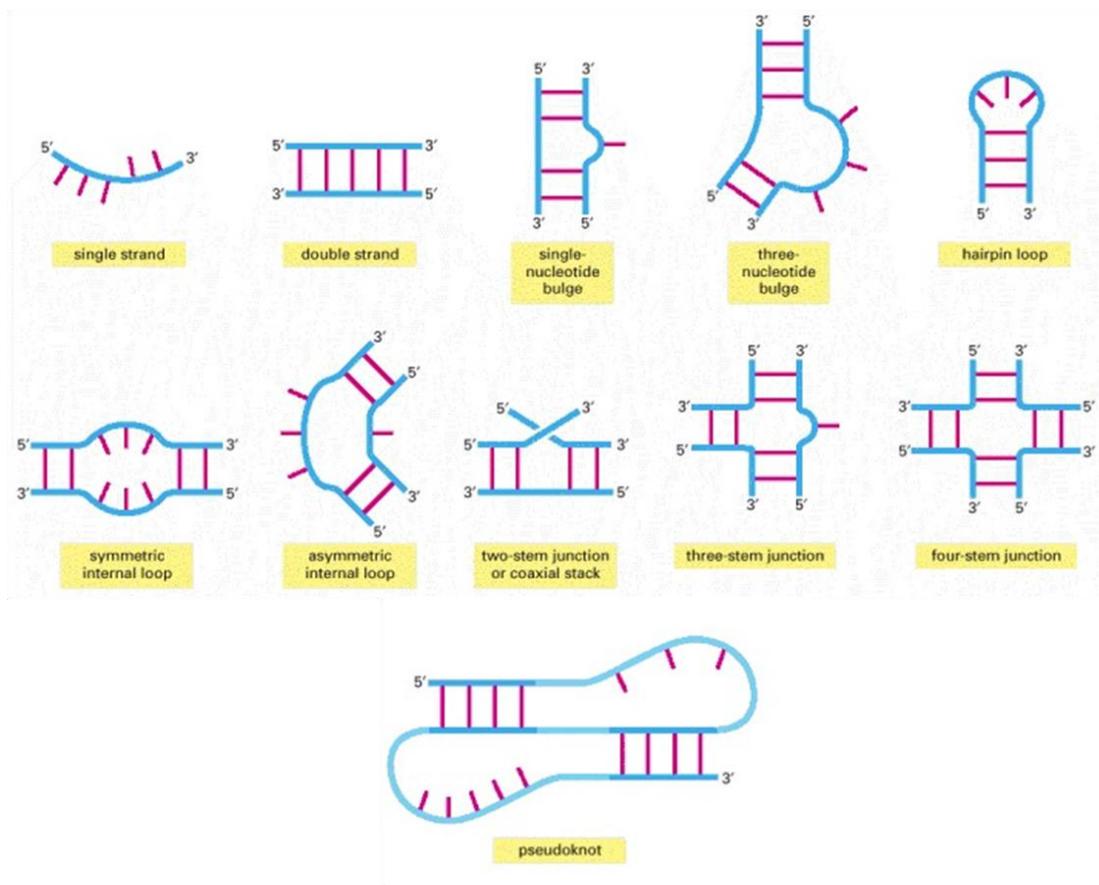


Figure 3. Common structural motifs of RNA³⁴.

From an evolutionary perspective, the conservation and diversity of the RNA structurome, defined as transcriptome wide structure profile, between species remains to be understood. Deng *et al.*, generated rice (*Oryza sativa*) *in vivo* structure-seq libraries to compare the conservation and divergence of *in vivo* RNA structurome between rice and previously published structures of *Arabidopsis*³⁶, and assess the evolutionary adaptation of RNA structure. This study found that *in vivo* RNA secondary structure conservation does not correlate with sequence conservation between rice and *Arabidopsis*, but that conservation and divergence in both sequence and RNA secondary structure are highly relevant with specific biological processes³⁷. This suggests that evolutionary pressure does not only affect sequence but also RNA structure to regulate gene expression³⁷.

2.1.5. Techniques used to study RNA structure

Precise folding into a complicated structure is a prerequisite for correct RNA function. Moreover, many RNAs require specific local structural flexibility to transition from one

conformation to another³⁸. Involvement of the RNA secondary structure is fundamental for cell processes like RNA synthesis, metabolism and pathway regulation, has caused an explosion in the need and interest to develop robust technologies to investigate RNA structure. Previous efforts employed techniques including nuclear magnetic resonance spectroscopy (NMR), X-ray crystallography, and cryo-electron microscopy (cryo-EM) that are restricted by small length, low abundance and poor yield limitations³⁹.

| | | |
|--------------|--|----------------------------|
| SHAPE-CE | Determine the RNA secondary structure of high abundant RNAs up to 400 nt length ⁴⁰ | <i>In vitro</i> |
| SHAPE-LMPCR | Determine the RNA secondary structure of low abundant RNAs ⁴¹ | <i>In vivo</i> |
| SHAPE-Seq | Determine the RNA secondary structure of long RNAs ⁴² | <i>In vitro</i> |
| SHAPE-MaP | Determine the RNA secondary structure of low abundant ⁴³ and long RNAs ⁴⁴ | <i>In vitro or in vivo</i> |
| icSHAPE | Determine genome-wide <i>in vivo</i> RNA secondary structure with chemical probing ⁴⁵ | <i>In vivo</i> |
| In-Gel SHAPE | Determine the RNA secondary structure of individual conformers from a mixed RNA population ⁸ | <i>In vitro</i> |
| XL-SHAPE | Determine both RNA structural changes and the sites of protein interaction within an RNA-protein complex ⁴⁶ | <i>In vitro or in vivo</i> |

Table 1. Variations of SHAPE technique for structural analysis of RNA

Traditional chemical and enzymatic reagents⁴⁷ and backbone-based cleavage⁴⁸ can be useful for mapping RNA secondary structure, but suffer from two important disadvantages including 1) use of reagents that cannot discriminate between single-stranded and base-paired regions accurately enough and 2) use of reagents that typically react with only a subset of the four RNA nucleotides or backbone sites, requiring multiple reagents that in turn make data analysis difficult. Selective 2'-Hydroxyl Acylation analysed by Primer Extension, or SHAPE, takes advantage of reagents that preferentially bind and modify the backbone of RNA in structurally flexible regions of unpaired nucleotides⁴⁹. SHAPE chemistry exploits hydroxyl-selective electrophiles used for 2'-hydroxyl acylation leading to the formation of 2'-O-adducts. Since every nucleotide has a 2'-OH group in their ribose backbone the use of this chemistry should provide structural information for every nucleotide of the RNA sequence. SHAPE can give single nucleotide level structural information for all nucleotides and hence determine the structure of any RNA under diverse, relevant states^{50 51 52 53}. Paired nucleotides remain unreactive whereas conformationally flexible nucleotides form 2'-O-adducts. These specific sites of modification can be determined by annealing a labelled primer to the RNA sequence of interest and then by extending to the nearest site of modification with a reverse transcription reaction. The position and degree of modification can be determined by reading the corresponding length and amount of cDNA labelled fragments. A negative control lacking the modifying reagent reaction is run in parallel. Mapping the measured reactivity to the RNA sequence is achieved by including two dideoxy sequencing ladders. In hSHAPE, each component of a SHAPE experiment uses the same primer sequence that is labelled with a different color-coded fluorophore. The resulting cDNAs are combined and resolved in a single capillary on a capillary electrophoresis sequencing instrument⁵⁴. SHAPEFinder software is used to measure single nucleotide reactivity by calculating the difference of reactivity between the modifying reagent reaction and the modifying reagent lacking reaction, while aligning each reactivity peak to a single nucleotide⁵⁵.

2.1.6. Structural packaging signals in the HIV-1 leader

The 5'-leader is the most conserved region of the HIV genome⁵⁶; it contains the *cis*-acting elements that regulate a variety of processes including transcriptional activation, reverse transcription, splicing, dimerisation, packaging and translation of HIV-1^{2 57}. Various groups have published on the structure of the packaging signal and on the HIV-1 leader as a whole, using different technical approaches including nuclear magnetic resonance and probing approaches (reviewed in^{8 57 58 59 60 61 62}). The intrinsic flexibility and instability of the RNA molecule, combined with technical limitations, can explain the differences in the plethora of these published structures. Indeed, it has been clarified that the HIV-1 leader is not a static structure, but can adopt several stable conformations⁵⁹. However, there is common agreement on the basic important features in the 5'-leader region, illustrated in Figure 4.

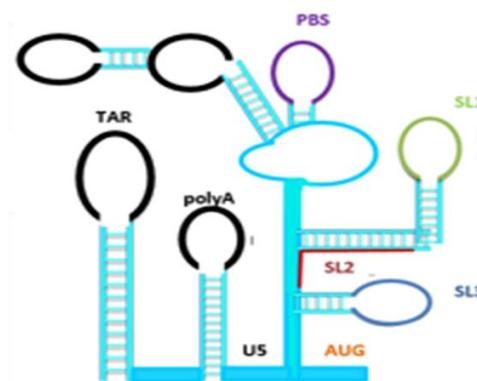


Figure 4. Secondary structural conformation of the one half of the dimeric HIV-1 leader (hemidimer).

The region shown in Figure 4 consists of the TAR loop, which is the binding site for the Tat protein leading to transcriptional activation⁶³, the polyA loop, which is suppressed in the 5'UTR⁶⁴, the primer binding site (PBS) loop bound by the tRNA^{Lys,3} primer for reverse transcription⁶⁵, followed by 4 stem-loops that are of significant importance to packaging. Stem Loop 1 (SL1) includes a conserved 6-base palindrome Dimerisation Initiation Site (DIS), which takes part in a loop-loop interaction with a complementary palindromic sequence from a second gRNA molecule to mediate dimerisation of the viral genomic RNA. SL2 contains the major splice donor and SL3 contains the most important packaging

determinants, discussed below. Finally, SL4, which is formed at the beginning of the *gag* ORF, is associated with the translational conformation of the RNA^{66 67 8}. More recently, evidence suggests the presence and importance of another structural element in the 5'-leader known as the U5-AUG duplex. The interaction between the 5' U5 and the residues around the beginning of the *gag* ORF promotes the dimerisation structure⁶¹ and may influence the process of translation by blocking the accessibility of the Gag start codon⁷. As alluded to above, packaging of the gRNA depends on structural motifs in the RNA rather than sequence recognition⁶⁸.

2.1.7. Structure of the HIV-1 RNA genome

The WT HIV-1 RNA sequence has evolved to be able to present critical replication signals in the context of a complex leader structure while the small intervening single-stranded segments passively contribute to proper RNA folding. Nearly any mutation in this RNA segment affects the leader RNA structure highlighting the importance of the selected WT HIV-1 sequence⁶⁹. The HIV-1 leader is proposed to adopt two conformations that differ in their ability to dimerise. Originally, Abbink *et al.*, have predicted a long-distance interaction (LDI) conformation in which the DIS nucleotides are paired to the polyA sequences restricting dimerisation. The branched multiple hairpin (BMH) conformation induced with NC binding is characterised by exposure of its DIS hairpin and its subsequent ability to form dimers⁵⁸. More recently, Lu *et al.*, proposed an alternative LDI conformation in which the DIS is occluded due to its intramolecular interaction with the lower U5 sequences. NC binding induces a shift to the BMH conformation that in this model finds the DIS free to interact with the DIS of another RNA molecule while the lower U5 sequences pair intramolecularly with the *gag* AUG⁶¹.

The aforementioned findings on the RNA structural switch are derived from *in vitro* analysis of the HIV-1 leader RNA. However, inside virions HIV-1 gRNA exists as a 5' capped, 3' polyadenylated, non-covalent dimer that is annealed to a host tRNA^{Lys,3} molecule⁷⁰. The architecture and the secondary structure of the entire HIV-1 gRNA extracted from virions were proposed based on the SHAPE technique and a thermodynamic structure prediction algorithm⁷¹. Their findings suggest that regions with low SHAPE reactivities and hence involved in base pairing interactions, are associated with regulatory functions. Watts *et al.*,

identified that peptide loops that link independently folded protein domains are encoded by highly structured RNA indicating that RNA structure encodes protein structure at a second level beyond the amino acid sequence⁷¹.

Since then, this model has been updated; initially by employing the improved SHAPE-MaP technology⁷². However, even the updated structural version suffers from lack of long-distance interactions and non-canonical base pairs. More recently, Sükösd *et al.*, presented the first global RNA secondary structure model for the HIV-1 genome without such distance constraints⁷³. Their model is the outcome of both a phylogenetic and a SHAPE data-driven approach. Their study showed that most of the HIV-1 genome appears to be relatively unstructured⁷³. Such results support the hypothesis, brought forward by Knoepfel and Berkhout, that HIV-1 does not belong to the group of viruses with a global genome-wide RNA structure (GORS)⁷⁴, the presence of which has been shown to correlate with persistent viral infection, potentially by facilitating escape from innate responses⁷⁵. No major virus replication defects were detected when RNA structure destabilising mutations were introduced, suggesting that the hairpin structures studied were not important for HIV-1 replication⁷⁴.

Lastly, comparison of HIV-1 and SIV genomic RNA structures showed that the RNA secondary structure is evolving at a much faster rate than its sequence. Structured regions seem to be enriched in guanosines, while unpaired regions have higher adenosine content, and functionally important structures have stronger base pairing than non conserved structures. Secondary structures of functional elements are stabilised by higher guanine content. That allows regions of structure to persist throughout evolutionary pressure while still allowing structures to evolve⁷⁶.

2.1.8. HIV-1 dimerisation

Retroviruses package two copies of gRNA into each of their virions. The packaging of their genetic information in a dimeric form allows for recombination events to occur during reverse transcription, increasing the diversity of the viral sequences⁷⁷. Also, packaging of a dimeric RNA can allow for the provision of functions of the viral RNA where one copy is damaged⁷⁸. Thus, this strategy positively affects the replication and evolution of retroviruses, by promoting the generation of viruses that can evade immune recognition

and also build resistance to antiretroviral treatments. The dimeric nature of retroviral RNA was first suggested when measurement of buoyant density was double to what was expected from a single gRNA copy⁷⁹, while gRNA heat treatment led to the expected buoyant density of monomeric RNA⁸⁰. RNA dimerisation has been studied by employing both *in vivo* and *in vitro* approaches. The *in vivo* approach involves RNA isolation from produced virions in tissue culture subsequently analysed by native Northern Blotting. Retroviral RNA analysed under non-denaturing conditions was detected as a single band whose migration pattern changes upon heat denaturation, presumably due to the shift of the dimeric RNA to its monomeric state upon denaturation⁸¹. The other method involves synthesis of short segments of viral RNA *in vitro*, and then studying the ability of these fragments to form dimers. The exact location where dimerisation takes place in the cell has been a mystery for a long time, but it was recently shown that dimerisation occurs in the plasma membrane before packaging¹⁰ and requires the presence of Gag¹⁰ and intact complementary sequences in the DIS region⁸². Both the DIS and the U5-AUG structures are phylogenetically conserved among retroviruses suggesting the importance of their functions⁸³.

SL1, a highly conserved hairpin from nt 236 to nt 282 (based on HXB2 strain) appears to be the most important motif for dimerisation of HIV-1 gRNA⁸⁴. Both its deletion⁸⁵ and antisense nucleotide targeting⁸⁶ prevent RNA dimerisation *in vitro*, making SL1 widely accepted as the dimerisation initiation site (DIS). It contains two internal bulges and an atypical loop of nine nucleotides, six of which form a palindrome. The palindromic nature of DIS allows it to form canonical Watson-Crick base pairs with the counterpart palindrome of another gRNA molecule; an interaction known as a kissing loop⁸⁷. Purine residues flanking this palindromic sequence were also shown to be important for this initial interaction⁸⁸. Not all palindromic sequences are equally efficient, with most known HIV-1 strains containing a DIS of either GCGCGC or GUGCAC sequences⁸⁹. The length of the palindromic sequence also seems to be specific with any numerical alterations quickly reverting to 6 nucleotides⁹⁰. Even with the 6 nucleotide restriction, the DIS could be of 64 (6^4) potential sequences, however competition experiments showed that only a few other palindromes could successfully substitute the two WT DIS sequences mentioned above, with most mutant viruses characterised by decreased infectivity⁹¹. It is unclear why the nature of DIS is so restrictive, as it has been

demonstrated to be dispensable for efficient viral replication despite its central role in RNA dimerisation⁹². Interestingly, when DIS viral mutants were placed in long term culture the compensatory mutations found were located within the *gag* sequences enhancing the cleaving ability of Gag^{93 94}. This observation suggests that DIS is not required for replication but for enhancement of gRNA dimerisation that could facilitate its recognition by Gag.

2.1.9. HIV-1 RNA packaging

Identification of a minimal HIV-1 packaging signal has been proven to be a challenging task for a number of reasons. First, the 5'-leader contains elements that might be unrelated to packaging but are essential for viral replication. Hence, mutagenesis studies designed to test effects on packaging can adversely affect other functions⁹⁵. Another complication relates to the fact that when packaging studies are conducted using transfection experiments, cellular concentration of mutant vector RNAs can be too high and thus mask potential packaging defects. Additionally, HIV-1 RNA packaging is dependent on correct Rev-mediated RNA localisation, so that even RNAs that contain intact packaging signals might be inefficiently packaged if the Rev-RRE pathway is disturbed⁹⁶. Lastly, as alluded to in section 2.1.6, efforts to determine the secondary structure of the HIV-1 5'-leader have given rise to multiple structural proposals and mechanistic predictions complicating the study of packaging structure-function relationship.

A specific region in the MLV genome has been identified as sufficient for RNA packaging because insertion of this sequence in a heterologous RNA leads to its packaging into viral particles. At this time, the necessary and sufficient HIV-1 packaging signal had not been defined^{97 98}. The first studies aiming at identifying sequences comprising the HIV-1 RNA packaging signal found that deletion of RNA sequences between the major splice donor (SD) and the *gag* coding region, defined as the ψ element, decreased the levels of genomic RNA packaged into virions^{19 99}. It was hypothesised that this region could be responsible for selective packaging of gRNA since these sequences are located downstream of the major 5' SD and can be found only in the full length gRNA. Subsequent studies revealed that RNA sequences upstream of the 5' SD site also affect RNA packaging¹⁰⁰. Meanwhile, several groups reported on the specific NC binding affinity and its requirement for retroviral RNA encapsidation^{99 101 102}.

Genome deletions were performed to trim down the sequences required for packaging of the HIV-1 gRNA. A minimal 159 nt RNA sequence that includes SL1–SL3 through the U5-AUG stem, but lacks TAR, PolyA and the upper PBS hairpin structure was shown to dimerise and bind NC *in vitro*¹⁰³. In addition, a 144 nt RNA sequence that contained the PBS hairpin, SL1, SL3 and SL4 proved to be sufficient to mediate intramolecular dimerisation when inserted into an ectopic position in the HIV-1 genome¹⁰⁴.

These findings combined with structural data described above, led to the establishment of a model for the HIV-1 ψ site made of four independent stem-loops named SL1-SL4^{105 106}. All stem loops of the 5'UTR are acknowledged to contribute towards gRNA packaging and other essential parts of the viral life cycle. For example the central role of SL1 in the initiation of dimerisation was discussed in detail in section 2.1.8. The existence of such overlapping functions for this RNA structure raises the possibility that these two functions, dimerisation and packaging, might be linked. The existence of such a link in HIV-1 is discussed in the next section.

2.1.10. The relationship between RNA dimerisation and packaging

HIV-1 uses *cis*-acting elements of the gRNA and *trans*-acting elements in Gag to achieve specific packaging of the viral genome¹⁰⁷. In an attempt to investigate the relationship between dimerisation and packaging, early *in vivo* studies assessed the infectivity efficiency of HIV-1 proviruses containing *psi* mutations in SL1 and SL3 that affect *in vitro* Gag binding, RNA dimerisation, or both. The outcome of this mutational analysis showed that deletion of SL1 alone, or SL3 plus adjacent flanking sequences, decreased gRNA packaging, while deletion of SL1 and SL3 simultaneously caused an even further reduction. Complete deletion of SL1 led to increased packaging of monomeric RNA, confirming once more the significance of SL1 for *in vivo* HIV-1 RNA dimerisation. However, these mutant genomes still managed to get encapsidated, suggesting that dimerisation is not a prerequisite for packaging but that it rather serves an independent function in the retroviral life cycle¹⁰⁸. In an attempt to separate the dimerisation and packaging functions Sakuragi *et al.*, created mutant constructs that contained duplications of approximately 1000 bp, including sequences in the encapsidation signal, in the dimerisation initiation site (DIS) and in the dimer linkage sequences (DLS), (E/DLS region). Duplication of the E/DLS region led to the packaging of

monomeric RNA in virus particles, which were likely generated by intramolecular interactions between the two E/DLS regions on the same RNA molecule¹⁰⁹.

In a different study DIS deletions were combined with various combinations of previously identified compensatory point mutations, including the Δ Loop (lacking the loop region of SL1) and Δ DIS (lacking the complete SL1) mutants that were characterised by defects in replication, RNA dimerisation, and packaging. Surprisingly, all functions, apart from dimerisation, were restored by compensatory point mutations in *gag*¹¹⁰. A more recent report exploiting advanced technology of labelled constructs that contained two packaging signals confirmed the formation of intramolecular dimers that can get packaged¹². Therefore, the authors of this study concluded that RNA packaging is promoted and regulated by the formation and recognition of this dimeric RNA.

The region required for RNA packaging and RNA dimerisation overlaps in HIV-1 with several studies suggesting that the RNA fragments containing the highly conserved 5'UTR extended to *gag* can dimerise *in vitro*^{111 112}. The link between dimerisation and packaging is a subject of ongoing debate^{113 114 115 106 116 117}. gRNA can be packaged as a monomer, or as a weak dimer that appears as monomeric on gels, but dimerisation defective mutant viruses generally do not replicate as well as WT viruses. It is possible that in the absence of DIS other sequences that affect dimerisation may lead to formation of a weaker than WT dimer that, however, allows gRNA to be adequately packaged¹¹⁷. To conclude, packaging of monomeric gRNA is possible, but dimerisation is likely to be a prerequisite for the efficient production of infectious HIV-1 viral particles¹¹⁸.

2.1.11. Packaging vs Translation: Maintaining the Balance

While dimerisation positively regulates packaging, translation of the genome is in opposition to its encapsidation, as alluded to above. Once the full-length RNA is exported to the cytoplasm it has two potential fates, it could either become the viral genomic RNA, serve as a template for translation or it could do both sequentially. In contrast to MLV that has two separate pools of RNA, one for translation and the other one for packaging^{119 120}, HIV-1 gRNA has the potential to serve both purposes¹²¹ and can be packaged with or without being translated first¹²². Cotranslational packaging poses a potential competition issue for

RNA between Gag and the ribosome, implying the requirement for a mechanism that would regulate these two processes.

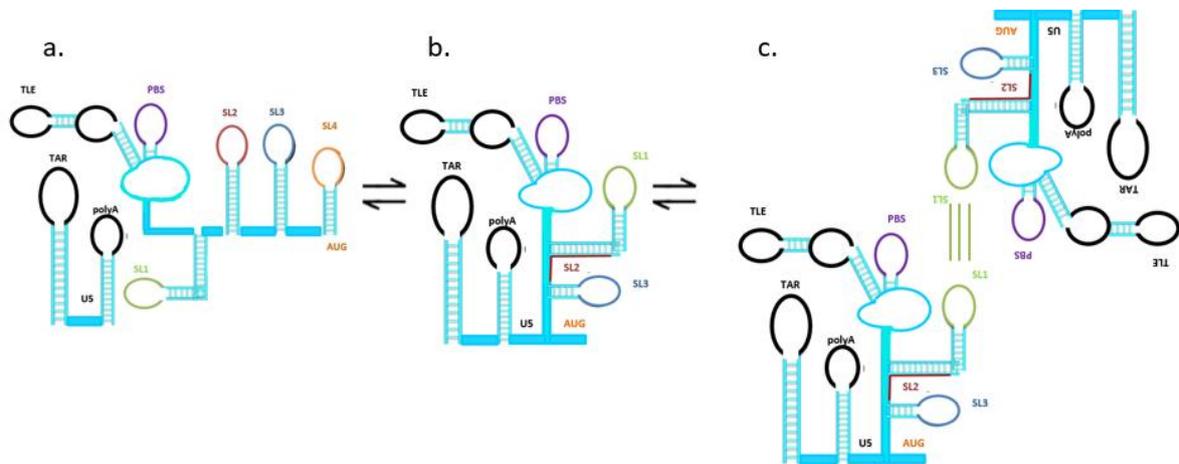


Figure 5. Conformational changes of the HIV-1 leader featuring a) a monomeric RNA structure, b) a dimeric RNA structure and c) a dimer interaction. Adapted from source:⁹

Evidence from NMR and RNA probing experiments suggest that the HIV-1 leader is able to adopt different conformations in order to regulate the processes of translation and packaging^{7 8 61}. As shown in the model in Figure 5, in the unspliced monomeric viral RNA, the DIS loop interacts with the U5 region in a way that leaves the AUG start codon of *gag* accessible to ribosomal translation. Once sufficient levels of Gag are produced, NC can bind to RNA, promoting it to adopt the BMH conformation and thus, promoting dimerisation and packaging^{59 58}. A conformational switch that leaves the DIS exposed while occluding the *gag* start codon, makes the same RNA template dimerisation competent⁹. This dimeric conformation is also characterized by the formation of another structural motif, the U5-AUG duplex. Mutations that strengthen the U5-AUG base pairing resulted in elevated levels of dimerisation^{61 7}. Also, gel electrophoresis revealed that the strengthened U5-AUG mutant migrates faster than WT, potentially reflecting a more compact structure⁷. These results suggest that the formation of the U5-AUG duplex promotes dimerisation, and thus packaging, while potentially inhibiting translation of the Gag protein due to occlusion of the *gag* start codon. More recently, a study suggests the involvement of full length HIV-1 RNA in

regulating translation and packaging through an epitranscriptomic switch. Demethylation of two conserved adenosine residues present within the 5'-UTR was required for the encapsidation of gRNA into viral particles, while their methylation was associated with increased Gag production and reduced gRNA encapsidation¹²³. Collectively, all data suggest that the gRNA plays an active role in its trafficking from the nucleus to the budding site and acts as a switch to control the processes of translation and packaging. Yet another possibility is that the RNAs are "marked" by host factors that sort the unspliced RNAs either for packaging or translation shortly after transcription. Identification of such marking factors was achieved by pulldown experiments on cell lysates using the HIV-1 5' UTR sequences as a bait and mass spectrometry to spot the interacting host proteins¹²⁴. Domain analysis of factors identified to interact with the HIV-1 RNA 5' UTRs included nucleotide binding factors, DNA/RNA helicases, double-stranded RNA binding proteins, heterogeneous nuclear ribonucleoproteins (hnRNPs), splicing factors, ribosomal proteins, scaffold proteins and proteins involved in the DNA damage response. In addition to host factors, cellular RNAs such as long non-coding RNAs (lncRNAs) may also play a role in determining the fate of viral RNAs¹²⁴. Nuclear paraspeckle assembly transcript 1 (NEAT1) lncRNA is one of several lncRNAs whose expression is changed by HIV-1 infection and was characterised to have a role in regulating HIV-1 replication by contributing to the integrity of the nuclear paraspeckle substructure¹²⁵. Such observations highlight the involvement of host proteins and noncoding RNAs in marking and sorting retroviral unspliced RNA for packaging or translation¹²⁶. Lastly, a more recent study challenges the dogma of a single HIV-1 RNA transcript serving both as a translation template and as the gRNA¹²⁷. This study revealed the existence of multiple HIV-1 RNA transcripts of slightly different lengths due to heterogeneous transcriptional start site usage in infected cells. RNAs that begin with a single capped guanosine seem to be preferentially selected for packaging, whereas those that begin with two or three capped guanosines are enriched on polysomes and used for translation. Such results suggest that transcript length drastically affects its structure, function and fate and can thus serve as a regulating mechanism that directs the function and structure of HIV-1 gRNA¹²⁷.

2.1.12. Established methods for RNA packaging measurement

The effect of RNA packaging on HIV-1 retroviral replication has led to the development of several methods to assess encapsidation of gRNA. Northern Blot analysis, ribonuclease protection assay (RPA) and quantitative reverse transcriptase-coupled polymerase chain reaction as well as the most recent sequencing technologies are different approaches for assessing RNA packaging¹²⁸. The choice of the most appropriate method to quantify RNA in cells and virions should be based on several factors like sensitivity, accuracy, relative quantification, RNA size determination and probe flexibility. Northern Blot is the only technique that can be used to visualise RNA size and address dimerisation efficiency. The main advantage of the RPA over the Northern Blot is its higher sensitivity. However RT-qPCR is the technique of the highest sensitivity and accuracy. FISH, a method for *in vivo* RNA visualisation can provide information on quantity, localisation and dynamics of viral RNA. Finally, if afforded, deep sequencing of viral RNA can be used for representation of multiple target molecules in the sample. Advantages and limitations of these techniques are briefly summarised in table 2. To address how efficiently RNA is encapsidated into virions one can measure RNA packaging efficiency and RNA packaging ability. Packaging efficiency is defined as the absolute amount of RNA per virus and can be measured with quantitative methods. Precise viral production calculation is achieved by the combination of ELISA (enzyme-linked immunosorbent assay) for Gag and measurement of RT activity. RNA packaging ability is defined as the ratio of gRNA amount in virions to gRNA in cells (gRNA virus/gRNA cell).

| <i>In Vitro</i> | Technique | Advantages | Disadvantages |
|------------------------|--|---|---|
| | Northern Blot | Information about the length and integrity of target RNA | Labour intensive, variable transfer efficiency, sensitivity based on RNA quality |
| | Ribonuclease protection assay (RPA) | Improved sensitivity and reproducibility | Specificity based on probe quality, no information of RNA length |
| <i>In Vivo</i> | RT-qPCR | Highest sensitivity and accuracy, high throughput, safe | Laborious, requirement of multiple controls for reliable detection, difficult to remove plasmid DNA |
| | FISH | Visualisation of RNA on a single virus/cell level, high specificity, detection of protein-RNA interactions, information on RNA localisation | Fixation and permeabilisation of cells are prerequisites, but can alter molecular conformations |
| | Next-generation-sequencing | Complete viral transcriptome analysis, relative expression of multiple RNAs at the same time | Costly, requires thorough normalisation and statistical analysis |

Table 2. Overview of methods to measure RNA packaging¹²⁹.

2.2. Lentiviral vectors for gene therapy

2.2.1. History of gene therapy

Gene therapy employs the introduction of therapeutic genes to either treat or prevent a disease. The concept of gene therapy, also commonly described as ‘using DNA as a drug’, was first explored during the 1960s and has been possible only after the identification of

DNA as the molecule that carries the genetic information. This section presents a brief history of the evolution of gene therapy and the discoveries that led to its inception.

Griffith's experiment in 1928 showed that the virulence of *Streptococcus pneumoniae* bacteria can change with the presence of a transforming principle¹²⁹. The exact nature of this transforming principle was later identified by the observations of Avery, McLeod and McCarty and by Hersey and Chase. The Avery-McLeod-McCarty experiment in 1944 suggested for the first time that DNA and not proteins carry the genetic information and cause bacterial transformation¹³⁰. In 1952, Hersey and Chase's experiments on viruses that infect bacteria, bacteriophages, supported Avery's, McLeod's and McCarty's conclusion. They used different radioactive isotopes to label proteins and DNA and observed that the DNA and not the proteins of T2 bacteriophages can enter the bacterial wall of *E.coli* bacterial cells¹³¹. The final confirmation of the suitability of DNA to serve as genetic material came from the important work performed by Franklin, Watson and Crick, that elucidated the DNA tertiary structure in a form of a double helix¹³².

The 1960s saw the birth of the 'Central dogma' of molecular biology, which describes the flow of information from DNA to proteins via an RNA intermediate. Between 1960-1980 scientific advancements, including the development of genetically marked cell lines¹³³, the clarification of mechanisms of cell transformation by SV40 and papoviruses, infection by RNA tumour viruses¹³⁴, the production of the first recombinant DNA molecule¹³⁵ as well as the use of calcium phosphate for cell transformation¹³⁶ led to the first human gene therapy study. Despite calcium phosphate transfection being too inefficient to be used in clinical applications, Martin Cline at UCLA reported successful introduction of the human globin gene into murine bone marrow cells and partial repopulation of the marrow of recipient mice with these genetically modified bone marrow cells¹³⁷. These surprising results encouraged Cline and his colleagues to embark on an epochal gene therapy clinical in 1980. Bone marrow cells from beta-thalassaemia patients in Italy and Israel were transfected *ex vivo* with plasmids encoding the human beta-globin gene and were then re-infused into patients. Cline had proceeded to this trial without having obtained permission from UCLA's Institutional Review Board. Even though the results of his study showed no adverse effects, his failure to seek and secure appropriate approval attracted severe criticism from the scientific, public policy and ethics committees¹³⁸. While this incident was disheartening for

the scientific community the controversy around it led to the establishment of the National Institutes of Health Gene Therapy Subcommittee for the regulation of genetic applications in humans in the United States in 1984. The Cline experiment also highlighted the need for more efficient gene delivery tools drawing much attention on the engineering of viral vectors.

Sir Peter Medawar once described viruses as *“bad news wrapped up in protein”*. However, advancements in the manipulation of DNA and molecular biology of viruses have enabled scientists to convert pathogens to therapeutic vehicles. Such advancements have led to blossoming of their use for disease treatment and more recently the commercialisation of several products. Glybera[®] was the first gene therapy product ever to receive approval from the European Medicines Agency (EMA). UniQure biopharma was given the green light by EMA to use Glybera[®], an adeno-associated virus serotype 1 (AAV1)-based product, to treat patients with lipoprotein lipase (LPL) deficiency under exceptional circumstances in 2012¹³⁹. However, five years after becoming the first gene therapy product to win approval in Europe, its limited usage in the EU market and its extreme cost, at an average of \$1 million per treatment, have led to Glybera[®]'s withdrawal. Upon publication of supportive results on the gene therapy treatment of 10 children with severe combined immunodeficiency (SCID) due to the lack of adenosine deaminase (ADA)¹⁴⁰, an alliance between San Raffaele, Molmed and GSK was formed in 2010 which resulted in the launch of Strimvelis in 2016. Strimvelis is the first *ex-vivo* gene therapy to gain EMA approval and its use is indicated for the treatment of ADA-SCID patients for whom no suitable human leukocyte antigen (HLA)-matched related stem cell donor is available. Strimvelis product contains autologous CD34+ cells transduced with a retroviral vector that encodes the human ADA cDNA sequence from human CD34+ cells. In 2018 GSK announced the sale of rare diseases drugs including Strimvelis, despite its approval by the National Institute for Health and Care Excellence (NICE)¹⁴¹. The success of gene therapy for the treatment of monogenic rare diseases and the need for personalised cancer treatments has given rise to a different gene therapy application. The University of Pennsylvania and Novartis first formed an alliance on 2012 that enabled the translation of Carl June's work on Chimeric Antigen Receptors (CARs)¹⁴² from bench to bedside. KYMRIAH, a CD19-directed genetically modified autologous T cell immunotherapy, was the first FDA approved CAR-T therapy. T-cells are transduced with a

lentiviral vector encoding an anti-CD19 chimeric antigen receptor (CAR) used for the treatment of patients up to 25 years of age with B-cell precursor acute lymphoblastic leukemia (ALL)¹⁴³.

Gene therapy has the potential to become an important broad approach for disease treatment. As described in this section, gene therapy has entered clinical reality with authorisations for the treatment of inherited diseases and cancers both in Europe and the U.S. Rational vector design has been the focus of many research groups and is shown to be essential for the successful development of gene therapy. The next chapter outlines the different categories of viral vectors and their basic properties.

2.2.2. Viral vectors as gene delivery tools

The 1990s saw the rise of gene therapy as an approach to cure the majority of human diseases. This exciting potential was impeded with the death of the 18 year old patient Jesse Gelsinger who participated in a gene therapy pilot study for the treatment of the non life threatening condition ornithine transcarbamylase (OTC) deficiency. Infusion of human adenovirus type 5 coding for the OTC cDNA caused systemic inflammatory response syndrome and multiple organ system failure that led to the death of this young patient within 98 h post treatment¹⁴⁴. Moreover, the initial success achieved in the treatment of children with SCID was replaced with severe scepticism when two different groups reported *LMO2* proto-oncogene triggered development of leukemia in some patients^{145 146}. The integration pattern and the enhancer activity of the vector used in this study led to transcriptional dysregulation of nearby genes like *LMO2*, and were thus found to be accountable for the observed insertional oncogenesis. These two setbacks associated with inexperience in the design of clinical trials and in the assessment of viral vectors' safety profile inevitably resulted in decrease of funding and conduct of gene therapy clinical trials. Luckily, it also led to a new round of viral vector engineering for improved safety and gene delivery which brought us the resurgence of gene therapy. The spectrum of viral vectors is broad and includes both DNA and RNA viruses, with either single stranded or double stranded genomes, developed for transient or permanent expression¹⁴⁷. This wide selection allows for the appropriate choice of viral vector type depending on the application for which they are required. As this thesis focuses on viral vector design, the use of non-viral vectors

will not be discussed any further; instead a brief overview of the most popular viral vectors for gene therapy is presented below.

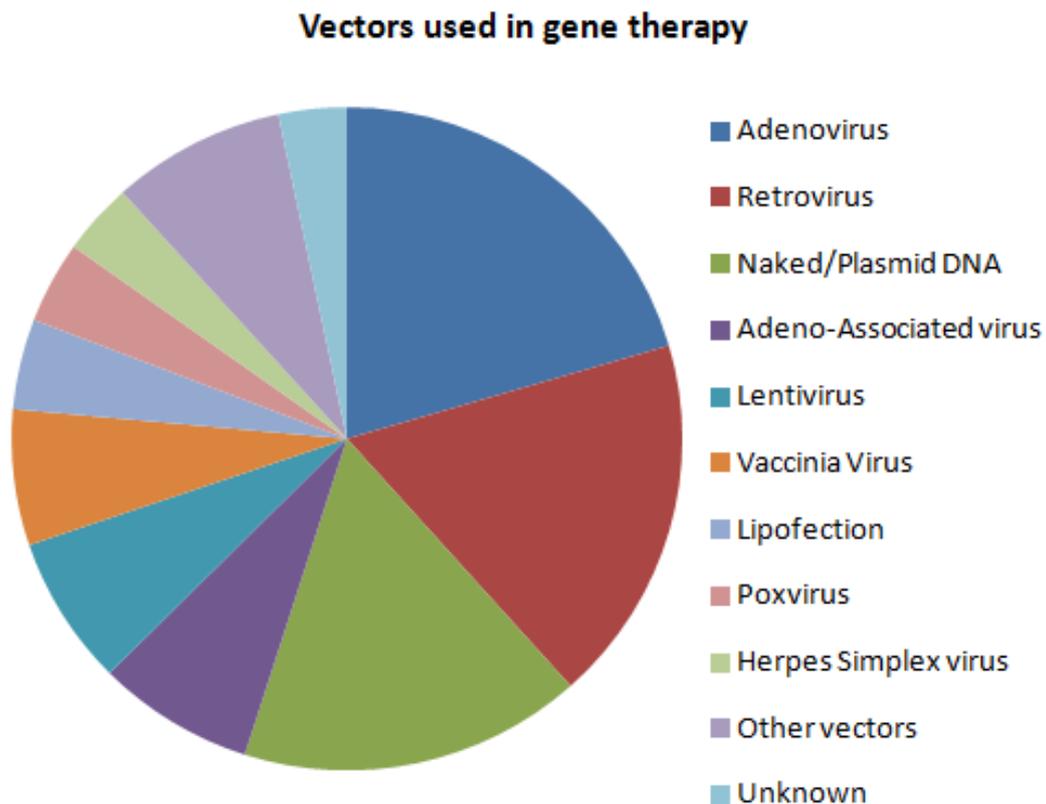


Figure 6. Representation of vectors used in clinical trials. Adapted from source:¹⁴⁹.

Human Adenovirus serotype 5-derived vectors are presently the most commonly used gene delivery vehicle in clinical trials. Its genome is dsDNA with an insert capacity of 8-36 kb¹⁴⁸, offering transient gene expression in a variety of host cells¹⁴⁹. The first generation of adenoviral vectors was causing strong immune responses as adenoviruses can naturally infect a variety of human tissues. That led to the most recent third generation design that contains big deletions showing higher genome capacity, reduced immunogenicity and extended expression periods¹⁵⁰.

Retroviruses are the next most frequently used type of viral vector in clinical applications due to the availability of stable cell lines for large scale production. Retroviruses are ssRNA enveloped viruses with genome sizes that vary between 7 and 11 kb¹⁴⁸ and which unlike that of adenovirus gets integrated in the DNA of the host, achieving sustainable

expression¹⁵¹. Unlike lentiviruses, retroviruses cannot pass through the nucleopore and as a consequence they can only infect mitotically active cells and are likely to integrate in cell cycle associated genes. More specifically, retroviruses like gammaretroviruses tend to integrate near transcription start sites due to their interaction with BET proteins¹⁵², a trait which has been associated with insertional deregulation of gene expression¹⁵³, while lentiviruses tend to integrate throughout the bodies of actively transcribed genes due to interaction with the LEDGF/p75 protein^{154 155}.

Adeno-Associated virus (AAV) is a popular choice for *in vivo* gene therapy and was the first type of viral vector to be commercially licensed. AAV vectors have a small ssDNA genome with a packaging capacity of only 4 kb that similar to retroviral vectors, provides sustainable long-term gene expression through chromosomal integration¹⁵⁶. Engineered dual vectors have been developed to address this insert capacity limitation associated with the use of AAV vectors¹⁵⁷. Even though AAV is generally considered a vector with a safe profile characterised with low pathogenicity and toxicity, multiple AAV serotypes, including AAV2, 3, 5, 6 and 8 have been developed to prevent potential immune responses triggered by repeated administrations¹⁵⁸.

As shown in Figure 6, other viral vectors frequently used in gene therapy clinical applications include lentiviruses, vaccinia virus, poxviruses and herpes simplex viruses. The next section discusses lentiviral vectors, the topic of this PhD project, in further detail.

2.2.3. Evolution of lentiviral vector design

2.2.3.1. First generation lentiviral vectors

The majority of lentiviral vectors are derived from HIV-1, which is pathogenic. Lack of cure for the disease it is causing, acquired immune deficiency syndrome (AIDS), posed significant considerations for its use as a gene transfer vehicle. Initially, the main focus was the preparation of an HIV-1 derived vector characterised by a decreased likelihood of RCL (replication competent lentivirus) generation. Additionally, HIV-1 envelope glycoprotein gives access only to CD4+ cells that contain the correct co-receptors, CXCR4 and/or CCR5, which would limit the potential therapeutic use of the vector system to a very small number of cell types¹⁵⁹. HIV-1 was converted into a replication deficient vector by separating its

genome into three independent plasmids, while broad tropism was achieved by VSVg pseudotyping¹⁶⁰. As shown in Figure 7.A, the first-generation lentiviral vector system consisted of these three independent elements: i) the packaging construct coding for the enzymatic, structural, regulatory and accessory proteins, ii) an envelope construct, while the HIV-1 envelope gene was deleted or mutated and iii) the transfer vector genome construct, carrying the minimum *cis*-acting elements (LTRs, ψ and RRE for transcription, encapsidation and RNA transport), and the therapeutic transgene under an internal promoter. Splitting the genome into 3 separate plasmids means that multiple independent recombination events would have to take place for replication competent lentiviruses to be generated. Additionally, substitution of HIV-1 *env* for the VSV glycoprotein gene further minimised the risk of recombination by reducing the homologous sequences¹⁶¹.

2.2.3.2. Second generation lentiviral vectors

In second generation vectors, multiply attenuated LV systems were achieved by removing accessory proteins. With five out of nine genes present in the parental virus deleted, the LV remained fully functional and able to transduce non dividing cells *in vivo* and *in vitro*¹⁶². These packaging constructs express Gag, Pol, Tat and Rev only (Figure 7.B)¹⁶³.

2.2.3.3. Third generation lentiviral vectors

In an attempt to further improve the safety of the lentiviral system, the U3 promoter of the 5'LTR was replaced with strong viral promoters, including those derived from Cytomegalovirus or Rous Sarcoma Virus, to remove the requirement for Tat protein whose role lies at the transcription level^{163 164}. Additionally, the *rev* open reading frame was moved to an independent fourth plasmid. Thus, third generation lentiviral vector systems contain only three out of nine HIV-1 genes and consist of four separate plasmids. Similar to previous systems, there is i) a packaging plasmid, which is now only expressing the Gag/Pol proteins, ii) an envelope construct typically coding for VSVg iii) an independent plasmid expressing the Rev protein and iv) a transfer vector construct that drives the expression of the transgene of interest (Figure 7.C).

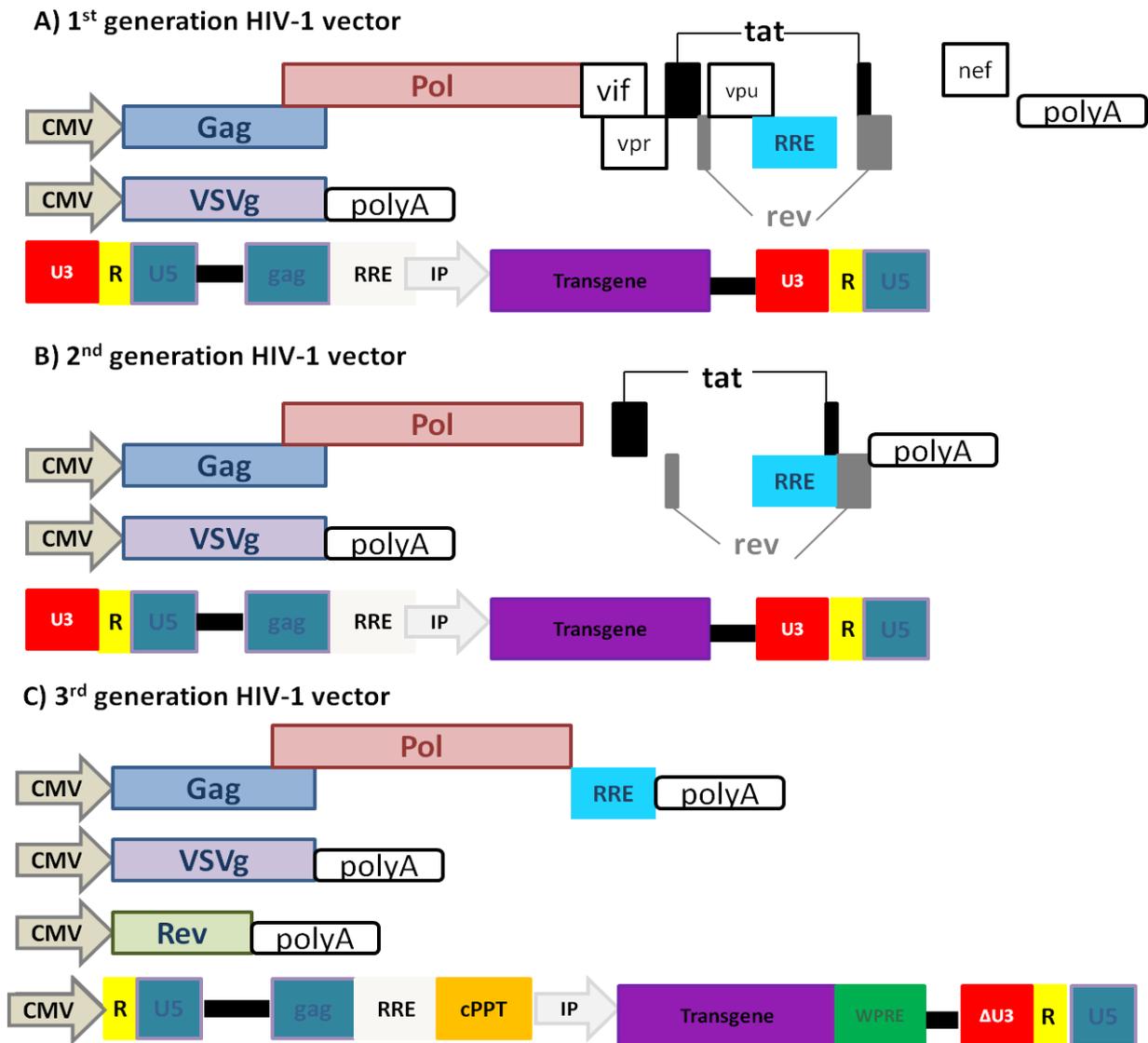


Figure 7. Schematic presentation of the evolution of lentiviral vector design.

[A] First generation lentiviral vector system consisting of a transfer vector, envelope and packaging construct. **[B]** For the second generation lentiviral vector system all accessory proteins are removed with the rest of the constructs remaining intact as presented in the first generation system. **[C]** Third generation lentiviral vector system consisting of four plasmids including a Tat-independent SIN transfer vector, an envelope construct, a Rev-expressing and a Gag-expressing construct. (A detailed explanation of transfer vector elements, including cPPT, IP and WPRE, is presented in section 2.2.7 below.)

2.2.4. Properties of lentiviral vectors

Moloney murine leukaemia virus (MoMLV) was the first gammaretrovirus to be used for the preparation of therapeutic vectors¹⁶⁵. The construction of a mutant MoMLV that was defective in the packaging of gRNA into virions was an important step towards the development of retroviral vectors¹⁶⁵. Until now, MoMLV-derived vectors have been the most frequently used retroviral vectors. Following the advancements in the retroviral vector design, the first lentiviral vector gene delivery systems based on HIV-1 were developed by the early 1990s¹⁶⁶. The inability of retroviral vectors to transduce non-dividing cells¹⁵¹ combined with the safer lentiviral vector integration profile¹⁶⁷ have increased the interest for the development of clinical grade lentiviral vectors for gene therapy applications.

Lentiviruses are a subclass of retroviruses characterised by a relative complex genome coding for a number of accessory proteins in addition to the canonical retroviral Gag/Pol and Env products. Retroviral particles are on average 100 nm in diameter and consist of 1% RNA, 60–70% protein, 30–40% lipid, the latter derived from the membrane of the producer cell and 2–4% carbohydrate, mainly from glycosylation of the envelope protein. This family of viruses is characterised by their ability to reverse transcribe and stably integrate their genetic information into the host genome hence when used as vectors, achieving sustained gene delivery and expression. Lentiviral vector replication includes receptor-mediated entry, capsid uncoating, reverse transcription, trafficking to the nucleus and integration into the host cell genome; common properties of the retroviral life cycle¹⁵¹.

Various properties of HIV-1 have been exploited or modified on different stages of its life cycle to convert it to an attractive vector for clinical applications. Lentivirus mediated transduction starts with the interaction of the envelope protein with receptors on the target host cells. One mechanism for altering or expanding cellular tropism is through the formation of phenotypically mixed particles, a process known as pseudotyping, that can occur during viral assembly in cells infected with two or more viruses¹⁶⁸. Several studies showed that WT HIV-1 produced in cells infected with xenotropic murine leukemia virus (MLV)¹⁶⁹, amphotropic MLV¹⁷⁰ or herpes simplex virus¹⁷¹ can give rise to phenotypically mixed virions with an expanded host range, suggesting that HIV-1 biology allows pseudotyping. These early observations were confirmed and extended in 1996 when three

independent groups showed that VSVg pseudotyped HIV-1 virions can be successfully produced^{172 173 174}. These pseudotyped vectors had the advantages of targeting a plethora of cells as VSVg appeared to interact with a ubiquitous cellular “receptor” on cells. The receptor for VSV has long been thought to be a ubiquitous membrane lipid, phosphatidylserine¹⁷⁵ until Finkelshtein *et al.*, identified the low-density lipoprotein receptor family (LDLR) as the VSV receptors on the membranes of cells¹⁷⁶. After the processes of entry and uncoating, the gRNA is reverse transcribed while forming the nucleus for the reverse transcription complex (RTC). The newly synthesised cDNA is associated with several cellular and viral proteins in the preintegration complex (PIC)¹⁷⁷. The ability of LVs to transduce non-replicating cells has been of great importance for the development of clinically useful gene therapy vectors. Since the PIC is too big to cross the nuclear pores, HIV must have found alternative ways to interact with the nuclear import machinery to infect non-dividing cells. The exact mechanism remains controversial, but the most recent studies by Emerman *et al.*, suggest the viral capsid to be the key determinant of HIV-1 infection of non-dividing cells.

Based on the properties described above, LVs have been proven to be an easy and useful system for stable expression of transgenes both for clinical and basic research applications¹⁷⁸. Lentiviruses, unlike Adenoviruses achieve sustainable gene expression via stable integration in the host genome and, unlike gammaretroviruses, lentiviruses have the ability to infect both dividing and non-dividing cells, making them attractive tools to transduce cells regardless of their mitotic state¹⁷⁹. HIV-1 derived vectors allow for pseudotyping with a plethora of envelopes, with VSVg being the most widespread one, offering broad tissue tropism. VSVg pseudotyping offers high vector particle stability, allowing for concentration by ultracentrifugation¹⁸⁰. Notably, LV components show low immunogenicity in humans compared to that of other viral gene transfer vectors^{158 181}. Lastly, most recent studies suggest a potentially safer integration site preference for lentiviral vectors compared to other retroviral vectors. Gammaretroviral vectors have a tendency to integrate near the transcription start site (TSS) of genes, a property associated with clonal expansion and carcinogenesis, whereas LVs show a preference to integration inside actively transcribed genes¹⁶⁷.

However, vectors that integrate their genome in the genetic material of the host always carry the possibility of transforming mutation¹⁸²; although the introduction of Self Inactivating (SIN) LVs aims to eliminate the likelihood of such events, see below¹⁸³. Despite being an easy system for manipulation and production, high vector doses and prolonged *ex vivo* culture conditions are required to achieve high transduction efficiencies of clinically relevant cellular targets as exemplified in recent attempts to use LVs in clinical trials¹⁸⁴. Currently, lentiviral vector production is inefficient and expensive¹⁸⁵. Therefore, advancement of the production system and increase of titres could significantly improve patient access to life changing gene therapy treatments. In the following paragraphs, I detail how vector design and safety has evolved and improved through successive generations as a prelude to our current approach targeting packaging efficiency.

2.2.5. Clinical use of lentiviral vectors

The first clinical trial using lentiviral vectors took place in 2003 and aimed to treat the acquired immunodeficiency syndrome. The vector used was a Tat and Rev dependent HIV-1 derived lentiviral vector that expressed an HIV env antisense RNA in CD4+ T cells¹⁸⁶. In 2006 and 2007, two additional clinical trials employing lentiviral vectors were designed for the treatment of β -thalassemia, caused by reduced production of β -globin in erythroid cell¹⁸⁷ and Adrenoleukodystrophy (ALD), a brain disorder caused by myelin destruction¹⁸⁸. The first attempt to treat β -thalassemia with a gene therapy approach used the LentiGlobin vector, a self-inactivating vector that contained large elements of the β -globin locus control region as well as chromatin insulators¹⁸⁷. The first clinical trial using a non-HIV-1 based lentiviral vector started in 2007 in France. An equine infectious anaemia virus (EIAV) based lentiviral vector called ProSavin was designed to express the three key dopamine biosynthetic enzymes (tyrosine hydroxylase, aromatic L-amino acid decarboxylase and GTP cyclohydrolase 1) by Oxford Biomedica for the treatment of Parkinson's disease¹⁸⁹. Ever since, multiple clinical therapies using lentiviral vectors have been announced and many more are to be expected. Currently, LVs are involved in 9.5% of all the gene therapy clinical trials worldwide and in 19% of those for monogenic diseases¹⁹⁰. As shown in table 3, inherited immunodeficiencies are important targets for gene therapy¹⁹¹. Of special interest was the development of a lentiviral vector for Wiskott-Aldrich syndrome (WAS) whose

transfer vector construct contained a promoter fragment derived from the WAS protein (WASP) gene to drive the expression of WASP cDNA¹⁹².

| Condition | Clinical Trial phase | NCT Number |
|---|----------------------|-------------|
| Transfusion-dependent β -thalassemia | 1/2 | NCT02453477 |
| | 3 | NCT02906202 |
| Cerebral adrenoleukodystrophy | 2/3 | NCT01896102 |
| Sickle cell disease | 1 | NCT02140554 |
| | 1 | NCT02193191 |
| Metachromatic leukodystrophy and adrenoleukodystrophy | 1/2 | NCT02559830 |
| Wiskott-Aldrich syndrome | 1/2 | NCT01347346 |
| | 1/2 | NCT01347242 |
| | 1/2 | NCT02333760 |
| X-SCID | 1/2 | NCT01306019 |
| | 1/2 | NCT01512888 |
| ADA-SCID | 1/2 | NCT02999984 |
| | 1/2 | NCT01380990 |
| Fanconi anemia | 2 | NCT02931071 |
| X-linked chronic granulomatous disease | 1/2 | NCT02234934 |

Table 3. Ongoing clinical trials using lentiviral vectors.Source:¹⁹⁴.

2.2.6. SIN (Self Inactivating) LVs

Following safety advancements in Murine Leukaemia Vectors (MLV), the SIN (Self Inactivating) LV was constructed. Initially, this was achieved by a 133bp deletion in the U3

region of the 3'LTR¹⁸³. Later in the same year, a different group described the production of a SIN LV by deleting 400bp of the 3'LTR¹⁹³. Upon reverse transcription, this deletion is transferred to the 5'LTR, leading to the transcriptional inactivation and immobility of the transfer vector, while maintaining the viral titres at a similar level. SIN vectors inactivate the 5'LTR and allow the internal promoters to work without upstream competition. As importantly, SIN vectors inactivate the 3'LTR and thus reduce the risk of it acting as a promoter that could potentially activate downstream cellular genes in the cell genome. Although retroviral infection can be associated with proto-oncogene activation, this has never been observed with HIV-1; a trait important for the clinical use of HIV-1 based vectors in gene therapy¹⁹⁴. However, an expanded cell clone which contained an integrated vector in the proto-oncogene HMG2A was noticed in a recent β thalassemia clinical trial¹⁹⁵. The use of SIN vectors offers the advantage of reducing the likelihood of any potential insertional activation of cellular oncogenes and also helps reduce the possibility of RCL production.

