



HAL
open science

Physical properties of R-Loops and viral capsids: a single molecule approach based on AFM high resolution imaging and nano-indentation

Yeraldinne Dennise Carrasco Salas

► To cite this version:

Yeraldinne Dennise Carrasco Salas. Physical properties of R-Loops and viral capsids: a single molecule approach based on AFM high resolution imaging and nano-indentation. Biological Physics [physics.bio-ph]. Université de Lyon, 2021. English. NNT: 2021LYSEN009 . tel-03280739

HAL Id: tel-03280739

<https://theses.hal.science/tel-03280739>

Submitted on 7 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Numéro National de Thèse : 2021LYSEN009

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée par

l'Ecole Normale Supérieure de Lyon

Ecole Doctorale N° 52

Physique et Astrophysique de Lyon

Spécialité de doctorat : Biophysique

Discipline : Physique

Soutenue publiquement le 06/04/2021, par :

Yeraldinne Dennise CARRASCO SALAS

Physical properties of R-loops and viral capsids: a single molecule approach based on AFM high resolution imaging and nano-indentation.

Propriétés physiques de R-loops et de capsides virales : une approche par molécules uniques basée sur la Microscopie à Force Atomique (AFM) par imagerie à haute résolution et nano-indentation.

Devant le jury composé de :

TARDIN Catherine, Maître de Conférences, Université Paul Sabatier, Rapporteur

CHARITAT Thierry, Professeur des Universités, Université de Strasbourg, Rapporteur

PICAS Laura, Chargée de Recherche, CNRS, Examinatrice

DE PABLO Pedro J., Professeur, Université Autonome de Madrid (Espagne), Examineur

VAILLANT Cédric, Directeur de Recherche, CNRS, Examineur

MOSKALENKO Cendrine, Maître de Conférences, ENS de Lyon, Directrice de thèse

Remerciements

Je remercie les rapporteurs de ma thèse, Catherine Tardin and et Thierry Charitat d'avoir pris le temps de lire et de corriger ce manuscrit. I also want to thank Laura Picas, Pedro J. de Pablo and Cédric Vaillant for agreeing to participate on this jury.

Je tiens à remercier chaleureusement Cendrine Moskalenko. Nous avons clairement eu des hauts et des bas au cours de ces quatre années, même si je n'en regrette aucun. Je le referais et bien sûr je recommande cette expérience. Je pense qu'avec toi j'ai eu l'occasion d'apprendre beaucoup de choses, imaginez que je ne me souvenais même pas des processus biologiques de l'ADN au début, j'ai dû repartir de zéro et tu m'as aidé à tout résoudre. Merci beaucoup!

Quiero agradecer a mi familia, en especial a mi compañero de vida, Cristóbal Oliver. No tengo palabras para describir lo agradecida que estoy y lo afortunada soy ya que pude conocerte. Gracias por todo el apoyo incondicional que me has brindado en mis peores momentos. La otra persona más importante para mí es mi hermana, Germina Carrasco, pienso que tampoco existen palabras que puedan describir el amor que se tienen dos hermanas, incluso si estamos separadas. Por supuesto, está mi abuela, Ana Velasquez, la persona que nos crió y nos dio todo lo que necesitamos para surgir como personas. Ustedes tres son las personas mas importantes de mi vida.

Aunque las personas que mencionaré estuvieron lejos durante mi doctorado, ya que ellos vivieron en Chile o simplemente en otra ciudad, quiero darles un especial agradecimiento. En particular, quiero dar un pequeño saludo a la señorita Carolina Robles, la mujer que ha experimentado de todo y espero poder estar ahí para ti cuando sigas experimentando cosas nuevas. Están mis amigos de Valparaíso con los cuales compartí momentos que son importantes para mi, entre ellos están: Veronica Gaete, Alexandra Ovalle, Leandro Nuñez, Marco Sepulveda, Eduardo Peters, Menka Stojanova y Felipe Olivares, Gabriel Ramos and Céliane Carmona.

Quiero dar un especial agradecimiento a los amigos que encontré en Lyon. Comenzando por los chilenos, Valentina Villalobos e Ignacio Andrade, fueron las primeras personas que conocí y nos acogieron sin siquiera conocernos. Quien pensaría que al otro lado del mundo puedes encontrar algún familiar, bueno eso le pasó a Valentina, su primo Matias Pinto y su esposo Clément Seité. Puedo decir que he compartido muy buenas experiencias con todos ellos, donde cada uno me ha entregado una parte de

su historia. Por supuesto, está Raphael Poryles, el francés con acento chileno indescriptible, el cual me ayudó a adaptarme en el laboratorio y con el cual compartimos varias cervezas tanto en Francia como en Chile. Gracias a esas cervezas, tuve la oportunidad de conocer a una gran chica, Angélique Richard. Debo decir que estoy super agradecida de haberte conocido y espero poder continuar con esta linda amistad. Más tarde, llegaron dos doctorandos muy especiales para mí, Alex Fontana y Lucas Menou. Lucas est une personne incroyable, avec qui nous avons travaillé ensemble pendant mon doctorat, je peux dire que peu importe quand et où, je pourrais toujours compter sur vous, merci ! I still remember when we shared an office with Alex, thanks to that, I was able to meet other incredible people: Alex Amato, Richard Pedurand and Edna Maria de la Laguna (known as Paola), Chiara Moretti and Clara Tosi. Thank you guys, you gave me beautiful moments!

By random situation in life, I met David (Chih-Wei Liu), the Englishman who loves hugs the most... we had fun times dancing. Later, we met Aneta Dvořáková ... who would think that you would become the most beautifully odd couple. Aneta enjoying the party until the last moment and you, David, the first person leaving the party. You both brought out the best in each other, I am happy for you.

Por supuesto, luego vino el chileno más atlético de todos, Marcelo Guzmán. Gracias por enseñarme muchas cosas y por compartir algunos momentos de tu vida. A pesar de que tuvieron todo tipo de emociones en nuestro viaje en bicicleta, desearía poder repertilo en algun momento.

Bien sûr, plus tard, j'ai rencontré d'autres étudiants qui m'ont bien accueilli, parmi lesquels se trouvent: Hadi Kabalane, Pauline Hussein, Géraldine Davis, Laura Lemahieu, Barbara Pascal, Raphaël Menu, Lavi Upreti, Jérôme Thibaut, Salvish Goomanee, Simon Vincent, Facundo Cabrera, Léa Chazot-Franguiadakis, Florencia Falkinhoff, Thomas Boquet, Bastien Molcrette, Lucas Méthivier and Eric Brillaux.

Résumé

En combinant l'imagerie AFM à haute résolution et les mesures mécaniques à l'échelle nanométrique nous nous sommes intéressés aux propriétés physiques de deux objets biologiques différents.

Les R-Loops sont des structures hybrides ADN/ARN résultant d'une transcription où l'ARN s'hybride avec sa matrice d'ADN, tandis que le brin d'ADN non matrice reste simple brin. Ces structures R-Loops à trois brins ont été associées à des phénotypes d'instabilité génomique *in vivo*. Nous avons visualisé ces structures générées *in vitro* caractérisé leur conformation. En s'intéressant à un certain nombre de paramètres physiques quantifiables et en comparant ses données pour plusieurs gènes générant des hybrides, nous avons montré que ces R-Loops diffèrent par leur architecture et que l'organisation du brin non modèle est une caractéristique fondamentale pour conduire à la formation d'un objet compact, qui pourrait alors générer des cassures d'ADN dépendantes de la réplication.

En combinant l'imagerie de ces capsides en milieu liquide et la mesure de leur réponse mécanique grâce à la nano-indentation AFM, nous avons étudié les propriétés mécaniques des vecteurs AAV (Adeno Associated Virus) et les capsides du virus de l'Hépatite B, en lien avec leur désassemblage. En comparant les données obtenues le long d'un profil spatial enregistré sur différentes capsides, nous avons mis en évidence des hétérogénéités spatiales dont l'origine est pour partie géométrique et résulte également d'une hétérogénéité intrinsèque liée à l'organisation de la capside. Également, nous avons pu comparer les propriétés mécaniques des capsides et leur capacité à libérer leur génome *in vitro* par décapsidation induite par la température. Enfin, nous avons étudié la structure des répétitions terminales inversées (ITR) de l'AAV. Ces ITR adoptent une structure fonctionnelle que nous avons caractérisée par l'imagerie AFM.

Abstract

By combining high-resolution AFM imaging and nano-mechanical measurements at the single-molecule level, we studied the physical properties of two different biological objects.

R-loops are DNA/RNA hybrid structures arising from the DNA transcription process. The RNA hybridises with its DNA template by Watson-Crick interactions while the non-template DNA strand remains single-stranded. The R-loops structures have been associated with genomic instability phenotypes *in vivo*. We characterised by AFM imaging the conformation of these structures generated *in vitro*. We showed that these R-loops, generated for different genes, differ in their architecture by analysing quantitative physical parameters. We concluded that the non-template DNA strand organisation is a fundamental characteristic to form a compact hybrid object, which could then generate replication-dependent DNA breaks.

We studied the physical properties of two viral capsids, the AAV vector and the Hepatitis B capsid, connected with their disassembly. By combining the imaging and mechanical response measurement in a liquid medium using AFM, we obtained quantitative information on their mechanical properties. An indentation profile over the capsid shows spatial heterogeneities, which origin is partly geometric and results from an intrinsic heterogeneity linked to the capsid organisation. We also compared the capsid mechanical properties to the ability to release their genome *in vitro* when induced by temperature. Finally, we studied the Inverted Terminal Repeats (ITRs) structure of AAV. These ITRs adopt a functional structure that we characterised by AFM imaging.

Contents

1	Introduction: An experimental study by AFM	19
1.1	Atomic Force Microscope	19
1.2	General principle	20
1.3	AFM modes	23
1.4	Imaging samples	24
1.5	Cantilever	25
1.6	Nanoindentation by AFM	28
I	Physical Characterisation of R-loops	31
2	Introduction: R-loops	32
	Summary	32
2.1	The biology of the DNA	33
	2.1.1 DNA structure	33
	2.1.2 DNA functions	38
2.2	R-loop	41
2.3	Factors that promote the R-loop formation	42
2.4	Physiological and pathological R-loops	44
2.5	DNA studies using AFM	45
	2.5.1 Visualisation of the DNA	46
	2.5.2 Analysing G-quadruplexes structure in DNA	47
	2.5.3 Single-Stranded DNA Binding Proteins	48
	2.5.4 Study the biological process in real time	49
2.6	Mechanical properties of DNA	50
2.7	Our work in this context	52
3	AFM image analysis to characterise R-loop formation in DNA	54
	Summary	54
3.1	R-loops purification	55
3.2	AFM deposition and imaging	56

3.3	AFM image analysis	57
3.3.1	Algorithm used to extract R-loops properties	57
3.3.2	Isolating DNA molecules	58
3.3.3	Selecting DNA molecules containing R-loop objects	59
3.3.4	DNA molecule skeletonization	61
3.3.5	Obtention of the R-loops properties	63
3.4	Conclusions	65
4	Quantitative study of R-loops formation in DNA using AFM	66
	Summary	66
4.1	<i>Airn</i> gene	67
4.1.1	R-loops visualisation	67
4.1.2	R-loops quantification	68
4.1.3	R-loops object morphology	70
4.1.4	R-loops object position on the DNA template	71
4.1.5	R-loop objects generate local bends in the DNA	72
4.2	R-loop do not result from G-quadruplex	76
4.3	The mutation affects objects but not hybrid RNA-DNA formation	79
4.4	New R-loop object position	79
4.5	<i>Sum3</i> gene	82
4.6	Conclusions	84
II	Physical properties of viral capsid	85
5	Introduction: Physical virology focused on nanoindentation experiments	86
	Summary	87
5.1	Viruses	87
5.1.1	Virus classification	88
5.1.2	Virus life cycle	91
5.2	The architecture of the viral capsid	93
5.2.1	Icosahedral Symmetry	94
5.2.2	Buckling transition	96
5.3	Mechanical stiffness of virus capsids using nanoindentation experiments	97
5.4	Models to describe the elastic response of viral capsid	100
5.5	Results in mechanical properties of viruses	101
5.5.1	Stiffness depending on the rotational symmetry axes	101
5.5.2	Genome increased the capsid stiffness	103
5.5.3	Stiffness depending on the protrusion on the capsid surface	104

5.5.4	Stability of viral capsid through capsid disassembly	105
5.5.5	Stiffness value for the symmetry axis	105
5.6	Adeno-associated virus AAV for gene therapy	109
5.6.1	Viral cycle of AAV	109
5.6.2	Serotypes	110
5.6.3	AAV genome	112
5.6.4	Inverted Terminal Repeat (ITR)	113
5.6.5	AAV capsid	116
5.6.6	AAV stability and mechanical properties	117
5.7	Hepatitis B virus HBV	119
5.7.1	Viral cycle of HBV	121
5.7.2	HBV genome	122
5.7.3	Structure of HBV capsid	124
5.7.4	HBV mechanical properties	126
5.8	Our work in this context	127
6	AFM protocol to determine the viral capsid stiffness	129
	Summary	129
6.1	Capsid purification	130
6.2	Capsid deposition in mica	131
6.3	AFM imaging	134
6.4	Image analysis in a population of capsids	135
6.5	AFM image analysis for a single capsid	139
6.6	Methods to determine the capsid stiffness from z - F curves	144
6.6.1	Method 2 spring in series	144
6.6.2	$F(\delta)$ method	145
6.7	Nanoindentation on the surface	149
6.7.1	Force curve slope on a rigid surface gives us the cantilever stiffness	150
6.7.2	Indentations on the surface using a higher indentation force . .	150
6.8	Nanoindentation on the capsid	152
6.8.1	One indentation point on the capsid centre	152
6.8.2	Indentation spatial profile over the capsid	154
6.9	Non-linear effects on force curves	157
6.10	Capsid average profile	161
6.11	Conclusion	161
7	Stiffness of icosahedral capsids	163
	Summary	163
7.1	AAV8 capsid	164
7.1.1	Experimental Stiffness profile	164

7.1.2	Stiffness profiles according to the values of central points	166
7.1.3	Simulations of indentation numerical on viral capsid by Lucas Menou and Martin Castelnovo	168
7.1.4	Cantilever spring constant effect	173
7.1.5	Cantilever tip shape effect	176
7.2	AAV2 and AAV9 capsids	179
7.3	HBV capsid	182
7.3.1	Morphology of HBV capsid	182
7.3.2	Stiffness profile according to the values of central points	183
7.3.3	Influence of capsid facet in the stiffness profile	186
7.4	Conclusions	188
8	Thermal stability of AAV and HBV capsids	190
	Summary	190
8.1	Deposition conditions and AFM imaging	191
8.2	Genome uncoating for HBV capsid	192
8.3	Genome uncoating of AAV capsids	194
8.4	Conclusions	196
9	ITR-DNA visualisation on AFM images	197
	Summary	197
9.1	Model DNA containing ITR	198
9.1.1	Plasmid purification	198
9.1.2	Deposition protocol and AFM imaging	199
9.1.3	ITR-DNA visualisation	200
9.1.4	Determination of ITR positions in the DNA molecule	201
9.2	ITR from rAAV DNA	204
9.2.1	Genome extraction of rAAV	205
9.2.2	ITR-DNA visualisation to optimise the purification protocol	206
9.2.3	ITR from ssAAV and scAAV DNA	207
9.2.4	Capsid stability: ITR linked to capsid	209
9.2.5	Conformations of the ITRs	212
9.3	Conclusions	213

List of Figures

1.1	The Atomic Force Microscopy AFM	21
1.2	Tip surface interaction	22
1.3	Tip sample system	23
1.4	AFM reveals topography of biological samples	25
1.5	Cantilever tip	26
1.6	Cantilever shapes	27
1.7	Convolution effect between the cantilever tip and the sample	28
1.8	Nanoindentation experiment on a viral particle	29
1.9	Two ways of representing the same z - F curve data	30
2.1	Structure of the DNA	34
2.2	DNA configurations	35
2.3	Structure of the DNA-DNA, DNA-RNA and RNA-RNA duplex	36
2.4	Non-B DNA structures	37
2.5	Replication of the DNA	38
2.6	Transcription of the DNA	40
2.7	Maturation process of the messenger RNA	41
2.8	Translation of the DNA	42
2.9	Scheme of factors promoting R-loop formation	43
2.10	Avoiding R-loop formation	44
2.11	R-loop as a factor physiological and pathological	45
2.12	High-resolution AFM of dsDNA molecules	46
2.13	DNA plasmidd deposited on mica for differents divalent cations	47
2.14	Characterisation of G-quadruplex structures by AFM imaging	48
2.15	AFM approach for studying protein-DNA interactions	49
2.16	AFM experiments to study the dynamics of the nucleosome particle	50
2.17	Models to determine polymer chains persistence length	53
3.1	Maps of the plasmids	55
3.2	In vitro transcription of R-loops	56
3.3	Sample protocol for the AFM	57

3.4	Flattened image using NanoScope Ana lysis program	58
3.5	Treatment to extract the DNA molecules in Matlab	59
3.6	Selection DNA molecules with R-loops formation	60
3.7	Treatment to obtain the skeletonisation of the DNA molecules in Matlab	61
3.8	Representation geodesic distance in a matrix	62
3.9	DNA molecules types	63
3.10	Properties obtained from the DNA that contains R-loop objects	64
4.1	<i>Airn</i> : AFM images	68
4.2	pUC57- <i>airn</i> : objects quantification and skeleton length for all molecules	69
4.3	pUC57- <i>airn</i> : R-loops objects types	70
4.4	pUC57- <i>airn</i> : correlation between R-loop objects and skeleton length	71
4.5	pUC57- <i>airn</i> : correlation between the R-loop objects serotypes and skeleton length	72
4.6	pUC57- <i>airn</i> : position of the R-loops objects	73
4.7	pUC57- <i>airn</i> : R-loop formation imposes mechanical constraints on the surrounding DNA	74
4.8	pUC57- <i>airn</i> : DNA persistence length without R-loop object	75
4.9	pUC57- <i>airn</i> : R-loop do not result from G-quadruplex	77
4.10	pFC53- <i>airn</i> : R-loop do not result from G-quadruplex	78
4.11	pUC57- <i>airn</i> : The mutation affects objects but not hybrid RNA-DNA formation	80
4.12	pFC53-small- <i>airn</i> : new R-loop position	81
4.13	pUC57- <i>sum3</i> : the extruded ssDNA does not form detectable R-loop objects	83
4.14	pUC57- <i>sum3</i> : DNA persistence length without R-loop object	83
5.1	Classification of viruses	89
5.2	Baltimore classification of viruses according to the genome	90
5.3	Life cycle of viruses	92
5.4	Mechanisms of viral nuclear entry	93
5.5	Gallery of icosahedral viruses	94
5.6	Orientations of icosahedral capsid	95
5.7	Scheme of icosahedral capsid with different triangulation number T	96
5.8	Buckling transition and the Föppl–von Kármán (FvK) number	97
5.9	Effect of faceting on some viral capsids	98
5.10	Scheme of a nanoindentation experiment and its z - F curves of MVM capsid	99
5.11	Scheme of a shell sphere	100
5.12	Stiffness of the MVM empty and full capsid	102

5.13	Genome length effect on phage capsid	104
5.14	Protrusion effect on mechanical properties of Nonovirus	105
5.15	Capsid disassembly into pentamers for Triatome virus capsid	106
5.16	Stiffness of bacteriophage T7 capsid	107
5.17	Local lateral stresses of some T-number	108
5.18	AAV viral cycle	110
5.19	Properties of AAV serotypes	111
5.20	Secuence homology for AAV serotypes	112
5.21	AAV2 genome organization	113
5.22	Structure of the ITR	114
5.23	Diagrams of ssAAV and scAAV viral DNA structures.	114
5.24	AAV2 genome organization	115
5.25	Cryo-EM reconstruction of AAV2 capsid	116
5.26	Comparison of AAV2, AAV8 and AAV9 structures	117
5.27	Thermal stability of the AAV capsids	118
5.28	Mechanical properties of AAV2 associated with 2-fold symmetry axis	119
5.29	AAV2 capsid stability as a function of genome length and self-complementarity	120
5.30	The replication cycle of hepatitis B virus	121
5.31	Schematic representation of the structure of HBV genomic DNA, RNAs, and proteins	123
5.32	Schematic representation of HBV particles	125
5.33	Assembly of HBV capsid	125
5.34	Cryo-electron micrographs of HBV capsids	126
5.35	AFM nanoindentation experiments for HBV capsid	127
6.1	Diagram of AAV production process <i>in vitro</i>	130
6.2	Dried sample deposition conditions	132
6.3	Nondried sample deposition conditions	134
6.4	Flattened image using NanoScope Analysis program	135
6.5	Image analysis steps	136
6.6	Capsid height and area Probability density for AAV8 capsid	137
6.7	Capsid height for AAV2, AAV8 and AAV9	138
6.8	Zeng model: capsid reduced height measured in the AFM is due by the deformation and adhesion energy	139
6.9	NanoScope software screenshot	140
6.10	Capsid centre from AFM images	141
6.11	Capsid apparent diameter determination	142
6.12	Apparent diameter for AAV2, AAV8 and AAV9	143
6.13	Scheme of the typical z - F curves when a capsid is deformed	144

6.14	Scheme showing two $z-F$ curves	145
6.15	Scheme showing two $z-F$ curves 2	147
6.16	Indentation curves interpolation	148
6.17	Surface-Mica Analysis: contaminated layer	149
6.18	Surface-Mica Analysis: capsid solution generates the contamination	150
6.19	Surface-Mica Analysis: $z-F$ curves	151
6.20	Single capsid indentation curves for different positions using can_1	153
6.21	Single capsid indentation curves for different positions using can_2	153
6.22	Indentation positions scheme on the capsid	154
6.23	The behaviour of the $z-F$ curves depends on the indentation position	155
6.24	Separating $z-F$ curves according to indentation position	156
6.25	$z-F$ curves on the AAV8 capsid top	158
6.26	$z-F$ curves on the AAV2 capsid top	159
6.27	$z-F$ curves on the AAV9 capsid top	160
6.28	Effective stiffness profiles alignment	162
6.29	Effective stiffness average profile	162
7.1	Stiffness profiles for AAV8 using the best reference curve	164
7.2	Stiffness profiles for AAV8 using created curve with slope value obtained from thermal noise	165
7.3	AAV8: three central points of the indentation profile	166
7.4	AAV8: stiffness profile according to the value of the central positions	167
7.5	Scheme of effective indentation for different positions and cantilevers	169
7.6	Numerical indentation of a capsid with triangular number $T=1$	172
7.7	$z-F$ curves on the AAV8 capsid top - softer cantilever-	174
7.8	AAV8: three central points of the indentation profile using can_2	175
7.9	AAV8: stiffness profile according to the value of the central positions using can_2	176
7.10	AAV8: three central points of the indentation profile using can_2	177
7.11	AAV8: stiffness profile according to the value of the central positions using can_{1a}	178
7.12	Stiffness profiles for AAV2, AAV8 and AAV9 capsids.	179
7.13	Stiffness profiles for AAV2, AAV8 and AAV9 capsids.	181
7.14	HBV capsid height	183
7.15	$z-F$ curves on the HBV capsid top	184
7.16	HBV: stiffness profile and the centre points of the profile	185
7.17	HBV: stiffness profile according to the value of the central positions	186
7.18	Comparing AAV and HBV capsids profiles	187
8.1	Stability Protocol	191

8.2	AFM imaging of heat-induced HBV capsids destabilisation 1	192
8.3	AFM imaging of heat-induced HBV capsids destabilisation 2	193
8.4	Comparison of thermal stability between AAV2, AAV8 and AAV9	194
8.5	Correlation between capsid maximum height and the equivalent diameter as a function of the temperature	195

General Introduction

In terms of experimental tools and microscopy, the technological advance of the last decades allowed the development of instruments capable of visualising and measuring materials mechanical properties on a nanometric scale. This point marked a new set of possibilities in understanding biological structures such as DNA, nucleo-protein complexes, viruses and cells, which can now be studied at the single-molecule level. This opens the possibility of describing and understanding biological processes through the intrinsic mechanical properties of the structures. In this work, we study two biological structures/processes from a physical point of view, using an atomic force microscope AFM.

In the first part of this thesis, we use AFM high-resolution imaging to study R-loop formations at the AIRN gene locus *in vitro*. We demonstrate that they adopt a defined set of three-dimensional conformations, characterised by distinct shapes and volumes, which we call R-loop objects. Interestingly, we will show that these R-loop objects impose specific physical constraints on the DNA, as revealed by the presence of stereotypical angles in the surrounding DNA.

In the second part, using an original combination of experimental approaches based on AFM imaging and nano-mechanical measurements, we study the physical properties of Adeno-Associated Virus (AAV) and Hepatitis B Virus (HBV). To define the parameters leading to capsid disassembly and ejection of the viral genome, AFM is used to probe the morphology of viral capsids with a very high spatial resolution and exert mechanical constraints on viral particles, this measuring their elastic response. We show that provided a careful analysis protocol, it is possible to extract the average stiffness of our viral capsids for the two small viruses AAV and HBV and compare their intrinsic stiffness heterogeneity associated with the icosahedral symmetry. In addition to this, we compare several AAV serotypes in thermodynamic equilibrium with the environment and relate their differences in terms of mechanical properties with their ability to uncoat their genome when destabilised by a temperature increase. Finally, we characterised the morphology of the Inverted Terminal Repeat (ITR). The main

goal is to establish a relationship between the ITR conformation and its efficiency in gene therapy by engineering the ITR of recombinant AAV vector and its biological role in AAV vector gene transfer.

Since the biological questions and objects we studied are different, the manuscript is divided into two parts related to one of these topics. For each topic, we first present an introduction of the biological object and state of the art. Then, we show the experimental approach we have developed. The results we obtained and their discussion we finally presented in the last section.

To start, we will begin with a brief introduction of the AFM and its experimental values. The results are then presented in two chapters, each beginning with a description of the methods and protocols used to perform the measurements and the data analysis.

Chapter 1

Introduction: An experimental study by AFM

Index

1.1 Atomic Force Microscope	19
1.2 General principle	20
1.3 AFM modes	23
1.4 Imaging samples	24
1.5 Cantilever	25
1.6 Nanoindentation by AFM	28

1.1 Atomic Force Microscope

The Atomic Force Microscope AFM is a combination of the Scanning Tunnelling Microscope STM principles, and the Stylus Profilometer SP and it was invented by Binnig, Quate and Gerber in 1986 [1]. The STM allows the scanning of the surface with a very sharp tip. As the tip approaches the surface, a voltage is applied to the tip to obtain surface images at atomic resolution. A condition of these instruments is that both the tip and sample should be conductive [2]. The SP is an instrument that analyses the surface texture of a sample measuring the surface topography of a sample [3]. Then, combining both methods, Binnig *et al.* opened the possibility of studying non-conductive systems, biological systems being a fundamental example [4].

In the last decade, the AFM is an instrument that has become an essential tool

for studying biological systems under physiologically relevant conditions, for example, in liquid. The AFM allows to map surfaces with a nanometric resolution and obtain mechanical properties such as the elasticity and stiffness of these surfaces. Currently, some AFMs allow studying dynamic processes in real-time, resulting in a greater possibility of studying biological processes. The great advantage of this instrument is that it is used to study biological systems that have a large variety of sizes, from DNA or RNA molecules, macromolecules such as proteins, or even larger system, such as viruses, bacteria or cells.

Next, we will explain the general principle of the AFM and the components involved in its operation.

1.2 General principle

Similarly to the Scanning Tunnel Microscope STM, The AFM is constituted by (i) the head containing the beam deflection system and (ii) the piezoelectric device that moves the sample in all three directions (x,y,z). Last, (iii) the base contains the electronic control and the vibration isolation system [Fig. 1.1(a)].

The forces between the cantilever tip and the sample surface generate deflection of the cantilever. Depending on the distance between the cantilever and the surface, the main contributions come from the Van der Waals force, coulombic interactions or electrostatic. A simple optical beam deflection setup reads the deflection of the cantilever. The collimated laser beam is focused on the cantilever, which is reflected in a mirror. This mirror redirects the laser beam to the photodetector [Fig. 1.1(b)]. To determine the laser point position on the photodetector, the sensor is divided into four quadrants [Fig. 1.1(c)]. Thus, the horizontal and vertical position can determine according to the following expression:

$$Hor = \frac{C - D}{C + D}, \quad (1.1)$$

and,

$$Vert = \frac{A - B}{A + B}. \quad (1.2)$$

The total intensity reflected SUM is the addition of the power retrieved by each quadrant. The SUM normalization of the Hor and Vert parameters prevents possible

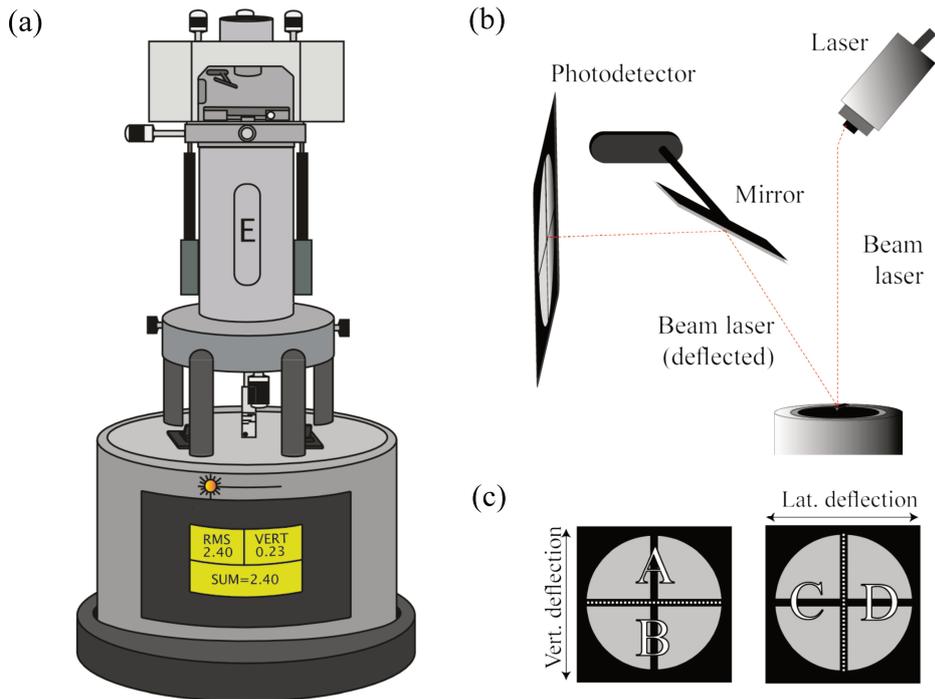


Figure 1.1: **The Atomic Force Microscopy AFM** - (a) Scheme of the AFM (MultiMode 8 from Bruker). (b) Optic system contains in the AFM head. (c) Scheme of the photodetector.

fluctuations in the initial laser intensity. During manipulation, this could be interpreted by the system as cantilever deflections. Figure 1.1(c) shows a schematic of the vertical deflection according to the position of the laser on the photodetector.

In response to the applied voltage, due to the cantilever deflections measured by the photodetector, the scanner (piezoelectric device) can move the sample depending on the sample topography. As a result of this, the closed-loop feedback can maintain the cantilever deflection constant, and the resulting three-dimensional movement tracks the sample surface, allowing the construction of a topographic image.

The sum of three potentials can describe the cantilever and sample system [5]: (i) the interaction potential between the tip and the sample $U_{cs}(D)$, where D is the tip-sample distance. (ii) The Hooke's elastic potential of the cantilever $U_c(\delta_c)$ and (iii) the potential that describes the sample deformation $U_s(\delta_s)$. The potentials that describe the deformation of the sample are defined as

$$U_c(\delta_c) = \frac{1}{2}k_c\delta_c^2 \quad (1.3)$$

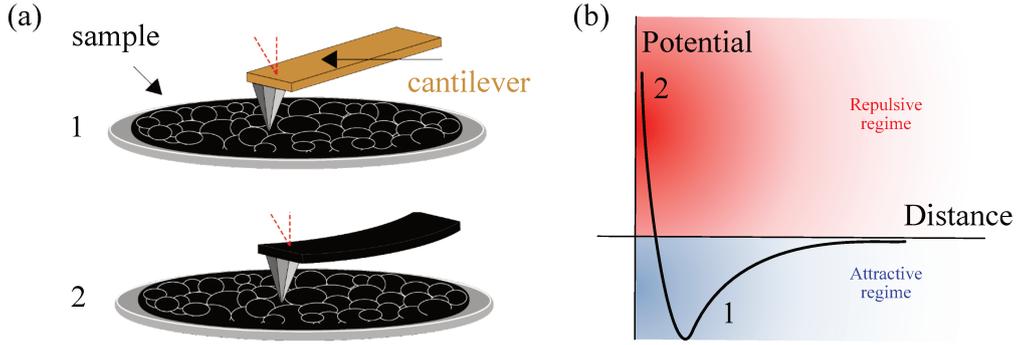


Figure 1.2: **Tip surface interaction** - (a) Scheme of the deflection of the cantilever as a function of whether it is in contact or not with the surface. (b) Potential as a function of the vertical displacement of the cantilever.

$$U_s(\delta_s) = \frac{1}{2}k_s\delta_s^2, \quad (1.4)$$

where k_c and k_s are the elastic constant of cantilever and sample. Generally and for simplicity, the tip-sample interaction potential is described by the Lennard-Jones interatomic potential,

$$U_{cs} = U(r) = -\frac{A}{r^6} + \frac{B}{r^{12}}, \quad (1.5)$$

where r is the tip-sample distance. The total interaction force F is defined as

$$F = -\frac{\partial(U_{cs} + U_c + U_s)}{\partial D}. \quad (1.6)$$

The interaction force F is attractive as the cantilever approaches the sample surface. The attractive short-range forces between the surface and the tip cause the cantilever to bend toward the surface [Fig. 1.2(a1,b1)]. The interaction force F becomes repulsive when the cantilever is close enough to the surface, inducing the cantilever to bend away from the surface [Fig. 1.2(a2,b2)].

The force-displacement curve of the AFM is the result of two contributions: the elastic force of the cantilever $F = -k_c\delta_c$ (see eq. 1.3) and the tip-sample interaction $F(D) = -\frac{\partial U_{cs}}{\partial D}$. We must mention that the Lennard-Jones force $F(D)$ only provides a qualitative description of the tip and sample interaction. The repulsive part of the force is generally more complex than the one modelled by the Lennard-Jones force [5]. The distance controlled during the measurement is not the actual tip-sample distance

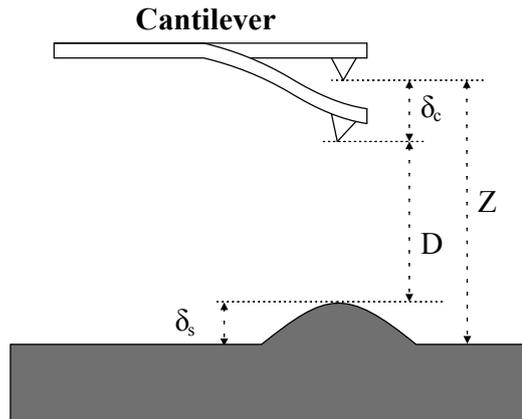


Figure 1.3: **Tip surface interaction** - D is the actual tip-sample distance, whereas Z is the distance between the sample and the cantilever rest position. These two distances differ because of the cantilever deflection δ_c , and because of the sample deformation δ_s [5].

D [Fig. 1.3], but the distance Z between the sample surface and the rest position of the cantilever. These two distances differ because of cantilever deflection δ_c and because of the sample deformation δ_s . These four quantities are related as follows:

$$D = Z - (\delta_c + \delta_s). \quad (1.7)$$

Therefore, in the experiment, the raw force curve obtained by AFM does not provide the tip-sample distance D but instead displays the vertical displacement of the cantilever Z .

1.3 AFM modes

Several imaging modes allow studying a wide variety of samples, such as rigid materials or biological samples. The most common are Contact Mode CM, Dynamic/Tapping Mode TP, Peak Force Mode PFM. This depends on how the sample is being studied in the air or the buffer solution and on the sample type. Below, we will briefly describe the three most popular modes.

- **Contact Mode CM:** this mode is the simplest operational method used by the AFM, and it was the first to be developed [1]. The tip is close to the surface or even in contact at each measurement pixel. The two methods of imaging in contact mode are in constant force or constant height. When the tip scans in

constant force, the tip is continually adjusted to maintain a specified cantilever deflection. On the contrary, in constant height, the sample must be relatively flat for the feedback loop to maintain control during scanning. However, this mode is not suitable for biological samples since it is an intrusive method and can damage the fragile biological objects due to lateral friction forces, which are important during scanning. The CM is also known as Friction Force Mode FFM since it can measure frictional forces on a surface.

- **Dynamic/Tapping Mode TP:** in this mode, the cantilever oscillates while the sample surface is being scanned. This means that the tip is quite close to the sample but not touching it; therefore, there is no contact. The feedback loop between this mode and the CM is similar, although in this case, the cantilever oscillates. The detection system is based on measuring changes in the resonant frequency or the cantilever amplitude as the interaction between the tip and the sample dampens the oscillation. This mode is also known as Intermittent-Contact Mode or Oscillating Mode. The surface topography is obtained with minimal intrusion since the oscillating tip only approaches the surface intermittently and always from top to bottom, improving lateral resolution. Additionally, lateral forces such as drag are practically eliminated.
- **Peak Force Mode:** this mode couples the two previous modes by controlling the interaction between the tip and the surface while maintaining intermittent contact. The cantilever oscillates at a very low frequency (around 2 or 8 kHz) compared to its resonant frequency (several hundred kHz). At a constant amplitude, this oscillatory mode does not cause the cantilever to bend. The feedback loop is performed by keeping the maximum interaction between the tip and the surface as the cantilever approaches. This mode is increasingly used to study viruses (and biological systems) because it has the advantage of being low intrusive with direct control of downforce, especially in liquid environments imaging.

1.4 Imaging samples

The images obtained by the AFM are iso-interaction maps; that is, each pixel (with x , y position) is associated with a z value, corresponding to the deflection of the cantilever. The image data are sampled digitally at equally spaced intervals. Typically, the obtained images are a square grid of data points, meaning that the number of lines is equal to the number of data points per line [Fig. 1.4(a)]. The representation of this matrix is an image associated with a scale that connects the colour of each pixel with the measured height [Fig. 1.4(b)].

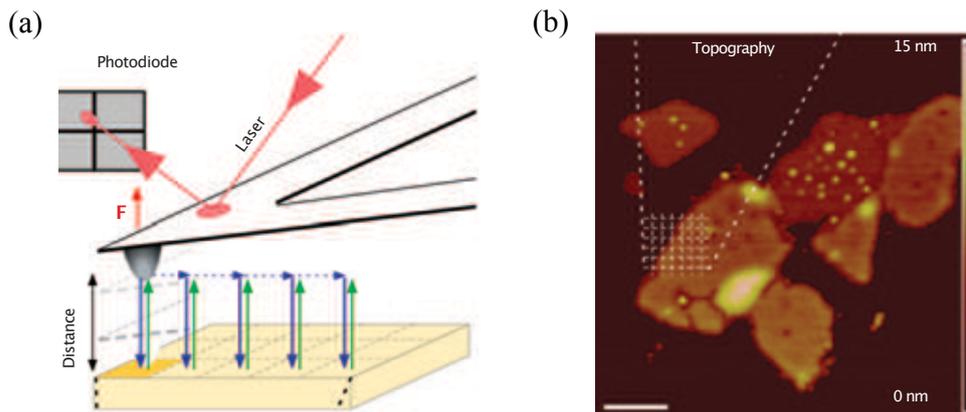


Figure 1.4: **AFM reveals topography of biological samples** - (a) The cantilever tip approaches and withdraws from the sample in a pixel-to-pixel manner. The high precision of the approach allows detection of pixel sizes $<1 \text{ nm}^2$ with a positional accuracy of $\sim 0.2 \text{ nm}$. (b) AFM image topography of native purple membrane adsorbed onto mica and imaged in buffer solution. Scale bars, 200 nm. Adapted from [6].

The AFM imaging depends on the control parameters such as surface dimension, image resolution, tip speed to make an image line and the efficiency of the feedback loop, all of which depend on the AFM mode. These mentioned parameters generally control the quality of the AFM image.

However, there is another critical parameter that must be considered to obtain AFM imaging: the cantilever. Next, we will describe the main characteristic of the cantilever and its implication in the sample topography.

1.5 Cantilever

The essential parameters of the cantilever are the sharpness of the apex and the whole tip aspect (length and width). The cantilever tip is an essential part of the AFM since it is the part that comes into contact with the sample. Currently, cantilevers are manufactured with silicon or silicon nitride due to the uniformity and reproducibility of results [7]. There are many tips types where the most common have spherical, conical or triangular shapes [8]. The choice of the cantilever tip shape depends on the surface analysed.

In addition to the different tip shapes (or probes), there are two designs for the cantilever; the "V" shaped type and the single-arm type. These show different torsional properties. The cantilever mechanical stiffness k_c varies between 0.03 N/m and 40 N/m ,

depending on the cantilever is length L , the width W and the thickness T . Although the width and length of the cantilever can be determined with great precision, the thickness presents difficulties to be measured [9]. Consequently, the manufactures usually give the cantilever spring constant with an error of 10-30 % and the user must then calibrate each cantilever by thermal noise or Sader methods [10].

The vertical resolution of the cantilever depends on two parameters [9]: (i) the noise of the photodetector since it defines the vertical movement of the piezoelectric device. The displacement dependence of the piezoelectric device is hysteretic, and it is affected by creep, which is a delay effect depending on temperature. This means that almost the entire displacement is performed at the beginning of the movement, but a little fraction is done later with a logarithmic time dependence [5]. The second problem that can affect the vertical resolution of the cantilever is (ii): the thermal noise. The cantilever oscillates at the resonance frequency with an amplitude.

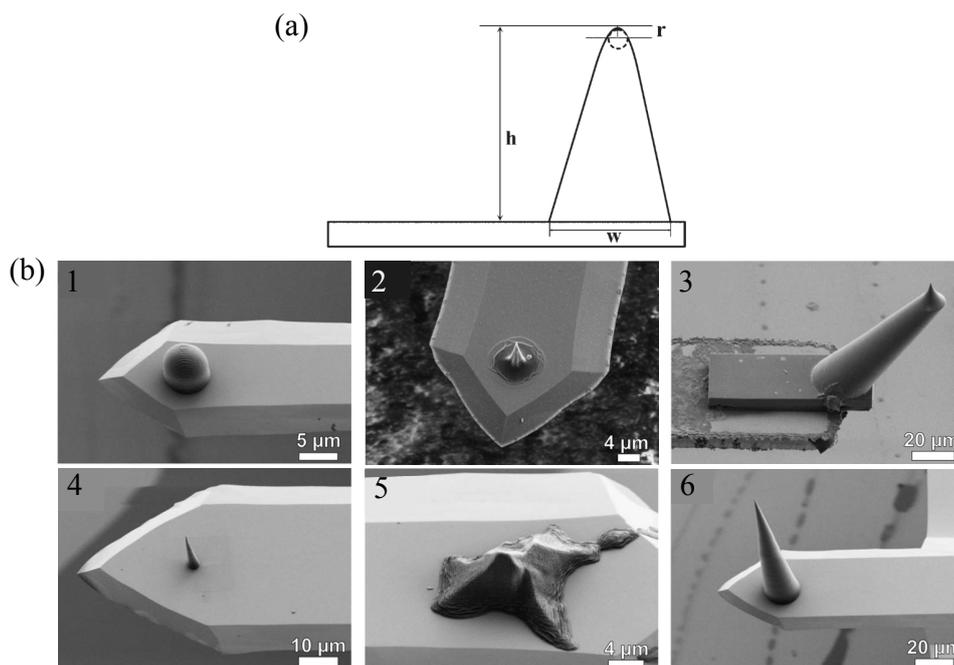


Figure 1.5: **Cantilever tip** - (a) Scheme of a cantilever where it are identify the radius of curvature r and the aspect ratio h and w [7]. (b) Several different tips: (1) spherical tip with diameter of $10 \mu\text{m}$, (2) shrunk conical tip after carbonization through pyrolysis, (3) protruding tip that is visible from the top during scanning, (4) high ($20 \mu\text{m}$) and thin ($5 \mu\text{m}$ at the base) tip, (5) triangular tip in the shape of the famous Swiss mountain “Matterhorn” and (6) Extremely long tip with a height of $100 \mu\text{m}$. Adapted from [8].

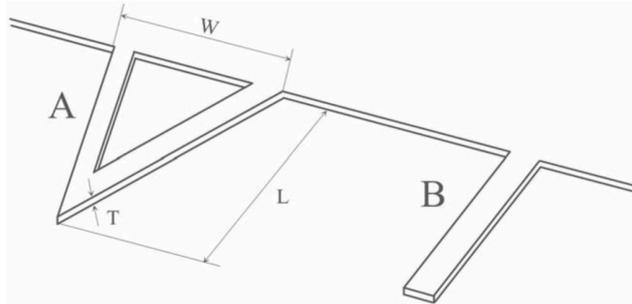


Figure 1.6: **Cantilever shapes** - Triangular and single-beam cantilevers. The force constant and resonant frequency depend on the values of width (W), length (L), and thickness (T) [7].

$$\Delta z = \sqrt{\frac{k_B T}{k_c}} \quad (1.8)$$

where k_B is the Boltzman constant, T is the absolute temperature and k_c the cantilever spring constant. For example, a cantilever with a spring constant of 0.02 N/m will have a noise of 5 Å approximately, at room temperature. One way to control thermal noise is by isolating the AFM with a chamber that allows humidity control.

The lateral resolution of AFM images is affected by artefacts that arise from the effects of convolution of the tip [11]. This effect is closely related to the tip geometry, so it does not depend on the AFM mode or if the sample is scanned in a liquid medium or air. Figure 1.7 shows two schemes where (a) the cantilever tip is larger than the object size and (b) when the round end of the tip is smaller than the object size. We notice that the object width measured by the cantilever tip W_{exp} would be greater than the actual width of the object. Case 1.7(b) shows that for triangular-shaped points, the width of the measured object W_{exp} will have an additional term that is related to the cone angle γ . Besides, Cannet *et al.*, [12] demonstrate that the convolution is increased with the object height for a constant tip radius, mainly for the narrowest objects.

Object areas measured through AFM images will always show an increased value. The only parameter that is not affected by the convolution effect between the tip and the object is the height. Thus, height is the parameter used mainly to characterise the morphology of objects.

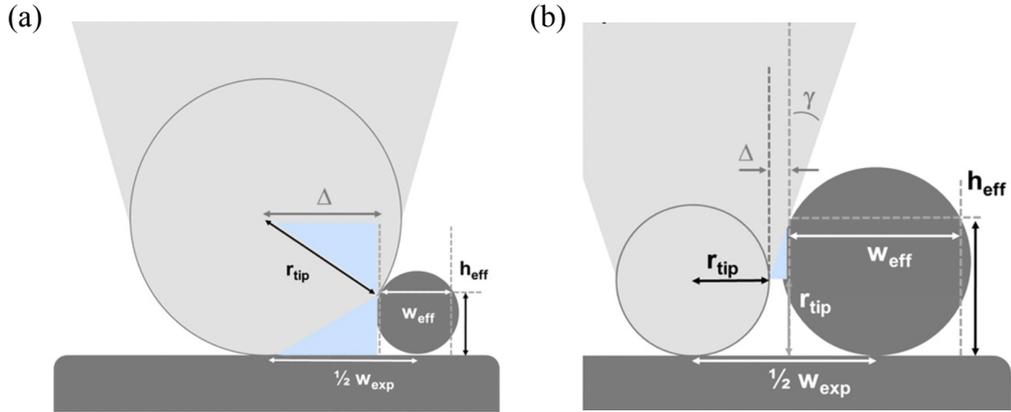


Figure 1.7: **Convolution effect between the cantilever tip and the sample** - Schematic of tip convolution effects described by trigonometric considerations for round tip end and object round. (a) The tip size is larger than the size of the object. (b) opposite case of (a) Adapted from [12].

1.6 Nanoindentation by AFM

Indentation experiments generally consist of understanding the type of cantilever-sample interaction and then extracting the sample deformation and the cantilever as two separate components. In this thesis, we have used this technique to measure the mechanical properties of viral capsids. Therefore, we here explain how to perform indentation experiments based on nanoindentation of viral capsids [13].

Unlike AFM imaging that moves the piezoelectric scanner in the x , y , z directions, nanoindentation experiments only move the scanner in the z -direction. Typically, measurements start at a sufficiently large z -distance to avoid cantilever-capsid interaction and breaking in the capsid. Once the piezo scanner begins to move up vertically, and before the cantilever tip touches the capsid, the deflection F is zero since the cantilever has not undergone any deflection [Fig. 1.8(a,d)]. When the cantilever touches the capsid and starts to deflect, it changes the quadrant photodiode signal [Fig. 1.8(b,e)]. In this case, the resulting force curve between the cantilever and capsid is approximately linear. This behaviour is elastic until the deformation is smaller than a critical value for which the capsid breaks [Fig. 1.8(c,f)]. This critical value is known as a breaking point, and it appears as a discontinuity in the force curve.

There are two classical ways to visualise the cantilever-capsid system deformation. (i) The force F curve is shown as a function of the vertical displacement of the cantilever z . In figure 1.9(a), the z - F glass curve is shifted laterally to have the contact point coincide with the contact point of the z - F capsid curve. The capsid indentation is the

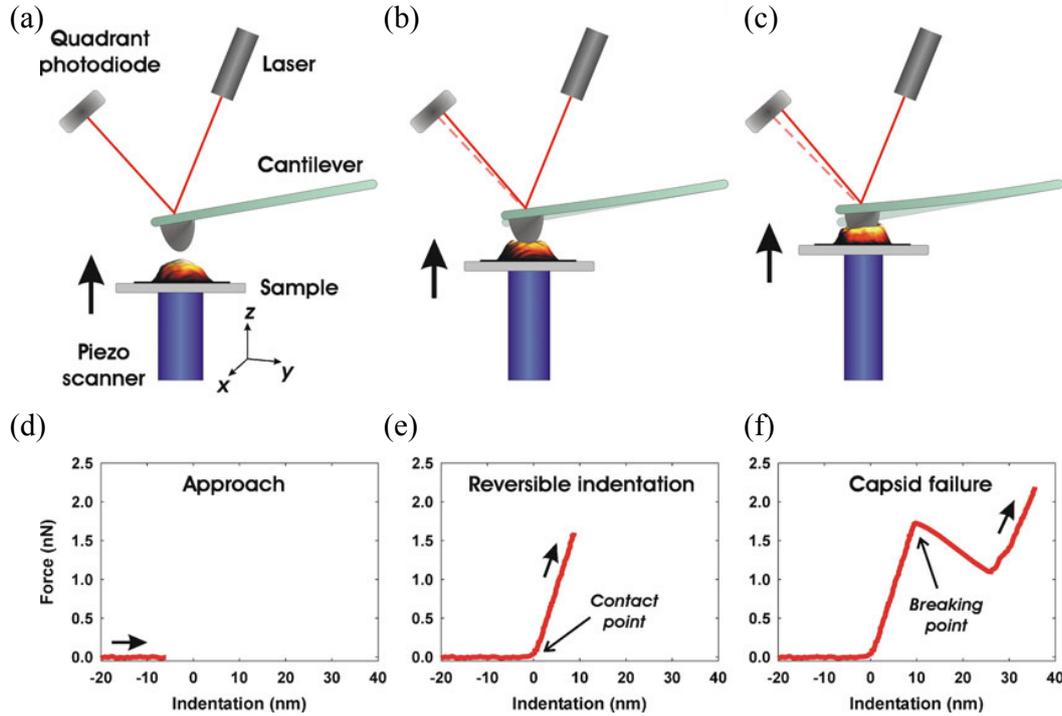


Figure 1.8: **Nanoindentation experiment on a viral particle** - (a,b,c) Schematic of an AFM nanoindentation experiment and (d–f) the corresponding force indentation curves [13].

difference between the z - F glass and the z - F capsid curve. The first, second, and third time pushing (FZ 1 Loading, FZ 2 Loading, and FZ 3 Loading), as well as the unloading curve after the first time indenting (FZ 1 Unloading), are shown. (ii) In a second way, the force F curve as a function of the indentation only provides the capsid deformation information [Fig. 1.9(b)]. All *indentation*- F curves show an incompressible behaviour at the end of indentation, indicating that one is pushing on the glass.

If the initial *indentation*- F behaviour is linear [Fig. 1.8(e)], the capsid spring constant can be deduced from the slope of the initial linear part. When the data is represented as a function of the indentation [Figures 1.8 and ??(b)], one can directly read out the approximate capsid spring constant from the curve. In contrast, if the data is represented as a function of the displacement z , one can not accurately determine the capsid spring constant from the graphs. As we mentioned, this curve contains the cantilever-capsid system deformation in which it can be described as two resorts in series. Using Hooke's law for two spring in series, the capsid spring constant can be determined as:

$$k_{capsid} = \frac{k_{total} \cdot k_{cantilever}}{k_{cantilever} - k_{total}} \quad (1.9)$$

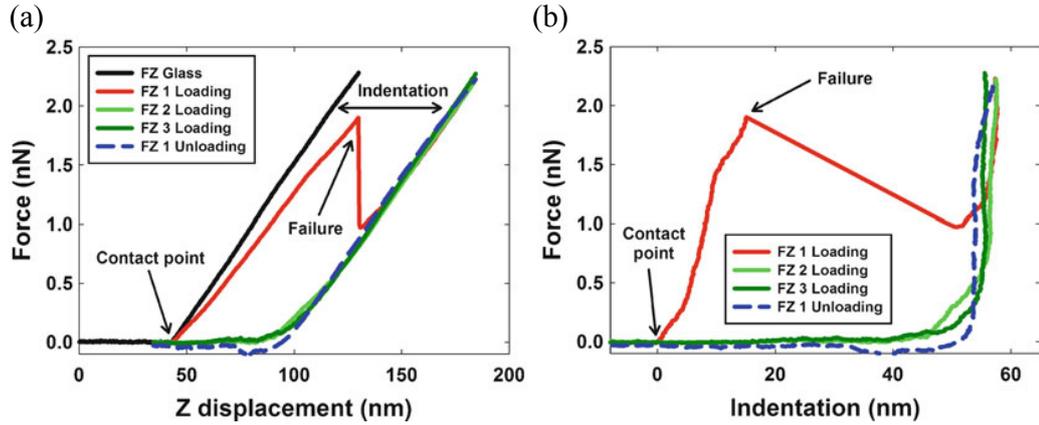


Figure 1.9: **Two ways of representing the same z - F curve data** - (a) The classical way to show the cantilever+capsid system deformation is by the force F as a function of the vertical displacement of the cantilever z . (b) The force F as a function of the indentation. The indentation of the capsid is the difference between the z - F glass and the z - F capsid curve. In this case, the force curves only shows the deformations relate to the capsid deformation [13].

where k_{total} and $k_{cantilever}$ are the slope of the linear part of the force curve for indentation on the capsid and surface, respectively.

In the next chapters, we will use the AFM, and the methods described in this section to achieve two goals: (i) characterise the R-loop objects by morphology and connect this to the pathological and physiological behaviour biological processes. (ii) study of the AAV and HBV capsids mechanical properties and their influence during the disassembly and genome release. (iii) finally, characterise the Inverted Terminal Repeats ITRs conformation *in vitro* to understand how the physical properties of this specific AAV viral DNA structure are related to its biological role *in vivo*.

Part I

Physical Characterisation of R-loops

Chapter 2

Introduction: R-loops

Index

Summary	32
2.1 The biology of the DNA	33
2.1.1 DNA structure	33
2.1.2 DNA functions	38
2.2 R-loop	41
2.3 Factors that promote the R-loop formation	42
2.4 Physiological and pathological R-loops	44
2.5 DNA studies using AFM	45
2.5.1 Visualisation of the DNA	46
2.5.2 Analysing G-quadruplexes structure in DNA	47
2.5.3 Single-Stranded DNA Binding Proteins	48
2.5.4 Study the biological process in real time	49
2.6 Mechanical properties of DNA	50
2.7 Our work in this context	52

Summary

R-loops are complex structures formed by a three-chain nucleic acid structure, which develop due to the DNA transcription process. The nucleic acid structure consists of RNA hybridised to a complementary DNA strand. Understanding the functions and

regulations of R-loops has been actively studied in recent years since these structures allow the characteristics of gene expressions and genome stability. Here, we will first explain the basic processes of gene expression: DNA replication, transcription, and translation. Then, we will explain how the R-loop is generated and its implication in genome stability.

2.1 The biology of the DNA

The DNA (Deoxyribonucleic acid) is the most important molecule in Biology because it can conserve and transfer genetic information from one generation to the next. Furthermore, it contains the characteristics of an organism and its functions [14–16]. Friedrich Miescher, in 1869, was the first person to describe DNA through the study of the leukocytes chemical composition. In experiments, he isolated an unknown substance (now known as DNA) from the nucleus of the cell, characterised by large amounts of phosphorus, named by him as nuclein. Subsequently, he demonstrated that the new substance was a characteristic component of all nuclei [17, 18]. Erwin Chargaff showed that the composition of the DNA base varies among species. DNA contains an equal number of nucleotide bases for one species: adenine-thymine and guanine-cytosine [?, ?]. James Watson and Francis Crick proposed the DNA structure model in 1953 [19], giving birth to modern molecular biology.

2.1.1 DNA structure

The DNA structure is constituted by two helical chains coiled around the same axis. The bases are complementary and located inside the helix, while the phosphates are located outside [Fig. 2.1(a)]. The backbone of each DNA strand is composed of two polymers consisting of a 5-carbon sugar [gray pentamer in figure 2.1(b)], a phosphate group (PO_4^-) [gray square in figure 2.1(b)], a heterocyclic nitrogen-containing base [colored structures in figure 2.1(b)] and three components that constitute a *nucleotide*. There are two types of bases: Purines and Pyrimidines. The purine bases in the DNA are guanine and adenine; their structure is a planer system of 9-atoms: 5-carbons and 4-nitrogens. The pyrimidine group contains the cytosine, thymine and uracil bases, and their structure is a system of 6-atoms: 4-carbons and 2-nitrogens. An N-glycosidic bond generates the connection between the sugar and the bases, forming: adenosine, guanosine, cytidine and thymidine, the four natural nucleosides. Nucleotides are joined together by phosphodiester linkage between 5'carbon of one deoxyribose group to the 3'carbon of the next to form nucleic acids. The ribonucleic acid RNA differs from

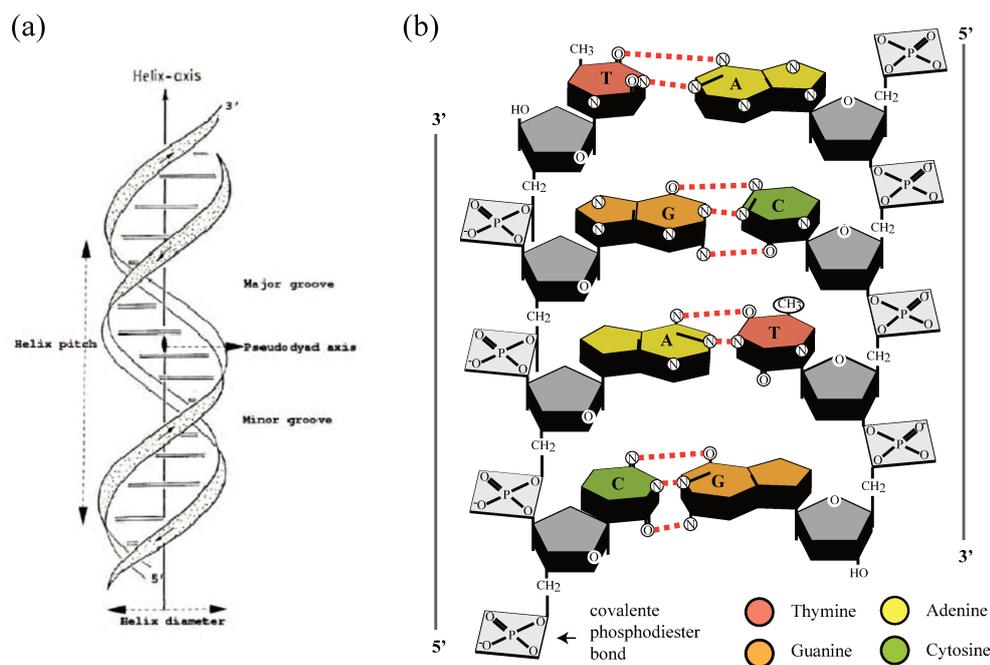


Figure 2.1: **Structure of the DNA** - (a) Structure of DNA by Watson and Crick [19]. (b) Chemical structure of DNA; hydrogen bonds shown as dotted red lines. The sugar-phosphate backbones are on the outside of the double helix. The direction of each strand is identified by numbering the carbons 3' and 5' in each sugar molecule.

the DNA by having a different sugar and base: ribose and uracil. Uracil in DNA is visualised when spontaneous deamination convert cytosine to uracil [20].

The double helix of DNA has certain characteristics. Two bases are separated by a distance of 3.4 Å, and the chain has a pitch of 34 Å [Fig. 2.1(a)]. The helix diameter is 20 Å; this parameter is used to characterise the DNA molecules in the AFM images. Each DNA strand has an end where there is a 3' and 5' type bond, conferring a direction to the strand. In a double helix, the nucleotides of each strand are complementary. That is, the strands are non-identical. During the transcription process, the synthesised RNA molecules are different depending on which of the DNA strands is transcribed. Antiparallel strands can be generated between a DNA strand and an RNA strand. These are called DNA-RNA hybrid strands.

2.1.1.1 Types of DNA helices (A-, B- and Z-DNA)

The DNA molecule is a flexible and dynamical structure. Due to its ability to rotate around the nucleotide bonds, DNA can take various forms. The three most popular configurations are A-DNA, B-DNA and Z-DNA. Through X-ray diffraction, Rosalind

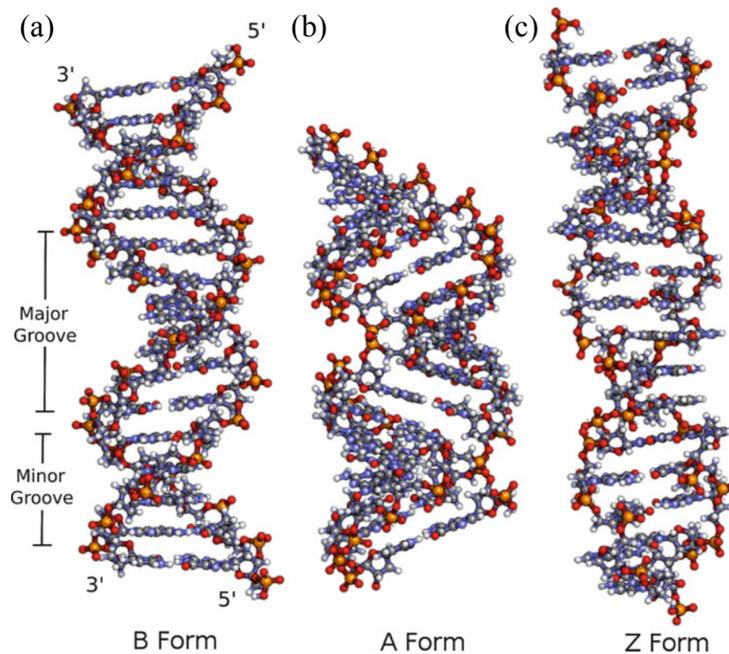


Figure 2.2: **DNA models** - The three most common types of DNA helices: (a) B-form DNA is right-handed, with 10.5 bp per helical turn. It is the normal, expected form of dsDNA under physiological-like solution conditions. The narrower minor groove and wider major groove are indicated. (b) A-form DNA is also right-handed and has a shorter more-compact helix with 10 bp turn. (c) Z-DNA is left-handed and slightly stretched with 12 bases per full turn of the helix [21].

Franklin and Maurice Wilkins [22] showed that DNA fibres become B-DNA type structures under high humidity conditions ($>75\%$). This corresponds to the average structure of DNA under physiological conditions. Also, The B-form is the closest to the original Watson-Crick model [19]

On the contrary, if the humidity is reduced to 75 %, the DNA adopts a conformation called A-DNA [22, 23]. Structurally, A-DNA is wider and shorter for the same number of nucleotides and has a greater number of bases per turn than B-DNA. Therefore, A-DNA has less rotation for each base and helix pitch. However, both the A-DNA and B-DNA structure have a right-handed helical sense [Fig. 2.2(a,b)]. The condition of A-type gives the double helix a greater stiffness than the B-type double helix. Bending the A-type configuration is energetically less favorable [24].

A third model, the Z-DNA model, arises from studying a double helix of the CGCGCG hexanucleotide through X-ray diffraction [25]. The Z-DNA model has different characteristics than forms A and B, the most important being the left-handed helical sense. Regarding the dimension, Z-DNA has a narrower helix, longer and with

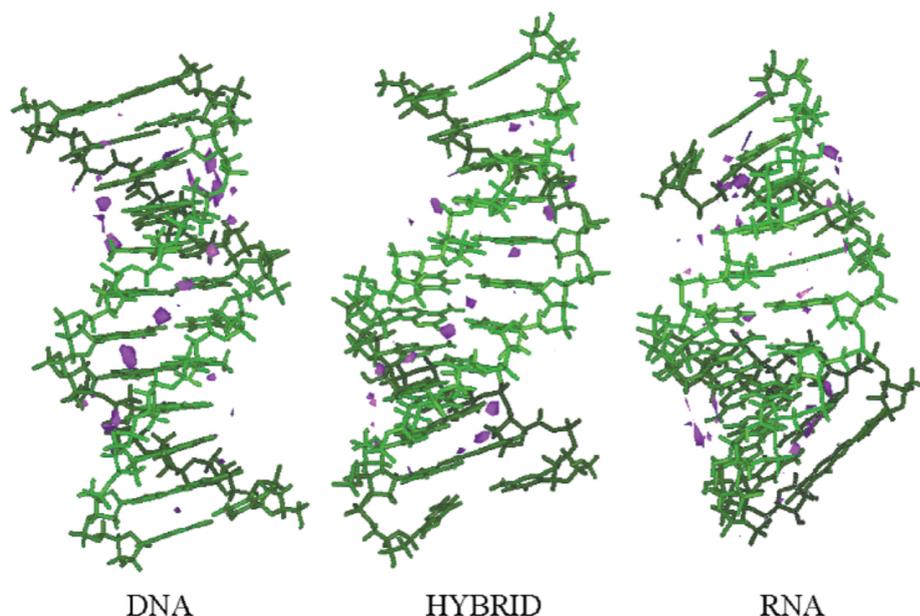


Figure 2.3: **Structure of the DNA-DNA, DNA-RNA and RNA-RNA duplexes** - Solvation map density for the DNA and RNA duplexes, and the DNA-RNA hybrid [30].

a greater number of bases per turn than B and A. Therefore, it is a more stretched helix [Fig. 2.2(c)].

The DNA-RNA hybrids tend to crystallise in the A-type form, but in solution or high relative humidity, their conformation shows an intermediate geometry between A- and B-type [26–35]. As RNA form an A-type double helix [30, 36–39], the hybrid is a combination of the two conformations [Fig. 2.3].

2.1.1.2 Non B-type

The nucleotides distribution in genomes is not random. Many DNA sequence patterns exist throughout genomes, such as direct repeats of homo-, di-, or tri-nucleotides, inverted repeats and mirror repeats. Repeated sequences have the capacity to adopt alternative conformations called non-B DNA structures.

To date, nearly a dozen types of non-B DNA structures have been described, including hairpins/cruciforms [Fig. 2.4(a)], Z-DNA [Fig. 2.4(b)], triplexes (H-DNA) [Fig. 2.4(c)], tetraplexes (G-quadruplex) [Fig. 2.4(d)], slipped and sticky DNA [Fig. 2.4(e)] [40]. Several studies have shown that the Non-B DNA-forming sequences in genomes affect DNA replication and transcription and contribute to genome instability [41–44]. Next, we will briefly explain the G-quadruplex structure and triplex

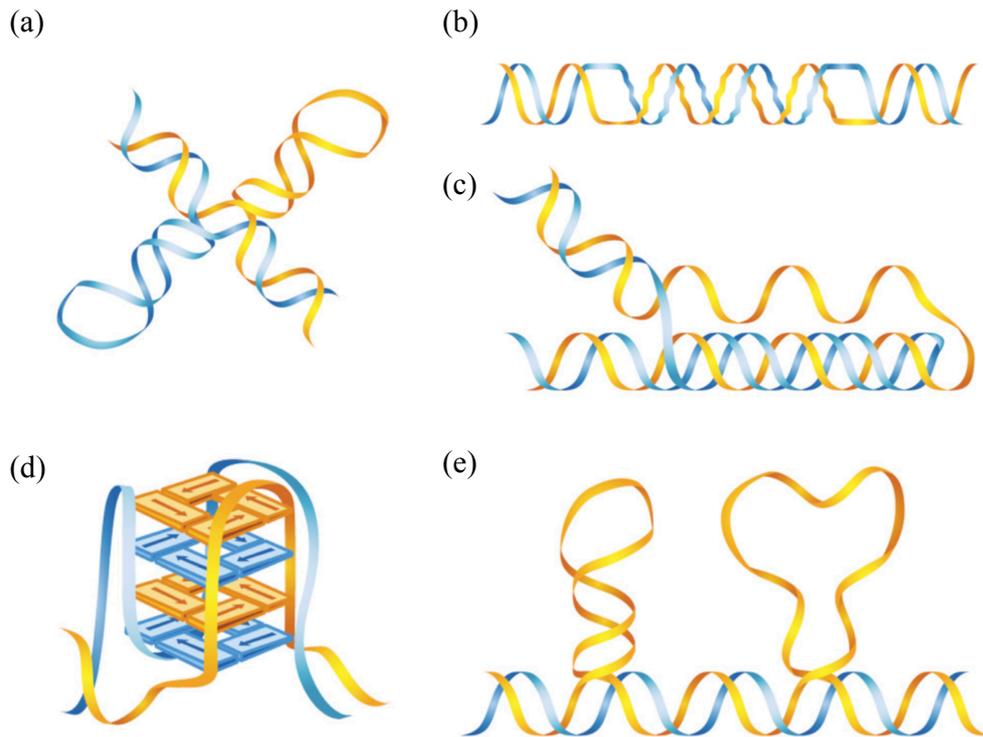


Figure 2.4: **Non-B DNA structures** - (a) Cruciform DNA, (b) Z-DNA, (c) H-DNA (triplex DNA), (d) G-quadruplex (tetraplex) DNA, and (e) Slipped DNA [40].

DNA because, during the transcription process, the non-template DNA strand can adopt these structures in many repetitive sequences. There are hints that this further stabilises the R-loops [45].

The H-DNA-forming sequences are very abundant in mammalian genomes [43]. This structure forms at regions containing mirror repeat symmetry, where one half of the tract can dissociate into single strands by using the energy provided by supercoiling. One of the single strands can then swivel its backbone parallel to the purine-rich strand in the remaining duplex, forming a three-stranded helix in this half of the region, leaving its complementary strand unpaired.

G-quadruplex DNA forms four-stranded, and the sequence can adopt a variety of folds, dependent in part on whether they are intra- or intermolecular. The quadruplexes contain the basic repeating motif, the G-quartet, which comprises four guanine bases held in a plane by Hoogsteen hydrogen bonding. Typically, three or four G-quartets are stacked one on top of each other within a quadruplex, held together by $\pi - \pi$ non-bonded attractive interactions [46].

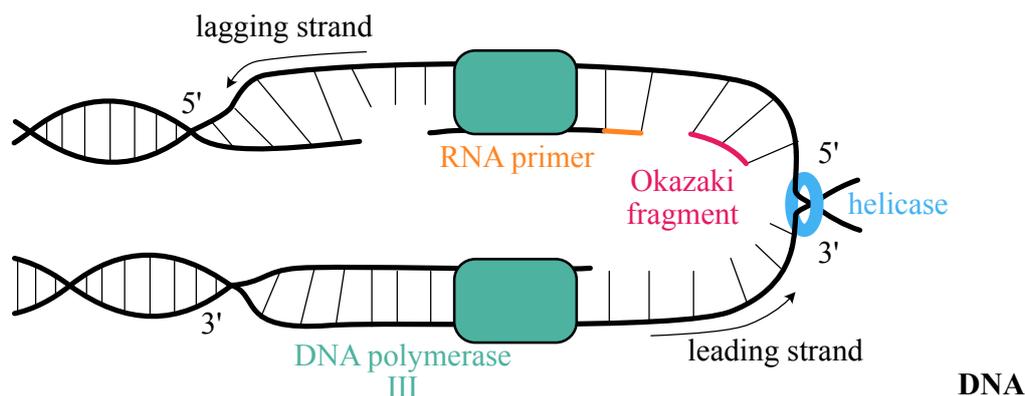


Figure 2.5: **Replication of the DNA** - The two strands of the dsDNA are pulled apart and each serves as a template for synthesis of a new complementary strand.

2.1.2 DNA functions

The DNA is responsible for containing and transporting genetic material. Different functions can be described: (i) the replication process, responsible for generating new DNA molecules, (ii) the transcription, which creates different RNA types that transport the genetic material. (iii) Moreover, the translation converts these RNA molecules into proteins. In the next sections, we will briefly describe these processes, and their impact on gene expression [47].