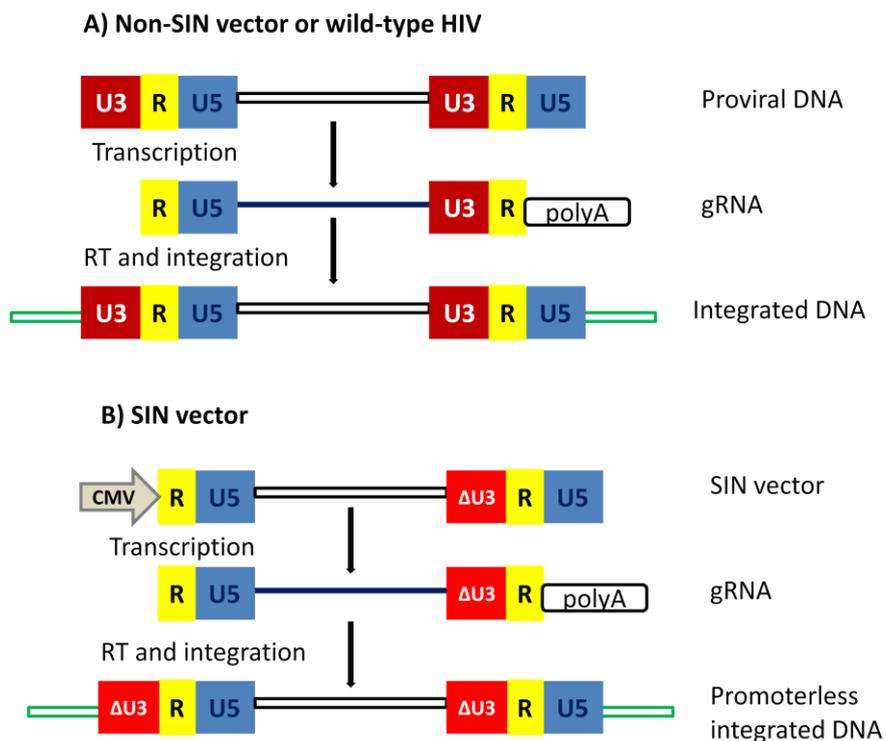


Figure 8. Schematic representation of non-SIN and SIN vector.

2.2.7. Transfer Vector design The current system used in clinical gene therapy applications is a 3rd generation SIN LV system with a heavily modified transfer vector to achieve maximum efficiency and safety (Figure 9). The strong CMV promoter replaces the U3 promoter/enhancer region of the 5'LTR¹⁶⁴. The *cis*-acting sequences from the R region up to the end of the 5'LTR are indispensable and are maintained intact. The SL3 stem loop with a 360 bp fragment of *gag* and other upstream sequences serve as the minimum core packaging signal¹⁹. Following the beginning of the *gag* ORF, there is a copy of the RRE required for successful transport of the unspliced transcript from the nucleus to the cytoplasm. There are multiple papers describing the efforts to make this system Rev/RRE independent in an attempt to reduce the amount of HIV-1 derived sequences; in one of them, the RRE element is substituted with alternative RNA export elements, primarily the constitutive transport element (CTE) from different Simian retroviruses. Despite the presence of data suggesting that the concatemerization of CTE elements could efficiently replace the Rev/RRE system¹⁹⁶, the HIV-1 derived lentiviral transfer vector established for reliable high titre production continues to be a Rev dependent one. The central polypurine tract (cPPT) is important for the initiation of plus strand synthesis during reverse transcription and leads to increased transduction efficiency via stimulation of nuclear import¹⁹⁷. Additionally, in vector systems that included the cPPT, a correlation between the dosage of the vector and transgene expression was noticed, which is in contrast with the plateau effect observed when using conventional non-cPPT including vectors¹⁹⁸. Studies have demonstrated that alteration of the vector's enhancer/promoter elements has a more significant effect on biosafety, with regards to transformation and expansion of transduced cells, than the retroviral insertion pattern *per se*¹⁹⁹. Thus, traditionally, gene therapy approaches make use of weak internal promoters. The latest *cis*-acting modification that was introduced for improved expression is the WPRE (woodchuck hepatitis virus post transcriptional regulatory element) sequence that increases the level of transgene expression significantly, by increasing the amount of both nuclear and cytosolic WPRE-containing unspliced transcripts. Based on the observations that HBV PRE (HPRE) and an intron were functionally equivalent, it was initially thought that WPRE enhanced gene expression via promoting nuclear export²⁰⁰. Further studies showed that its post transcriptional effects do not result from increased RNA export, or an increased

transcription rate, or a longer RNA half-life, but rather from improved RNA processing²⁰¹. Although the exact mechanism of the WPRE remains unknown, it is thought to act by driving efficient 3' processing of RNA during the early synthesis stage²⁰². Despite the great advantage of including WPRE in the vector design, concerns arose due to the presence of an ORF within this element that codes for a truncated peptide derived from the WHV X gene that is associated with liver oncogenesis²⁰³. The safety of this element has been improved with the introduction of mutations that abrogate the WHV X expression while still maintaining high transgene expression²⁰².

Despite decades of study and iterative improvements in vector design, LV transduction efficiency remains a limitation to their successful translation to clinical use.

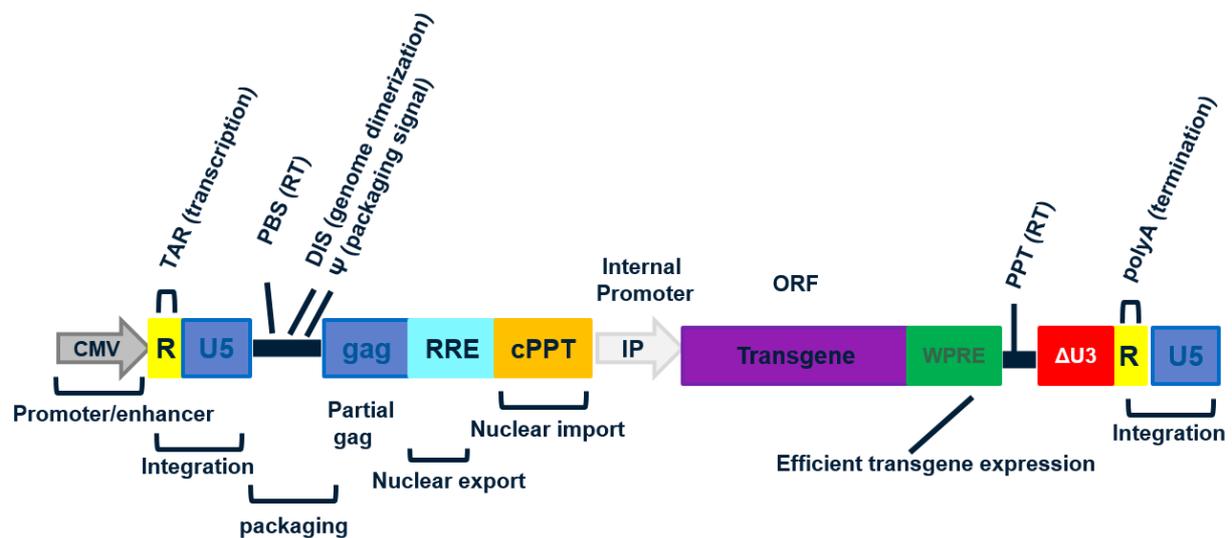


Figure 9. Components of a typical 3rd generation lentiviral transfer vector.

2.2.8. Provisional lentiviral vector design modifications

Introduction of the Chicken β -Globin Locus HS4 Insulator Element (cHS4) was used to address the safety issues regarding vector insertional mutagenesis. The full 1.2 kb insulator element negatively affected vector titre and transgene expression, whereas a 0.25 kb core, part of the insulator, managed to rescue vector titres by improving reverse transcription. Introduction of this core element was also shown to act as a barrier by reducing variability of expression due to position effects and by blocking enhancer activity²⁰⁴. However, even though a single cHS4 insulator 0.25 kb core was found to be effective in lentiviral vectors²⁰⁴, its inclusion in the extended β -globin locus control region of a β -globin lentiviral vector²⁰⁵

did not manage to prevent cell expansion when used for the treatment of human β -thalassaemia¹⁹⁵. Similarly, the enhancerless novel human *HNRPA2B1-CBX3* UCOE (A2UCOE) was assessed for its ability to overcome insertion-site position effects. A2UCOE belongs to the category of ubiquitously acting chromatin opening elements (UCOEs). UCOEs are genetic regulatory elements known for their ability to confer resistance to transcriptional silencing and thus achieving stable, reproducible, high levels of gene expression²⁰⁶. Results suggested that the A2UCOE reaches therapeutically relevant levels of gene expression in the absence of enhancer sequences adding an extra layer of safety to clinical vector use²⁰⁷. Another approach that addressed the safety concerns associated with lentiviral integration led to the development of novel lentiviral vectors (LTR1) designed to prevent transfer of HIV-1 packaging sequences to patient cells. LTR1 vectors reverse transcribe their genomes with a single strand transfer instead of the usual two, resulting in integrated proviruses that lack packaging sequences. LTR1 vectors are suggested to offer remobilisation resistance, reduced frequency of splicing into human genes and higher transgene expression²⁰⁸.

The instability sequences present in the coding region of *gag/pol* and the codon bias of the HIV-1 genome, which unlike the human one, is AU-rich, are considered accountable for the Rev dependence of the lentiviral system²⁰⁹. To address poor Gag/Pol protein expression Kotsopoulou *et al.*, created a codon optimised Rev-independent construct²¹⁰. Codon optimisation is an approach used to substitute rare codons by frequent codons according to the genomic codon usage in a host organism and thus improve protein expression²¹¹. Additionally, the use of this codon optimised Rev independent construct offers improved safety by minimising the homology between the partial *gag* included in the transfer vector and the beginning of the *gag* ORF of the packaging construct. A similar codon optimisation approach was followed years later to address safety concerns associated with the production of replication competent lentiviruses. It was shown that Rev-Independent *gag/pol* eliminates *psi-gag* recombination, hence its inclusion in the lentiviral system offers a safety advantage for clinical use²¹². A codon optimised *gag/pol* construct was also used for the production of a stable cell for clinical vector production²¹³. Stable cell line development is generally associated with the advantages of stable and scalable vector production. However, the viral titre of the aforementioned stable cell line was only of 10^6 transducing units/mL, while transient transfections can yield supernatant of up to 10^8 transducing

units/mL. Additionally the expression of certain therapeutic transgenes has been observed to negatively impact vector production by affecting cell metabolism and/or vector virion itself. To overcome this challenge Oxford Biomedica developed the Transgene Repression In vector Production (TRiP) system that blocks mRNA translation. Introduction of the TRiP system in a lentiviral vector coding for the Cyclo-oxygenase-2 enhanced titres by 600-fold²¹⁴.

2.2.9. Transduction efficiency-What are the challenges?

Despite the tremendous potential of hematopoietic gene therapy, Hematopoietic Stem and Progenitor cells (HSPCs), the target cells of gene therapy applications, are poorly permissive, requiring multiple rounds of high vector doses and prolonged *ex vivo* culture to reach adequate transduction levels^{215 216}. This poor permissiveness can be explained by the retroviral origin of LVs combined with the high viral particle load used in LV transduction compared to a natural LV infection. Viruses evolve to compete against their hosts which in turn develop sophisticated defensive mechanisms against them. Along with LV transduction resistance, HSPCs have also proved to be resistant to viral and bacterial infections^{217 218}. Recent research has revealed the expression of mammalian antiviral factors known as restriction factors (RFs)^{219 220}. Viral arsenals include several accessory proteins whose major role is to counteract RFs and avoid innate immune activation²²¹. LV lack of accessory proteins make LV-mediated gene transfer inefficient, highlighting the importance of improved LV design to increase transduction efficiency.

The size, sensitivity to inhibitors, particle complexity and instability of lentiviral vectors can further affect vector integrity and functionality after bioprocessing²²². Low titres and maintenance of infectivity are the biggest challenges for upstream and downstream processing respectively. Retroviral particles typically have a half-life of 6-7 hours at 37°C. The reduction of LV infectivity at 37°C was shown to be associated with loss of reverse transcriptase activity²²³. Instability of Env due to dissociation of SU and TM contributes to particle instability; a problem which as explained in section 2.2.3, is now resolved with replacement of Env with VSVg. Additional challenges contributing to reduction of functional titres include the presence of reported cell culture media components that act as viral inhibitors^{224 225}. Host cell-derived proteoglycans and glycosaminoglycans have also been reported as inhibitors of transduction²²⁶.

It has been estimated that early phase clinical trials for *ex vivo* therapy require 10^{11} - 10^{12} functional particles²²⁷. Titres from transiently transfected HEK293T cells are typically close to 10^7 TU/mL²²⁸. It is generally assumed that the quality control consumes half the produced batch and that after downstream processing, only approximately 10% of the batch will be recovered. That leads to a requirement of harvesting of big volumes of viral supernatant which restricts gene therapy from taking off as a treatment for more common diseases.

To address these current challenges facing the field of gene therapy, research groups working with LVs have focused their recent efforts on improving the current transduction protocols. Such efforts include ultracentrifugation and culturing conditions optimisation²²⁹ and use of new transduction enhancers^{230 231}. However, such approaches are cell specific and are likely to increase costs and processing. Additionally, transient pharmacologic immune system modulation achieved by certain transduction enhancers²³¹ may not be a genetically safe approach. Therefore, here, I suggest the improvement of LV infectious titres by modifying the basic backbone of the 3rd generation transfer vector currently used in clinical applications, through generating new packaging signal RNAs with improved dimerisation and encapsidation properties.

2.3. Aims of the project

LV titres are a limiting factor for the wide use of lentiviral vectors in clinical applications. Many stages of the HIV-1 life cycle could be limiting vector titres, with packaging being a potential limiting factor candidate. As mentioned above, the LDI-BMH model suggests competition between the processes of packaging and translation⁷, which could be eliminated via rationally designing mutations in the transfer vector construct to increase packaging.

To increase LV infectious titre through generating packaging signal RNAs with more efficient dimerisation and encapsidation properties.

This project focused on improving the efficiency of lentiviral vector dimerisation and packaging. The tight relationship between these two processes had been shown by Sakuragi *et al.*, and has already been discussed¹⁰⁹. I designed dimerisation promoting mutations by targeting the polyA, U5-AUG and SL1 structural motifs in an attempt to evaluate their impact on vector infectivity and packaging efficiency. As an improvement to previous work I worked with the full-length gRNA encoded by the transfer vector construct and with a 690 nt long *in vitro* transcribed RNA fragment, instead of focusing solely on the 5'UTR. Furthermore, unlike previous studies, I assessed the dimerisation ratio and the encapsidation efficiency of gRNA produced in the cell rather than in a *in vitro* context. I then investigated whether the improvement of these two processes could also increase the transduction efficiency of these modified vectors. Identification and validation of mutations that increase LV packaging and infectious titre have the potential to impact the design of LV for use in basic and translational research.

3. Materials and Methods

3.1. Lentivirus production by transient transfection

293T cells were maintained in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum (GIBCO) at 37 °C. For transfection 5×10^6 cells were seeded under sterile conditions in 10 mL medium in a 10 cm² dish. The medium was changed 18 h after seeding and transfected 2-3 h later. For each transfection plate, 25 µg DNA (25 µg total, composition of vector mix is detailed in each experiment) were added to the first tube with 250 mM CaCl₂ in a 500 µL final volume. Air was bubbled through 500 µL of 1M, pH 7.0 HEPES Buffered Solution (HBS) while adding the CaCl₂-DNA solution drop-wise. The solution was incubated at room temperature for 30 min. The solution was mixed by pipetting up and down and immediately added dropwise, evenly across the entire plate, followed by swirling to ensure even distribution. Cells were incubated at 37°C overnight. Medium was replaced with 8 mL fresh medium 16 h post transfection. Virus-like particles were harvested 40 h post transfection, 1 mL aliquots of supernatant were filtered through a 0.2 µm filter, aliquoted into cryovials, and stored at -80°C. Transfection efficiency was assessed by flow cytometry.

3.2. Vector titration with flow cytometry

293T cells were seeded at 3×10^5 per well of a 6-well plate in a final volume of 2 mL. Viral supernatants were thawed and used to infect cells 18 h post-seeding at 10^{-1} , 10^{-2} and 10^{-3} dilutions in a final volume of 500 µL containing 8 µg/mL polybrene. Plates were returned to a 37°C CO₂ incubator for 2 h, followed by the addition of 1 mL medium to each well. The vector can be titrated 48 h post transduction by flow cytometry. 250 µL of TrypLE™ Express (1x) were added to cells, which were incubated with the dissociation reagent at 37°C for 5 min. 50 µL of the TrypLE™ Express treated cells were resuspended in 500 µL of PBS to perform flow cytometric analysis to determine the percentage of eGFP positive cells. The transduction efficiency of each produced lentiviral vector was calculated by multiplying the number of seeded cells by the percentage of eGFP positive cells and by the vector dilution factor. This value was then divided by the volume of lentiviral vector used for HEK293T cell transduction. Values used for titration were between 4% - 25% eGFP positive cells. Flow cytometry analysis was performed using the BD Accuri™ C6 Plus system (excitation at 488 nm, filter at 491 nm and

emission at 530 nm, filter at 533±30 nm)²³². Non transduced HEK293T cells were used to draw a live gate on a SSC vs FSC plot. Such plots were used to distinguish cell populations based on their forward and side scatter properties that gave an estimation of the cell size and granularity cells respectively. eGFP was detected in FL1 by using the standard C6 filter configuration (FL1 = 530/30 BP)²³². Hence, a count vs FL1 histogram was used to study the distribution of fluorescence intensities of transduced cells and calculate the number of eGFP positive cells. The FL1-X axis represented fluorescence intensity and the Y axis showed the cell count. The non transduced HEK293T sample was also used to set an FL1 threshold bar to differentiate eGFP expressing HEK293T cells to non-eGFP expressing HEK293T cells. The live gate and the FL1 bar were then applied to all samples to positively select populations for further examination. 10.000 events were collected for each sample of interest.

3.3. Design and cloning of mutants

3.3.1. Mutagenesis primer design

Primers to create mutations in pCCL-eGFP were designed using the Agilent QuikChange Primer Design Program available online²³³.

| Primer ID | Primer Sequence (5'-3') |
|------------------|---|
| uc2m Fw1 | gagtgcctcaagtagtgcgcgcccgtctgttgtgtga |
| uc2m Rv1 | tcacacaacagacgggcgcgccactacttgaagcactc |
| uc2m Fw2 | ctagaaggagagagacgggcgcgagagtgtcagtattaagcgg |
| uc2m Rv2 | ccgcttaatactgacactctcgcgcccgtctctctccttctag |
| U5sFw | gccttgagtgcctcaagtagtgcgcacccatctgttgtgtgactctggtaa c |
| U5s Rv | gttaccagagtcacacaacagatgggtgcgccactacttgaagcactcaagg c |
| NCSL4s Fw | cttcaagtagtggtgctgctgtttgttgtgtgactctggta |
| NCSL4s Rv | taccagagtcacacaacaacaggcacacactacttgaag |

| | |
|-------------------|---|
| NCSL4s Fw | tgggtgcgagagcgccagtattaagcggg |
| NCSL4s Rv | cccgcttaataactggcgctctcgaccca |
| S1 Fw | gagtgcttcaagtagtgtgcacccatctgttgtgtgactctggt |
| S1 Rv | accagagtcacacaacagatgggtgcacactacttgaagcactc |
| LU5AUG Fw1 | ctagcggaggctagaaggagagagacggggcacaagagcgtcagtat |
| LU5AUG Rv1 | atactgacgctcttgrgccgctctctctccttctagcctccgctag |
| LU5AUG Fw2 | gctagaaggagagagacggggcacaagaaaaaaggattattaagcgggggagaa ttagatcgc |
| LU5AUG Rv2 | gcgatctaattgtcccccgcttaataaccttttttcttgtgccgctctctct ccttctagc |
| polyAs Fw | cgtactcaccagtcgctggactcgctcttgccgtgc |
| polyAs Rv | gcacggcaagaggcgagtccaggcgactggtgagtacg |
| J9Gag+ Rv1 | ccgagtcctgcgccgagagagctcc |
| J9Gag+ Fw1 | ggagctctctcggcgcaggactcgg |
| J9Gag+ Rv2 | ctcaccagtcggcgccccctcgcc |
| J9Gag+ Fw2 | ggcgagggggcgccgactggtgag |
| SIVS1 Fw | tcggcttgctgaagtgagcacggcaagaggc |
| SIVS1 Rv | gcctcttgccgtgctcacttcagcaagccga |
| polyAw Rv | cagacgggcacacactgcttaagcactcaagg |
| polyAw Fw | ccttgagtgcttaagcagtgtgtgcccgtctg |
| ΔP1 Del Fw | cggcgactggtgagtacgcc ggctagaaggagagagatggggtg |
| ΔP1 Del Rv | ccatctctctccttctagcc ggcgtactcaccagtcgccg |

Table 4. List of primers used for SDM.

Mutations could not be directly introduced into the U5 regions of the WT pCCL-eGFP plasmid as this region is present both in the 5' and 3'LTRs. Thus a subcloning strategy was designed to isolate the 5'LTR via PCR and subsequently clone this fragment into the Blunt-TOPO vector (Invitrogen). Nucleotides 280-3100 (pCCL-eGFP plasmid map provided by Conrad Vink, GSK) were amplified using Fw Subcloning Primer 5'-TACGGTAAACTGCCCACTTG-3' and Rv Subcloning Primer 5'-TAGGTCAGGGTGGTCACGAG-3', including the 5'LTR and flanking sequences. PCR reactions contained 0.5 U Accuprime™ *Pfx* Supermix (22 U/mL of *Thermococcus* species KOD thermostable Polymerase, which leaves blunt ends, complexed with anti-KOD antibodies, 66 mM Tris-SO₄ (pH 8.4), 30.8 mM (NH₄)₂SO₄, 11 mM KCl, 1.1 mM MgSO₄, 330 μM dNTPs, AccuPrime™ proteins, and stabilisers, 250 nM of each primer and 50 ng of the plasmid template in a final volume of 25 μL. The thermal cycling conditions were: 95°C for 5 min, 35 cycles of: 95°C for 15 sec, temperature gradient 60-65°C for 30 sec, 68°C for 3 min and a final extension step at 68°C for 7 min. Following this, 1 μL of the PCR product, 1 μL of the TOPO-Blunt vector (Invitrogen) and 3 μL of water were incubated at 24°C for 30-40 min and 2 μL of the resulting products were used to transform Stbl3 *E.coli* competent cells. Cells were incubated on ice for 30 min, and heat-shocked for 45 sec at 42°C. 250 μL of pre-warmed S.O.C Medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, and 20 mM glucose) were added to each vial and cells were placed in a shaking incubator at 37°C for 1 h at 225 rpm, before the whole transformation suspension was spread on a selective plate and incubated overnight at 37°C. Insert-containing clones were identified by diagnostic digest with *EcoRI*. Mutagenesis was performed using the Site-Directed-Mutagenesis (SDM) XLII Agilent kit. Reactions contained 2.5 U *Pfu*, 200 μM each primer, 200 μM dNTP mix, 1x Buffer (50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate and 100 μg/mL BSA) and 50 ng of plasmid in a total volume of 50 μL. Cycling conditions were: 95°C for 1 min, 35 cycles of: 95°C for 50 sec, 60°C for 50 sec, 68°C for 5 min 30 sec and a final extension step at 68°C for 7 min. The SDM products were analysed by gel electrophoresis to ensure the presence of the correct PCR product size which was then treated with 10 U of *DpnI* to remove the parental methylated plasmid template and used to transform XL10-Gold® Ultracompetent Cells. Colonies were screened by diagnostic digest and mutations verified by Sanger sequencing.

3.3.3. Transfer of mutant inserts into pCCL

Transfer vector pCCL backbone (1µg) and mutated TOPO-vectors (1.5 µg) were digested with 20 U *NdeI* and 20 U *MfeI*-HF enzymes in a 20 µL reaction in 1x CutSmart Buffer (50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate and 100 µg/mL BSA) at 37°C for 2 h. Digestion products separated on a 1% agarose gel and bands of interest were excised and DNA extracted using a Gel extraction Kit (Qiagen). Ligation reactions contained 50 ng of each mutant insert and backbone and 5 U T4 expressLink ligase (Invitrogen) in a final volume of 20 µL and were incubated at room temperature for 2h 30min. Stbl3 E. Coli cells were transformed with 2 µL ligation reaction and cultures were amplified and screened as above.

3.4. Intracellular RNA purification

RNA from transfected cells was extracted using the Qiagen RNeasy kit. Cells were washed twice with 3 mL of sterile PBS and lysed in 1 mL RLT buffer at RT for 20 min with occasionally swirling. Lysates were homogenised by vortexing; 350 µL of each cell lysate sample were mixed with 1 volume of chilled 70% ethanol in a new Eppendorf tube and transferred to an RNeasy Mini Spin column, followed by centrifugation at 16,000 x *g* for 1 min. Columns were washed once with Buffer RW1 (700 µL), twice with Buffer RPE (500 µL) and eluted in 40 µL of RNase-free water.

3.5. Extracellular RNA purification

Vector supernatant (8 mL) was centrifuged for 10 min at 3000 x *g* and 1 mL of it was added to 400 µL 8.4% Opti-Prep in sterile PBS and centrifuged at 22000 x *g* for 2 h at 4°C. Pellets were resuspended in 100 µL PBS and incubated in 200 µL protease K extraction buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM EDTA, 1% SDS, 100 µg/mL Protease K (Ambion) and 100 µg/mL yeast tRNA (Sigma-Aldrich) at 37°C for 30 min. An equal volume of acidic phenol-chloroform (5:1) was added and the mixture was centrifuged for 2 min at 1600 x *g*. The upper layer was mixed with an equal volume of chloroform-isoamyl alcohol (24:1) and was centrifuged for 2 min at 16000 x *g*. The upper layer was mixed with 1/10 volume of 3M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol. The precipitate was incubated at

-80°C for 1 h and then centrifuged at 16000 x *g* at 4°C for 20 min. The pellet was washed with 200 µL ice-cold 70% ethanol followed by centrifugation at 16000 x *g* at 4°C for 20 min. The pellet was resuspended in 10 µL nuclease-free water and stored at -80°C.

3.6. RT-qPCR assays

3.6.1. WPRE RT-qPCR

3.6.1.1. DNase digestion

Extracted RNA (800 ng) was digested with 9 U Turbo DNase in a total volume of 40 µL at 37°C for at least 2 h. The volume was brought to 110 µL with RNase-free water and an equal volume of acidic phenol-chloroform was added. The mixture was vigorously vortexed and centrifuged at 16,000 x *g* for 2 min. Supernatant was carefully removed to a new Eppendorf tube. An equal volume of chloroform-isoamyl alcohol was added and the mixture was vortexed and centrifuged again. The supernatant was carefully removed to a new Eppendorf tube and mixed with 1/10 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol. The mix was incubated at -80°C for 1 h and then centrifuged at 16,000 x *g* at 4°C for 20 min. The supernatant was decanted, and the pellet was washed with 200 µL 70% ethanol and centrifuged at 16,000 x *g* at 4°C for 20 min. The pellet was air-dried, resuspended in 10 µL nuclease-free water and stored at -80°C.

3.6.1.2. Reverse Transcription

RNA samples isolated from the previous step were normalized to 100 ng per sample and used for cDNA synthesis using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems). Reactions contained 1x Buffer, 1x RT Random Primers, 40 mM dNTPs, 50 U Multiscribe Reverse Transcriptase and 20 U RNase Inhibitor in 20 µL final volume. Reactions lacking RT were included as controls. Thermal cycler conditions were: 25°C for 10 min, 37°C for 120 min and 85°C for 5 min.

3.6.1.3. Real Time qualitative PCR (qPCR)

The cDNA products were quantified by qPCR using primers targeting the WPRE element (Fw 5'-CCGTTGTCAGGCAACGTG-3' Rv 5'-AGCTGACAGGTGGTGGCAAT-3', and probe 5'-FAM-TGCTGACGCAACCCCCACTGGT-BHQ1-3'²³⁴). After optimisation the final reaction conditions were 100 nM probe, 400 nM each primer, 1x Fast Advanced buffer (Applied Biosystems) and 10 ng cDNA in 10 µL reaction volume. Cycling conditions were 95°C for 5 min and 40 cycles of 95°C for 15 sec, 60°C for 15 sec and 72°C for 15 sec. The cellular levels of β-actin cDNA were measured using the SYBR Green technology and 40 nM of the following primers: Fw 5'-GAGCGGTTCCGCTGCCCTGAGGCACTC-3', Rv 5'-GGGCAGTGATCTCCTTCTGCA TCCTG-3'. The reaction conditions were 40 nM each primer, 1x SYBR Green Fast Advanced buffer (Applied Biosystems) and 10 ng cDNA in 10 µL reaction volume. Cycling conditions were 50°C for 2 min, 95°C for 20 sec and 40 cycles of 95°C for 1 sec and 60°C for 20 sec.

3.6.2. Competitive RT-qPCR BglG and MSL assay

3.6.2.1. Cloning of the unique sequences BglG and MSL into the pCCL backbone

Oligos containing the sequences of BglG: 5'-GGATTGTTACTGCATTTCGCAGGCAAAACC-3' and MS2: 5'-GCATGAGGATCACCCATGT-3' flanked between the *PvuI* recognition sites were designed and then used in an annealing reaction that contained 100 µM Fw Oligo, 100 µM Rv Oligo, and 1x ExpressLink™ T4 Buffer (Invitrogen) in a 100 µL final volume reaction that was held at 90°C for 10 min. Reagents were allowed to cool down to RT. Meanwhile 1 µg of the pCCL backbone was digested with 20 U of FastDigest *PvuI* (Thermo Fisher Scientific) and dephosphorylated with 1 U of FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific) in a 20 µL reaction in 1x Fast Digest Buffer (Thermo Fisher Scientific) at 37°C for 1 h. Digestion products were separated on a 1% agarose gel and the bands of interest were excised and DNA extracted using the Gel extraction Kit (Qiagen). Ligation reactions contained 35 ng of annealed oligos, 100 ng backbone and 5 U of T4 expressLink ligase (Invitrogen) in a final volume of 20 µL and were incubated at room temperature for 2h 30min. Stbl3 E. Coli cells were transformed with a 2 µL ligation reaction and cultures amplified and screened as above.

3.6.2.2. RNA preparation

WT and Mutant transfer vectors were then used for co-transfections as described above. Intracellular and Extracellular RNAs were isolated as described in sections 3.4. and 3.5. Removal of contaminant plasmid DNA was achieved using the Turbo DNA-free™ kit (Invitrogen) in a reaction containing 8 U of Turbo DNase™ Enzyme, 1x Turbo DNase™ Buffer and the total amount of RNA isolated. Reaction was left at 37°C for 2 h. 0.2 volumes of the inactivation reagent were added and samples were incubated at RT for 5 min. Clean RNA was recovered after centrifugation at 10,000 x *g* for 2 min.

3.6.2.3. RT-qPCR

The reverse transcription step was performed as above. The cDNA products were then quantified by qPCR using a common set of primers Fw 5'-GAATTCTGCAGTCGACGGTA-3', Rv 5'-TCCAGAGGTTGATTGCGA-3' and Probes targeting either the unique sequence of the BglG 5'-FAM-TCGATCGGGATTGTTACTG-BHQ1-3' or the MS2 sequence 5'-HEX-CATGGGTGATCCTCATGCCGAT-BHQ2-3'. After optimisation the final reaction conditions were 750 nM BglG probe or 250 MS2 probe, 550 nM each primer, 1x Fast Advanced buffer (Applied Biosystems) and 3 µL cDNA in a 10 µL reaction volume. Cycling conditions were 95°C for 5 min and 40 cycles of 95°C for 15 sec, 60°C for 15 sec and 72°C for 15 sec.

3.7. Production of a DIG-labelled RNA probe for Northern Blotting

A 744 nucleotide PCR fragment including the eGFP ORF and the start of the WPRE sequence was amplified by PCR using primers Fw 5'-GCTCCCTCGTTGACCGAATC-3' and Rv 5'-TAA TACGACTCACTATAGGGTCGTCCATGCCGAGAGTGATC-3'. PCR reaction contained pCCL-eGFP template (100ng), 1x Phusion HF Buffer, 12.5 µM each primer, 10 mM dNTPs, 3% DMSO and 1 U HF Phusion DNA Polymerase (NEB). The thermal cycling conditions were: 98°C for 30 sec, 35 cycles of: 98°C for 10 sec, temperature gradient 58°C for 30 sec, 72°C for 1 min and a final extension step at 72°C for 10 min. The PCR fragment was electrophoresed on a 1% agarose gel, visualised by UV and extracted using the Qiagen gel extraction kit according to the accompanying protocol. *In vitro* transcription was carried out using the MEGAScript T7 transcription kit (Ambion). Reactions contained extracted PCR product (1

µg), 1x transcription buffer, 10 mM each ATP, CTP and GTP, 6 mM UTP, 7 mM DIG-UTP, and 2 µl of T7 Polymerase enzyme mix in a final volume of 20 µL. The product was treated with 6 U Turbo DNase for 1h at 37°C. RNA was visualised on a 1% Ethidium Bromide stained agarose gel alongside Low Range (LR) and High Range (HR) RNA ladders (Thermo Fisher Scientific). DNA template and the 1 kb Invitrogen ladder. The RNA probe was purified by lithium chloride precipitation, resuspended in RNase-free water and stored at -80°C.

3.8. Northern Blotting

Vector encapsidated RNA samples extracted from 1 mL of culture supernatant were purified as in section 3.5. and were analysed on ethidium bromide-stained 1% agarose gels in Tris-borate-EDTA (TBE) (Thermo Fisher Scientific) and electrophoresed at 60 V for at least 4 h. Samples were transferred from the gel to a Hybond N+ nylon membrane (GE healthcare) by capillary transfer for 2 h in UltraPure™ 20x SSC (Thermo Fisher Scientific). The membrane was briefly washed in 0.5x TBE, and crosslinked by baking at 80°C for 30 min. The membrane was then transferred to a 50 mL Falcon containing 4 mL pre-warmed UltraHyb™ (Ambion) and rotated at 68°C for 30 min. RNA probe (75 ng) was hybridised to the membrane in 1 mL of pre-warmed UltraHyb for 18 h at 68°C. The membrane was washed 2 x 5 min in low stringency wash buffer (75 µM NaCl, 8 µM Sodium Citrate and 0.1% SDS) at room temperature, and then 2 x 15 min in high stringency wash buffer (50 µM NaCl, 400 µM Sodium Citrate and 0.1% SDS) at 68°C. Bound probe was visualised using the DIG luminescent detection kit (Roche). Briefly, the membrane was washed for 5 min in 50 mL 1x washing solution and then blocked for 1 h in 50 mL blocking solution. The anti-DIG-AP fragments antibody was diluted 1:20,000 in 20 mL blocking solution and incubated with the membrane for 1 h. The membrane was washed twice for 15 min in 50 mL 1x washing solution and then incubated for 5 min in 20 mL detection buffer. RNA was visualised with the addition of 500 µL CDP-Star and detected with photographic film.

3.9. In-Gel SHAPE

3.9.1. *In Vitro* production of RNA

A 690 nucleotide PCR fragment including the sequences between TAR and the partial *gag* was amplified by PCR using primers Fw 5'-TAATACGACTCACTATAG

GGTCTCTCTGGTTAGACCAGATCTG-3', Rv 5'- CTTGCTGTGCGGTGGTCT-3'. Fw primer contains the T7 Polymerase promoter sequences to be used downstream for *in vitro* transcription. PCR reaction contained pCCL-eGFP template (100 ng), 1x Phusion HF Buffer, 12.5 μ M each primer, 10 mM dNTPs, 3% DMSO and 1 U HF Phusion DNA Polymerase (NEB). The thermal cycling conditions were: 98°C for 30 sec, 35 cycles of: 98°C for 10 sec, 64°C for 30 sec, 72°C for 1 min and a final extension step at 72°C for 10 min. The PCR fragment was electrophoresed on a 1% agarose gel, visualised by UV and extracted using the NEB Monarch^R extraction kit according to the accompanying protocol. The PCR product was eluted in a 15 μ L final volume. *In vitro* transcription was carried out using the MEGAScript T7 transcription kit (Ambion). Reactions contained extracted PCR product (14 μ L), 1x Transcription buffer, 10 mM each ATP, TTP, CTP and GTP, and 4 μ L of T7 Polymerase enzyme mix in a final volume of 40 μ L. The product was treated with 6 U Turbo DNase for 1h at 37°C. The RNA product was purified by lithium chloride precipitation, resuspended in RNase-free water and stored at -80°C.

3.9.2. In-gel chemical acylation of RNAs

45 μ g of RNA were resuspended in 160 μ L of Renaturation Buffer (10 mM Tris pH 8, 100 mM KCL and 0.1 mM EDTA) heated at 85°C for 5 min and slow-cooled to room temperature. 40 μ L of 5x Refolding Dimerisation Buffer were added to adjust the buffer to 40 mM Tris pH 8, 5 mM magnesium chloride (MgCl₂), 130 mM KCl and 50mM sodium cacodylate pH 7.5. The mixture was incubated at 37°C for 30 min. 60 μ L of Orange G native loading dye (1 x TBM (89 mM Tris Base, 89 mM Boric acid and 0.1 mM MgCl₂), 5% (w/v) glycerol, 0.1% (w/v) orange G dye) were added and 20 μ L of the treated RNA were loaded in each lane of a 20 cm x 18.5-cm 4% (w/v) Acrylamide/Bis 19:1 non denaturing gel prepared with 1x TBM. Gel electrophoresis was set at 90 V for 1h and 100 V for 4 additional hours. One fragment corresponding to the LR RNA ladder (Thermo Fisher Scientific) and one RNA sample were excised and stained for 10 min with 1.3 μ M ethidium bromide in 1x TBM and visualized under UV to detect the position of the dimeric and monomeric RNA. The ethidium bromide-stained gel fragments were aligned to the remaining of the unstained gel. The gel parts corresponding to the dimeric and monomeric RNA were excised using a scalpel. Both the dimeric and monomeric gel pieces were divided into two equal parts, with one fragment

incubated in 1× TBM containing 10% (v/v) dimethyl sulphoxide, and the other one in 1× TBM containing 10% (v/v) dimethyl sulphoxide and 10 mM *N*-methylisatoic anhydride. The 4 gel pieces were placed at RT for 1 h and were then washed three times in 1× TAE (40 mM Tris-acetate and 1 mM EDTA pH 8.3). The cleaned gel pieces were then minced and the RNA was electroeluted using the BIO-RAD model 422 Electro-Eluter at 40mA for 1h 30min. RNA was precipitated with 300 mM sodium acetate and 2.5 volume of ethanol and recovered by centrifugation, washed with 70% ethanol and resuspended in 10 µl of water. RNA concentration was determined by spectrophotometry.

3.9.3. Reverse Transcription and Structural analysis

1000 ng of each RNA sample were resuspended in 12 µl of 2.1 mM Tris pH 8.0, 42 µM EDTA. 5 nmol of 6FAMTM-labelled primers described in section 3.9.1 (Applied Biosystems) were added to the dimeric and monomeric NMIA-treated samples and 5 nmol of VIC[®]-labelled primer (Applied Biosystems) were added to the dimeric and monomeric Negative control samples. Primers were annealed to the RNA at 85°C for 1 min, 60°C for 5 min and 35°C for 5 min. 8 µL of the Reverse transcription mix (100 U of Superscript III reverse transcriptase (RT), 1.5× SSIII RT buffer, 12.5 mM dithiothreitol (DTT), 5M Betaine, 1.25 mM Deoxycytidine triphosphate (dCTP), deoxyadenosine triphosphate (dATP), deoxyuridine triphosphate (dUTP) and 7-deaza-deoxyguanosine triphosphate (dGTP)) were added to each sample and were incubated at 55°C for 50 min. RNA was degraded with 200 mM NaOH at 95°C for 5 min and ice cooling for 5 min. The samples were then treated with 200 mM HCl and were precipitated as described above. The cDNA produced was combined with two sequencing ladders. Sequencing ladders were produced using the Thermo Sequenase Cycle Sequencing Kit (Applied Biosystems) and primers of the same sequence as before, but which were now labelled with NEDTM and PET[®] (Applied Biosystems). Each cycle sequencing reaction contained 300 ng of DNA template, 1.3 nmol NEDTM- or PET[®]-labelled primer, sequencing buffer, ddNTP of choice and enzyme as per manufacturer's instructions. The thermal cycling conditions were: 95°C for 2 min, 40 cycles of: 95°C for 30 sec, 50°C for 30 sec, 72°C for 1 min and a final extension step at 72°C for 10 min. Sequencing ladders were precipitated with 300 mM sodium acetate and 2.5 volume of ethanol and recovered by centrifugation, washed with 70% ethanol and resuspended in 30 µl of water. RNA samples were mixed with the

sequencing ladders in a 1:1:1 concentration ratio and were submitted for sequencing. SHAPEFinder software⁵⁵ was used to align the peaks to the nucleotides of the sequence of interest. The 4 labelled primers were used for a control experiment to ensure mobility shift normalisation by running a sequencing reaction using each primer with the same ddNTP, ddATP. Reactivity was then calculated by subtracting the negative control peaks to the NMIA treated peaks. Single level nucleotide structural modelling was performed using the RNAstructure software²³⁵ where RNA secondary structure is predicted using SHAPE data and free energy constraints. Structures were drawn using XRNA²³⁶.

3.10. Western Blotting of viral supernatants

Vector supernatant (8 mL) was centrifuged for 10 min at 3000 x *g* to remove cell debris. Purified viral supernatant (1 mL) was subsequently centrifuged at 22,000 x *g* for 2.5 h at 4°C. The pellet was resuspended in 20 µL of PBS and the proteins were denatured by heating to 95°C in 1x Laemmli sample loading buffer (Sigma-Aldrich). Samples (1 µL) were separated in a 0.75 mm 15% SDS-PAGE gel, at 100 V for 4 h. Proteins were transferred to a nitrocellulose membrane by electroblotting at 100 V for 90 min in 1x transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol (v/v) and 0.1% SDS, before blocking in 5% (w/v) skimmed milk powder in PBS for 30 min. The primary antibody ARP313 (NIBSC) for p24 was diluted 1:20,000 in 5 mL 5% milk in PBS. The membrane was incubated with the antibody overnight at 4°C, followed by 3 x 10 min washes with PBST (0.1% Tween in PBS). The membrane was incubated with secondary antibody (horse anti-mouse HRP, Abcam) at 1:2,000 for 2 h. The membrane was washed as previously and bound secondary antibody was visualised by using 1 mL ECLTM prime Western blotting detection reagent (GE healthcare) and by exposing to photographic film. Films were scanned and bands quantified using ImageJ²³⁷.

3.11. Cell lysate preparation and Western Blotting

Supernatant was removed 48 h post transfection and cells were harvested by adding 1mL ice cold PBS directly into half of the surface of a 10 cm² plate. Harvested cells were transferred to a 1.5 mL microcentrifuge tube and were centrifuged at 6,000 x *g* for 5 min. Pellets were washed once in ice-cold PBS, re-centrifuged at 6,000 x *g* for 5 min and stored at

-80°C for later analysis. Cells were lysed in 250 µL RIPA Lysis and Extraction Buffer (Thermo Fisher Scientific) containing 1x HALT protease inhibitor cocktail (Thermo Fisher Scientific) by rocking on ice for 15 min. Lysates were clarified by centrifugation at 6,000 x g for 5 min at 4°C. Protein quantification was performed using the BCA Protein assay kit (Pierce) according to the manufacturer's instructions. Extracted cell lysates were heated in 1x Laemmli sample loading buffer (Sigma-Aldrich) at 95°C for 10 min. 20 µg protein per sample were analysed by SDS-PAGE as described above. GAPDH, used as a loading control, was detected with the use of anti-GAPDH (1:10,000 Abcam) and donkey anti-rabbit antibody (1:2,000 Abcam).

3.12. ELISA p24 assay

High protein binding 96 well plates were coated overnight with 25 µL per well of 10 µg/mL anti-HIV-1-p24 D7320 coating antibody. Plates were blocked for 1 h with 5% bovine serum albumin in 1x Tris-Buffered Saline (TBS) (100 µL per well). Samples were diluted in 1x TBS 0.05% Empigen (Sigma-Aldrich) and a serial dilution of standard was prepared using the HIV-1-p24 antigen AG6054 (Aalto Bio Reagents). Blocking solution was removed by washing plates four times in 1x TBS on a plate washer, before the plate was loaded with 25 µL of sample and standard. Plates were incubated in a shaking mode for 90 min and washed as above. 9.25 µL of the mouse monoclonal anti-HIV-1 p24 BC 1071 (alkaline phosphatase conjugated, Aalto Bio Reagents) solution was added in each well and the plate was incubated for 1 h on a plate shaker. The plate was then washed manually four times with PBS containing 0.1% Tween-20. 25 µL Lumiphos Plus reagent (Lumigen) were added to each well and the plate was incubated in the dark for 30 min. Luminescence levels were measured with the Glomax luminometer and p24 levels interpolated from the linear portion of the standard curve.

3.13. ELLA p24 assay

ELLA is the branded name for next generation automated ELISA developed by proteinsimple²³⁸. The objective of this analytical method was the measurement of the intracellular and extracellular HIV-1 p24 levels to measure Gag expression and Gag budding using the ELLA automated microfluidic GNRs (Glass Nano Reactors) based cartridge system. ELLA is a fully automated ELISA that quantifies p24 protein in triplicate for each sample well and all results are determined from the factory-calibrated internal standard curve.

Samples were thawed at RT for 30 min and 500x dilutions were performed using the provided 0.5% Triton X Lysis buffer. Meanwhile, the ELLA Instrument was allowed to perform the verification test before testing samples. 1 mL of the Wash Buffer was loaded in each of the designated wells and 50 μ L of the diluted samples were loaded on the ELLA cartridge. The run lasted 90 min and the software automatically generated the final p24 results in pg/mL based on the raw data and the inbuilt standard curve.

4. Rational design and biochemical characterisation of transfer vector mutants

Despite the tremendous potential of gene therapy the main target cells are poorly permissive, requiring multiple rounds of high vector doses and prolonged *ex vivo* culture to reach adequate transduction levels^{215 216}. Here, I aimed to address this challenge by focusing on improving the lentiviral vector efficiency.

The HIV-1 gRNA can act as a template for translation but can also be packaged without having to be translated first^{121 122}. Comparison studies of gRNA sorting mechanisms revealed that unlike MuLV gRNA, HIV-1 gRNA also has the potential to first be translated and then packaged¹²¹. As mentioned in section 2.1.11, the gRNA of HIV-1 is known to act as a switch between the monomeric conformation that is associated with translation and the dimeric conformation linked with packaging. While dimerisation positively regulates packaging^{108 109}, translation of the genome is in opposition to its encapsidation. Based on this observation I aimed to create transfer vector mutants that were more likely to adopt the dimeric conformation and therefore be packaged. I targeted regions in the 5'UTR that play important roles in the process of dimerisation including the DSL, the U5-AUG duplex and the polyA stem loop.

The effects of these mutations were assessed by measuring infectious titres with flow cytometry and physical titres with ELISA p24 and ELLA p24. Furthermore, Western Blotting and ELLA p24 were performed to investigate any potential effects of these 5'UTR mutations on *in trans*-Gag processing and budding.

4.1. Verification of integrity of 3rd generation lentiviral plasmids

pSYNGP packaging plasmid, pVSVg envelope plasmid, pRev and pCCL-eGFP transfer vector plasmid, obtained from GlaxoSmithKline, were used to transform One Shot Stbl3TM chemically competent *E.coli* cells. Test digestions were performed to verify the integrity of the isolated plasmids. The expected band sizes for the *EcoRI* digestion of pSYNGP were 4.3 kb, 4.1 kb and 401 bp, of pVSVg 4.1 kb and 1.6 kb and of pRev 3.6 kb and 311 bp. Expected

fragment sizes of the pCCL-eGFP plasmid's digestion with *Hind*III were 3.3 kb, 2 kb, 584 bp, 556 bp, 499 bp and 313 bp. Uncleaved controls of the plasmids were also loaded to verify the correct plasmid size. Figure 10 shows the plasmid maps with their restriction sites and expected band sizes, while Figure 11 confirms the correct identity of all plasmids comprising the 3rd generation lentiviral vector system.

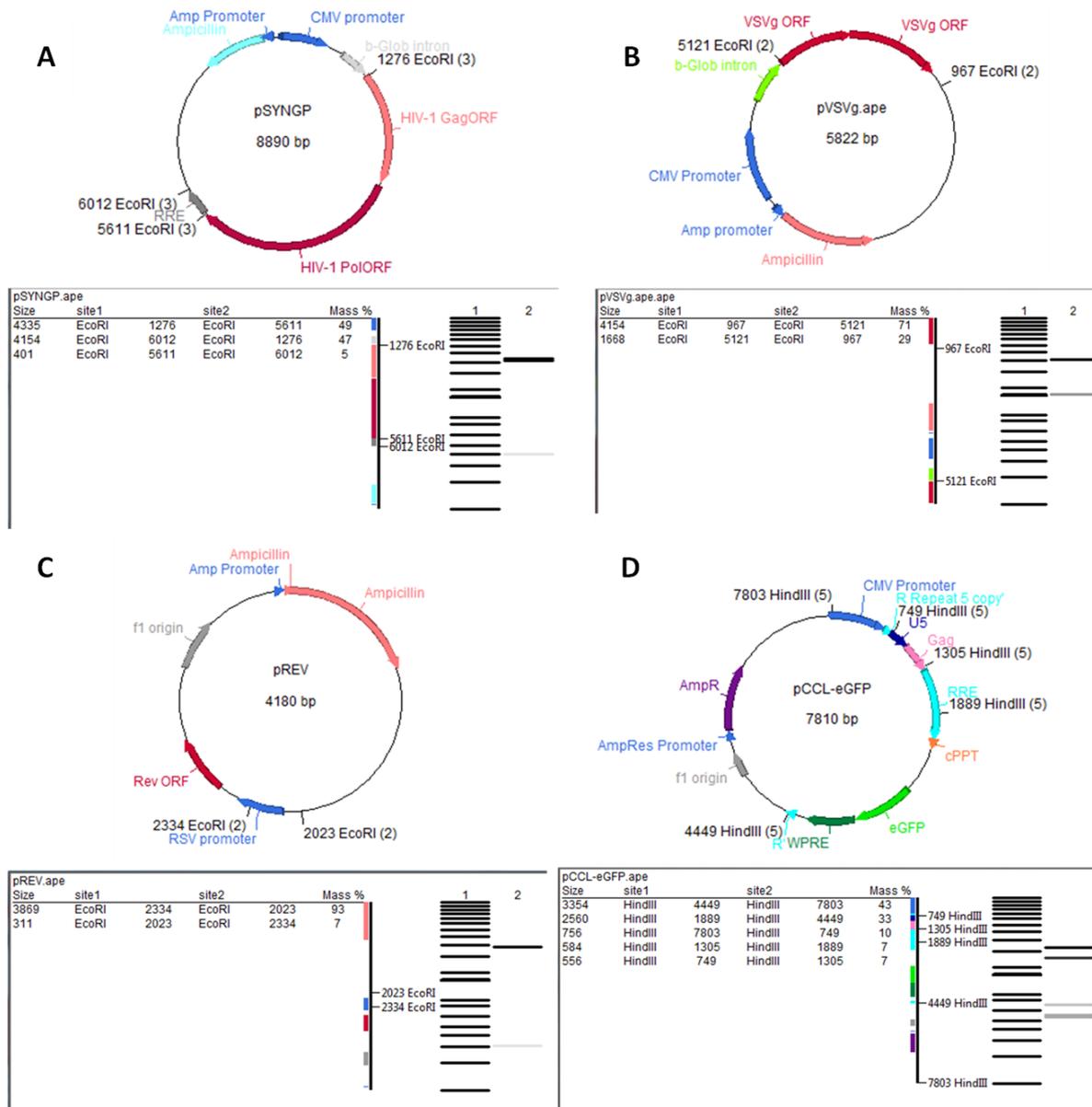


Figure 10. Plasmid maps of constructs comprising the 3rd generation lentiviral vector system.

[A] *pSYNGP* plasmid map with its *EcoRI* restriction sites. *EcoRI* digestion of *pSYNGP* produces three bands of 4.3 kb, 4.1 kb and 401 bp. **[B]** *pVSVg* incubation with *EcoRI* is expected to create 2 fragments of 4.1 kb and 1.6 kb length. **[C]** *pRev* plasmid map accompanied with its *EcoRI* restriction sites. The size of the expected bands are 3.6 kb and 311 bp. **[D]** Plasmid map of the *pCCL-eGFP* transfer vector and its *HindIII* digestion sites. Expected fragment sizes of the *pCCL-eGFP* digestion with *HindIII* are 3.3 kb, 2 kb, 584 bp, 556 bp, 499 bp and 313 bp.

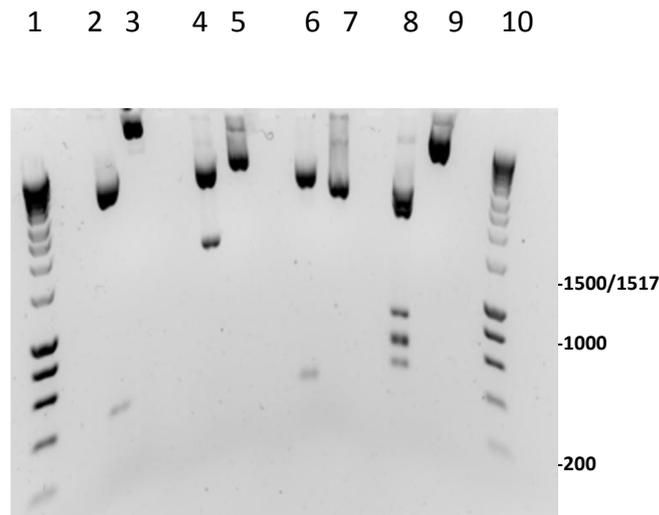


Figure 11. Agarose gel electrophoresis of 3rd generation lentiviral plasmids.

1% agarose gel electrophoresed in TBE and stained with SYBR Safe. *pSYNGP* (lane 2), *pVSVg* (lane 4) and *pRev* (lane 6) were digested with *EcoRI* and *pCCL-eGFP* (lane 8) with *HindIII*. Lanes 3, 5, 7 and 9 represent undigested *pSYNGP*, *pVSVg*, *pRev* and *pCCL-eGFP* respectively. The fragment size of digested plasmids was estimated via the use of HyperLadder 1Kb (Bioline) (lanes 1 and 10).

4.2 Transfection optimisation

In order to study the packaging efficiency of the various transfer vector constructs, HEK293T cells were co-transfected with the plasmids described in section 4.1. Initially, a transfection protocol for lentiviral vector production was established, as described in the following subsections 4.2.1 and 4.2.2.

4.2.1. Calculation of DNA mass

Dull *et al.*,¹⁶⁴ described a 3rd generation lentiviral vector system that consists of 4 plasmids including a packaging construct, a Rev expressing construct, an Envelope construct and a eGFP containing transfer vector construct. The number of template copies was calculated using the plasmid DNA mass and length based on the assumption that the average weight of a base pair (bp) is 650 Daltons. The formula used was: number of copies = (amount * 6.022x10²³) / (length * 1x10⁹ * 650)²³⁹. The copy number required for each plasmid was calculated in table 5, using the aforementioned equation and the details provided from the Dull *et al.*, paper. Maintaining the molar ratio 1:1:1:3 that came out of Dull's research, table 6 explains the way that the final DNA mass of each plasmid was calculated for transfections.

| Plasmid | µg per 10cm dish | plasmid length (bp) | Copy number |
|---------|------------------|---------------------|-------------|
| p3Rev | 2.5 | 5118 | 4.53E+11 |
| p3VSVg | 3.5 | 7364 | 4.40E+11 |
| p3GP | 5 | 9406 | 4.92E+11 |
| p3TF | 14 | 8504 | 1.5.E+12 |

Table 5. Calculation of plasmid copy number for transfections.

The amounts of plasmid used per 10cm dish come from the research scale vector production protocol established at GlaxoSmithKline which was shared via personal communication. Plasmids presented here were developed by the ATD team at GlaxoSmithKline.

| Plasmid | plasmid length (bp) | Copy number | Calculated plasmid amount (μg) |
|-----------|---------------------|-------------|---|
| pRev | 4100 | 4.5.E+11 | 2.0 |
| pVSVg | 5800 | 4.4.E+11 | 2.8 |
| pSYNGP | 8900 | 4.9.E+11 | 4.7 |
| pCCL eGFP | 8000 | 1.5.E+12 | 13.2 |

Table 6. Calculation of plasmid DNA mass for transfections.

pRev was ordered from Addgene²⁴⁰, pSYNGP and a newly synthesised pCCL eGFP transfer vector were obtained from GlaxoSmithKline, while pVSVg was part of the plasmid repertoire of the Lever group. The plasmids presented here were the ones consistently used in this thesis. The plasmid mass ratio used for high efficiency transfections was 1:1:1:3.

4.2.2. Calculation of transfection and transduction efficiencies

293T cells were transfected with the transfer vector and packaging plasmids to produce lentiviral vectors. 12 h post transfection cells were viewed under light and fluorescent microscopy.

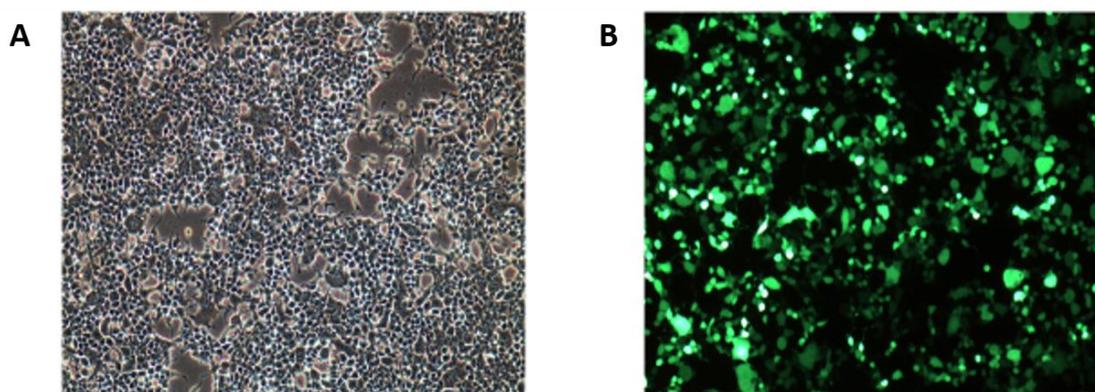


Figure 12. Measurement of Transfection efficiency with fluorescent microscopy.