2.1.2.1 Replication of DNA

The replication process occurs during cell division, and then the double-strand DNA is copied, generating two identical DNA molecules. The replication is a complicated process in which several enzymes participated. The set of proteins responsible for DNA replication is known as the replisome. The essential steps that are involved in this process are [Fig. 2.5]:

- (i) the separation of the two strands. The topoisomerase is in charge of unscrewing the DNA, while the helicase breaks the hydrogen bridges that join the bases to open the chains.
- (ii) Once the chains are opened, a primase enzyme is responsible for starting the duplication process by creating a small RNA molecule called a primer chain.
- (iii) The DNA polymerase III enzyme bind to the primer RNA adds deoxyribonucleotide acids to create the new strand. However, this enzyme can only add

deoxyribonucleotide acids in a 3' to 5' direction; therefore, the replication is bidirectional. Given this configuration, there are two types of chains, a continuous chain (leading strand) and a discontinuous one (lagging strand). In the case of the continuous chain, only one primer is needed. Nevertheless, in the leading strand, several RNA primers are synthesised. Next, the DNA III polymerase synthesising DNA fragments between one primer and another: these fragments are called Okazaki fragments.

- (iv) The exonuclease removes all the RNA primers, and then the DNA polymerase fills those spaces.
- (v) Finally, DNA ligase connects all the DNA fragments in both strands. Topoisomerase returns the coiled state of the DNA molecule.

2.1.2.2 Transcription of DNA

The transcription process consists of copying the information contained into the DNA and storing it in an RNA molecule. Within the transcription process, a messenger, transfer and ribosomal RNA can be obtained. There is a specific RNA polymerase to synthesise each RNA molecule: RNA polymerase I, II and III, respectively. The messenger RNA, created in the cell nucleus, crosses the nuclear membrane to the cytoplasm to synthesise proteins. We should mention that transcription and translation occur separately in eukaryotic cells since transcription occurs in the nucleus and translation in the cytosol. In prokaryotic cells, the transcription and translation process can co-occur in the cytosol. The transfer RNA is responsible for transporting specific amino acids to a protein in formation. The ribosomal RNA combines with proteins to form ribosomes, and its function is to create peptide bonds between amino acids and proteins. The overall process of obtaining proteins from DNA information is known as gene expression.

The DNA strand contains a specific sequence known as a gene used as a template to create the messenger RNA. The complementary strand is the non-template or coding strand since this chain has the same sequence as the RNA chain. Nevertheless, the newly made RNA strand has U nucleotides instead of T nucleotides.

The template strand also contains a sequence that allows the initiation of transcription called promoter. Specific proteins (transcription factors) adhere to the promoter allowing RNA polymerase to bind to this promoter and therefore initiate transcription. The RNA polymerase is the essential transcription enzyme since it is responsible for transcribing the DNA into RNA. Once the RNA polymerase is located in the promoter

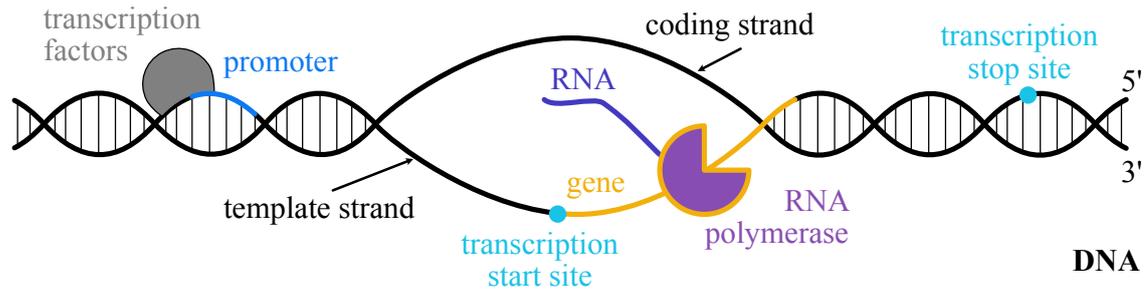


Figure 2.6: **Transcription of the DNA** - Segments of the DNA sequence are used to synthesise molecules of RNA.

position, the polymerase advances while copying the template strand sequence, and then the RNA begins to grow thanks to the addition of new nucleotides. As seen in the diagram in figure 2.6, the promoter is not transcribed; the transcription starts at the beginning of the gene (transcription start site) and reaches a termination sequence (transcription stop site).

2.1.2.3 Maturation process of the messenger RNA

The newly created RNA molecule is not ready to leave the nucleus to the cytoplasm; it must go through a maturation process. The messenger RNA molecule that we will call pre-mRNA undergoes a series of modifications so that ribosomes can recognise it and prevent it from being degraded by nucleases in the cytoplasm [Fig. 2.7].

First, a 7-methylguanosine cap molecule is added at the 5' end and is responsible for binding the mRNA to ribosomes in the cytoplasm. At the other 3' end, one adenine chain known as the poly-A chain is added in a polyadenylation process [Fig. 2.7(b)]. In this new molecule, the sequences that are not translated are eliminated. The process by which introns are removed is known as splicing. Consequently, the new messenger RNA molecule will be shorter than the initially translated one [Fig. 2.7(c)]. Finally, the messenger RNA is ready to go out into the cytoplasm and be translated, which would be the next step in gene expression.

2.1.2.4 Translation of DNA

The goal of translation is to generate a chain of peptide-linked amino acids. Then, the post-translational modifications will give a specific shape to the protein. The translation process is carried out thanks to the action of ribosomes.

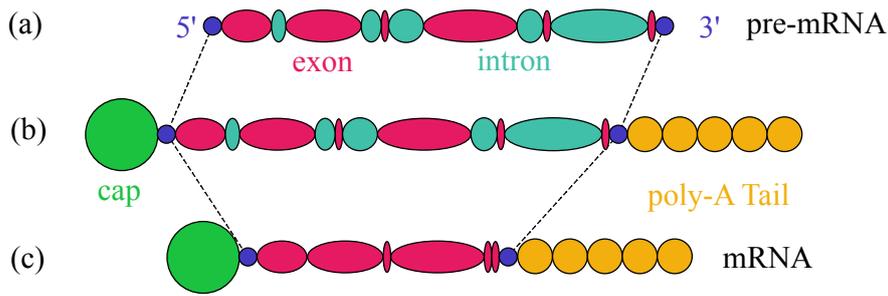


Figure 2.7: **Maturation process of the messenger RNA** - Scheme of mRNA maturation process in the nucleus.

The genetic code is based on the reading of the bases present in the messenger RNA. The four bases of RNA are uracil, adenine, guanine, and cytosine (U A G C). A triplet of coding nucleotides (bases) makes up a codon structure; RNA contains 64 coding codons. The codon sequence codes for single amino acid and different codons can encode each amino acid.

Before the translation process begins Fig. 2.8, translation factors bind to the cap molecule located at the 5' end of the RNA, allowing the lower part of the ribosome to bind to the messenger RNA. The lower part of the ribosome will advance to locate the initiation codon AUG, which encodes the amino acid methionine. Next, the upper subunit of the ribosome is attached, and the translation process begins. The molecule in charge of synthesising amino acids is transfer RNA, where the codon of the messenger RNA binds with its specific bases triplet (anticodon) of the transfer RNA. Then, elongation factors, which allow the ribosome to advance over messenger RNA, are attached. For the next codon of the messenger RNA, another transfer RNA is attached, which has a complementary anticodon and carries another amino acid. The two amino acids are joined by peptide bonds, thanks to the peptidyl-transferase enzyme. The ribosome will advance until it finds the stop codon. Translation termination proteins will then enter, allowing the two ribosome subunits to disassemble. Finally, the amino acid chain is released.

2.2 R-loop

The R-loop is a structure made up of three single strands, two of which form a DNA-RNA hybrid, and the third is a single strand of DNA that is similar to the strand

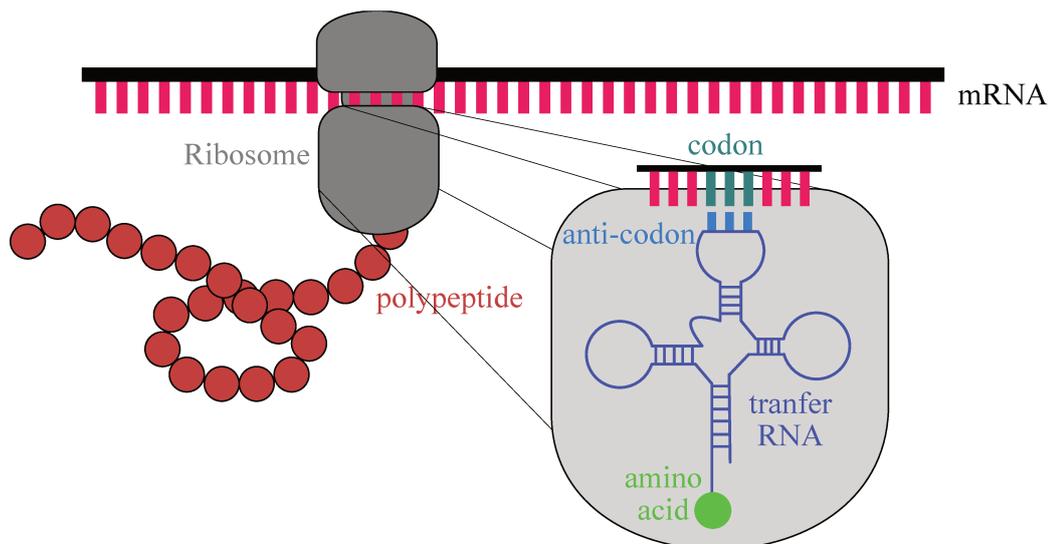


Figure 2.8: **Translation of the DNA** - During the translation of DNA, the ribosome generate a chain of peptide-linked amino acids. The molecule responsible for the synthesis of amino acids is the transfer RNA, and this is carried out by mediating the union between the codons of the messenger RNA and the anticodons of the transfer RNA.

of RNA. The R-loop is formed during the DNA transcription when the nascent RNA remains hybridised with its DNA template through Watson-Crick interactions, whilst the non-template DNA strand remains single-stranded [Fig. 2.9]. While DNA-RNA hybrids are structures that arise naturally during replication and transcription, R-loops do not occur very frequently [48]. These structures were not considered a subject of study but rather by-products of the transcription process until later studies revealed that R-loops could be connected to human diseases such as cancer and neurological disorders [49–53].

2.3 Factors that promote the R-loop formation

Several factors can promote R-loops formation [54], the most frequent of which occurs in GC or purine-rich regions. Both *in vitro* and *in vivo* observations show that the capacity of transcription to form R-loops is determined at least in part by the sequence of the nontemplate strand. In particular, the nontemplate strands containing stretches of consecutive G residues increases the probability of forming co-transcriptional R-loops [55–60].

DNA topology can play a role in the formation of the R-loop. During the transcription process *in vitro*, DNA elongation generates negative supercoiling at the end of the

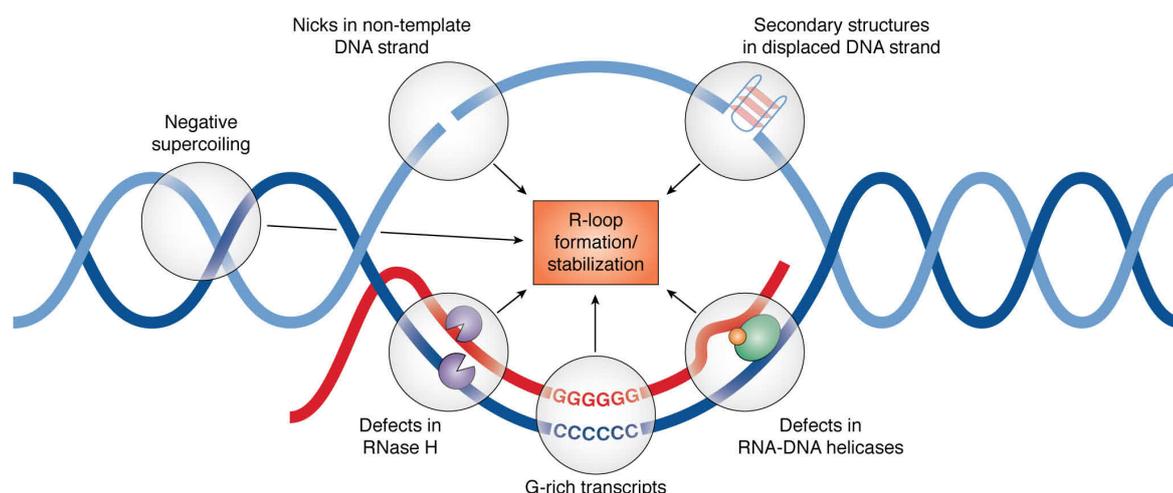


Figure 2.9: **Scheme of factors promoting R-loop formation** - The factors that can promote the formation of R-loops are: negative supercoiling, nick or secondary structures (G-quadruplex) in the nontemplate strand, G-clusters on the RNA transcript and defects in the RNase H and RNA-DNA helicases [54].

hairpin, promoting the formation of the RNA-DNA hybrid [61]. Through a statistical mechanical equilibrium model, Stolz *et al.*, 2019 [62] suggest that the supercoiling-induced strand separation may not be strictly necessary for R-loop initiation. Instead, R-loops may absorb negative supercoiling, relieving stress on the DNA molecule, contributing to genome stability. Nicks in non-template strands can also favour R-loops formation even in regions without rich G [55]. In this reference, the authors suggest that the non-template strand instability in hybridising with the template strand is due to the loss of strength of the non-template strand at the nick. Secondary structures formed at the non-template ssDNA, such as G-quadruplex (G4), can generate a less stable re-annealing of the two DNA strands. Then it can promote hybridisation between the RNA and the template single-strand DNA [63].

The defects or absence of enzymes implicated in the R-loops manipulation can also be a cause of their stability [64, 65]. The enzymes that help to prevent the formation of R-loops and avoid their accumulation are the RNase H, RNase H1 and RNase H2 enzymes [66–68]. Also, R-loop *in vivo* removal is achieved by DNA-RNA helicases that unwind the hybrid [Fig. 2.10(a)] [69]. Li and Manley in 2005 [68] proposed that in eukaryotic cells, DNA-RNA hybrids are prevented by the coating of the nascent RNA molecule with proteins.

As we have mentioned, the negative supercoiling generated behind the RNA polymerase favours the formation of the DNA-RNA hybrid bacteria, yeast and human cells [67, 69–71]. Topoisomerase 1 can resolve the negative supercoiling and thus pre-

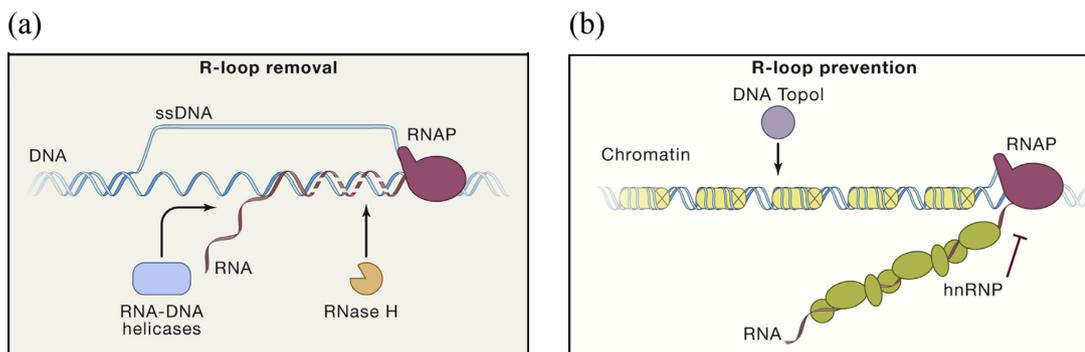


Figure 2.10: **Avoiding R-loop formation** - (a) The RNase H enzyme degrades the hybrid, and DNA-RNA helicases are unwinding the hybrid, thus eliminating the R-loop. (b) R-loop formation is prevented by specific RNA-binding proteins that are involved in RNA biogenesis (hnRNP), by topoisomerase 1 (Topo I) that resolves the negative supercoiling behind the elongating RNA polymerase II (RNAP), and by chromatin. Adapted from [69].

vent the R-loop formation. In eukaryotic cells, chromatin is another factor that can protect the genome from R-loop formation [69].

2.4 Physiological and pathological R-loops

The RNA-DNA hybrids are structures that form naturally, emerging as important regulators of genome stability [72], the modulation of gene expression and chromatin state [73], chromosome compaction [74], transcription termination [60, 75–77], DNA repair [78–81] and transcription regulation of specific genes [61, 82, 83]. A specific gene can be regulated through an R-loop and modulate physiological processes of the flowering of plants in response to temperature changes [63]. R-loops have been shown to form in G-rich transcription termination regions, promoting RNA polymerase to pause before termination [60]. R-loops have also been associated with DNA repair. For example, when DNA is broken, the RNA-DNA hybrid can be an alternative structural element to repair DNA [84]. Keskin *et al.*, [85] demonstrated that yeast, *S. cerevisiae*, can repair a restriction enzyme-mediated break using an mRNA as a template.

Adversely, studies have shown that R-loops could also be associated with pathological DNA damage, such as double-stranded DNA breaks and genome instability [61, 82]. Since the single strand of DNA is less chemically stable than the double strand of DNA, the R-loops induce the non-template ssDNA to be exposed to possible mutagenesis associated with transcription, and recombination [86–88]. Modifications of ssDNA generated by mammalian cell proteins produce abasic sites (loss of purine or pyrimi-

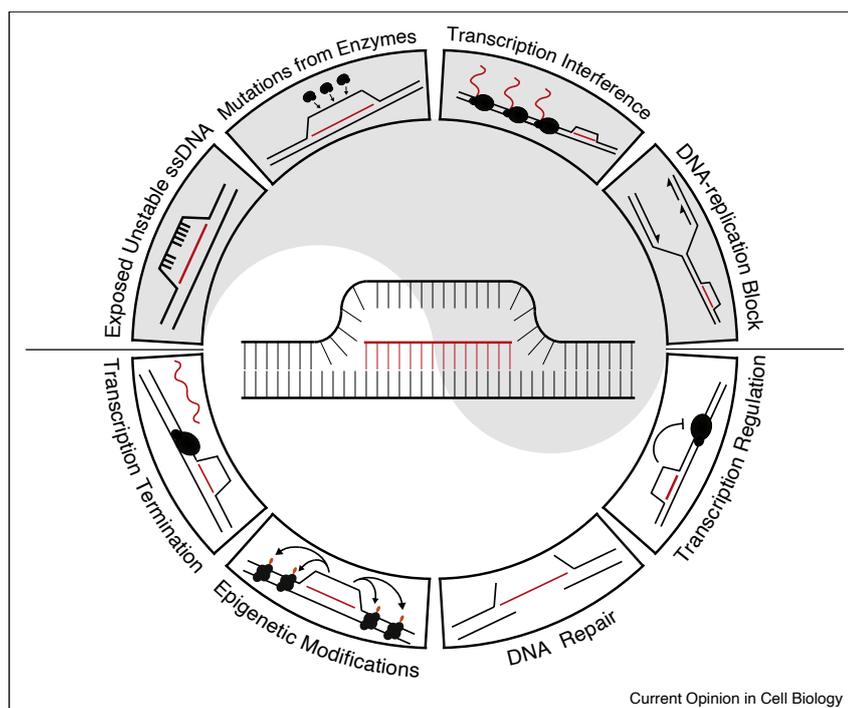


Figure 2.11: **R-loop as a factor physiological and pathological** - Different process by which the R-loop can impact on the genome stability, but also contributes positively in regulating cellular processes [82].

dine base) that lead to base substitution or nicks in ssDNA [76, 89, 90]. R-loops can cause DNA breakage due to hairpin collapse during the replication and transcription processes, as well as blocking the transcription and interfering with the transcription factor binding [48, 91–95, 95–99]. Skourti-Stathaki *et al.*, [76] showed that the presence of R-loops led to antisense transcription, formation of double-stranded RNA, and activation of the RNA interference machinery.

2.5 DNA studies using AFM

As we mentioned at the beginning of this thesis, the AFM is a fundamental tool for imaging and characterising DNA molecules, viruses and cells, and, most importantly, to study dynamical processes in real-time. We will show some examples of DNA parameters that could be obtained by AFM imaging.

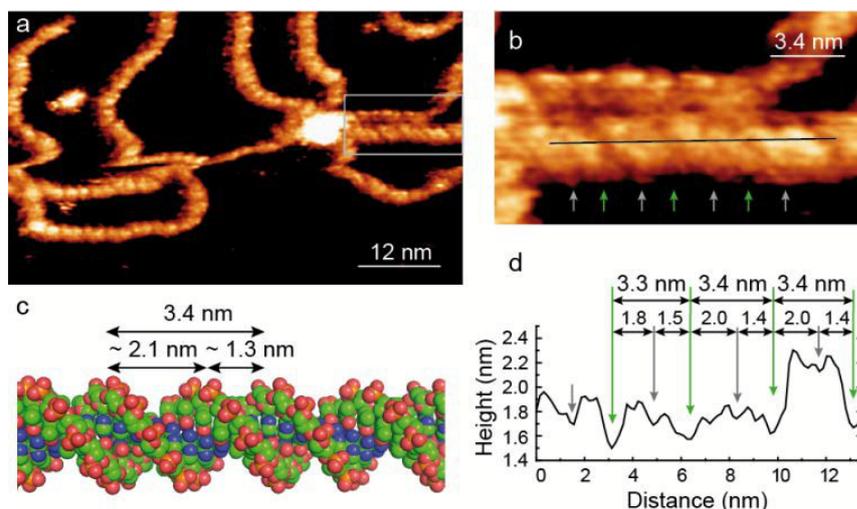


Figure 2.12: **High-resolution AFM of DNA molecules** - High-resolution imaging of dsDNA using AM-AFM. (a) A periodic corrugation and a helical structure along the molecule is observed. (b) Detail at higher magnification of (a) that highlights the double-band corrugation corresponding to the major (green arrow) and minor (gray arrow) grooves. (c) Model of dsDNA showing relevant dimensions. (d) Height profile taken along the line in (b). Color scale (from dark to bright) in (a) and (b) was adjusted to enhance the corrugation observed along the dsDNA (2.9 nm total range in (a) and 1.7 nm in (b)) [102].

2.5.1 Visualisation of the DNA

Imaging double-stranded DNA (with a diameter of roughly 2-nm) in ambient conditions with a tip with a radius of curvature of 10 nm yields a DNA image of ~ 5.7 nm. This is 3 times wider than the sample. Recent advances in the fabrication of ultrasharp AFM tips with a radius of curvature ~ 1 nm allows high-resolution studies of nucleic acids of different types, and conformations [100, 101]. However, occasionally, high-resolution images have been obtained with regular tips, suggesting that these tips were terminated with sharp spikes (asperities) that provide the necessary high-resolution topography.

Figure 2.12 show an dsDNA AFM image in air using a tip with a radius of 8 nm [102]. The authors achieved imaging the dsDNA at a surprisingly high resolution, identifying a periodic corrugation of the dsDNA structure, which correspond to the helix major and minor grooves.

Several studies have shown that it is possible to visualise DNA molecules through AFM imaging. However, the deposition of DNA molecules on the surface can generate a distortion of its structure. Such distortion depends on the physical and chemical characteristics of the surface and the DNA-surface interactions. Some methods using

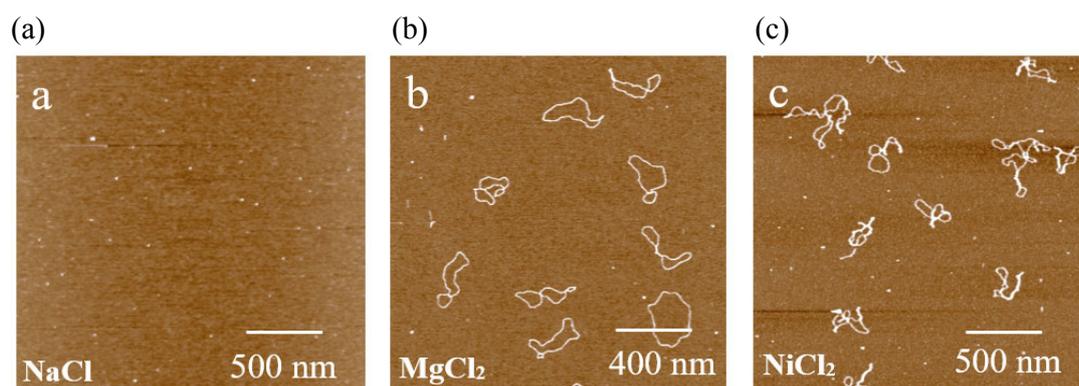


Figure 2.13: DNA plasmid deposited on mica for different divalent cations - AFM images of mica showing the conformation of absorbed plasmid DNA with different background ions: (a) NaCl, (b) MgCl₂ and (c) NiCl₂ [108].

divalent cations work very well for imaging linear DNA [103–106]. Ellis *et al.*, [107] showed that it is possible to immobilise the DNA on mica by monovalent cations. However, a rather long incubation period is required for the DNA immobilisation step.

Recently, Freeman *et al.*, [108] showed that DNA conformation in mica is affected by background ionic potential. Thus, the higher the ionic potential, the more twisting of the DNA plasmid. Figure 2.13 shows three AFM images of the DNA imaging in the air where the DNA plasmid was adhered to the surface using NaCl, MgCl₂ and NiCl₂. When NaCl is used as a background electrolyte, DNA does not adsorb to mica [Fig. 2.13(a)]. In the MgCl₂ case, the entire plasmid is adsorbed, and it is generally adsorbed in the form of a ring or with one or two turns [Fig. 2.13(b)]. When using NiCl₂, DNA supercoils and circular rings are not observed. The latter indicates a greater number of stronger absorption sites that prevent the structure relaxation during absorption [Fig. 2.13(c)]. Hence, they suggest that the local mineral surface charge affects the DNA absorption, particularly ion bridging.

2.5.2 Analysing G-quadruplexes structure in DNA

As we have mentioned previously, G-quadruplex structures that favour the formation of R-loops. Furthermore, many natural proteins have been found to interact specifically with G-quadruplexes [109]. The formation of G-quadruplexes in G-rich regions of DNA is believed to affect DNA transcription, and replication [110, 111, 111, 112]. Given its potential involvement in cellular function, it is important to understand the mechanism of G-quadruplex formation and the perturbations in the DNA structure involved in its generation.

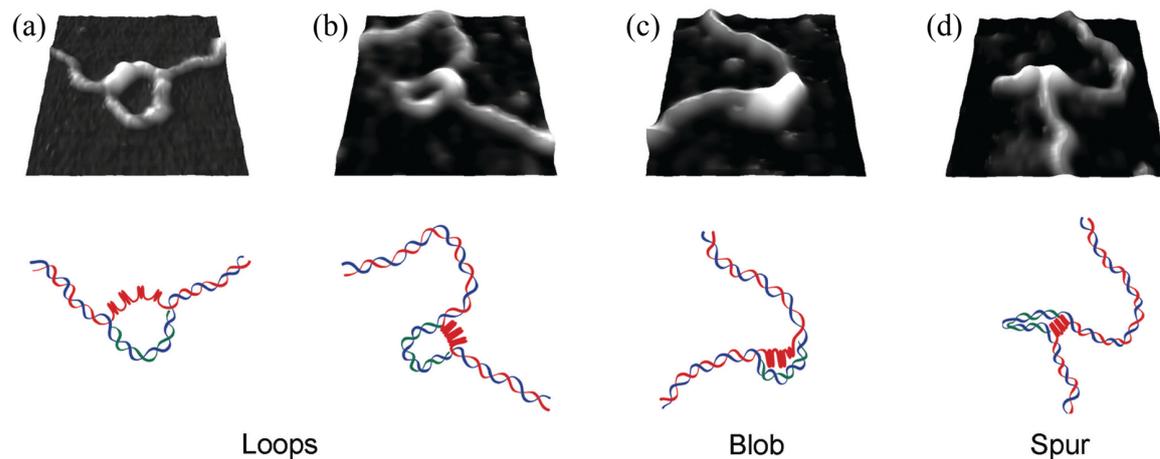


Figure 2.14: **Characterisation of G-quadruplex structures by AFM imaging** - (top) Zoomed AFM images of G-quadruplex structures with corresponding diagrammatic representations of the DNA arrangement. Areas ($140 \times 140 \text{ nm}^2$) showing regions of transcribed plasmids containing loops (A, B), a blob (C) and a spur (D). The blue lines in the diagrams represent the non-coding strand of the DNA; the red lines represent the G-rich coding DNA strand and the green lines represent hybridized mRNA [113].

Through an analysis of AFM images of stable RNA-DNA hybrid loops generated by the transcription of the plasmid pPH600, which contains a fragment of the murine immunoglobulin $S_{\gamma 3}$ switch regions, Neaves *et al.* [113] showed that the non-RNA-containing portion folds into G-quadruplexes. Following *in vitro* transcription, the plasmid exhibits an asymmetric loop [Fig. 4.9(a,b)], a large asymmetric blob [Fig. 4.9(c)] or a spur-like projection at the appropriate position on the DNA contour [Fig. 4.9(d)]. The loops disappeared following the treatment of the transcribed plasmid with RNase H. In the presence of K^+ in the transcription buffer, the G-quadruplex structure stability falls, which is consistent with the known effects of monovalent cations on G-quadruplex stability [114].

2.5.3 Single-Stranded DNA Binding Proteins

The single-stranded DNA has a high affinity to SSB proteins (Single-stranded binding protein), and as the ssDNA participate in all the genetic processes, it is important to elucidate the mechanism of SSB-DNA interaction. The high flexibility and smaller filament height of single-stranded DNA complicate the reliable visualisation of ssDNA-SSB complexes [117]. Shlyakhtenko *et al.*, [115, 116] proposed to study the SSB-DNA interaction by which the DNA construct consists of a relatively long DNA duplex, to which ssDNA of the desired length is attached [Fig. 2.15(a)]. In the air,

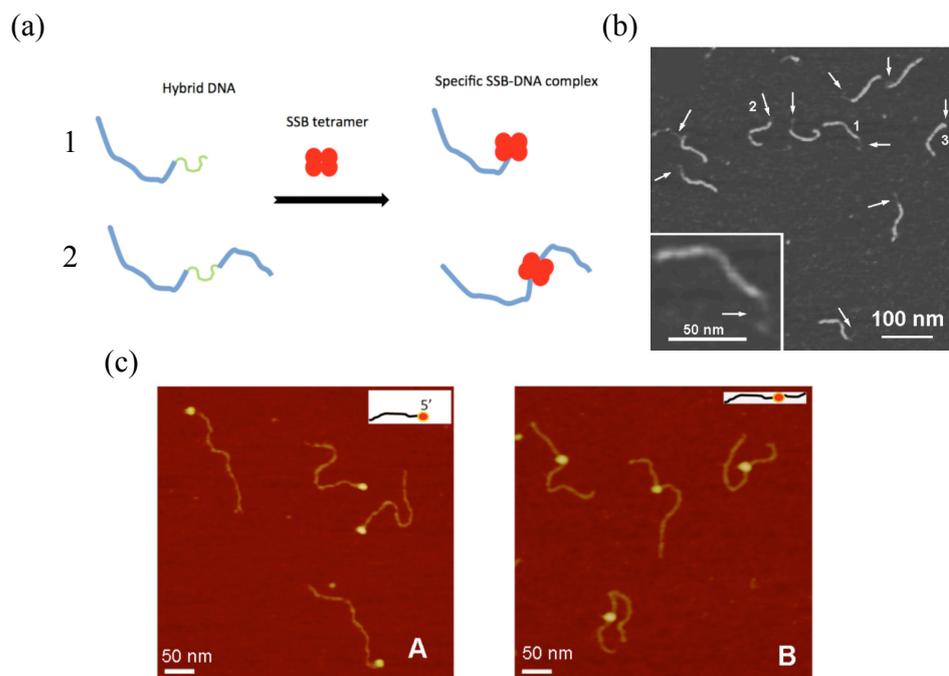


Figure 2.15: **AFM approach for studying protein-DNA interactions** - (a) Schematic presentation of the formation of the complex between the SSB tetramer and (1) the tail-DNA substrate and (2) the gap-DNA substrate. DNA duplexes are shown as blue lines, ssDNA is depicted as green lines, and SSB is shown as red spheres. (b) AFM image of dsDNA with a region containing ssDNA at its end [115]. (c) AFM images for complexes of SSB with hybrid DNA of different types. (1) and (2) are AFM images of complexes with 5'-tail-DNA and gap-DNA, respectively. The insets show the schematic presentations of the complexes [116].

AFM images clearly show the difference between dsDNA and ssDNA through the difference in height and thickness [Fig. 2.15(b)]. The SSB-DNA interaction changes for the ssDNA region in the middle or at the helix end. If the protein binds specifically to tail DNA, the complex attaches only to one end of the hybrid DNA [Fig. 2.15(c,left)], while specific interactions with gap-DNA lead to the complex formation in which the protein is located inside the design [Fig. 2.15(c,right)].

2.5.4 Study the biological process in real time

Another interesting application of AFM is to power biological processes in real-time. For example, Lyubchenko *et al.*, [101] studied the dynamics of protein-DNA complexes using time-lapse AFM. Through quantitative analysis of the nucleosome core particle AFM images [Fig. 2.16(a)], they found that the wrapping DNA turns around the histone cores were not constant, indicating spontaneous dynamics of nucleosomes. To

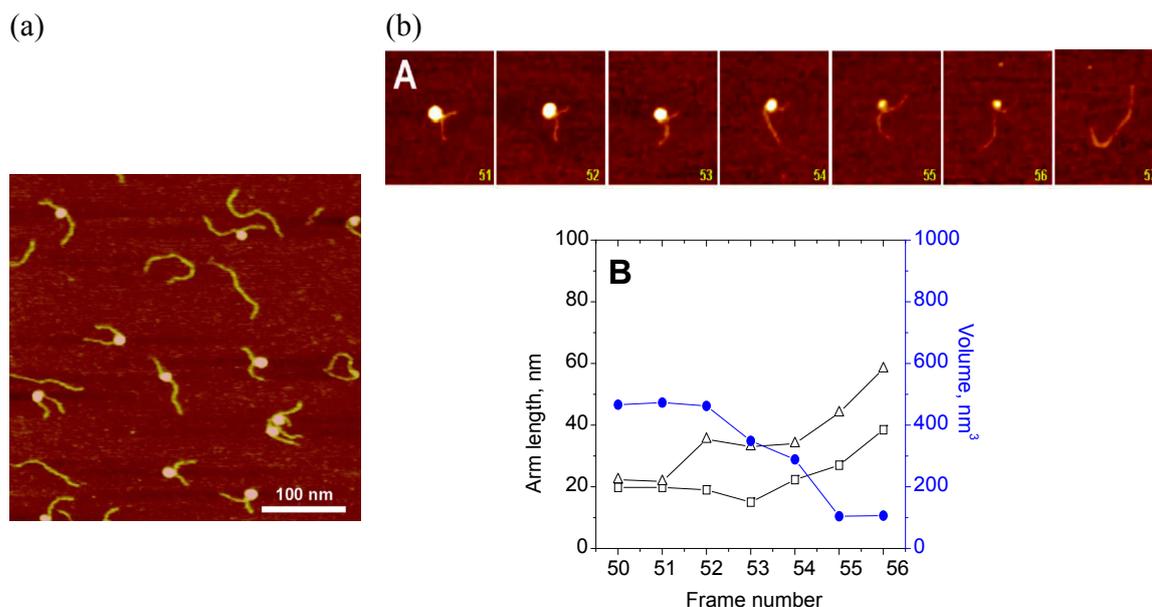


Figure 2.16: **AFM experiments to study the dynamics of the nucleosome particle** - (a) AFM image of nucleosomes in air. The image represents nucleosomes with different amounts of DNA wrapped around the core particle. (b-top) Time-lapse AFM experiments for seven consecutive frames illustrating the dynamics of the nucleosome particle. The scan size is 150 nm. (b-down) The variability of arm length (left y-axis, black) and protein volume (right y-axis, blue) is shown with respect to the imaging frame number (x-axis) [101].

understand the chromatin dynamic feature, they observed the interaction between the nucleosome core particle in which DNA is slightly longer than needed for wrapping around the histone core [Fig. 2.16(b)]. Finally, they observed that the nucleosome core particle dynamics are accompanied by the elongation of both DNA arms and the decrease of the blob size.

2.6 Mechanical properties of DNA

Stretching individual biomolecules can be achieved by a variety of techniques [118] including flow stress, AFM, micro-needles, optical tweezers, and magnetic tweezers. These techniques allow the measurement of forces from 10 femtonewtons to hundreds of piconewtons. Different biological molecules have now been analysed, and the accuracy of these techniques has sufficiently improved so that the theoretical models used to analyse force curves must be refined. In particular, Bustamante *et al.*, [119] have shown that the force-extension diagram of a DNA molecule is well described by a worm-like chain (WLC) model, allowing the theoretical modelling of elastic properties of DNA.

The WLC model was developed in polymer physics and is the more complex model describing ideal polymer chains. Other models describing polymer chains are:

- Free-jointed chain model: The chain is composed by rigid segments. The connection points are ideal, so no energy is needed to change the position of each segment. All segments, positions and rotations are equally probable. A random walk can describe the state of the chain.
- Free-rotating chain model: In this model, the segments are free to rotate, but the angle between segments is fixed, given the first approximation to the chemical bonds inside the molecule.
- Hindered rotation model: assumes that potential energy modulates the torsion angle. This makes the probability of each torsion angle proportional to the Boltzmann constant. A statistical mechanics description of the system is needed.
- Rotational isomeric state model, the allowed torsion angles are determined via the minimal energy state in the rotational energy potential. Bond lengths and bond angles are constant.
- Worm-like chain model: In this model, polymers are not completely flexible; bending them requires energy. This additional criterion defines a new parameter, the persistence length P . At a chain length scale below P , the polymer behaves more or less like a rigid rod.