[A] The cell monolayer viewed by light microscopy with a 10X magnification. **[B]** Image taken by fluorescence microscopy with a 20X magnification in order to estimate the transfection efficiency of the plasmid combination used.

Using table 6 as a guide, a 1:1:1:2 and a 1:1:1:3 mass ratio of the four plasmids was used for transfections. Additional combinations of plasmid mass ratios were tried, including the plasmid concentrations from the Dull *et al.*, 1998 paper and the concentrations used at the GSK research scale vector production protocol (Dr Vink, personal communication). Furthermore, two different CaCl₂ molarities were tried with all the aforementioned plasmid combinations. A final 125 mM of CaCl₂ gave higher transfection efficiencies compared to 150 mM CaCl₂, and was thus chosen for the final protocol. The harvested supernatant was diluted 10³ times and was used for 293T cells transductions. Vector titre was calculated as the seeded number of cells transduced x %eGFP+ cells x dilution factor / volume. The 1:1:1:3 plasmid ratio generated the highest titre vectors of 2.4x10⁷ TU/mL presented in Figure 13.E.

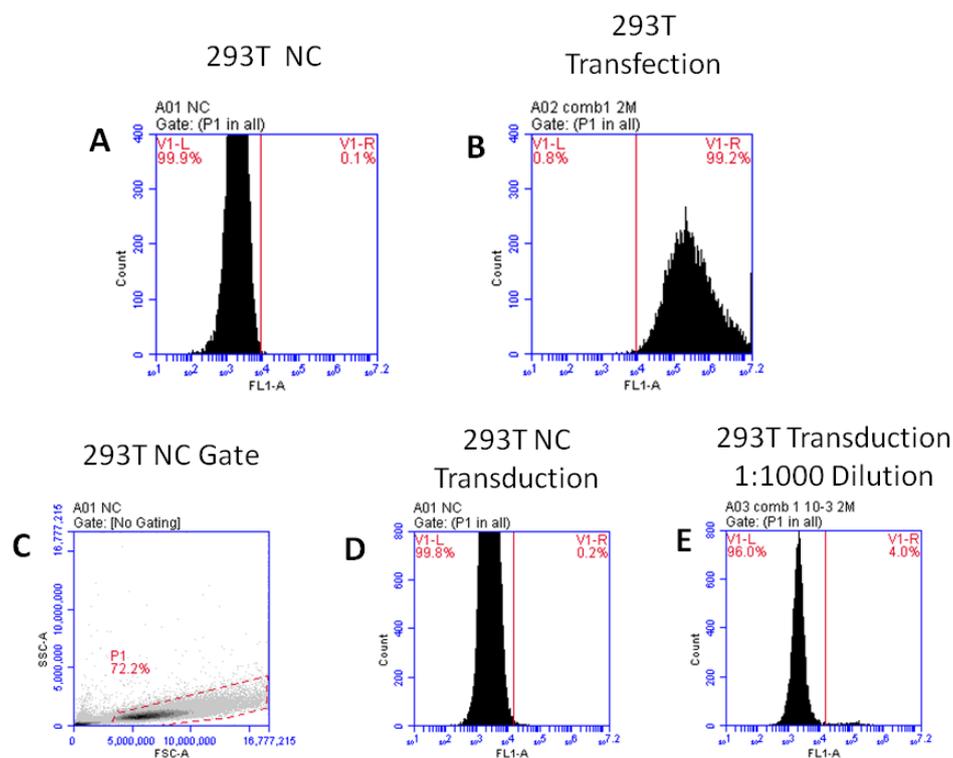


Figure 13. Measurement of transduction efficiency.

Flow cytometry performed on cells harvested 42 h post transfection; the transfection and transduction efficiencies were calculated using the eGFP reporter gene included in the transfer vector. All cellular populations were lysed in TrypLE and harvested in PBS. [A] Non transfected 293T cells were used to create a histogram showing their distribution using the

FL1 parameter. **[B]** Flow cytometry graph of cells described in Figure 13.B, with a transfection efficiency of 99.2%. **[C,D]** Non transduced 293T cells showing gate associated with live cells (C) and [D] threshold to distinguish the transduced from the non-transduced cells. **[E]**. Flow cytometry graph showing the percentage of cells successfully transduced with the use of one in a thousand diluted viral supernatant. Flow cytometry set up, including gating of live cells and FL1 bar allocation, was performed as explained in section 3.2.

4.3. Design and cloning of U5-AUG mutants

The rationale behind the construct design was the introduction of mutations that would either strengthen or weaken the U5-AUG base pairing. As discussed in the introduction chapter, it has been shown that a stabilised U5-AUG interaction is associated with increased dimerisation, which is associated with efficient packaging. The structures of the monomeric and dimeric RNAs are proposed to differ significantly. In gel probing of RNA conformers has revealed a dimerisation structural switch in the HIV-1 leader⁸.

Initially the design of the mutant transfer vectors was based on computer assisted structural prediction with mFold²⁴¹. mFold structures were based on calculations of minimal free energy. Apart from the structural stability of the RNA, other parameters had to be taken into consideration including the pseudo-knot structure. SHAPE data as well as NMR studies⁶¹ suggest the formation of a pseudo-knot via the base-pairing between DIS and U5, which cannot be predicted with mFold. A pseudoknot is a secondary structure in which half of one stem is intercalated between the two halves of another stem. Because of the overlapping nature of the base pairing, mFold is unable to predict the pseudoknot, but will instead predict the more stable of the two stems. Taking these two criteria into consideration, the following 3 mutants were designed in order to test whether increased stability of the U5-AUG motif will promote dimerisation and hence packaging.

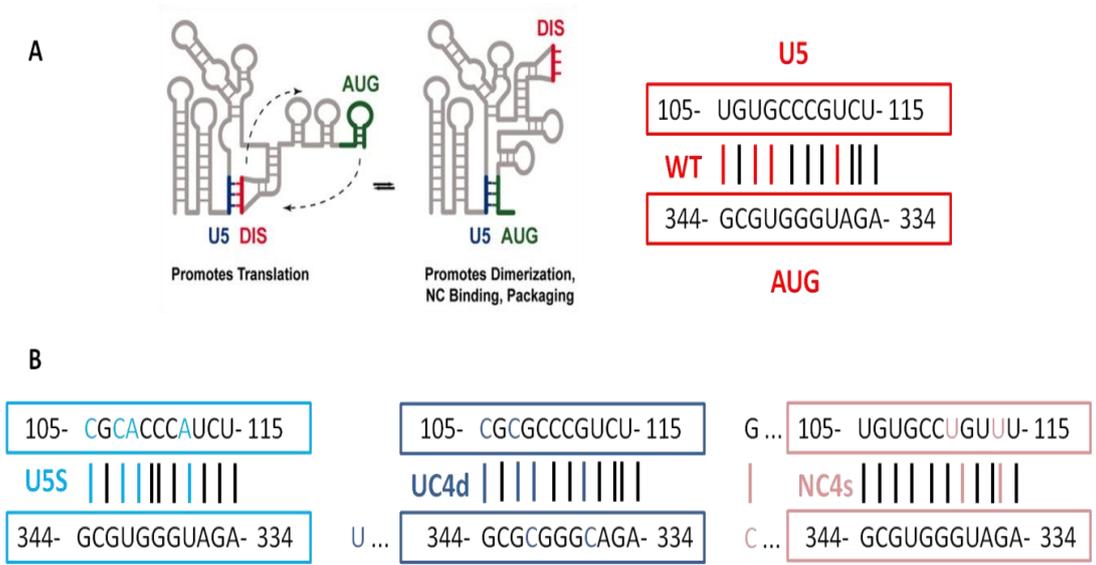


Figure 14. Design of mutants with either strengthened or impaired U5-AUG base pairing.

[A] The U5-AUG duplex is an important structural motif of dimeric RNA. It is formed via the base pairing of 4 U-Gs, 2 U-As and 5 G-Cs. Adapted from:⁷ **[B]** The **U5s** mutant was created by introducing mutations in the U5 region that strengthened the U5-AUG base pairing. 4 U-G non-canonical pairs were converted into two A-U and two C-G pairs, increasing the number of hydrogen bonds by 6. **UC4d** mutant included the conversion of four U-G pairs to C-G pairs, increasing the number of hydrogen bonds by 8, and thus the stability of the U5-AUG base pairing. Also, this mutant included a mutation that destabilised SL-4 by the introduction of a mutation that converted a G-C pair to U-C. **NC4s** included the substitution of two A-U pairs by A-G that disrupted canonical base pairing in the U5-AUG and the introduction of a G-C base pair that strengthened SL4.

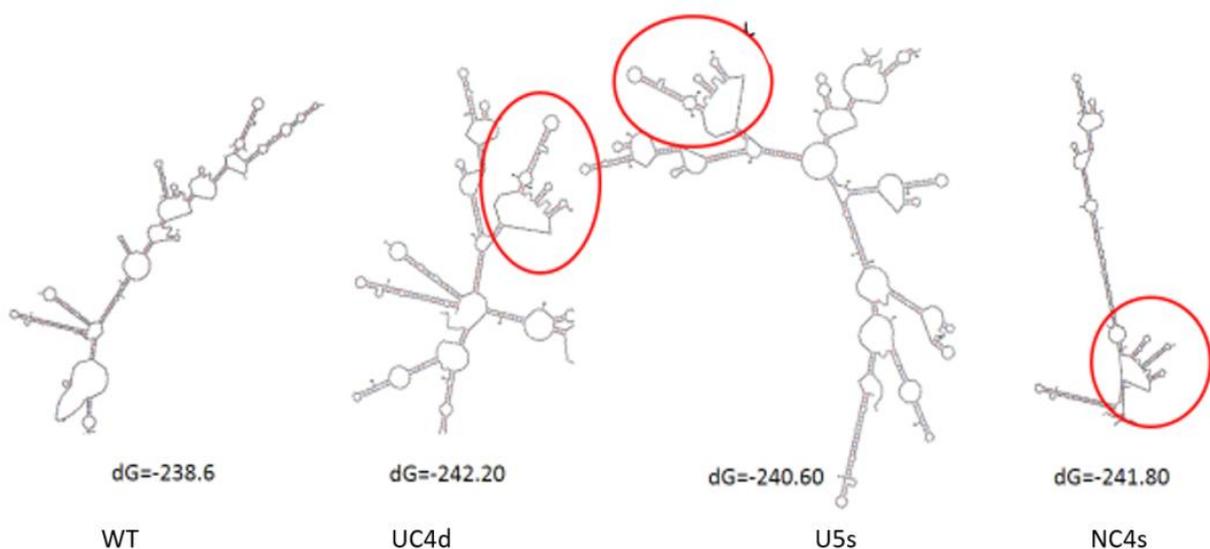


Figure 15. mFold predicted secondary structures of WT and mutants.

The first 800 bp of the WT and mutated pCCL-eGFP constructs, that include the HIV-1 leader sequences, were inputted into mFold using the default parameters of the RNA Folding form in order to predict their secondary structures based on minimum free energy. mFold indicates that the global structure of the main packaging signal region of the HIV-1 leader (SL1-3), shown in the red circles, was predicted to be maintained intact upon introduction of these mutations (Figure 15). **[UC4d]**: 2 U to Gs and 2 C to Gs conversions, expected to lead to U5-AUG duplex stabilisation and polyA formation. One of these mutations was designed to destroy the start codon of gag; furthermore this mutant included a destabilising mutation in the SL-4 loop. **[U5s]**: U to G, 2 C to G and 2 A to U conversions aimed to lead to U5-AUG duplex stabilisation but destruction of the structure containing the polyA signal. **[NC4s]**: Mutations in this construct intended to destabilise the U5-AUG base pairing and stabilise the SL4 loop.

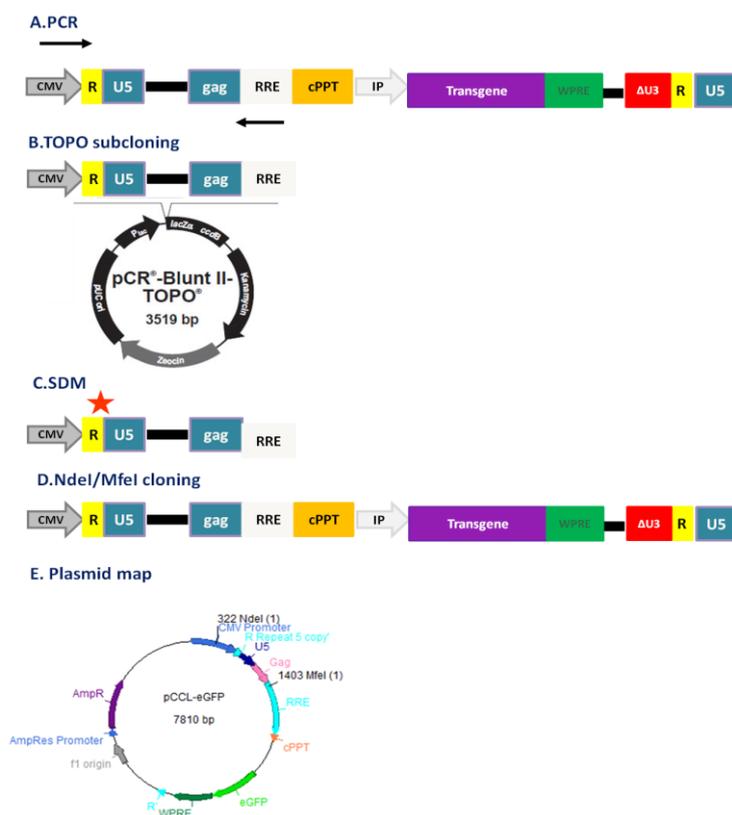


Figure 16. Intermediate steps of the cloning process.

[A] PCR amplification of a DNA fragment spanning the CMV and RRE sequences **[B]** Blunt TOPO subcloning of the amplified DNA fragment shown in A, that includes the *NdeI* and *MfeI* restriction sites **[C]** Site directed mutagenesis reactions to incorporate mutations of interest only in the 5'LTR introduced in the Blunt-TOPO vector **[D]** Digestion of mutated TOPO vector and WT pCCL-eGFP vector with *NdeI* and *MfeI*-HF and ligation of the mutated 5'LTR digested fragment with the pCCL-eGFP backbone to produce rationally designed mutated lentiviral vectors. **[E]** Plasmid map of pCCL-eGFP vector depicting the *NdeI* and *MfeI* restriction sites.

Results from this cloning strategy are illustrated in Figure 17. The pCCL-eGFP vector was used as a template for PCR using a gradient temperature and various high fidelity polymerases. Generation of the correct size products was confirmed by electrophoresing 5 µl of each PCR product on a 1% agarose gel (shown in Figure 17.A). In particular; lanes 1-7 represent the Accuprime™ *Pfx* PCR products using a gradient temperature from 60-65°C. Bands 4 and 5 were excised and purified for use in downstream reactions. Lane 8 includes the Q5 Polymerase PCR product. The rest of the lanes, which are blank, represent the unsuccessful 60-64 °C gradient PCRs with the use of the Trustart Hot Start Polymerase^R.

Both the gel extracted and non- gel extracted PCR products amplified with Accuprime™ *Pfx* and Q5 high fidelity^R polymerases were used for blunt-end TOPO-subcloning. 8 colonies were selected and their extracted DNA was used for test digestion with *BamHI*-hf and *NdeI* (shown in Figure 17.B). The expected sizes of a successful TOPO-subcloning reaction are 3265, 2568 and 512 bp, which were visualised in lanes 2 to 5 and 9 to 12.

Upon confirmation of the successful TOPO-cloning reaction, purified DNA from colony 4 (Figure 17.B lane 5) was selected and used as a template for site directed mutagenesis (SDM) reactions. The 3 SDM products were used for bacterial transformation. Sequentially the extracted and purified plasmid DNA was digested with *NdeI* and *BamHI*-hf to identify plasmids that contained the correct insert (shown in Figure 17.C). Depending on the direction in which the inserts were ligated the outcome of the test digestion would be different. If the insert ligated in the forward direction, flanked between *NdeI* and *BamHI*-hf, then the products would be of sizes: 3695, 2568 and 81 bp, whereas if the insert was ligated

in the reverse direction, flanked between *Bam*HI-hf and *Nde*I, then the produced fragments would be of the sizes: 3520, 2566 and 256 bp.

Generation of the UC4d and NC4s mutants required another round of SDM. Incorporation of the mutations of interest was confirmed by sequencing. Upon confirmation of appropriate mutation introduction, the transfer vector pCCL-eGFP backbone and the mutated TOPO-vectors were digested with *Nde*I and *Mfe*I-HF. The digestion products were separated and the bands of interest were excised and purified to be used in the final step that included the ligation of the TOPO-insert in the pCCL-eGFP backbone. *Hind*III test digestions were performed to verify the completion of cloning (shown in Figure 17.D). All 3 successfully cloned mutants gave rise to the expected size fragments which were 3.3 kb, 2 kb, 584 bp, 556 bp, 499 bp and 313 bp long. The discrepancy between the expected number of bands and the number of bands on Figure 17.D was due to the inability to differentiate a 584 bp band from a 556 bp band on a 1% agarose gel.

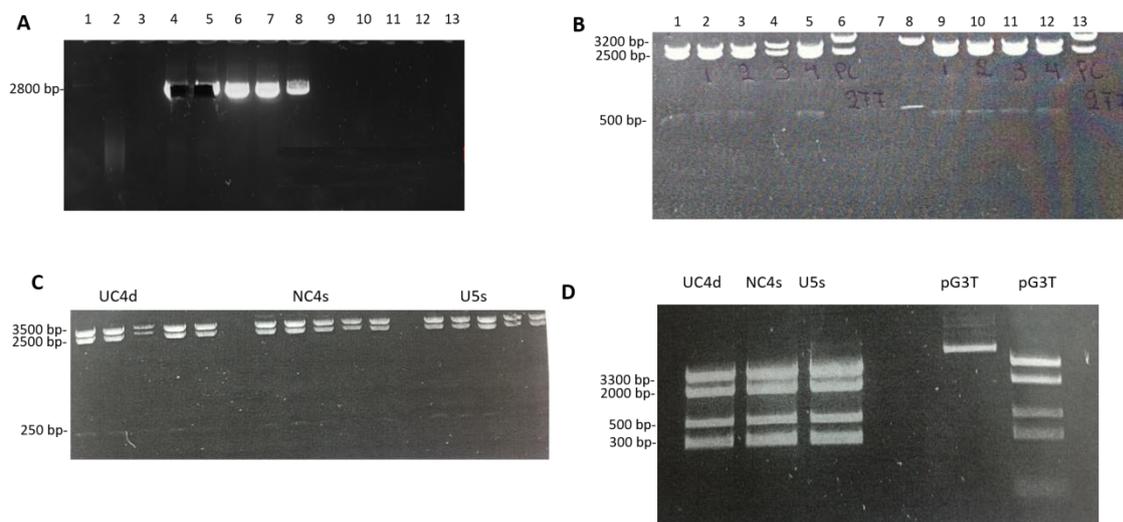


Figure 17. Agarose gel analysis of intermediates during the cloning process.

[A] 1% agarose gel showing PCR products from a temperature gradient PCR of pCCL-eGFP using primers Fw-Subcloning and Rv-Subcloning. Lanes 1-7 show PCR product formed using Accuprime-Pfx mix at 60°C (lane 1), 61°C (lane 2), 61.9°C (lane 3), 63°C (lane 4), 64°C (lane 5), 64.6°C (lane 6), and 64.8°C (lane 7). Lane 8 shows product formed with Q5 high fidelity Polymerase at 64°C. **[B]** Plasmids purified from 8 clones (lanes 2 to 5 and 9 to 12) and digested with *Bam*HI-hf and *Nde*I. **[C]** *Nde*I and *Bam*HI-hf digestion of clones following site directed mutagenesis to produce UC4d (lanes 1-5, NC4s (lanes 7-11), U5s (lanes 13-17). **[D]**

HindIII digested plasmids purified from clones resulting from a second round of site-directed mutagenesis for UC4d (lane 1) and NC4s (lane 2). Products of U5s digestion with *HindIII* are presented in lane 3.

4.4. Cloning of S1 (Abbink *et al.*, 2003) and LU5-AUG (Lu *et al.*, 2011) constructs

Abbink *et al.*, and Lu *et al.*, described the presence and importance of the U5-AUG duplex and its role in promoting a structure that favours dimerisation (Figure 18). These two papers included structural and biochemical characterisation of the U5-AUG duplex from the *in vitro* transcribed RNA of the HIV-1 leader sequence. The S1 and LU5AUG mutations, described in the aforementioned papers (Figure 18.A and 18.C), were introduced into the pCCL-eGFP transfer vector construct as controls to test their behaviour in a *in virio* context.

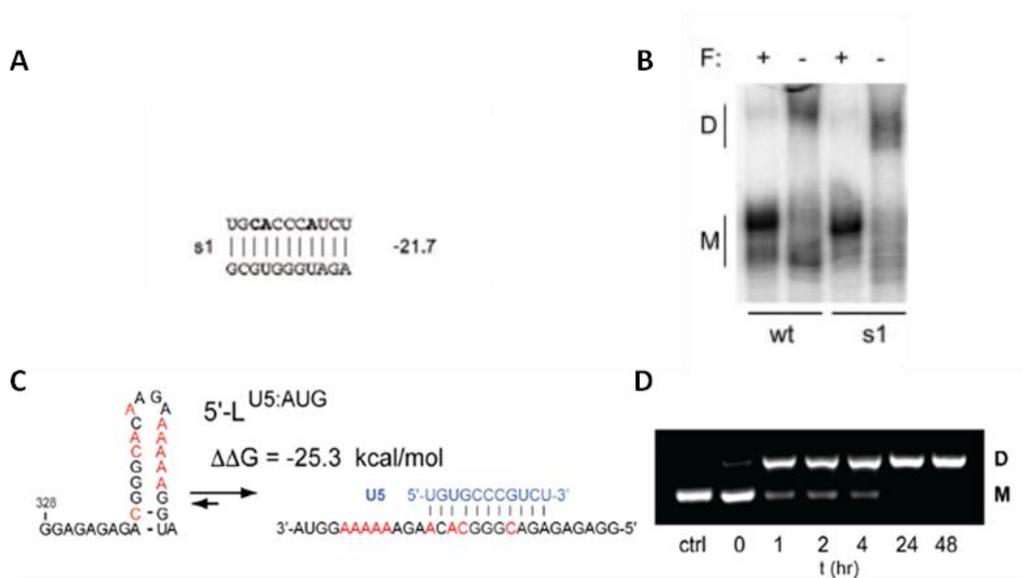


Figure 18. Design and *in vitro* dimerisation propensity of S1 and LU5AUG mutants.

[A] Rational design of S1 mutant. **[B]** Dimerisation propensity of S1 mutant against WT in the presence and absence of formamide. Taken from Abbink *et al.* **[C]** Rational design and free energy details of LU5AUG mutant. Mutated nucleotides are shown in red. The left part of this figure shows that SL4 can no longer form, while the right figure shows the newer, stronger U5-AUG interaction. **[D]** Time dependent dimerisation assay of LU5AUG mutant. Taken from Lu *et al.*

The dimerisation properties of the transfer vector RNA expressed in a cellular context were compared to those of the *in vitro* transcribed RNA of the HIV-1 leader sequence. These mutants were cloned to study the effect of sequence size and environmental conditions on dimerisation and hence packaging. The rationale behind the creation of these U5-AUG mutants from the literature in the context of the pCCL-eGFP transfer vector system was to confirm the results described in these studies and use these and other similar mutations to improve packaging efficiency of lentiviral vectors, or identify potential parameters that could be affecting the dimerisation properties of the HIV-1 gRNA in an *in virio* context.

The intermediate steps of the creation of S1 and LU5AUG transfer vector mutants are presented in Figure 19. The cloning strategy followed was exactly as presented in section 4.3.

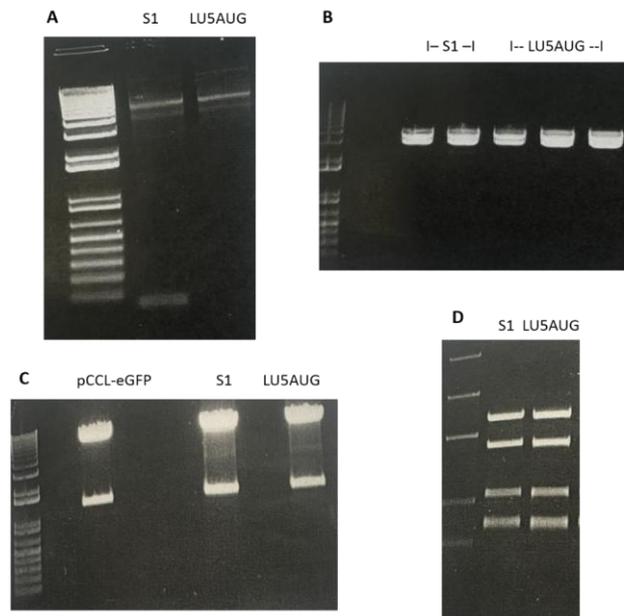


Figure 19. Stages of cloning for S1 and LU5AUG transfer vector mutants.

[A] 1% agarose gel showing PCR products from the site SDM reactions for S1 and LU5AUG mutants **[B]** EcoRI digestion of clones to investigate the results of TOPO-subcloning. Confirmation of S1 and LU5AUG insertion in the Blunt-TOPO® vector **[C]** WT pCCL-eGFP backbone and S1 and LU5AUG plasmids purified and digested with MfeI-hf and NdeI to be used for downstream isolation, extraction and ligation. **[D]** HindIII digested plasmids purified from final clones for both U5-AUG mutants.

4.5. Rational design and cloning of dimerisation mutants

The idea behind this third round of construct design was the introduction of mutations in other regions of the HIV-1 leader that have been documented to affect dimerisation. Recent findings suggest that destabilisation of the polyA motif increases the ratio of monomeric to dimeric RNA¹²⁷. Based on these findings I was interested in investigating whether stabilisation of the lower part of the polyA stem loop could lead to increased dimerisation. That led to the creation of two polyA transfer vector mutants: polyAsw and polyAs. J9+Gag was designed based on a study conducted by Van Bel *et al.*, which showed that DIS extension could increase *in vitro* dimerisation²⁴². Lastly, the potential of an alternative dimerisation palindrome sequence was investigated by creating mutant SIVS1 that contained the GUGCAC dimerisation palindromic sequence of SIV-1^{83 266}.

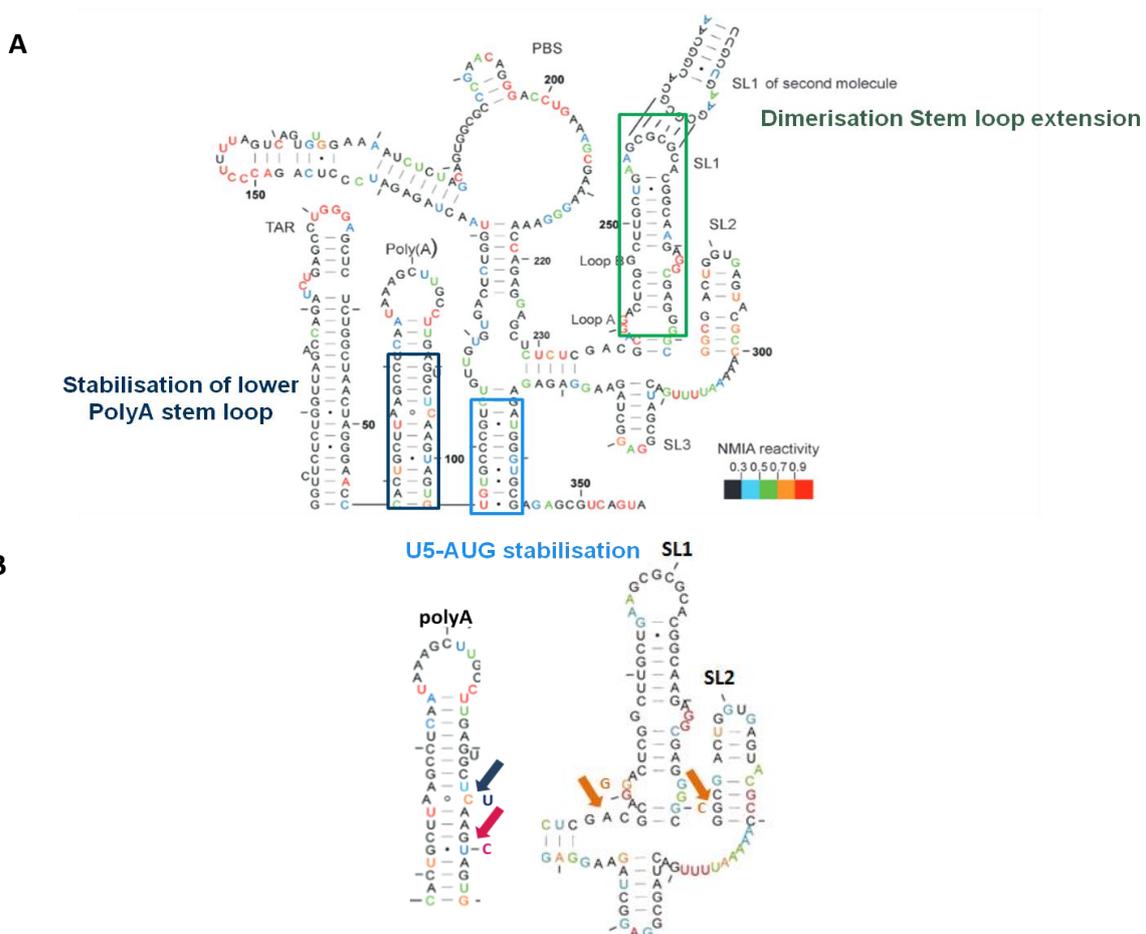


Figure 20. Depiction of dimerisation mutant design targeting the polyA, U5-AUG and SL-1 structural motifs. Adapted from source:⁸

[A] Indication of all targeted regions of the HIV-1 leader for generation of dimerisation promoting transfer vector mutants. **[polyAs]:** polyA stabilisation by conversion of the non canonical A-C pair, which leads to the formation of an RNA bulge in the lower part of the polyA stem, to an A-U Watson-Crick base pair. **[J9+Gag]:** Introduction of two point mutations resulted in two extra G-C pairs in the lower part of SL-1 achieving extension of the dimerisation stem loop, while maintaining the reported Gag binding sites⁴⁶. **[PolyAsw]:** A single point mutation that converts a G-U pair to a G-C pair was designed to stabilise the lower polyA stem loop. **[SIVS1]:** A mutant transfer vector that contained the dimerisation palindrome sequence of SIV-1 GUGCAC. **[B]** Closer look at the targeted nucleotides of the polyA and SL1 stem loops. Arrows indicate the exact position of the introduced point mutations. In further detail, mutated nucleotides for the generation of polyAw, polyAs and J9+Gag are presented in pink, dark blue and orange respectively.

Results from this cloning strategy are illustrated in Figure 21. The intermediate TOPO-LTR plasmid generated in 4.3. was used as a template for SDM reactions. The 4 SDM products were used for bacterial transformation; sequentially the extracted and purified plasmid DNA was digested with *EcoRI* to identify plasmids that contain the correct size insert (shown in Figure 21.C). *EcoRI* digestion of the TOPO vector containing a modified insert would produce fragments of the sizes: 3500bp (TOPO vector) and 2800bp (insert). Both intermediate TOPO vector plasmids containing the modified inserts and the pCCL-eGFP plasmid were digested with *NdeI* and *MfeI*-hf and the produced fragments were visualised under the ultraviolet in a 1% agarose gel. The fragments of interest were excised and extracted. The final step included the ligation of the TOPO-insert in the pCCL-eGFP backbone. *HindIII* test digestions were performed to verify the completion of cloning (shown in Figure 21.F). All 4 successfully cloned mutants gave rise to the expected size fragments which were 3.3 kb, 2 kb, 584 bp, 556 bp, 499 bp and 313 bp long. As previously noted, the discrepancy between the expected number of bands and the number of bands on Figure 21.F was due to the inability to differentiate a 584 bp band to a 556 bp band on a 1% agarose gel.

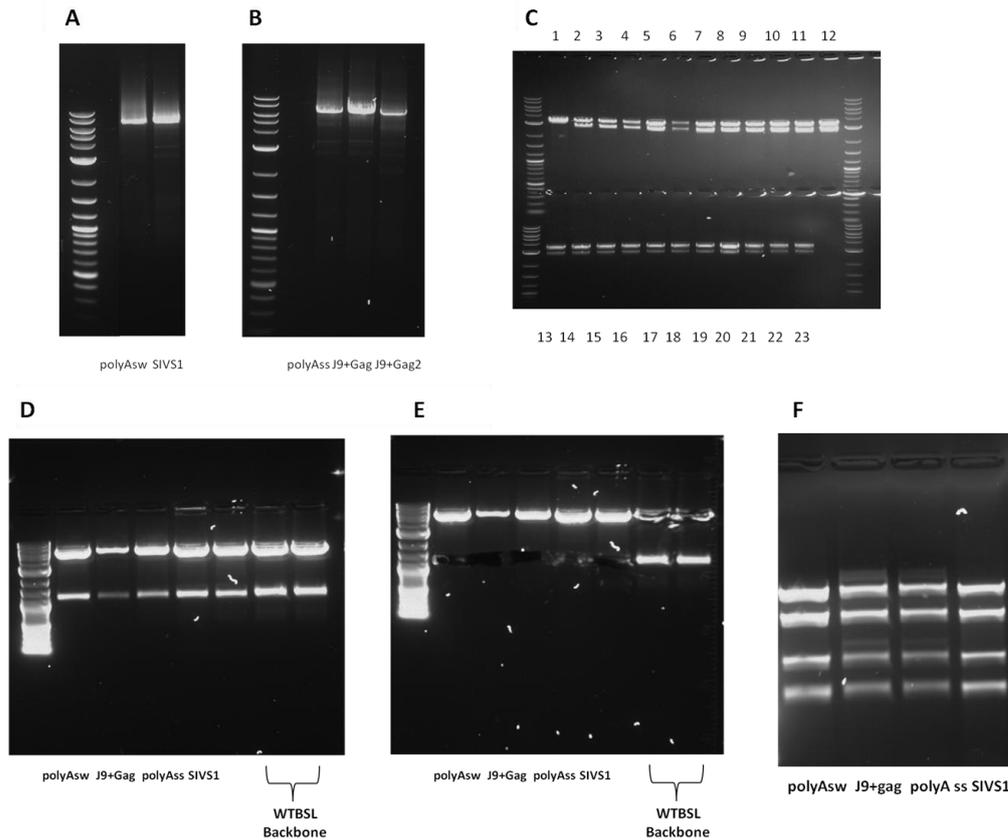


Figure 21. Agarose gel analysis of intermediates during the cloning process of the dimerisation mutants.

[A], [B] 1% agarose gel showing PCR products from the site directed mutagenesis reaction for all dimerisation mutants **[C]** EcoRI digestion of clones for confirmation of mutagenic insert of polyAsw (lanes 1-6), J9+Gag (lanes 7-12), polyAs (lanes 13-18) and SIVS1 (lanes 19-23). **[D]** Plasmids purified and digested with MfeI-hf and NdeI. **[E]** Bands of interest extracted. **[F]** HindIII digested plasmids purified from final clones.

4.6. Measurement of viral titres via Flow cytometry

The ability of the rationally designed transfer vectors to successfully transduce cells and express the eGFP transgene they carry was investigated with flow cytometry. 293T cells were transduced with serial dilutions of vector supernatant in order to assess the transduction efficiency of each transfer vector construct. In an attempt to understand the relationship between packaging efficiency and transduction efficiency, analysis of functional titres for a plethora of transfer vector mutants was performed. All transfer vector constructs

express eGFP under the control of the hPGK promoter that can be used as a marker of successful transductions. 72 h post transduction, cells were harvested and injected into the flow cytometer instrument to measure the successfully transduced eGFP expressing cells. Infectious titres were measured as transduction units per mL of viral supernatant (TU/mL). Transduction units were defined as the number of functional viral particles in the supernatant that were capable of transducing a cell and successfully expressing the transgene.

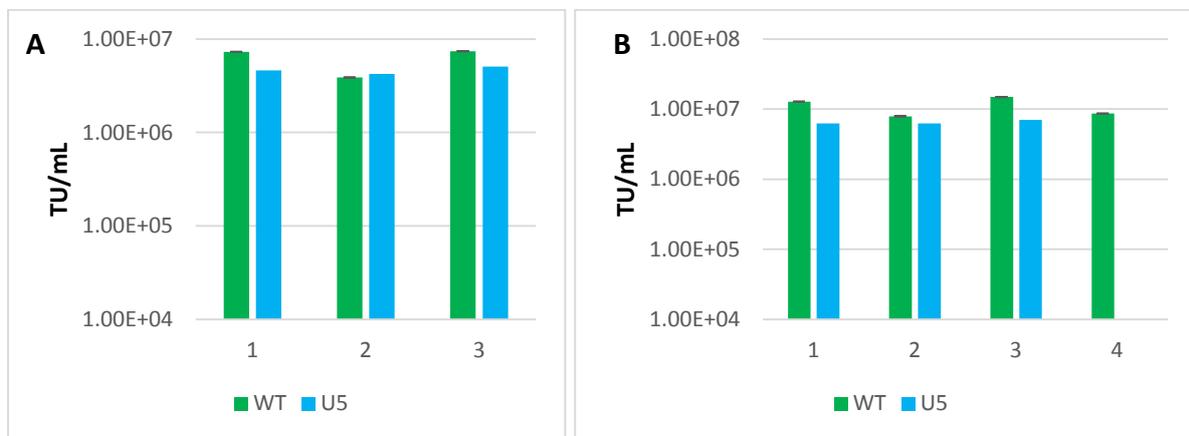


Figure 22. Comparison of infectious titres between WT and U5s viral vectors.

[A] Determination of infectious titres from 3 independent transfection replicates for WT and U5s from the same round of transfection experiments. Infectious titres measured in transduction units per mL (TU/mL). **[B]** Graphs presenting infectious titres measured in TU/mL from 4 independent transfection replicates for WT and 3 independent transfection replicates for U5s mutant. These independent replicates were derived from a second round of transfection experiments.

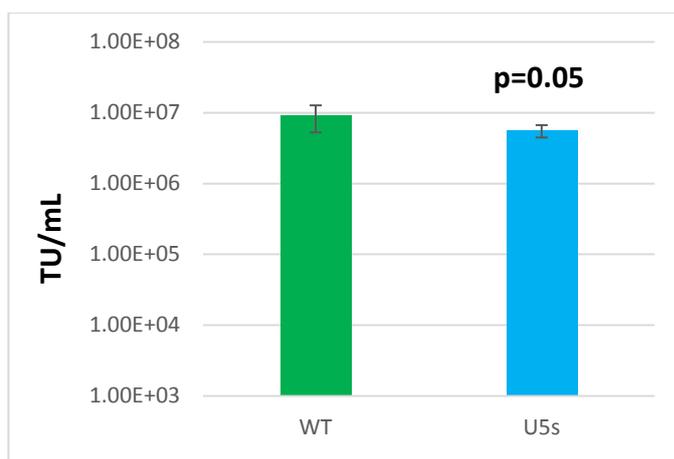


Figure 23. Combined comparison of infectious titres between WT and U5s viral vectors.

Columns showing the combined average transduction efficiency from 7 independent transfection replicates for WT and 6 independent transfection replicates for U5s. Column 1 represents the average TU/mL for WT, and column 2 represents the average TU/mL for U5s. Error bars represent standard deviations. The T-test performed showed no statistical significance.

HEK293T cells were transfected with the respective transfer vector and packaging constructs and viral supernatant was harvested 40 h later. Serial dilutions of harvested lentiviral supernatant were used to transduce HEK293T cells and the functional titres of WT and mutant lentiviral vectors were measured by multiplying the number of successfully transduced cells with the number of cells seeded and the viral dilution used, divided by the final supernatant volume.

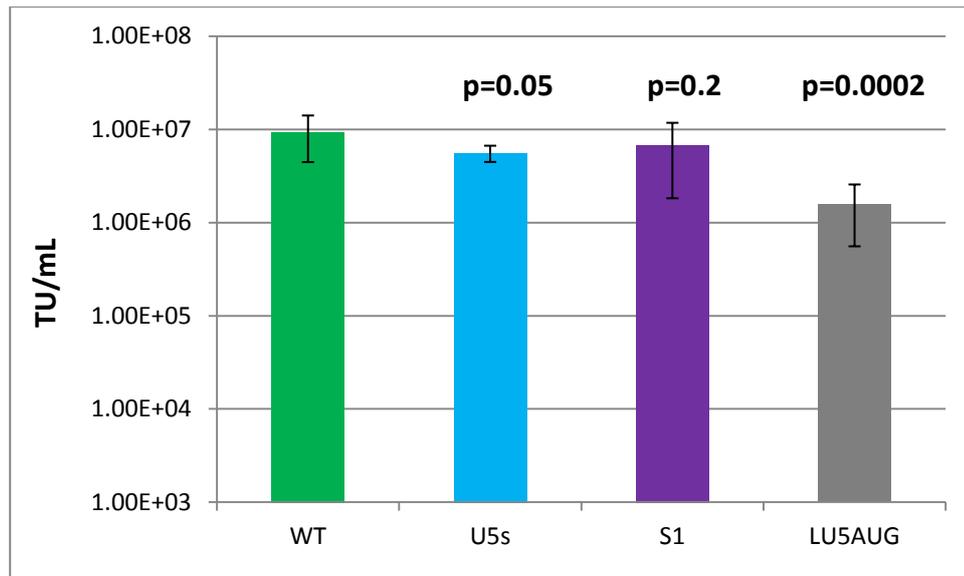


Figure 24. Comparison of infectious titres between WT, U5s viral vectors and the S1 and LU5AUG U5-AUG mutants.

Infectious titres of S1 and LU5AUG mutants were determined following the same methodology as described for WT and U5s above. Columns showing the average transduction efficiency from independent transfection replicates for each construct from 5 different experiments. Average WT is the average of 20 transfection experiments, while U5s average value is derived from 6 independent transfections, S11 from 14 and LU5AUG from 6. Column 1 represents the average transduction units per mL (TU/mL) for WT, column 2 represents the average TU/mL for U5s, column 3 represents the average TU/mL for S1 and column 4 represents the average TU/mL for LU5AUG. Error bars represent standard deviations and p values represent comparisons with WT.

U5s, S1 and LU5AUG mutants were designed to increase the U5-AUG base pairing that is associated with a conformation documented to dimerise more efficiently^{7 61}. PolyAs and PolyAsw were designed to have a stabilised lower polyA stem while both J9+Gag and SIVS1 contained a modified SL1. All of these mutations aimed to promote adoption of a dimerisation efficient conformation. Flow cytometry was performed to investigate the ability of such constructs to successfully transduce cells. Functional analyses were performed to study the effect of these mutations not only on their dimerisation and

packaging efficiency, but also on their infectivity. Interestingly, a wide range of transduction efficiencies was observed with mutants like S1 and PolyAsw having functional titres comparable to WT, while J9+Gag and PolyAs were characterised by a 50 fold decrease (Figure 25).

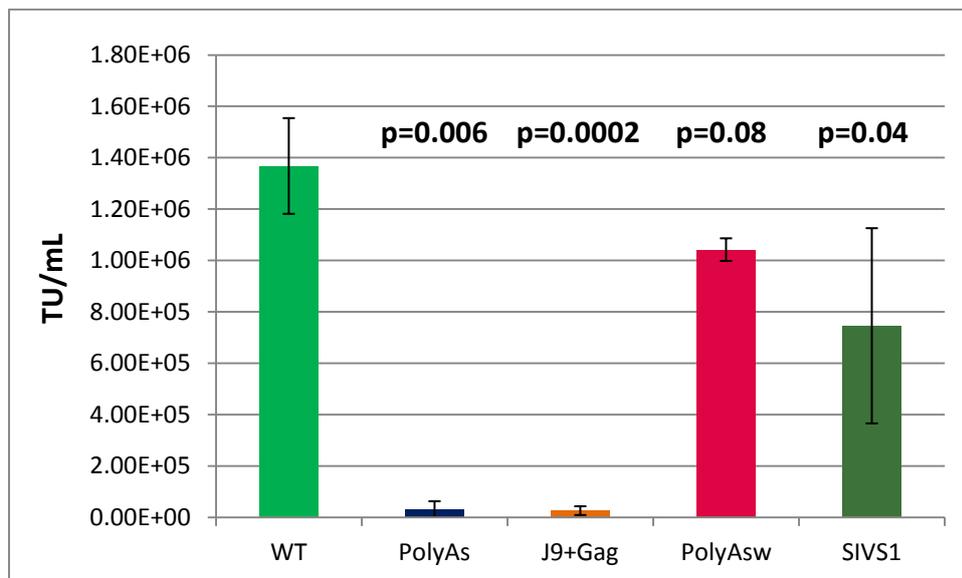


Figure 25. Comparison of infectious titres between WT and the dimerisation vector mutants designed: PolyAs, J9+Gag, PolyAsw and SIVS1.

Transfections and transductions were performed as previously described. Columns showing the average transduction efficiency from 7 independent transfection replicates for each construct. Column 1 represents the average transduction units per mL (TU/mL) for WT, column 2 represents the average TU/mL for PolyAs, column 3 represents the average TU/mL for J9+ Gag, column 4 represents the average TU/mL for PolyAsw and column 5 represents the average TU/mL for SIVS1. Error bars represent standard deviations and p values represent comparisons with WT.

4.7. Measurement of virion production

ELISA p24 was performed to measure Capsid-p24 protein concentration and thus Gag protein expression, which is associated with physical vector titre. WT, U5s and UC4d constructs generated similar amounts of p24 as seen in Figure 26.A, although U5s expression was generally higher. HEK293T cells were transfected with the packaging

constructs and the transfer vector of interest. Viral supernatant was harvested 40 h post transfections and was used for measurement of physical titres. The standard deviation of NC4s results was high, but on average this construct produced the least amount of p24. Fluorescence microscopy analysis revealed that the WT plasmid was initially giving lower transfection efficiencies (data not shown). Another plasmid purification (Maxi-prep) was performed for WT and used for 3 independent transfections alongside the U5s plasmid. When transfection efficiencies were comparable (fluorescence microscopy, data not shown), supernatants produced with the use of these plasmids were harvested and treated for p24 measurement by ELISA. p24 expression for WT and U5s was very similar, with an average value of 189 ng/mL and 204 ng/mL respectively. Thus it can be concluded that WT and U5s constructs led to similar virion production, possibly because, unlike in WT HIV-1, Gag was provided *in trans*.

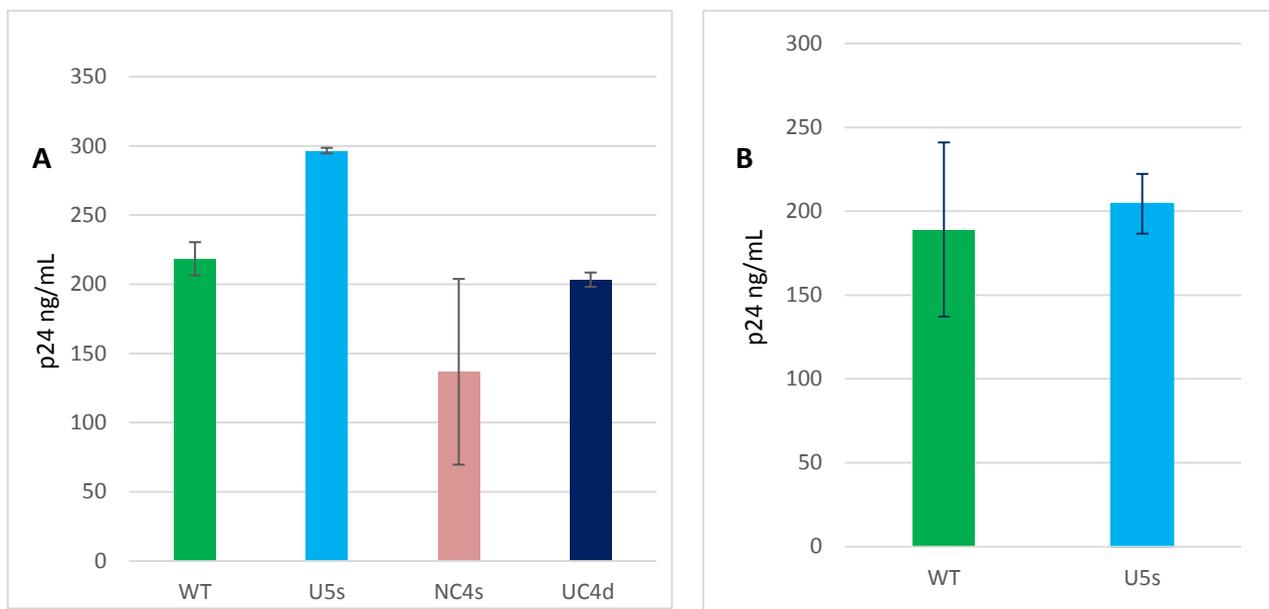


Figure 26. Comparison of virion production between the various mutants.

Graphs show the amount of p24 produced per mL and measured with ELISA. Error bars represent standard deviation. **[A]** Graph representing the average amount of p24 per mL for WT and the 3 designed mutants in viral supernatant from 3 transfection replicates. **[B]** A comparison of virion production between WT and U5s using supernatant generated from three independent rounds of transfections with the use of different plasmid purifications (maxi-preps).

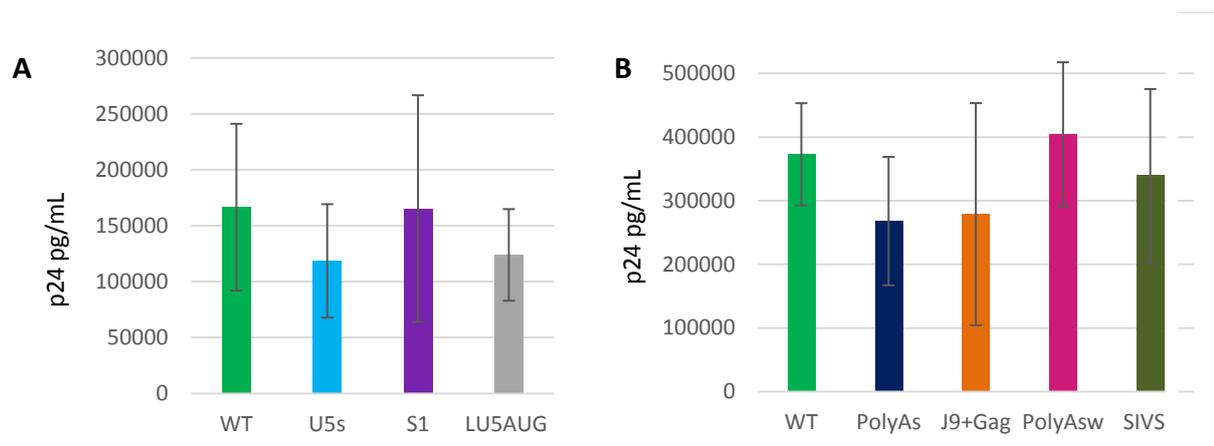


Figure 27. Measurement of virion production for the U5-AUG and dimerisation mutants.

Graphs show the amount of p24 produced per mL and measured with ELLA p24 with error bars that represent standard deviation. **[A]** Graph representing the average amount of p24 per mL for WT and the 3 U5-AUG mutants in viral supernatant from 21 transfection replicates per mutant. **[B]** A comparison of virion production between WT and the 4 dimerisation mutants using supernatant generated from three independent rounds of transfections.

Upon their translation, Gag and Gag/Pol multimerise alongside the gRNA. HIV-1 virus like particle (VLP) assembly can be achieved independently of the packaging of the gRNA, it is however facilitated by it²⁴³. Similarly to ELISA p24, ELLA p24 (GSK), an automated immunoassay, was performed to measure Capsid-p24 expression but in a more accurate way. The ELLA p24 run contained a pre-loaded on a cartridge p24 immunoassay with an internal calibration curve. The only non-automated part included the sample and buffer loading. Each sample ran through a microfluidic channel that contained 3 Glass Nano Reactors (GNRs) coated with a capture antibody. The final values were calculated based on the pre-loaded standard curve and the triplicate data generated per sample. Hence, ELLA performed the same p24 immunoassay with ELISA, but without the accompanying errors of traditional immunoassays.

WT, U5s, S1 and LU5AUG produced physical titres analogous to their infectious titres, suggesting that mutations in the U5-AUG motif of the 5'UTR did not influence Gag expression when provided *in trans*. Measurement of the p24 expression levels from

supernatant produced by polyAs, J9+Gag, polyAsw and SIVS1 mutants, revealed a similar pattern. p24 expression and thus Gag expression and virion production for all dimerisation mutants was related to their respective ability to transduce cells successfully.

4.8. Effect of mutations in 5'UTR on Gag processing and budding

As shown in Figure 28, Gag expression levels were equivalent between the WT and mutated constructs. Interestingly, the p55/p24 antibody detected uncleaved p55 product in viral supernatants which suggests the presence of immature particles, or might be reflecting the presence of VLPs lacking gRNA in the supernatant. The conversion of p55 to p24 is one of the late stage processing events and a sensitive step in the whole process²⁴⁴. Westerns on cell lysates were performed in order to look into the p24/p25 ratio. The blots did not reveal any significant differences in the amount of total Gag products or in Gag processing. Both the Gag expression levels and the p24/p25 ratio appeared very similar for all constructs.

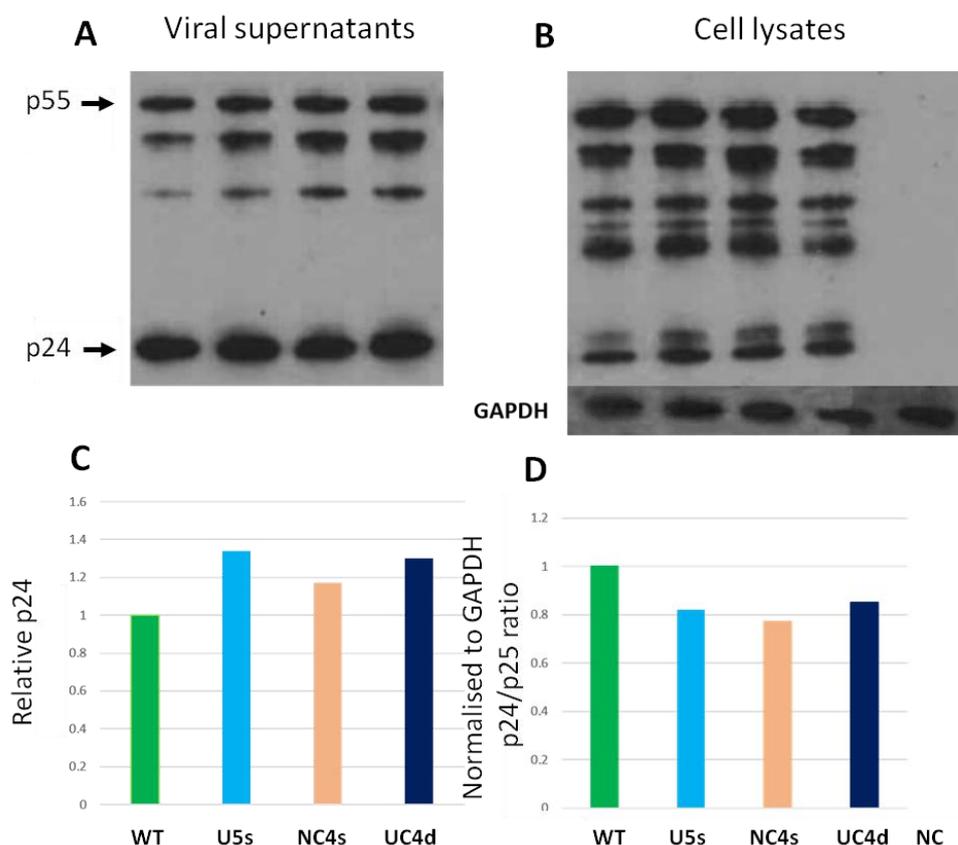


Figure 28. Investigation of Gag processing of first group of transfer vector mutants.

Anti-p55/p24 Western blots of purified virions and cell lysates were performed to quantify p24 expression and to also investigate any potential effect of mutations in the U5 region of gRNA in Gag processing in a trans system. **[A]** Western blots of purified virions. **[B]** Western blots of cell lysates. The same blot was stripped and re-used for GAPDH detection and sample normalisation. **[C]** Quantification of p24 on purified virions by ImageJ. **[D]** Quantification of p24/p25 ratio on cell lysates by ImageJ.

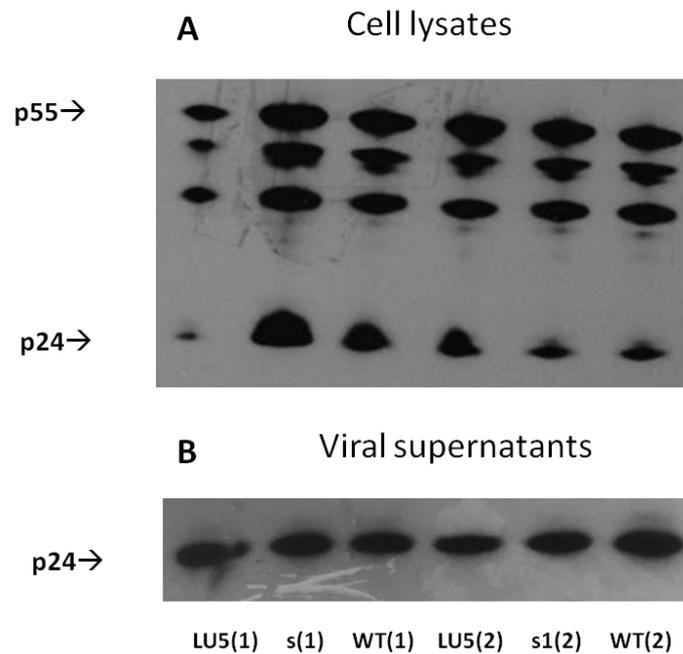


Figure 29. Gag processing study of U5-AUG mutants.

[A] Western blots of cell lysates. This time the blot was not re-used for GAPDH detection, but sample normalisation was achieved by loading an equal amount of total protein as described in 3.11. **[B]** Western blots of purified virions.

Anti-p55/p24 Western blots of cell lysates and purified virions were performed to compare p24 expression between Western Blots and ELLA and to also identify any changes in Gag processing caused by mutations in the U5-AUG motif. This was an important experiment to perform as ELLA, like ELISA, detects three dimensional structures, and the structures within Gag are known to change during particle maturation.

Finally, ELLA p24 was also performed to measure intracellular Capsid-p24 expression levels to study any potential impact of the various 5'UTR mutations on intracellular Gag expression

and/or its ability to bud from the cell. 20 replicates per U5-AUG transfer vector construct from 4 independent rounds of transfections, and 7 repeats per dimerisation transfer vector mutant from 2 independent rounds of transfections were analysed in Figures 30.A and 30.B respectively. The extracellular to intracellular Gag amounts (p24 V/C) measured did not show any major differences between WT and the U5-AUG mutants while as suggested by the t-test statistical analysis, the calculated p24 V/C ratio was relatively higher for the dimerisation mutants polyAs, polyAw and SIVS1 compared to WT.

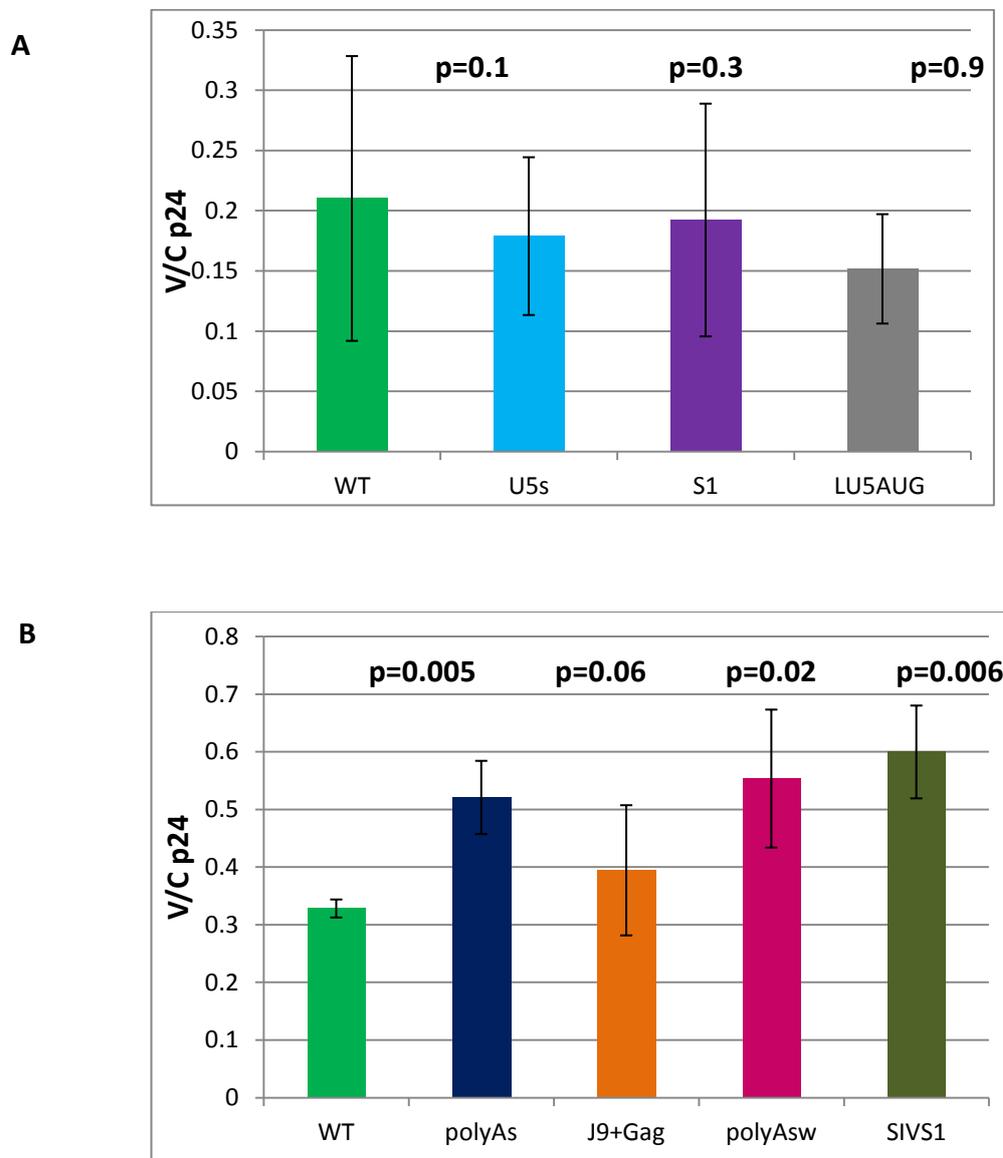


Figure 30. Effects of 5'UTR mutations on Gag budding.

[A] Intracellular to extracellular p24 expression ratio for the U5-AUG mutants **[B]** Intracellular to extracellular p24 expression ratio for the dimerisation mutants.

4.9. Chapter conclusions

Dimerisation promoting transfer vector mutants were designed by targeting regions in the 5'UTR that play important roles in the process of dimerisation, including the DSL, the U5-AUG duplex and the polyA stem loop. The mutants were firstly assessed for their infectious titres. Amongst all mutants, U5s and S1, designed to have strengthened U5-AUG base pairing, showed comparable, but slightly lower infectious titres to WT. However, the LU5AUG mutant that was also designed with a strengthened U5-AUG interaction, showed a 14 fold infectious titre reduction. Interestingly, the dimerisation mutants showed a wide range of infectious titres, with mutants PolyAsw having infectious titres comparable to WT, while J9+Gag and PolyAs were characterised by a 50 fold decrease. This result could be reflecting the multifunctionality of the targeted areas. Mutations in the DSL, U5-AUG and polyA are likely to affect not only dimerisation and packaging, but also other important parts of the HIV-1 life cycle essential for the infectivity of the newly synthesised mutant viral particle. The amount of p24 Capsid in the viral supernatant of WT and mutants was measured next to assess the effect of the rationally designed mutations on physical titres. It was observed that physical titres follow the same trend as infectious titres. This observation might suggest that a certain amount of Gag was able to bud away independently of the presence of gRNA, in the form of VLPs, but that the presence of gRNA further facilitated Gag assembly and budding. Western Blots were performed to investigate whether the 5' UTR mutations can affect the processing of Gag provided *in trans*. No change in the Gag processing pattern was observed between WT and mutants. Finally, intracellular Capsid-p24 expression levels were measured to investigate any potential impact of the various 5' UTR mutations on intracellular Gag expression and/or its ability to bud from the cell. The extracellular to intracellular Gag amounts (p24 V/C) measured did not show any differences between WT and the U5-AUG mutants while, with the exception of J9+Gag, the p24 V/C ratio was observed to be relatively higher for the dimerisation mutants compared to WT ($p < 0.5$).

This chapter discussed the biological characterisation of the transfer vector mutants. These mutants were rationally designed based on the premise that dimerisation promoting mutations should increase packaging and thus transduction efficiency. Vector titration

revealed the overall negative effect of these mutations on infectious titres. However, such assays could not provide information on RNA dimerisation or packaging efficiency. To assess this, the next chapter discusses assays performed to measure the total amounts of gRNA packaged in virions and its propensity to dimerise, but also the efficiency of the mutant RNAs to dimerise *in vitro*.

5. RNA dimerisation propensity

Lentiviral vector particles contain two copies of the full length transfer vector RNA. HIV-1 gRNA is known to act as a switch between the monomeric conformation that is associated with translation and the dimeric conformation linked with packaging⁸. The dimerisation and packaging properties of lentiviral vectors were targeted with the aim of improving their efficiency based on the hypothesis that WT HIV-1 regulates genome encapsidation tightly by recognising dimeric RNA¹². Mutations introduced in the 5'UTR were assessed in the previous chapter for their effect on viral vector infectivity. In chapter 5, I investigated the effect of the same mutations on RNA dimerisation efficiency as well as any potential link between RNA dimerisation and viral vector infectivity. The propensity of RNA to dimerise was studied *in virio* with Northern Blotting and with *in vitro* dimerisation assays. Northern Blotting was chosen as the most appropriate technique due to its advantage of measuring both the quantity of total gRNA levels and the stability of its dimeric form. *In vitro* dimerisation assays were employed to further study the propensity of RNA to dimerise for mutants whose gRNA was not detectable with Northern Blotting due to low encapsidation levels.

5.1. Probe generation

The different kinetic properties of RNA and DNA were used to visualise the conversion of the PCR DNA to an RNA product via *in vitro* transcription, by running these two nucleic acid molecules alongside each other in a 1% agarose RNase-free gel. It can be seen that both PCR product 1 (lane 5) and PCR product 2 (PCR product band extracted from agarose gel, lane 6) used for *in vitro* transcription, generated probes of the right size and without any degradation (Figure 31). The RNA probe 2 (lane 3), which was produced from the gel-extracted PCR template, was selected as the probe of choice for use in Northern Blotting to ensure that there were no other by-products in the probe solution.

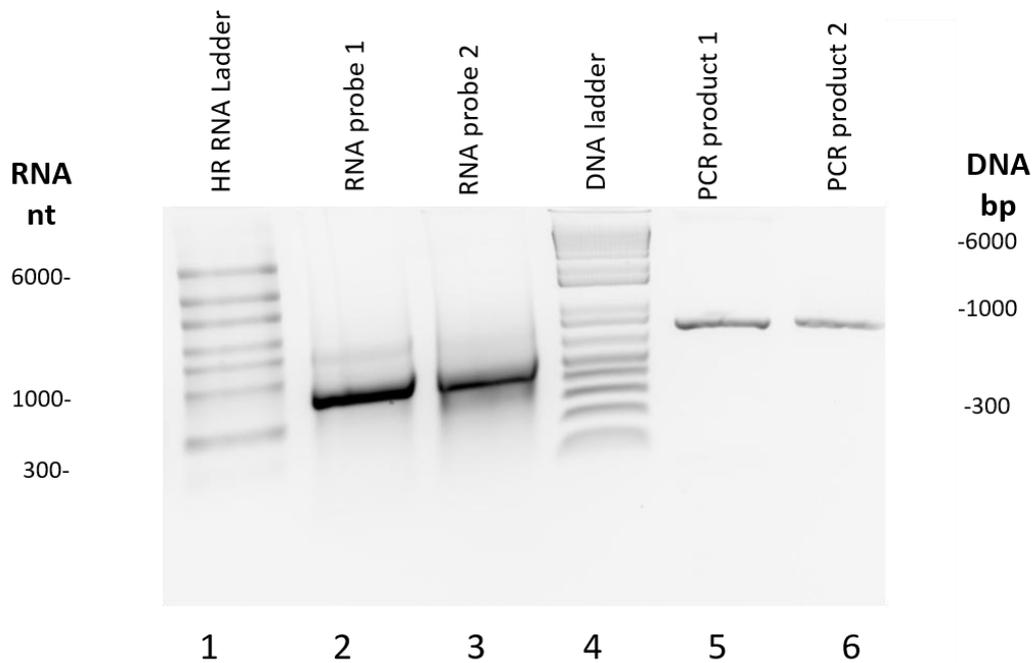


Figure 31. Production of an RNA probe that targets eGFP and WPRE sequences.

1% agarose gel showing the 744 nt RNA probe (lane 2) transcribed from a gel extracted PCR product template (lane 6) and another RNA probe (lane 3) transcribed from a PCR product template that had not been gel extracted (lane 3).

5.2. Non-denaturing (TBE) Northern Blots

The expected size of the monomeric transcript of the transfer vector is 3.7 kb, thus the size of the genomic RNA in its dimeric form was expected to be 7.4 kb. RNA extracted from 1 mL of concentrated supernatant was loaded on a 0.8% agarose gel in TBE and was electrophoresed to allow different sizes of encapsidated gRNA to be separated. RNA was then transferred on a membrane on which was fixed with heat. The membrane was hybridised with the DIG-labelled probe from Figure 31 that spans the eGFP and WPRE region. RNA extracted from transfected cells (cRNA) was used as a control for the transfer step. WT and U5s RNAs electrophoresed in this blot were extracted from viral supernatants generated from either different transfection experiments, or independent replicates from the same round of transfection. RNA of the right sequence was labelled and visualised on X-ray film. In all cases, based on visual inspection of band intensity shown in Figure 32, more

WT than U5s virion RNA was detected. NC4s RNA could not be detected in this blot, and UC4d RNA levels were lower than WT. A dimer/monomer RNA ratio could not be calculated due to the saturation of the dimer RNA band.

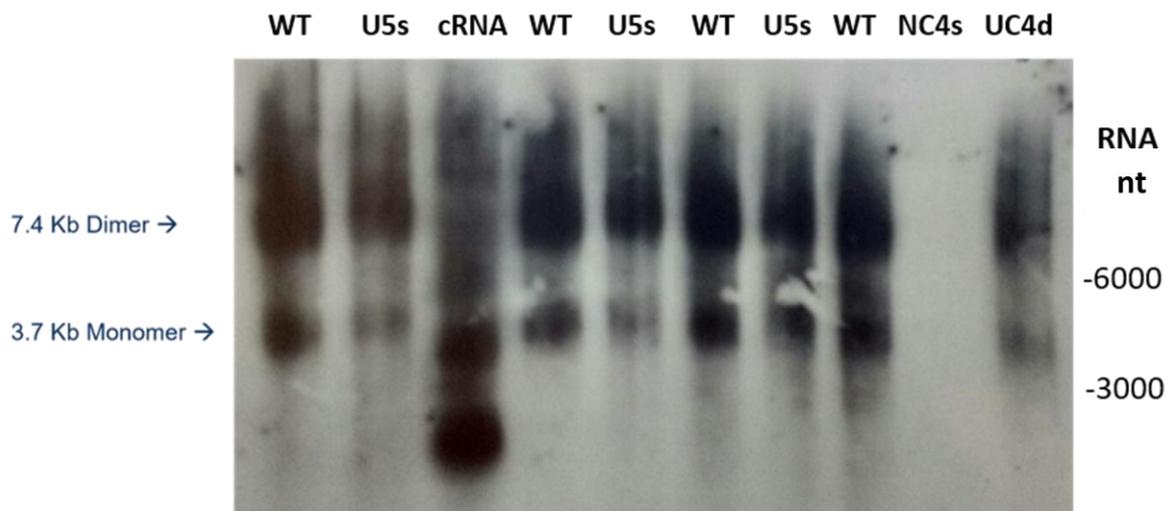


Figure 32. Visualisation of virion RNA of WT and mutant samples with Northern Blotting.

Native Northern Blot of purified transfer vector RNA probed with the 744 nt probe. 0.8% agarose gel was run in TBE for 4 h to allow separation of monomeric and dimeric RNAs. The membrane on which the separated RNAs were transferred to was incubated overnight with 75 ng of a DIG-labelled RNA probe. Arrows indicate the bands corresponding to the full length dimeric and monomeric RNA for WT and mutants. Total levels of encapsidated gRNA were lower for all designed transfer vector mutants than for WT.

The experiment was repeated and RNA extracted from WT and U5s supernatants from independent transfection experiments was electrophoresed in a native 0.8% agarose gel, Northern Blotted and used to calculate dimerisation efficiency. A t-test was performed to compare the dimerisation efficiency of WT and U5s, as Figure 33 indicated that U5s has a higher dimerisation efficiency. This observed trend was not found to be statistically significant ($p=0.07$). With the exception of WT2 and U5s2 repeats, the total amount of dimeric RNA and the total amount of RNA encapsidated appeared to be higher for WT virions.

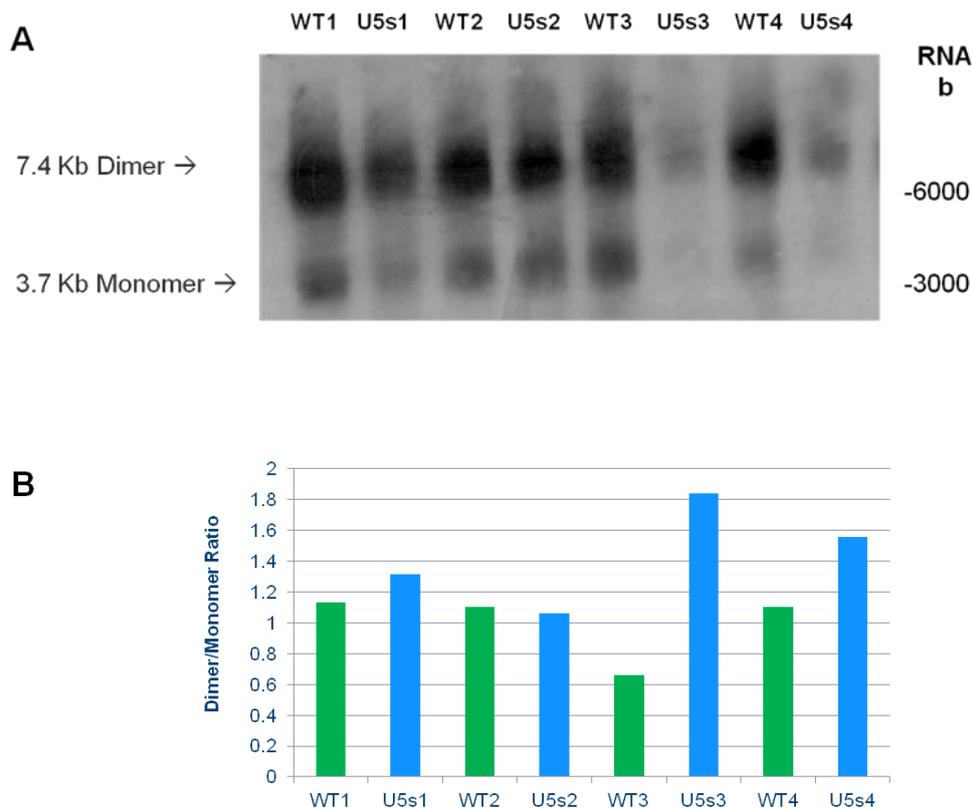
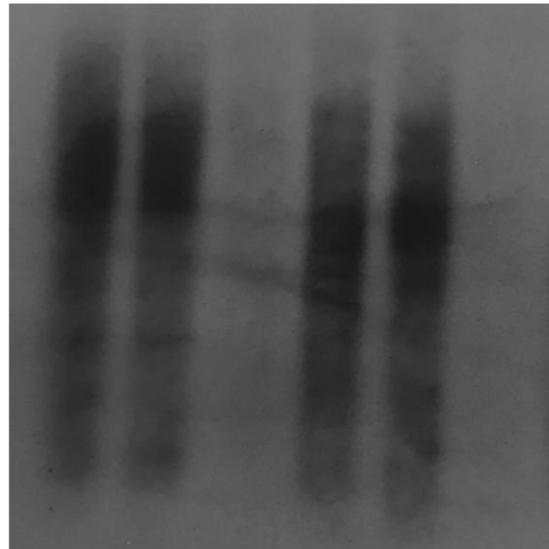


Figure 33. Visualisation of virion RNA of WT and mutant samples with Northern Blotting.

[A] Native Northern Blot of WT and U5s purified transfer vector RNAs probed with the 744 nt probe targeting WPRE-eGFP. Arrows indicate bands corresponding to the full length genomic dimeric and monomeric RNA for WT and U5s replicates. **[B]** Calculation of Dimer/Monomer genomic RNA calculated via ImageJ.

WT, S1 and LU5AUG virion RNA was extracted from two independent replicates from the same round of transfections and electrophoresed in a 0.8 % native agarose gel, then transferred to a positively charged Amersham™ Hybond™-N+ membrane (GE healthcare), and blotted using the 744 nt DIG-labelled probe. Visualisation of the RNA bands could not differentiate between dimeric and monomeric RNA, hence this blot was mainly used for a relative comparison of the total RNA levels. Band intensity was equivalent for WT and S1 virions whereas no LU5AUG RNA could be detected in this blot.



WT(1) S1(1) LU5AUG(1) WT(2) S1(2) LU5 AUG(2)

Figure 34. Visualisation of virion RNA of WT and U5-AUG mutants.

Native Northern Blot showing the total amount of RNA packaged and extracted from WT, S1 and LU5AUG mutants. This Northern Blot was characterised by low resolution and it was thus used for a relative comparison of the total amounts of RNA packaged by each type of viral vector produced.

WT and S1 virion RNA was extracted from three independent replicates from the same round of transfections and electrophoresed on a 0.8% native agarose gel, transferred and blotted as described above. S1 was chosen for further studies due to the higher viral yields produced by this mutant that would allow for its viral RNA to be detected with Northern Blotting, but also due to its interesting functional phenotype. Northern Blotting (Figure 35) revealed that S1 RNA from virions appears to be characterised by a higher propensity to dimerise compared to WT. Dimeric and monomeric RNA band intensity analysis with ImageJ showed that S1(1) and S1(3) repeats have a 3 times higher dimerisation efficiency ratio (D/M) compared to the respective WT(1) and WT(3) repeats. However, the total levels of encapsidated RNA extracted from virions and detected here were lower for S1 than for WT. T-tests performed on both dimerisation efficiency ($p=0.2$) and packaging efficiency ($p=0.08$) did not reveal statistical significance, possibly due to the small number of repeats that could be assessed.

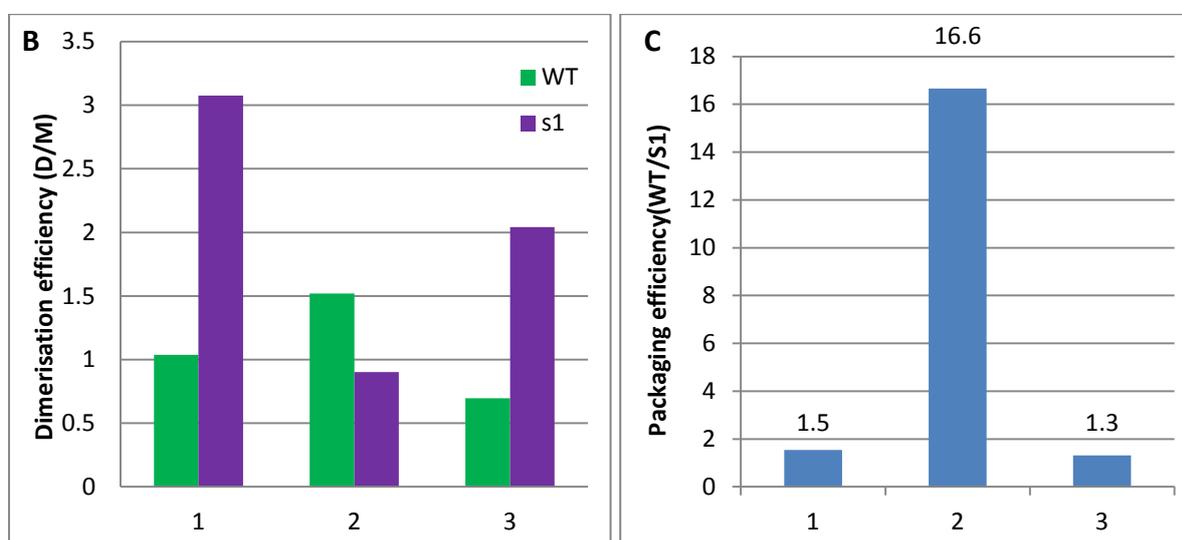
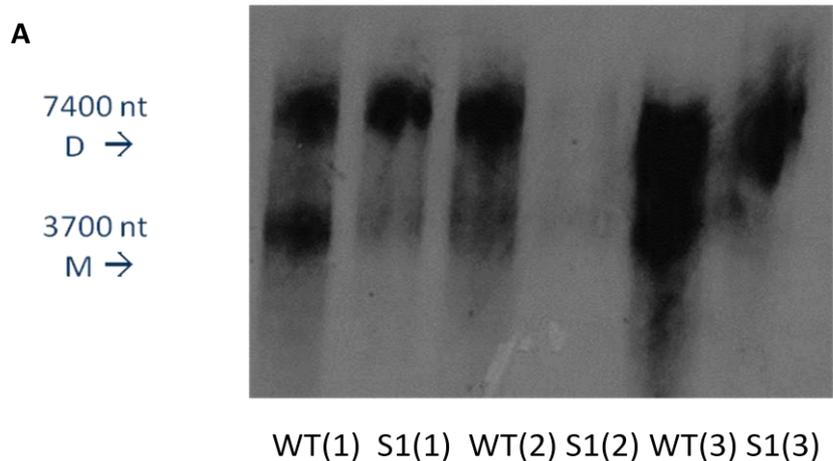


Figure 35. Visualisation of virion RNA of WT and S1 with Northern Blotting.

[A] Native Northern Blot of WT and S1 purified transfer vector RNAs probed with the 744 nt probe targeting WPRE-eGFP sequences. The visualised bands correspond to the full length genomic dimeric and monomeric RNA for WT and S1 replicates. **[B]** Graph presenting a comparison between WT and S1 virion RNA dimerisation levels. Calculation of transfer vector dimerisation efficiency, defined as the ratio of dimeric to monomeric full length genomic RNA, was performed with ImageJ. **[C]** Graph showing the relative packaging efficiency of WT and S1 virions. Band intensity of both dimeric and monomeric RNA bands was measured with ImageJ. The dimeric and monomeric RNA values generated by densitometry were added to calculate the total amount of RNA packaged in virions and were then used to measure the relative packaging efficiency of WT to S1 transfer vectors.

5.3. *In Vitro* Dimerisation RNA Studies

5.3.1. *In vitro* synthesis of RNA for dimerisation studies

690 nt RNA fragments that included the sequences of the HIV 5' R, U5 and partial *gag* included in the transfer vector constructs were produced to 1) study the effect of the U5-AUG and dimerisation mutations on the RNA structure and its propensity to dimerise *in vitro* and 2) investigate the effect of partial *gag* sequences on the stability of the packaging conformation. PCR products spanning the aforementioned sequences were created using a Forward 5' Primer that contained the T7-Pol Transcription start site (TSS). The 50 µl PCR reactions were loaded in a 1% agarose gel and the correct size products were excised and purified. Lack of cross contamination was ensured by loading a PCR negative control (NC), which was made of plasmid DNA template-free PCR mix, alongside the mutant plasmid containing reactions. The total amount of the gel extracted PCR products was then used as template for large scale *in vitro* transcription reactions (scaled up 4 times than the manufacturer's suggested protocol). 2 µg of each RNA fragments were then visualised under the UV in a 1% ethidium bromide stained agarose gel to assess conversion of DNA to RNA and ensure correct length and integrity of the *in vitro* synthesised RNA products.

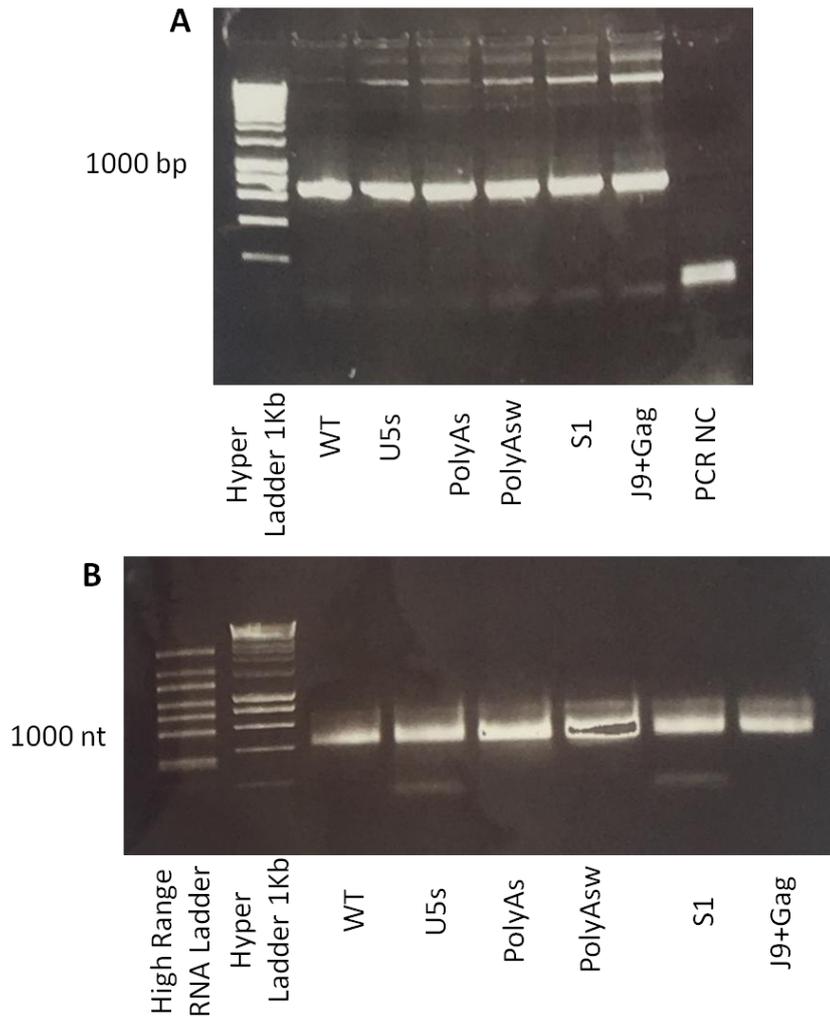


Figure 36. Production of *in vitro* transcribed 690 nt RNA fragments of interest.

[A] PCR reactions were loaded in a 1% TBE agarose gel to visualise PCR products that should contain the T7 TSS for *in vitro* transcription of U5-AUG and dimerisation mutants. The correct size products were excised, purified and used as *in vitro* transcription templates. **[B]** Confirmation of synthesis of a 690 nt RNA fragment containing the *cis*-acting elements of 5' R, U5 and partial gag, for a selection of U5-AUG and dimerisation mutants. *In vitro* transcribed products were mixed with 2x Ambion RNA loading dye and were electrophoresed in a 0.8% agarose gel in TBE to assess correct size, lack of degradation and potential existence of byproducts.

5.3.2. Optimisation of *in vitro* RNA dimerisation

An *in vitro* RNA dimerisation protocol was optimised to study the dimerisation efficiency of WT and mutant RNAs, as well as the structural conformation of their respective dimers and monomers. As mentioned earlier and unlike previous studies that had focused only on the HIV-1 leader sequences, a longer 690nt RNA containing the HIV 5' U5, R and partial *gag* was synthesised and studied. As shown by Sinck *et al.*, *in vitro* RNA dimerisation is inhibited with increasing RNA size²⁴⁵, hence a new protocol had to be established to enable dimerisation of the longer RNA of interest. Similar to Sinck *et al.*, a refolding buffer that contained sodium cacodylate and increased MgCl₂ concentration was used to investigate whether sodium cacodylate addition and increased MgCl₂ concentration could positively affect RNA dimerisation *in vitro*.

Two independent rounds of *in vitro* transcription were performed using WT and U5s PCR gel extracted templates. Samples were heated at 95°C for 2 min, placed on ice for 2 min and then electrophoresed on a 1% agarose TBE gel with a native loading dye and visualised with EtBr staining and UV transillumination. As shown in Figure 37.A, under these renaturation conditions, both WT and U5s RNAs existed predominantly in the monomeric state. Figure 37.B depicts WT RNA studied in a 4% polyacrylamide TBM gel, which is normally used for in-gel SHAPE structural analysis experiments. WT RNA was mixed with 1x Renaturation buffer (10 mM Tris pH 8, 100 mM KCL and 0.1 mM EDTA) and heated for 5 min at 95°C. The sample was allowed to reach 28°C and was then mixed with 5x refolding buffer (40 mM Tris pH 8, 4 mM MgCl₂ and 130 mM KCl) for 30 min at 37°C. Gel electrophoresis took place for 4h allowing the dimeric and monomeric RNAs to separate. Similar to the 1% TBE agarose gel the WT RNA studied here was mainly in the monomeric conformation with only a very small proportion of it adopting the dimeric conformation.

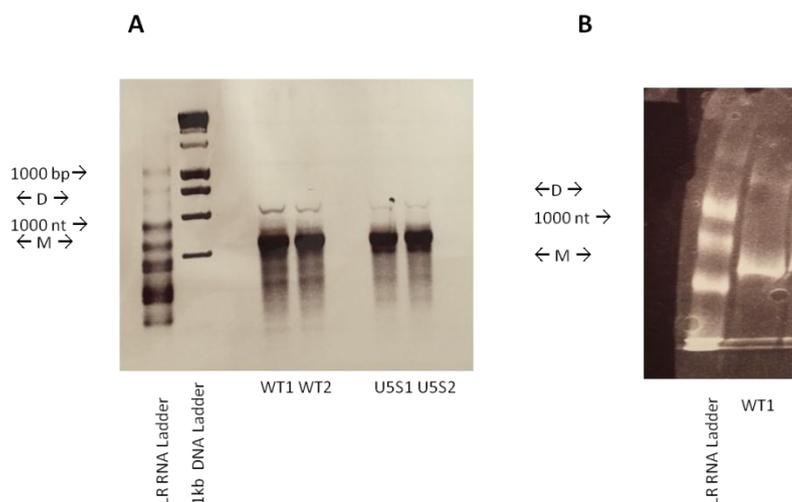


Figure 37. *In vitro* dimerisation studies of WT and U5s RNAs visualised with the use of different types of gels.

[A] *In vitro* dimerisation study of WT and U5s RNAs visualised in a 1% agarose TBE gel. **[B]** *In vitro* dimerisation of WT RNA visualised in a 4% polyacrylamide TBM gel.

With the aim of enabling dimerisation in this length of RNA, the effect of sodium cacodylate and increased $MgCl_2$ concentration on enhancement of RNA efficiency was studied. The 5x Refolding buffer (Rb): 40 mM Tris pH 8, 4 mM $MgCl_2$ and 130 mM KCl^8 was compared against the 5x Dimerisation buffer (Db): 50mM sodium cacodylate pH 7.5, 40 mM Tris pH 8, 5 mM $MgCl_2$ and 130 mM KCl^{245} . Confirming the results previously shown in Figure 37.A., both WT and U5s 690 nt long RNA fragments in Rb dimerised poorly (Figure 38 lanes 1 and 2 respectively). Incubation of WT and U5s RNAs in Db caused a dramatic shift towards the dimeric conformation, with both RNAs mainly adopting the dimeric conformation (Figure 38 lanes 3 and 4 respectively). Finally, addition of these reagents had an effect on the electrophoretic RNA kinetics with a notable non linear smear.

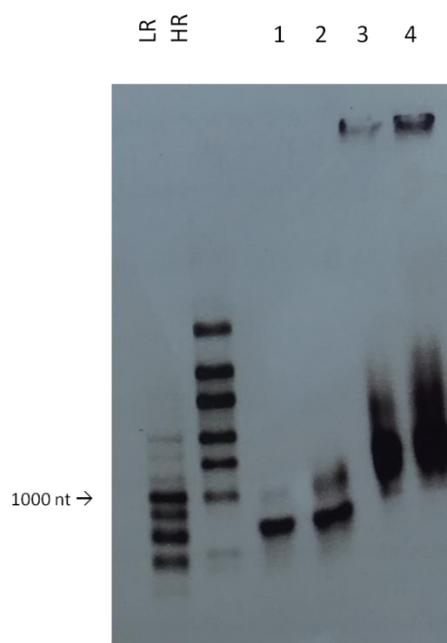


Figure 38. Optimisation of the refolding buffer conditions to enhance *in vitro* RNA dimerisation efficiency.

WT RNA in Rb (lane 1), U5s RNA in Rb (lane 2), WT RNA in Db (lane 3) and U5s RNA in Db (lane 4) were visualised in a 1% agarose TBM gel containing 0.1 mM MgCl₂²⁴⁵.

The RNA smear observed in Figure 38. was addressed by decreasing the MgCl₂ concentration in the TBM buffer. MgCl₂ concentration was reduced to 0.03mM⁸. 3 µg of U5s RNA were electrophoresed in a 4% non-denaturing polyacrylamide TBM gel. U5s was chosen over WT to assess the impact of MgCl₂ concentration reduction on electrophoresis kinetics, as a higher level of smear and dimerisation efficiency was previously observed for this mutant. As shown in Figure 39 a good ratio of dimer to monomer RNA was achieved without the presence of any noted smear. Based on these observations Db buffer and 1x TBM buffer containing 0.03mM MgCl₂ were chosen as optimised conditions for all further *in vitro* dimerisation assays and in-gel SHAPE experiments performed. This optimised protocol was also used for a quality control experiment using both the 356 nt long RNA fragment previously studied⁸ and the longer 690 nt fragment interrogated in this thesis (Figure 40).

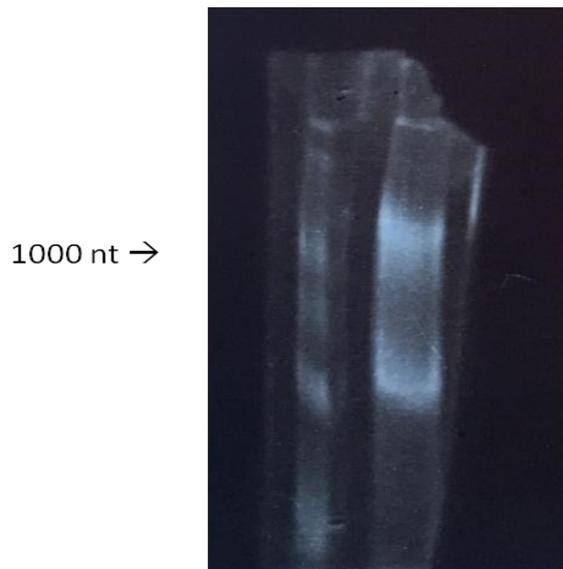


Figure 39. *In vitro* dimerisation of U5s RNA incubated in the optimised Db visualised in a 4% polyacrylamide TBM gel.

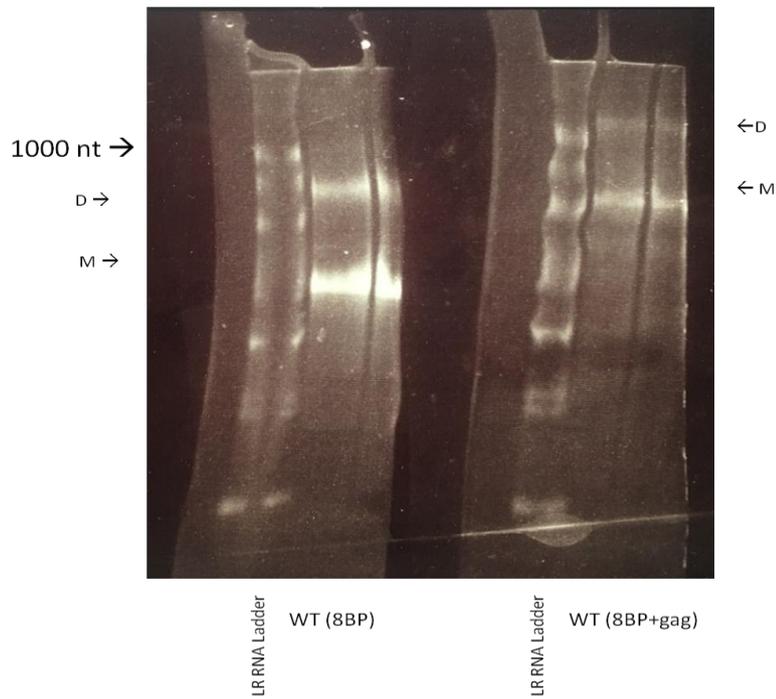


Figure 40. Side by side comparison of WT RNA of two different lengths in a 4% polyacrylamide TBM gel used for in-gel SHAPE experiments.

356 nt long RNA fragment (8BP⁸) kinetics and dimerisation propensities were compared to the 690 nt long RNA fragment that contains the additional partial gag sequences.

5.3.3. *In vitro* dimerisation of WT and mutant RNA

Previous studies from the literature have shown S1 to dimerise 60% better *in vitro* than WT. Hence, an *in virio* to *in vitro* study for S1 was performed above to study its dimerisation properties in our system (Figure 35). *In vitro* studies of the 690 nt long RNA fragments presented in Figure 41 revealed a similar result with previously published data. S1 RNA synthesised *in vitro* had a higher dimerisation propensity compared to the same length WT *in vitro* synthesised RNA fragment.

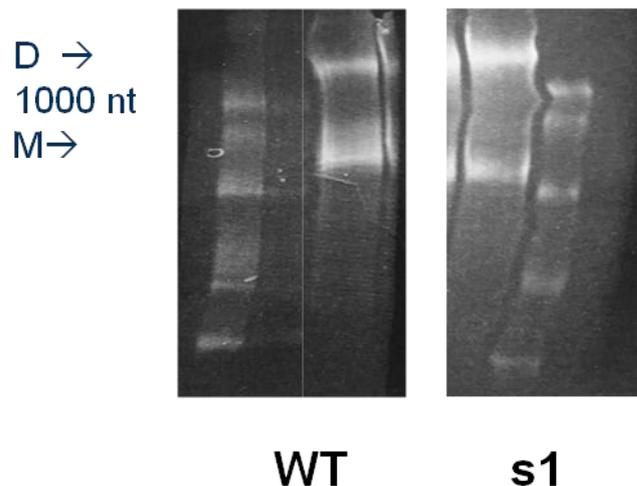


Figure 41. *In vitro* dimerisation assay assessing the dimerisation efficiency of WT and S1 transcribed RNA fragments produced with the optimised protocol developed in 5.3.b.

Arrows mark the 690 nt monomeric fragment and the 1380 nt dimeric fragment. Low Range Riboruler was the ladder of choice to confirm right size of RNA products.

690 nt long RNAs were produced *in vitro* as described in 3.9.1. for WT, U5-AUG mutants S1 and U5s, and the dimerisation mutants J9+Gag, polyAsw and polyAs. The produced RNAs were incubated under the same renaturing and refolding conditions, as previously described in 5.3.2, to directly compare the effect of the designed mutations on their *in vitro* RNA dimerisation efficiency. Both U5-AUG mutants, S1 and U5s, had a higher dimer to monomer RNA ratio than WT. Amongst them, S1 had the highest dimerisation efficiency, confirming the Northern Blot data from virion RNA shown in Figure 35.A. The dimerisation mutants, J9+Gag, polyAsw and polyAs had a dramatic increase of their *in vitro* RNA dimerisation

efficiency, with only a small portion of the RNA adopting the monomeric conformation. Overall, the rationally designed mutations presented here, had a positive effect on the propensity of RNA to dimerise *in vitro*.

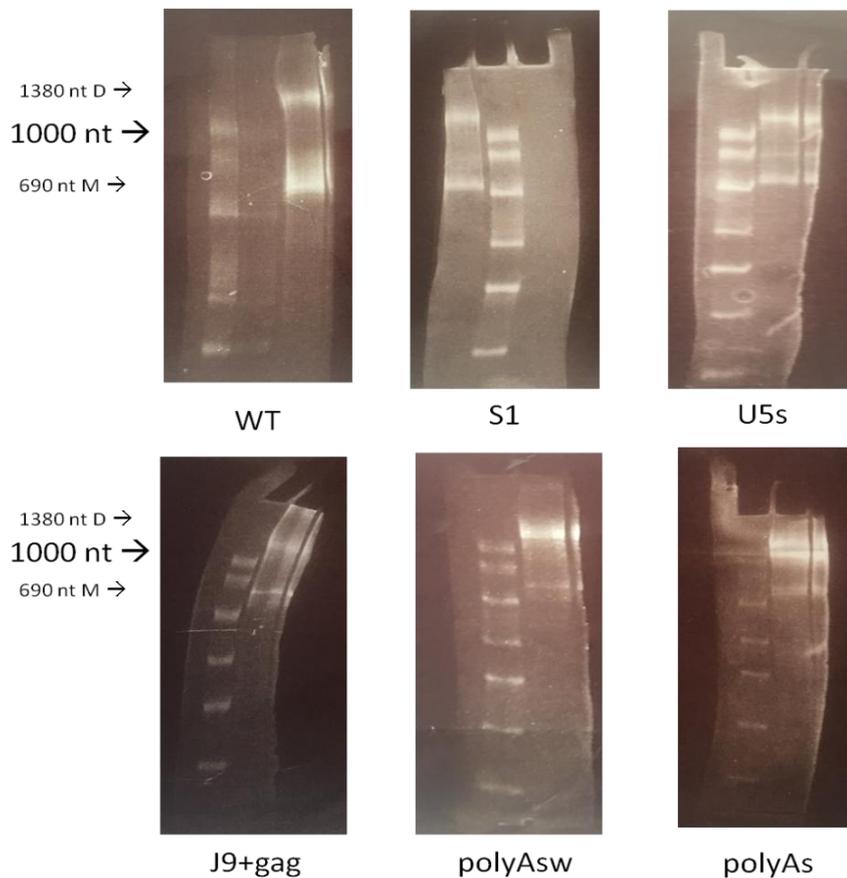


Figure 42. Side by side comparison of *in vitro* RNA dimerisation efficiencies of WT to a range of U5-AUG and dimerisation mutants in 4% polyacrylamide non denaturing TBM gels.

Low Range Riboruler, whose longest RNA fragment is 1000 nt was loaded in all gels to evaluate RNA sizes.

5.4. Chapter conclusions

Rationally designed mutations introduced in the 5'UTR targeted the dimerisation and packaging properties of lentiviral vectors with the aim of improving their efficiency. The effect of these mutations on infectivity was investigated in chapter 4, while in this chapter I aimed to characterise the dimerisation efficiency of transfer vector mutants and attempt to investigate a potential link between transduction efficiency and RNA dimerisation efficiency. Northern Blotting revealed an increased RNA dimerisation propensity trend for S1 and U5s. However, it should be noted that this observed phenotypic difference between the S1 and U5s mutants and WT did not achieve statistical significance. *In vitro* dimerisation assays showed an overall positive effect of the introduced mutations on the propensity of RNA to dimerise *in vitro*. Therefore, these observations indicate that enhancing dimerisation does not automatically lead to improvement of vector infectivity. Northern Blotting was chosen as an appropriate technique to visualise the monomeric and dimeric RNA extracted from virions and measure the quantity of total gRNA levels packaged in virions. However, high yields and lack of degradation throughout the whole process were prerequisites for measurement of total gRNA levels. RT-qPCR was chosen as the best assay to overcome these technical hurdles and study in detail the extent on which the rationally designed mutations affect packaging *per se*. RT-qPCR was used to measure the levels of gRNA successfully encapsidated into virions in comparison to the gRNA found on the producer cells and thus address the next scientific question that is: “what is the relative packaging efficiency (RPE) of WT HIV-1 derived lentiviral vectors to the designed mutant transfer vectors?”

6. Measurement of relative packaging efficiency (RPE) by RT-qPCR

In cells, positive strand RNA viruses like HIV-1, need to selectively recognise their full-length gRNA among similarly modified 5'-capped, 3'-polyadenylated abundant cellular RNAs to successfully assemble and release infectious viral particles⁷⁰. Since RNA packaging is known to be linked with several aspects of retroviral replication, including RNA dimerisation¹², it was rendered essential to determine the RNA packaging efficiency of the rationally designed transfer vector mutants. Northern Blot analysis, ribonuclease protection assay (RPA) and quantitative reverse transcriptase-coupled polymerase chain reaction as well as the most recent sequencing technologies are different approaches for assessing RNA packaging¹²⁸. RT-qPCR was the method of choice to assess packaging efficiency due to its advantage of quantifying RNA expressed at very low levels. Additionally RT-qPCR is highly specific and much more sensitive than the methods mentioned above. In this chapter, packaging efficiency measured with RT-qPCR was employed to investigate whether the dimerisation enhancing mutations promoted packaging as hypothesised. As seen in chapter 4, all designed transfer vector mutants had lower infectious titres. However, this observation did not necessarily mean that the mutants were also of lower packaging efficiency, as the introduced mutations could have improved packaging but hindered a different step of the HIV-1 lifecycle. Lastly, this assay was employed to study the relationship between packaging and RNA dimerisation efficiencies.

6.1. Measurement of the packaging efficiency of rationally designed U5-AUG mutants with a WPRE-targeting RT-qPCR assay

Lizee *et al.*, developed a qPCR assay that targets the WPRE element²³⁴, a sequence inserted in the backbone of lentiviral vectors due to its ability to enhance transgene expression²⁰¹. This assay does not rely on reporter genes but targets a sequence unique to the lentiviral vectors which should thus secure assay specificity. Additionally, unlike most assays available in the literature that target sequences in the 5'UTR region of HIV-1^{246 247 248 249 250}, the assay designed by Lizee *et al.*, targets WPRE which is located in the 3'UTR. Assays targeting sequences in the 5'UTR would be inappropriate for this study because all dimerisation

enhancing mutations introduced in the transfer vectors (described in chapter 4) are located in the 5'UTR and could potentially interfere with assay efficiency. According to Lizee *et al.*, their system allows the quantitation of lentiviral vector preparations at three levels: vector genomic RNA, integrated proviral DNA, and cellular lentiviral vector-mediated gene (mRNA) expression. For the purposes of this thesis the WPRE qPCR assay was employed to measure gRNA levels in both transfected cells and virions. Transfection efficiencies and cellular RNA extraction efficiencies were normalised by normalising to the expression levels of β -actin²⁵¹.

Based on the preliminary data shown in Figure 43, the U5s construct was characterised by a higher packaging efficiency compared to WT and all other mutants. Data were presented in Figure 43 as the average values of the replicates with error bars showing standard deviation. In specific, U5s mutant packaged RNA approximately 4 times more efficiently than WT, whereas the NC4s and the UC4d constructs were characterised by a 10 times and a 2 times lower packaging efficiency respectively.

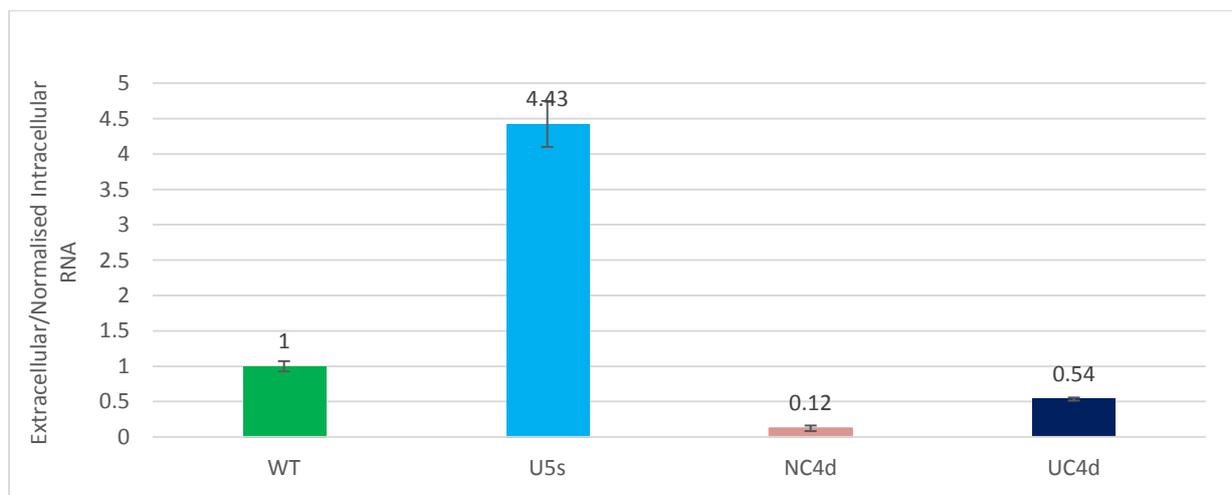


Figure 43. qPCR measuring relative packaging efficiency of mutant vectors against WT.

RT-qPCR using the cDNA from 3 transfection replicates per construct was performed to calculate the relative packaging efficiency, which was defined as the ratio of the extracellular virion RNA to the normalized to β -actin intracellular RNA. RNA was extracted from both virions and cells. Bar graphs show the average packaging efficiency of each construct; error bars represent standard deviation.

RT-qPCR using the cDNA from 12 transfection replicates per construct showed that on average, the packaging efficiency of U5s was 3 times higher than that of WT (Figure 44.A). This result was statistically significant, with a p value of 0.006 using the Student's t test. Similar transfection efficiency between all replicates was achieved via another round of plasmid extraction and purification. Figure 44.B, shows that the data were indicative of enhanced packaging, but that there was also observed variance. This variability was believed to be caused by several technical issues. Performed gRNA extraction was a multiple step process that employed yeast tRNA as a coprecipitant to aid recovery of gRNA, but lacked an internal control necessary to assess gRNA recovery efficiency. Additionally, transfer vector plasmid traces have been challenging to be removed and could potentially interfere with RT-qPCR. These technical issues leading to the observed variability raised concerns about the assay's reproducibility.

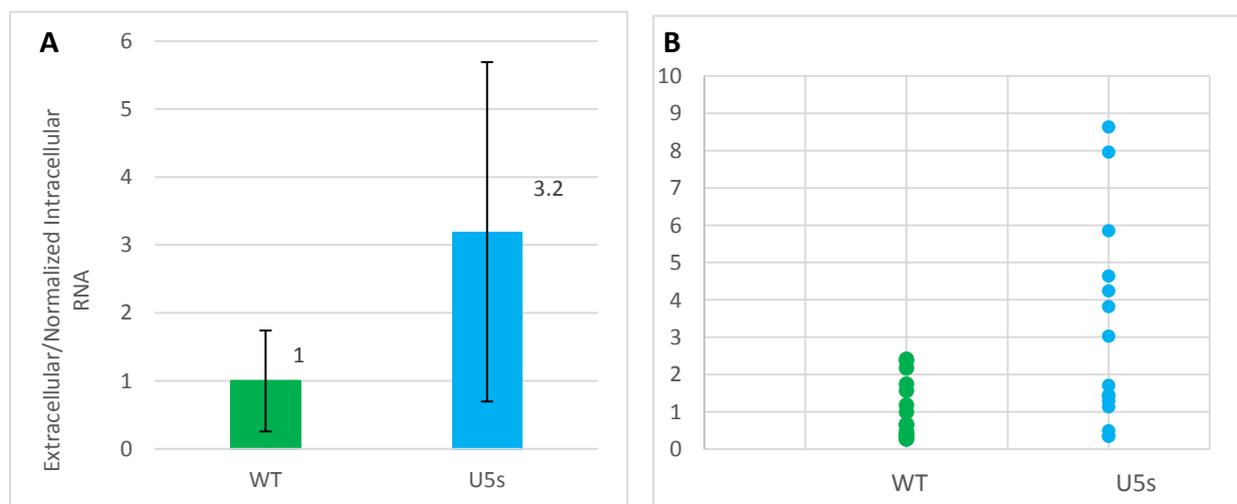


Figure 44. Relative packaging efficiency comparison between WT and U5s.

RT-qPCR experiment comparing the relative packaging efficiency between 20 WT and 18 U5s independent replicates from separate experiments. [A] data presented as an average of all samples. [B] Data presented as individual experiments. Error bars represent standard deviation.

Similarly, RT-qPCR was performed to assess the packaging efficiency of mutant NC4s compared to WT. Using the Student's t test, the results were found to be statistically significant with a p value of 0.012. The NC4s mutant was designed so that the pseudoknot conformation will be favoured while the dimerisation one will be hindered. As expected the RNA encapsidation efficiency of this mutant was severely decreased compared to WT. A similar set of RT-qPCR experiments was set up to compare the packaging efficiency of 6 UC4d replicates against 7 WT replicates. The UC4d mutant was found to be approximately half as efficient as WT in encapsidating RNA into the virion (Figure 45.B). The two sets of data differed significantly, with a p value of 0.04.

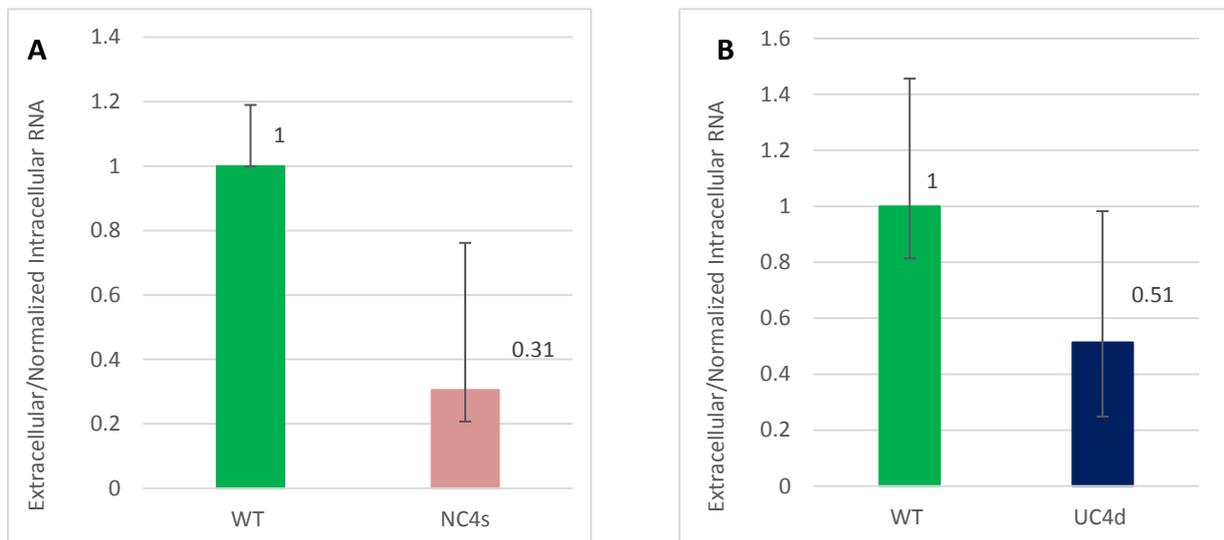


Figure 45. Relative packaging efficiency of WT, NC4s and UC4d with the use of a WPRE-targeting RT-qPCR assay.