In particular, the WLC model has been used in several articles to describe elastic properties of DNA, and RNA strands [105, 118, 120–123]. In the WLC model, the polymer is treated as a straight, relatively stiff rod made up of a continuous, homogeneous, isotropic material. The chain elastic behaviour is assumed to be purely entropic, i.e., at finite temperatures, thermal fluctuations induce local curvature in the polymer, making it bend and deviate smoothly from the straight configuration.

An important result of the WLC model is that the average directional correlation between two segments in a polymer decreases exponentially with their separation [118]:

$$\langle \cos(\theta) \rangle = \langle \vec{u}(s) \cdot \vec{u}(s') \rangle = \exp^{-|s'-s|/P} \quad (2.1)$$

where $\vec{u}(s)$ and $\vec{u}(s')$ are the unit vectors tangent to the chain at positions s and s' , respectively. P acts as the decay length through which the memory of the polymer initial orientation persists. The persistence length is a statistical mechanical quantity

whose value depends on the mechanical properties of the chain through its bending rigidity κ , and on the magnitude of the statistical deformations induced in the polymer by the energy of the thermal bath: $P = \kappa/k_B T$ [124].

In particular, the correlation between orientations of different segments vanishes for a far enough segment: after a certain length, known as P , the molecule can bend or twist easily due to thermal fluctuations energy. As a practical example, we can think about the garden hose, which is easy to fold and roll on itself. However, if we cut a short enough segment, it will be difficult or even impossible to bend it. In this thesis, we developed an automatised code in Matlab capable of measuring the molecules persistence length from the AFM images. The procedure is similar to the one followed by Lamour *et al.*, [121]. In this article, they analysed AFM images of the individual amyloid fibril.

The persistence length P of a sample of individual polymeric chains can be obtained via three distinct measures derived from the worm-like chain model (WLC) for semi-flexible polymers:

1. From the angle θ between segments of the same length l [Fig. 2.17(a)].
2. From measuring the end to end distance on the molecules [Fig. 2.17(b)].
3. From the mean square of the deviation δ to secant midpoints, as a function of l [Fig. 2.17(c)].

These three different methods lead to correlations between the length of the segment l and a second parameter having information about the orientation, in particular $\langle \cos(\theta) \rangle$, $\langle R^2 \rangle$, $\langle \delta^2 \rangle$, of each segment as we can see in figure 2.17(d,e,f).

In chapter 4.1.5 we will measure the persistence length for DNA molecules with/without R-loops objects in order to describe how these objects change the mechanical properties of the DNA. Also, we will try to correlate the results with the pathological or physiological behaviour of the R-loop object.

2.7 Our work in this context

How the same R-loop can be either positive regulators or have a negative impact inducing pathological DNA strand remains an open question. Although strong determinants of R-loop formation have been identified, little is known about their architecture. In

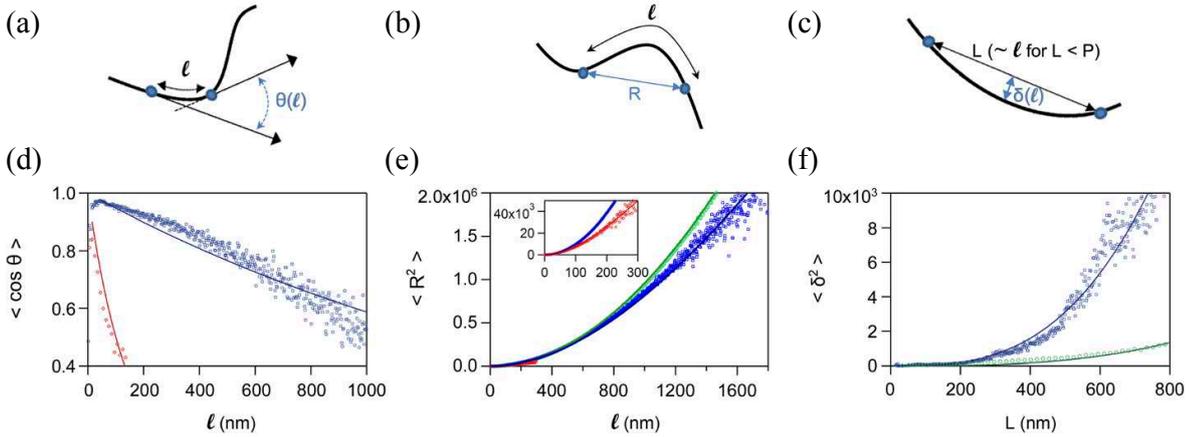


Figure 2.17: **Models to determine polymer chains persistence length** - The data were generated from three different amyloid fibrils, and they are represent in red diamonds, blue squares, and green circles. The persistence length P is obtained via three distinct measures derived from the worm-like chain model (WLC) for semi-flexible polymers: (a,d) Correlation between the angle (θ) and the length of the segments l , (b,e) mean square of the end-to-end distances R , and (c,f) mean square of the deviations δ to secant midpoints, as a function of l [121].

particular, how the single-stranded DNA is organised and established has yet to be fully addressed experimentally. It is conceivable that it could recruit specific proteins and/or form secondary structures, opening the possibility that the R-loop architecture is variable and dependent on various parameters such as size or sequence. In this context, we can ask the following questions:

- Can we visualise an R-loop, can we characterise their morphology and extract their mechanical properties using AFM imaging?
- What is the conformation of R-loops? Can we quantify the amount of DNA contained inside the object and its position in the chain?
- Is the DNA sequence an important parameter in the R-loop structure?
- What is the difference between the positive regulators and pathological R-loops?

In the next chapters of this thesis, I will present the experiments and analyses to answer those questions.

Chapter 3

AFM image analysis to characterise R-loop formation in DNA

Index

Summary	54
3.1 R-loops purification	55
3.2 AFM deposition and imaging	56
3.3 AFM image analysis	57
3.3.1 Algorithm used to extract R-loops properties	57
3.3.2 Isolating DNA molecules	58
3.3.3 Selecting DNA molecules containing R-loop objects	59
3.3.4 DNA molecule skeletonization	61
3.3.5 Obtention of the R-loops properties	63
3.4 Conclusions	65

Summary

This chapter focuses on characterising R-loop formation in transcribed DNA by Atomic Force Microscopy (AFM). We will show in detail the image analysis performed on the AFM images to characterise these structures. The analysis of the images was carried out using Matlab, allowing a semi-automatic treatment. The next chapter will show the results obtained for different conditions, such as the R-loop morphology for two genes capable of generating these structures.

3.1 R-loops purification

The formation, purification and biochemical characterisation of all the samples was carried out *in vitro* by Vincent Vanoosthuyse and his team at the Laboratoire de Biologie et Modélisation de la Cellule (LBMC), ENS in Lyon. We analysed two gene types capable of generating R-loops during the transcription process: the *airn* gene and the *sum3* gene. The DNA that contains these genes are, initially, circular plasmids. The *airn* gene was analysed using the plasmid pUC57 (4050 bp) and the plasmid pFC53 (3991 bp). In contrast, the *sum3* gene was only analysed using plasmid pUC57 (4641 bp) [Fig. 3.1]. In the following, we will explain the main steps to produce and purify the R-loops after DNA transcription *in vitro*.

The transcription process was realized by incubating 1875 ng of circular plasmids, 50 U of T3 RNA polymerase (Promega) in a transcription buffer (40 mM Tris-HCl, pH 7.9, 6 mM MgCl₂, 2 mM spermidine, NaCl 10mM, 20 mM DTT, 0.05% Tween-20) containing 0.25 mM rATP, rCTP, rUTP and rGTP (Promega) for 30 min at 37° C [Fig. 3.2(a)]. The T3 polymerase was then quenched by incubating the reaction at 65° C for 10 minutes [Fig. 3.2(b)]. Then, 750 ng of transcribed plasmids were incubated in the CutSmart buffer (New England Biolabs) containing appropriate restriction enzymes and 2U of RNase H to remove the RNA-DNA hybrid (from *E. coli*, ref. 10786357001 Roche) [Fig. 3.2(c)]. Restriction enzymes identify specific particular sequences and cut DNA at that point. For the plasmid pFC53 we used the enzyme ApaLI, the plasmid pUC57-*airn* with the enzymes KpnI and HindIII, and finally, in the plasmid pUC57-*sum3* have used the enzymes BamHI and XbaI [Fig. 3.1]. After 140 minutes at 37° C, the NaCl concentration was brought to 500 mM, and 0.4 /*mug* of RNase A was added over 20 minutes at 37° C to digest soluble RNA [Fig. 3.2(d)]. Finally, the DNA was purified using chloroform and isopropanol precipitation and resuspended in 10mM

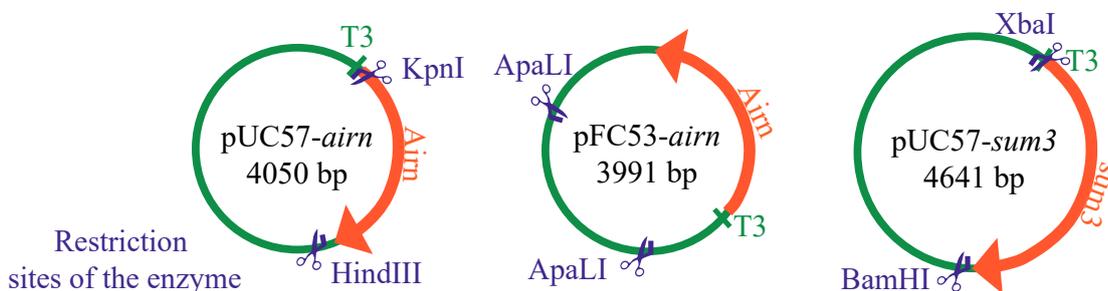


Figure 3.1: **Maps of the plasmids** - pUC57-*airn* was digested with KpnI and HindIII. pFC53-*airn* was digested with ApaLI. pUC57-*sum3* was digested with BamHI and XbaI. The scissors in the diagrams represent the points where the plasmid was cut.

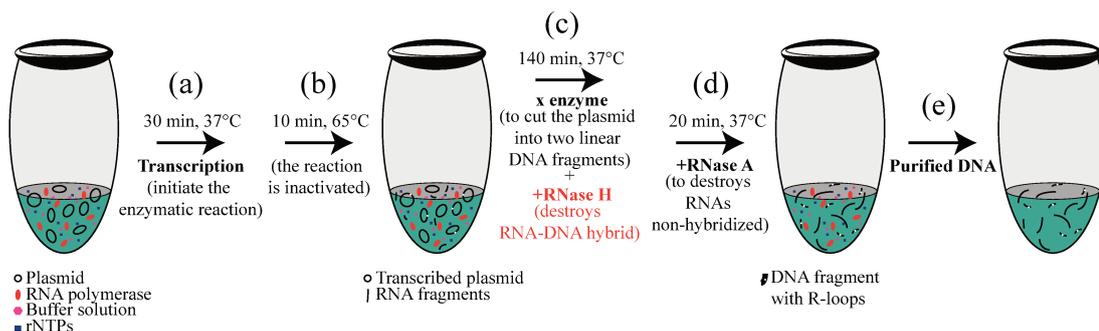


Figure 3.2: **In vitro transcription of R-loops** - (a) Incubation of 1875 ng of circular plasmids, 50 U of T3 RNA polymerase (Promega) in a transcription buffer (40 mM Tris-HCl, pH 7.9, 6 mM MgCl₂, 2 mM spermidine, NaCl 10mM, 20 mM DTT, 0.05% Tween-20) containing 0.25 mM rATP, rCTP, rUTP and rGTP (Promega) for 30 min at 37° C. (b) The T3 polymerase was then quenched by incubating the reaction at 65° C for 10 minutes. (c) 750 ng of transcribed plasmids were incubated in the CutSmart buffer (New England Biolabs) containing restriction enzymes and 2U of RNase H to remove the RNA-DNA hybrid (from *E. coli*, ref. 10786357001 Roche). (d) After 140 minutes at 37° C, the NaCl concentration was brought to 500 mM, and 0.4 μg of RNase A was added over 20 minutes at 37°C to digest soluble RNA. (e) The DNA was purified using chloroform and isopropanol precipitation and resuspended in 10mM Tris.

Tris. DNA was run on 0.8% agarose gels prepared in 1x TBE with no intercalating agent [Fig. 3.2(d)].

3.2 AFM deposition and imaging

It is necessary to observe the R-loops at the individual molecule level to address this study. Since the double helix of the DNA molecule is approximately 2 nm in diameter, we use an AFM, allowing us to obtain images of isolated molecules with a resolution at the nanometer scale.

To dilute the concentration and adhere the DNA molecules to the mica, we diluted 5 ng of purified DNA in TM buffer (10mM Tris, pH 7.4, 5mM MgCl₂). Then, the sample was loaded onto mica for 2 min [Fig. 3.3(a)]. The DNA adhesion was successful since the mica disk was previously cleaved using scotch tape to obtain a charged and flat surface. Then, the sample was washed with 1 ml of ultrapure water and dried gently using a nitrogen flow [Fig. 3.3(b-c)].

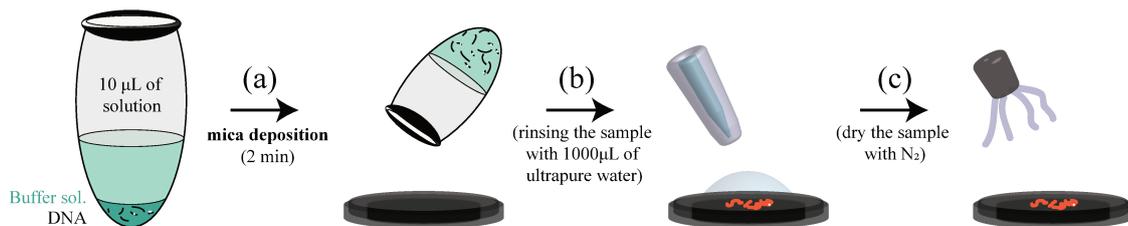


Figure 3.3: **Sample protocol for the AFM** - (a) 5 ng of purified DNA was diluted in TM buffer (10mM Tris, pH 7.4, 5mM MgCl₂). Then, the sample was loaded onto mica for 2 min. After that, the sample was washed with 1 ml of ultrapure water (b) and dried gently using a nitrogen flow (c).

The images have been obtained using Tapping Mode TM in the air. The TM technique maps topography by lightly tapping the surface with an oscillating probe tip, reducing lateral friction forces, and imaging fragile molecules such as DNA. The cantilever used, TESP-300, have a nominal stiffness of 40 N/m with a resonant frequency of ~ 300 kHz. The images are 3 μm wide, a resolution of 512×512 pixels and a scan frequency of 2 Hz.

3.3 AFM image analysis

All images were edited using the NanoScope Analysis program before treating them in Matlab. This program allows us to remove the tilt and low noise frequency in the image, which appears as horizontal shifts or stripes [Fig. 3.4(a)]. The Flatten command eliminates non-thermal-mechanical drift in the piezoelectric scanner. It uses all portions of scan lines to calculate individual least-square fit polynomials for each line and thus obtain a flat image [Fig. 3.4(b)]. Once the image inclination and noise have been eliminated, we analyse the images with the Matlab program.

3.3.1 Algorithm used to extract R-loops properties

We have made a quantitative analysis of the R-loops formation properties in the DNA molecules, working with the automatised program that I developed during my PhD. It is possible to get a complete selection of the DNA molecules and study its properties. The algorithm has relevant four section: **obtain the DNA molecules, selection DNA molecules with R-loops formation, skeletonisation of the DNA molecules** and finally, **obtention of the properties of R-loops formation**.

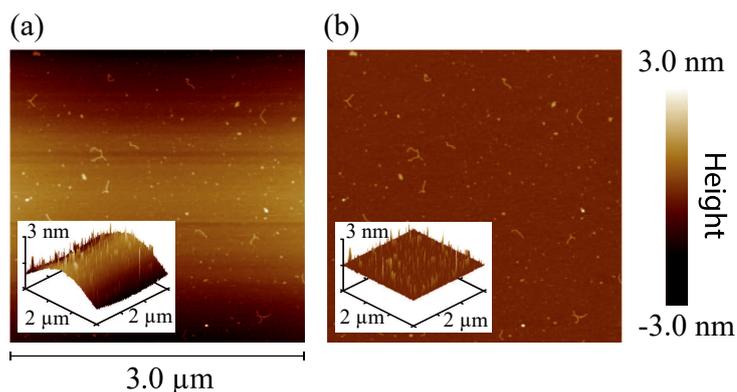


Figure 3.4: **AFM images of DNA** - (a) Original image. (b) Flattened image using NanoScope Analysis program.

3.3.2 Isolating DNA molecules

The first edition in the image is an interpolation; this process allows smoothing the molecule profile and increasing image resolution [Fig. 3.5(b)]. We used the interpolation type "spline". The image size was increased, doubling the number of pixels, optimising the resolution without demanding too much processor memory and not increasing the calculus time. Although this process allows us to have a well-defined DNA molecules profile, it is not enough to isolate the surface DNA molecule. A salt layer is present in the image [126], reducing the edge definition on the molecule. As a consequence, we used a spatial filter (Bitonic filter [125]) to remove the image noise keeping the details of the edge of the structures [Fig. 3.5(c)]. This filter advantage is that the distribution peak (representing the molecules concerning the surface peak) is well-differenced. To establish the threshold value between the molecules and background peak, we used *Otsu's method*¹ [127] of the function *graythresh* of Matlab or fixed manually, depending on the image quality [Fig. 3.5(d)]. It performed an area filter using the function *bwareaopen*. This function allows us to select structures with different areas, so we only studied structures with a minimum area of 2000 nm² and maximum area 25000 nm² [Fig. 3.5(e)]. Since we only want to study free and undisturbed molecules, it is necessary to remove those touching the image edge because they are not complete. Besides, some impurities are detected (it is possible to see them in the images 3.5 (a), (b) and (c) as yellow dots [Fig. 3.5]) which have a greater intensity than the molecules of DNA. These impurities and any molecule that are in contact

¹Otsu's thresholding method involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels on each side of the threshold, i.e., the pixels that either fall in foreground or background. The aim is to find the threshold value where the sum of foreground and background spreads is at its minimum.

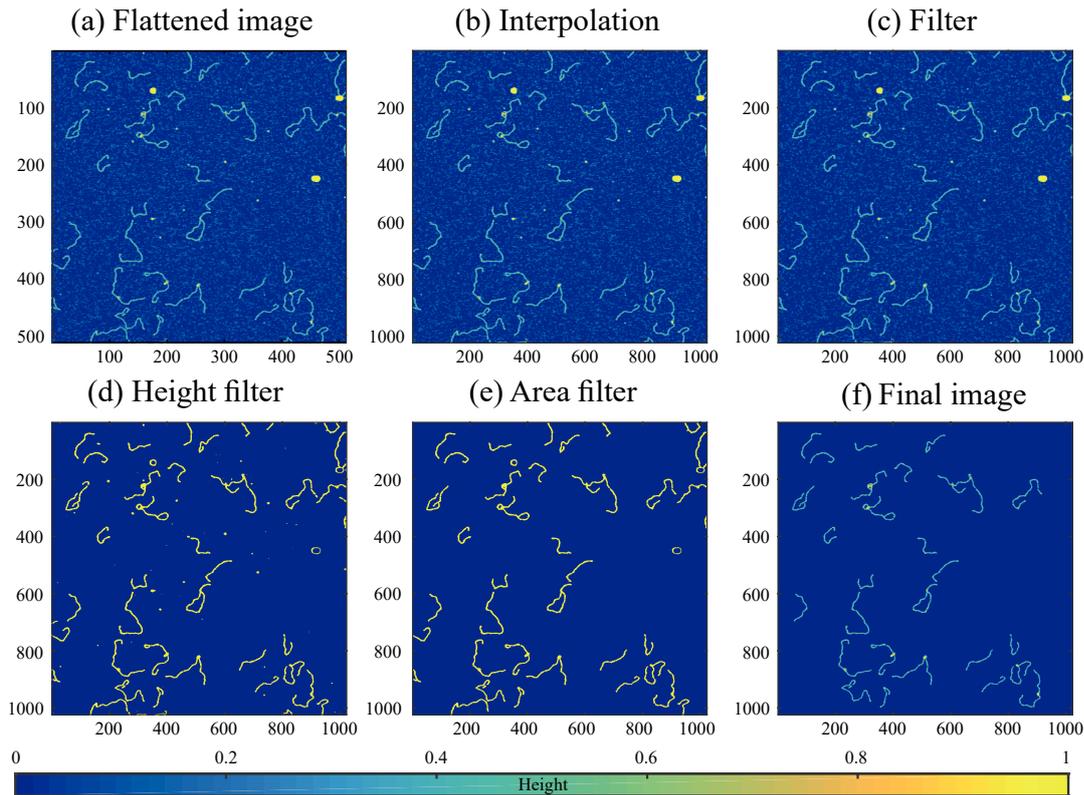


Figure 3.5: **Treatment to extract the DNA molecules in Matlab** - (a) Flatted image (b) Interpolation of the flattened image, where we double the number of pixels. (c) A spatial filter was applied, reducing the image noise and establish a great difference between the molecules and background in the image [125]. (d) Height filter of ~ 0.2 nm applied to select the DNA molecules. (e) Area filter applied to remove structures smaller than 2000 nm^2 or bigger than 25000 nm^2 (for short Matlab, the minimal and maximal area is 1000 nm^2 and 10000 nm^2 , respectively). (f) We remove the molecules that are touching the edge of the image. Then, the binary image is multiplied by the original morphological image.

with it were removed. When the molecules are finally isolated, the binary image was multiplied by the original morphological image ([Fig. 3.5(b)]), resulting in an image that has the topography of the original image [Fig. 3.5(f)], but does not have noise and/or impurities.

3.3.3 Selecting DNA molecules containing R-loop objects

As we can see in the final image that has been previously treated [Fig. 3.5(f)], the DNA molecules contain objects (in yellow) with a height greater than DNA. These objects are identified as R-loop. To select the DNA molecules with R-loop, we coupled a height and area filter. We look at the intensity distribution of each molecule [Fig. 3.6(a)]. In

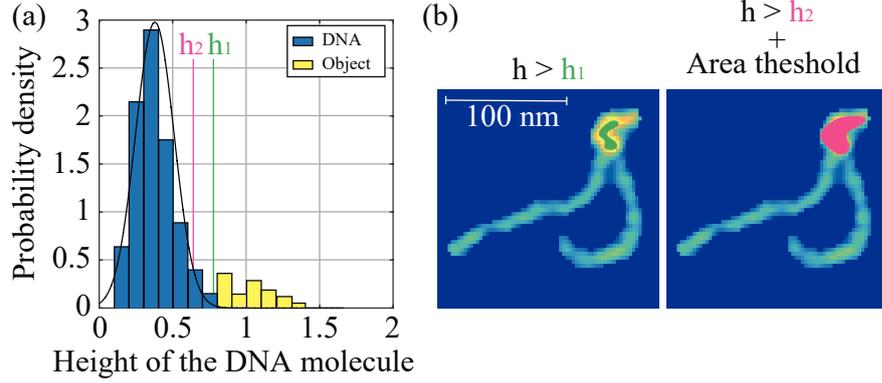


Figure 3.6: **Selection DNA molecules with R-loops formation** - (a) A DNA molecule without R-loops is described by Normal distributions, and the binomial distributions describe molecules with R-loop objects, where the first and second peak represent the height of the DNA and R-loop object height, respectively. (b) Example of a molecule containing an R-loop object. The green (left) and pink (right) region represent all pixels with height greater than h_1 and h_2 , respectively. An area threshold is applied to remove small objects and noise.

the first place, the DNA molecules distribution without objects is a normal distribution, and when the DNA molecule has objects, a binomial distribution is observed. In order to unambiguously identify R-loop, a series of thresholds was applied.

First, we used a hysteresis thresholding protocol: the height distribution was first fitted with a single Gaussian function to obtain $hg = mg + 2sg$, where mg and sg are respectively the mean and standard deviation of the Gaussian fit. Every pixel with height $h > hg$ was then removed from the height distribution, and a new Gaussian function was fitted to the new height distribution (mean value m and standard deviation) [black line of Fig. 3.6(a)]. A first height threshold ($h_1 = m + 3s$) was applied, and only pixels with $h > h_1$ were selected [green object Fig. 3.6(b)]. To identify the whole object, pixels linked to those objects with height $h > h_2$ ($h_2 = m + 2s$) were also considered as part of the object [pink object Fig. 3.6(b)].

We only considered R-loops whose minimal area was more than 150 nm^2 . This threshold was determined to strictly discriminate real R-loop objects from the height fluctuations observed on DNA molecules adsorbed on a flat surface.

In some cases, we noticed that DNA molecules contain two R-loop objects. The analysis of the R-loop object position in the DNA molecule (later chapter) shows that these two objects are located in two characteristic positions of the R-loop in the plasmid configuration. However, rare cases are visualised ($\approx 1\%$ of molecules), so it is impossible to perform a quantitative analysis. Therefore, DNA molecules with two

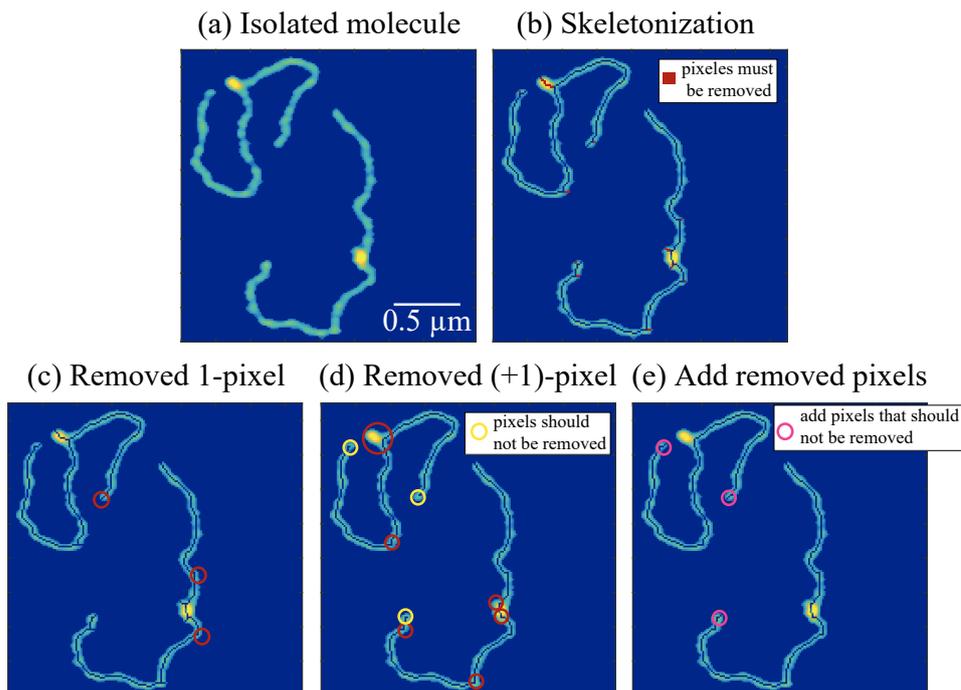


Figure 3.7: **Treatment to obtain the skeletonisation of the DNA molecules in Matlab** - (a) Original DNA molecules. (b) Skeleton of the molecule with spines (red colour). (c) Spines of 1-pixel removed (this process is performed because, on some occasion, the skeleton is perfect except for 1-pixel spines). The red circles indicate the place where 1-pixel spines were removed. (d) We remove spines with a length greater than 1 pixel. This process generates a reduction of the total length of the skeleton (yellow circles). (e) Spines connected only to the ends of the skeleton are relocated to correct the previous operation.

R-loop objects are excluded from the analysis.

The type of object was assigned manually [see Fig.3.9 and section 3.3.5]. By comparing all the AFM experiments results, we established that we could not estimate the relative proportion of each type of objects with better accuracy than 10% (standard deviation). For each selected R-loop object, the topographic information extracted from the image allowed us to measure its height, area, and volume precisely [see section 3.3.5].

3.3.4 DNA molecule skeletonization

To measure the DNA length and/or the R-loop object position, we use the DNA molecule skeleton. The skeletonisation process is characterised mainly by recursive and conditional operations instead of morphological operations such as dilation, ero-

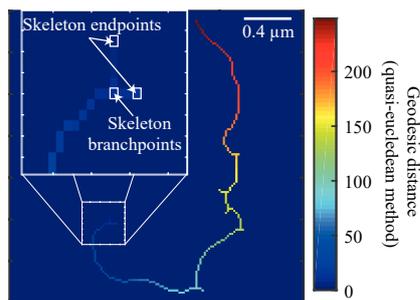


Figure 3.8: **Representation geodesic distance in a matrix-** The matrix represent geodesic distance of the skeleton between a reference pixel (endpoints) and the others.

sion, opening, and closing. These morphological operations reduce the total molecule skeleton length and make it difficult to obtain a one-pixel skeletonisation width, resulting in problems obtaining the total skeleton length. The steps we used to optimise the skeletonisation process are described here:

- **Skeletonization:** the skeletonization process is developed using the *bwmorph* function with the “*skel*” operation in Matlab. The problem to perform this operation is that the skeleton is not perfect; spines are created in the process [Fig. 3.7(b)].
- **Remove 1-pixel and more than 1-pixel spines:** to identify skeleton spines, we need to find the skeleton end-points and branch-points. Then we used the “*bwmorph*” function with the “*endpoints*” and “*branchpoints*” operations. We measured the distances from each end-points to each branch-points using the “*bwdistgeodesic*” function with the “*quasi-euclidean*” method [Fig. 3.8].

Then, when the distance between the endpoint and branchpoint is 1 pixel, the spine is removed. We perform this operation first because, on some occasions, the skeletonisation of the molecule is almost perfect except for 1-pixel spines. At the end of this operation, we verified that the skeleton is perfect, and then the molecule is free of the next edition. Otherwise, we repeat this operation but for spines with a length higher than 1 pixel.

- **Add deleted longer spines to the skeleton:** when spines with a length greater than 1 pixel were removed, one pixel at each end of the skeleton is also removed [Fig. 3.7(d)] causing the skeleton length to be less than the real length of the molecule. Therefore, we must do the inverse process and add the spines in the skeleton endpoints to recover the original length [Fig. 3.7(e)]. After, we verify that the skeleton has only two endpoints and is free of spines. If the skeleton still has spines, the step is repeated.

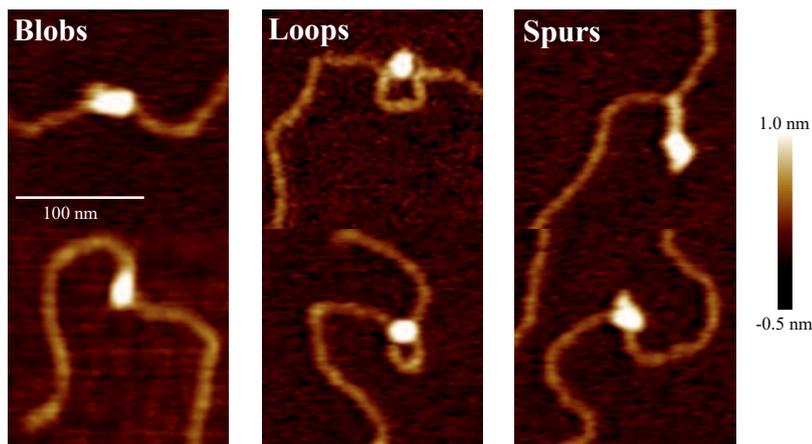


Figure 3.9: **DNA molecules types** - Images obtained with AFM. Three types of R-loop objects are identified: Blobs (left). Loops (centre) and Spurs (right) objects.

We should mention that the skeletonisation process is complicated because of the different form or size of the R-loop objects and how they are combined with the DNA molecules. Consequently, we developed seven types of skeletonisation to cover all types of molecules. However, the strategy explained before was successful in obtaining a perfect DNA skeleton. We identified three types of molecular shapes: *blobs*, *spurs* and *loops*. Blobs represent R-loop objects aligned on the DNA molecule in the main axis, while spurs come away from this axis; loops correspond to objects formed when a blob is located at the base of a loop of DNA regular height [Fig. 3.9].

3.3.5 Obtention of the R-loops properties

The ideal case to analyse the molecules is when no perturbations are present in the DNA.

The script identifies the free molecules when the skeleton has only two endpoints and when the DNA is not coiling and touching it. In any other case, the molecules were excluded from our analysis. The analysis could be presented in three main steps:

- Distance: The skeleton length is measured using the quasi-Euclidean distance but excluding the skeleton part containing the objects [Fig. 3.10(a), white line on the object]. It is also possible to calculate the distances to the nearest and the furthest DNA end using the skeleton length and the object position [Fig. 3.10(b), purple and orange line, respectively].
- Angle: The DNA skeleton was fitted with a straight line up to 5-pixel contour

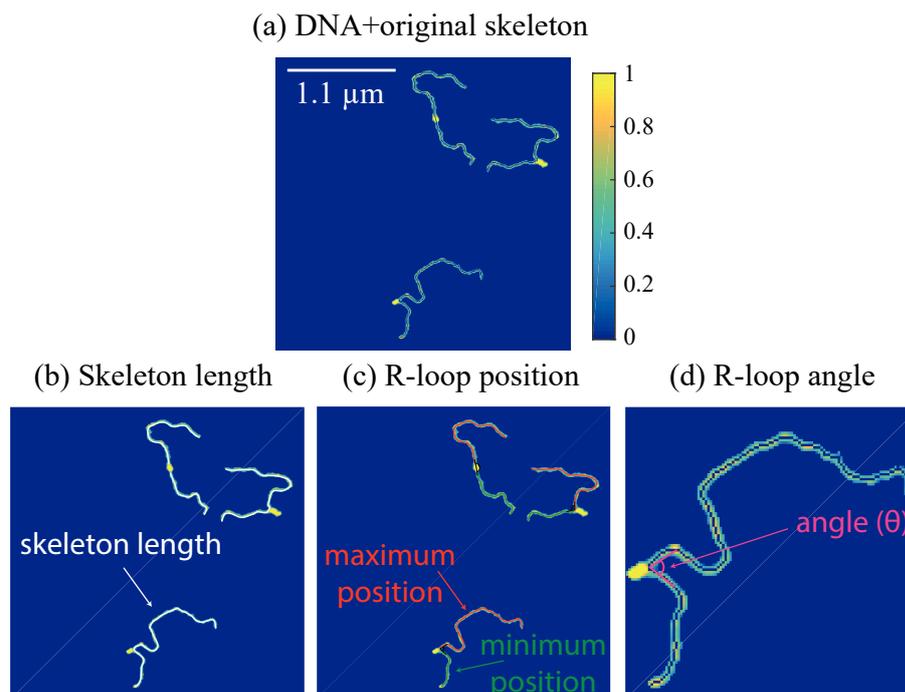


Figure 3.10: **Properties obtained from the DNA that contains R-loop objects -** (a) Image show the DNA with R-loop objects and their skeleton in black. From the DNA skelonzation we obtained (b) the length of the DNA skeleton, (c) the position of the R-loop object on the DNA and (d) the angle that this object generates in the DNA.

length distance (15 nm) emanating from a single point on the edge of the object (maximum height value along with the skeleton object). This determines the angle formed between the DNA strands entering and exiting an object. The angle θ between the two lines was then measured using the Al-Kashi formula: $c^2 = a^2 + b^2 + 2ab \cos(a)$, where a and b are the 2 DNA arm lengths and c the length of the triangle third side.

- **Morphology:** The average of 5 angles measured, along with one molecule skeleton at random positions, was considered for each DNA molecule containing no R-loop objects (not transcribed sample) to obtain the reference angle. The morphological properties were measured directly from the R-loop object structure, such as the area (numbers of pixels size) and the volume (adding the number total of pixels; each pixel has a number associated with the height). Other properties were measured; for example, the R-loop is the maximum height and the angle that the R-loop generates in the molecules [Fig. 3.10(d)].

3.4 Conclusions

Here, we developed an algorithm in Matlab that obtains the DNA molecule skeleton length and selects the DNA molecule R-loop objects. Additionally, we determine the position in which the R-loop object is inserted into the DNA molecule. Three types of conformations were identified, which we called blobs, spurs, and loops. Next, we will apply this analysis to characterise the R-loop morphology, its effect on DNA and compare different conformations that these structures can generate.