[A] Relative packaging efficiency of WT and NC4s mutant qPCR data showing relative packaging efficiency of 7 WT and 6 NC4s samples. Error bars represent standard deviation.

[B] Relative packaging efficiency of UC4d mutant. qPCR data showing relative packaging efficiency of 7 WT and 6 UC4d samples. Error bars represent standard deviation.

6.2. Design of a novel competitive qPCR assay to measure RPE

The WPRE qPCR assay described above was characterised by large variability potentially due to lengthy processing of samples and lack of an internal control for the extracellular RNA. Although retroviruses, including HIV-1, are known to recruit host cell RNAs into virions, the

spectrum of RNAs encapsidated, the amount of specific RNAs packaged and the mechanisms by which they are recruited remain largely unknown; hence the lack of a trustworthy internal control for virion gRNA. An alternative qPCR approach needed to be employed in order to conclusively assess the packaging efficiency ratio for each construct. To achieve this, I designed a competitive qPCR assay by introducing unique sequences in the WT and mutant transfer vectors to allow for simultaneous quantification of their gRNAs in a co-transfection environment. The unique sequences of MS2 (ACATGAGGATCACCCATGT) and BglG (GGATTGTTACTGCATTCGCAGGCAAAACC) were introduced into the WT and mutant sequences and primer/probe pairs were designed using the Oligo 7.0, Molecular Biology Insights software²⁵². The MS2 and BglG sequences were chosen because various studies have employed these stem loops without any noted effect on viral replication^{12 253}. The unique MS2 and BglG sequences were introduced in a multiple cloning site (MCS) between the eGFP and WPRE elements. Cloning of these unique sequences in each of the mutant and WT constructs allowed for independent quantification of their gRNA levels in a co-transfection context. The primer pair was designed to be common for both BglG and MS2 constructs, while differently labelled probes were designed to target the unique MS2 and BglG sequences. This design was important because it could allow for simultaneous detection of WT and mutant transfer vector gRNA levels in a one-well reaction. Each transfer vector construct was cloned twice, to include an MS2 and a BglG version so that the efficiency of each probe could be assessed in a reciprocal system where the results should be the same whichever way around the sequences were tested (MS2-WT and BglG-mutant or vice-versa). The intention of these preliminary results was to assess the validity of the system; when the MS2 and BglG qPCR assays were optimised to equal efficiency, the MS2-containing WT transfer vector and the BglG-containing mutant transfer vectors of interest were used in a competitive environment for further analysis. The experimental design was as follows: WT and mutant gRNA vectors were co-transfected into 293T cells and RT-qPCR targeting MS2 and BglG sequences was performed on each sample enabling a comparison of their encapsidation efficiencies to be performed in a controlled manner. The introduction and targeting of unique sequences in the WT and mutant transfer vectors aimed to resolve the observed variability (potentially due to different sample processing and lack of an internal control for the extracellular RNA).

BglG

REV

5' –

aggagaaaatgaaagccatacgggaagcaatagcatgatatacaaggcattaagcagcgtatccacat
agcgtaaaaggagcaacatagttaagaataccagtcfaatctttcacaattttgtaatccagaggttg
attgcatcgGGTTTTGCCTGCCAATGCAGTAACAATCCcgaatcgagtactcacgtgtaattaagat
atcttatctagatccgggtggatcccgggcccgcggtaccgtcgactgcagaattcgaagcttgagctc
gagatctgagtcggacttgtacagctcgtccatgcccagagtgatcccggcggcggtcacgaactcc
agcaggaccatgtgatcgcgcttctcgttgg–3'

FWD

5' –

ccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgcgcgggatcactctcggcatg
gacgagctgtacaagtccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaaccg
gggccgggatccaccggatctagataagatatcttaattaacacgtgagtactcgatcgGGATTGTT
ACTGCATTGCAGGCAAAACCcgatcgcaatcaacctctggattacaaaatttgtaagattgactg
gtattcttaactatggtgctccttttacgctatgtggatcgcgtgctttaatgcctttgtatcatgct
attgcttcccgtatggctttcattttctcct–3'

MS2

REV

5' –

gtaaaaggagcaacatagttaagaataccagtcfaatctttcacaattttgtaatccagaggttgatt
gcatcgACATGGGTGATCCTCATGCcgaatcgagtactcacgtgtaattaagatatcttatctagat
ccggtggatcccgggcccgcggtaccgtcga–3'

FWD

5' –

gacgagctgtacaagtccggactcagatctcgagctcaagcttcgaattctgcagtcgacggtaaccg
gggccgggatccaccggatctagataagatatcttaattaacacgtgagtactcgaatcgGCATGAGG
ATCACCCATGTcgatcgcaatcaacctctggattacaaaatttgtaagattgactggtattcttaa
ctatggtgctccttttacgctatgtggatcgcgtgctttaatgcctttgtatcatgc–3'

Figure 46. Design of a multiplex qPCR assay consisting of a common primer set and two TaqMan probes targeting the unique sequences inserted within the transfer vector.

The BglG and MS2 assays were designed using the Oligo 7.0, Molecular Biology Insights software²⁵². More specifically forward primer was 20 nt long, had a T_m of 55°C and was of 50% GC context. The reverse primer was 18 nt long, had a T_m of 54.6°C and was of 50% GC context. The sequence fragment presented here included part of the eGFP and WPRE sequences, and the area of the MCS within which the unique sequences was inserted. The BglG and MS2 sequences were introduced following the same cloning strategy so that a common set of primers could be developed. This sequence fragment of interest is presented

both in its forward and reverse direction. The primer sequences are highlighted in yellow, the inserted BglG and MS2 sequences are highlighted in purple and transcribed in capital letters, and the designed probes spanning the unique BglG and MS2 sequences are underlined, italicised and in bold. The BglG probe was designed to be 23 nt long, had a T_m of 65.3°C and was of 48% GC context, while the MS2 probe was 22 nt long, had a T_m of 64.9°C and was of 55% GC context. The assay sequences designed are: Fw 5'-GAATTCTGCAGTCGACGGTA-3', Rv 5'-TCCAGA GGTGATTGCGA-3', PBglG 5'-[6'FAM]-TGCGAATGCAGTAACAATCCCGA-[BHQ1]-3', PMS2 5'-[HEX]-CATGGGTGATCCTCAT GCCGAT-[BHQ1]-3'.

6.3. Cloning of transfer vectors containing the unique BglG and MS2 sequences

1 µg of WT pCCL-eGFP and 1 µg of U5s pCCL-eGFP backbones were incubated with 20 U of FastDigest *PvuI* and 1 U of FastAP Thermosensitive Alkaline Phosphatase in a 20 µL reaction in 1x Fast Digest Buffer at 37°C for 1 h, to achieve simultaneous dephosphorylation and digestion. Digestion products were separated on a 1% agarose gel and bands of interest were excised and extracted (Figure 47). Meanwhile BglG and MS2 phosphorylated oligo sequences flanked between the *PvuI* recognition sites were used in an annealing reaction that contained 100 µM Fw Oligo, 100 µM Rv Oligo, and 1x ExpressLink™ T4 Buffer in a 100 µL final volume reaction that was held at 90°C for 10 min. The extracted bands of interest derived upon digestion of the transfer vectors with *PvuI* were treated similarly. Four ligation reactions of the following combinations were set up: WT pCCL-eGFP and BglG oligos, WT pCCL-eGFP and MS2 oligos, U5s pCCL-eGFP and BglG oligos and U5s pCCL-eGFP and MS2 oligos. Each ligation reaction contained 35 ng of annealed oligos, 100 ng backbone and 5 U of ExpressLink™ T4 DNA Ligase in a final volume of 20 µL. 1 µL of each ligation reaction containing a combination of backbone and oligos was used to transform StbI3™ *E.coli* cells. Successful incorporation of the BglG and MS2 unique sequences in both transfer vectors was confirmed with sequencing (data not shown).

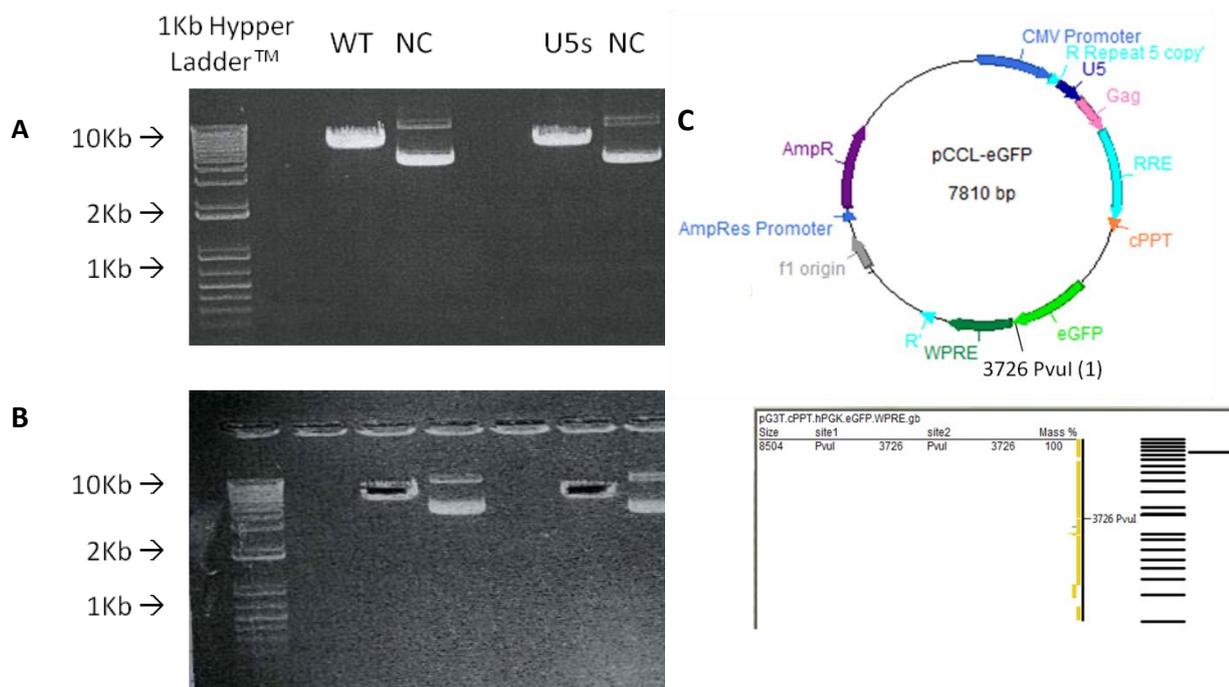


Figure 47. Cloning of transfer vectors containing the unique BglG and MS2 sequences.

[A] *PvuI* digested WT and U5s pCCL-eGFP vector plasmids were loaded alongside control undigested (NC) WT and U5s pCCL-eGFP vector plasmids in a 1% agarose gel. **[B]** *PvuI* digested WT and U5s pCCL-eGFP fragments were excised and purified. **[C]** pCCL-eGFP transfer vector plasmid map presented with the single *PvuI* restriction site located between eGFP and WPRE. *PvuI* cuts the pCCL-eGFP vector plasmid at nucleotide 3726 producing an 8.5 kb long linear DNA fragment.

6.4. Optimisation of a competitive qPCR assay to measure RPE

According to the The MIQE Guidelines (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) PCR efficiency is correlated with robust and precise qPCR assays²⁵⁴, and can be calculated based on the slope of the standard curve using the Thermo Fisher qPCR efficiency calculator²⁵⁵. The linear dynamic range should extend to 5 or 6 log₁₀ concentrations and must include the interval for the target nucleic acids being quantified. Correlation coefficients (R^2 values) need to be calculated to ensure a strong linear relationship between two variable points of the standard curve and should be >0.99²⁵⁶. The limit of detection (LOD) is defined as the lowest concentration at which 95% of the positive samples are detected. Empirically, samples should be of Ct<30 to be accurately measured,

as qPCR precision decreases with the copy number, and samples of Ct>35 will be considered negative²⁵⁶.

TaqMan® Assays exploit the 5' nuclease chemistry of polymerases. Fluorogenic probes are used to detect specific PCR products as they accumulate during PCR. Such assays contain: a pair of unlabeled primers, a TaqMan probe commonly 5' labelled with a FAM™ or VIC™ dye, a minor groove binder (MGB) and nonfluorescent quencher (NFQ) on the 3' end. The process begins with a temperature raise to denature the double-stranded cDNA. At this stage the signal from the 5' fluorescent dye of the TaqMan probe is quenched by the presence of a NFQ on the 3' end. The next stage involves decrease of temperature to enable primers and probes to anneal specifically to their target sequences. On the last step of extension, Taq DNA Polymerase synthesises new products with the use of unlabeled primers and the template whose expression is to be studied. When the polymerase reaches the annealed TaqMan probe, its endogenous 5' nuclease activity cleaves it resulting in separation of the dye from the quencher. Product amplification leads to more dye molecule release, resulting in an increase in fluorescence intensity proportional to the amount of amplicon synthesized²⁵⁷.

Calibration of primer concentration was the first step towards optimisation of the new qPCR assay. The MS2 probe concentration was kept constant at 250 nM, while a gradient of 100 nM-900 nM primer concentration combinations was used to create a matrix, presented below, to identify the optimal primer concentration. Optimal primer concentration was chosen based on the sensitivity which was assessed based on how early the sequence of interest was detected in the qPCR cycles, but also based on the specificity of primers which was addressed by the negative control reactions.

| Fw (100nM) 1:1000 dil from stock | | | | | |
|----------------------------------|-----|-----|-----|-----|-----|
| Rv (100nM) | 100 | 300 | 500 | 700 | 900 |
| 100 | | | | | |
| 300 | | | | | |
| 500 | | | | | |
| 700 | | | | | |
| 900 | | | | | |
| 250 nM MS2 probe | | | | | |

| per reaction (ul) | |
|-------------------|------|
| FA Buffer | 5 |
| Fw Primer | |
| Rv Primer | |
| Probe | 0.25 |
| Plasmid DNA | 2 |
| dH2O | 1.75 |

C

| | | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 100/100 | 100/100 | 100/300 | 100/300 | 100/500 | 100/500 | 100/700 | 100/700 | 100/900 | 100/900 |
| 300/100 | 300/100 | 300/300 | 300/300 | 300/500 | 300/500 | 300/700 | 300/700 | 300/900 | 300/900 |
| 500/100 | 500/100 | 500/300 | 500/300 | 500/500 | 500/500 | 500/700 | 500/700 | 500/900 | 500/900 |
| 700/100 | 700/100 | 700/300 | 700/300 | 700/500 | 700/500 | 700/700 | 700/700 | 700/900 | 700/900 |
| 900/100 | 900/100 | 900/300 | 900/300 | 900/500 | 900/500 | 900/700 | 900/700 | 900/900 | 900/900 |
| 100/100 | 100/300 | 100/500 | 100/700 | 100/900 | 300/100 | 300/300 | 300/500 | 300/700 | 300/900 |
| 500/100 | 500/300 | 500/500 | 500/700 | 500/900 | 700/100 | 700/300 | 700/500 | 700/700 | 700/900 |
| 900/100 | 900/300 | 900/500 | 900/700 | 900/900 | | | | | |

Figure 48. Primer optimisation process for the competitive MS2/BSL RT-qPCR assay.

[A] Table presenting the matrix used to determine optimal primer concentration while using the HEX-labelled MS2 probe at a constant concentration of 250 nM. Numbers presented in this table represent [Forward primer concentration/Reverse primer concentration] measured in nM. **[B]** Reaction mixture for primer concentration optimisation. **[C]** Plate presenting the positions of the different primer concentration combinations. 5000000 copies of the pCCL-eGFP-MS2 transfer vector were used as a template for each reaction (shown in black), while 2 µl of water were used for the NC reactions (shown in blue). DNA samples were loaded in the black labelled wells, while water was loaded in the blue labelled ones.

By independently varying the primer concentrations, one could identify the optimal primer set concentration to achieve the best qPCR assay performance. Based on observations from these experiments the [500/500] ratio was found to be the optimal primer concentration that achieved both a low Ct value and lack of product formation in the respective NC (data not shown). Using this primer set concentration; probe concentrations were then optimised. 250 nM, 400 nM and 600 nM were the primer concentrations trialled for TaqMan assay optimisation. WT-pCCL-eGFP-MS2 and WT-pCCL-eGFP-BglG plasmids were used as standard

curve templates in independent monoplex reactions to interrogate the efficiency and linearity of the assays. A six \log_{10} concentration standard curve between 5000000 and 500 copies was created following the protocol published by Applied Biosystems²⁵⁸. As shown in Figure 49, all BglG probe concentrations showed variability between 500000 and 5000 copy numbers, without any of the probe concentrations achieving an acceptable level of linearity for BglG; all probe concentrations generated an R^2 that was lower than 0.99. On the contrary, all MS2 probe concentrations had an $R^2 > 0.99$, with [250 nM] giving an $R^2=1$! The PCR efficiency was $\sim 98\%$ for all MS2 reactions and 97-100% for the BglG reactions.

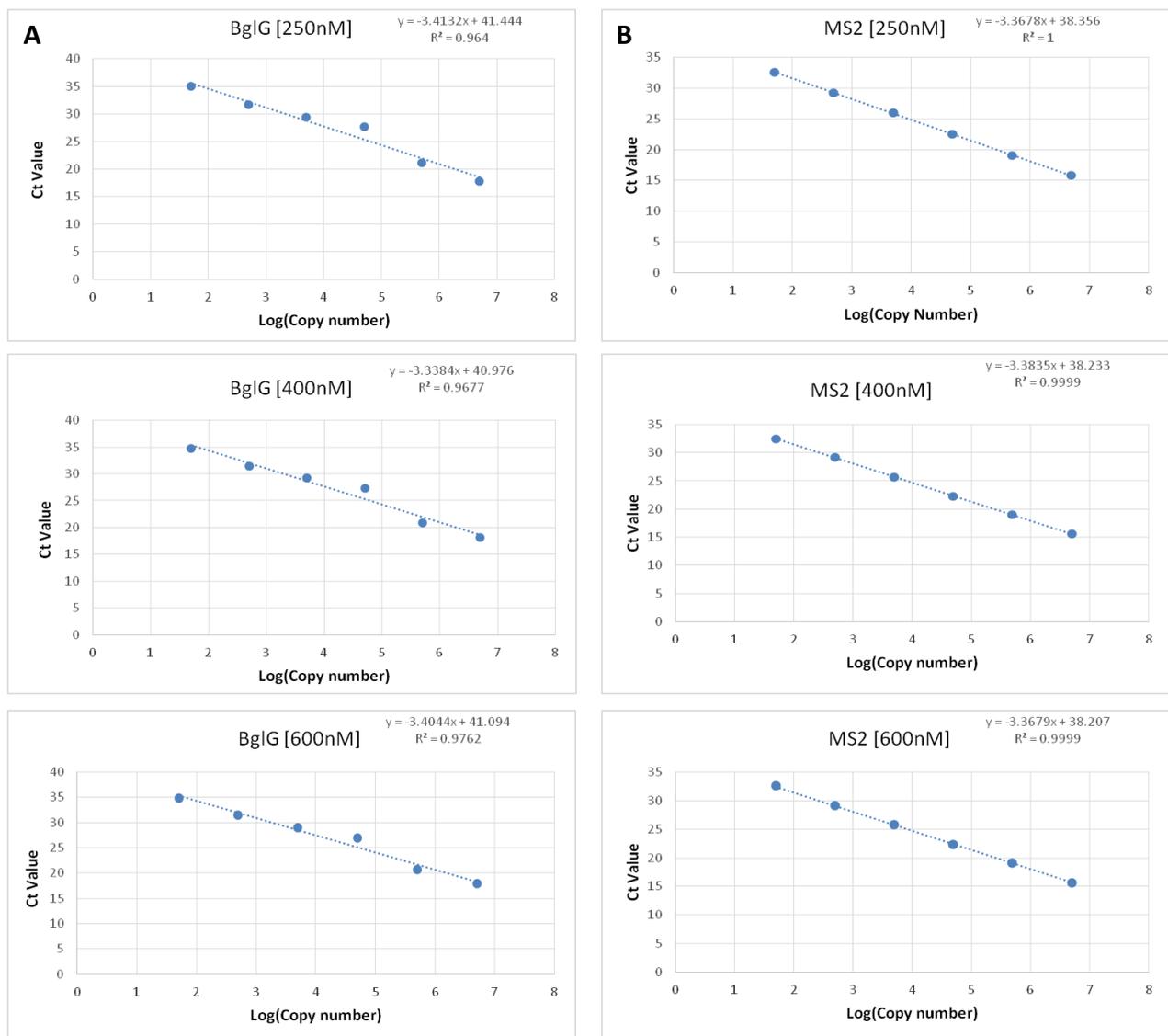


Figure 49. Optimisation of BglG and MS2 probes' concentration.

[A] Assessment of BglG probe concentration was performed trying three different concentrations, including 250 nM, 400 nM and 600 nM. Primer concentration was maintained at [500/500] for all reactions. **[B]** The effect of MS2 probe concentration on qPCR efficiency and linearity was studied by setting reactions with variant MS2 probe concentration including 200 nM, 400 nM and 600 nM while maintaining primer concentration at 500 nM.

The effect of the BglG and MS2 probes in a multiplex reaction was investigated next. A range of 5000000 to 500 copies of both WT-pCCL-eGFP-BglG and WT-pCCL-eGFP-MS2 plasmids were used as standard curve templates in a multiplex reaction containing 250 nM MS2 probe, 600 nM BglG probe, 500 nM Forward primer, 500 nM Reverse primer and 5 µl of Fast Advanced TaqMan buffer to assess the linearity of the BglG and MS2 standard curves and the efficiency of the respective assays. The presence of pCCL-eGFP-MS2 template was shown to negatively affect the linearity of the BglG qPCR assay (Figure 50), with the R² dropping from 0.97 to 0.51. Additionally, the efficiency of the assay was also observed to decrease. Given the high efficiency of the MS2 assay, it was concluded that it was the BglG probe, rather than the primers, whose sequence needed redesigning to optimise this multiplex reaction.

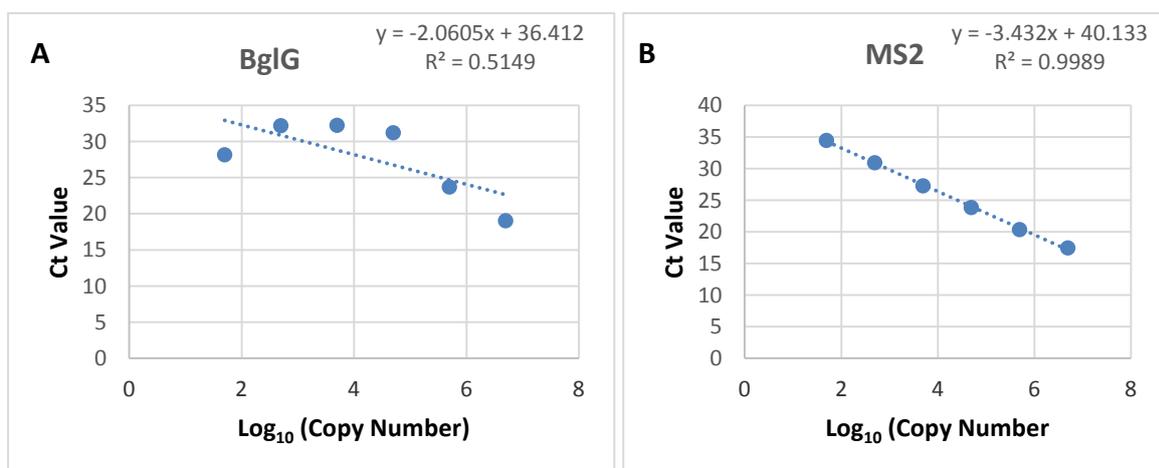


Figure 50. Investigation of the efficiency and linearity of the duplex BglG-MS2 qPCR assay.

Multiplex reaction set up with 600 nM BglG probe, 250 nM MS2 probe, 500 nM Forward primer and 500 Reverse primer for a range of 5000000 to 500 copies of both WT-pCCL-eGFP-BglG and WT-pCCL-eGFP-MS2 plasmids.

The Custom TaqMan™ MGB Probe design service by Applied Biosystems²⁵⁹ was used to design a new FAM-labelled BglG probe. TCGATCGGGATTGTTACTG was the new probe sequence used to optimise a new BglG assay, using a concentration of 750 nM as directed by the manufacturer. Keeping the primer concentration at 500 nM as previously established, independent monoplex reactions for MS2 (250nM) and new AB BglG (750 nM) were repeated following the protocol presented in Figure 48.B. As shown in Figure 51, the new probe sequence had a significant positive effect on the assay's linearity, with its R² improved to be higher than 0.99. The linearity of the MS2 was maintained higher than 0.99 as shown previously. The PCR efficiency was lower than previously described for both assays, but was noted to be comparably reduced between BglG and MS2, hence it was assumed that this was due to a technical error. Nevertheless, overall the new AB BglG probe design had a positive effect on the assay and was used to further optimise the multiplex assay.

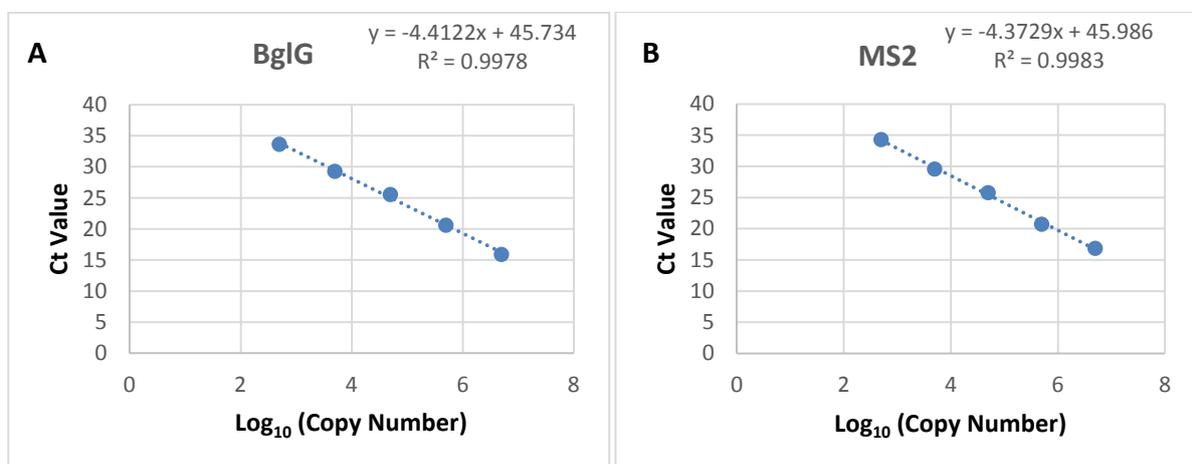


Figure 51. Investigation of new BglG assay linearity and efficiency.

[A] Effect of the new BglG probe sequence on assay linearity and efficiency. Monoplex reaction set up with 750 nM new AB BglG probe, 500 nM Forward primer and 500 Reverse primer. 5000000 to 500 copies of WT-pCCL-eGFP-BglG plasmid were used as standard curve templates. **[B]** Repeat of the previously established MS2 monoplex assay conditions to ensure reproducibility and direct comparison to the new BglG assay efficiency.

Next, it was necessary to investigate the effect of primer concentration on the multiplex assay, as both BglG and MS2 assays were previously optimised in monoplex reactions. I compared the effect of 500 nM primer concentration, which is the optimal primer concentration established in the monoplex assays previously described, against 1000 nM primer concentration. The data was inconclusive as the linearity of the MS2 assay (red) increased with the increase of primer concentration whereas its efficiency decreased. On the contrary, the efficiency of the new AB BglG assay (blue) increased with the increase of the primer concentration, while the standard curve's linearity was noted to decrease.

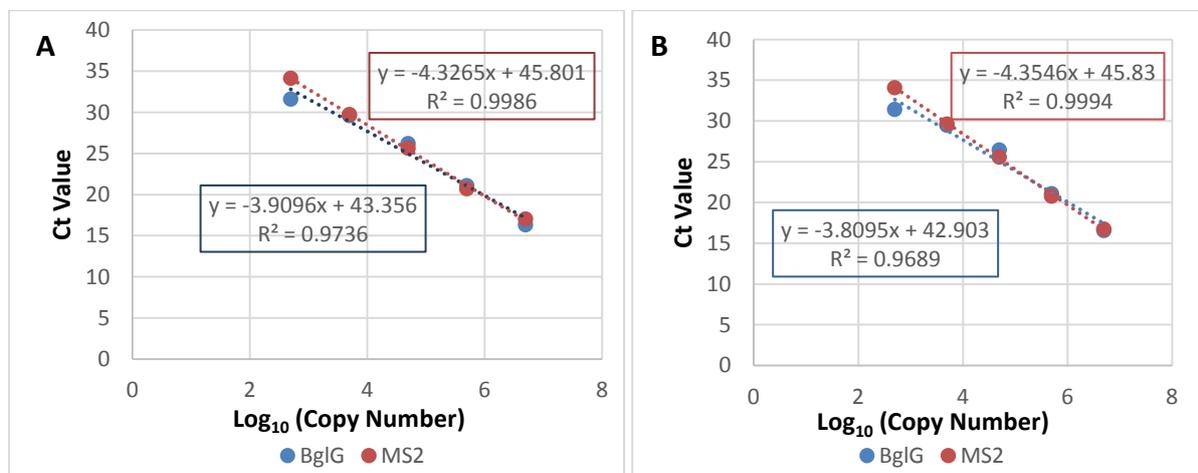


Figure 52. Effect of primer concentration on the multiplex assay.

[A] Multiplex reaction set up with 750 nM new AB BglG probe, 250 nM MS2 probe, 500 nM Forward primer, 500 nM Reverse primer and 5 μ l Fast Advanced Buffer. **[B]** Multiplex reaction set up with 750 nM new AB BglG probe, 250 nM MS2 probe, 1000 nM Forward primer, 1000 nM Reverse primer and 5 μ l Fast Advanced Buffer.

Given the lack of conclusive data, the experiment was repeated with an additional reaction of increased primer concentration to 1500 nM. The data presented in Figure 53 was inconsistent with the data presented in Figure 52. The increase of primer concentration to 1500 nM affected both assays' linearity positively but both assays' efficiency negatively. A possible interpretation of this result could be that detection of MS2 and BglG containing sequences was stochastic in a multiplex environment.

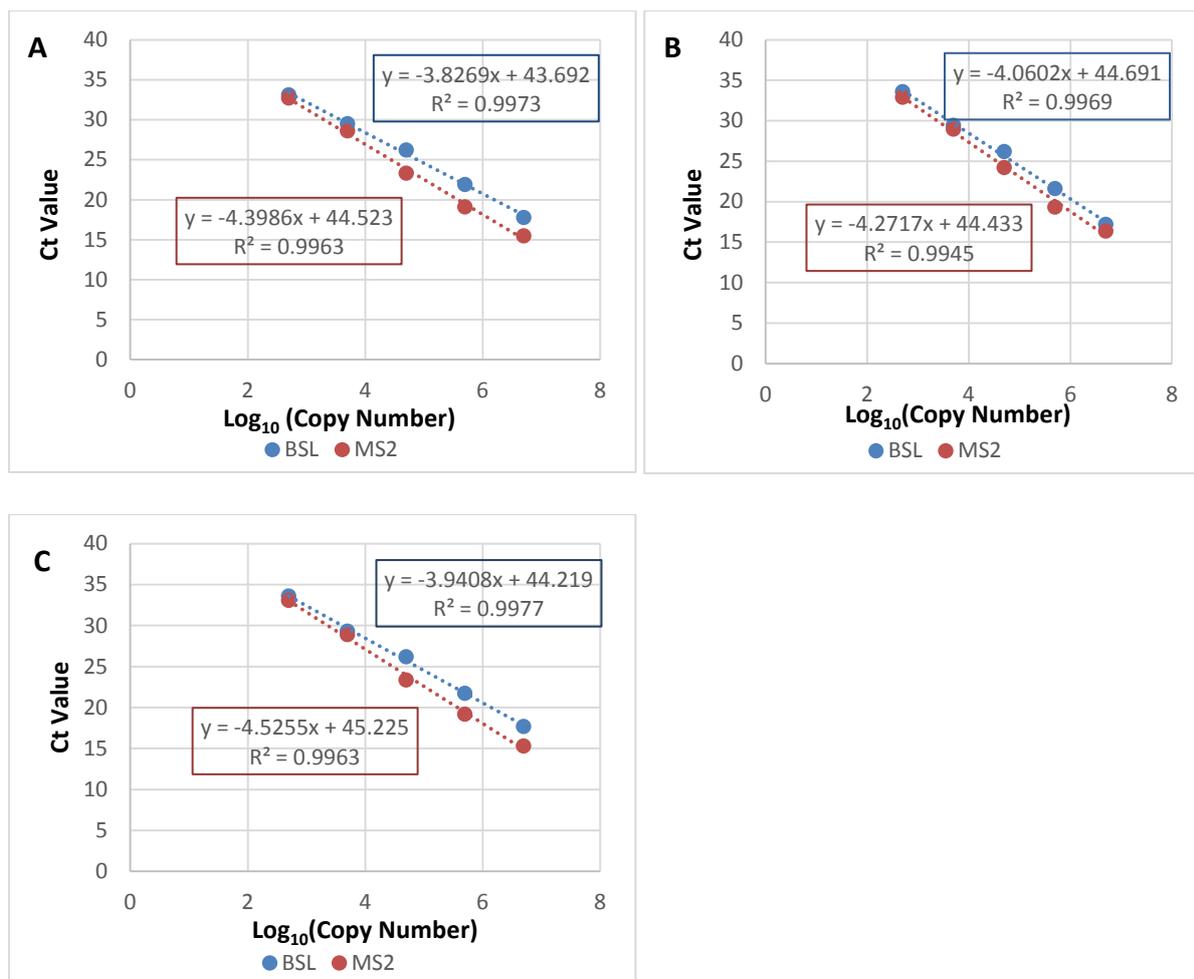
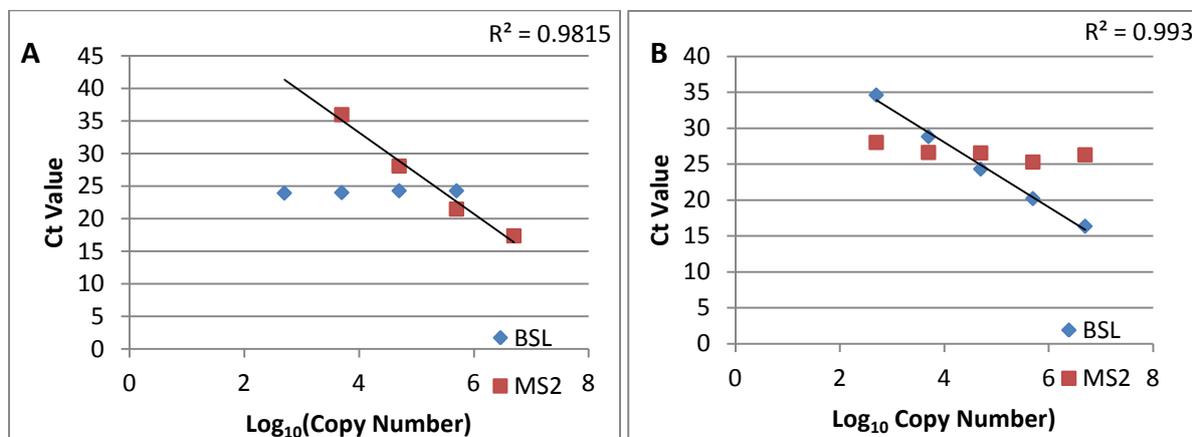


Figure 53. Effect of primer concentration on the multiplex assay.

[A] Multiplex reaction set up with 750 nM new AB BglG probe, 250 nM MS2 probe, 500 nM Forward primer, 500 nM Reverse primer and 5 μ l Fast Advanced Buffer. **[B]** Multiplex reaction set up with 750 nM new AB BglG probe, 250 nM MS2 probe, 1000 nM Forward primer, 1000 nM Reverse primer and 5 μ l Fast Advanced Buffer. **[C]** Multiplex reaction set up with 750 nM new AB BglG probe, 250 nM MS2 probe, 1500 nM Forward primer, 1500 nM Reverse primer and 5 μ l Fast Advanced Buffer.

To address this challenge, a new set of experiments was designed maintaining the plasmid concentration of WT-pCCL-eGFP-BglG at a constant of 50000 copies in the presence of a \log_{10} WT-pCCL-eGFP-MS2 standard curve and vice versa. The linearity of the assays was assessed with the calculated R^2 and the Δ Ct of each standard curve. This time the efficiency of each assay was investigated by the Ct consistency of the plasmid whose DNA mass was kept constant. The Δ Ct of a 10x sample dilution was expected to be around 3.2. As shown in Figure 54, the Δ Ct values of the MS2 standard curve in the presence of a constant amount of

50000 BglG copy number were between 4.1 and 7.9, which was of very poor efficiency. Respectively, the Ct difference of values of the BglG standard curve in the presence of a constant amount of 50000 MS2 copy number was between 3.8 and 5.7, which was also determined to be of poor efficiency. As expected, the Ct value of the BglG plasmid in the presence of 50000 copies was maintained at 24 between 5000000 and 500 MS2 copies. Similarly, the Ct value of the MS2 plasmid kept at 50000 copies was measured between 25.3 and 26.6 between 5000000 and 5000 BglG copies but at 28 in the presence of 500 BglG copies.



| C | Log ₁₀ (CP) | MS2 | ΔCt | BglG | Log ₁₀ (CP) | MS2 | BglG | ΔCt |
|---|------------------------|-----------|-----------|-----------|------------------------|-----------|-----------|-----------|
| | 6.69897 | 17.377204 | 4.1000261 | - | 6.69897 | 26.307735 | 16.372656 | 3.842514 |
| | 5.69897 | 21.47723 | 6.6052074 | 24.289837 | 5.69897 | 25.295114 | 20.21517 | 4.1127205 |
| | 4.69897 | 28.082438 | 7.8894901 | 24.295522 | 4.69897 | 26.572425 | 24.32789 | 4.5126457 |
| | 3.69897 | 35.971928 | | 24.012357 | 3.69897 | 26.632078 | 28.840536 | 5.7923031 |
| | 2.69897 | - | | 23.936703 | 2.69897 | 28.02824 | 34.632839 | |

Figure 54. Effect of constant and variable plasmid concentrations on MS2 and BglG assays.

[A] Effect of WT-pCCL-eGFP-BglG plasmid presence on a log₁₀ WT-pCCL-eGFP-MS2 standard curve. **[B]** Effect of WT-pCCL-eGFP-MS2 plasmid presence on a log₁₀ WT-pCCL-eGFP-MS2 standard curve. Multiplex reactions presented in A and B were set up with 750 nM new AB BglG probe, 250 nM MS2 probe, 1000 nM Forward primer, 1000 nM Reverse primer. **[C]** Raw data of MS2 and BglG assays presented in graphs A and B to study the assays' efficiency and linearity.

Given the data presented in Figure 54, I addressed the aforementioned challenges by comparing directly the linearity and efficiency of the multiplex assay and the two respective monoplex assays in the same plate.

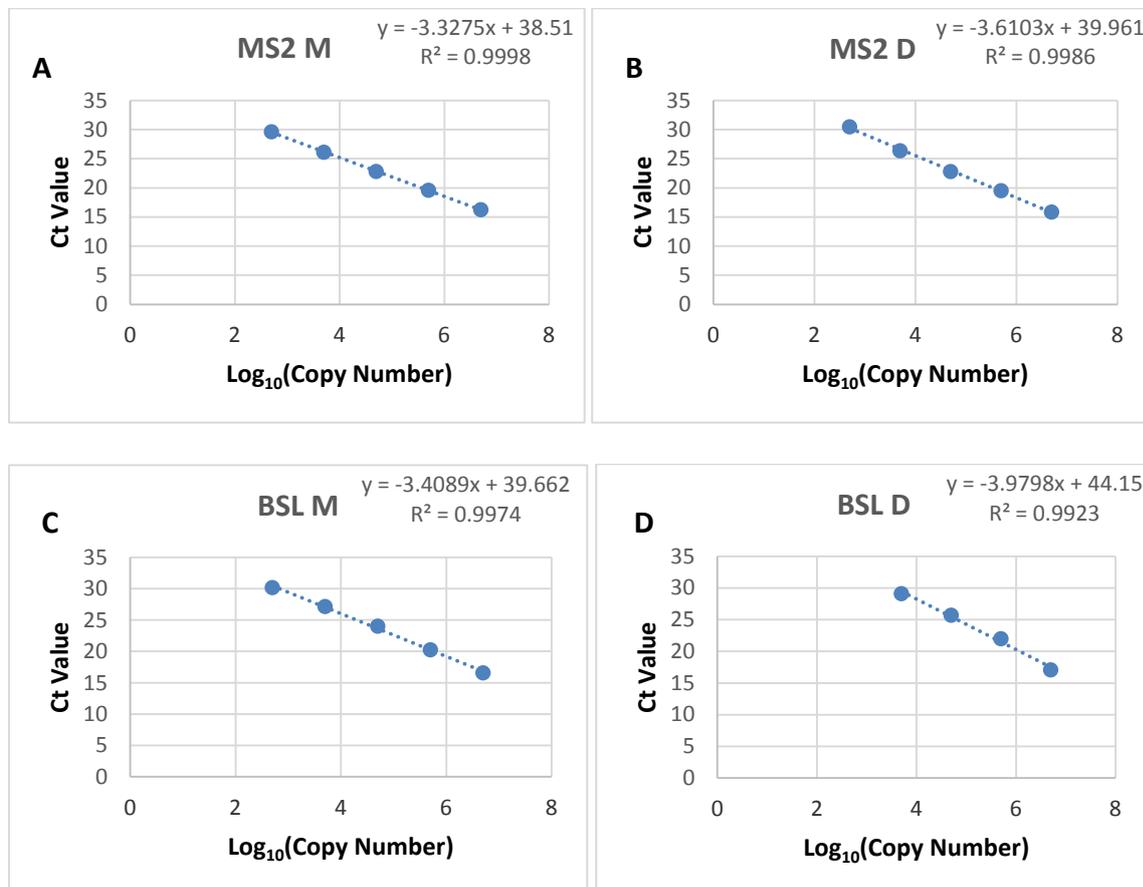


Figure 55. Comparison of monoplex and duplex reactions for MS2 and BglG assays.

[A] Graph showing a five log₁₀ standard curve of the MS2 assay in a monoplex reaction containing 250 nM MS2 probe, 500 nM Forward primer and 500 nM Reverse primer. **[B]** Graph showing a five log₁₀ standard curve of the MS2 assay in a multiplex BglG-MS2 reaction containing 750 nM new AB BglG probe, 250 nM MS2 probe, 1000 nM Forward primer and 1000 nM Reverse primer. **[C]** Graph showing a five log₁₀ standard curve of the BglG assay in a monoplex reaction containing 750 nM new AB BglG probe, 500 nM Forward primer and 500 nM Reverse primer. **[D]** Graph showing a five log₁₀ standard curve of the BglG assay in a BglG-MS2 multiplex reaction containing 750 nM new AB BglG probe, 250 nM MS2 probe, 1000 nM Forward primer and 1000 nM Reverse primer. Reactions presented in B and D contain both MS2 and BglG plasmid templates, while A and C contain only MS2 and BglG plasmid template respectively.

The MS2 monoplex qPCR efficiency was calculated to be 99.77% with an amplification factor of 2 and an R^2 of 0.999, while the equivalent MS2 multiplex qPCR efficiency dropped to 89.23% with an amplification factor of 1.89 and an R^2 of 0.998. Similarly, The BglG monoplex qPCR efficiency was 96.49% with an amplification factor of 1.96 and an R^2 of 0.997, while the equivalent BglG multiplex qPCR efficiency was decreased to 78.35% with an amplification factor of 1.78 and an R^2 of 0.992. Collectively, the data suggest that the equivalent monoplex assays were more efficient than the multiplex reaction. The multiplex assay was initially of preference as it would have been time and cost efficient but was not found as accurate as the monoplex assays. Therefore, the independent use of the optimised monoplex MS2 and BglG assays was chosen for accurate measurement of MS2- and BglG- sequence containing genetic material derived from the same co-transfection plate. The next section discusses the employment of these two monoplex assays (conditions described in Figure 55.A and 55.C) to measure MS2- and BglG- containing gRNA and the methodology followed to optimise the protocol for genetic material extraction and data analysis.

6.5. Optimisation of a qPCR protocol for genetic material extraction and data analysis to measure RPE

The main advantage offered by this new PCR approach was the ability to measure WT and mutant gRNA levels in the same sample, and thus an equal amount of WT and mutant plasmid DNA mass was used to co-transfect HEK293T (Figure 56). 6 μ g WT pCCL-eGFP-MS2, 6 μ g mutant pCCL-eGFP-BglG, 2.5 μ g pRev, 3.5 μ g pVSVg and 5 μ g pSYNGP were used to produce lentiviral vector following the protocol described in section 3.1. Intracellular and extracellular RNA was extracted from 1 mL cellular lysate and 1 mL of viral supernatant respectively, and used as template for reverse transcription. qPCR targeting the MS2 and BglG sequences was then performed to measure the relative packaging efficiency of WT to mutant vector.

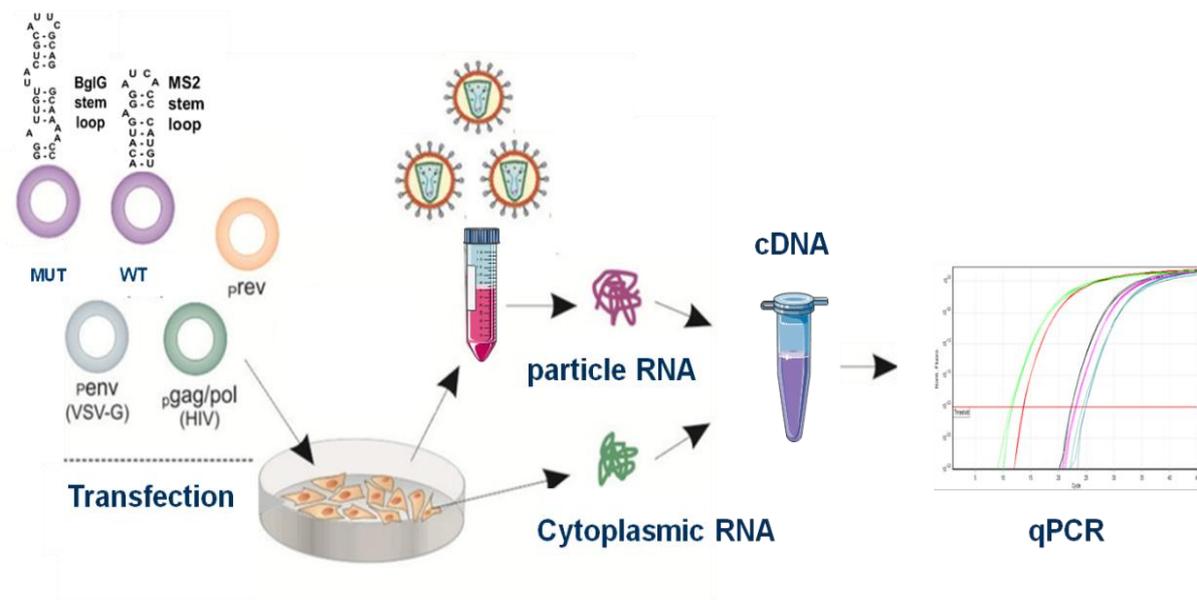


Figure 56. Representation of qPCR workflow to measure relative packaging efficiency.

Relative packaging efficiency was defined as the ratio of extracellular to intracellular WT gRNA divided by the ratio of extracellular to intracellular mutant gRNA derived from the same co-transfection plate. This method overcame the necessity to normalise samples to a housekeeping gene as the genetic material from both WT and mutant vectors came from the same transfection environment and was already normalised to the transfection efficiency of each vector by dividing the encapsidated gRNA to the intracellularly expressed gRNA.

The efficiency of the BglG and MS2 assays was previously studied in section 6.4 with the use of plasmid DNA. I then sought to investigate whether these assays were also equally efficient at detecting BglG- and MS2-containing cDNA produced from extracted gRNA. This was important to ensure unbiased and accurate measurement of WT and mutant gRNA packaging regardless of the sequence “attached” to them. To study this, HEK293T cells were co-transfected with equal amounts of WT-pCCL-eGFP-BglG and WT-pCCL-eGFP-MS2 transfer vector plasmids. The WTMS2/WTBglG RPE should be of value 1, as both plasmids contained identical WT sequences and were therefore expected to have the same packaging efficiency.

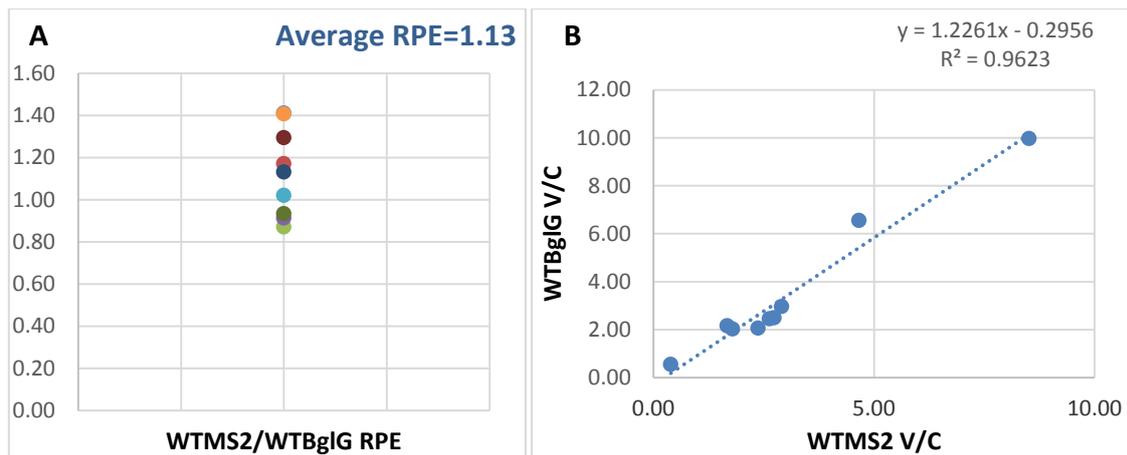


Figure 57. Measurement of WTMS2/WTBglG RPE.

[A] Graph showing the measured WTMS2/WTBglG RPE values from 8 independent transfection replicates from two rounds of experiments. Average RPE was calculated to be 1.13. **[B]** Plot of WTBglG Vector RNA/Cellular RNA to WTMS2 Vector RNA/Cellular RNA showing linear distribution of samples.

Upon confirmation that WTMS2/WTBglG RPE indeed equalled to a value close to 1 when studied and analysed with the optimised qPCR monoplex assays (Figure 57), I proceeded to calculating the relative packaging efficiency of a known packaging deficient mutant to address the LOD of the assays. Lever *et al.*, had created $\Delta P1$, a mutant with a 19 bp deletion in the Ψ region, whose packaging efficiency was measured to be less than 2% of that of the WT virus by RNA dot blots. $\Delta P1$ -pCCL-eGFP transfer vector was created exploiting the ability of *E.coli* to recombine DNA with homologous ends. Primer $\Delta P1$ Rv ccactctctctccttctagcc**ggcgctactcaccagtcgccc** was designed as the reverse complement of the sequence corresponding to 20 bases upstream of the plasmid DNA to be deleted, followed by 20 bases equal to the downstream sequence. Primer $\Delta P1$ Fw cggcgactggtgagtagcgc**ggctgaaggagagagatgggtg** was designed in the same way but corresponding to the complementary strand. These primers were used in a 50 μ l reaction in a final concentration of 0.2 μ M also containing 10 ng WT-pCCL-eGFP-BglG, 2 U *Pfu* Polymerase, 1x Buffer with $MgSO_4$ and 200 μ M dNTP mix²⁶⁰. The thermal cycling conditions were: 95°C for 2 min, 35 cycles of: 95°C for 1 min, 60°C for 30 sec, 68°C for 4 min and a final extension step at 68°C for 5 min. 10 μ l of the reaction were treated with *DpnI* for 1 h at 37°C to degrade any unwanted plasmid template. 1 μ l of the *DpnI* treated PCR mix was used for transformation of XL10-Gold® Ultracompetent Cells as per the manufacturer's

instructions. Plasmid DNA was extracted from grown colonies with the use of the Qiagen mini-prep kit. Deletion of the 19 bp sequence was confirmed with sequencing (data not shown) and the $\Delta P1$ -pCCL-eGFP-BglG plasmid was used for co-transfections. Analysis of WTMS2/ $\Delta P1$ BglG revealed that the relative packaging efficiency ratio of WT to $\Delta P1$ was 64.4 or alternatively that $\Delta P1$ had 1.6% the packaging efficiency of WT. This result confirmed the previously published data on the packaging efficiency of $\Delta P1$ (Figure 58).

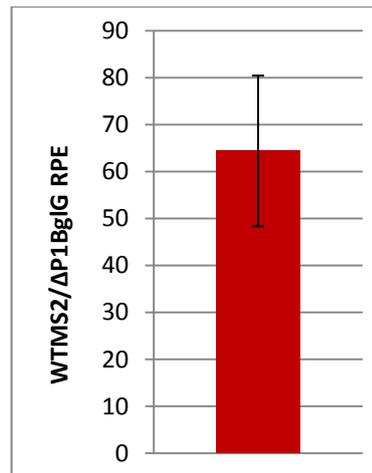


Figure 58. Measurement of WTMS2/ $\Delta P1$ BglG RPE.

Graph showing the WTMS2/ $\Delta P1$ BglG RPE as measured with the developed MS2 and BglG qPCR assays. $\Delta P1$ RPE was measured to be 1.6% that of WT confirming previously published data on $\Delta P1$.

Transfections with packaging constructs and either WT-pCCL-eGFP-MS2 or WT-pCCL-eGFP-BglG transfer vectors were performed to investigate any potential effects of the insertion of these unique sequences on transfection efficiency. The flow cytometry histograms showed no negative effect on transfection efficiency.

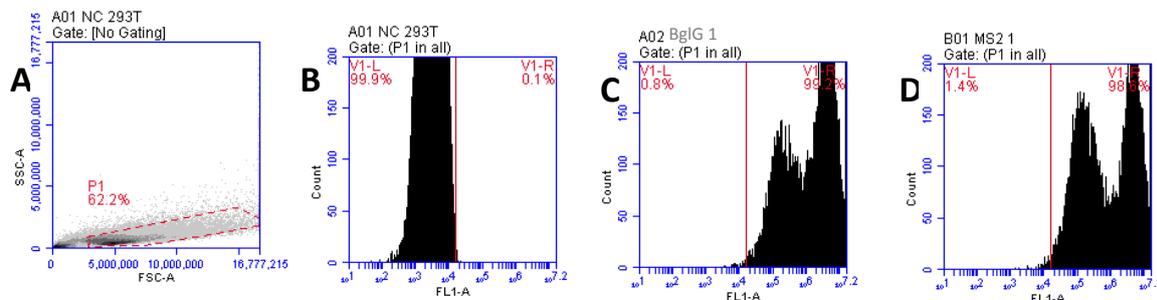


Figure 59. Effect of MS2 and BglG sequence introduction on viral vector transduction efficiency.

[A] Gating of live HEK293T cell population. **[B]** Non transfected HEK293T cells used to create a histogram showing their distribution using the FL1 parameter. **[C]** Transfection efficiency of WT-pCCL-eGFP-Bglg transfer vector. **[D]** Transfection efficiency of WT-pCCL-eGFP-MS2 transfer vector. Flow cytometry set up, including gating of live cells and FL1 bar allocation, were performed as explained in section 3.2.

Upon completion of all these aforementioned control experiments WT-pCCL-eGFP-MS2 and mutant-pCCL-eGFP-BglG plasmids of interest were used for co-transfections following the protocol presented earlier in the beginning of this section.

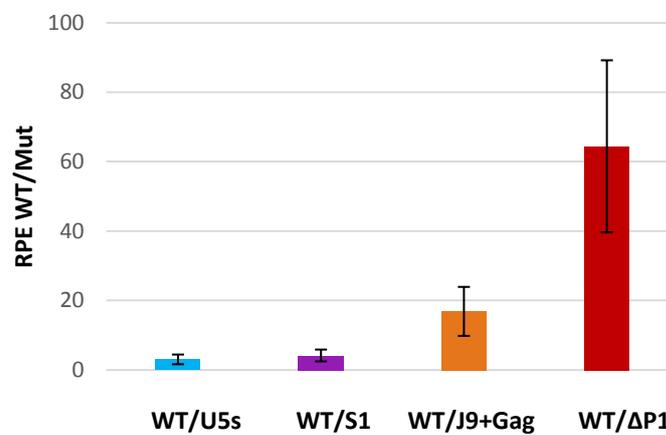


Figure 60. Measurement of RPE of U5-AUG and dimerisation mutants of interest.

A more accurate look at the raw data used to generate the graph of Figure 60 led to the observation that the Ct values from extracellular gRNA were relatively high, close to Ct 30, and did not always fit in the limits of the designed 5 \log_{10} standard curve. A standard curve was set to determine the copy numbers of unknown samples by comparing them to a set of standard samples of known copy number. However, when the unknown samples fell outside of the known linear part of the standard curve limits, I could not be certain of the accuracy of the measured RPE. Therefore, efforts were focused on increasing the qPCR template load by enhancing cDNA synthesis with the addition of betaine, a reagent known to resolve secondary RNA structures by decreasing its melting temperature²⁶¹, and/or increasing the amount of supernatant used for gRNA extraction.

6 μg WT pCCL-eGFP-MS2, 6 μg WT pCCL-eGFP-BglG, 2.5 μg pRev, 3.5 μg pVSVg and 5 μg pSYNGP were used to produce lentiviral vector. This time, extracellular RNA was extracted from 2 mL of viral supernatant instead of 1 mL. The downstream process remained the same as described before. Viral Ct values were observed to shift from 28-30 to 25-28 as aimed and thus improving the reliability of the reading. However, measurement of WTMS2/WTBglG revealed an effect of the processed volume on RPE, which was now calculated to be 2.1 instead of 1.13 as previously shown (Figure 61.C).