Chapter 4

Quantitative study of R-loops formation in DNA using AFM

Index

Summary	66
4.1 <i>Airn</i> gene	67
4.1.1 R-loops visualisation	67
4.1.2 R-loops quantification	68
4.1.3 R-loops object morphology	70
4.1.4 R-loops object position on the DNA template	71
4.1.5 R-loop objects generate local bends in the DNA	72
4.2 R-loop do not result from G-quadruplex	76
4.3 The mutation affects objects but not hybrid RNA-DNA formation	79
4.4 New R-loop object position	79
4.5 <i>Sum3</i> gene	82
4.6 Conclusions	84

Summary

We quantitatively characterise the R-loops objects formed in the *airn* and *sum3* model genes through the image analysis. We compare the morphological characteristics of the R-loops formed for different DNA constructions of the *airn* genes template. The

main properties that we studied are the DNA length in the R-loop presence, the R-loop position over DNA and the fold in the DNA molecule generated by this structure. In this chapter, the results are shown in the following order: (i) We analyse the R-loops formed on the *airn* gene for the mentioned parameters. (ii) We verify that the studied objects are not due to G-quadruplex conformations. (iii) We probe conformational changes in the R-loop object caused by mutations in the *airn* gene. (iv) Finally, we compared the *sum3* gene, which can generate R-loop during the transcription process, but no objects.

4.1 *Airn* gene

As mentioned in the previous section, we studied two genes capable of generating R-loops during the transcription process: the *airn* and *sum3* genes. As seen by biochemical methods. In the following section, we show the analysis for the *airn* gene.

4.1.1 R-loops visualisation

The *airn* gene was encoded in two plasmids, the plasmids pUC57 and pFC53, cut to produce two fragments of difference size, one containing the *airn* gene, where the R-loop is produced, and a backbone DNA. For the two constructions, the R-loop object is in the short (pUC57) / long (pFC53) fragments, whereas the backbone DNA control fragment is long (pUC57) / short (pFc53). Figure 4.1(a) and 4.1(b) shows the AFM images of the two plasmids, respectively. In an AFM image, the colour represents a height difference in the sample. Therefore the position of the R-loop object is easily seen as the bright part of the molecule. We study the results obtained with the plasmid pUC57 since this configuration generated many objects displayed in the AFM images and then allowed us to develop a more effective analysis. The pUC57 plasmid configuration is made up of two fragments, the short fragment of DNA associated with the encoded region of the *airn* and the long fragment associated with the plasmid backbone. The latter used as a control reference for the same sample.

The analysis is carried out in three different configurations, (i) Not transcribed, (ii) transcribed and (iii) transcribed and treated with the RNase H enzyme. Sample (i) [Fig 4.1(a)] visualises homogeneous DNA molecules, that is, without height differences on each molecule. In this case, the DNA molecules were not transcribed, and we used them as a reference. Besides, it is possible to visualise two sizes, long and short fragments, as expected. The AFM image of the sample (ii) Figure 4.1(b) shows that DNA molecules

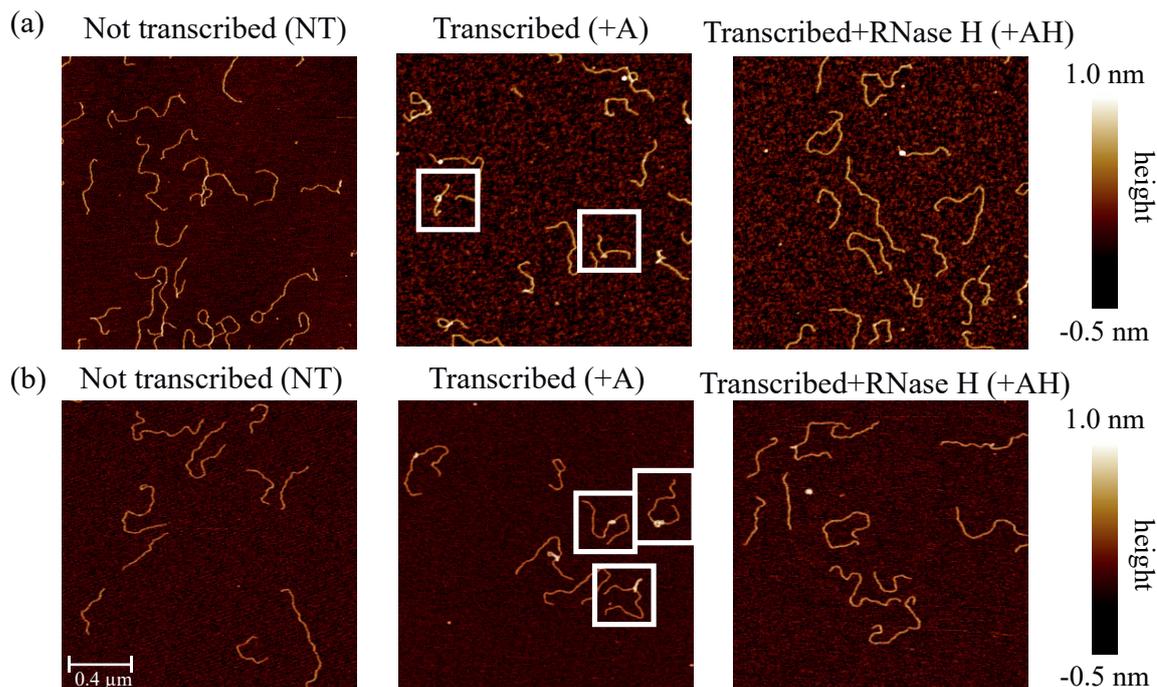


Figure 4.1: *Airn*: **AFM images** - AFM images for not transcribed, transcribed and transcribed+AH cases. (a) pUC57-*airn* plasmid and (b) pFC53-*airn* plasmid. The white squares show some DNA molecules containing R-loop-objects. The colour difference represents a height difference of the sample.

have a specific area and a height greater than the mean DNA (displayed as a colour difference in the image). Some of those molecules were marked in white. Finally, in the sample (iii), the DNA molecules were transcribed, but an enzyme capable of digesting the RNA-DNA hybrid was incorporated. The R-loop object size is reduced or and we noticed that there are fewer DNA molecules with R-loop objects.

4.1.2 R-loops quantification

Once we have identified the R-loops objects in the AFM images, we perform a thorough analysis of these structures. The appearance quantification of these objects is shown in figure 4.2(a). Higher regions relative to DNA height are observed when DNA molecules have not been transcribed, 10% approximately. This difference could be related to small fluctuations when the sample is being scanned, generating noise artefacts smaller than the R-loops size. In the second case, when the sample is transcribed, around 80% of the fragments contain R-loops. A decrease in R-loops frequency, 20% of the fragments, occurs when the RNase H enzyme is present. From this point, the R-loops visualised

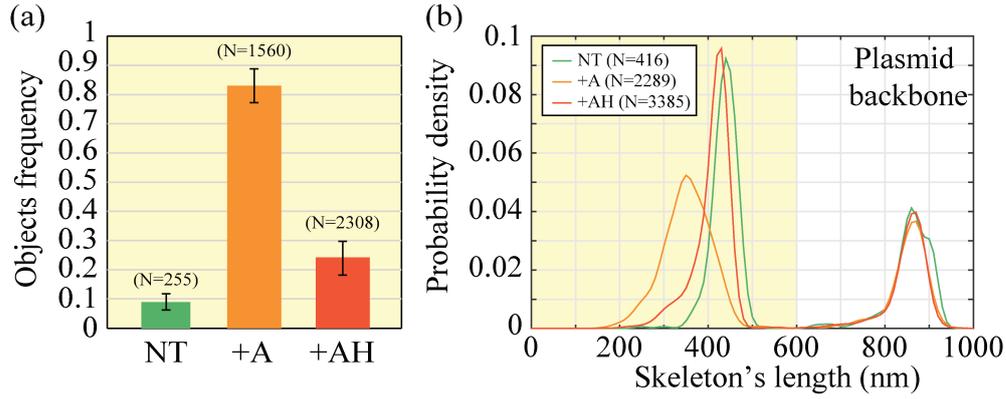


Figure 4.2: **pUC57-*airn*: skeleton length for all molecules** - (a) Object frequency in the DNA molecules for not transcribed (NT), transcribed (+)A and transcribed+AH (+AH) cases ($n = 4$ independent experiments; error bars represent the standard deviation). In pUC57-*airn* case, *airn* gene is present on the short fragment. (b) Probability density of the DNA skeleton length of all the molecules analysed (both long and short fragments) in the indicated conditions (with and without objects).

in the AFM images are called "R-loop objects".

According to the analysis protocol described above, we accurately measure the skeleton length of each DNA molecule, both long and short fragments, with and without R-loop objects [Fig. 4.2(b)]. The long fragment length is used as a control parameter between the three conditions, untranscribed, transcribed and transcribed with the enzyme RNase H (green, orange and red lines, respectively). We can see the long fragments distribution (or the main plasmid backbone) are overlap, and it is consistent with the fact that no objects should be observed on this part of the plasmid. Additionally, the short fragment, specifically for the untranscribed sample, is used to determine the conversion factor between nm and bp units. The length of the short fragment measured using the AFM images is 440 nm, where the value obtained is equivalent to 1375 bp; therefore, 1 bp is 0.32 nm approximately. If we look at the transcribed short fragments, there is a shortening in the skeleton length (orange line), although this length tends to recover if the sample has the enzyme that eliminates the RNA-DNA hybrid (red line).

Among the 1283 DNA molecules containing R-loop objects, we identified three configurations of these objects on the DNA [Figure 3.9]. Structures that have their centre of mass on the skeleton of DNA molecules are called "Blobs". On the contrary, structures that have their centre of mass outside the skeleton of DNA molecules are called "Spurs". Finally, structures that present a loop with DNA are called "Loops". We found that 45% of the objects are spurs objects, followed by blobs objects, with 30%, and finally, loops objects with 25% [Fig 4.3(a)]. The objects spurs and blobs

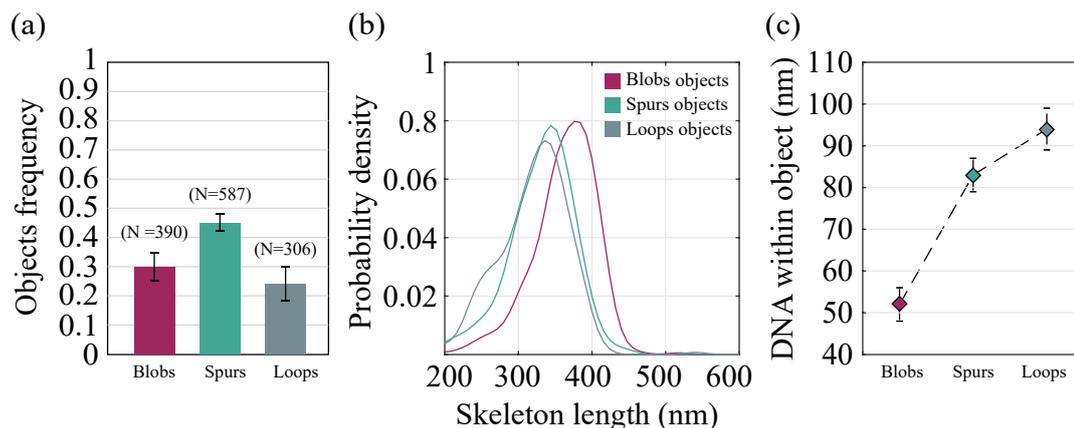


Figure 4.3: **pUC57-airn: R-loops objects types** - (a) Probability of the different types of R-loop objects ($n = 4$ independent experiments; error bars represent the standard deviation). (b) Probability density for blob, spur, and loops objects. (c) Estimated amount of DNA caught in each type of R-loop objects (mean \pm standard error).

present a more significant reduction in the DNA length [Fig 4.3(b)]. As mentioned in the previous section, the total DNA length is mainly calculated as the sum of the arms from the point closest to the object centre of mass. However, there is a section of the DNA in loops that we did not consider in the calculation of the DNA total length, that is, the DNA loop roll over the object [Fig 3.9]. Hence, DNA reduction can be seen as the amount of DNA in the R-loop object. Figure 4.3(c) shows the average and standard deviation values of DNA length contained in the object, where the amount of DNA in blobs is ~ 50 nm, smaller than the other two types of objects, the amount of DNA contained in loops is ~ 95 nm higher than the amount contained in spurs ~ 82 nm.

4.1.3 R-loops object morphology

Using the R-loop objects morphology analysis, we studied the correlation between the R-loop object volume and skeleton length to understand the length reduction of the DNA when the object is present. Figure 4.4 shows the correlations on a 2D probability density graph and his respective 1D projections. We noticed that higher volume R-loop objects generate a reduction in skeleton length, establishing that the larger the object, the more DNA is contained. On the other hand, observing the treated sample with the RNase H enzyme, the population or section with the most significant probability presents a smaller volume than the population of the transcribed sample (without the RNase H enzyme), consistent with the peak of the skeleton length distribution being

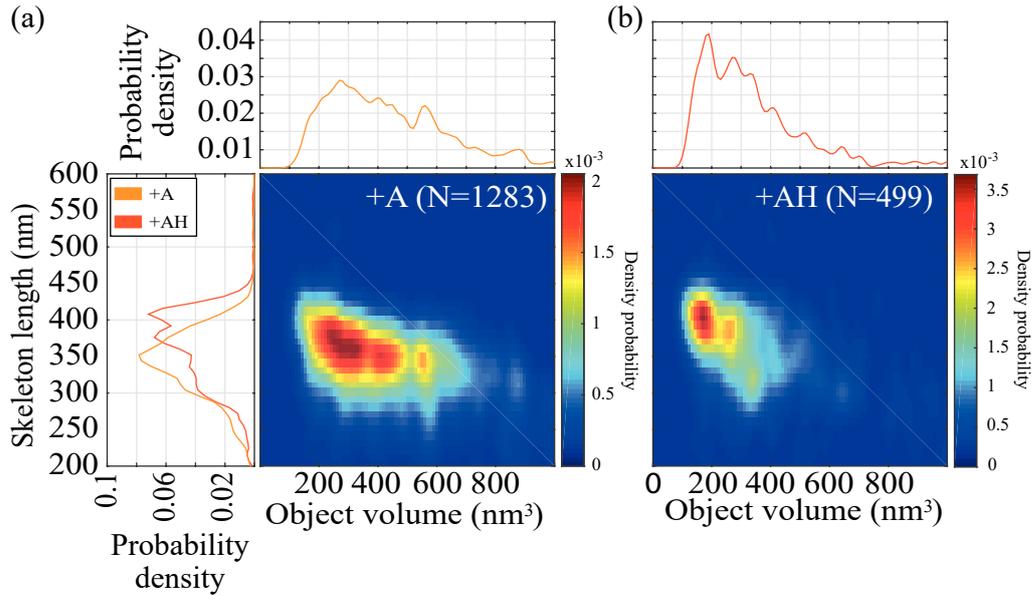


Figure 4.4: **pUC57-*airn*: correlation between R-loop objects and skeleton length** - (a) 2D probability density plots showing the correlation between the skeleton length and the objects volume for +A sample. The 1D probability density plots of the skeleton length and object volume are the respective projections of the 2D probability density. (b) Same as (a) but for the +AH sample.

close to the peak of the non-transcribed sample.

To understand the relationship between the DNA reduction and the R-loop object type, we analysed the correlation between the DNA length and the object volume [Fig 4.5]. We found that the average value of the population, concerning the object volume, is lower than the other two types of objects, generating molecules with blob objects showing less DNA length reduction. If we compare the objects spurs and loops, we see that the volumes of the objects are different, although the reduction in the DNA length is the same, this is consistent with the fact that in the case of loops, they contain a loop of naked DNA of regular height also.

4.1.4 R-loops object position on the DNA template

The skeletonisation of the DNA molecules allowed us to calculate each molecule length and the R-loop object position over the DNA. We defined two parameters to represent the position: the distance to the (i)nearest and (ii)furthest end of the DNA. We showed that the position of the object is not symmetric [Fig 4.6(a)].

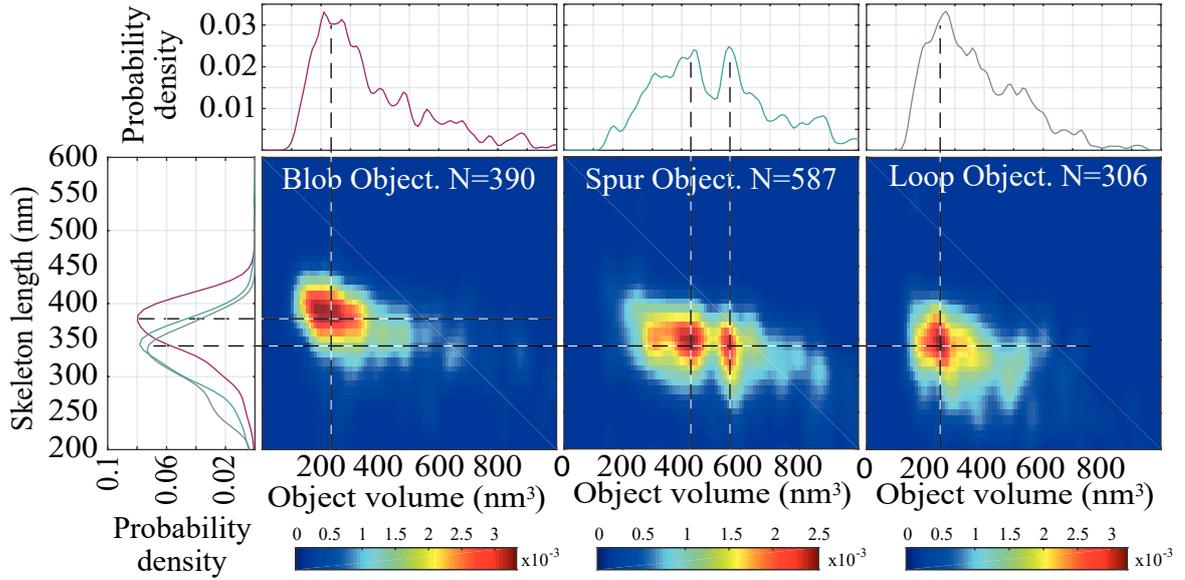


Figure 4.5: **pUC57-*airn*: correlation between the R-loop objects serotypes and skeleton length** - 2D probability density plots showing the correlation between the skeleton length and the objects volume for each category of R-loop objects. The respective 1D projections of each probability density function are also shown.

The R-loops object position is compared with the position obtained using single-molecule R-loop footprinting, performed by F. Chedin [128]. Even though the AFM does not allow to identify of the DNA orientation, the data sets have a better agreement when it is assumed that the distance to the nearest end represents 5' of the *airn* [Fig 4.6(b)]. Analysing the position of the R-loop object, differentiating the type of object, we found that the distance to the nearest end is the same for the three objects, but not for the distance to the furthest 3' end [Fig 4.6(c)]. This observation is consistent with the idea that the starting position is conserved between objects, while their end position varies depending on their size.

4.1.5 R-loop objects generate local bends in the DNA

R-loop objects generate a marked angle in the DNA molecule, suggesting that these objects mechanically restrict circulating DNA [Fig. 4.7]. We noticed that spurs objects generate a larger marked angle than blob objects, implying that the internal organisation differs between the two R-loop forms. To formalise these observations, we have determined the persistence length in the absence/presence of the R-loop objects. The persistence length is a quantity that allows measuring the flexibility of a polymer submitted to thermal fluctuations. In other words, the elasticity of double-stranded DNA

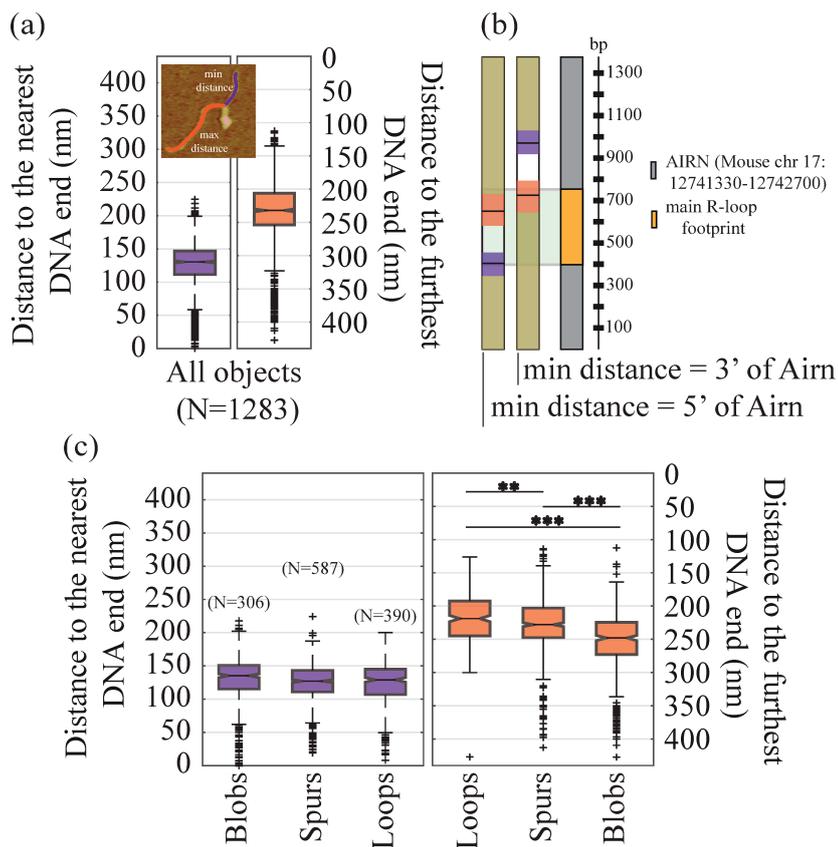


Figure 4.6: **pUC57-*airn*: position of the R-loops objects** - (a) Distance to the nearest DNA end (purple) and to the furthest DNA end (orange) measured for all objects. (b) The position of R-loop footprints was compared to the positions of R-loop objects detected by AFM. The overlap between the two sets of measurements is best when one assumes that the nearest DNA end represents the 5' of *airn*. (c) Distance to the nearest DNA end (purple) and the furthest DNA end (orange) measured for each R-loop object and represented as boxplots. The notches in the boxplots indicate the 5% uncertainty mark around the median. P-values were determined using the Wilcoxon Mann Whitney statistical test (***) $P < 0.001$; (**) $P < 0.01$).

can be described by its persistence length. Therefore, determining these mechanical parameters can be fundamental to understanding the interaction between R-loops and DNA molecules.

In particular, we used the method described in [Fig. 2.17(a)] to obtain the persistence length from our AFM images. The molecule is divided into several segments of the same length, and we measured the angles between all possible segments permutations for lengths varying between 3% and 25% of the total molecule length. We performed the measurements over long and short molecules; long molecules have a length between 800 and 950 nm, while short ones have lengths between 300 and 600 nm. The measurements were carried out in molecules with and without R-loops ob-

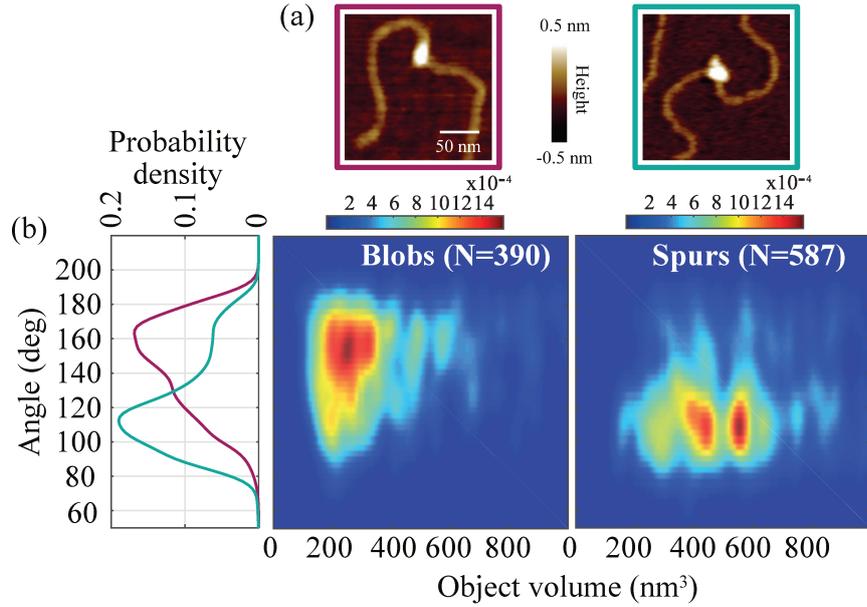


Figure 4.7: **pUC57-airn: R-loop formation imposes mechanical constraints on the surrounding DNA** - (a) AFM images of molecules identified as blob (pink) and spur (green) objects. (b) 2D probability density plots showing the correlation between the angle and the object volume for the two types of the object mentioned. The angle formed between the DNA strand going into the object and the DNA strand coming out of the object was measured. The distribution of angles for molecules without objects was also established.

jects. In figure 4.8(a,c) we show the correlation between the angle $\cos(\theta)$ and segment length l . Theoretically, this relation obeys:

$$\langle \cos(\theta) \rangle = \exp^{-\frac{l}{2P}}, \quad (4.1)$$

where P is the persistence length and θ is the angle between two segments of length l in the chain. Note that this equation is only valid in 2D and for the linear part of the figure 4.8(a,c). To obtain the persistence length, we should determine until which value of l there is a correlation between the variables. To estimate this value, we calculated the ratio of the even moments, i.e., the kurtosis of the distribution of θ angle [120,121]. If the chains are fully equilibrated in 2D, then the θ distribution is Gaussian [121] and in the range where angles θ are still fully correlated, the kurtosis results in:

$$\frac{\langle \theta^4(l) \rangle}{\langle \theta^2(l) \rangle^2} = 3. \quad (4.2)$$

As shown in figure 4.8(b,d), the ratio between the even momentums is not exactly 3. To

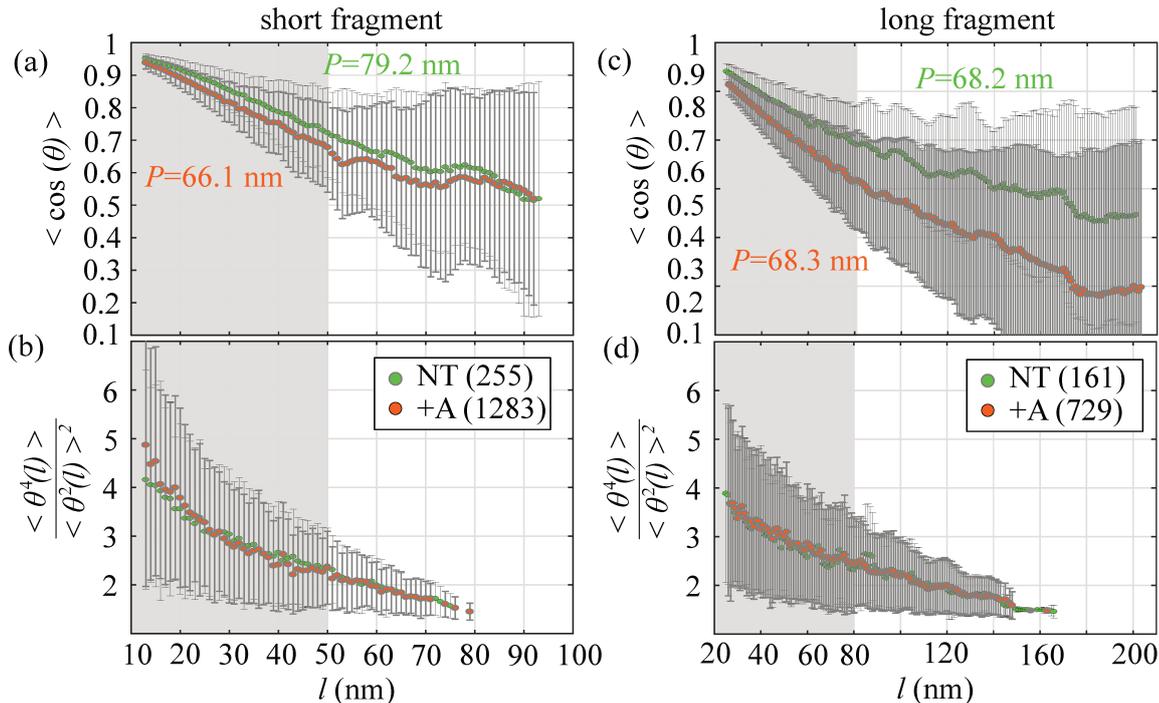


Figure 4.8: **pUC57-airn: DNA persistence length without R-loop object** - (a,c) Orange and green circles represent data for transcribed and not transcribed DNA where 1283 and 255 molecules were analysed for the short fragment and 729 and 161 molecules for a long fragment, respectively. Both graphs show the decay of tangent-tangent correlations, and P was derived from the worm-like chain model to the data. (b) Kurtosis parameter for not transcribed and transcribed DNA where the short fragment of the DNA was analysed. (d) Similar to (b) but for the long fragment of the DNA. The grey region for the $\langle \cos(\theta) \rangle$ and $\frac{\langle \theta^4(l) \rangle}{\langle \theta^2(l) \rangle^2}$ were determined from the median value of the kurtosis, i.e, the expected value 3.

found the cutting point, we minimised the difference between the expected value 3 and the kurtosis medium value as a function of l . The limit for short and long molecules is around 50 nm and 80 nm, respectively. We perform the calculation of the persistence length over not transcribed molecules, used as a reference, and transcribed molecules presenting R-loops. Using the cutting points and the respective fits over the curves, we obtained a persistence length $P = 79.2$ nm and $P = 68.2$ nm for short and long not transcribed molecules. In figure 4.8, the circles and error bars represent the median and the standard deviations of the data.

For transcribed short molecules, we found $P = 66.1$ nm and $P = 68.3$ for transcribed long molecules. A total around of 400 reference molecules were measured, and around 2000 transcribed molecules. The persistence length that we found for the reference molecules agrees with [120, 129] in similar conditions. The measurements show that R-loops objects generate a change in the persistence length for short molecules,

becoming 16% smaller. On the other hand, long molecules have a negligible chance of 0.14% in the persistence length, comparing transcribed and not transcribed ones.

As persistence length is a property independent of the chain length, we expected to find the same value for short and long not transcribed molecules. The difference between them could be explained, like the effect of the environment surrounding the molecules where long molecules could present a lack of space and an induced rollover itself, generating a smaller artificial persistence length. Hence, a comparison between long and short molecules result inaccurate. Nevertheless, the persistence length for not transcribed long molecules and transcribed long molecules are the same; this has sense if we noted from the purification process explained in figure 3.1 that long transcribed and not transcribed molecules are the same because the R-loop object always is contained in the short segment. The difference found between short molecules could be attributed to the R-loops object, but a more detailed analysis should describe the mechanism related to P shortening.

4.2 R-loop do not result from G-quadruplex

Similar objects to what we studied were precisely visualised for *airn* and characterised through AFM images [113]. However, in this case, the RNA-DNA hybrid loops were generated by *in vitro* transcription of the murine immunoglobulin S γ 3 switch region. The formation of objects was then attributed to the extruded ssDNA assembly into G-quadruplexes (G4s).

Since the G-quadruplexes object stability is affected when the ionic composition of the buffer changes. Replacement of K^{2+} with Na^{2+} and then with Lithium caused a reduction in the percentage of the DNA that contains these structures, indicating the G4s stability decline [113] [114].

We performed experiments in the presence of Li^{2+} to check the objects we visualised are not formations of G-quadruplexes. Then, Vincent Vanoosthuyse performed an *in vitro* transcription in a total absence of NaCl, which was replaced with 40 mM LiCl, to destabilise the formation of G4. This variant of the protocol was carried out in the initial stage of the purification [Fig. 3.2(a)]. Figure 4.9(a) shows the frequency of objects in the presence and absence of Li^{2+} . The formations of objects show no alteration; we count the same amount of DNA molecules with structures for both cases. If we look at the DNA length as a function of the object volume, we see that the object size does not change; therefore, the reduction of the DNA length as a consequence of

these objects is the same [Fig. 4.9(b)]. The position of R-loops objects in DNA is also constant, regardless of the type of the object [Fig. 4.9(c)].

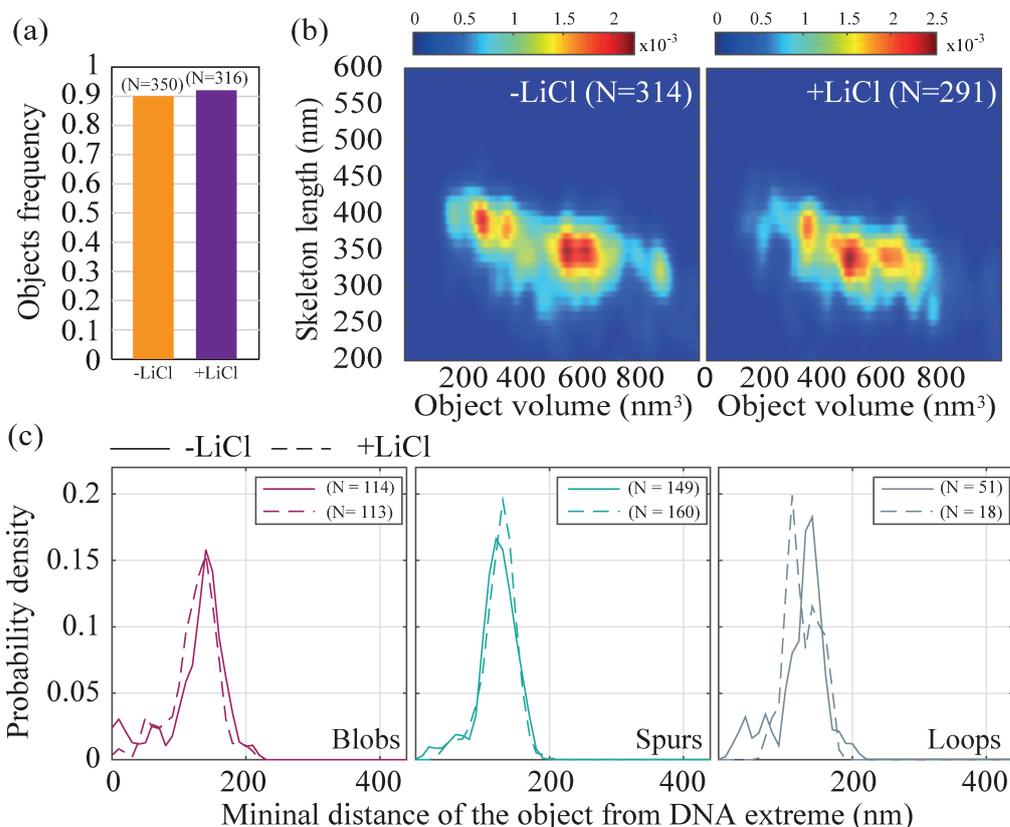


Figure 4.9: **pUC57-*airn*: R-loop do not result from G-quadruplex** - *In vitro*, the circular plasmid pUC57-*airn* was transcribed in the presence (+LiCl) and absence (-LiCl) of 40 mM LiCl before restriction enzymes were used to separate the plasmid backbone from the *airn* template. (a) Probability of R-loop objects. The proportion of the different types of R-loops objects is also indicated. (b) 2D probability density plots showing the correlation between the skeleton length and the object volume in the indicated conditions. (c) Distance between R-loop objects and the nearest DNA end in the indicated conditions and for all objects. (d) The position of R-loop footprints was compared to the positions of R-loop objects detected by AFM.