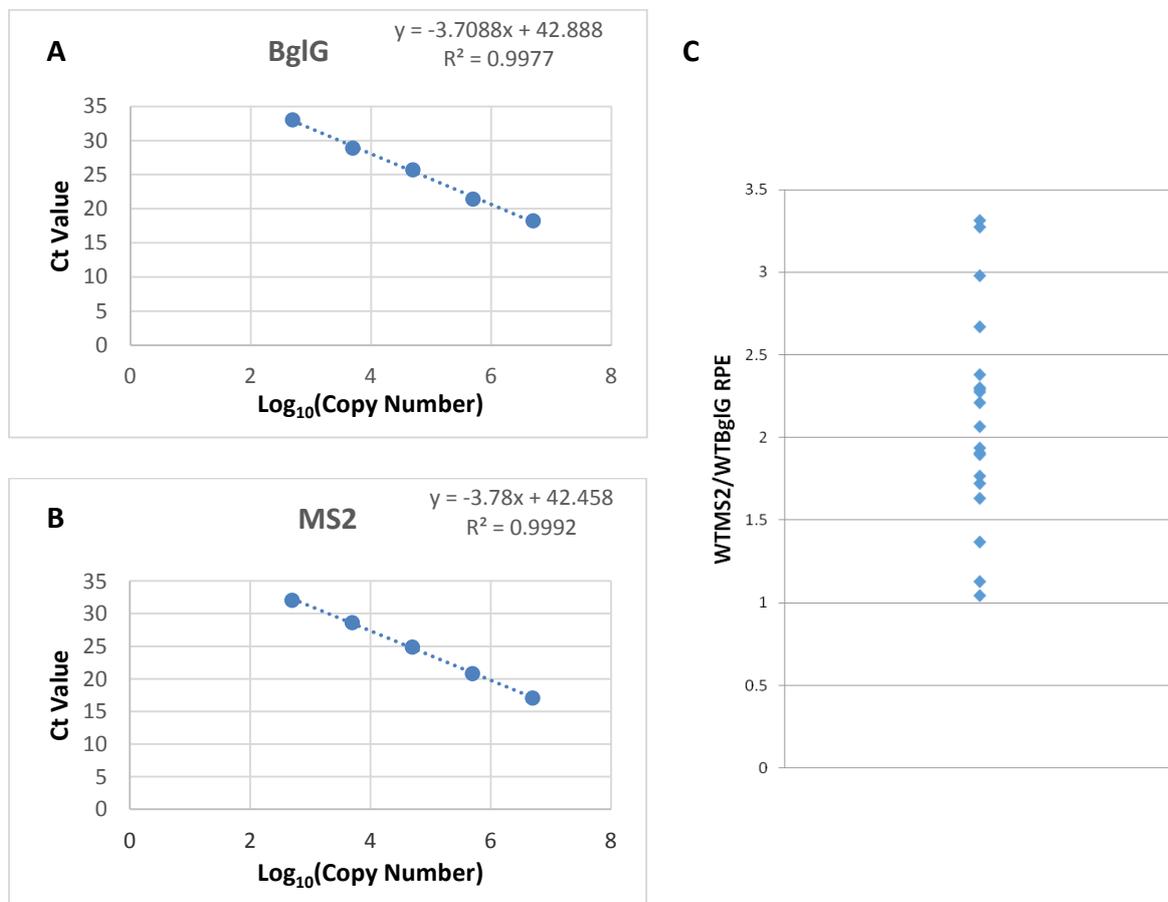


Figure 61. Assessment of WTMS2/WTBglG RPE with the use of a modified protocol.

[A] Graph showing a five \log_{10} standard curve of the BglG assay in a monoplex reaction with $R^2 > 0.99$. **[B]** Graph showing a five \log_{10} standard curve of the MS2 assay in a monoplex reaction with $R^2 > 0.99$. **[C]** Graph showing the measured WTMS2/WTBglG RPE values from 20 independent transfection replicates. gRNA was extracted from 2 mL harvested viral supernatant. Average RPE was calculated to be 2.1.

Based on this result, MS2 was concluded to be a more sensitive assay than BglG. To investigate further, HEK293T cells were co-transfected with the two following combinations of transfer vectors: 1) WTMS2 and U5sBglG, and 2) WTBglG and U5sMS2. 7 co-transfection replicates for each transfer vector combination were performed and analysed as described in the beginning of section 6.5. Measured WTMS2/U5sBglG RPE equalled 3.1 while the calculated U5sMS2/WTBglG RPE equalled 0.85. If the two assays were equally efficient, given that WTMS2/U5sBglG RPE was measured to be 3.1, the expected U5sMS2/WTBglG RPE should have been 0.3 ($=1/3.1$). Thus, the MS2 assay was found to be approximately 2 times more sensitive than the BglG assay.

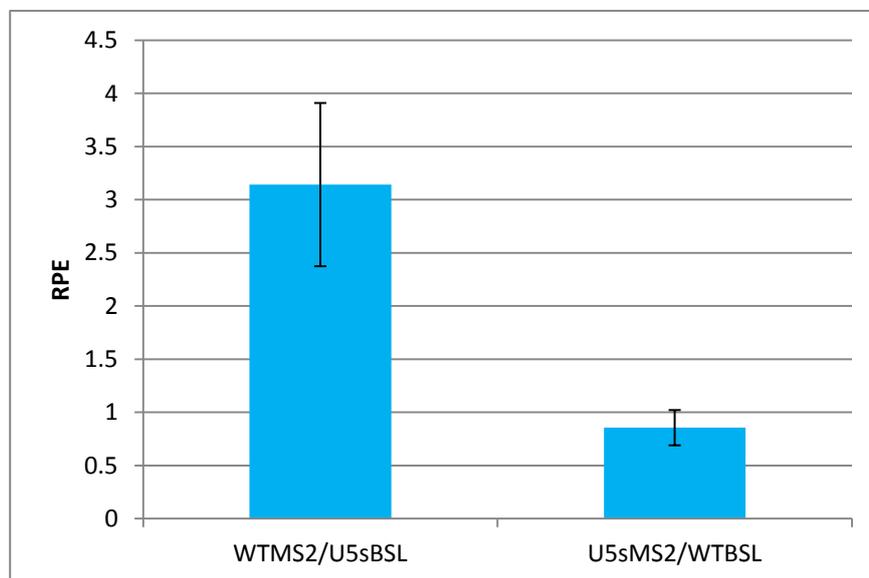


Figure 62. Measurement of WTMS2/U5sBglG and U5sMS2/WTBglG RPEs.

Graph showing the calculated WTMS2/U5sBglG and U5sMS2/WTBglG RPEs. WTMS2/U5sBglG RPE was found to be 3.1 while U5sMS2/WTBglG RPE calculation equals with 0.85. Error bars show standard deviation.

The next experiment was designed to assess the impact of both processed viral supernatant volume and betaine addition (0.5 M final concentration per reverse transcription reaction) on the measured RPE. The transfection protocol remained the same, but the downstream processing and the reverse transcription reaction were altered according to the following four combinations: 1) 1.4 mL viral supernatant + Betaine, 2) 2 mL viral supernatant + Betaine, 3) 1.4 mL viral supernatant - Betaine, 4) 2 mL viral supernatant - Betaine. Analysis

of RPEs from 3 replicates per condition revealed that both increase of processed volume and betaine addition on the reverse transcription level had positive effects on the qPCR readout. The qPCR protocol was updated to accommodate these observations. Viral gRNA was extracted from 2 mL of viral supernatant and all reverse transcription reactions contained 0.5 M betaine in all following experiments. A final round of WT-pCCL-eGFP-MS2/WT-pCCL-eGFP-BglG co-transfections was performed and all data were combined to assess the ratio of MS2/BglG assay sensitivity. gRNA extracted from viral supernatant and their respective transfected cells was analysed from 33 independent transfection replicates. The average MS2/BglG efficiency of the new protocol was 1.85 (Figure 63). This ratio was taken into account in all analysis presented below to compensate for the lower BglG assay efficiency. This adjustment enabled us to accurately measure the relative packaging efficiency of several biological mutants of interest while accounting for any technical biases.

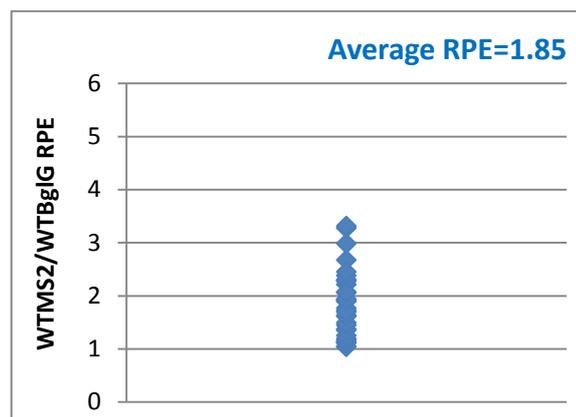


Figure 63. Measurement of WTMS2/WTBglG RPE with the use of an optimised protocol.

Graph showing the measured WTMS2/WTBglG RPE values from 33 independent transfection replicates from three rounds of experiments. Average RPE reflecting the MS2/BglG assay sensitivity was calculated to be 1.85.

Co-transfections of WT-pCCL-eGFP-MS2/U5s-pCCL-eGFP-BglG and U5s-pCCL-eGFP-MS2/WT-pCCL-eGFP-BglG were performed to evaluate whether the qPCR protocol was optimal to accurately measure the biological significance of mutations on relative packaging efficiency without the interference of technical biases. As shown in Figure 64 the average WTMS2/U5sBglG RPE was 3.2 when taking into consideration the 1:1.85 MS2/BglG technical bias explained above. If upon optimisation the two assays were equally efficient, given that

WTMS2/U5sBglG RPE was measured to be 3.2, the expected U5sMS2/WTBglG RPE should be 0.3 ($=1/3.2$). The average U5sMS2/WTBglG was measured to be 0.34 when accounting for the aforementioned technical bias (Figure 64). These results validated the use of this novel qPCR assay to accurately measure relative packaging efficiency of lentiviral vectors in a competitive environment. This qPCR methodology was used in the following section to study the effect of rationally designed mutations on RPE.

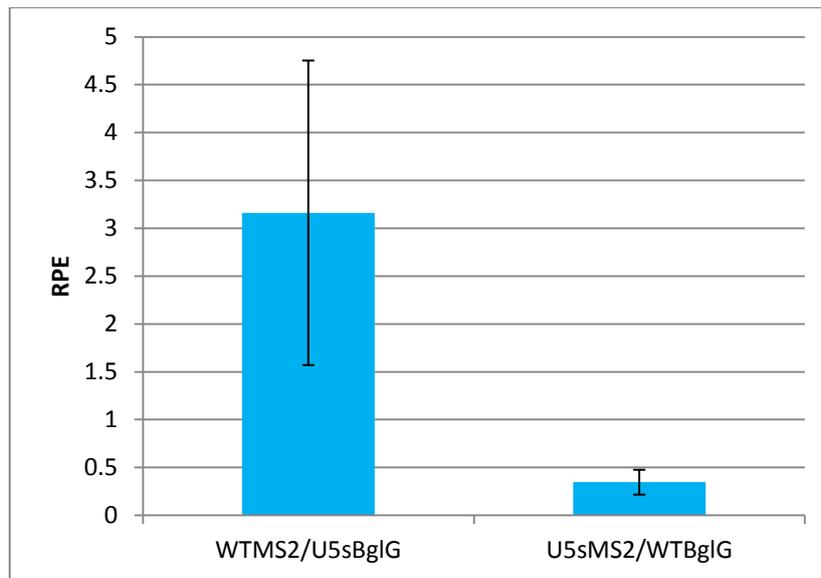


Figure 64. Measurement of WTMS2/U5sBglG and U5sMS2/WTBglG RPEs with the use of an optimised protocol.

Graph showing the calculated WTMS2/U5sBglG and U5sMS2/WTBglG RPEs. WTMS2/U5sBglG RPE was found to be 3.2 upon analysis of 7 independent transfections. The U5sMS2/WTBglG RPE calculation from 6 transfection replicates equals with 0.34, validating the methodology developed to accurately measure relative packaging efficiency. Error bars show standard deviation.

6.6. Use of a novel qPCR protocol to measure relative packaging efficiency of lentiviral vectors in a competitive environment

6 µg WT pCCL-eGFP-MS2, 6 µg mutant pCCL-eGFP-BglG, 2.5 µg pRev, 3.5 µg pVSVg and 5 µg pSYNGP were used for lentiviral vector production. Extracellular RNA was extracted from 2 mL of viral supernatant while intracellular RNA was extracted from 900 µl transfected cell lysates. Extracted RNAs were treated with the Turbo DNA-freeTM kit to achieve plasmid carry over removal. Post digestion, gRNAs were used as templates for reverse transcription reactions containing 0.5 M Betaine, 1x Buffer, 1x RT Random Primers, 40 mM dNTPs, 50 U Multiscribe Reverse Transcriptase and 20 U RNase Inhibitor in 20 µL final volume. No RT NC reactions were also set up to ensure lack of plasmid carry over. qPCR reactions contained 750 nM BglG probe or 250 MSL probe, 500 nM of each primer, 1x Fast Advanced buffer and 3 µL cDNA in a 10 µL reaction volume. Data analysis was performed by adjusting the BglG readout 1.85 times to compensate for the assay's sensitivity discrepancy compared to MS2.

U5s, S1, J9+Gag and ΔP1 lentiviral vector mutants whose backbone was engineered to contain the unique BglG sequence, were chosen to be studied further based on their biochemical and dimerisation phenotypes. U5s mutant was previously studied extensively with the use of the WPRE qPCR assay. S1 mutant produced infectious titres equivalent to those of WT. Northern Blot and *in vitro* dimerisation analysis revealed that S1 had a higher propensity to dimerise than the WT construct. J9+Gag was characterised by a severe decrease in the production of infectious titres. Northern Blotting was not possible due to the poor encapsidation yield (data not shown), but *in vitro* dimerisation assays indicated an increased dimerisation propensity. ΔP1 is a packaging deficient mutant whose RPE was measured to assess the LOD of the developed assay. The transfection protocol, downstream processing and RT-qPCR conditions described in the beginning of this section were applied to study the relative packaging efficiency of U5s, S1, J9+Gag and ΔP1 lentiviral vector mutants and thus investigate the relationship between relative packaging efficiency, transduction and dimerisation efficiency.

Average RPE from analysis of 14 replicates derived from the two independent rounds of transfections presented in Figure 65 was 2.4 for WTMS2/U5sBglG and 0.4 for U5sMS2/WTBglG. This result suggests once more the suitability of the developed qPCR methodology for accurate and unbiased measurement of packaging efficiency of lentiviral vector mutants in relation to WT.

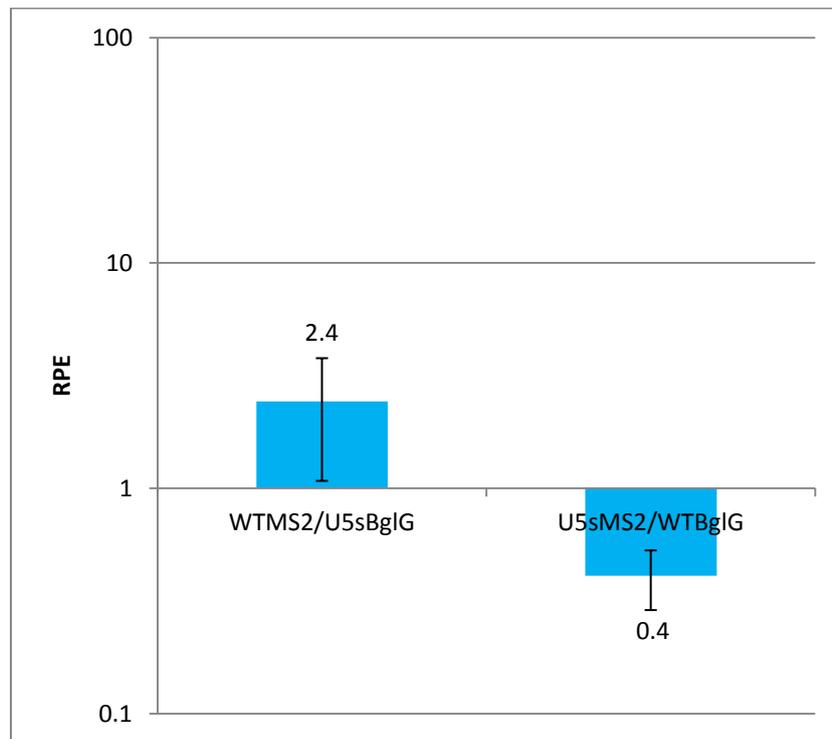


Figure 65. Relative packaging efficiency measurement of U5s mutant.

Average WTMS2/U5sBglG and U5sMS2/WTBglG RPEs presented on a logarithmic scale. Error bars show standard deviation.

Δ P1 was the next lentiviral vector mutant to be analysed. Quantification of the relative packaging efficiency of Δ P1 was important to define the assay's limits of detection. From the 13 independent transfections, only 7 were taken into consideration for further analysis. The other 6 replicates were excluded as their Ct values were as high as 37; too close to background control levels to be analysed; potentially due to the established low packaging efficiency of this mutant. The average WTMS2/ Δ P1BglG RPE of the analysed replicates was 85.7 suggesting that in comparison to WT, only 1.1% of the Δ P1 gRNA was successfully packaged.

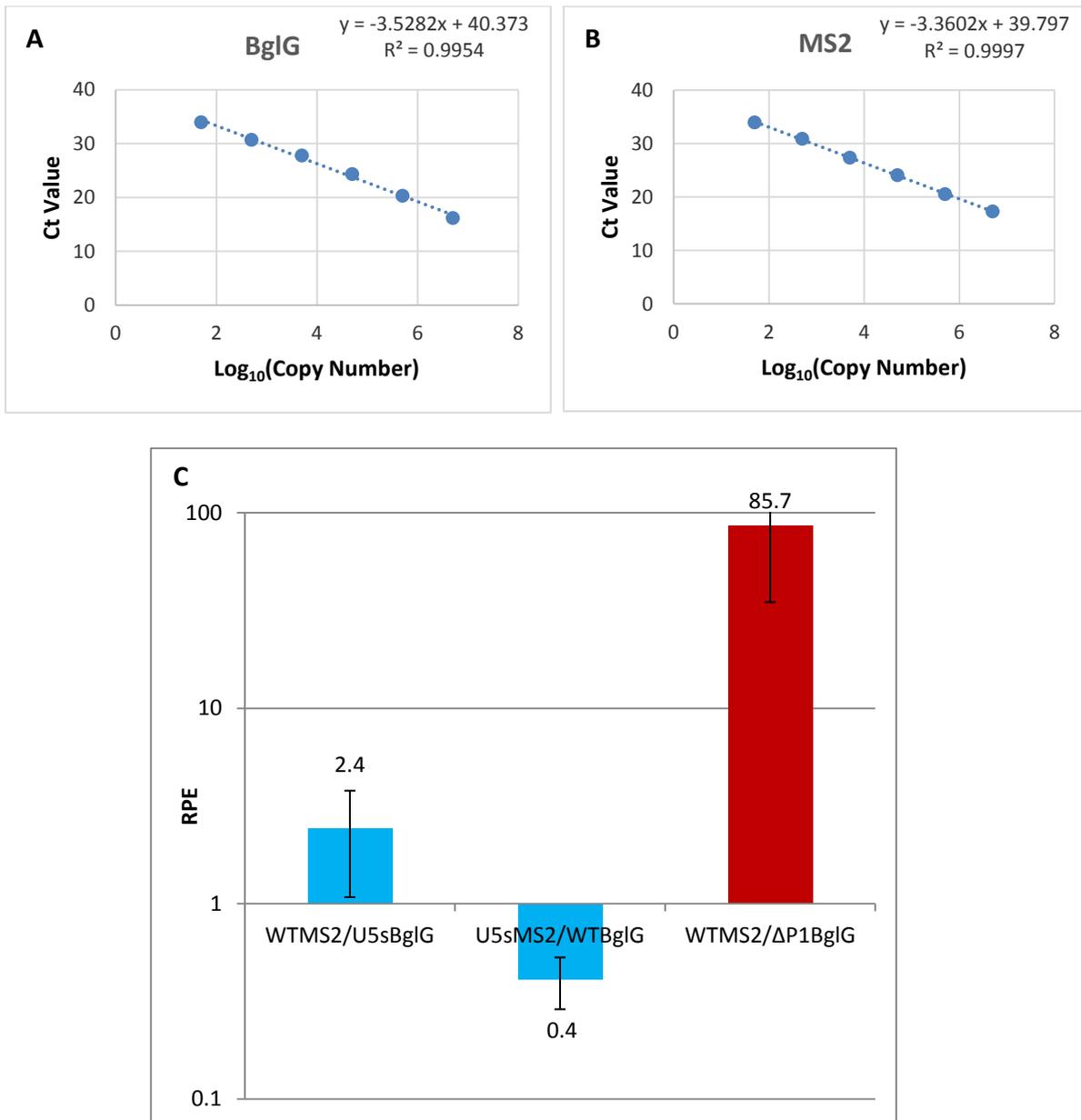


Figure 66. Relative packaging efficiency comparison between U5s and $\Delta P1$ mutants.

[A] Graph showing a six \log_{10} standard curve of the BglG assay with an $R^2 > 0.99$, assay efficiency 92.04% and amplification factor of 1.92. **[B]** Graph showing a six \log_{10} standard curve of the MS2 assay with an $R^2 > 0.99$, assay efficiency 98.4% and amplification factor of 1.98. **[C]** Chart confirming assay's accuracy and LOD. Assay's accuracy was confirmed by comparing the average RPE of WTMS2/U5sBglG and U5sMS2/WTBglG, while the LOD of the assay was defined by measuring the RPE of WTMS2/ $\Delta P1$ BglG. Error bars show standard deviation.

Upon definition of the assay's LOD, the assay was used to investigate whether the 50 fold decrease in the infectious titres of J9+Gag mutant was due to poor packaging. The RPE of WTMS2/J9+GagBglG was measured to study whether this dramatic infectivity reduction was indeed because of the effect of the designed J9+Gag mutations on the encapsidation process. The average WTMS2/J9+GagBglG RPE was found to be 11. However, viral production of this round was measured to be relatively low. Both MS2 the BglG Ct values were thus affected, resulting on BglG Ct values being close to or slightly outside the limits of the standard curve.

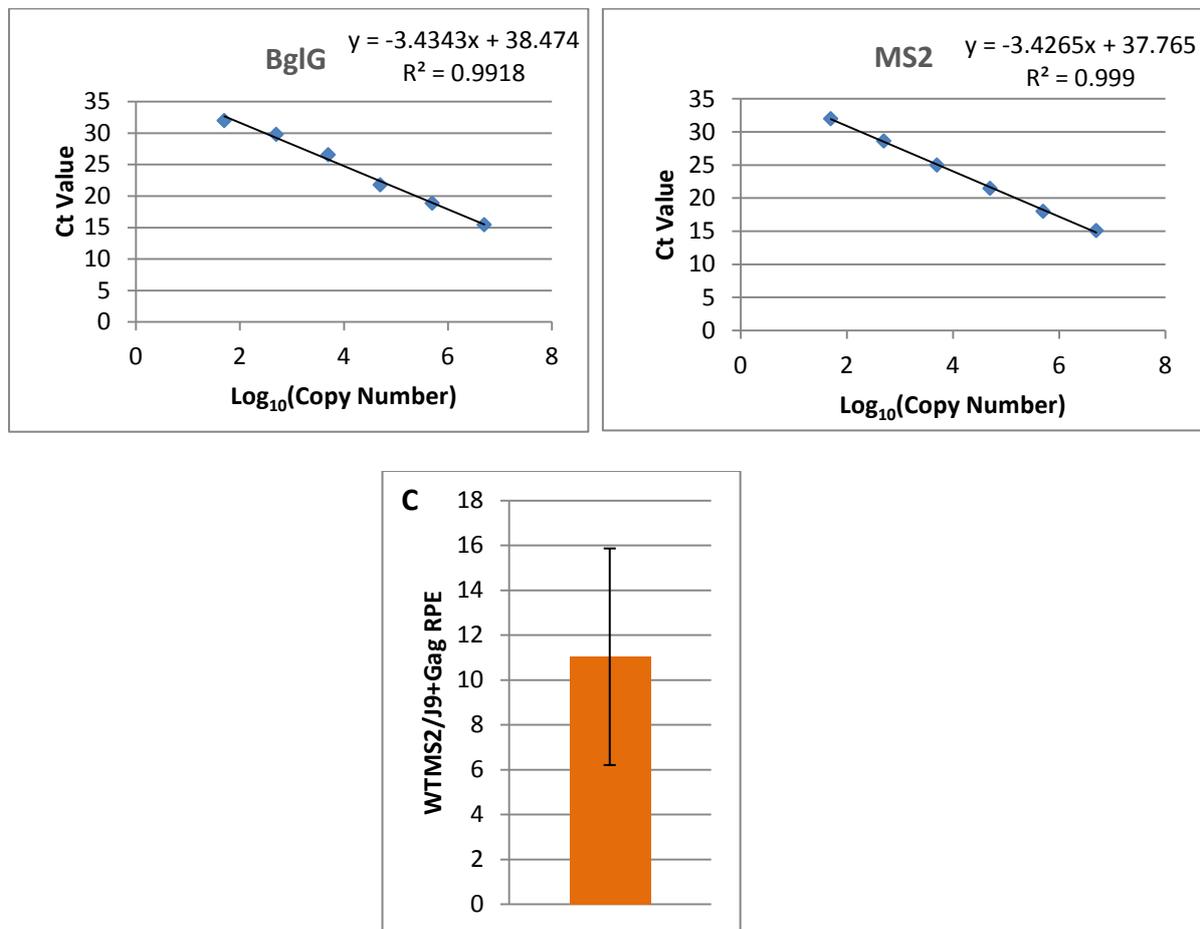


Figure 67. Relative packaging efficiency measurement of J9+Gag mutant.

[A] Graph showing a six \log_{10} standard curve of the BglG assay with an $R^2 > 0.99$, assay efficiency 95.52% and amplification factor of 1.96. **[B]** Graph showing a six \log_{10} standard curve of the MS2 assay with an $R^2 > 0.99$, assay efficiency 95.8% and amplification factor of 1.96. **[C]** Average WTMS2/J9+Gag RPE analysis from 9 independent transfection replicates. Error bars show standard deviation.

The S1 mutant was designed to have a strengthened U5-AUG base pair interaction and was shown to have increased dimerisation efficiency *in vitro*⁷. *In vitro* and *in vivo* experiments revealed an increased dimerisation efficiency of S1 lentiviral vector in our system (Figures 35 and 41). The transduction efficiency of this mutant was found to be comparable to that of WT, but slightly lower. Interestingly, as shown in Figure 68, the relative packaging efficiency of S1 was also found to be comparable to that of WT, but slightly higher.

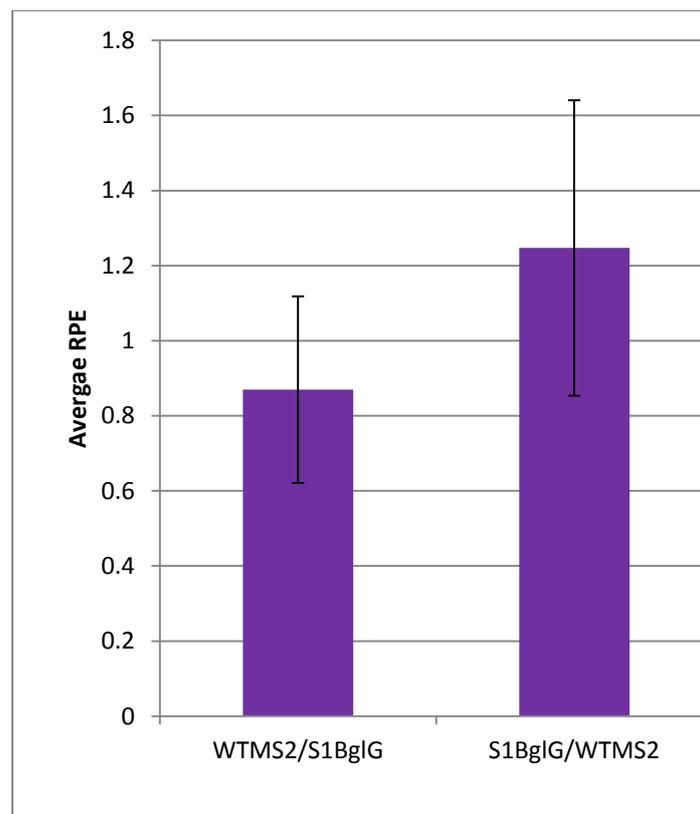


Figure 68. Relative packaging efficiency measurement of S1 mutant.

Graph presenting average RPE of WTMS2/S1BglG and average RPE of S1BglG/WTMS2. Average RPEs were calculated from 11 independent replicates from two different rounds of transfections. Error bars show standard deviation.

6.7. Chapter conclusions

The introduction of rationally designed mutations in the 5'UTR aimed to create vectors whose RNA was more likely to adopt the dimeric conformation and therefore be packaged. To evaluate this, an assay targeting the WPRE element of the 3rd generation lentiviral vectors, developed by Lizee *et al.*, was initially used. The observed variability resulted in reassessment of the assay of choice and design of a novel assay to measure packaging efficiency accurately. Here, I presented a novel competitive qPCR assay designed such that it removes any technical bias by measuring WT and mutant RPE in a co-transfection environment. The unique MS2 and BglG sequences were introduced in the backbone of WT and mutant lentiviral vectors. In the final analysis, MS2 was associated with the WT backbone and BglG with each of the mutants chosen to be studied. The development of a competitive assay aimed to remove biases introduced by potential differences in transfection and/or gRNA extraction efficiencies. Additionally, this novel assay is time and cost efficient as the chosen methodology analysis eliminates the need to normalise samples to an internal housekeeping gene. Initially, I had planned to develop a multiplex assay for measurement of relative packaging efficiency. However, both the efficiency and linearity of the MS2 and BglG assays were observed to decrease in multiplex reactions compared to individual monoplex reactions. The qPCR assay was optimised and then used to assess the relative packaging efficiency of U5s, S1, J9+Gag and Δ P1 lentiviral vector mutants (Figure 69).

The observed overall positive effect of the introduced mutations on RNA dimerisation efficiency did not correlate with RPE improvement. With the exception of S1, all other assessed mutants were characterised by a decrease in their packaging efficiency. Additionally, comparison between the RPE of mutants and their infectivity, revealed a negative effect of the dimerisation enhancing mutations on functional titres. In the case of U5s transfer vector, the reduction of RPE was proportional to its infectivity reduction. However, in the case of J9+Gag, the decrease in viral titres was disproportional to the reduction in RPE, and in the case of S1, vector infectivity was decreased despite the improvement of its packaging efficiency. This data reflects the multifunctionality of the

targeted regions and suggests that RNA flexibility might be more important than RNA dimerisation for vector infectivity.

The assay was used to measure a broad range of relative packaging efficiencies of mutants. This range covered mutants comparably efficient to WT to mutants that encapsidated at approximately 1% of WT levels. I aimed to optimise HIV-1 derived lentiviral vectors through structure informed genome modifications and thus such modifications were introduced based on structural information rather than sequence importance. Even though the effect of these mutations on functionality, dimerisation and packaging efficiency was addressed, it still remained unclear how these rationally designed mutations affect RNA structure directly. In-gel SHAPE was performed to study the structures of lentiviral vector RNA mutants but also to explore the structure of WT *psi* in our vector RNAs, in particular studying the influence of regions adjacent to *psi* on dimerisation and packaging. The results of these studies are presented in the next chapter.

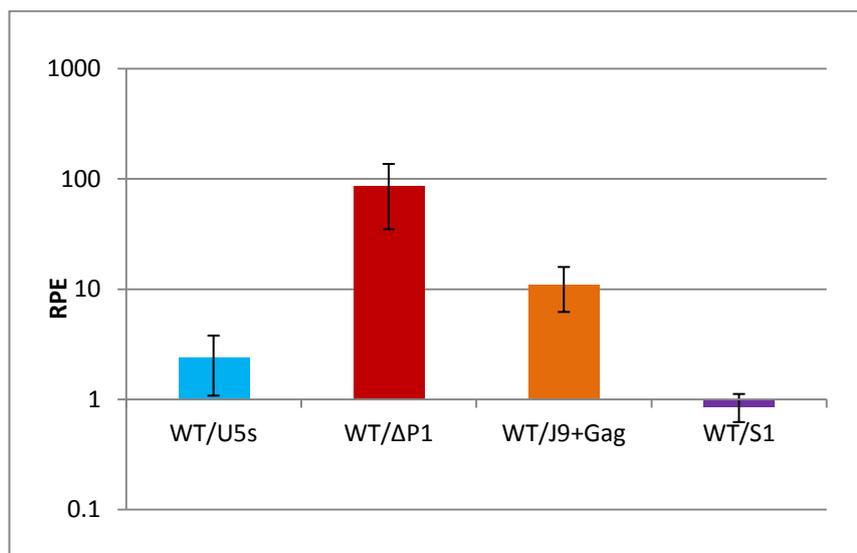


Figure 69. Relative packaging efficiency measurement of U5s, ΔP1, J9+Gag and S1 mutants.

Graph presenting the average RPEs of all mutants studied with the use of the developed competitive qPCR assay. In comparison to WT, only 1.1% of the ΔP1 gRNA was successfully packaged. WT was measured to have a 2.4 and 11 times higher packaging efficiency than the U5s and J9+Gag mutants respectively. Lastly, S1 was the only mutant with a slightly improved, 1.2 times higher than WT's packaging efficiency.

7. RNA structural analysis with in-gel SHAPE

Combined data from previous chapters showed that dimerisation enhancing mutations did not automatically lead to better packaging of vector RNA. These dimerisation enhancing mutations were created based on previous structural studies and our mFold structural predictions; however, the structures of the developed gRNA mutants have yet to be explored. The structure of the WT HIV-1 leader (figure 70.A) has been extensively studied with different techniques^{8 60 61 62}, but the biological role of the downstream *gag* sequences remains unknown. Recent findings supported their importance on packaging, structure and function^{262 263}. Efforts to remove *gag* sequences from the transfer vector backbone have been unproductive and *gag* deletion experiments revealed that in addition to the HIV-1 leader sequences, a significant portion of *gag* was required in *cis* for efficient vector propagation²⁶⁴, and according to more recent evidence to mediate efficient RNA packaging²⁶². The 3rd generation advanced lentiviral transfer vector genome was established to contain 355 bp of the 5' *gag* gene (figure 70.B) based on the work of Naldini *et al.*, despite its exact biological role being unknown¹⁷³.

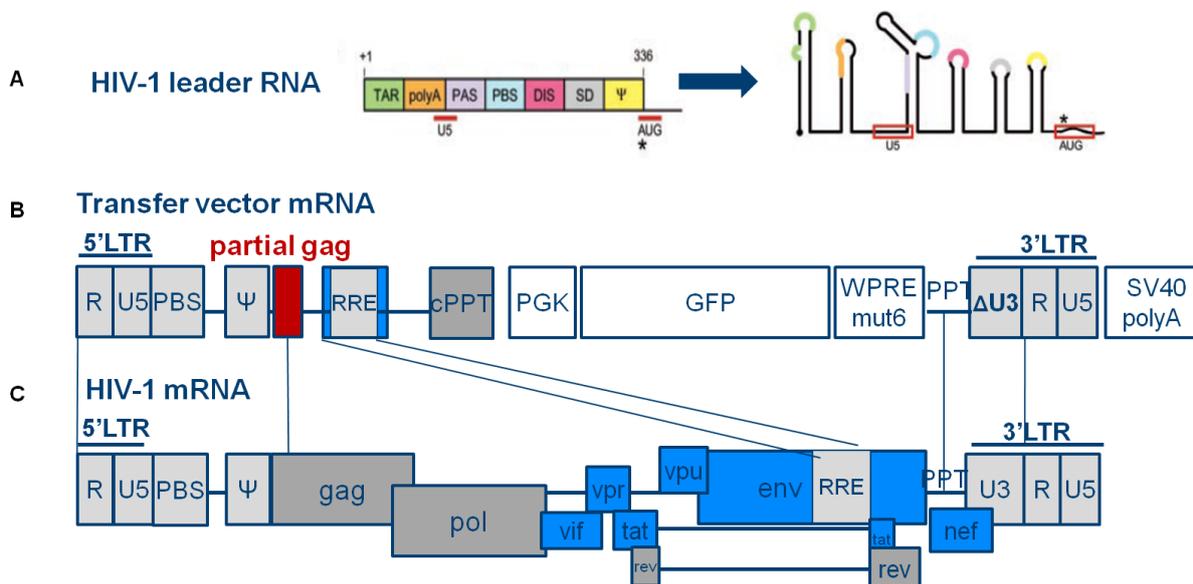


Figure 70. Comparison of transfer vector and HIV-1 mRNAs.

[A] HIV-1 leader sequences and their respective structural motifs⁷. **[B]** gRNA sequences of a 3rd generation HIV-1 derived lentiviral vector. **[C]** gRNA sequences of WT HIV-1.

In-gel SHAPE was employed to study the effect of the introduced mutations on the monomeric and dimeric conformation as well as the influence of the 355 nt long *gag* region on the structure of *psi* on our vector system. SHAPE is a technique that can be used to identify RNA backbone flexibility, an indication of whether nucleotides are base-paired or not. In-gel SHAPE was specifically chosen due to the ability of the technique to determine the RNA secondary structure of individual conformers from a mixed RNA population and was shown to be an appropriate technique for examination of conformational switches or dimerisation processes⁸. In-gel SHAPE involved the separation of dimeric and monomeric structural population of HIV-1 RNA by non-denaturing gel electrophoresis followed by individual probing of the two different conformers within the gel matrix⁸. The following section discusses optimisation steps followed to develop an appropriate in-gel SHAPE methodology to study the 690 nt long RNAs of interest.

7.1. In-gel SHAPE methodology

690 nt long RNA fragments including the packaging signal and the adjacent *gag* sequences were prepared by *in vitro* transcription as described in 3.9.1. RNA was renatured, refolded in a buffer that was optimised to contain 50 mM sodium cacodylate and then fractionated by native gel electrophoresis to separate the monomeric and dimeric RNA. A portion of the gel was stained with ethidium bromide to visualise the level of bands that correspond to the 690 nt long monomer and the 1380 nt long dimer. The remaining unstained monomeric and dimeric RNAs were excised and probed with NMIA before electroelution. Optimisation of the refolding conditions was discussed in 5.3.b and was performed to enhance dimerisation efficiency of this longer RNA fragment to the previously published 356 nt long HIV-1 leader RNA fragments. More specifically, in this chapter, the 356 nt long RNA fragments that contained the HIV-1 leader sequences and 21 nt of the *gag* region was defined as the 8BP length, named after a previously published primer set designed to study the HIV-1 leader region⁸, while the 690 nt long RNA fragment that contained the HIV-1 leader sequences and the 355 nt long partial *gag* region was called 8BP+*gag*.

Improvement of *in vitro* RNA dimerisation efficiency was critical for isolation of adequate RNA amounts to be studied with in-gel SHAPE. Upon successful optimisation of these steps, the reverse transcription reaction had to be optimised next to proceed to data analysis. As

presented in Figure 71.A the reverse transcription protocol followed⁸ produced poor quality data; the signal quality for both cDNA and sequencing ladders was observed to decrease throughout the sequencing traces. It was hypothesised that the RNA secondary structure of the longer 690 nt RNA fragment could potentially be affecting the efficiency of reverse transcription. Increase of the RNA template reaction to 1000 ng and addition of 5 M (final concentration) betaine in the reverse transcription reaction were shown to improve the quality of data (Figure 71.B). The new reverse transcription conditions contained 1000 ng RNA template, 5 M Betaine, 100 U of Superscript III reverse transcriptase (RT), 1.5× SSIII RT buffer, 12.5 mM dithiothreitol (DTT), 1.25 mM deoxycytidine triphosphate (dCTP), deoxyadenosine triphosphate (dATP), deoxyuridine triphosphate (dUTP) and 7-deaza-deoxyguanosine triphosphate (dGTP). 6FAMTM-labelled primers were used for the NMIA-treated samples and VIC[®]-labelled primers were used for the negative control samples.

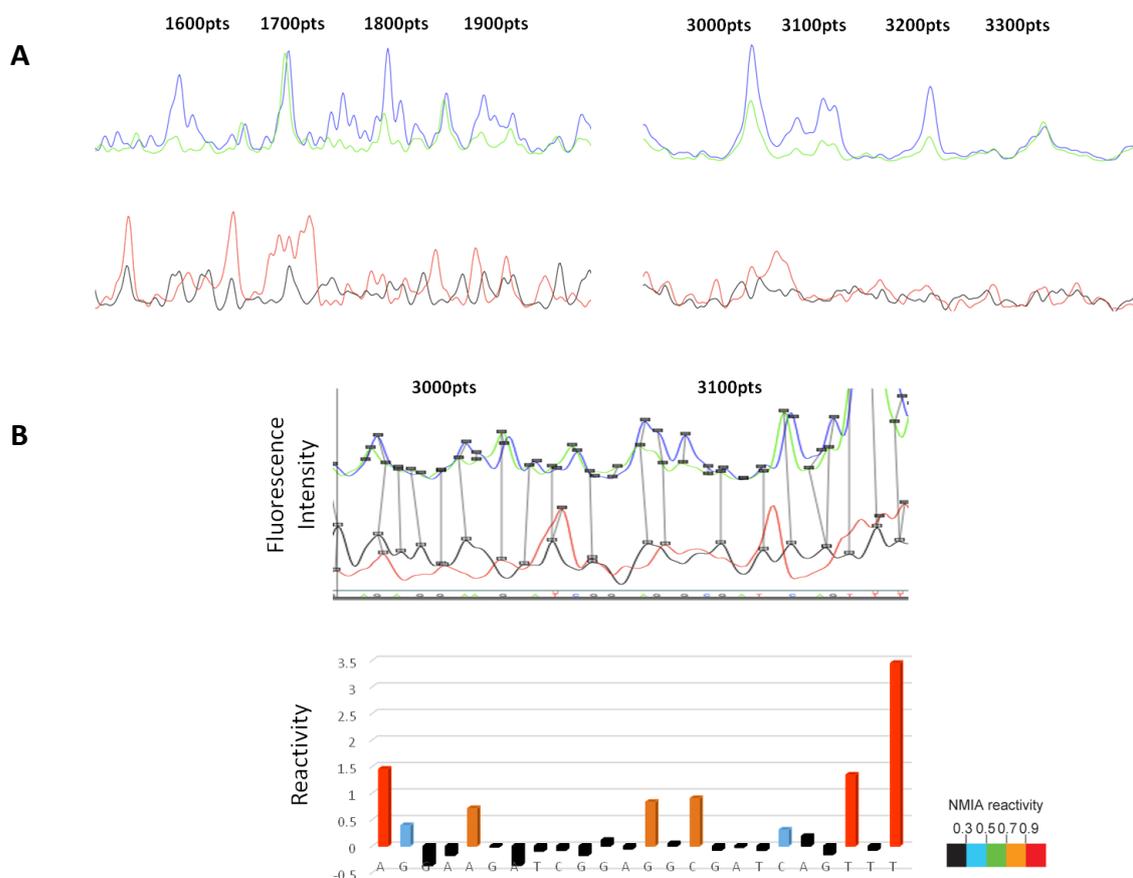


Figure 71. Example of in-gel SHAPE data analysis with the SHAPEfinder software.

[A] WT RNA (8BP+gag) was used as a template for reverse transcription data analysis with sequencing. The blue trace represents the NMIA treated RNA under investigation while the green trace represents its untreated counterpart. Red and black traces represent sequencing ladders used to align each nucleotide to a reactivity peak. The numbers presented in the graph represent the scale along the sequencing capillary. Analysis of different parts of the chosen RNA fragments revealed a decrease in the quality of sequencing peaks throughout the sequence. **[B]** Use of the same RNA template in an optimised reverse transcription reaction including 1000 ng of RNA template and addition of 5 M betaine improved signal quality and allowed for data analysis. Conversion of fluorescence intensity to single level nucleotide reactivity is shown in the bottom graph. Nucleotide reactivity levels were assessed with the use of the NMIA reactivity ladder⁸.

The products of the reverse transcription reactions were 5'-end-labeled cDNA fragments whose length and amount corresponded to the position and degree of modification for every nucleotide. The negative control reverse transcription reaction contained the RNA of interest that lacked NMIA treatment (-) as a template to assess RNA degradation and position-dependent processivity in parallel to the NMIA treated RNA (+) reactions. The resulting (+) and (-) cDNAs were then combined and resolved in a single capillary on a capillary electrophoresis sequencing instrument. The peaks were located, aligned, and quantified for every position in the (+) and (-) reagent channel using the SHAPEfinder software⁵⁵. The difference in the fluorescence intensity between the (+) and (-) reactions was converted to chemical reactivity that was used to study backbone flexibility for each nucleotide. Two dideoxy sequencing reactions were set up with either NED or PET labelled primers and were used to map reactivity to the RNA sequence.

Data were an average of 3–4 replicates at each nucleotide position. For the purposes of this chapter a replicate was defined as an independent repeat of the in-gel SHAPE process (as described above), including all steps from RNA renaturation to reverse transcription of the differently probed RNAs and analysis of their respective results. Differences in reactivity between 1) the monomer and dimer or 2) WT and mutant RNAs, were tested for statistical significance with the use of a paired t-test. The reactivity peaks for 10 nt at both ends of the structure were discounted as data for nucleotides closest to the reverse transcription primer

are more variable. The average NMIA reactivities at individual nucleotides were mapped using the RNAstructure software that predicts secondary structure based on the following prediction and analysis algorithms: calculating a partition function, predicting a maximum free energy (MFE) structure, finding structures with maximum expected accuracy, and pseudoknot prediction²³⁵. Upon prediction of RNA structure, XRNA software was used for RNA structure illustration.

The same conditions and methodology described in this section were employed to study the effect of *gag* sequences on *psi* by comparing the previously published shorter WT 8BP length RNA (356 nt)⁸ to the WT 8BP+*gag* length RNA (690 nt) for both monomeric and dimeric RNAs, as well as the effect of introduced mutations on the monomeric and dimeric conformations. The next section discusses the modelled structures of RNA fragments under investigation and regions of interesting NMIA reactivities. The single nucleotide level colouring in all structures presented in this chapter was according to the NMIA reactivity ladder presented in Figure 71.

7.2. Structural analysis of the monomeric WT HIV-1 RNA

A 356 nt long RNA fragment of the WT HIV-1 leader sequence was *in vitro* transcribed and the 356 nt size corresponding to the monomeric conformation was studied with in-gel SHAPE following the protocol described in section 7.1. Average SHAPE reactivity data from three replicates were analysed with the RNAstructure software and nucleotides were mapped on the predicted structure as presented in Figure 72. Data from the monomeric RNA mapped accurately onto the TAR structure with high reactivity observed in the unpaired nucleotides of the loop and low reactivity observed throughout the TAR stem.

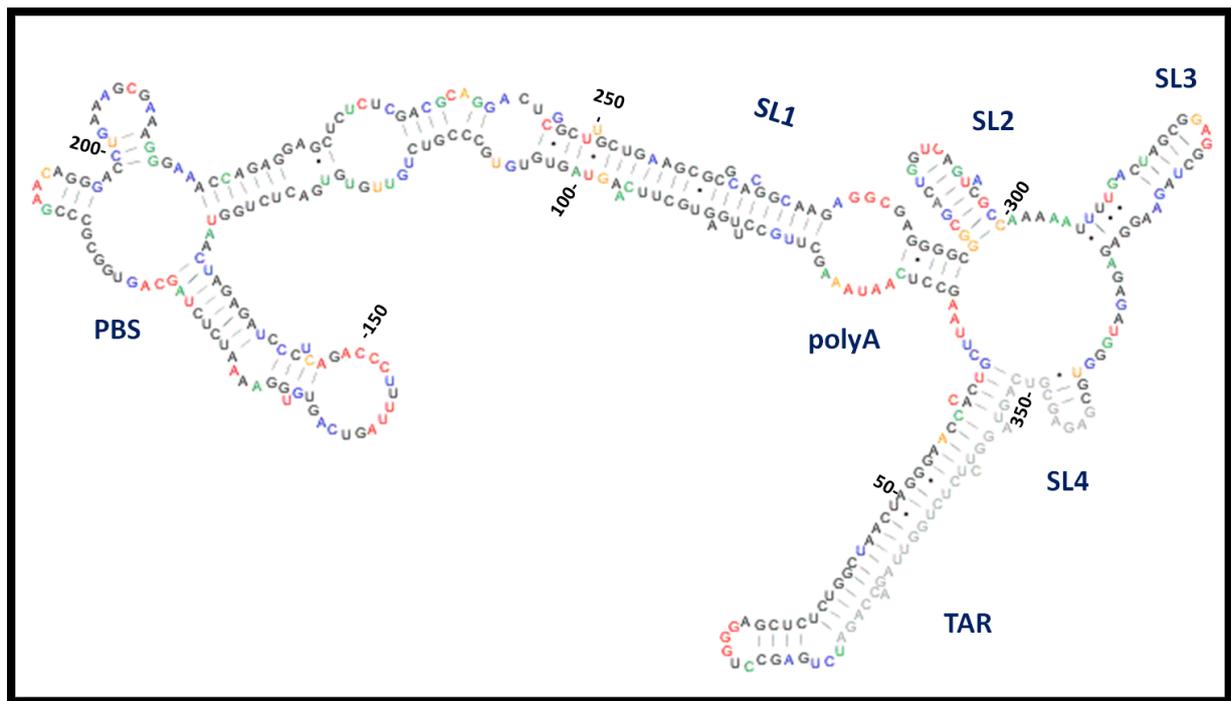


Figure 72. Structure Illustration of a 356 nt long monomeric RNA fragment corresponding to the HIV-1 leader.

In-gel SHAPE was performed as described in section 7.1 and the data from monomeric HIV-1 RNA were mapped onto the proposed structure. Predicted structure includes TAR, PBS and SL2-SL4 stem loops. Numbering is noted every 50 nucleotides.

Similar to Kenyon *et al.*, the lower part of the TAR stem was characterised by intermediate reactivity, suggesting a potential involvement of the TAR base in regulating the monomer-dimer structural switch⁸. Unlike the proposed monomeric structures from Kenyon *et al.*, and Lu *et al.*, there was no formed polyA stem loop, but interaction of the 79-104 nucleotides of the polyA sequences with nucleotides 248-270 of the SL1 sequences. A similarly described model known as LDI was described by Abbink *et al.*⁷. Similarly to the structure presented in Figure 72, the LDI proposed structure contained a long distance base-pairing interaction that occluded the DIS region.

Highly reactive nucleotides 73-78 of the polyA formed a loop with the highly reactive nucleotides 271-274 of the SL1. While in the monomeric HIV-1 leader RNA structure proposed by Kenyon *et al.*, and Lu *et al.*, the SL1 was observed to form a pseudoknot with the 5' of the U5-AUG motif, which was formed in their proposed structure of the dimeric RNA^{8 61}, in the monomeric structure predicted from work performed in this thesis, the 3'

stem of SL1 was shown to interact with sequences in the polyA. Interaction between the polyA and SL1 sequences suggests a mechanism for both polyA suppression and SL1 occlusion in the 5'UTR of the monomeric RNA to prevent RNA polyadenylation and RNA dimerisation. Sequences downstream of the polyA that have been reported to form the U5-AUG motif in the dimeric structure of Kenyon *et al.*, and Lu *et al.*, were observed to participate in the formation of two loops with sequences downstream of the PBS and sequences that would belong to the 5' stem of SL1. NMIA reactivities for the PBS stem-loops mapped accurately onto previously published structures^{8 61}. An exception was the formation of an additional stem loop of nucleotides 199-213, which were also reported to be highly reactive in the monomeric structure by Kenyon *et al.* Reactivities mapped accurately onto SL2 and SL3. The existence and function of SL4 has been the topic of debate^{7 8 60 265 266}, but our predicted model supported its formation in the monomeric RNA structure. The nucleotides around the *gag* start codon have also been documented to form an extended hairpin structure in the LDI conformation, preventing their engagement in a long distance interaction with the U5 sequences known to take place in the WT HIV-1 8BP dimeric RNA⁷.

Similarly, a 690 nt long monomeric RNA fragment, including the WT HIV-1 leader sequence and the partial *gag* sequences, was produced and studied with in-gel SHAPE following the protocol described in section 7.1. As shown in Figure 73, addition of *gag* to the RNA sequence had no observed effect on the structure of the HIV-1 leader, with sequences from TAR to SL4 adopting the same structural conformation as in Figure 72, suggesting that *gag* sequences do not play a role in regulating structural motif formation of the monomeric RNA conformation. No direct interaction was observed between the sequences of the WT HIV-1 leader and *gag*. High reactivity was noted throughout the *gag* sequences resulting in the formation of only five short stem loops and one more extended one, while the majority of *gag* sequences were predicted to remain single stranded (Figure 73).

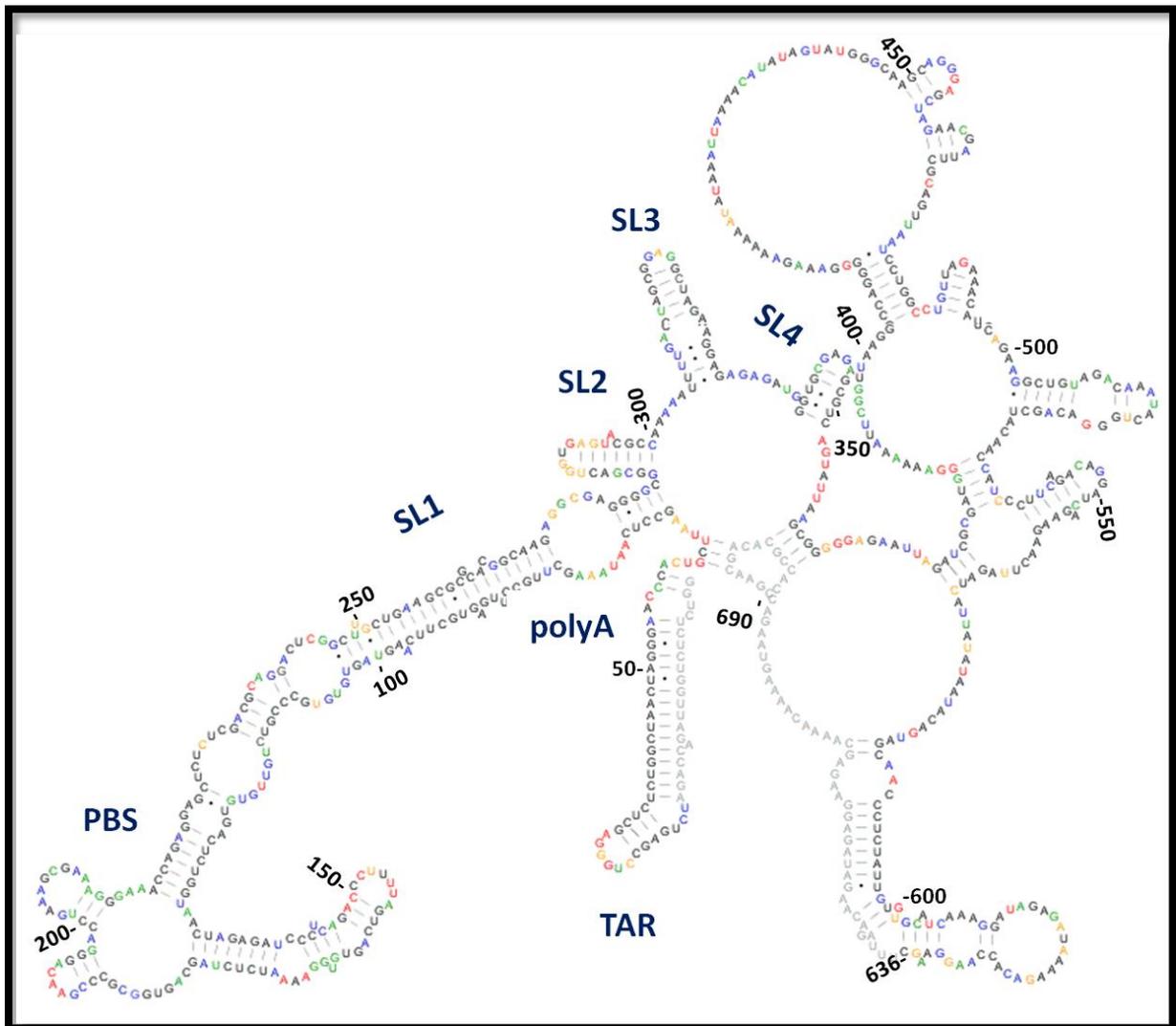


Figure 73. Structure Illustration of a 690 nt long monomeric RNA fragment including sequences of the HIV-1 leader and partial *gag* of the 3rd generation HIV-1 derived lentiviral transfer vector.

Four rounds of in-gel SHAPE were performed as described in section 7.1 and average values from monomeric HIV-1 RNA were mapped onto the proposed structure with the use of RNA structure and XRNA software. No direct interaction between the gag sequences and the HIV-1 leader was observed. Predicted structure of the HIV-1 leader sequences included the same stem loops depicted in Figure 72 and five short stem loops and one extended stem loop formed by the gag sequences. Shape reactivity values could not be retrieved for 637-690 nt, hence -999 was used as a value for structural simulation.