Additionally, we did this analysis for a different plasmid, resizing the DNA that contains the *airn* gene. As explained previously, the plasmid puC57 contains two fragments, one long and one short, where the short fragment contains the *airn* gene. In this new case, the situation is the opposite; the long fragment contains the *airn* gene, hence the R-loop object. The long fragment contains 2745 bp, where the gene is located between base pairs 542 and 1925. The first difference that we can see is the number of DNA molecules that contain R-loops objects [Fig. 4.10(a)]. However, the addition of Li²⁺ does not impact the frequency of objects, as well as the shape. The correlation

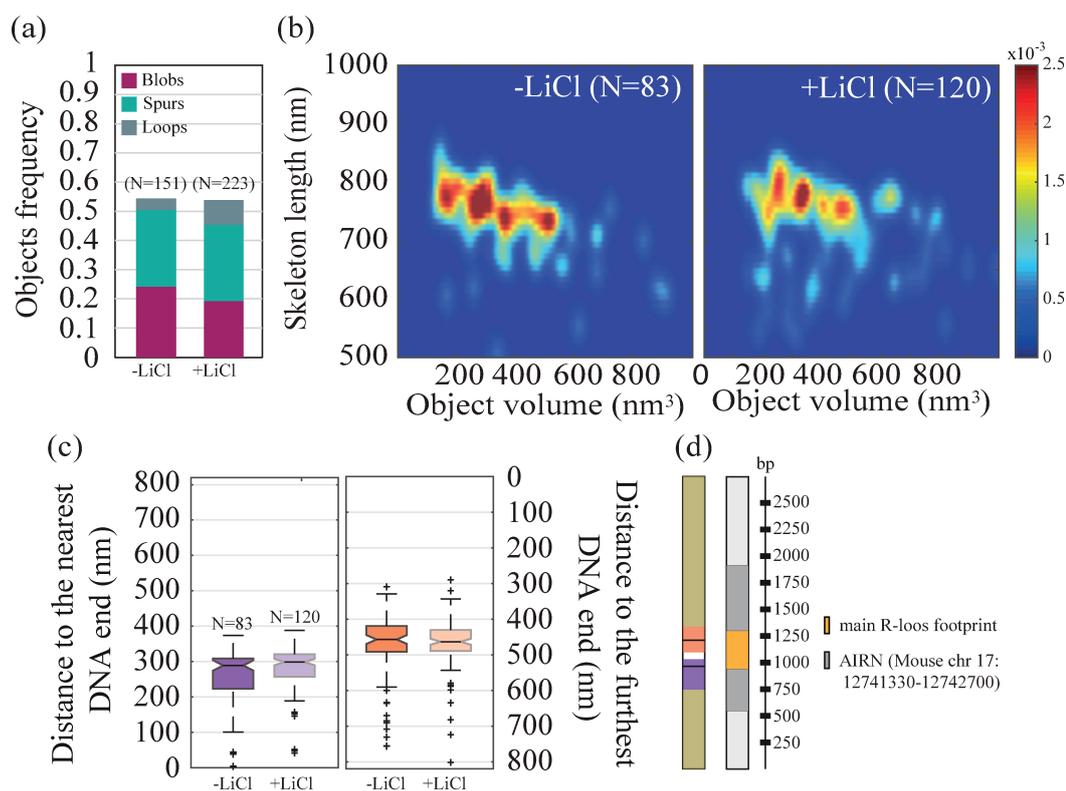


Figure 4.10: **pFC53-*airn*: R-loop do not result from G-quadruplex** - The circular plasmid pUC57-*airn* was transcribed in vitro in the presence or not of 40 mM LiCl, before restriction enzymes were used to separate the plasmid backbone from the *airn* template. (a) Probability of R-loop objects. The proportion of the different types of R-loops objects is also indicated. (b) 2D probability density plots showing the correlation between the skeleton length and the objects volume in the indicated conditions. (c) Distance between R-loop objects and the nearest DNA end in the indicated conditions and for all objects. (d) The position of R-loop footprints was compared to the positions of R-loop objects detected by AFM.

between the length of the DNA molecule and the object volume remains constant in both cases 4.10(b). On the other hand, the R-loop object position is similar; there is no Li^{2+} dependency [Fig. 4.10(c)]. If we compare with the position obtained using R-loops footprints, we see an agreement between both results [Fig. 4.10(d)]. Finally, these results in both *airn* plasmid suggest that the objects that we visualise in the AFM images are not formations of G4s but are the result of formations of lithium-resistant structures in the extruded ssDNA.

4.3 The mutation affects objects but not hybrid RNA-DNA formation

To evaluate the possibility that R-loop formation would allow the G-stretch to form secondary structures by interacting with other parts of the gene through either Watson-Crick or Hoogsteen bonds, this motif was mutated without interfering with the R-loop forming ability of *airn*. Then, the introduction of five G to A mutations in the G stretch [Fig. 4.11(a)] did not affect the amount of DNA:RNA hybrids produced after *in vitro* transcription [see the agarose gel results in [130]]. However, differences related to the R-loop objects are observed. Figure 4.11(b) shows a decrease of at least half of the observed objects when the gene is mutated and declining to 10% when the RNase H enzyme is present. These observations suggest that the stretch of 14 G residues is somehow involved in forming these secondary structures, and small changes in the number of G residuals change the size of the secondary structure formed.

Comparing the correlation between the skeleton length and the object volume for wild-type and G stretch mutant case, we can see a reduction concerning the object size [Fig. 4.11(c)], although the mutant template was still able to produce blobs, spurs and loops object. Alterations in object positions were also observed, mainly for blobs and loops objects; a new position increased its probability of being generated [Fig. 4.11(d)]. The R-loops were distributed mainly into two clusters, with 22% and 88% of the R-loop fingerprint mapping starting at position 100 bp and 400 bp approximately.

Additionally, the new position overlapped with the R-loop footprint minor cluster position seen by the SMRF-seq method [130]. Frédéric Chédin's group carried out the SMRF-seq analysis. This main cluster or position matched well with a region of predicted high R-loop favourability [62]. The new position in cluster 1 can also be seen in the figure 4.6(a). The purple sample boxplot shows the main R-loop initial position, but it is also possible to visualise several points between 0-50 nm corresponding to R-loops generated in cluster 1, nevertheless with a lower frequency.

4.4 New R-loop object position

The pFC53 plasmid fragment that contains the *airn* gene was cut in the region where the R-loops appear more frequently. As a consequence, the only possible region where an R-loop could be generated will be in cluster 1 [yellow region Fig. 4.12(a)]; we will call this construction as *smairn*. The new small fragment length is 584 pb, and the *smairn*

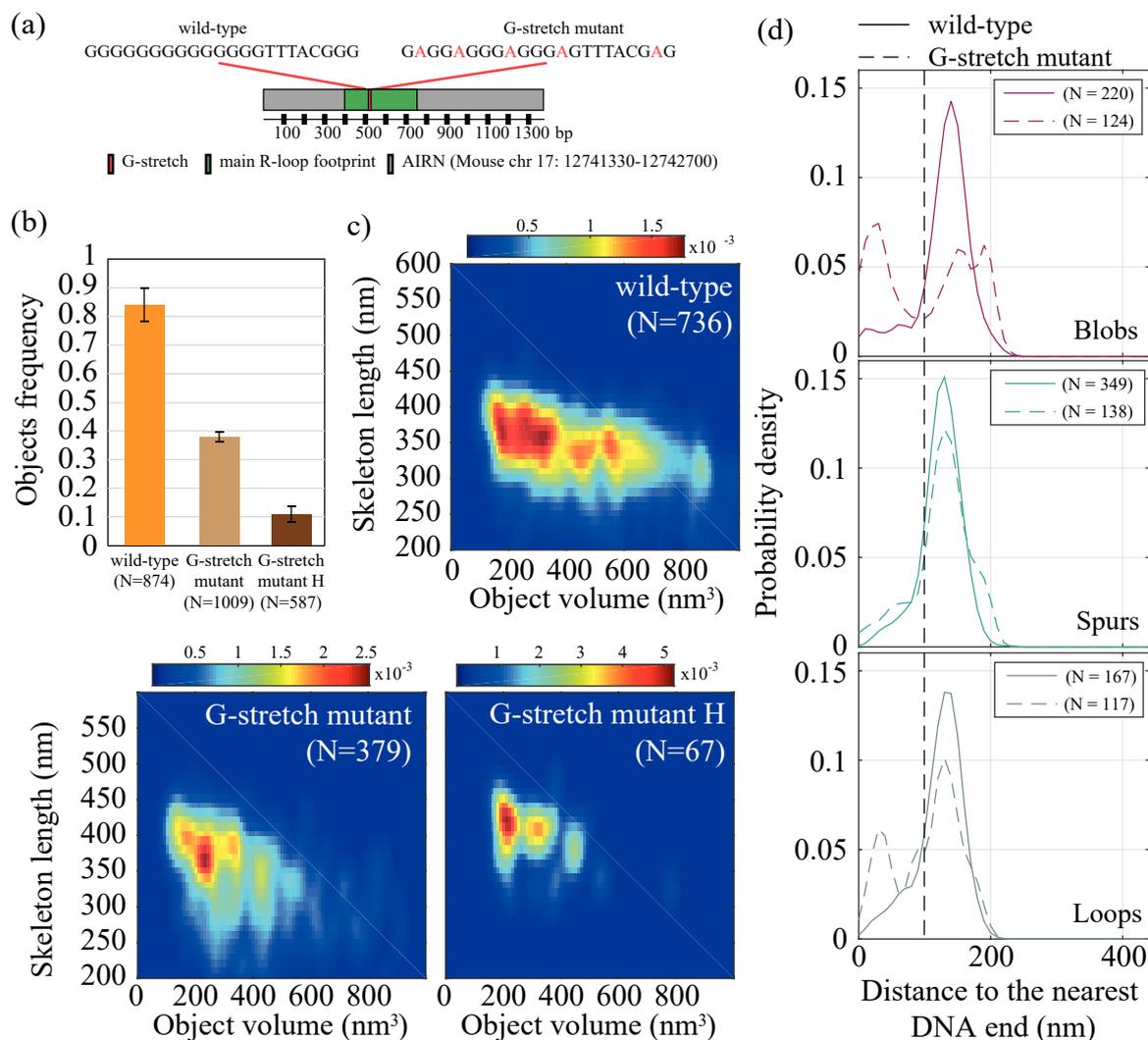


Figure 4.11: **pUC57-*airn*: The mutation affects objects but not hybrid RNA-DNA formation** - (a) Schematic representation of the mutated *airn* locus. (b) Probability of R-loop objects (mean \pm standard deviation, $n = 2$). (c) 2D probability density plots showing the correlation between the skeleton length and the objects volume in the indicated conditions. (d) Distance between R-loop objects and the nearest DNA end in the indicated conditions and for the different types of objects.

gene is between 36 pb and 504 pb. R-loops were formed on this *smairn* WT gene. The large-scale AFM images for the three conditions (NT, +A and +AH) are shown in figure 4.12(b). The transcribed sample images clearly show the R-loops objects; two molecules are magnified to better visualise these objects. Non-objects are identified in the NT sample, and rare objects are visualised in the +AH sample. Besides, a reduction in size are presented in the last condition mentioned [single molecules for the NT and +AH condition in figure 4.12(b)]. Confirming the AFM images qualitative analysis, no molecules with R-loop objects have been counted in the NT sample [Fig. 4.12(c)].

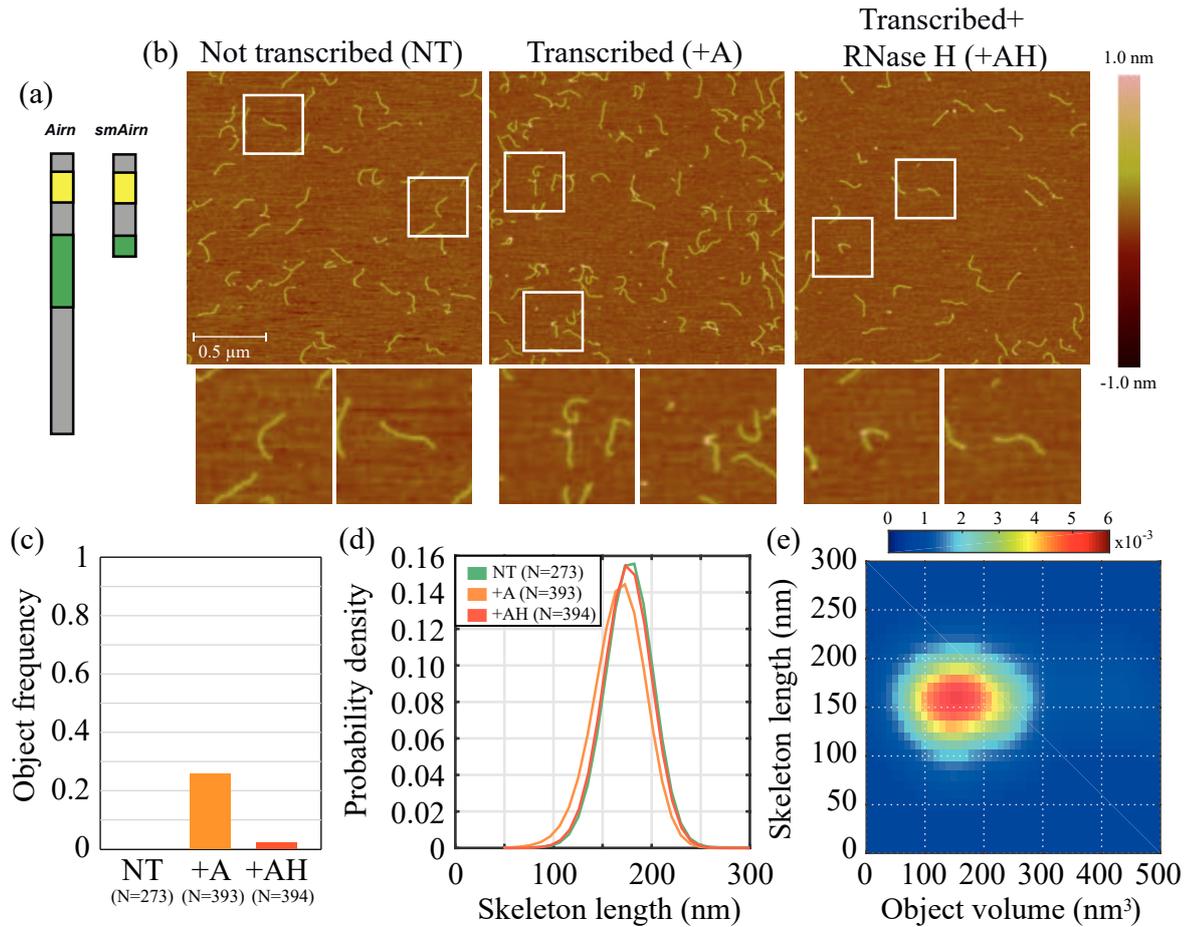


Figure 4.12: **pFC53-small-airn: new R-loop position** - (a) Scheme representing *airn* and *smAirn*. (b) AFM images of the not transcribed, transcribed and transcribed+AH templates. Individual DNA molecules are molecules zoomed that are marked with white boxes in the large images. (c) Probability of objects. (d) Probability density of the skeleton length of all the molecules analysed in the indicated conditions. (e) 2D probability density showing the correlation between the skeleton length and the volume of the object of 393 molecules.

From 393 molecules analysed, approximately 30% contains R-loops objects when the sample is transcribed, and less than 5% of the molecules present R-loops object when the RNase H is present. These results suggest that the R-loop generation in cluster 1 and cluster 2 are not correlated. The probability of finding R-loop objects only in cluster 1 (*smAirn*) is the same as when the entire gene is present in the molecule [Fig. 6 in article [130]]. The probability density of the DNA skeleton length shows a slight difference when the R-loop objects are present in the DNA molecule [Fig. 4.12(d)]. Although it is possible to differentiate between blobs and spurs, there are no differences in size: we see this effect in the correlation between skeleton length and volume for all types of R-loop object; no differences are identified according to the R-loop type [Fig. 4.12(e)].

In conclusion, these observations demonstrate the substantial agreement between AFM mapping data and chemical footprinting data [130], establishing that the *airn* contains two R-loop formation sites, where one of them is more stable, coinciding in addition to the G-stretch ones. Mutations in the main cluster, involving only five residues of 1375 nt long, generated an impact on the R-loop object architecture and the position in which the R-loop is formed (secondary cluster).

4.5 *Sum3* gene

The analysis carried out on R-loops in the *airn* gene until now suggests that its architecture is a sequence-dependent secondary structure within the non-template strand. Consequently, we evaluated the possibility that their architecture can distinguish R-loops between different genes. Therefore, we analysed another gene for the purpose to compare both structures. Vincent and Frédéric groups identify R-loops previously in the fission yeast can form a comparable amount of R-loops as *airn* upon *in vitro* transcription [131]. The *sum3* on agarose gel and dot blot analysis shows that R-loop is present in this configuration [see figure 7a-b in [130]]. However, AFM images did not show colour differences in DNA molecules, suggesting that although R-loops are present, secondary structures are not [Fig. 4.13(a)]. We corroborated by looking at the DNA molecule lengths, as it is not affected [Fig. 4.13(b)] as the *airn* gene does [Fig. 4.2(b)]. These observations show that the extruded ssDNA at *sum3* gene does not form detectable R-loop objects.

In the same way, the persistence length did not show changes in the presence and absence of R-loop objects [short fragment in figure 4.14(a)]. Therefore, and as we see in the AFM images [Fig. 4.13(a)], the presence of a non-detectable R-loop does not generate mechanical constrains in the DNA molecule.

To conclude, our results demonstrate that R-loop formation does not necessarily lead to the formation of R-loop objects detectable by AFM, suggesting that the extruded non-template strand determine the three-dimensional architecture of R-loops *in vitro*.

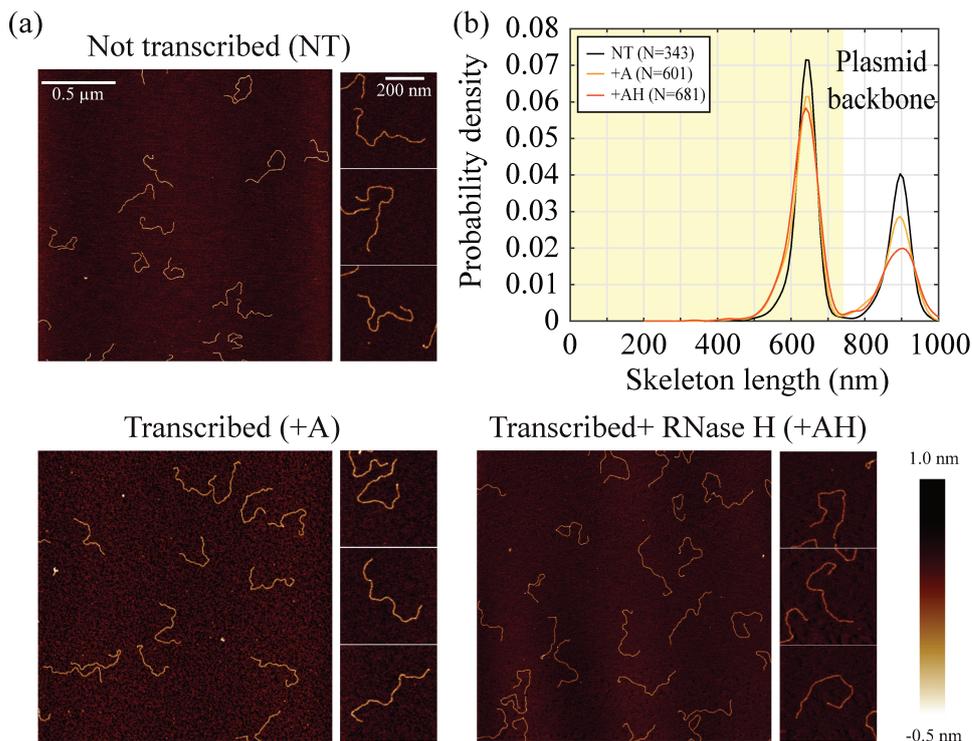


Figure 4.13: **pUC57-*sum3*: the extruded ssDNA does not form detectable R-loop objects** - (a) AFM images for NT, +A and +AH conditions. The *sum3* gene are contained in the short fragment. (b) Probability density of the DNA skeleton length of all the molecules analysed in the indicated conditions.

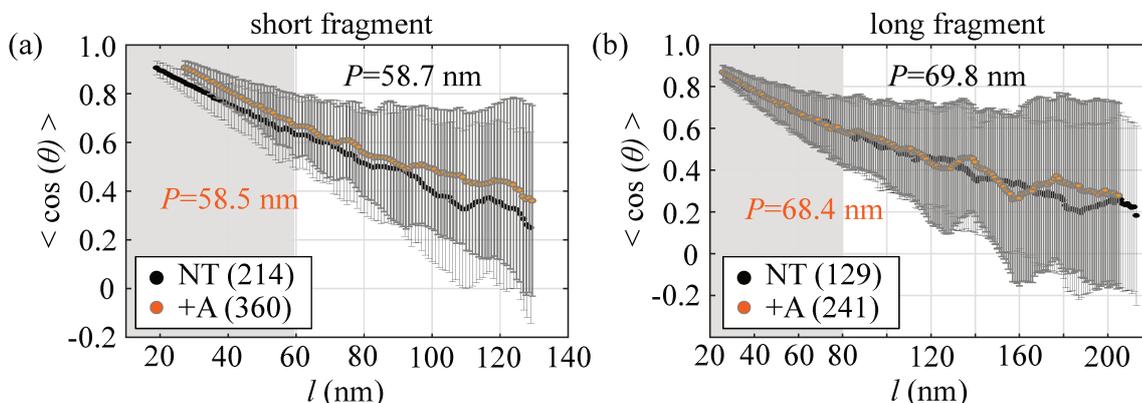


Figure 4.14: **pUC57-*sum3*: DNA persistence length without R-loop object** - (a,b) Orange and black circles represent data for transcribed and not transcribed DNA where 214 and 360 molecules were analysed for the short fragment and 129 and 241 molecules for long fragment, respectively. Both graph show the decay of tangent-tangent correlations and P was derived from the worm-like chain model to the data. The grey region were determined from the median value of the kurtosis, i.e, the expected value 3.

4.6 Conclusions

We have studied the R-loops objects resulting from the *in vitro* transcription of AIRN gene using the combination of AFM imaging in the air at very high resolution and automated image analysis a large population of DNA molecules. Theses R-loops objects disappear when RNase H is applied to digest DNA/RNA hybrids.

We have studied the R-loops objects resulting from the *in vitro* transcription of AIRN gene using the combination of AFM imaging in the air at very high resolution and automated image analysis of a large population of DNA molecules. Theses R-loops objects disappear when RNase H is applied to digest DNA/RNA hybrids.

The R-loops formations generate a G-stretch interaction with other parts of the gene through Watson-Crick or Hoogsteem links, generating secondary structures. Mutations in the gene (introduction of five G to A) did not affect the DNA-RNA hybrid amount; however, differences related to the R-loop objects are observed. Hence, the stretch of 14 G is somehow involved in the formation of these secondary structures. The R-loop objects new position is observed at 100 bp approximately, with less probability of appearance. We show no correlation between the two possible R-loop object position and the appearance frequency related to G's amount in the stretch. Finally, we compared the R-loop structure generated in the airn gene with the sum3 gene. Non-visible R-loops object is identified through AFM images, suggesting that the extruded ssDNA at sum3 gene does not form detectable R-loop objects.

The variation of the persistence parameter in the presence of R-loop objects globally changes the DNA molecule mechanical properties. On the contrary, when R-loop objects are not detectable in the AFM images, the DNA persistence did not undergo structural changes.

Our R-loop formation analysis at single-molecule resolution demonstrates that the extruded non-template strand sequence determine R-loops three-dimensional architecture *in vitro*. Consistent with this, different R-loops adopt different conformations and mutating the DNA sequence without interfering with R-loop formation resulted in significant alterations to the 3D R-loop objects detected by AFM.

Part II

Physical properties of viral capsid

Chapter 5

Introduction: Physical virology focused on nanoindentation experiments

Index

Summary	87
5.1 Viruses	87
5.1.1 Virus classification	88
5.1.2 Virus life cycle	91
5.2 The architecture of the viral capsid	93
5.2.1 Icosahedral Symmetry	94
5.2.2 Buckling transition	96
5.3 Mechanical stiffness of virus capsids using nanoindentation experiments	97
5.4 Models to describe the elastic response of viral capsid . .	100
5.5 Results in mechanical properties of viruses	101
5.5.1 Stiffness depending on the rotational symmetry axes	101
5.5.2 Genome increased the capsid stiffness	103
5.5.3 Stiffness depending on the protrusion on the capsid surface .	104
5.5.4 Stability of viral capsid through capsid disassembly	105
5.5.5 Stiffness value for the symmetry axis	105
5.6 Adeno-associated virus AAV for gene therapy	109
5.6.1 Viral cycle of AAV	109

5.6.2	Serotypes	110
5.6.3	AAV genome	112
5.6.4	Inverted Terminal Repeat (ITR)	113
5.6.5	AAV capsid	116
5.6.6	AAV stability and mechanical properties	117
5.7	Hepatitis B virus HBV	119
5.7.1	Viral cycle of HBV	121
5.7.2	HBV genome	122
5.7.3	Structure of HBV capsid	124
5.7.4	HBV mechanical properties	126
5.8	Our work in this context	127

Summary

The second part of this thesis is based on the physics of viruses. Our research focused on studying the mechanical properties of viral capsids of the Adeno-Associated Virus and the Hepatitis B Virus and how these properties are related to capsid thermal stability. As the main tool, we used the Atomic Force Microscope AFM. Through imaging and nanoindentation experiments, we measured the morphology and mechanical properties of the capsids. We have also characterised the AAV genome morphology; specifically, the ITRs structures found at the genome ends. To understand both the biological and physical aspects of viral capsids, we briefly introduce the virus biology and the mechanical aspects of capsids.

5.1 Viruses

Generally, viruses are associated with diseases since they were discovered as infectious agents such as yellow fever or rabies. However, some viruses are nonpathogenic, such as the Adeno-Associated Virus AAV currently used for gene therapy. Viruses are supramolecular complexes composed of RNA or DNA genome, single-stranded (ss) or double-stranded (ds). The genome viral is encapsulated in a capsid made of proteins, whose role is to protect the genetic material. The viruses can not generate metabolic energy for their reproduction; they depend on a living cell to replicate their nucleic acid and synthesise proteins. The process of multiplication and propagation of viruses

varies according to the type. However, there are certain fundamental principles in the events of the infection cycle typical to all viruses. Understanding the mechanics of the processes involved, such as the assembly or disassembly of the capsid, entry into the cell or trafficking in the cytoplasm, can be essential to develop more effective treatments for viral diseases. Here, we will address topics from the point of view of biologies, such as the viral cycle or the capsid structure and the physics of viruses, to analyse the viral capsid mechanical properties.

5.1.1 Virus classification

There are different virus classifications according to the shape, genome, enveloped or non-enveloped and life cycle. The viral genome can be single- or double-stranded, linear or circular. Most DNA viruses contain a single genome of linear dsDNA, ranging from 4-6 kb in length. Viruses that contain DNA as genetic material are called desoxyriboviruses or DNA viruses and included Adenovirus, Herpes, Hepatitis, and Smallpox Virus. In contrast, viruses that contain RNA are called riboviruses or RNA viruses. The most common viruses are Tobacco Mosaic, Yellow Fever, Rabies, Dengue, Measles, and Hepatitis (A, C).

Figure 5.1 shows a classification according to the genome type and morphological features, i.e., the capsid is enveloped or non-enveloped. However, the most popular Classification is D. Baltimore's Classification (American scientist), since it is based on the nucleotide sequences and their replication, allowing a more precise classification between the species of viruses. In the Baltimore classification, there are seven Groups (I, II, III, IV, V, VI, VII) where each group shows the genome type that the virus presents and how it is converted into messenger RNA, which synthesises viral proteins [Fig. 5.2] [132].

- In **Group I**, the messenger RNA is directly transcribed from double-stranded DNA. From this messenger RNA, proteins are synthesised, and finally, the DNA replicates and gives rise to new viral particles.
- **Group II** consists of viruses with single-stranded DNA and is converted to double-stranded DNA, generally by the proteins of the host cell. The subsequent process is similar to group I. The Adeno Associated Virus AAV belongs to the parvovirus family.
- **Group III** presents positive double-stranded RNA. The viruses contain an RNA-dependent RNA polymerase responsible for the transcription of double-stranded

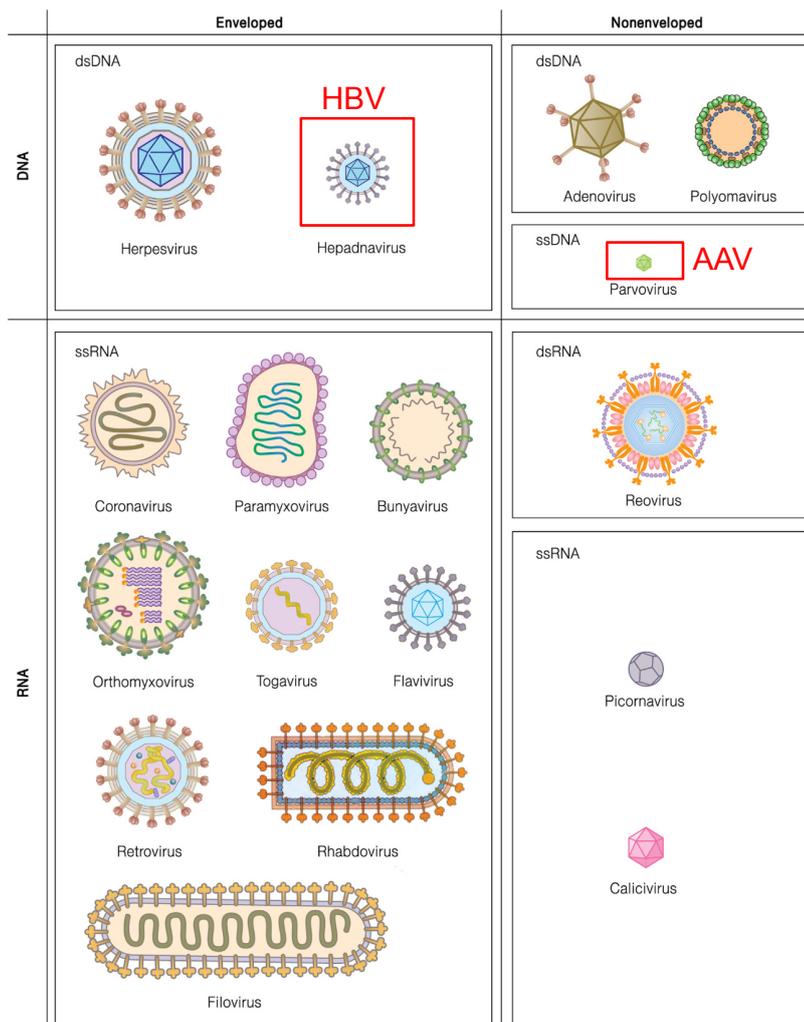


Figure 5.1: **Classification of viruses** - The viruses are classified by the nature of the nucleic genome (i.e., DNA or RNA) and their morphological features (non-enveloped and enveloped). Family names are given on each virus drawing [132]. AAV belongs to the parvovirus family with a non-enveloped capsid and ssDNA genome. On the other hand, HBV belongs to the hepadnavirus family with an enveloped capsid and dsDNA genome.

RNA to messenger RNA, and then the transcription to proteins is carried out. Genome replication begins with the assembly of the positive single-stranded RNA and the viral proteins in immature virions, and then transcription of the single-stranded RNA to double-stranded RNA inside virion is performed. Finally, all components are assembled, forming the new virus.

- In **Group IV**, positive single-stranded RNA viruses can act directly as messenger RNA and give rise to viral proteins. For replication, messenger RNA encodes RNA replicase, responsible for making copies of the viral RNA and duplicating itself.

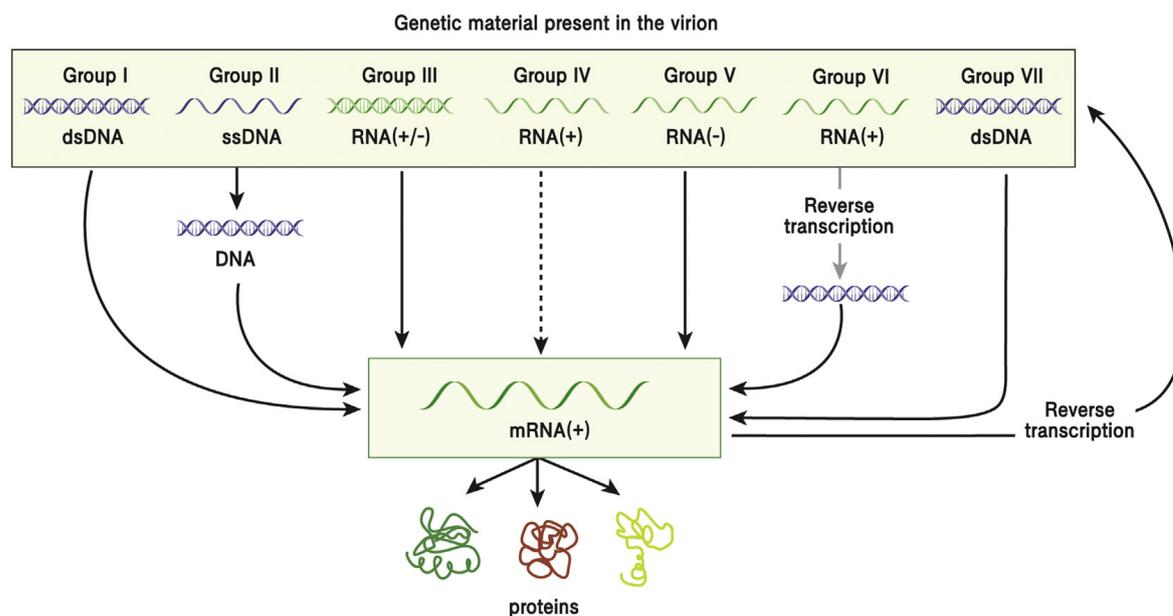


Figure 5.2: **Baltimore classification of viruses according to the genome** - Viruses are divided into DNA viruses (Group I and II), RNA viruses (Group III, IV, and V), and Reverse Transcription viruses (Group VI and VII). The relationship of the genome to mRNA is indicated either by solid line (transcription) or dotted line (no transcription). Terminology: ss: single-strand and ds: double-strand, polarity (plus or minus) [132].

- The negative single-stranded RNA of the **Group V** can not act directly as messenger RNA. An enzyme of the virus synthesises the positive messenger RNA from the negative single-stranded RNA, giving rise to viral proteins. The positive single-stranded RNA is then synthesised, generating negative single-stranded RNA and then the new copies of the viruses.
- In **Group VI** are the viruses are called retroviruses, and they are known as reverse-transcribed single-stranded RNA viruses. Unlike group IV, this positive single-stranded RNA does not act directly as messenger RNA. In this case, the virus integrates a reverse transcriptase by converting the RNA into single-stranded DNA and then double-stranded DNA. The formation of a DNA-RNA hybrid characterises this last step before becoming dsDNA. The DNA is integrated into the cell, and then a transcription generates the messenger RNA that allows for protein synthesis and is also the original viral genome, and the cycle is restarted.
- Finally, **Group VII** contains reverse-transcribed double-stranded DNA, where the double-stranded DNA is transcribed into messenger RNA and thus lead to viral proteins. However, to replicate the genome, the messenger RNA uses a viral reverse transcriptase. Viral proteins interact with messenger RNA and assemble

into a structure called a provirion. A DNA-RNA hybrid is synthesised from this structure before being converted into double-stranded DNA like the original genome. This group can find the Hepatitis B virus (HBV), which belongs to the hepadnavirus family, and their genome is a partially double-stranded circular DNA.

5.1.2 Virus life cycle

The viral cycle is represented mainly by the replication process. Virus replication is the process by which viruses infect cells and generate new copies to infect new cells, and it depends on whether the virus is non-enveloped or enveloped and the type of nucleic acid, DNA or RNA. In the following, we will explain the more general process of the virus life cycle before focusing on the nuclear entry step.

In the life cycle, the **binding process** between the virus and the cell is generated by receptors found on the cell surface [Fig. 5.3(a)]. So, the **process of entry or penetration** into the cell depends on whether the nucleocapsid (viral genome and capsid) is enveloped or naked [Fig. 5.3(b)]. Generally, the non-enveloped virus enters by stimulating receptor-mediated endocytosis. An inward fold of the plasma cell membrane or endosome is generated, allowing the virus to enter the cell. Two other cases are possible, the virus crosses the membrane without generating endocytosis, or the virus injects the genetic material directly into the cell.

On the other hand, the enveloped virus enters the cell through two mechanisms. The first mechanism is by fusion with the membrane, the envelope of the nucleocapsid fuses with the cell plasmatic membrane, and then the nucleocapsid is directly delivered to the cytoplasm: viruses contain RNA are representative of this behaviour. The other possibility is endocytosis, like non-enveloped viruses.

Following, the **capsid uncoating and genome release** [Fig. 5.3(c)]. Once the viral genome is released, it is replicated. The **replication process**, as we explained before, depends on the family type. Then the genome is translated, and thus, the capsid proteins are synthesised [Fig. 5.3(d)]. These processes occur in the cell nucleus or the cytoplasm, usually for viruses that contain DNA or RNA, respectively. Subsequent, the **capsid proteins and nucleic acid are assembled** into full capsids, and these two processes occur sequentially or simultaneously in a coupled manner. [Fig. 5.3(e)]. Finally, the new **virion is released from the host cell**. The non-enveloped viruses are released by cell lysis (cell breakdown); consequently, the cell dies. On the contrary, enveloped viruses emerge by exocytosis. In the latter case, the virus is surrounded by

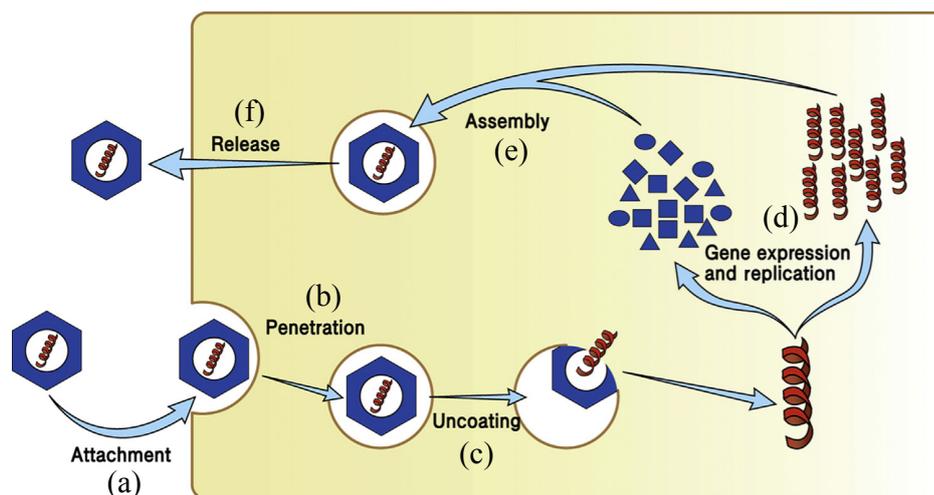


Figure 5.3: **Life cycle of viruses** - The viral cycle could be divided into attachment, penetration, uncoating, gene expression and replication, assembly, and release. Here, the schematic steps of entry and release shown for a non-enveloped virus [133].

a lipid membrane; the cell plasma membrane and the cell does not die immediately but remains infected; it is called the cytopathic effect [Fig. 5.3(f)].

5.1.2.1 Nuclear entry processes

For efficient replication, most viruses must enter the nucleus, and this process depends on the interaction with the nuclear envelope. The viruses pass through the nucleus either by passive entry, nuclear pore complex (NPC) shuttling, or active nuclear envelope (NE) remodelling [134]. Passive entry refers to the mechanism when viruses do not dispose of active mechanisms to access the nucleus. So, it depends on a natural mitotic nuclear envelope breakdown event: the nuclear envelope disassembles during a meticulously regulated process [Fig. 5.4(a)]. The Human Papillomavirus (HPV) and Murine Leukaemia Virus (MuLV) are examples of this category. On the contrary, Herpes Simplex Virus type 1 (HSV-1) and Hepatitis B virus (HBV), for example, have active mechanisms to enter the nucleus, entering through the nuclear pore complex shuttling. This pathway does not disturb the nuclear envelope and therefore preserves host cell homeostasis [Fig. 5.4(b)].

Unlike herpesviruses (HSV-1), hepadnaviruses (HBV) are much smaller enveloped viruses, and nuclear entry could occur directly through the complex nuclear pore since the size of HBV is within the range of the channel diameter of the nuclear pore. However, their transport is also assisted by transport receptors [135]. Finally, the viruses that induce herniations and full-blown ruptures of the nuclear envelope are

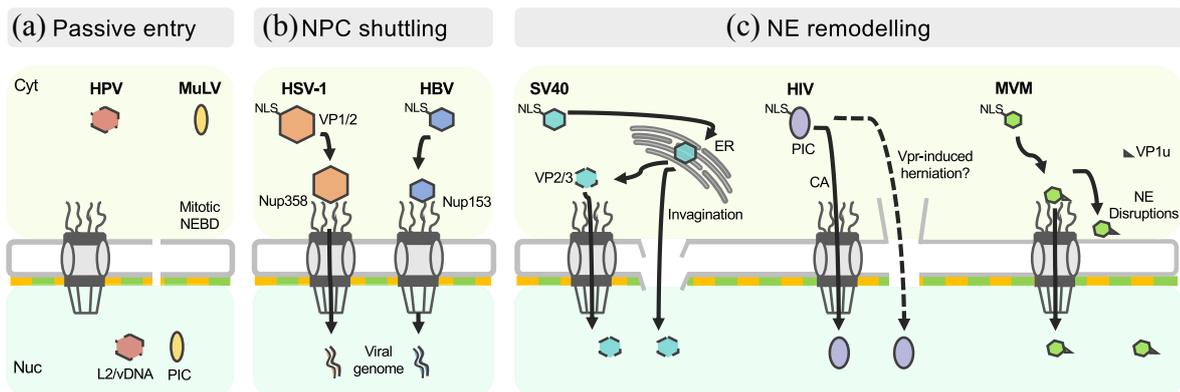


Figure 5.4: **Mechanisms of viral nuclear entry** - The viruses pass through the nucleus either by passive entry (a), NPC shuttling (b) or through active NE remodelling (c). The human papillomavirus (HPV) and murine leukaemia virus (MuLV) depend on mitotic nuclear envelope breakdown. Herpes Simplex Virus type 1 (HSV-1) and Hepatitis B Virus (HBV) will expose a nuclear localisation sequence at their capsid to guide their transport to the nuclear pore complex and HSV-1 dock at nuclear pores complex to release their genome into the nucleus. In contrast, HBV capsids can enter the nuclear pore complex basket ring, disassemble inside the NPC and release their viral genome and capsid subunits into the nucleus. Simian Virus 40 (SV40), Human Immunodeficiency Virus (HIV) and Minute Virus of Mice (MVM) will remodel and perturb the nuclear envelope to facilitate entry, even though they expose a nuclear localisation sequence. [134].

defined as nuclear envelope remodelling [Fig. 5.4(c)]. The nuclear entry remodelling is the mechanism by which most parvoviruses function [136, 137]. The Parvoviridae is a family of small (18-26 nm in size), non-enveloped, icosahedral viruses with linear single-stranded genomes by which they may be able to cross the nuclear pore complex by diffusion. However, most like Minute Virus of Mice (MVM), cause a transient loss of the integrity in the nuclear envelope [136, 137]. Particularly, certain parvoviridae, such as MVM and AAV, enters the nucleus and remain intact through the nuclear pore complex without the help of nuclear receptors [138–140].

5.2 The architecture of the viral capsid

Viruses present various shapes and sizes according to each viral family. In general, virus geometry is based on capsomeres distribution and classified into polyhedral, helical, complex, or mixed. For example, Adenovirus or Herpes Virus have polyhedral geometry because capsomeres form faces by arranging themselves as a polyhedron. Viruses such as Tobacco Mosaic, Influenza or Ebola are composed of capsomeres adhered to nucleic

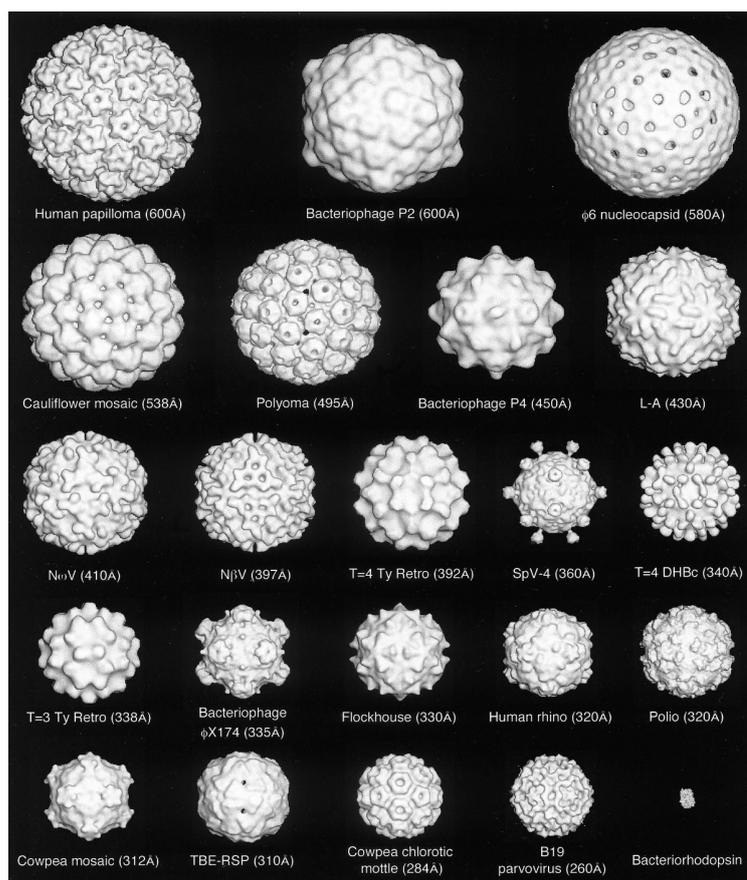


Figure 5.5: **Gallery of icosahedral viruses** - 3D image reconstruction of several icosahedral viruses by using cryo-EM [141].

acid, forming a helix or spiral that follows the virus longitudinal axis. The viruses with mixed geometry are a combination of icosahedral and helical symmetry. For example, Bacteriophages have a head with icosahedral geometry containing the genome and a helical tail.

The two most common geometries are those with a helical and spherical morphology (usually icosahedral symmetry). This section will delve into the capsid with icosahedral symmetry since it is the symmetry-type of the capsids studied in this thesis. Figure 5.5 shows different viruses with icosahedral symmetry where we see that the Human Papilloma has a size of 600 Å and B19 Parvovirus of 260 Å in size, for example.

5.2.1 Icosahedral Symmetry

The icosahedron structure exhibit three rotational symmetry axes: 2-, 3- and 5-fold, referencing the edges, faces and vertices of an icosahedral. [Fig. 5.6]. The symmetry

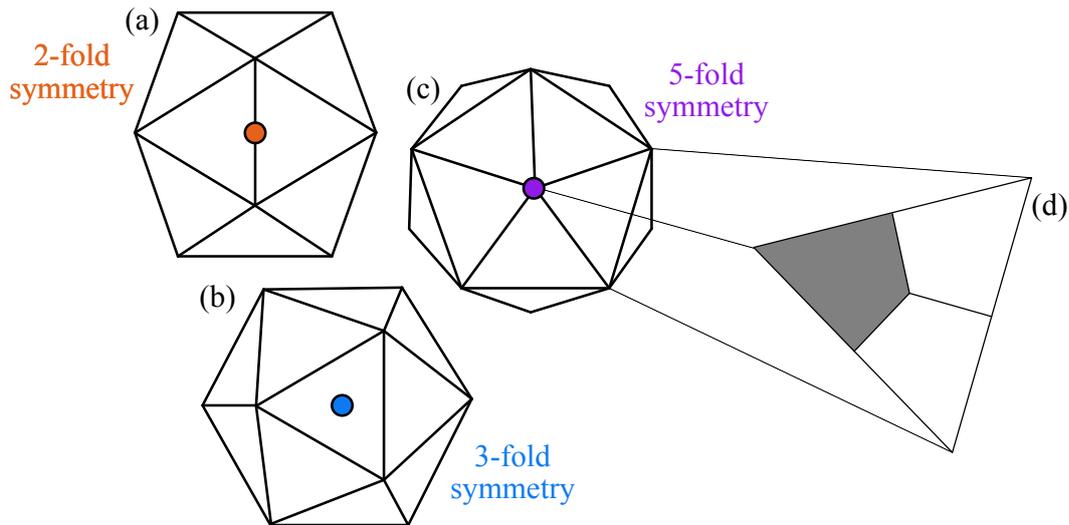


Figure 5.6: **Orientations of icosahedral capsid** - Orientation 2 (a), 3 (b) and 5 (c). The example shows a capsid with structure $T = 1$, that is, it contains 60 asymmetric units (1 unit is 1 capsid protein). A unit is represented in gray in figure (d).

architecture is related to the conformation of 60 asymmetric units (one unit is one protein), where a unit has grey colour in the example of the figure 5.6(d)]. However, not all viruses are built by 60 asymmetric units. The construction of a stable structure from asymmetric units allowing maximising the volume per unit area to contain the viral genome is then limited.

Crick and Watson showed that symmetry was compatible with capsids made of different numbers of subunits [142]. The subunits or capsomeres are arranged in pentamers and hexamers [Fig. 5.7(a)]. A pentamer is always in exactly 12 places on the capsid surface, determining the global shape of the capsid [143]. Visually, the pentamer contains five proteins, while a hexamer has six proteins. Also, the hexamer is always positioned at the centre of three pentamers. The amount of hexamers varies between groups of viruses. On the other hand, the "quasi-equivalence" principle of Caspar and Klug [144] explains how proteins can adopt positions in the capsid, repeating local configurations across the capsid surface. Thus, it is necessary to identify the closest pair of pentamers [white circles in figure 5.7(a)] and the distribution of proteins between these two pentamers. The number of capsomeres and its symmetry is determined by the triangulation value T :

$$T = h^2 + h * k + k^2. \quad (5.1)$$

The variables h and k are integer values and represent the numbers of hexamers

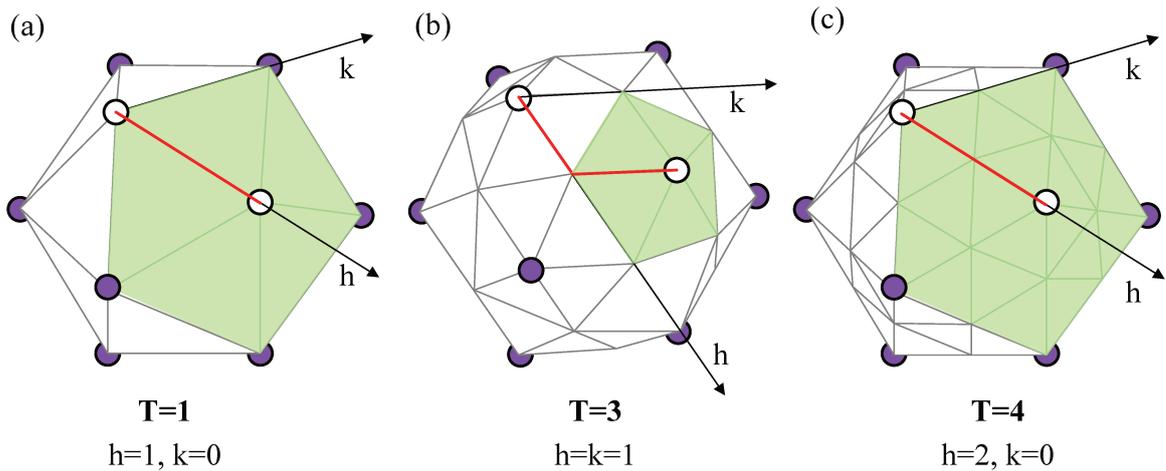


Figure 5.7: **Scheme of icosahedral capsid with different triangulation number T** - Capsids with triangulation number (a) $T=1$, (b) $T=3$ and (c) $T=4$. The white circles are positioned at the pentamers, and the path to get from one pentamer to its nearest neighbouring one is highlighted in red.

and pentamers through the shortest path. That is, starting at a pentamer [white circles in figure 5.7(b-d)], h represents the hexamer numbers in straight line and k the pentamer number after changing direction [Fig. 5.7(b-d)]. For $T=1$ and $T=4$, the shortest distance between two pentamers is given for one capsomer ($h = 1, k = 0$ or $h = 0, k = 1$) and two proteins ($h = 0, k = 2$) in a single direction, respectively [Fig. 5.7(b,d)]. For $T=3$, we move one hexamer and then a pentamer after having changed direction ($h = k = 1$) [Fig. 5.7(c)]. The examples show that the triangulation value can take values $T=1, 3$ and 4 , but for larger viruses, T can have values of $7, 9, 12, 13, 16, 19, 21, 25$, and so on. This principle is based on the fact that shell always contains 12 pentamers plus $10(T - 1)$ hexameres, hence, we can determine the total amount of capsomers as the sum of the pentamers and hexamers that a capsid contains:

$$N_{capsomers} = 10T + 2.$$

5.2.2 Buckling transition

The diameters of the viruses span 10 to 100 nm, where small viruses as HBV are said to be spherical viruses. On the other hand, large spherical viruses can exhibit faceted geometry such as bacteriophage T4 and HK97 [Fig. 5.8(a)]. Lidmar *et al.*, [145] explained the faceting of the capsids as the buckling transition resulting from the competition between the stretching energy associated with the pentamers (in-plane energy) in the lattice of capsomers and the bending elasticity of the viral capsid (out of plane energy), i.e., the Föppl-von Kármán (FvK) number [Fig. 5.8(b)]. Figure 5.9

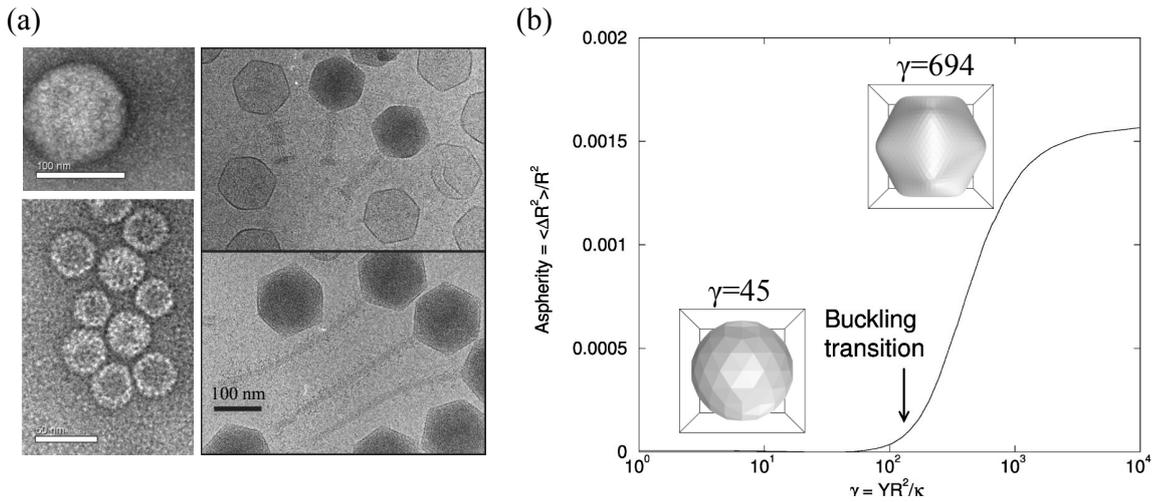


Figure 5.8: **Buckling transition and the Föppl–von Kármán (FvK) number** - (a) Viruses can have different level of faceting as we see for HSV-1 (top-left), HBV (down-left), bacteriophage T4 (top-right) and HK59 (down-right). Adapted from [146, 147]. (b) Mean-square asphericity as a function of YR^2/κ for many different icosahedral shells. The two inserts show numerically calculated shapes with (h,k) indices 2,2 and 8,8. The Föppl–von Kármán numbers γ for these shapes are 45 and 694. Adapted from [145].

show some virus illustrations that for the same triangulation value T , the capsid can show a marked faceted geometry.

5.3 Mechanical stiffness of virus capsids using nanoindentation experiments

As we mentioned, viruses present a great variety of genetic material type, shape, or geometry, which can significantly impact the capsid morphology and stability. Therefore, studying the mechanical effects of viral shells can help understand the virus biological behaviours through their viral cycle. The Atomic Force Microscopy is an instrument that allows measuring both the mechanical and morphological properties of viruses and any other micro or nanoscopic object. The main properties that are used to characterise the capsid are (i) the capsid stiffness (k), (ii) structural strength when a force is applied, this is, the maximal force that the capsid support before breaking F_{break} and (iii) resistance to material fatigue; progressive damage when the capsid is subject to a series of controlled low-force indentations [149].

If the indented object is non-deformable, as is the surface where the virus capsid is

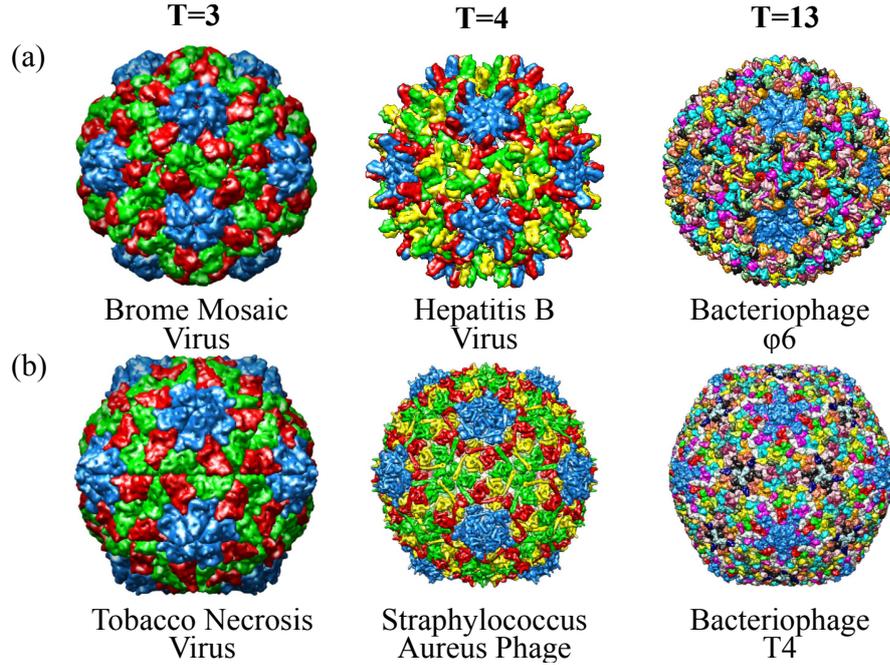


Figure 5.9: **Effect of faceting on some viral capsids** - (a) Illustrations of viruses with spherical capsid for $T=3$, $T=4$ and $T=13$. (b) Illustrations of viruses with spherical capsid but that have a marked faceted geometry for the same triangular numbers. Representations were obtained from *VIPERdb* [148].

deposited, only the cantilever is deformed. Hence, the displacement of the cantilever is given by Hooke's law: $F = -kz$, where k is the elastic constant of the deformed object, in this case, the elastic constant of the cantilever k_c [gray line in $z-F$ plot, figure 5.10(a)]. On the contrary, if the virus capsid indented is mechanically deformable; thus, the force curve contains cantilever-capsid system deformation. If the cantilever and the capsid behave like two ideal springs, the cantilever-capsid system can be considered as two springs in series [scheme and black line in $z-F$ plot, figure 5.10(a)]. Consequently, the spring constant of the virus capsid can be determined from the slope of the corresponding $z-F$ curve (k_{eff}), according to the equation: $\frac{1}{k_{eff}} = \frac{1}{k_c} + \frac{1}{k_v}$, where k_c is the cantilever spring constant [149].

Another method to obtain the deformation of the virus capsid is by subtracting the $z-F$ curve over the surface and capsid-cantilever system from the point of contact. The new force curve as a function of indentation only contains the virus capsid deformation [Fig. 5.10(b)]. Finally, the capsid stiffness is the slope of the indentation-F curve [13, 151]. This thesis determined the stiffness of AAV and HBV capsids, mainly using this second method.

The sudden fracture in the indentation-F curves represents the point where the

capsid structural strength is lost caused by mechanical failures [150] [Fig. 5.10(c)]: figure 5.10(d) shows AFM images of an intact capsid (A) and a capsid that has lost one capsomer after being indented (B). The drop in the curve also allows calculating the indentation depth [point indicated with a black arrow in figure 5.10(c)]. The structural strength of the viral capsid sometimes exhibits particular properties. For example, the norovirus capsid (icosahedral and nonenveloped capsid, T=3) persists at large deformations, and after several indentations, the force curves eventually show drops, suggesting a mechanical failure in the capsid. However, later curves did not show discontinuities, suggesting a possible recovery of the capsid [152]. Also, a self-recovery of the T7 bacteriophage virus shell (tailed icosahedral and nonenveloped capsid, T=7) has been experimentally shown when fractures and cracks are small enough not to alter the original shape of the virus shell. Nevertheless, this type of event is rarely seen and needs more research to control and understand the virus self-recovery of protein shell [153].

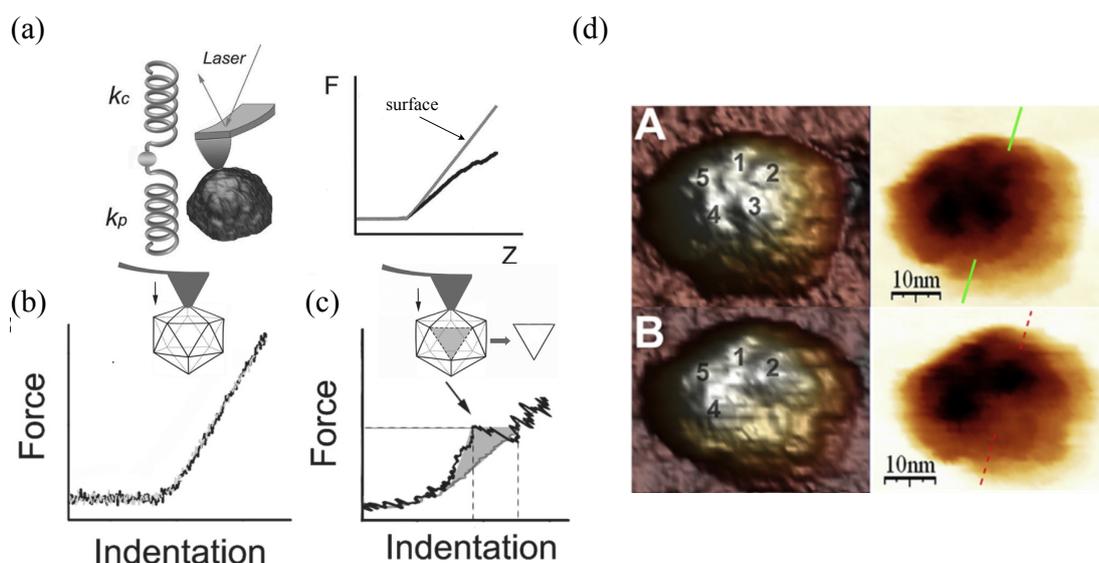


Figure 5.10: **Scheme of a nanoindentation experiment and its z - F curves of MVM capsid** - (a) Scheme of the AFM cantilever and virus capsid as two ideal springs in series and z - F curve obtained by indentation over the surface (grey curve) and the virus capsid (black curve). (b,c) z - F curve on the MVM capsid where the black and grey curves represent the trace and retrace curves, respectively. (b) Indentation force curve of an intact capsid. (c) Force curve shows a nonlinear behaviour due to material failure of the virus capsid. The force curve show extends and retracts tip displacement. Adapted from [149]. (d) AFM images before (A) and after (B) indentation, where one capsomer was removed in the latter case. Adapted from [150].

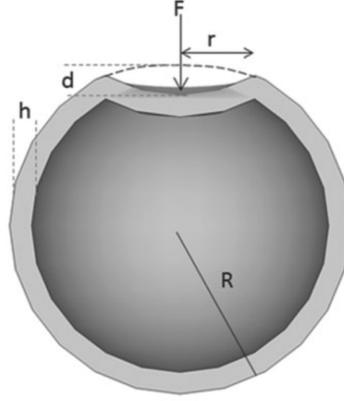


Figure 5.11: **Scheme of a shell sphere** - (a) Scheme of a spherical shell [4].

5.4 Models to describe the elastic response of viral capsid

When a spherical shell of radius R and thickness $h \ll R$ is deformed on a distance d by an indenting point force F in the regime where also $d \ll R$, the shell is locally both stretched and bent [Fig. 5.11]. Thus, the capsid stiffness can be interpreted as the ratio between the applied mechanical force F and the indentation length d [4]. Introducing the stretching and bending energy as

$$E_{str} \sim K \left(\frac{d}{R} \right)^2 r^2 \quad (5.2)$$

$$E_{ben} \sim \kappa \left(\frac{d}{r^2} \right)^2 r^2, \quad (5.3)$$

where K and κ are the stretchings and bending modulus defined as

$$K = \frac{Eh}{2(1 - \mu)} \quad (5.4)$$

$$\kappa = \frac{Eh^3}{12(1 - \mu^2)}, \quad (5.5)$$

where, E is the three-dimensional Young modulus and μ is the Poisson ratio of the material. Within standard 3D elasticity for a shell sphere, the total energy can write the sum of bending and stretching energies as:

$$E_{total} \sim E \frac{h^2}{R} \left[\frac{hR}{r^2} + \frac{r^2}{hR} \right] d^2. \quad (5.6)$$

The size of the bend region r can be determined by imposing the equilibrium condition $dE_{total}/dr = 0$, then, $r \sim \sqrt{hR}$. From the derivate of the energy to vertical deformation d , the stiffness k of the shell can be obtained,

$$k \sim E \frac{h^2}{R}. \quad (5.7)$$

This equation is valid only for small deformations $d \ll R$. Furthermore, k can be written in terms of the dimensionless number associated to stretching/bending balance, which is known as the Föppl–von Kármán (FvK) number defined as:

$$\gamma = \frac{\kappa}{K} R^2. \quad (5.8)$$

Then, shell sphere stiffness can be written based on this parameter:

$$k \sim \left(\frac{\kappa K}{R^2} \right)^{1/2} \sim \kappa \gamma^{-1/2}. \quad (5.9)$$

5.5 Results in mechanical properties of viruses

This section focuses on the mechanical properties of viral capsids, specifically capsid stiffness. The mechanic stiffness of some viral capsids will be analysed according to the capsid architecture and stability.

5.5.1 Stiffness depending on the rotational symmetry axes

The first work presenting in detail the mechanical properties of viral capsids was registered by Ivanovska *et al.*, [154]. The authors showed that the mechanical response of bacteriophage $\Phi 29$ (ellipsocylindrical, icosahedral and non-enveloped capsid, T=3) behaves linearly when the tip of the cantilever vertically deforms the capsid. For large deformations, a force of 2.5 nN, the shell undergoes irreversible fractures. Furthermore, they found a bimodal distribution for capsid stiffness. Later, work on the Cowpea Chlorotic Mottle Virus (icosahedral and non-enveloped capsid, T=3) pointed

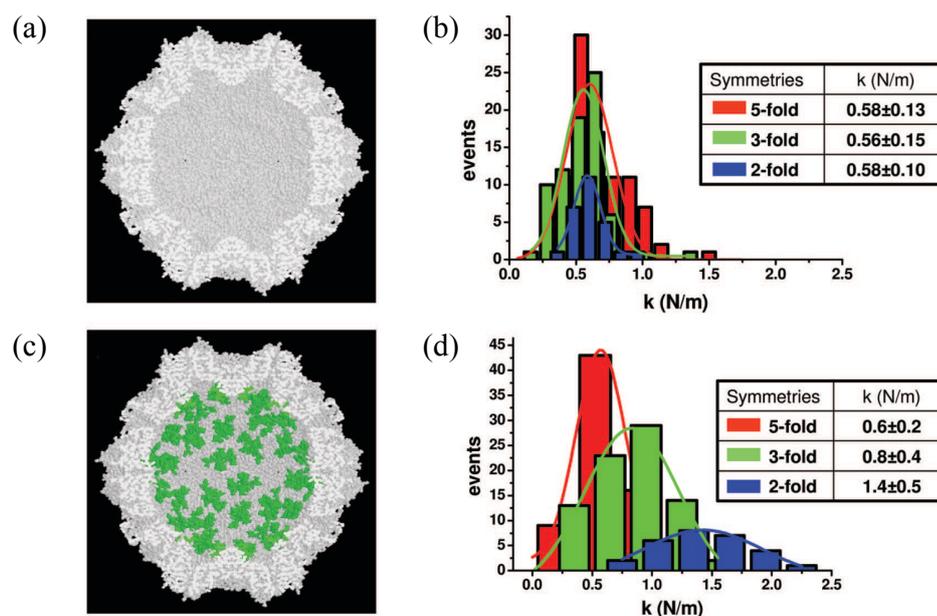


Figure 5.12: **Stiffness of the MVM empty and full capsid** - Crystallographic structures of the empty capsid (a) and DNA-filled virion (c) Histogram of stiffness for empty capsid (b) and DNA-filled virion (d) according the orientation axe of the capsid [156].

out that the elastic response could vary between hexamers and pentamers, resulting in a stiffness bimodal distribution [155].

Focusing on small parvovirus, Carrasco *et al.* [156] showed that the virion of Minute Virus of Mice (icosahedral and non-enveloped capsid, $T=1$) presents a stiffness dependent on the orientation axis of the capsid [Fig. 5.12(c,d)]. Another study by the same group determined the same value of stiffness for the 5-fold symmetry axis [157]. Interestingly, in the absence of the viral genome, the stiffness of the capsid does not show variations according to the orientation axis [Fig. 5.12(a,b)]. These results suggested that anisotropy is due to the capsid-bound DNA. Later, the capsid protein was mutated, eliminating the DNA-capsid interaction and confirmed the theory about the anisotropic reinforcement given by DNA. In particular, they found no variations in the 5-fold symmetric axis [158]; the authors suggested that the MVM is not reinforced on this axis to allow DNA released during infection. Some work showed that the 5-axis pore is the genome encapsidation site for viruses belonging to the parvovirus family and other viruses with icosahedral symmetry. Thus, the pore size is wide enough, ~ 1.5 nm for MVM [159], to allow the transit of DNA in its linear conformation [160–167]. The pore is characterized by being narrow, so small modifications can generate inefficient encapsidation of the genome [162,165]. At the capsid mechanical properties level, these structural changes could be interpreted as changes in the pore stiffness.

Subsequent studies have worked on the axis orientation-dependent capsid stiffness; however, some results appear to show some discrepancies. For example, van Rosmalen *et al.*, 2018 reported orientational dependence on the human Adenovirus stiffness (icosahedral and non-enveloped capsid, T=25) [168] unlike Snijder *et al.*, [169]. Additionally, differences in the stiffness value are also present between each study [168–171]. The authors suggest that the stiffness differences could be due to a difference in particle selection or the ratio between the cantilever tip and capsid size. Cuellar *et al.*, 2010 and Baclayon *et al.*, 2011 do not mention the dependence of stiffness according to the orientation for the Norovirus capsid (icosahedral and non-enveloped capsid, T=3) [152, 172], neither did Michel *et al.*, 2006 for Cowpea Chlorotic Mottle Virus (icosahedral and non-enveloped capsid, T=3) [155]. The authors suggest that measurements made with sharper tips may have a chance to obtain variations concerning the capsid orientation.

Aversely, a study on Herpes Simplex Virus type 1 (icosahedral capsid and envelope, T = 16), no changes in the capsid mechanical properties were observed in the presence of DNA. [173, 174]. Utrecht *et al.*, [175] and Arkhipov *et al.*, [176] reached the same conclusion through nanoindentation experiments on HBV capsid (icosahedral and enveloped capsid, T= 3, 4).

5.5.2 Genome increased the capsid stiffness

The genome not only generates anisotropy in the stiffness of the viral capsid but also causes the capsids to become more rigid as found on MVM (ssDNA), phage (dsDNA), CCMV(ssRNA) and AdV (dsDNA) capsids [155, 156, 158, 177–181]. For example, Ivanovska *et al.*, [179] analysed the effect of dsDNA internal pressure on bacteriophage λ (tailed icosahedral and non-enveloped capsid, T=21) in response to an external deformation 5.13. Figure 5.13(a) show two-dimensional projections of cryo-electron microscopy reconstruction of the 100%, 94% and 78% DNA-filled and empty phage capsids. Then, the partially filled capsids responded to external deformation like an empty capsid since the stiffness capsid were similar, on contrary, the WT capsid (100% DNA filled) is twice stiffer thus, more resilient to external forces Figure 5.13(b).

However, recent work on Simian Virus 40 (icosahedral and nonenveloped capsid, T=7) with circular double-strand DNA genome did not exhibit any effect of the genome on capsid stiffness [182]. Similar results were obtained for Herpes Simplex Virus type 1 (icosahedral and enveloped capsid, T=16) that contains dsDNA genome [173]. Zeng *et al.*, [183] also experimentally found no significant differences in the presence or absence of the ssDNA genome for Adeno Associated Virus 2 (icosahedral and non-enveloped

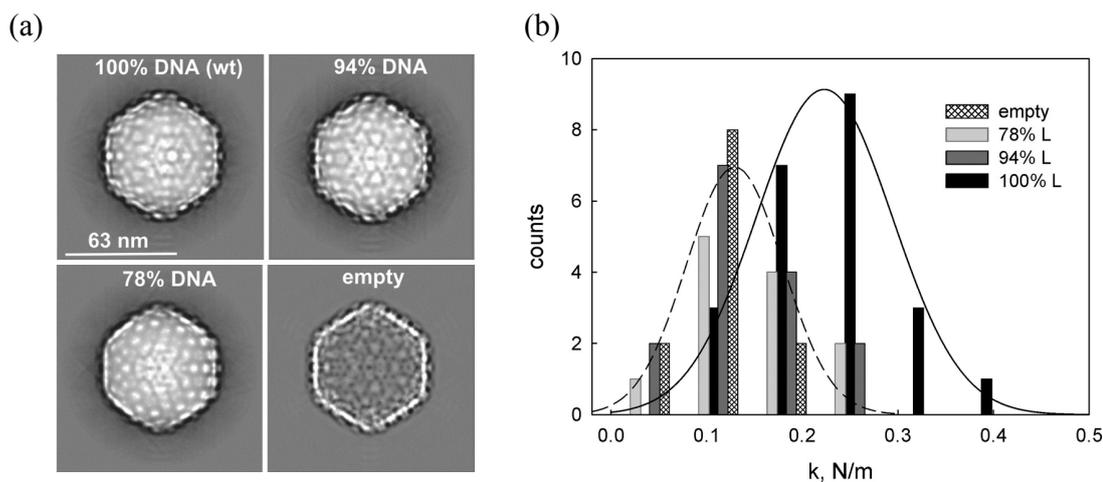


Figure 5.13: **Genome length effect on phage capsid** - (a) Two-dimensional projections of cryoelectron microscopy reconstruction of the 100%, 94% and 78% DNA-filled and empty phage capsids. (b) Histogram of spring constants for the conditions mentioned. [179].

capsid, $T=1$). However, we should mention that this analysis was carried out only for the 2-fold symmetry axis. In this same study, capsid stiffness was compared according to the type of DNA, in which no significant differences were found between ssDNA and scDNA experimentally.