7.3. Structural analysis of the dimeric HIV-1 leader

The 712 nt long RNA fragment corresponding to the dimeric conformation adopted by the dimerisation of the 356 nt WT HIV-1 leader was excised and analysed following the same process as for the monomeric RNA described above. Each type of monomeric RNA studied in section 7.2 and dimeric RNA studied in this section were excised from the same three rounds of in-gel SHAPE to ensure comparability. Similarly, the RNAstructure software was used to predict structures based on the calculated average reactivity values of each nucleotide and free energy constraints. In order to model the structure using RNAstructure, the dimeric RNA was modelled as a single strand containing a duplicated RNA sequence separated by 30 uracil residues that were forced to be single-stranded. The DIS palindrome was forced to be single-stranded as otherwise the initial models generated did not pair DIS with DIS and were hence unlikely to be physiologically relevant. Overall the dimeric structure presented in Figure 74 was similar to previously proposed structural models. Examination of the dimeric RNA structure revealed a TAR stem loop with low reactivity throughout the stem and moderate reactivity of the GGG A nucleotides that comprised the loop. High reactivity was observed in two nucleotides located in a bulge in the bottom of the TAR stem, consistent with previous studies that have identified high reactivity in the closing parts of helices as a common structural trait²⁶⁷. Unlike in the monomeric structure, the formation of the polyA stem loop was observed. Nucleotides 58-104 formed the polyA stem loop which, similarly to TAR, was characterised by low reactivity throughout the stem and moderate to high reactivity in the nucleotides of the loop. High reactivity nucleotides 91 and 96 formed two bulges in the middle of the polyA stem while the two Us in the bottom of the stem were found to be reactive in U-A pairs. Similarly to previously proposed models by Lu *et al.*, and Kenyon *et al.*, the dimeric structure of the 356 nucleotide long HIV-1 leader was indeed characterised by the formation of the U5-AUG motif. There are a number of limitations of the model due to the way in which it was modelled, using one RNA sequence duplicated, rather than two that interacted: In the proposed dimeric structure of the HIV-1 leader presented in Figure 74, the U5-AUG motif was formed between nucleotides 106-115 of the U5 of one RNA molecule with nucleotides 720-729 around the *gag* AUG start codon of the second RNA molecule, however this would be likely to be representative of the intramolecular interaction occurring. There was also a proposed interaction between the A/U

rich sequence between SL2 and SL3, which could be serving as an additional contact site. The DIS-DIS pairing was not shown in the model either, as due to modelling constraints the DIS palindrome had to be forced to be single-stranded. Despite PBS being the UTR region with the greatest structural heterogeneity²⁶⁸, similar NMIA reactivities were observed in the PBS region of our model with these presented in the model by Kenyon *et al.* Low reactivity of the semi-palindromic sequence 181-GUGGCGC-187 both in the dimer and the monomer of our system did not suggest a role of this sequence as a potential second dimerisation contact like previous studies have suggested⁸. However, the prediction of these nucleotides to engage in base pairing could be explained by modelling limitations that could be masking their *in vivo* biological relevance.

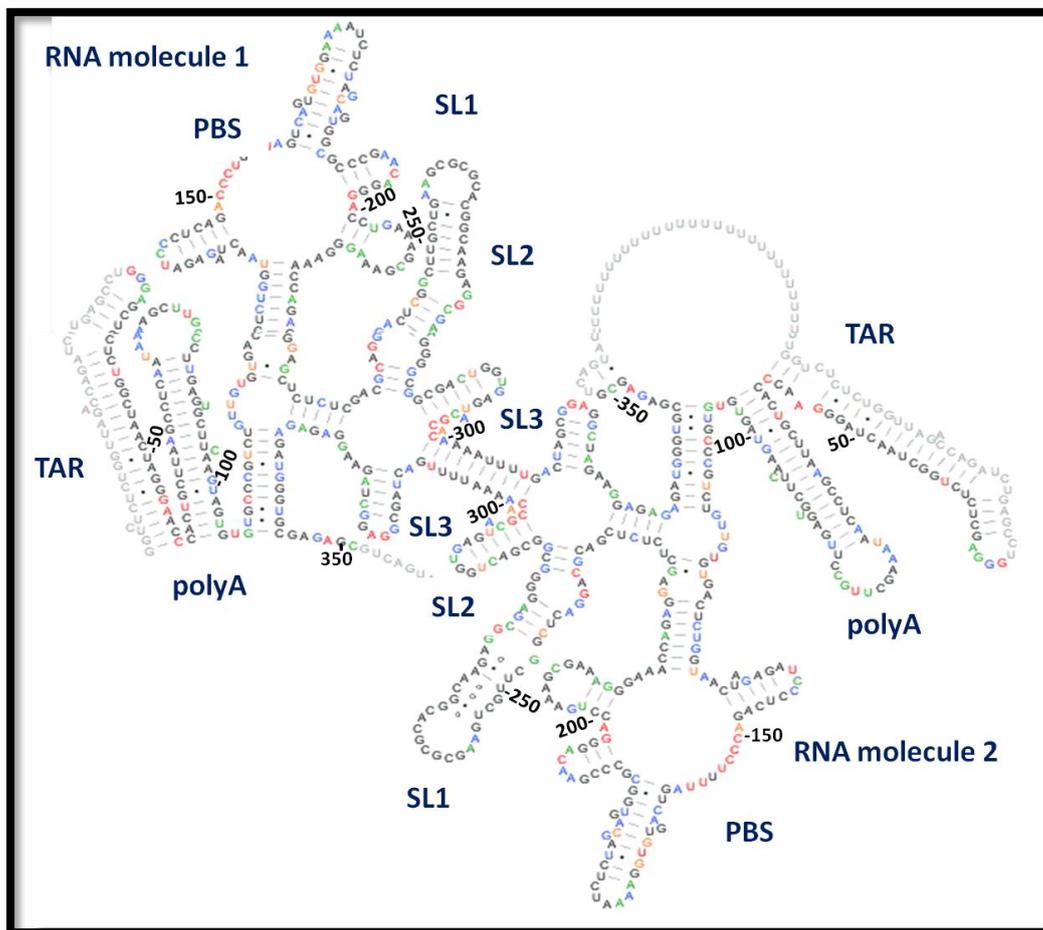


Figure 74. Structure Illustration of a dimeric WT HIV-1 leader RNA fragment.

Average NMIA reactivity values from three rounds of *in-gel* SHAPE experiments were used for prediction of the dimeric structure of the WT HIV-1 leader RNA. NMIA reactivities mapped accurately on the structure predicted by Lu *et al.*, and Kenyon *et al.*, and included stem loops of TAR, polyA, PBS, SL1-SL3 as well as the structural U5-AUG motif.

Nucleotides 229–233 from RNA molecule 1 paired with nucleotides 329–333 from RNA molecule 2 to form a 5-bp helix. Nucleotides 229–233 were unreactive in the dimeric structure, with NMIA values less than 0.3, whereas these nucleotides appeared to be highly reactive in the monomeric structure and participated in the formation of a loop. Dimer destabilisation during probing could be accountable for this phenomenon as any loose dimer could be converted to a monomer and thus locally affect average NMIA reactivity. Previously performed 3D analysis of the region had suggested a possible interaction with the neighbouring helix at nts 125–131 and 217–223²⁶⁹. Finally, SL1, SL2 and SL3, which comprise the major packaging signal, fitted the SHAPE data accurately.

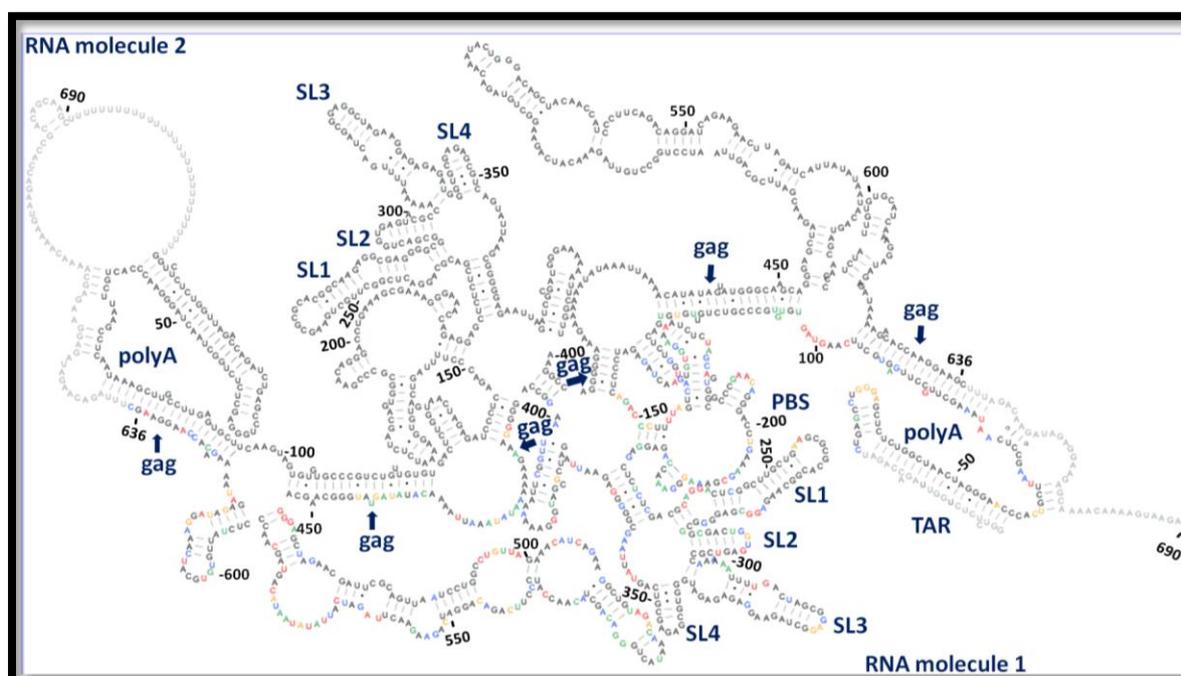


Figure 75. Structure Illustration of a 690 nt long dimeric RNA fragment including sequences of the HIV-1 leader and partial *gag* of the 3rd generation HIV-1 derived lentiviral transfer vector.

Average NMIA reactivity values from four rounds of in-gel SHAPE experiments were used to study the effect of gag on the dimeric structure of the WT HIV-1 leader RNA. Direct interaction between sequences of the gag region and the polyA, U5 and PBS were observed as indicated by the arrows. Shape reactivity values could not be retrieved for 639-690 nt, hence -999 was used as a value for structural simulation.

A 1380 nt long RNA fragment corresponding to the dimeric conformation adopted by the WT HIV-1 leader followed by the partial *gag* region was excised and analysed as previously described from four rounds of in-gel SHAPE. This experiment was designed to investigate the effect of extended *gag* on the packaging conformation. Structural analysis was performed as described for the shorter dimer above, again using a duplicated RNA sequence separated by 30 single-stranded U residues, and forcing the DIS palindrome to be single-stranded. Similar to the dimeric RNA fragment of the WT HIV-1 leader that lacked more extensive *gag* sequences (8BP-*gag*), examination of the dimeric RNA structure followed by the *gag* region (8BP+*gag*), revealed a TAR stem loop with low reactivity throughout the stem and moderate reactivity of the GGG nucleotides that comprised the loop. However, unlike previously described for 8BP-*gag*, no reactivity was observed in the bottom of the TAR stem of 8BP+*gag*, suggesting that the presence of *gag* promoted a more rigid dimeric conformation. As presented in Figure 75, unlike the formation of the polyA stem loop noted in the 8BP-*gag* structure, sequences of the polyA (72-95nt) were observed to pair with nucleotides of the *gag* region (1340-1360 nt) in the 8BP+*gag* structure.

| | | | | | | | | | | | | | | | | | | | | | | |
|---------------------|-----|------|-----|-----|------|------|-----|------|------|------|-----|------|-----|-----|------|------|------|-----|------|------|-----|-----|
| | T | A | A | A | G | C | T | T | G | C | C | T | T | G | A | G | T | G | ---- | C | T | T |
| polyA 72-95 | 99% | 99% | 98% | 97% | 97% | 97% | 97% | 96% | 96% | 97% | 98% | 98% | 98% | 98% | 98% | 97% | 98% | 98% | | 94% | 99% | 98% |
| <i>gag</i> 1360- | A | T | T | T | C | G | A | A | ---- | G | G | A | A | C | ---- | C | A | C | A | G | A | A |
| 1340 | 99% | 100% | 68% | 80% | 100% | 100% | 94% | 100% | | 100% | 75% | 100% | 99% | 97% | | 100% | 100% | 96% | 100% | 100% | 86% | 73% |

Table 7. Sequence conservation analysis of the regions of polyA (72-95nt) and *gag* (1340-1360 nt).

Sequence conservation analysis of the regions of polyA (72-95 nt) and gag (1340-1360 nt) predicted their direct interaction based on NMIA reactivity values from four independent rounds of in-gel SHAPE structural analysis. Nucleotide preservation analysis was performed by comparing the available HIV-1 strain sequences from the Los Alamos database²⁷⁰.

To investigate the likelihood of this interaction, sequence conservation analysis was performed for nucleotides 72-95 and 1340-1360 that took part in the suggested polyA-*gag* interaction, by comparing available HIV-1 strains from the Los Alamos database sequence compendium database 2018²⁷⁰. As expected, nucleotides of the *cis*-acting element of polyA

were highly conserved. Interestingly, half of the *gag* nucleotides taking part in this interaction were measured to be 100% conserved, higher than even the highly conserved sequences of the *cis*-acting element polyA (Table 7). Additionally, analysis of the base pairing with the RNAstructure software supported the likelihood of this interaction with a higher than 99% probability (Figure 7.8).

The presence of the U5-AUG motif in the dimeric RNA of the WT HIV-1 leader (8BP-*gag*) has been supported by several groups^{7 8 60}. However, there was no U5-AUG formation in the presence of the extended *gag*, but a direct interaction between U5 and *gag* nucleotides further at the 3' end of the sequence under interrogation. As shown on table 8, structure preservation analysis based on sequencing variability supported the predicted U5-*gag* interaction. Nucleotide 112 of the U5 highly conserved region is a guanosine in the WT HXB2 sequence and in 95% of HIV-1 strain sequences but an adenosine in the remaining 5% of HIV-1 strains. This sequence variability further supported the aforementioned U5-*gag* interaction, as the (U5) G-U (*gag*) pair of the WT HXB2 sequence would become a paired (U5) A-U (*gag*) with two hydrogen bonds. Notably, 13/19 nucleotides of the involved *gag* sequence were 100% conserved, whereas only 1/19 nucleotides of the involved U5 region, known to be a highly conserved region of the HIV-1 leader, were 100% conserved. Additionally, nucleotide 1156 is a uracile in the WT HXB2 sequence and in 81% of HIV-1 strain sequences but a cytosine in the remaining of the 19% of HIV-1 strains. This sequence variability further supported structure conservation due to the conversion of the (U5) G-U (*gag*) non canonical pair to a now canonical (U5) G-C (*gag*) pair. The only exception was 1157 nucleotide that was highly variable. As shown in Figure 76, the *gag* nucleotides predicted to interact with the U5 region were shown to be highly conserved compared to the rest of *gag* sequences not involved in direct engagement with sequences of the HIV-1 leader. Lastly, the 90-99% base pair probability calculated by the RNAstructure software also supported the likelihood of this (103-122) U5-*gag* (1172-1154) interaction (Figure 77).

| | | | | | | | | | | | | | | | | | | | | | |
|-----------|-----|------|------|------|------|------|------|------|------|------|-----|------|-----|-----|------|-----|-----|-----|------|------|--|
| U5 | | | | | | | | | | G-> | | | | | | | | | | | |
| 103-122 | T | G | T | G | T | G | C | C | C | A! | T | --- | C | T | G | T | T | G | T | G | |
| | 98% | 99% | 99% | 99% | 99% | 99% | 98% | 98% | 98% | 95% | 99% | | 98% | 99% | 99% | 99% | 96% | 96% | 100% | 96% | |
| gag | | | | | | | | | | | | | | | | | A-> | T-> | | | |
| 1154-1172 | A | C | G | A | A | C | G | G | G | T | A | T | G | A | T | --- | C/T | C! | A | C | |
| | 99% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 94% | 100% | 97% | 96% | 100% | | 26% | 81% | 100% | 100% | |

Table 8. Sequence conservation of the regions of U5 (103-122 nt) and gag (1553-1172 nt).

Sequence conservation of the regions of U5 (103-122 nt) and gag (1553-1172 nt) predicted base pairing of these nucleotides instead of engagement of the U5 nucleotides in the previously reported U5-AUG motif. NMIA reactivity values were retrieved from four independent rounds of in-gel SHAPE structural analysis. Nucleotide preservation analysis was performed by comparing the available HIV-1 strain sequences from the Los Alamos database²⁷⁰.

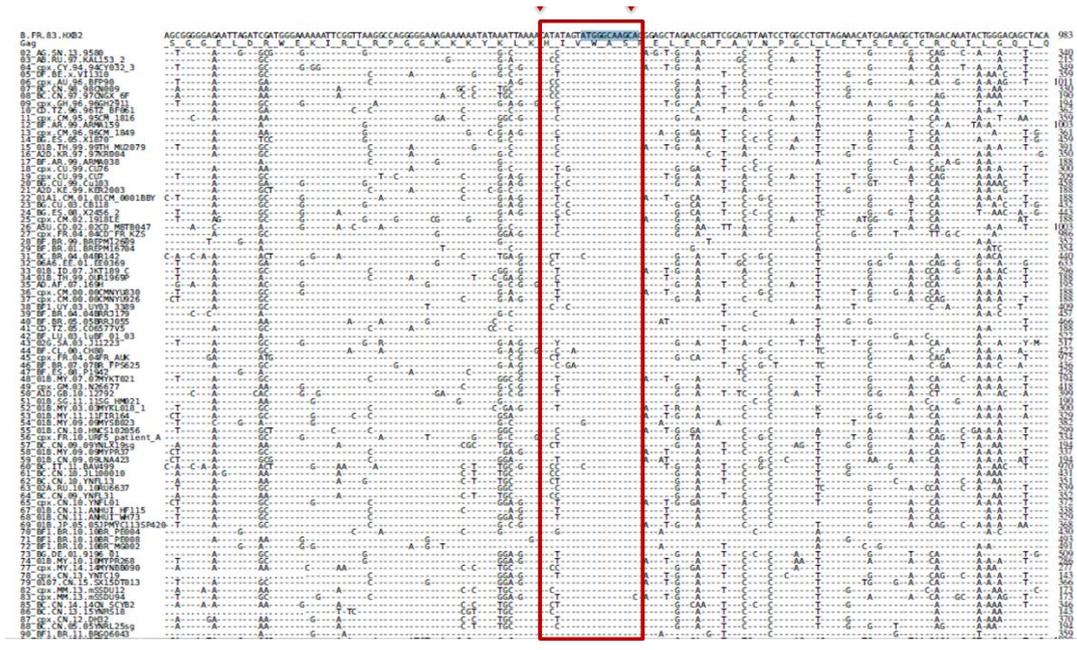


Figure 76. Visual representation of gag region sequence variability.

HIV-1 strain sequences presented here were derived from the Los Alamos database²⁷⁰. gag nucleotides predicted to interact with the U5 region were placed in the red box. Comparison of the gag nucleotides contained in the red box with gag nucleotides on each side, showed high sequence conservation of the gag sequences under investigation, a phenomenon linked with structure preservation.

The formation of a 6 bp helix followed the U5-*gag* interaction which was in turn followed by the prediction of another HIV-1 leader-*gag* interaction. Nucleotides 142-147 within the PBS were predicted to pair with the nucleotides 1127-1132 in the *gag* region. Structural preservation analysis presented in table 9 was in relative favour of this interaction, although it needs to be noted that this was a 6 bp interaction only. Furthermore, the probability calculation by RNA structure estimated that the engagement of the (142-147 nt) PBS-*gag* (1127-1132 nt) was only around 60% (Figure 77).

| | | | | | | |
|-------------------|----------|----------|-------------|----------|-------------|-------------|
| PBS | T | C | C | C | T | C |
| 150-155 | 99% | 99% | 99% | 99% | 100% | 100% |
| <i>gag</i> | A | G | G | G | G | G |
| 1127-1132 | 93% | 99% | 100% | 85% | 99% | 100% |

Table 9. Sequence conservation of the regions of PBS (142-147 nt) and *gag* (1127-1132 nt).

*Potential PBS-*gag* interaction was predicted based on NMIA reactivity values from four independent rounds of in-gel SHAPE structural analysis. Nucleotide preservation analysis was performed by comparing the available HIV-1 strain sequences from the Los Alamos database²⁷⁰.*

Addition of *gag* sequences on the RNA level did not affect the structural motifs PBS and SL1-SL3, which were formed as described above. However, a profound effect on the reactivity values of SL3, denoted as the ψ element, was noticed. In this predicted dimeric structure, SL3 was formed by nucleotides 289-338. Almost half (18/41) of the 41 nucleotides forming SL3 had a decrease in their measured reactivity value, five of them were characterised by an increase in their measured NMIA reactivity, while the remaining were not affected. This observation suggests that *gag* sequences stabilise the ψ element in the dimeric RNA to promote its packaging. Finally, addition of the downstream *gag* sequences promoted the creation of SL4, which did not appear in the dimeric WT 8BP structure due to the formation of the U5-AUG motif.

The model derived from our data showed a direct interaction between sequences of the polyA, U5 and PBS, *cis*-acting elements of the HIV-1 leader, with sequences within the partial *gag* region of the dimeric 8BP+*gag* RNA, for the first time. In Figure 73, the majority

of *gag* nucleotides were part of extensive single stranded regions in the WT 8BP+*gag* monomeric structure, whereas in the WT 8BP+*gag* dimeric structure (Figure 75) *gag* was seen to engage in interactions with the aforementioned regions of the HIV-1 leader. This interesting finding in the *gag* base pairing engagement difference might be part of a structural switch mechanism between the monomeric and dimeric RNA. These HIV-1 leader-*gag* interactions were closely visualised and presented in Figure 77 with the use of the RNAstructure software. Analysis of the full length 690 nt fragment was performed by probing the extracted RNA and performing reverse transcription with the use of primers that spanned the 8BP length (1-356 nt) and *gag* region independently.

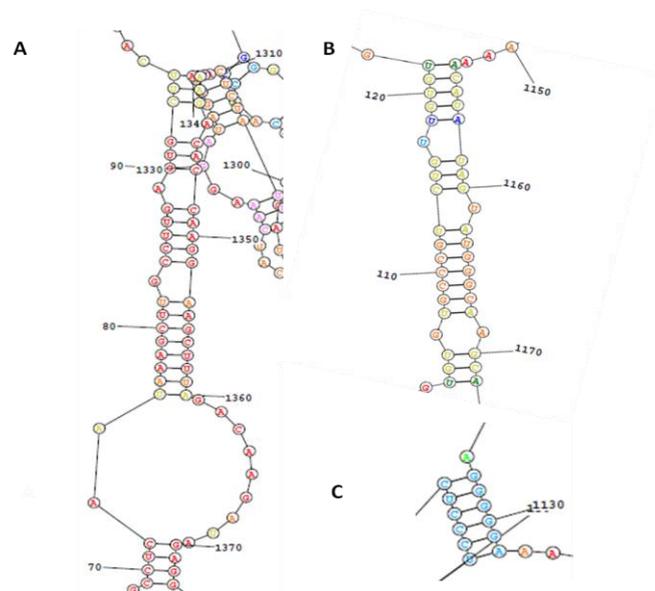


Figure 77. Visualisation of modelled HIV-1 leader-*gag* interactions with the RNAstructure software.

Close up visualisation of **[A]** polyA-*gag*, **[B]** U5-*gag* and **[C]** PBS-*gag* interactions with RNAstructure software. Colouring of nucleotides represents the likelihood of the interacting nucleotides to form a base pair, as calculated by the RNAstructure software. In specific: red $\geq 99\%$, orange 99% to $\geq 95\%$, yellow 95% to $\geq 90\%$, dark green 90% to $\geq 80\%$, light green 80% to $\geq 70\%$, light blue 70% to $\geq 60\%$, dark blue 60% to $\geq 50\%$ and pink $< 50\%$ probability.

The effect of *gag* sequences on the packaging conformation was studied by creating a 690 nt long RNA fragment that contained the WT-HIV sequences and the partial *gag* region which was processed as described above.

The SHAPEfinder software used for secondary RNA prediction included alignment of reactivity peaks to the sequencing ladder traces on a single nucleotide level faced certain trace length limitations. The maximum number of nucleotides that could be analysed with this technology was ~350 nucleotides⁵⁵, which is half the number of nucleotides of the *gag* including RNA fragment of interest. Hence, two sets of primers had to be designed and used for independent reverse transcription and sequencing reactions to cover the 8BP and *gag* sequences separately. This limitation combined with the aforementioned challenges due to the nature of the longer RNA fragment, made the process extremely laborious and time costly. Given the interesting findings on the effect of *gag* on the dimeric RNA packaging conformation, it was desirable to keep the *gag* region in the RNA sequences of the mutants but investigate if probing only for the 8BP 356 nt long region would still accurately predict any potential HIV-1 leader-*gag* interactions. The NMIA values of the dimeric 8BP+RNA were loaded on the RNAstructure software for nucleotides 1-356 of the HIV-1 leader, while the -999 value, indicative of no data present, was loaded for nucleotides 357-690 belonging to the *gag* region. The generated predicted structure was identical to the one presented in Figure 75 in which NMIA values were used for all 690 nucleotides. This result suggests that use of the full length 690 nt RNA but probing only of the HIV-1 leader sequences was adequate to predict interactions between the 5'UTR sequences and *gag* with the same accuracy as with the inclusion of probing data for the 3' half of the nucleotides. Based on this observation, S1 and J9+Gag transfer vector mutants were used as templates for *in vitro* transcription of 1-690 nucleotides containing the respective mutated HIV-1 leader sequences and the WT HXB2 *gag* sequence. However, probing was performed only for nucleotides 1-356. S1 and J9+Gag structures were generated by using the calculated NMIA values for nucleotides 1-356 and the simulating value -999 for nucleotides 357-690 of the *gag* region. The predicted structures for S1 and J9+Gag are discussed in the following sections.

7.4. Structural analysis of the J9+Gag mutant

A 690 nt long RNA fragment containing the J9+Gag mutants described in section 4.5 was *in vitro* transcribed and used for in-gel SHAPE analysis.

As shown in Figure 42, the J9+Gag mutant, designed to have an extended SL1 and characterised by a severe drop in functional titres, mainly adopted the dimeric conformation. Excision of adequate amounts of monomeric J9+Gag RNA has not been possible, hence only the dimeric RNA could be analysed for this mutant. Dimeric J9+Gag RNA was excised from three rounds of in-gel SHAPE. Average calculated NMIA reactivity values for nucleotides 1-356 were loaded on the RNAstructure software while using the -999 simulating value for the remaining of nucleotides belonging to the *gag* region. Structural modelling was again conducted based upon a single RNA sequence that contained two monomer sequences separated by 30 Us, and with the DIS forced to be single-stranded. The predicted RNA secondary structure was illustrated with XRNA as presented in Figure 78.

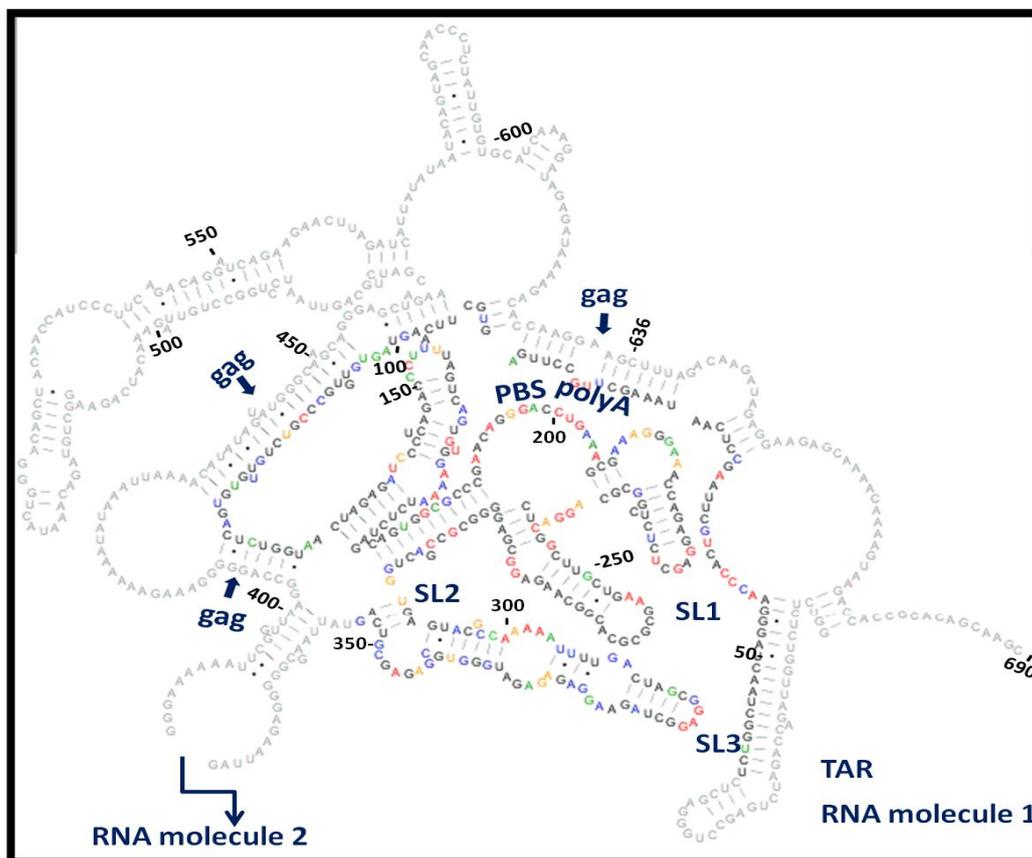


Figure 78. Structure Illustration of a 690 nt long dimeric RNA fragment including sequences of the J9+Gag HIV-1 leader and partial *gag* of the 3rd generation HIV-1 derived lentiviral transfer vector with XRNA.

Average NMIA reactivity values from three rounds of in-gel SHAPE experiments were used to study the secondary structure of the J9+Gag transfer vector mutant dimeric RNA. The position of the second RNA molecule is indicated with an arrow.

TAR stem loop, polyA-*gag* and U5-*gag* interactions mapped accurately on the WT 8BP+*gag* dimeric structure presented in Figure 75. The PBS motif was characterised by great variability to the previously presented PBS structural motifs. Nucleotides 148-158 were of high reactivity in the WT 8BP+*gag* RNA construct and were predicted to be single stranded in Figure 75, whereas more than half of these nucleotides were of low reactivity in the J9+Gag 8BP+*gag* RNA fragment and were predicted to form an independent extended stem loop (Figure 78). Other differences in the PBS motif between WT and J9+Gag could be explained by notable differences in NMIA reactivity of nucleotides 165-172 and 183-187, with nucleotides of WT being mainly unreactive and of J9+Gag being moderately to highly reactive. Surprisingly, SL1 also mapped accurately to the previously described WT structures and was not extended as initially designed. SL2 was not formed at all, but was instead observed to interact with sequences of the PBS. Lastly, SL3 was observed to be extended and to contain the sequences around the Gag start codon, which were previously described to either interact with sequences of the U5 (WT 8BP) or to independently form SL4 (WT 8BP+*gag*).

The J9+Gag mutations were introduced based on available data from WT structures of the 8BP length⁸. However, as shown in section 7.3, addition of *gag* sequences had a profound effect on the dimeric structure of the HIV-1 leader with three observed interactions between sequences of the HIV-1 leader and the *gag* region. These structural differences could not have been predicted before studying the effect of *gag* on the WT packaging conformation and were therefore not taken into consideration when the transfer vector mutants were first designed. The newly studied HIV-1 leader-*gag* interactions could explain the unexpected effect of the J9+Gag mutations on the predicted RNA structure as according to data from section 7.3, nucleotides of the HIV-1 leader are engaged in different structural motifs than previously thought. The profound disruption of the J9+Gag introduced mutations on the structural motifs of the packaging signal could explain the negative effect of these introduced mutations on packaging and transduction efficiencies. However, neither the structural modelling, which predicted shortening of SL1 and lengthening of SL3, nor the NMIA reactivity values and free energy calculations that indicated a decrease in J9+Gag

structural stability, could explain the increased propensity of this mutant to dimerise *in vitro*.

7.5. Structural analysis of the S1 mutant

S1 mutant, originally designed by Abbink *et al.* to have a strengthened U5-AUG base pair interaction⁷, was the only rationally designed mutant that was measured to have equivalent packaging and transduction efficiencies to WT. This mutant was furthermore characterised by an increased dimerisation propensity both *in vitro* and *in vivo*. Hence based on its interesting phenotype, S1 mutant was chosen for further investigation. Four independent rounds of in-gel SHAPE were performed and *gag* containing S1 dimeric and monomeric RNAs were excised, probed with 8BP primers and studied as described in earlier sections.

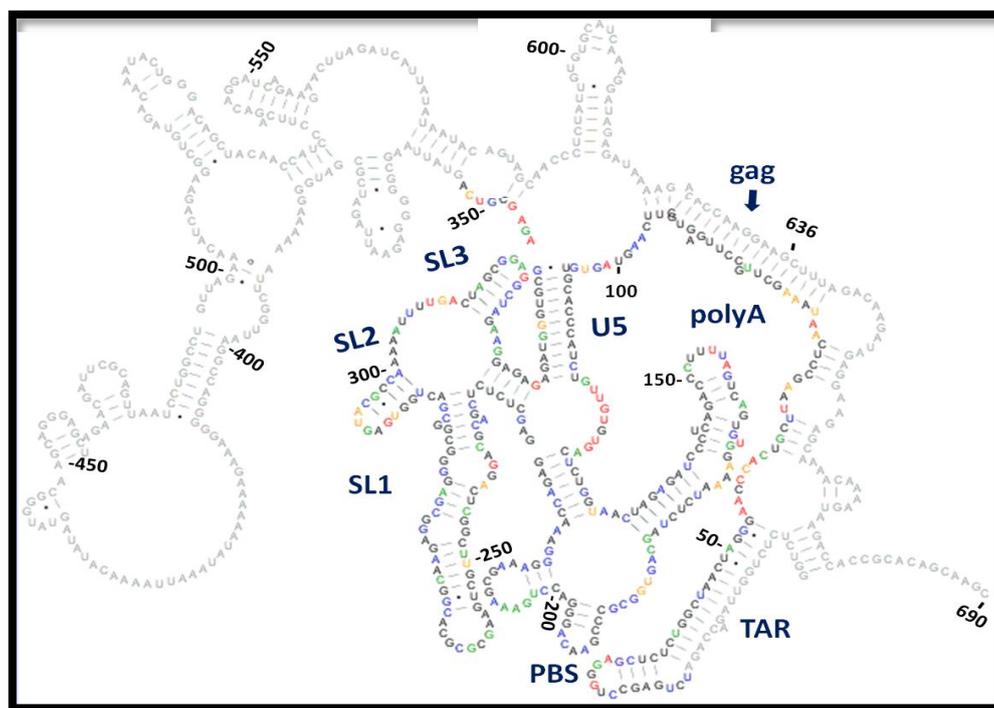


Figure 79. Structure Illustration of a 690 nt long monomeric RNA fragment including sequences of the introduced S1 mutations in the HIV-1 leader and partial *gag* of the 3rd generation HIV-1 derived lentiviral transfer vector with XRNA.

Average NMIA reactivity values were retrieved from four rounds of in-gel SHAPE experiments. NMIA reactivities for nucleotides 1-356 and the -999 simulating value for nucleotides 357-690 were loaded on RNAstructure software. The predicted interactions were used as a template for structure illustration with XRNA.

Monomeric S1 8BP+*gag* RNA exhibited several structural similarities with the dimeric WT RNAs. As per usual NMIA reactivity data mapped accurately on previously described TAR structures. However, unlike the monomeric WT 8BP+*gag* RNA that was characterised by the formation of polyA, the monomeric S1 8BP+*gag* was characterised by a polyA-*gag* interaction, a common feature of the *gag*-containing dimeric RNAs. Downstream of the polyA-*gag* interaction, S1 8BP+*gag* RNA saw the formation of the U5-AUG structural motif, previously described in the WT 8BP dimeric RNA. Additionally, SL1 formation was observed in the S1 monomeric RNA, which was previously documented to be mainly formed in dimeric RNAs (with or without the presence of *gag*). Data mapped accurately onto the PBS, SL2 and SL3 structures of the monomeric S1 8BP+*gag* RNA.

The formation of the U5-AUG and SL1 structural motifs observed in monomeric S1 8BP+*gag* RNA have so far mainly been described as characteristic features of the secondary structures of dimeric RNAs. The predicted structure suggests that the introduced S1 mutations indeed strengthened the U5-AUG interaction and promoted characteristics of dimerisation, including the formation, exposure and availability of SL1.

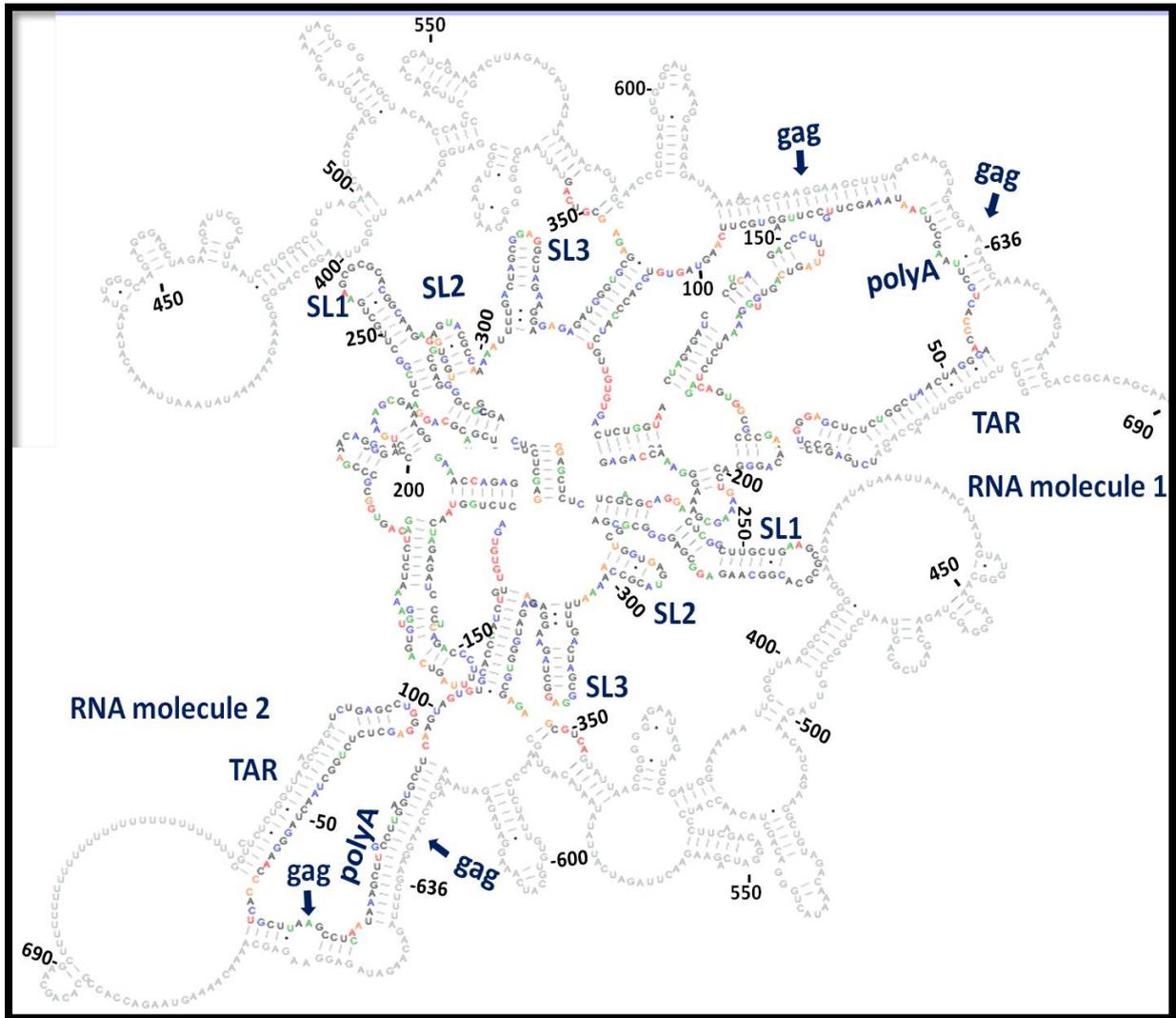


Figure 80. Structure Illustration of a 690 nt long dimeric RNA fragment including sequences of the introduced S1 mutations in the HIV-1 leader and partial *gag* of the 3rd generation HIV-1 derived lentiviral transfer vector with XRNA.

Average NMIA reactivity values were retrieved from four rounds of in-gel SHAPE experiments. NMIA reactivities for nucleotides 1-356 and the -999 simulating value for nucleotides 357-690 were loaded on RNAstructure software. Predicted interactions between HIV-1 leader-gag are highlighted with an arrow.

S1 dimeric 8BP+*gag* RNA NMIA reactivity data mapped accurately on the structure of TAR and on the polyA-*gag* interaction as predicted for the WT 8BP+*gag* dimeric RNA secondary structure. However, unlike the U5-*gag* interaction observed in the WT 8BP+*gag* dimeric RNA, S1 dimeric 8BP+*gag* RNA was characterised by the formation of a U5-AUG structural motif. Notably, the S1 mutant was designed to have a strengthened U5-AUG interaction,

which as presented in Figures 79 and 80 was formed both in the monomeric and dimeric S1 8BP+*gag* RNAs. PBS, SL1-SL3 fitted the WT 8BP+*gag* RNA SHAPE data accurately. Finally, SL4 formation was not observed as the nucleotides around the *gag* start codon were engaged in the U5-AUG interaction. S1 and WT dimeric 8BP+*gag* RNAs were shown to have very similar secondary structures with the exception of the U5-*gag* interaction of WT being substituted with the formation of the U5-AUG structural motif in S1, a structural motif previously described in the WT dimeric 8BP-*gag* RNA.

7.6. Chapter conclusions

In-gel SHAPE was performed to study the secondary RNA structures of transfer vector mutants and assess the effect of the partial *gag* sequences which are included in all transfer vector constructs on RNA structure and particularly on the packaging conformation of dimeric RNAs. Protocol optimisation had to be performed due to the inefficient *in vitro* dimerisation of longer RNAs. The structure of the WT HIV-1 leader, previously studied by several groups, was re-evaluated with the use of the newly optimised protocol. It was shown that it resembled the LDI-BMH model suggested by Abbink *et al.*, including the formation of the U5-AUG structural motif, the target region of several of our rationally designed mutants, in WT 8BP-*gag* dimeric RNA but not in the WT 8BP-*gag* monomeric RNA.

Addition of *gag* sequences had no effect on the monomeric WT RNA, but a profound effect on the conformation of the dimeric RNA. Direct interactions between the polyA, U5 and PBS elements of the HIV-1 leader with *gag* were observed and validated both with sequence conservation analysis and base pairing probability calculations. The U5-AUG motif observed in the dimeric WT 8BP-*gag* RNA was substituted by a U5-*gag* interaction. Furthermore, addition of *gag* sequences downstream of the HIV-1 leader was accompanied by a decrease in SHAPE reactivity in the packaging signal, suggesting that *gag* had a stabilising effect on the structure of *psi*. The existence and importance of the suggested interactions between U5, polyA, PBS and *gag* could be examined with the use of antisense oligonucleotides designed to block the interactions between the aforementioned regions of the HIV-1 leader and the involved *gag* sequences *in vitro*. Incubation of the 690 nt long RNA fragment with a series of different antisense oligonucleotides sequences and concentrations, followed by gel electrophoresis and visualisation could be used to assess the importance of the interactions under investigation²⁷¹.

The J9+Gag mutant, designed to have an extended SL1, mainly adopted the dimeric conformation hence investigation of the monomeric conformation was rendered impossible. Structural analysis with in-gel SHAPE, revealed extension of SL3 rather than SL1 extension. Furthermore, J9+Gag 8BP+*gag* dimeric RNA structure was disturbed compared to the predicted WT 8BP+*gag* structure, as it had a disformed PBS, that could potentially be negatively affecting the process of reverse transcription leading to the observed reduction

in infectious titres, and lacked SL2. The disturbance of structural motifs that comprise the packaging signal could explain the negative effect of the J9+Gag introduced mutations on packaging and infectivity. The last mutant whose structure was interrogated was S1, which was first designed by Abbink *et al.*, to have strengthened U5-AUG base pairing. Interestingly, S1 monomer had many similarities with the dimeric RNAs, including the formation of SL1 and the U5-AUG motif. Dimeric S1 8BP+*gag* RNA was rather similar to dimeric WT 8BP+*gag* RNA, apart from the substitution of the U5-*gag* interaction (WT) with U5-AUG motif formation (S1). The predicted structure revealed that S1 mutations successfully stabilised the U5-AUG interaction and promoted the dimeric conformation. Notably, even the monomeric S1 8BP+*gag* RNA contained structural motifs, like U5-AUG and SL1 formation, previously observed mainly in dimeric RNAs. Overall, the predicted S1 structures suggest that the formation of the U5-AUG motif was important for promoting dimerisation and enhancing packaging efficiency as hypothesised. However, the biological contribution of the U5-AUG interaction could not be concluded, as its structural importance was assessed with in-gel SHAPE analysis that was performed on static structures of 690 nt long *in vitro* transcribed RNAs. A different approach would be to investigate the existence of the observed interactions between regions of the HIV-1 leader and *gag* in vivo, exploiting techniques like the recently published COMRADES, that is designed to determine interactions and structures *in vivo*²⁷². Nevertheless, despite the aforementioned limitations, combined data would suggest that RNA flexibility to adopt U5-AUG, U5-*gag*, and, potentially other interactions *in vivo* could be one of the most critical parameters in performing the different functions needed to maintain packaging and transduction efficiencies. In particular, S1 data that implied the importance of the U5-AUG interaction for improved packaging which however failed to translate in enhanced infectivity support this argument.

8. Discussion

8.1. Conclusions

Work performed for this PhD thesis targeted the dimerisation and packaging properties of HIV-1 derived lentiviral vectors to improve lentiviral vector efficiency. HIV-1 gRNA is known to act as a switch between the monomeric conformation, associated with translation, and the dimeric conformation, linked with packaging. While dimerisation positively regulates packaging^{108 109}, translation of the genome is in opposition to its encapsidation. Rationally designed transfer vector mutants that were more likely to adopt the dimeric conformation and therefore get packaged were created by targeting regions in the 5'UTR, including SL1 extension, U5-AUG duplex and polyA stabilisation (Table 10). A variety of assays were designed to assess the effects of the introduced mutations on different parts of the vector life cycle. ELLA p24 was performed to measure physical titres while flow cytometry was performed to measure functional titres. Western Blotting and ELLA p24 were performed to investigate any potential effects of these 5'UTR mutations on *in trans*-Gag processing and budding. Northern Blotting and *in vitro* dimerisation were used to investigate the propensity of WT and mutant transfer vector RNA to dimerise *in vivo* and *in vitro* respectively, while prediction of the *in vitro* transcribed RNA secondary structure was achieved with in-gel SHAPE. Lastly, the effect of mutations on the packaging efficiency of mutant transfer vectors was calculated with a novel in-house designed competitive RT-qPCR assay. Table 10 lists the names of the rationally designed transfer vector mutants and their respective targeted regions.

| Transfer Vector | Targeted region |
|------------------------|------------------------|
| U5s | U5-AUG |
| NC4s | U5-AUG |
| UC4d | U5-AUG |
| S1 | U5-AUG |
| LU5AUG | U5-AUG |
| polyAs | polyA |
| polyAsw | polyA |
| SIVS1 | SL1 |
| J9+Gag | SL1 |
| Δp1 | SL3 |

Table 10. List of rationally designed transfer vector mutants.

The mutants were first assessed for their transduction efficiency. Amongst all mutants, U5s and S1, designed to have strengthened U5-AUG base pairing, showed comparable, but slightly lower infectious titres to WT. However, the LU5AUG mutant that was also designed with a strengthened U5-AUG interaction, showed a big reduction in transduction efficiency. The dimerisation mutants showed a wide range of infectious titres, with mutant PolyAsw having functional titres comparable to WT, while J9+Gag and PolyAs were characterised by a dramatic decrease. ELISA p24 was performed next, to assess the effect of the aforementioned mutations on physical titres and budding efficiency. Extracellular ELISA p24 analysis revealed that physical titres followed the same trend as infectious titres, while intracellular ELISA p24 showed no differences between WT and the U5-AUG mutants but a relatively higher p24 V/C ratio for the remaining dimerisation mutants compared to WT.

Biochemical vector characterisation was completed with Western Blotting used to investigate whether the effect of 5' UTR mutations impacted the processing of Gag provided *in trans*. Western Blotting visualisation showed no change in the Gag processing pattern of mutants compared to WT.

It was concluded that the variety of observed effects on functional titres reflected the multifunctionality of the targeted areas. Mutations in the DSL, U5-AUG and polyA were likely to affect not only the dimerisation and packaging properties of HIV-1 derived lentiviral vectors, but also other important parts of the HIV-1 life cycle essential for the infectivity of the newly synthesised mutant viral particle. Vector titration revealed an overall negative effect of the rationally designed mutations on vector infectivity. These experiments were performed using the optimised transfection protocol presented in section 4.2. A different approach would be to perform these infectivity experiments under conditions in which availability of dimerised genomic RNA would be limiting for packaging. This could be achieved by performing a number of independent transfections with the use of the packaging plasmids and a series of reduced WT pCCL-eGFP transfer vector DNA mass. The use of a 'buffer plasmid' is recommended to maintain the total amount of DNA at 25 µg to avoid negatively affecting transfection efficiency. Infectivity experiments could then be performed under the identified limiting conditions. The use of dimerisation promoting mutants under limiting conditions might enhance their biological effects on infectivity and improve our understanding of the mutant phenotypes.

The performed biochemical assays could not provide any information on the impact of mutations on RNA dimerisation efficiency, which was the focus of this PhD thesis. To assess this, Northern Blotting was performed to measure the total amounts of encapsidated gRNA packaged and its propensity to dimerise. Northern Blotting revealed an increased RNA dimerisation propensity trend for S1 and U5s. Northern Blotting was initially chosen as the most appropriate technique due to its advantage of measuring both the quantity of total gRNA levels and the stability of its dimeric form. However, *in vitro* dimerisation assays had to be employed to further study the propensity of RNA to dimerise for mutants whose gRNA was not detectable with Northern Blotting due to low encapsidation levels. *In vitro* dimerisation assays showed an overall positive effect of the introduced mutations on the propensity of RNA to dimerise *in vitro*. Overall findings suggest that enhancing dimerisation

did not automatically lead to improvement of vector infectivity. In general, even though *in vitro* dimerisation information was significant for understanding the RNA biology of dimerisation, it was not possible to assess a potential link between gRNA dimerisation and vector infectivity from data that was not derived from *in virio* assays. The requirement for high yields and lack of degradation throughout the whole process challenged measurements of packaged gRNA levels for transfer vectors whose virion production and/or RNA stability could have been affected by the introduced mutations. RT-qPCR has the advantage of quantifying RNA expressed at very low levels and was therefore chosen as the best assay to overcome these technical hurdles and study in detail the extent to which the rationally designed mutations affected packaging *per se*.

A novel competitive RT-qPCR assay was designed and employed over previously published qPCR assays, to investigate whether the dimerisation enhancing mutations promoted packaging as originally hypothesised and to study the relationship between packaging efficiency and RNA dimerisation propensity. The qPCR assay was optimised and then used to assess the relative packaging efficiency of U5s, S1, J9+Gag and Δ P1 lentiviral vector mutants. Δ P1 mutant, previously published to package less than 2% of WT gRNA, was used as a negative control to assess the LOD of our assay. WT/ Δ P1 RPE was calculated to be 85, which meant that the packaging efficiency of Δ P1 was 1% to WT's, confirming the previously published data. WT/U5s RPE ratio was measured to be 2.4, while the equivalent transduction efficiency ratio was calculated to be 1.7, suggesting that impact of introduced mutations on packaging directly affected the functional titres of the U5s vector. Interestingly, even though the J9+Gag mutations caused only an 11 times decrease in packaging efficiency, they caused a 50 fold decrease in transduction efficiency. This result implied that the J9+Gag mutations, in addition to packaging, must have also affected other parts of the HIV-1 vector life cycle; potentially reverse transcription. Finally, S1 was the only mutant that had a packaging efficiency equivalent of WT's and more specifically slightly higher. This result was unexpected given the previously measured slightly lower functional titres of S1 in comparison to WT, and reinforced once again the multifunctionality of the targeted regions. Based on this observation, it was concluded that despite the packaging advantage of S1 transfer vector mutant, the introduced mutations must have negatively affected different to packaging processes of the HIV-1 vector life cycle. This topic is analysed

further in section 8.2.3. Overall, our in house designed and optimised competitive qPCR assay could be used to measure a broad range of relative packaging efficiencies of mutants accurately. This range covered mutants comparably efficient to WT to mutants that encapsidated at approximately 1% of WT levels. Nevertheless it might be advantageous to further assess this competitive assay with digital droplet PCR (dPCR). dPCR is performed by dividing the sample into separate reaction droplets, so that each reaction droplet would contain either only 1 or 0 target molecules providing this way an improved approach for sensitive absolute quantification²⁷³. Lastly, with further optimisation this assay could also be used for measuring vector copy number of transduced cells.

I had aimed to optimise HIV-1 derived lentiviral vectors through structure informed genome modifications. In other words, mutations were designed and introduced based on structural information rather than sequence importance. It was therefore essential to finally study the direct impact of these rationally designed mutations on the RNA structure. In-gel SHAPE was performed to study the structures of lentiviral vector RNA mutants and to explore the influence of adjacent to *psi* sequences on its structure. Previous efforts to remove *gag* sequences from the transfer vector backbone have been unproductive and *gag* deletion experiments had revealed that in addition to the HIV-1 leader sequences, a significant portion of *gag* was required to mediate efficient RNA packaging²⁶². Even though the structure of the WT HIV-1 leader had been extensively studied^{8 60 61 62}, the biological role of the downstream *gag* sequences remained unknown. Addition of *gag* sequences had no effect on the monomeric WT RNA, but a profound effect on the packaging conformation of the dimeric RNA. Direct interactions between elements of the HIV-1 leader, including the polyA, U5 and PBS, with *gag* were documented here for the first time. Potential interactions of these aforementioned *cis*-acting elements and *gag* were further validated with sequence conservation analysis and base pairing probability calculations. The U5-AUG motif previously observed in the dimeric WT RNAs was substituted by a U5-*gag* interaction, while the formation of an independent polyA stem loop was substituted with a polyA-*gag* interaction in the WT 8BP+*gag* RNA fragment. Furthermore, addition of *gag* sequences downstream of the HIV-1 leader was accompanied by a decrease in reactivity in the packaging signal, suggesting that *gag* promotes a more compact and stable *psi* structure. Lastly, the *gag* sequences were mainly observed to be single stranded in the monomeric RNA, whereas the

regions of the *gag* sequences were predicted to form base pairs with sequences of the HIV-1 leader, implying the potential participation of *gag* in regulating the role of HIV-1 gRNA to switch between alternative conformations that serve different functions. My data supported recent studies that had suggested a role of *gag* as a *cis*-acting element, important for RNA packaging, structure and function^{262 263}. J9+Gag and S1 mutants were chosen for structural analysis based on their overall interesting phenotypes. Briefly, the J9+Gag mutant displayed a very disturbed RNA secondary structure with an extended SL3 instead of an extended SL1 as originally designed, a differently formed PBS and the absence of SL2. The effect of the introduced mutations on multiple structural motifs could be a potential explanation for the discrepancy between vector infectivity and the relative packaging efficiency of the J9+Gag transfer vector. Based on the predicted dimeric structure of J9+Gag, it was hypothesised that these mutations had a negative effect on packaging due to the disturbance of SL2 and SL3 that are part of the packaging signal and on transduction efficiency potentially because of an impact on reverse transcription due to PBS malformation. S1 was the last mutant whose RNA secondary structure was interrogated. The predicted S1 structure revealed that mutations successfully stabilised the U5-AUG interaction and promoted the dimeric conformation. Notably, even the monomeric S1 8BP+*gag* RNA contained structural motifs, like U5-AUG and SL1 formation, previously mainly observed in dimeric RNAs. Overall, the predicted S1 structures suggest that the formation of the U5-AUG motif was important for promoting dimerisation and enhancing packaging efficiency as hypothesised. However, this observation was made by analysing a single static RNA structure. The ability of RNA flexibility to adopt U5-AUG, U5-*gag*, and, potentially other interactions *in vivo*, could be one of the most critical parameters for maintaining packaging and transduction efficiencies. In support of this argument is the observation that the increase in dimerisation and packaging efficiency of S1 did not translate to an increase of functional titres as expected. S1 and J9+Gag structural findings suggest that the ratio of monomeric to dimeric RNA and RNA's flexibility to alternate between its two states and fates must be essential for vector infectivity. Lastly, the discovery of the polyA-*gag* interaction on the RNA level made the polyAs and polyAsw of great interest for structural analysis. However due to time constraints this has not been possible, but interrogation of the RNA structure of polyAs and polyAsw would be of great significance to further validate the polyA-*gag* interaction and to understand the significant infectivity difference between

these two polyA mutants that were designed with the same rationale before the finding of the polyA-*gag* interaction. Additionally, it might be beneficial to perform a new round of rationally designed dimerisation promoting mutants after taking into consideration these new interesting structural findings, including the interactions of the HIV-1 leader sequences with *gag* sequences. A potential strategy is further analysed in 8.2.1.

To conclude, this thesis presented the design, generation, biochemical and structural characterisation of dimerisation promoting transfer vector mutants based on the hypothesis that WT HIV-1 regulates genome encapsidation tightly by recognising dimeric RNA¹². These dimerisation enhancing transfer vector mutants, whose RNA should be more likely to adopt the dimeric conformation and therefore be packaged, were created based on previously published structural studies and our mFold structural predictions and were employed to investigate a relationship between dimerisation, packaging and transduction efficiencies. Biochemical characterisation showed that transduction and packaging efficiencies were highly influenced by the region where the mutations were introduced. Both Northern blots and *in vitro* dimerisation assays confirmed that the propensity of mutated vector RNA to dimerise had indeed increased in the mutants as hypothesised. However, relative packaging efficiency data, acquired with the use of a novel, competitive, in-house developed RT-qPCR assay, suggest that enhancing dimerisation does not automatically lead to better packaging of vector RNA and that the introduced mutations might be affecting multiple vector processes, reflecting the multifunctionality of targeted regions. Finally, I explored by in gel-SHAPE the structure of *psi* in our transfer vector RNAs, in particular by studying the influence of regions adjacent to *psi* on dimerisation and packaging. This single nucleotide level structural analysis revealed that the presence of *gag* sequences stabilise the *psi* element of the dimeric RNA by direct interactions with the polyA, U5 and PBS of the HIV-1 leader, suggesting their role in supporting a stable RNA conformation that can be packaged and offering a potential explanation for their requirement in the transfer vector plasmid for maintenance of infectious titres. Finally, structural analysis of mutants confirmed the importance of the U5-AUG duplex in the dimeric RNA conformation. These findings will give us better insights into the biology of lentiviral vectors and enable us to design more efficient vectors for the demanding future of gene therapy.

8.2. Future directions

8.2.1. *In vivo* SELEX for the production and selection of best replicating variants

This thesis employed the introduction of specific, rationally designed mutations, which target the dimerisation and packaging properties of lentiviral vectors. However this approach of rationally designed mutations can only allow for the production of a limited number of different transfer vectors. The *in vivo* SELEX (systematic evolution of ligands by exponential enrichment) approach involves the creation of HIV-1 lentiviral vector variants libraries in which targeted segments of interest are randomized for selection of the best replicating variants²⁷⁴. Thus, *in vivo* SELEX could serve as an alternative way to produce and select for mutated lentiviral constructs with enhanced genome encapsidation properties in a larger scale. Briefly, this process involves either the design of various oligos bearing random mutations in selected regions or error prone PCR performed in a template of interest including sequences that are known to be important for RNA encapsidation. Once the targeted segment is mutated, a mass-transformation would be performed at the level of the transfer vector. The produced transfer vector library would be assessed with sequencing to confirm both the variety of mutants and the lack of dominant mutants. Once the generation of the transfer vector library is achieved, the variety of mutated constructs would be used for co-transfections with the packaging plasmids for vector production. Generated vectors would be harvested and used for multiple rounds of transductions, giving the opportunity to the system to select and purify out the vectors carrying the sequences that lead to higher transduction efficiency. Multiple rounds of transduction aim to place the library under selective pressure *in vivo*. *In vivo* SELEX has previously been used in HIV-1 research to study the importance of the few well-conserved single-stranded sequences that connect the hairpins of the 5' UTR²⁷⁵.

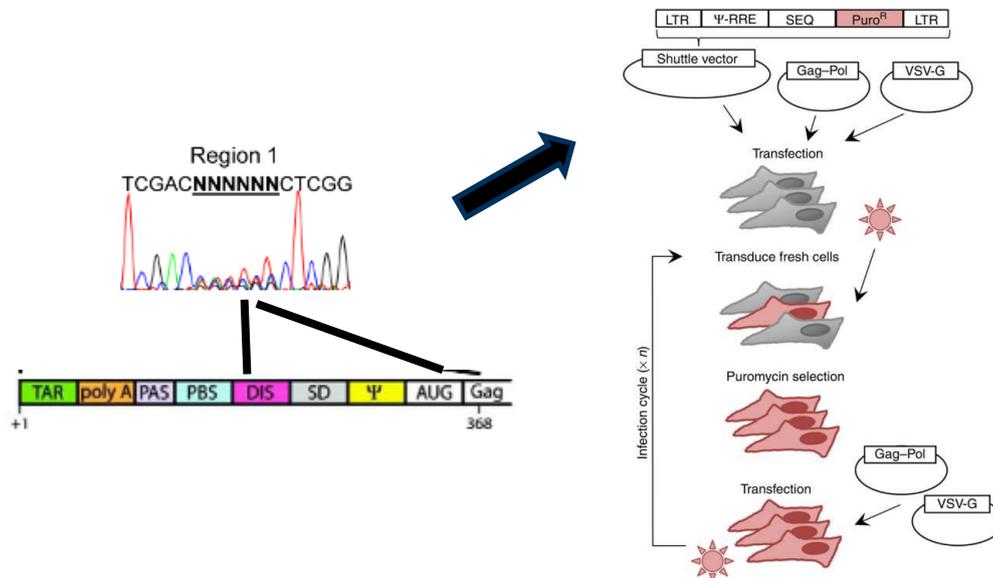


Figure 81. In vivo SELEX approach for the creation of HIV-1 derived lentiviral vector libraries to select the best replicating variants. Adapted from source²⁷⁶.

Sequences of partial *gag* can serve as a potential region of interest, since this thesis has shown the direct interaction between *gag* and 5'UTR sequences. In our system all mutations introduced in the 5'UTR of HIV-1 have been associated with a negative effect on vector transduction efficiency. The 5'UTR is the most conserved part of the HIV-1 RNA genome²⁷⁷ and is known to promote packaging, regulate transcriptional activation, splicing, primer binding during reverse transcription, and dimerisation⁷⁰. On the contrary, partial *gag* sequences are relatively variable and could therefore be potentially targeted with a directed evolution experiment.

The transfer vector of this lentiviral system is a 3rd generation Self Inactivating Vector (SIN), whose 5' U3 is substituted with the CMV promoter and its 3' U3 is partially deleted. Upon reverse transcription, the incorporated genome will be promoter-less. The carry-out of the described SELEX experiment, that requires active competition of the vectors, needs a mobile vector. To achieve this, an option would be to clone the WT-HIV-1 3'U3 back to the transfer vector construct and combine it with the use of a 2nd generation *tat-gag/pol* construct to attain efficient levels of transcription.

8.2.2. Improvement of gRNA stability

RNA is a single stranded molecule made of ribose sugars which contain a hydroxyl group at the 2' position. These traits render this molecule chemically unstable and thus more prone to alkaline hydrolysis, making RNA an intrinsically unstable molecule²⁷⁸. Auto-hydrolysis, self-cleavage of RNA that can occur spontaneously without the presence of an enzyme, is more likely to occur when the RNA is single-stranded and when the solution of the pH is basic. It could be hypothesised that during the various laboratory manipulations performed to process the vector samples, a certain proportion of RNA would get hydrolysed due to the slightly basic pH of the protease buffer, thus affecting the availability of intact extracted gRNA molecules to be assessed with techniques like Northern Blotting. Furthermore, HIV-1 is known to contain instability sequences (INS) in *gag/pol*²⁷⁹ and RRE²⁸⁰ that cause inefficient expression by impairing nucleocytoplasmic transport and stability that result to transcript degradation²⁸⁰. Hence, addition of a stability element could alleviate both any potential impact of vector processing on gRNA stability and the negative effect of INS on posttranscriptional regulation. Lastly, Nikolaitchik *et al.*, speculated that larger inserts affect viral fitness due to inefficient expression and lower RNA stability rather than decreased packaging efficiency as initially hypothesised¹². Therefore, introduction of a stability element could potentially reduce the fitness cost caused by a larger genome and hence improve the transduction efficiency of lentiviral vector expressing larger therapeutic transgenes.

The aforementioned challenges caused by the nature of HIV-1 transfer vector gRNA could be overcome by engineering their basic backbone by introducing stability elements from other viruses. Palusa *et al.*, described the identification of an RNA element in Rabies Virus that stabilises transcripts via interacting with the cellular poly(rC)-binding PCBP2 protein²⁸¹. The authors had initially noticed the differential expression of the glycoprotein mRNA of the virus compared to the rest of the transcripts. This observation was followed by identification of a conserved region in the 3'UTR of the glycoprotein mRNA that was found to act as a stability element. Isolation and cloning of this region into reporter constructs revealed an approximate 2-fold stabilisation of the transcript compared to the control RNA²⁸¹.

Based on these encouraging findings, this Rabies' stability element or an equivalent human analogue could be cloned into the basic backbone of the transfer vector. Ideally, the stability

element of choice should be cloned in various locations within the transfer vector, to investigate any potential positional effects. These engineered vector constructs can then be used for co-transfections with the packaging plasmids and their extracellular and intracellular RNAs can be extracted to evaluate the impact of the stability element in comparison to the control WT sequences of the current lentiviral vector system.

8.2.3. Investigation of the effect of secondary gRNA structure on other parts of the HIV-1 life cycle

The secondary structure of viral RNAs is known to have an impact on the efficiency of reverse transcription²⁸² and on immune response induction through the RIG-I pathway²⁸³. This thesis has addressed the effect of the 5'UTR mutations on dimerisation, packaging and transduction efficiency. However, it was observed that the relationship between these processes was not direct for all mutants. For example, despite S1 mutant being shown to have higher dimerisation and packaging efficiencies than WT, its transduction efficiency was measured to be lower than WT's. This unorthodox result implicates the contribution of pathways different to the ones involved in dimerisation and packaging. Given that mutant transfer vector gRNAs of this thesis were shown to adopt secondary structures different to WT's, it could be hypothesised that these differences in the secondary structure of RNA are accountable for this observed phenomenon due to their effects on reverse transcription efficiency and/or triggering of immune sensing. *In vitro* reverse transcription can be performed to assess the effect of different RNA templates on the efficiency of reverse transcriptase by measuring the incorporation of digoxigenin-labelled dUTP in newly synthesized DNA²⁸⁴. Cell based signalling assays can be employed to measure RIG-I activation by gRNAs of different structures²⁸⁵. Investigation of the effects of structure informed genome modifications on other parts of the life cycle will elucidate the importance of alternative pathways for lentiviral vector optimisation.

9. Bibliography

1. German Advisory Committee Blood (Arbeitskreis Blut), Subgroup 'Assessment of Pathogens Transmissible by Blood', G. A. C. B. (Arbeitskreis & Blood', S. 'Assessment of P. T. by. Human Immunodeficiency Virus (HIV). *Transfus. Med. Hemother.* **43**, 203–22 (2016).
2. HIV Databases Review Article. Available at: <https://www.hiv.lanl.gov/content/sequence/HIV/REVIEWS/WIGDAHL2001/Wigdahl.html>. (Accessed: 10th August 2019)
3. Freed, E. O. HIV-1 replication. *Somat. Cell Mol. Genet.* **26**, 13–33 (2001).
4. Landmarks of the HIV genome. Available at: <https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html>. (Accessed: 10th August 2019)
5. Seelamgari, A. *et al.* Role of viral regulatory and accessory proteins in HIV-1 replication. *Front. Biosci.* **9**, 2388–413 (2004).
6. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254–257 (1989).
7. Abbink, T. E. M. & Berkhout, B. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 rna occludes the gag start codon. *J. Biol. Chem.* **278**, 11601–11611 (2003).
8. Kenyon, J. C. & Lever, A. M. L. In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. *Nucleic Acids Res.* **41**, 1–11 (2013).
9. Olson, E. D. & Musier-Forsyth, K. New structure sheds light on selective HIV-1 genomic RNA packaging. *Viruses* **7**, 4826–4835 (2015).
10. Chen, J. & Hu, W.-S. HIV-1 RNA genome dimerizes on the plasma membrane in the presence of Gag protein. *Proc. Natl. Acad. Sci.* **113**, E201-8 (2016).
11. Larson, D. R., Ma, Y. M., Vogt, V. M. & Webb, W. W. Direct measurement of Gag-Gag interaction during retrovirus assembly with FRET and fluorescence correlation spectroscopy. *J. Cell Biol.* **162**, 1233–1244 (2003).
12. Nikolaitchik, O. A. & Hu, W. Dimeric RNA Recognition Regulates HIV-1 Genome Packaging. *Plos Pathog.* **9**, e1003249. doi:10.1371/journal.ppat.1003249 (2013).

13. VerPlank, L. & Carter, C. A. Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7724–9 (2001).
14. Liang, C. & Wainberg, M. A. Deletion Mutagenesis within the Dimerization Initiation Site of Human Immunodeficiency Virus Type 1 Results in Delayed Processing of the p2 Peptide from Precursor Proteins. *J. Virol.* **73**, 6147–6151 (1999).
15. Kemler, I., Meehan, A. & Poeschla, E. M. Live-Cell Coimaging of the Genomic RNAs and Gag Proteins of Two Lentiviruses. *J. Virol.* **84**, 6352–6366 (2010).
16. Lévesque, K. *et al.* Trafficking of HIV-1 RNA is mediated by heterogeneous nuclear ribonucleoprotein A2 expression and impacts on viral assembly. *Traffic* **7**, 1177–93 (2006).
17. Jouvenet, N., Simon, S. M. & Bieniasz, P. D. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc. Natl. Acad. Sci.* **106**, 19114–19119 (2009).
18. Kutluay, S. B. & Bieniasz, P. D. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell* **159**, 1096–1109 (2014).
19. Lever, A. & Sodroski, J. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* **63**, 4085–7 (1989).
20. Chen, J. & Hu, W.-S. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13535–40 (2009).
21. Poon, D. T., Wu, J. & Aldovini, A. Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J. Virol.* **70**, 6607–16 (1996).
22. Aldovini, A. & Young, R. A. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* **64**, 1920–6 (1990).
23. Rein, A. RNA Packaging in HIV. *Trends Microbiol.* **27**, 715–723 (2019).
24. Levin, J. G., Guo, J., Rouzina, I. & Musier-Forsyth, K. Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein: Critical Role in Reverse Transcription and Molecular Mechanism. in *Progress in nucleic acid research and molecular biology* **80**, 217–286 (2005).

25. Rein, A. Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol.* **7**, 700–705 (2010).
26. Poon, D. T. K. & Aldovini, A. Charged Amino Acid Residues of Human Immunodeficiency Virus Type 1 Nucleocapsid p7 Protein Involved in RNA Packaging and Infectivity. *J. Virol.* **70**, 6607–6616 (1996).
27. Luban, J. & Goff, S. P. Mutational Analysis of cis-Acting Packaging Signals in Human Immunodeficiency Virus Type 1 RNA. **68**, 3784–3793 (1994).
28. Gorelick, R. J. & Rein, A. Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **64**, 3207–11 (1990).
29. Rein, A. & Musier-Forsyth, K. Diverse interactions of retroviral Gag proteins with RNAs. *Trends Biochem. Sci.* **36**, 373–380 (2011).
30. Molecular Biology Review. Available at: https://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioReview/central_dogma.html. (Accessed: 10th August 2019)
31. Neveu, M., Kim, H.-J. & Benner, S. A. The “Strong” RNA World Hypothesis: Fifty Years Old. *Astrobiology* **13**, 391–403 (2013).
32. The RNA World, 2nd Ed.: The Nature of Modern RNA Suggests a Prebiotic RNA World (1999, Volume 37). Available at: <https://cshmonographs.org/index.php/monographs/issue/current>. (Accessed: 10th August 2019)
33. Alberts, B. *et al.* The RNA World and the Origins of Life. (2002).
34. Bevilacqua, P. C., Ritchey, L. E., Su, Z. & Assmann, S. M. Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.* **50**, 235–266 (2016).
35. Harcourt, E. M., Kietrys, A. M. & Kool, E. T. Chemical and structural effects of base modifications in messenger RNA. *Nature* **541**, 339–346 (2017).
36. Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).
37. Deng, H. *et al.* Rice In Vivo RNA Structurome Reveals RNA Secondary Structure Conservation and Divergence in Plants. *Mol. Plant* **11**, 607–622 (2018).
38. Robertson, M. P. & Joyce, G. F. The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **4**, (2012).
39. Lengyel, J., Hnath, E., Storms, M. & Wohlfarth, T. Towards an integrative structural

- biology approach: combining Cryo-TEM, X-ray crystallography, and NMR. *J. Struct. Funct. Genomics* **15**, 117–124 (2014).
40. Hawkes, E. J. *et al.* COOLAIR Antisense RNAs Form Evolutionarily Conserved Elaborate Secondary Structures. *Cell Rep.* **16**, 3087–3096 (2016).
 41. Kwok, C. K., Ding, Y., Tang, Y., Assmann, S. M. & Bevilacqua, P. C. Determination of in vivo RNA structure in low-abundance transcripts. *Nat. Commun.* **4**, 2971 (2013).
 42. Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci.* **108**, 11063–11068 (2011).
 43. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965 (2014).
 44. Smola, M. J. *et al.* SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the *Xist* lncRNA in living cells. *Proc. Natl. Acad. Sci.* **113**, 10322–10327 (2016).
 45. Spitale, R. C. *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
 46. Kenyon, J. C., Prestwood, L. J. & Lever, A. M. L. A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Sci. Rep.* **5**, 14369 (2015).
 47. Ehresmann, C. *et al.* Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 9109–9128 (1987).
 48. Soukup, G. A. & Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308–25 (1999).
 49. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
 50. Badorrek, C. S. & Weeks, K. M. RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat. Chem. Biol.* **1**, 104–111 (2005).
 51. Badorrek, C. S., Gherghe, C. M. & Weeks, K. M. Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization. *Proc. Natl. Acad. Sci.* **103**, 13640–13645 (2006).

52. Duncan, C. D. S. & Weeks, K. M. SHAPE Analysis of Long-Range Interactions Reveals Extensive and Thermodynamically Preferred Misfolding in a Fragile Group I Intron RNA[†]. *Biochemistry* **47**, 8504–8513 (2008).
53. Vicens, Q., Gooding, A. R., Laederach, A. & Cech, T. R. Local RNA structural changes induced by crystallization are revealed by SHAPE. *RNA* **13**, 536–48 (2007).
54. Mortimer, S. A. & Weeks, K. M. A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. *J. Am. Chem. Soc.* **129**, 4144–4145 (2007).
55. Vasa, S. M., Guex, N., Wilkinson, K. A., Weeks, K. M. & Giddings, M. C. ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14**, 1979–1990 (2008).
56. HIV Sequence Database: 2016 Compendium. Available at: <https://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2016compendium.html>. (Accessed: 10th August 2019)
57. Baudin, F. *et al.* Functional Sites in the 5′ Region of Human Immunodeficiency Virus Type 1 RNA Form Defined Structural Domains. *J. Mol. Biol.* **229**, 382–397 (1993).
58. Ooms, M., Huthoff, H., Russell, R., Liang, C. & Berkhout, B. A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J. Virol.* **78**, 10814–9 (2004).
59. Huthoff, H. & Berkhout, B. E. N. Two alternating structures of the HIV-1 leader RNA. *RNA* **(1)**, 143–157 (2001).
60. Keane, S. C. & Summers, M. F. RNA Structure. Structure of the HIV-1 RNA packaging signal. *Science (80-.)*. **22;348(623)**, doi: 10.1126/science.aaa9266. (2015).
61. Lu, K. & Summers, M. F. NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. *Science* **334**, 242–5 (2011).
62. Keane, S. C. & Summers, M. F. NMR Studies of the Structure and Function of the HIV-1 5′-Leader. *Viruses* **8**, (2016).
63. Roy, S., Delling, U., Chen, C. H., Rosen, C. A. & Sonenberg, N. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev.* **4**, 1365–1373 (1990).
64. Russell, R. S., Hu, J., Laughrea, M., Wainberg, M. A. & Liang, C. Deficient Dimerization of Human Immunodeficiency Virus Type 1 RNA Caused by Mutations of the U5 RNA

- Sequences. *Virology* **303**, 152–163 (2002).
65. Das, A. T., Koken, S. E. C., Oude Essink, B. B., van Wamel, J. L. B. & Berkhout, B. Human immunodeficiency virus uses tRNA^{Lys,3} as primer for reverse transcription in HeLa-CD4⁺ cells. *FEBS Lett.* **341**, 49–53 (1994).
 66. Berkhout, B. E. N. & Wamel, J. L. B. V. A. N. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA* **6(2)**, 282–295 (2000).
 67. Harrison, G. P. & Lever, A. M. L. The Human Immunodeficiency Virus Type 1 Packaging Signal and Major Splice Donor Region Have a Conserved Stable Secondary Structure. **66**, 4144–4153 (1992).
 68. Clever, J. L., Miranda, D. & Parslow, T. G. RNA Structure and Packaging Signals in the 5' J Leader Region of the Human Immunodeficiency Virus Type 1 Genome. *J. Virol.* **76**, 12381–12387 (2002).
 69. van Bel, N., Ghabri, A., Das, A. T. & Berkhout, B. The HIV-1 leader RNA is exquisitely sensitive to structural changes. *Virology* **483**, 236–252 (2015).
 70. Coffin, J. M., Hughes, S. H. & Varmus, H. E. *Retroviruses*. *Retroviruses* (Cold Spring Harbor Laboratory Press, 1997).
 71. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–6 (2009).
 72. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965 (2014).
 73. Sükösd, Z. *et al.* Full-length RNA structure prediction of the HIV-1 genome reveals a conserved core domain. *Nucleic Acids Res.* **43**, 10168–79 (2015).
 74. Knoepfel, S. A. & Berkhout, B. On the role of four small hairpins in the HIV-1 RNA genome. *RNA Biol.* **10**, 540–552 (2013).
 75. Simmonds, P., Tuplin, A. & Evans, D. J. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *RNA* **10**, 1337–51 (2004).
 76. Pollom, E. *et al.* Comparison of SIV and HIV-1 Genomic RNA Structures Reveals Impact of Sequence Evolution on Conserved and Non-Conserved Structural Motifs. *PLoS Pathog.* **9**, e1003294 (2013).
 77. Poole, A. & Penny, D. Retroviral recombination : what drives the switch ? *Nat. Rev.*

- Mol. Cell Biol.* **2**, 1–5 (2001).
78. Hu, W.-S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med.* **2**, a006882 (2012).
 79. Robinson, W. S., Robinson, H. L. & Duesberg, P. H. Tumor virus RNA's. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 825–34 (1967).
 80. Duesberg, P. H. Physical properties of Rous Sarcoma Virus RNA. *Proc. Natl. Acad. Sci. U. S. A.* **60**, 1511–8 (1968).
 81. Khandjian, E. W. & Méric, C. A procedure for Northern blot analysis of native RNA. *Anal. Biochem.* **159**, 227–232 (1986).
 82. Chin, M. P. S. & Hu, W.-S. Identification of a major restriction in HIV-1 intersubtype recombination. *Proc. Natl. Acad. Sci.* **102**, 9002–9007 (2005).
 83. Tran, T. & Summers, M. F. Conserved determinants of lentiviral genome dimerization. *Retrovirology* **29;12:83**, 1–16 (2015).
 84. Paillait, J.-C., Skripkin, E., Ehresmann, B., Ehresmann, C. & Marquet, R. The use of chemical modification interference and inverse PCR mutagenesis to identify the dimerization initiation site of HIV-1 genomic RNA. *Pharm. Acta Helv.* **71**, 21–28 (1996).
 85. Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B. & Ehresmann, C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4945–9 (1994).
 86. Muriaux, D., Girard, P.-M., Bonnet-Mathonière, B. & Paoletti, J. Dimerization of HIV-1_{Lai} RNA at Low Ionic Strength. *J. Biol. Chem.* **270**, 8209–8216 (1995).
 87. Muriaux, D., Fossé, P. & Paoletti, J. A Kissing Complex Together with a Stable Dimer Is Involved in the HIV-1_{Lai} RNA Dimerization Process *in Vitro*[†]. *Biochemistry* **35**, 5075–5082 (1996).
 88. Mujeeb, A., Parslow, T. G., Zarrinpar, A., Das, C. & James, T. L. NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA. *FEBS Lett.* **458**, 387–392 (1999).
 89. St. Louis, D. C. *et al.* Infectious Molecular Clones with the Nonhomologous Dimer Initiation Sequences Found in Different Subtypes of Human Immunodeficiency Virus Type 1 Can Recombine and Initiate a Spreading Infection In Vitro. *J. Virol.* **72**, (1998).
 90. Berkhout, B. & van Wamel, J. L. Role of the DIS hairpin in replication of human

- immunodeficiency virus type 1. *J. Virol.* **70**, 6723–32 (1996).
91. Laughrea, M., Shen, N., Jetté, L. & Wainberg, M. A. Variant Effects of Non-Native Kissing-Loop Hairpin Palindromes on HIV Replication and HIV RNA Dimerization: Role of Stem–Loop B in HIV Replication and HIV RNA Dimerization [†]. *Biochemistry* **38**, 226–234 (1999).
 92. Hill, M. K. *et al.* The dimer initiation sequence stem-loop of human immunodeficiency virus type 1 is dispensable for viral replication in peripheral blood mononuclear cells. *J. Virol.* **77**, 8329–35 (2003).
 93. Liang, C., Rong, L., Russell, R. S. & Wainberg, M. A. Deletion mutagenesis downstream of the 5' long terminal repeat of human immunodeficiency virus type 1 is compensated for by point mutations in both the U5 region and gag gene. *J. Virol.* **74**, 6251–61 (2000).
 94. Liang, C. *et al.* Mutations within four distinct gag proteins are required to restore replication of human immunodeficiency virus type 1 after deletion mutagenesis within the dimerization initiation site. *J. Virol.* **73**, 7014–20 (1999).
 95. Russell, R. S., Liang, C. & Wainberg, M. A. Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology* **1**, 23 (2004).
 96. Brandt, S. *et al.* Rev Proteins of Human and Simian Immunodeficiency Virus Enhance RNA Encapsidation. *PLoS Pathog.* **3**, e54 (2007).
 97. Adam, M. A. & Miller, A. D. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J. Virol.* **62**, 3802–6 (1988).
 98. Banks, J. D., Kealoha, B. O. & Linial, M. L. An Mpsi-containing heterologous RNA, but not env mRNA, is efficiently packaged into avian retroviral particles. *J. Virol.* **73**, 8926–33 (1999).
 99. Aldovini, A. & Young, R. A. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* **64**, 1920–6 (1990).
 100. Kim, H.-J., Lee, K. & O'Rear, J. J. A Short Sequence Upstream of the 5' Major Splice Site Is Important for Encapsidation of HIV-1 Genomic RNA. *Virology* **198**, 336–340 (1994).
 101. Gorelick, R. J. *et al.* Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **64**, 3207–11 (1990).
 102. South, T. L. & Summers, M. F. Zinc- and sequence-dependent binding to nucleic acids

- by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci.* **2**, 3–19 (1993).
103. Heng, X. *et al.* Identification of a Minimal Region of the HIV-1 5'-Leader Required for RNA Dimerization, NC Binding, and Packaging. *J. Mol. Biol.* **417**, 224–239 (2012).
 104. Sakuragi, J.-I., Sakuragi, S. & Shioda, T. Minimal region sufficient for genome dimerization in the human immunodeficiency virus type 1 virion and its potential roles in the early stages of viral replication. *J. Virol.* **81**, 7985–92 (2007).
 105. McBride, M. S. & Panganiban, A. T. The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J. Virol.* **70**, 2963–73 (1996).
 106. Clever, J., Sasseti, C. & Parslow, T. G. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J. Virol.* **69**, 2101–9 (1995).
 107. Kuzembayeva, M., Dilley, K., Sardo, L. & Hu, W.-S. Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* **454–455**, 362–370 (2014).
 108. Clever, J. L. & Parslow, T. G. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* **71**, 3407–14 (1997).
 109. Sakuragi, J. & Shioda, T. Duplication of the Primary Encapsidation and Dimer Linkage Region of Human Immunodeficiency Virus Type 1 RNA Results in the Appearance of Monomeric RNA in Virions. *J. Virol.* **75**, 2557–2565 (2001).
 110. Russell, R. S. *et al.* Effects of a single amino acid substitution within the p2 region of human immunodeficiency virus type 1 on packaging of spliced viral RNA. *J. Virol.* **77**, 12986–95 (2003).
 111. Awang, G. & Sen, D. Mode of dimerization of HIV-1 genomic RNA. *Biochemistry* **32**, 11453–11457 (1993).
 112. Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F. & Barré-Sinoussi, F. Cis elements and Trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* **216**, 689–699 (1990).
 113. Laughrea, M. *et al.* Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J. Virol.* **71**, 3397–406 (1997).

114. Berkhout, B. & van Wamel, J. L. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J. Virol.* **70**, 6723–32 (1996).
115. Feng, Y. X. *et al.* HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 7577–81 (1996).
116. Sakuragi, J., Iwamoto, A. & Shioda, T. Dissociation of genome dimerization from packaging functions and virion maturation of human immunodeficiency virus type 1. *J. Virol.* **76**, 959–67 (2002).
117. Russell, R. S. *et al.* Sequences downstream of the 5' splice donor site are required for both packaging and dimerization of human immunodeficiency virus type 1 RNA. *J. Virol.* **77**, 84–96 (2003).
118. Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* **2**, 461–472 (2004).
119. Levin, J. G., Grimley, P. M., Ramseur, J. M. & Berezsky, I. K. Deficiency of 60 to 70S RNA in murine leukemia virus particles assembled in cells treated with actinomycin D. *J. Virol.* **14**, 152–61 (1974).
120. Levin, J. G. & Rosenak, M. J. Synthesis of murine leukemia virus proteins associated with virions assembled in actinomycin D-treated cells: evidence for persistence of viral messenger RNA. *Proc. Natl. Acad. Sci.* **73**, 1154–1158 (1976).
121. Dorman, N. & Lever, A. Comparison of viral genomic RNA sorting mechanisms in human immunodeficiency virus type 1 (HIV-1), HIV-2, and Moloney murine leukemia virus. *J. Virol.* **74**, 11413–7 (2000).
122. Butsch, M. & Boris-Lawrie, K. Translation is not required To generate virion precursor RNA in human immunodeficiency virus type 1-infected T cells. *J. Virol.* **74**, 11531–7 (2000).
123. Pereira-Montecinos, C. *et al.* An epitranscriptomic switch at the 5'-UTR controls genome selection during HIV-1 genomic RNA packaging. *bioRxiv* 676031 (2019). doi:10.1101/676031
124. Stake, M. *et al.* HIV-1 and two avian retroviral 5' untranslated regions bind orthologous human and chicken RNA binding proteins. *Virology* **486**, 307–320 (2015).
125. Zhang, Q., Chen, C.-Y., Yedavalli, V. S. R. K. & Jeang, K.-T. NEAT1 Long Noncoding RNA and Paraspeckle Bodies Modulate HIV-1 Posttranscriptional Expression. *MBio* **4**, e00596-12 (2013).

126. Kaddis Maldonado, R. & Parent, L. Orchestrating the Selection and Packaging of Genomic RNA by Retroviruses: An Ensemble of Viral and Host Factors. *Viruses* **8**, 257 (2016).
127. Kharytonchyk, S. *et al.* Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13378–13383 (2016).
128. Ferrer, M., Henriët, S., Chamontin, C., Lainé, S. & Mougel, M. From Cells to Virus Particles: Quantitative Methods to Monitor RNA Packaging. *Viruses* **8**, (2016).
129. Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond)*. **27**, 113–59 (1928).
130. Avery, O. T., MacLeod, C. M. & McCarty, M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *J. Exp. Med.* **79**, 137 (1944).
131. Hershey, A. D. & CHASE, M. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. *J. Gen. Physiol.* **36**, 39–56 (1952).
132. WATSON, J. D. & CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
133. Borenfreund, E. & Bendich, A. A STUDY OF THE PENETRATION OF MAMMALIAN CELLS BY DEOXYRIBONUCLEIC ACIDS. *J. Biophys. Biochem. Cytol.* **9**, 81–91 (1961).
134. Oncogenicity and cell transformation by papovavirus SV40: the role of the viral genome. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/4333789>. (Accessed: 11th August 2019)
135. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proc. Natl. Acad. Sci.* **70**, 3240–3244 (1973).
136. Graham, F. L. & van der Eb, A. J. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52**, 456–467 (1973).
137. Mercola, K., Stang, H., Browne, J., Salser, W. & Cline, M. Insertion of a new gene of viral origin into bone marrow cells of mice. *Science (80-.)*. **208**, 1033–1035 (1980).
138. Wade, N. Gene therapy caught in more entanglements. *Science (80-.)*. **212**, 24–25 (1981).
139. Burnett, J. R. & Hooper, A. J. Alipogene tiparvovec, an adeno-associated virus encoding the Ser(447)X variant of the human lipoprotein lipase gene for the

- treatment of patients with lipoprotein lipase deficiency. *Curr. Opin. Mol. Ther.* **11**, 681–91 (2009).
140. Aiuti, A. *et al.* Gene Therapy for Immunodeficiency Due to Adenosine Deaminase Deficiency. *N. Engl. J. Med.* **360**, 447–458 (2009).
 141. South, E., Cox, E., Meader, N., Woolacott, N. & Griffin, S. Strimvelis® for Treating Severe Combined Immunodeficiency Caused by Adenosine Deaminase Deficiency: An Evidence Review Group Perspective of a NICE Highly Specialised Technology Evaluation. *Pharmacoeconomics - Open* **3**, 151–161 (2019).
 142. Kalos, M. *et al.* T Cells with Chimeric Antigen Receptors Have Potent Antitumor Effects and Can Establish Memory in Patients with Advanced Leukemia. *Sci. Transl. Med.* **3**, 95ra73–95ra73 (2011).
 143. Tisagenlecleucel (Kymriah) for ALL. *Med. Lett. Drugs Ther.* **59**, 177–178 (2017).
 144. Raper, S. E. *et al.* Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Mol. Genet. Metab.* **80**, 148–158 (2003).
 145. Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* **118**, 3132–3142 (2008).
 146. McCormack, M. P. & Rabbitts, T. H. Activation of the T-Cell Oncogene *LMO2* after Gene Therapy for X-Linked Severe Combined Immunodeficiency. *N. Engl. J. Med.* **350**, 913–922 (2004).
 147. Lundstrom, K. Viral Vectors in Gene Therapy. *Dis. (Basel, Switzerland)* **6**, (2018).
 148. Lee, C. S. *et al.* Adenovirus-mediated gene delivery: Potential applications for gene and cell-based therapies in the new era of personalized medicine. *Genes Dis.* **4**, 43–63 (2017).
 149. Schiedner, G. *et al.* Genomic DNA transfer with a high-capacity adenovirus vector results in improved in vivo gene expression and decreased toxicity. *Nat. Genet.* **18**, 180–183 (1998).
 150. Dormond, E., Perrier, M. & Kamen, A. From the first to the third generation adenoviral vector: What parameters are governing the production yield? *Biotechnol. Adv.* **27**, 133–144 (2009).
 151. Coffin, J. M., Hughes, S. H. & Varmus, H. *Retroviruses*. (Cold Spring Harbor Laboratory Press, 1997).

152. De Riick, J. *et al.* Bromodomain and extra-terminal (BET) proteins target Moloney murine leukemia virus integration to transcription start sites. *Retrovirology* **10**, O20 (2013).
153. Cavazza, A., Moiani, A. & Mavilio, F. Mechanisms of Retroviral Integration and Mutagenesis. *Hum. Gene Ther.* **24**, 119–131 (2013).
154. Cherepanov, P. *et al.* HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells. *J. Biol. Chem.* **278**, 372–381 (2003).
155. Busschots, K. *et al.* The Interaction of LEDGF/p75 with Integrase Is Lentivirus-specific and Promotes DNA Binding. *J. Biol. Chem.* **280**, 17841–17847 (2005).
156. Samulski, R. J. & Muzyczka, N. AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annu. Rev. Virol.* **1**, 427–451 (2014).
157. McClements, M. E. & MacLaren, R. E. Adeno-associated Virus (AAV) Dual Vector Strategies for Gene Therapy Encoding Large Transgenes. *Yale J. Biol. Med.* **90**, 611–623 (2017).
158. Mingozzi, F. & High, K. A. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood* **122**, 23–36 (2013).
159. Wilen, C. B. & Doms, R. W. HIV: Cell binding and entry. *Cold Spring Harb. Perspect. Med.* **2**, 1–14 (2012).
160. Akkina, R. K. & Chen, I. S. Y. High-efficiency gene transfer into CD34+ cells with a human immunodeficiency virus type 1-based retroviral vector pseudotyped with vesicular stomatitis virus envelope glycoprotein G. *J. Virol.* **70**, 2581–5 (1996).
161. Naldini, L. & Trono, D. In vivo gene delivery and stable transduction of post mitotic cells by a lentiviral vector. *Science (80-.).* **272**, 263–267 (1996).
162. Zufferey, R. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat Biotechnol* **15**, 871–875 (1997).
163. Kim, V. N. & Kingsman, A. J. Minimal requirement for a lentivirus vector based on human immunodeficiency virus type 1. *J. Virol.* **72**, 811–816 (1998).
164. Dull, T. & Naldini, L. A Third-Generation Lentivirus Vector with a Conditional Packaging System. *J. Virol.* **72**, 8463–8471 (1998).
165. Mann, R., Mulligan, R. C. & Baltimore, D. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* **33**, 153–159 (1983).

166. Page, K. A., Landau, N. R. & Littman, D. R. Construction and use of a human immunodeficiency virus vector for analysis of virus infectivity. *J. Virol.* **64**, 5270–6 (1990).
167. Mitchell, R. S. & Bushman, F. D. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2**, 1127–1137 (2004).
168. Zavada, J. The Pseudotypic Paradox. *J. Gen. Virol.* **63**, 15–24 (1982).
169. Canivet, M., Hoffman, A. D., Hardy, D., Sernatinger, J. & Levy, J. A. Replication of HIV-1 in a wide variety of animal cells following phenotypic mixing with murine retroviruses. *Virology* **178**, 543–51 (1990).
170. Chesebro, B., Wehrly, K. & Maury, W. Differential expression in human and mouse cells of human immunodeficiency virus pseudotyped by murine retroviruses. *J. Virol.* **64**, 4553–7 (1990).
171. Zhu, Z. H., Chen, S. S. & Huang, A. S. Phenotypic mixing between human immunodeficiency virus and vesicular stomatitis virus or herpes simplex virus. *J. Acquir. Immune Defic. Syndr.* **3**, 215–9 (1990).
172. Akkina, R. K. *et al.* High-efficiency gene transfer into CD34+ cells with a human immunodeficiency virus type 1-based retroviral vector pseudotyped with vesicular stomatitis virus envelope glycoprotein G. *J. Virol.* **70**, 2581–5 (1996).
173. Naldini, L. *et al.* In Vivo Gene Delivery and Stable Transduction of Nondividing Cells by a Lentiviral Vector. *Science (80-.)*. **272**, 263–267 (1996).
174. Reiser, J. *et al.* Transduction of nondividing cells using pseudotyped defective high-titer HIV type 1 particles. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 15266–71 (1996).
175. Schlegel, R., Tralka, T. S., Willingham, M. C. & Pastan, I. Inhibition of VSV binding and infectivity by phosphatidylserine: is phosphatidylserine a VSV-binding site? *Cell* **32**, 639–46 (1983).
176. Finkelshtein, D., Werman, A., Novick, D., Barak, S. & Rubinstein, M. LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. *Proc. Natl. Acad. Sci.* **110**, 7306–7311 (2013).
177. Miller, M. D., Farnet, C. M. & Bushman, F. D. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* **71**, 5382–90 (1997).
178. Escors, D. & Breckpot, K. Lentiviral Vectors in Gene Therapy: Their Current Status and

- Future Potential. *Arch. Immunol. Ther. Exp. (Warsz)*. **58**, 107–119 (2010).
179. Yamashita, M. & Emerman, M. Retroviral infection of non-dividing cells: Old and new perspectives. *Virology* **344**, 88–93 (2006).
 180. Burns, J. C., Friedmann, T., Driever, W., Burrascano, M. & Yee, J. K. Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc. Natl. Acad. Sci.* **90**, 8033–8037 (1993).
 181. Annoni, A., Gregori, S., Naldini, L. & Cantore, A. Modulation of immune responses in lentiviral vector-mediated gene transfer. *Cell. Immunol.* **342**, 103802 (2019).
 182. Modlich, U. & Baum, C. Leukemia induction after a single retroviral vector insertion in Evi1 or Prdm16. *Leukemia* **22**, 1519–28 (2008).
 183. Miyoshi, H. & Verma, I. M. Development of a self-inactivating lentivirus vector. *J. Virol.* **72**, 8150–8157 (1998).
 184. Aiuti, A. & Naldini, L. Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome. *Science (80-.)*. **341**, 1233151 (2013).
 185. Segura, M. M., Mangion, M., Gaillet, B. & Garnier, A. New developments in lentiviral vector design, production and purification. *Expert Opinion on Biological Therapy* **13**, 987–1011 (2013).
 186. Dropulic, B. & June, C. H. Gene-Based Immunotherapy for Human Immunodeficiency Virus Infection and Acquired Immunodeficiency Syndrome. *Hum. Gene Ther.* **17**, 577–588 (2006).
 187. BANK, A., DORAZIO, R. & LEBOULCH, P. A Phase I/II Clinical Trial of β -Globin Gene Therapy for β -Thalassemia. *Ann. N. Y. Acad. Sci.* **1054**, 308–316 (2005).
 188. Cartier, N. & Aubourg, P. Hematopoietic stem cell gene therapy in Hurler syndrome, globoid cell leukodystrophy, metachromatic leukodystrophy and X-adrenoleukodystrophy. *Curr. Opin. Mol. Ther.* **10**, 471–8 (2008).
 189. 777. ProSavin, a Gene Therapy for Parkinson's Disease. *Mol. Ther.* **18**, S302 (2010).
 190. Gene Therapy Clinical Trials Worldwide. Available at: <http://www.abedia.com/wiley/index.html>. (Accessed: 11th August 2019)
 191. Cavazzana-Calvo, M. & Fischer, A. Gene therapy for severe combined immunodeficiency: are we there yet? *J. Clin. Invest.* **117**, 1456–1465 (2007).
 192. Dupré, L. *et al.* Efficacy of Gene Therapy for Wiskott-Aldrich Syndrome Using a WAS

- Promoter/cDNA-Containing Lentiviral Vector and Nonlethal Irradiation. *Hum. Gene Ther.* **17**, 303–313 (2006).
193. Zufferey, R. *et al.* Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J. Virol.* **72**, 9873–80 (1998).
 194. Craigie, R. & Bushman, F. D. HIV DNA Integration. 1–18 (2012).
 195. Cavazzana-Calvo, M. *et al.* Transfusion independence and HMGA2 activation after gene therapy of human β -thalassaemia. *Nature* **467**, 318–22 (2010).
 196. Oh, T., Bajwa, A., Jia, G. & Park, F. Lentiviral vector design using alternative RNA export elements. *Retrovirology* **4**, 38 (2007).
 197. Zennou, V. & Charneau, P. HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101**, 173–185 (2000).
 198. Zennou, V. & Charneau, P. The HIV-1 DNA flap stimulates HIV vector-mediated cell transduction in the brain. *Nat. Biotechnol.* **19**, 446–50 (2001).
 199. Modlich, U. & Baum, C. Insertional Transformation of Hematopoietic Cells by Self-inactivating Lentiviral and Gammaretroviral Vectors. *Mol. Ther.* **17**, 1919–1928 (2009).
 200. Huang, Z. M. & Yen, T. S. Role of the hepatitis B virus posttranscriptional regulatory element in export of intronless transcripts. *Mol. Cell. Biol.* **15**, 3864–9 (1995).
 201. Zufferey, R., Donello, J. E., Trono, D. & Hope, T. J. Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J. Virol.* **73**, 2886–92 (1999).
 202. Zanta-boussif, M. A. & Hope, T. J. Validation of a mutated PRE sequence allowing high and sustained transgene expression while abrogating WHV-X protein synthesis : application to the gene therapy of WAS. *Gene Ther.* **16**, 605–619 (2009).
 203. Kingsman, S. M., Mitrophanous, K. & Olsen, J. C. Potential oncogene activity of the woodchuck hepatitis post-transcriptional regulatory element (WPRE). *Gene Ther.* **12**, 3–4 (2005).
 204. Hanawa, H., Yamamoto, M., Zhao, H., Shimada, T. & Persons, D. A. Optimized lentiviral vector design improves titer and transgene expression of vectors containing the chicken beta-globin locus HS4 insulator element. *Mol. Ther.* **17**, 667–74 (2009).
 205. Hanawa, H. *et al.* Extended β -globin locus control region elements promote consistent therapeutic expression of a β -globin lentiviral vector in murine β -thalassemia. *Blood*

- 104**, 2281–2290 (2004).
206. Antoniou, M. *et al.* Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. *Genomics* **82**, 269–79 (2003).
 207. Zhang, F. *et al.* Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells. *Blood* **110**, 1448–57 (2007).
 208. Vink, C. A. *et al.* Eliminating HIV-1 Packaging Sequences from Lentiviral Vector Proviruses Enhances Safety and Expedites Gene Transfer for Gene Therapy. *Mol. Ther.* **25**, 1790–1804 (2017).
 209. Schneider, R., Campbell, M., Nasioulas, G., Felber, B. K. & Pavlakis, G. N. Inactivation of the human immunodeficiency virus type 1 inhibitory elements allows Rev-independent expression of Gag and Gag/protease and particle formation. *J. Virol.* **71**, 4892–903 (1997).
 210. Kotsopoulou, E., Kim, V. N., Kingsman, A. J., Kingsman, S. M. & Mitrophanous, K. A. A Rev-independent human immunodeficiency virus type 1 (HIV-1)-based vector that exploits a codon-optimized HIV-1 gag-pol gene. *J. Virol.* **74**, 4839–52 (2000).
 211. Brule, C. E. & Grayhack, E. J. Synonymous Codons: Choose Wisely for Expression. *Trends Genet.* **33**, 283–297 (2017).
 212. Tareen, S. U. *et al.* A Rev-Independent gag/pol Eliminates Detectable psi-gag Recombination in Lentiviral Vectors. *Biores. Open Access* **2**, 421–30 (2013).
 213. Sanber, K. S. *et al.* Construction of stable packaging cell lines for clinical lentiviral vector production. *Sci. Rep.* **5**, 9021 (2015).
 214. Maunder, H. E. *et al.* Enhancing titres of therapeutic viral vectors using the transgene repression in vector production (TRiP) system. *Nat. Commun.* **8**, 14834 (2017).
 215. Biffi, A. *et al.* Lentiviral Hematopoietic Stem Cell Gene Therapy Benefits Metachromatic Leukodystrophy. *Science (80-.)*. **341**, 1233158–1233158 (2013).
 216. Aiuti, A. *et al.* Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome. *Science (80-.)*. **341**, 1233151–1233151 (2013).
 217. Griffin, D. O. & Goff, S. P. HIV-1 Is Restricted prior to Integration of Viral DNA in Primary Cord-Derived Human CD34+ Cells. *J. Virol.* **89**, 8096–100 (2015).
 218. Kolb-Maurer, A., Wilhelm, M., Weissinger, F., Bröcker, E.-B. & Goebel, W. Interaction

- of human hematopoietic stem cells with bacterial pathogens. *Blood* **100**, 3703–3709 (2002).
219. Bieniasz, P. D. Intrinsic immunity: a front-line defense against viral attack. *Nat. Immunol.* **5**, 1109–1115 (2004).
220. Towers, G. J. & Noursadeghi, M. Interactions between HIV-1 and the cell-autonomous innate immune system. *Cell Host Microbe* **16**, 10–8 (2014).
221. Schwefel, D. *et al.* Structural basis of lentiviral subversion of a cellular protein degradation pathway. *Nature* **505**, 234–238 (2014).
222. Andreadis, S. T., Roth, C. M., Le Doux, J. M., Morgan, J. R. & Yarmush, M. L. Large-Scale Processing of Recombinant Retroviruses for Gene Therapy. *Biotechnol. Prog.* **15**, 1–11 (1999).
223. Kajaste-Rudnitski, A. & Naldini, L. Cellular Innate Immunity and Restriction of Viral Infection: Implications for Lentiviral Gene Therapy in Human Hematopoietic Cells. *Hum. Gene Ther.* **26**, 201–209 (2015).
224. Slingsby, J. H. *et al.* Brief Report: Analysis of 4070A Envelope Levels in Retroviral Preparations and Effect on Target Cell Transduction Efficiency. *Hum. Gene Ther.* **11**, 1439–1451 (2000).
225. de las Mercedes Segura, M., Kamen, A. & Garnier, A. Downstream processing of oncoretroviral and lentiviral gene therapy vectors. *Biotechnol. Adv.* **24**, 321–337 (2006).
226. Le Doux, J. M., Morgan, J. R., Snow, R. G. & Yarmush, M. L. Proteoglycans secreted by packaging cell lines inhibit retrovirus infection. *J. Virol.* **70**, 6468–73 (1996).
227. Kamen, A. Recent progress in lentiviral vector mass production. *Biochem. Eng. J.*
228. Merten, O.-W. *et al.* Large-Scale Manufacture and Characterization of a Lentiviral Vector Produced for Clinical *Ex Vivo* Gene Therapy Application. *Hum. Gene Ther.* **22**, 343–356 (2011).
229. Cribbs, A. P., Kennedy, A., Gregory, B. & Brennan, F. M. Simplified production and concentration of lentiviral vectors to achieve high transduction in primary human T cells. *BMC Biotechnol.* **13**, 98 (2013).
230. Delville, M. *et al.* A Nontoxic Transduction Enhancer Enables Highly Efficient Lentiviral Transduction of Primary Murine T Cells and Hematopoietic Stem Cells. *Mol. Ther. Methods Clin. Dev.* **10**, 341–347 (2018).

231. Ozog, S. *et al.* Resveratrol trimer enhances gene delivery to hematopoietic stem cells by reducing antiviral restriction at endosomes. *Blood* **129**, 4000–4004 (2019). doi:10.1182/blood.2019000040
232. BD Accuri™ C6 Plus System | BD Biosciences-US. Available at: <https://www.bdbiosciences.com/us/instruments/research/cell-analyzers/bd-accuri/bd-accuri-c6-plus-system/c/1294935>. (Accessed: 8th November 2019)
233. QuickChange Primer Design. Available at: <https://www.agilent.com/store/primerDesignProgram.jsp>. (Accessed: 10th October 2019)
234. Lizée, G. *et al.* Real-Time Quantitative Reverse Transcriptase-Polymerase Chain Reaction as a Method for Determining Lentiviral Vector Titers and Measuring Transgene Expression. *Hum. Gene Ther.* **14**, 497–507 (2003).
235. Welcome to the Predict a Secondary Structure Web Server. Available at: <https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>. (Accessed: 11th October 2019)
236. XRNA. Available at: <http://rna.ucsc.edu/rnacenter/xrna/>. (Accessed: 11th October 2019)
237. ImageJ. Available at: <https://imagej.nih.gov/ij/index.html>. (Accessed: 11th October 2019)
238. Ella - Simple Plex :: ProteinSimple. Available at: <https://www.proteinsimple.com/ella.html>. (Accessed: 6th November 2019)
239. dsDNA copy number calculator. Available at: <http://cels.uri.edu/gsc/cndna.html>. (Accessed: 10th October 2019)
240. Addgene: Homepage. Available at: <https://www.addgene.org/>. (Accessed: 6th November 2019)
241. The Mfold Web Server | mfold.rit.albany.edu. Available at: <http://unafold.rna.albany.edu/?q=mfold>. (Accessed: 10th October 2019)
242. van Bel, N., Das, A. T., Cornelissen, M., Abbink, T. E. M. & Berkhout, B. A short sequence motif in the 5' leader of the HIV-1 genome modulates extended RNA dimer formation and virus replication. *J. Biol. Chem.* **289**, 35061–74 (2014).
243. Burniston, M. T., Cimarelli, A., Colgan, J., Curtis, S. P. & Luban, J. Human immunodeficiency virus type 1 Gag polyprotein multimerization requires the

- nucleocapsid domain and RNA and is promoted by the capsid-dimer interface and the basic region of matrix protein. *J. Virol.* **73**, 8527–40 (1999).
244. Bell, N. M. & Lever, A. M. L. HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol.* **21**, 136–144 (2013).
 245. Sinck, L. *et al.* In vitro dimerization of human immunodeficiency virus type 1 (HIV-1) spliced RNAs. *RNA* **13**, 2141–50 (2007).
 246. Scherr, M., Battmer, K., Blömer, U., Ganser, A. & Grez, M. Quantitative Determination of Lentiviral Vector Particle Numbers by Real-Time PCR. *Biotechniques* **31**, 520–526 (2001).
 247. Sastry, L., Johnson, T., Hobson, M. J., Smucker, B. & Cornetta, K. Titering lentiviral vectors: comparison of DNA, RNA and marker expression methods. *Gene Ther.* **9**, 1155–1162 (2002).
 248. Butler, S. L., Hansen, M. S. T. & Bushman, F. D. A quantitative assay for HIV DNA integration in vivo. *Nat. Med.* **7**, 631–634 (2001).
 249. Charrier, S. *et al.* A lentiviral vector encoding the human Wiskott–Aldrich syndrome protein corrects immune and cytoskeletal defects in WASP knockout mice. *Gene Ther.* **12**, 597–606 (2005).
 250. Van Maele, B., De Rijck, J., De Clercq, E. & Debyser, Z. Impact of the central polypurine tract on the kinetics of human immunodeficiency virus type 1 vector transduction. *J. Virol.* **77**, 4685–94 (2003).
 251. Duffy, S. & Cochrane, A. Analysis of HIV-1 RNA Splicing. in *Alternative pre-mRNA Splicing* 438–448 (Wiley-VCH Verlag GmbH & Co. KGaA, 2012).
doi:10.1002/9783527636778.ch41
 252. Rychlik, W. OLIGO 7 Primer Analysis Software. in *Methods in molecular biology (Clifton, N.J.)* **402**, 35–59 (2007).
 253. Moore, M. D. *et al.* Probing the HIV-1 Genomic RNA Trafficking Pathway and Dimerization by Genetic Recombination and Single Virion Analyses. *PLoS Pathog.* **5**, e1000627 (2009).
 254. Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–22 (2009).
 255. qPCR Efficiency Calculator - GR.
 256. Real-Time PCR: Understanding Ct - UK.

257. How TaqMan Assays Work - UK.
258. Essentials of Real-Time PCR - UK.
259. TaqMan™ MGB Probe. Available at:
<https://www.thermofisher.com/order/catalog/product/4316034#/4316034>.
 (Accessed: 11th October 2019)
260. Hansson, M. D., Rzeznicka, K., Rosenbäck, M., Hansson, M. & Sirijovski, N. PCR-mediated deletion of plasmid DNA. *Anal. Biochem.* **375**, 373–375 (2008).
261. Spiess, A.-N. & Ivell, R. A Highly Efficient Method for Long-Chain cDNA Synthesis Using Trehalose and Betaine. *Anal. Biochem.* **301**, 168–174 (2002).
262. Liu, Y. *et al.* HIV-1 Sequence Necessary and Sufficient to Package Non-viral RNAs into HIV-1 Particles. *J. Mol. Biol.* **429**, 2542–2555 (2017).
263. Kharytonchyk, S. *et al.* Influence of gag and RRE Sequences on HIV-1 RNA Packaging Signal Structure and Function. *J. Mol. Biol.* **430**, 2066–2079 (2018).
264. Buchschacher, G. L., Panganiban, A. T. & Panganiban, A. T. Human immunodeficiency virus vectors for inducible expression of foreign genes. *J. Virol.* **66**, 2731–9 (1992).
265. Deborah J. Kerwood, Michael J. Cavaluzzi, and & Borer*, P. N. Structure of SL4 RNA from the HIV-1 Packaging Signal†,‡. (2001). doi:10.1021/BI0111909
266. Amarasinghe, G. K. *et al.* Stem-loop SL4 of the HIV-1 Ψ RNA packaging signal exhibits weak affinity for the nucleocapsid protein. structural studies and implications for genome recognition. *J. Mol. Biol.* **314**, 961–970 (2001).
267. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 97–102 (2009).
268. Lu, K., Heng, X. & Summers, M. F. Structural Determinants and Mechanism of HIV-1 Genome Packaging. *J. Mol. Biol.* **410**, 609–633 (2011).
269. Stephenson, J. D. *et al.* Three-dimensional RNA structure of the major HIV-1 packaging signal region. *Structure* **21**, 951–62 (2013).
270. HIV Sequence Database: 2018 Compendium. Available at:
<https://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2018compendium.html>.
 (Accessed: 30th October 2019)
271. Reyes-Darias, J. A., Sánchez-Luque, F. J. & Berzal-Herranz, A. HIV RNA dimerisation interference by antisense oligonucleotides targeted to the 5' UTR structural elements. *Virus Res.* **169**, 63–71 (2012).

272. Ziv, O. *et al.* COMRADES determines in vivo RNA structures and interactions. *Nat. Methods* **15**, 785–788 (2018).
273. Quan, P.-L., Sauzade, M. & Brouzes, E. dPCR: A Technology Review. *Sensors (Basel)*. **18**, (2018).
274. Cooper, T. A. *In vivo* SELEX Strategies. in *Handbook of RNA Biochemistry* 1207–1220 (Wiley-VCH Verlag GmbH & Co. KGaA, 2014). doi:10.1002/9783527647064.ch54
275. van Bel, N., Das, A. T. & Berkhout, B. In Vivo SELEX of Single-Stranded Domains in the HIV-1 Leader RNA. *J. Virol.* **88**, 1870–1880 (2014).
276. Geller, R. *et al.* The external domains of the HIV-1 envelope are a mutational cold spot. *Nat. Commun.* **6**, 8571 (2015).
277. HIV Databases. Available at: <https://www.hiv.lanl.gov/content/index>. (Accessed: 11th October 2019)
278. Elliott, D. & Lodomery, M. *Molecular biology of RNA*.
279. Maldarelli, F., Martin, M. A. & Strebel, K. Identification of posttranscriptionally active inhibitory sequences in human immunodeficiency virus type 1 RNA: novel level of gene regulation. *J. Virol.* **65**, 5732–43 (1991).
280. Nasioulas, G. *et al.* Elements distinct from human immunodeficiency virus type 1 splice sites are responsible for the Rev dependence of env mRNA. *J. Virol.* **68**, 2986–93 (1994).
281. Palusa, S., Ndaluka, C., Bowen, R. A., Wilusz, C. J. & Wilusz, J. The 3' Untranslated Region of the Rabies Virus Glycoprotein mRNA Specifically Interacts with Cellular PCBP2 Protein and Promotes Transcript Stability. *PLoS One* **7**, e33561 (2012).
282. Zhang, Y.-J., Pan, H.-Y. & Gao, S.-J. Reverse Transcription Slippage over the mRNA Secondary Structure of the LIP1 Gene. *Biotechniques* **31**, 1286–1294 (2001).
283. Wang, M. Q., Huang, Y. L., Huang, J., Zheng, J. L. & Qian, G. X. RIG-I detects HIV-1 infection and mediates type I interferon response in human macrophages from patients with HIV-1-associated neurocognitive disorders. *Genet. Mol. Res.* **14**, 13799–811 (2015).
284. Eberle, J. & Seibl, R. A new method for measuring reverse transcriptase activity by ELISA. *J. Virol. Methods* **40**, 347–56 (1992).
285. Devarkar, S. C. *et al.* Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proc. Natl. Acad.*

Sci. **113**, 596–601 (2016).

286. Hellmund, C. & Lever, A. M. L. Coordination of Genomic RNA Packaging with Viral Assembly in HIV-1. 1–13 (2016). doi:10.3390/v8070192