5.5.3 Stiffness depending on the protrusion on the capsid surface

The capsids of viruses are not completely smooth; capsids present protrusions that are generally responsible for cell union during the first stage of the infectious cycle. Baclayon *et al.*, [172] analysed the Norovirus particles having similar morphology and antigenicity as the native Norovirus (icosahedral and non-enveloped capsid, $T=3$), which exhibits hollows and protrusions on the capsid surface. Mutations allowed eliminating these protrusions obtaining softer particles, 10 nm smaller than the WT capsid [Fig. 5.14(a)]. The results suggest that the protrusions force the capsid to a prestressed state since the capsid shows a reduction in stiffness when the protrusions are removed [Fig. 5.14(b)]. Besides, the protrusion layer generates an increase in the particle radius and, consequently, a general and isotropic strengthening of the particle.

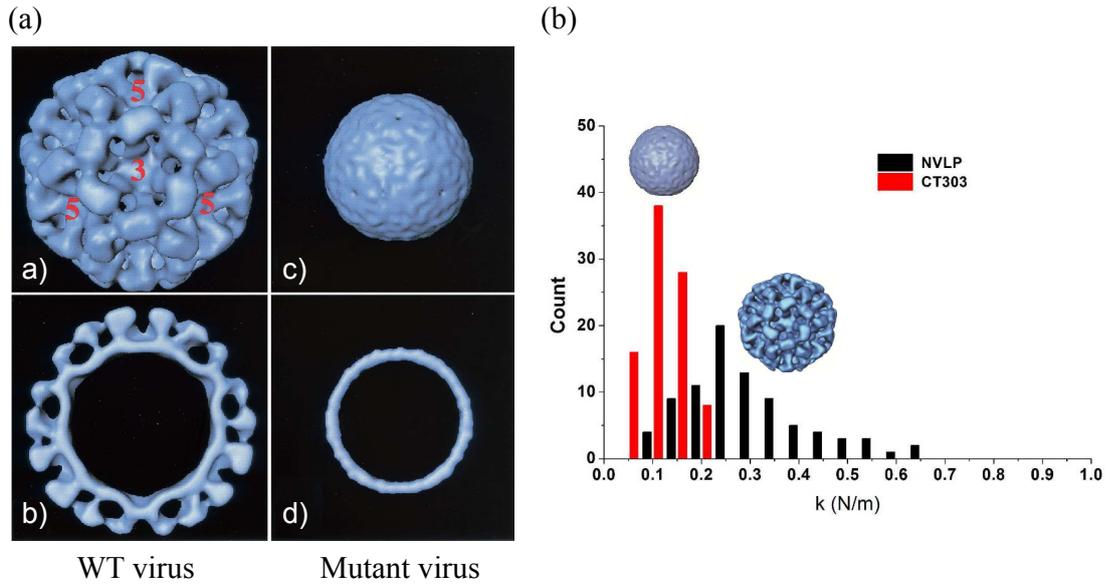


Figure 5.14: **Protrusion effect on mechanical properties** - (a) Electron microscopy reconstructions (surface and cross section) of WT (left) and mutant with no protrusions (right) norovirus capsids. (b) Stiffness distribution of two cases mentioned [172].

5.5.4 Stability of viral capsid through capsid disassembly

A study on the stability of the *Triatoma* Virus capsid (icosahedral and non-enveloped capsid, $T=3$) was carried out by Snijder *et al.*, [178]. They show that the capsid can divide into pentamers when the capsid is subjected to high forces by AFM. Figure 5.15(a) shows AFM images before and after indentation, where the pentamers are visible in the second case [Fig. 5.15]. Additionally, the same results were obtained when DNA released was induced at high pH. Ortega *et al.*, [184] reached the same conclusion studying the Adenovirus capsid. The pentons are released first, followed by complete disassembly. In this case, the experiments are performed in the YOYO-1 dye molecule, allowing the released DNA visualisation. For bacteriophage 22 (icosahedral capsid and non-enveloped, $T = 7$), Llauro *et al.*, 2016 [185] show that the pentamers were released at the vertices inducing a destabilisation of the capsid, implicating the edges as points of mechanical fracture.

5.5.5 Stiffness value for the symmetry axis

In the example shown above [Fig. 5.12], we see that the stiffness of the capsid follows the following scaling order according to the axis of symmetry: $k_5 < k_3 < k_2$ [156]. Experimentally, the same order of stiffness was found for the bacteriophage T7 (icosa-

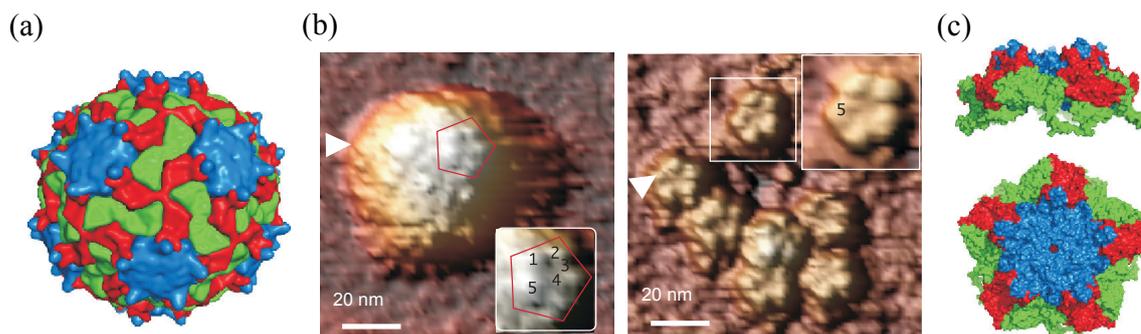


Figure 5.15: **Capsid disassembly into pentamers for Triatome virus capsid** - (a) Surface renderings of intact capsid and pentons (left). AFM images before and after nanoindentation (right). Reconstructions show the individual points of maximal force in function of the capsid stiffness at pH 6.8, pH 8 and pH 9 for single virions. Adapted from [178].

hedral and non-enveloped capsid, $T=7$) [Fig. 5.16(a)]. To reproduce these results, the authors implemented Finite Element simulations of increasing complexity [186]. In the simplest case, the capsid of bacteriophage T7 was modelled as a thick shell icosahedron, and the order of the effective stiffness was found numerically to be $k_2 < k_3 < k_5$, which differs from the experimental measurements [Fig. 5.16(b)]. It was then implemented a more realistic inhomogeneous finite element simulation based on the electron microscopy (EM) structure of the bacteriophage T7 empty head. In this way, they analysed the elastic response of inhomogeneities in the capsid thickness and the real topography of bacteriophage T7. The effective stiffness of k_2 and k_3 was found identical, and k_5 was bigger than the other two. Thus, the order of the effective stiffness also differs from the experimental results [Fig. 5.16(c)]. These simulations results suggest that the mechanical response of T7 cannot be modelled in the framework of the continuum elasticity.

Other works also showed differences according to the method used. Using an elastic network normal mode analysis, in other words, an analysis that allows describing the flexible states that the capsid adopts around an equilibrium position, Tama *et al.*, [188] indicated that the capsid shell is not mechanically uniform, especially for viruses of higher complexity such as $T=7$ and $T=13$ viruses. In most cases, the pentamer displays greater flexibility than the other subunits, which is in agreement with an experiment performed on the Cowpea Chlorotic Mottle Virus (icosahedral and non-enveloped capsid, $T=3$) and determined that pentamers are more flexible than hexamers [189]. On the contrary, Hesperheide *et al.*, [190] proposed a higher stiffness for the pentamer than the hexamer using a method that maps the chemical bond of the protein in a generic 3D graph, in which the edges represent distance constraints between atoms, i.e., the atoms are restricted to a certain distance. This graph is then decomposed into

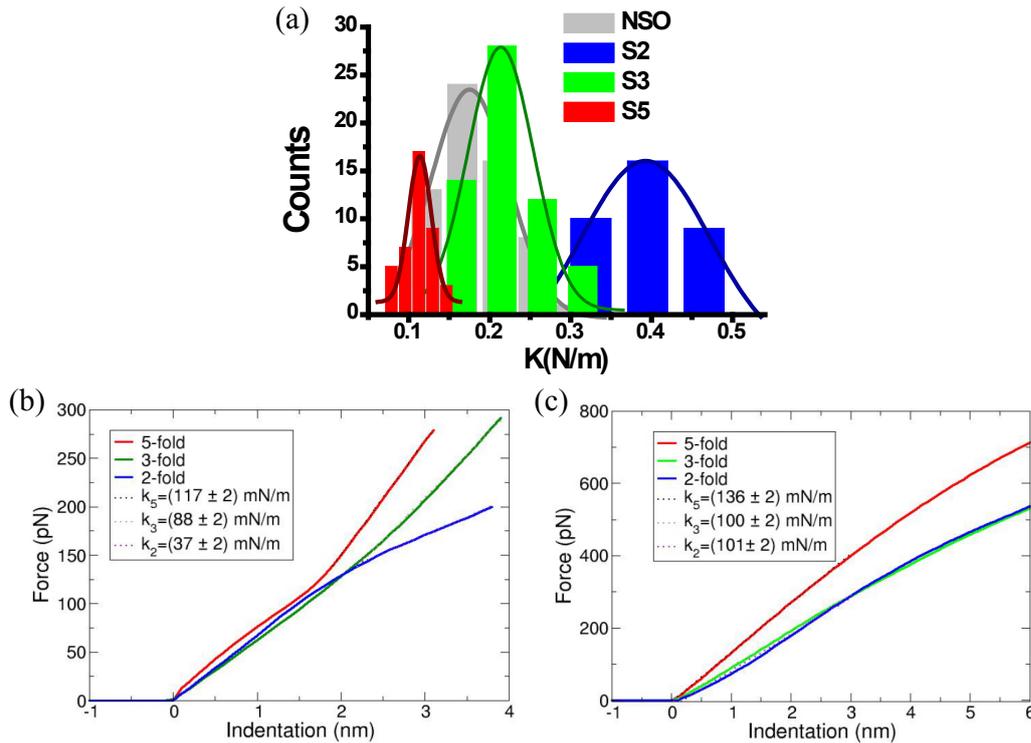


Figure 5.16: **Stiffness of bacteriophage T7 capsid** - (a) Histogram of capsid stiffness where blue, green and red colors correspond to capsids adsorbed on 2-fold, 3-fold and 5-fold symmetry axes, respectively. Grey color depicts stiffness data from particles with no specific orientation (NSO). (b) Indentation curves for the thick finite elements model. (c) Indentation curves for the inhomogeneous finite elements model [186].

rigid and flexible regions. Zandi *et al.*, [187] reached the same conclusion. The author determined the local stresses and pressures from continuous modelling. The compressive stress on the pentamers increases with the capsid size, as predicted by continuum elasticity theory [Fig. 5.17].

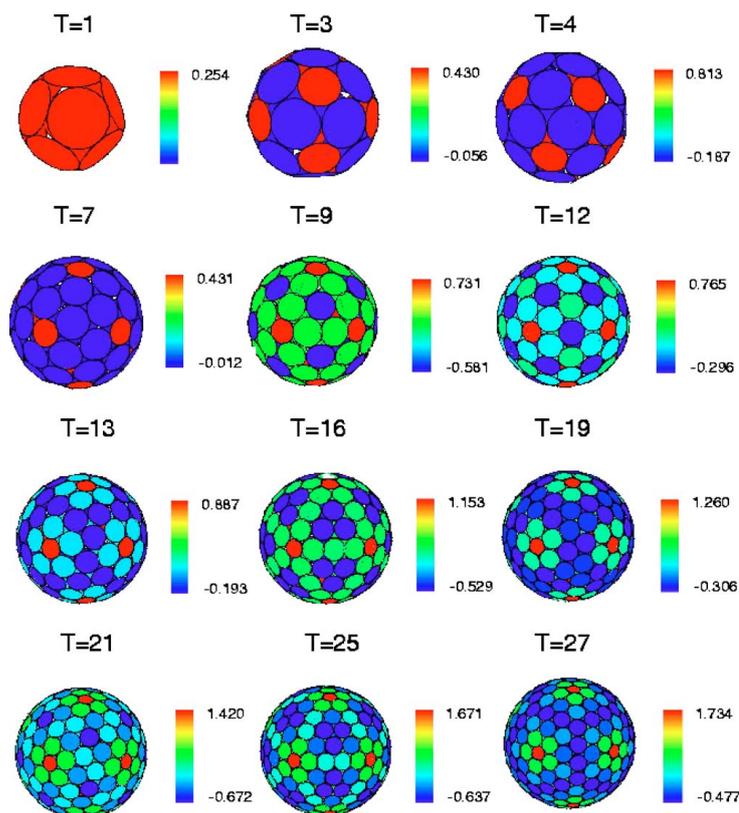


Figure 5.17: **Local lateral stresses of some T-number** - The color bar represents the lateral stress distribution for some of the T-number structures and the different colors indicate different levels of stress. Pentamers and hexamers have different values of stress by which pentamers present consistently more positive stress than hexamers [187].

5.6 Adeno-associated virus AAV for gene therapy

In vivo gene therapy delivers genetic material directly into the target organs or tissues of patients. This technique is the simplest way to introduce genetic material into target cells via non-viral or viral vehicles to treat or prevent diseases by correcting or supplementing defective genes. The viruses used for this purpose are known as recombinant vectors. A deficiency causes most diseases in a specific gene; for example, haemophilia generate mutations in genes encoding clotting factors and can be corrected by delivering the missing gene. Since 2012 AAV vectors are well established tools in clinical trials for *in vivo* gene therapy [191, 192]. Several AAV serotypes have been isolated and studied from which 13 natural types and more than 100 variants have continuously been generated mutating the genome or the capsid to optimise the use of AAV in gene therapy [191].

It seems interesting to see if the mechanical properties of AAV capsids can play a role in how these viral vectors can deliver the genetic material, mainly to know if there are mechanical differences according to the different serotypes. Before addressing this problem, we will briefly introduce the viral cycle, the differences according to serotypes, genome type and AAV capsid.

5.6.1 Viral cycle of AAV

The first step of the AAV viral cycle is binding primary receptors and co-receptors on the cell surface [Fig. 5.18(1)], where the union between the receptors and the capsid is given by amino acid residues on the capsid surface. The interaction with the target cell constitutes a critical parameter that determines the serotype tropism (tropism refers to the way viruses present particular preferences to specific cell tissues), generating that the receptor nature depends on the AAV serotype [194–203].

In vitro studies support that binding to the cell surface triggers endocytosis in the plasma membrane, forming an endosome whose content is to transport the capsid to the nucleus [192, 204]. Next, the capsid is released by the endosome before it arrived in the nucleus [Fig. 5.18(2)]. Evidence suggests that the endocytic vesicles could play an important role in preparing the capsid for nuclear transport and the uncoating process (capsid maturation process occurs within the endosome). Indeed, the transduction efficiency is intricately linked to the intracytoplasmatic trafficking [205–207].

The entire capsid enters the nucleus by crossing the nuclear pore complex [Fig. 5.18(3)] [138–140], and then the capsid releases the viral genome [Fig. 5.18(4)]. The

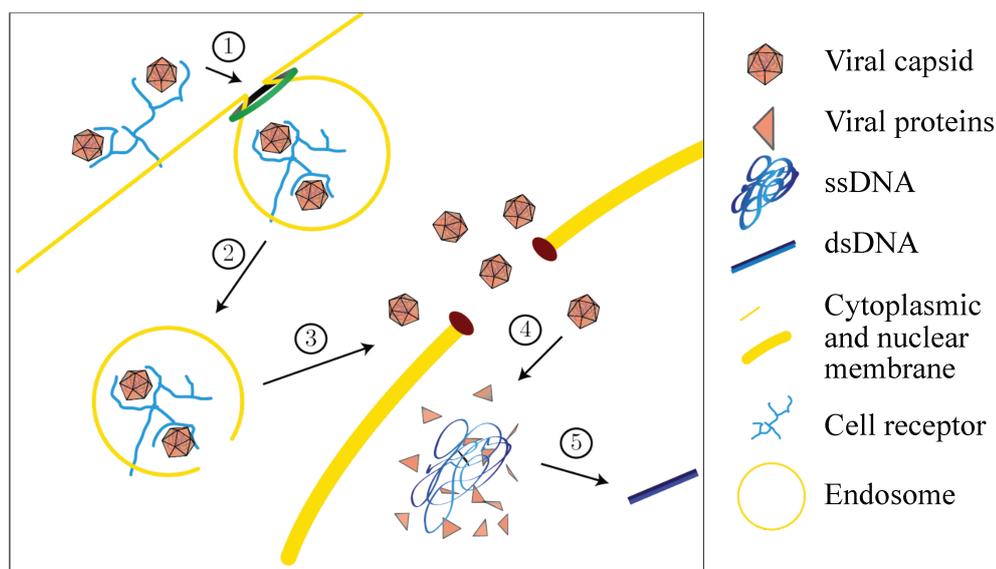


Figure 5.18: **AAV viral cycle** - (1) Virus attaches to the cell membrane of the host cell, and entry to the cell membrane triggers their endocytosis into endosomes. (2) The virus is transported toward the cytoplasm until the perinuclear area. (3) Intact virus capsid crosses the nuclear membrane through the nuclear pore complex. (4) The single-stranded DNA is released and converted into double-stranded DNA (5) (Scheme from Julien Bernaud thesis [193]).

mechanism of genome release, to date, is not clear. The single-stranded genome is converted to a double-stranded due to cellular DNA-polymerase [Fig. 5.18(5)]: this process is necessary to gene expression [208–210].

However, AAV requires helper virus such as Herpes Simple Virus or Adenovirus for replication process [194,211–215]. In the absence of a helper virus, the genome remains in a lysogenic or latent phase. In the presence of helper viruses, AAV passes into a lytic phase, allowing its replication. The efficiency of translation, which also occurs in the nucleus [207], depends on the serotype and the cell type [216,217]. Once the capsid is assembled, AAV can packaging either the plus or minus ssDNA strand with equal efficiency [218]. The multiplication of the virus also depends on the translation of the viral dsDNA genome. The translation leads to producing the viral proteins that form the capsid and the enzymatic proteins for the virion assembly.

5.6.2 Serotypes

To date, at least 13 natural serotypes have been isolated and studied from which they have different binding receptors and tissue tropisms [Fig. 5.19]. For example, AAV9 is the only serotype capable of crossing the blood-brain barrier and traduce the central

AAV serotype	Origin of isolation	Primary receptor	Co-receptor	Tissue tropism
AAV1	Monkey	Sialic acid	AAVR	Muscle, CNS, heart
AAV2	Human	Heparin	Integrin, FGFR, HGFR, LamR, AAVR	Liver, CNS, muscle
AAV3	Human	Heparin	FGFR, HGFR LamR, AAVR	Muscle, stem cells
AAV4	Monkey	Sialic acid	Unknown	Eye, CNS
AAV5	Human	Sialic acid	PDGFR, AAVR	CNS, lung, eye
AAV6	Human	Heparin, sialic acid	EGFR, AAVR	Muscle, CNS, heart, lung
AAV7	Monkey	Unknown	Unknown	Muscle, CNS
AAV8	Monkey	Unknown	LamR, AAVR	Liver, muscle, pancreas, CNS
AAV9	Human	Galactose	LamR, AAVR	Every tissue
AAV10	Monkey	Unknown	Unknown	Muscle
AAV11	Monkey	Unknown	Unknown	Unknown
AAV12	Human	Unknown	Unknown	Nasal

Figure 5.19: **Properties of AAV serotypes** - [191].

nervous system [219], and also is the serotype that has been associated with all tissue tropisms. Comparison of AAV1-9 revealed that serotypes are remarkably similar in structure [220] while differing in amino acid sequence [Fig. 5.20] [195]. Considering the structure similarity, we found it interesting to study the mechanical response of the capsids of AAV2, AAV8 and AAV9 and, if possible, to understand their behaviours at the tropism level.

AAV mutants have continuously been generated to optimise the use of AAV for gene delivery due to the specificity, and non-pathogenic behaviour [220–223]. One possible way to increase the efficiency of the AAV transduction is genetically modifying the AAV vector; this can be by optimising the transgene or engineering the AAV capsid [191, 224–228].

To describe AAV main characteristics, we will use the AAV2 serotype since it is the most extensively studied serotype of AAV and is used as the reference.

Sequence identity matrix of representative primate AAVs. Percentage sequence identity based on alignments of the capsid amino acid sequences									
	AAV1	AAV2	AAV3	AAV4	AAV5	AAV6	AAV7	AAV8	AAV9
AAV1	ID	83%	86%	63%	57%	99%	84%	84%	81%
AAV2		ID	87%	60%	57%	83%	81%	82%	81%
AAV3			ID	62%	57%	87%	83%	85%	83%
AAV4				ID	51%	63%	63%	63%	61%
AAV5					ID	57%	56%	57%	55%
AAV6						ID	84%	84%	81%
AAV7							ID	87%	80%
AAV8								ID	84%
AAV9									ID

Figure 5.20: Sequence homology for AAV serotypes - [195].

5.6.3 AAV genome

AAV contains a linear single-stranded DNA genome of 4.7 kb long [229], where the AAV genome packaged into its capsid is either the sense or antisense strand. The genome contains two open reading frames ORFs, *rep* and *cap* flanked by two inverted terminal repeats (ITRs) [Fig. 5.21(a)]. The ORFs produced the following viral proteins:

(i) The common element between AAV genes is a polyadenylation signal (polyA) required for the synthesis of messenger RNA (mRNA), and their expression is under control of three different promoters p5, p19 and p40, which is in charge of the transcription profiles.

(ii) The proteins Rep78-Rep68 and Rep52-Rep40 encoded by the *rep* gene are transcribed from the p5 and p19 promoters, respectively [230] [Fig. 5.21(b)]. The proteins four are related to the functional process as such as the genome replication, transcription regulation, site-specific integration and packaging of AAV genome into the capsid [166, 231–234].

(iii) The *cap* gene encodes the VP1, VP2 and VP3 C-terminally overlapping capsid proteins [Fig. 5.21(c)] responsible of the assembly of the capsid [235–237]. Hence, they are responsible for forming an icosahedral capsid composed of 60 copies VP1, VP2 and VP3 in a 1:1:10 ratio [238, 239]. The p40 promoter regulates the proteins encoded by the *cap* gene. Interestingly, due to the proteins overlapping in the VP3 region, a new protein is created: Assembly-Activating Protein (AAP). The AAP is responsible for the assembly of the capsid for most AAV serotypes. [240–243].

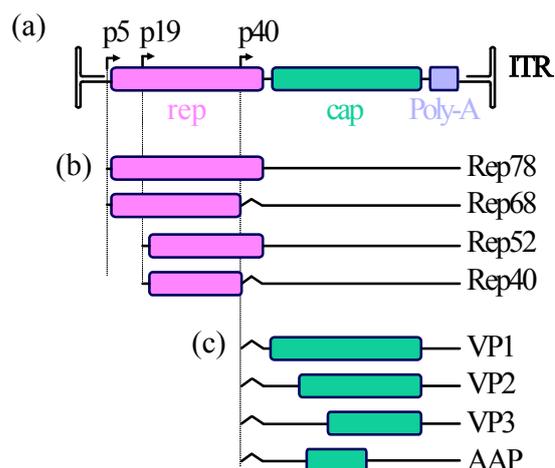


Figure 5.21: **AAV2 genome organization** - (a) The AAV genome contains two open reading frames (ORFs): *rep* and *cap* flanked by two inverted terminal repeats (ITRs). Three promoters were identified within the viral genome: p5, p19 and p40. (b) The *rep* gene encodes four proteins required for replication (Rep78, Rep68, Rep52 and Rep40) and (c) the *cap* gene encodes three structural proteins for the viral capsid (VP1, VP2 and VP3).

5.6.4 Inverted Terminal Repeat (ITR)

The AAV ITRs are 145 bp long, and the palindromic sequence has a T-shaped structure, forming two small internal inverted repeat sequences (BB' and CC') separated by a single unpaired nucleotide. A second and larger flanking inverted repeat sequence, AA', is located at the centre to form the three-way junction [Fig. 5.21(a)]. The D sequence is related to the AAV genome and does not participate in the T structure sequence. Two ITR forms have been identified: flip and flop sequences. The flop sequence formed is the inverse complement of flip [Fig. 5.22 (b) and (a), respectively]. The flip ITR has BB' arm closest to the 3' end, and flop ITR has its CC' closest to its 3' end [244]. ITRs contain a Rep binding element (RBE) and a terminal resolution site (trs), which interact with the Rep proteins [245–248].

ITRs are the only genome element that contains cis-elements required for viral genome replication and encapsidation [250] playing an essential role for AAV vector production [218, 251–253]. An increasing number of studies tend to demonstrate that host cell elements recognise the ITRs in gene therapy. For example, several innate immune responses DNA sensors were shown to recognise specific elements present in viral DNA [254–258]. Consequently, the recognition of recombinant AAV (rAAV) DNA as foreign to the cell can trigger an innate immune response leading to the destruction of the targeted cell and finally a lack of treatment efficiency. This is a question that we also address in this thesis. We think that characterising these structures can help

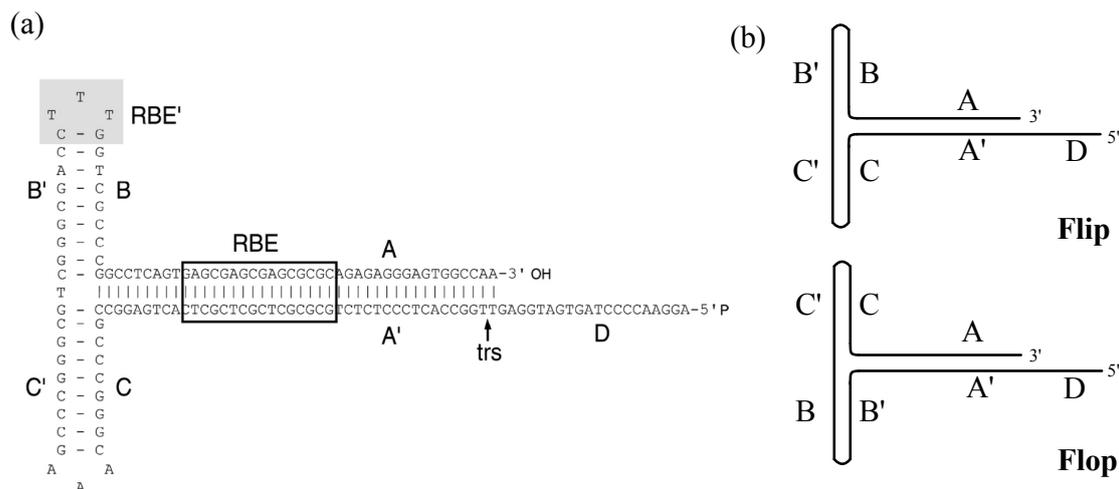


Figure 5.22: **Structure of the ITR** - (a) The T-shape of the ITR is formed by two small inverted repeats BB' and CC' and a larger inverted repeat AA' from which it is located at the centre. The D sequence is separated by the genome and does not participate in the T structure sequence, thus remaining single-stranded. The boxed motif corresponds to the Rep-binding element RBE, where the AAV Rep78 and Rep68 proteins bind. The terminal resolution site trs are required for the recognition and cleavage by the Rep proteins [249]. (b) The ITR can acquire two configurations: flip and flop. The flop sequence represents an inversion of the A-A' portion of the ITR.

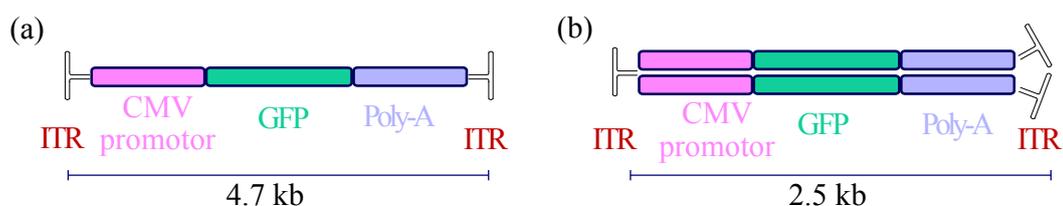


Figure 5.23: **Diagrams of ssAAV and scAAV viral DNA structures** - (a) Conventional ssAAV vector (4.7 kb) consists of two wild-type ITR framing the expression cassette. (b) The scAAV vector contains a mutation in the terminal resolution site of one ITR, while the second ITR is intact, generating a recombinant viral genome limited to 2.5 kb in size of a ssDNA viral vector genome.

to understand their role in gene therapy. For them, our study was based on analysing the morphology and the different conformations of the ITR through AFM images.

5.6.4.1 Vector genome

The two methods by which rAAV DNA can be converted into double-stranded DNA are (i) synthesis of a new complementary strand and (ii) via two-stranded complementarity with opposite polarity. However, it can be completely avoided by packaging both strands as a single molecule (self-complementary AAV vector) as the ssAAV vector (4.7 kb) consists of two wild-type ITRs [Fig 5.23(a)]. The scAAV vector contains a mutation in the terminal resolution site of one ITR, while the second ITR is intact, generating a recombinant viral genome limited to 2.5 kb in size [Fig 5.23(b)]. Although the scAAV vector has a small capacity for containing the genome, it presents a high efficiency of gene expression since conversion to a double-stranded DNA is not necessary anymore [259].

Horowitz *et al.* in 2013 analysed the viral genome behaviour, ssDNA or scDNA when the capsid is brought under thermal instability [260]. Tungsten-shadowing EM images show the released single-stranded DNA from the capsid is highly compact, and the self-complementary DNA is linear [Fig. 5.24]. Furthermore, they showed that internal pressure might be more important if the genome is of the self-complementary type since it decreases its thermal stability. For this reason, comparing the stiffness and thermic stability of ssAAV and scAAV could be an interesting question.

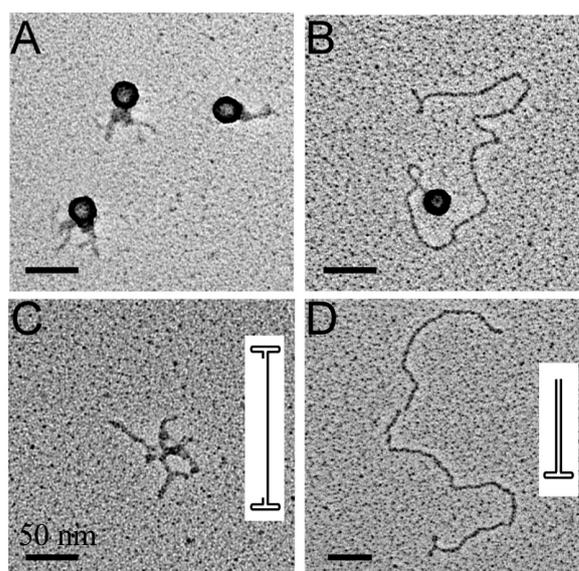


Figure 5.24: **AAV2 genome organization** - EM images of ssDNA (A,C) and scDNA (B,D) capsids. The capsids were heated to 65°. Adapted from [260].

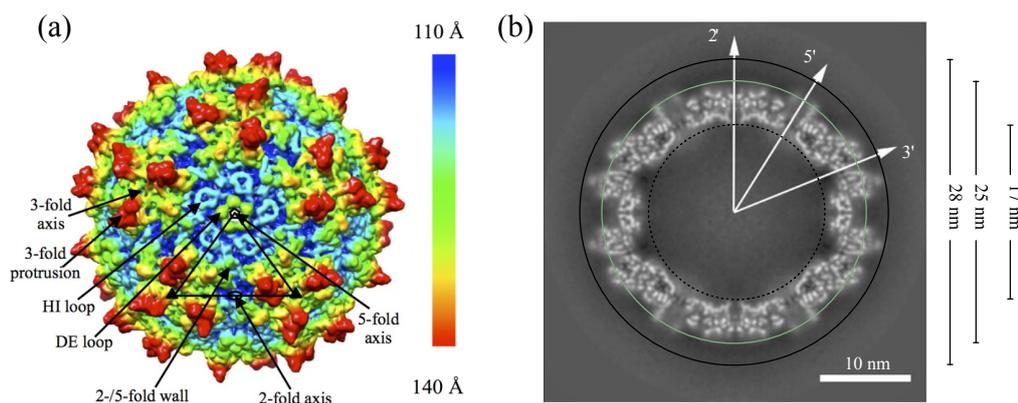


Figure 5.25: **Cryo-EM reconstruction of AAV2 capsid-** (a) Radially colored sensitivity maps at 5.0 Å. The 2-, 3-, and 5-fold symmetry axes are as well as the AAV capsid surface features are indicated by the arrows [270]. (b) Cryo-EM map of the unbound AAV2. Adapted from [195].

5.6.5 AAV capsid

AAV is a small and non-enveloped virus characterised by a structure with $T=1$ icosahedral symmetry [261]. The tridimensional structure of several AAV serotypes has been resolved by cryo-electron microscopy and/or X-ray crystallography [262–268, 268–278], showing that all these structures has a β -barrel core decorated by extended surface loop forming protrusions at the 3-fold where have 3 nm in size. Also, it is identified as a channel at 5-fold symmetry axes that connect the interior and exterior of the capsid [269]. The channel served as a route of viral genome packaging and genome release; consequently, it could play an important role during the viral replication [279]. Figure 5.25(a) shows the AAV2 capsid structure obtained by Cryo-EM and the 2-, 3-, and 5-fold symmetry axes are as well as the AAV capsid surface features are indicated by the arrows [270]. The protrusions are the most exposed regions of the capsid and are those involved with receptor binding, thus being critical sites of interaction with the host cell [195, 196, 200, 201, 222]. The inner and outer radius of the capsid is around 8 and 14 nm [dotted and continuous black line in figure 5.25(b)]. Typically, the channel size is 4-5 Å approximately, and this value determines the capsid width [194]. The average size that characterises this type of viral capsid is 25 nm in diameter [light green line in figure 5.25(b)].

Next, we will compare the structure of the capsids studied in this thesis (AAV2, AAV8 and AAV9) according to work carried out by DiMattia *et al.*, [263]. We see that the main difference between capsids is the size and shape of the protrusions. The

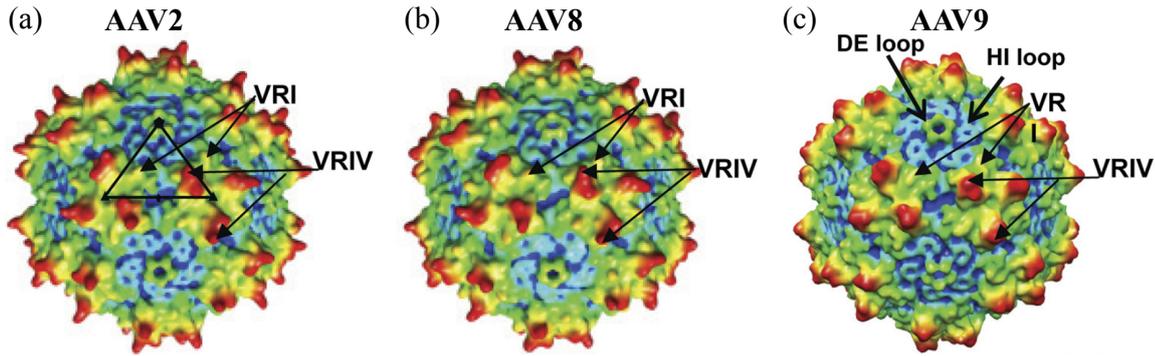


Figure 5.26: **Comparison of AAV2, AAV8 and AAV9 structures** - A surface density map of the pseudoatomic model built into the (a) AAV2, (b) AAV8 and (c) AAV9 cryo-EM reconstructed density map. The capsid surfaces shown are radially depth cued with colours ramping from red, yellow, green, to cyan, and finally to dark blue. The radial range for the depth cueing was expanded to enhance the internal surface features. The triangles (a) depict an asymmetric viral unit bounded by a five-fold axis and two three-fold axes. Arrows indicate density for the DE and HI loops in (c). Arrows indicate example capsid surface regions corresponding to VR-I and VR-IV. Adapted from [263]

“pointed finger-like” tips of the AAV2 and AAV8 capsids protrusions are not present in AAV9. The AAV9 capsid has shorter protrusions, indicating that the corresponding surface loops in the AAV9 VP3 subunit must adopt different conformations. Density at the base of each protrusion, which creates a raised surface region (a “two/five-fold wall”) between the depressions at the two-fold axes and surrounding the five-fold axes, differs three serotypes. According to the differences in the surface of AAV capsids, it would be interesting to identify, in the same way, differences at the level of mechanical properties. As we have mentioned previously, capsid protrusions are an essential factor during the AAV infection process since they involved with receptor binding, thus being critical sites of interaction with the host cell.

5.6.6 AAV stability and mechanical properties

As the AAVs have to cross the plasmatic and nuclear membrane barriers and protect the viral genome during the external and intracellular trafficking, the capsid performs a wide variety of structural and biological functions required during the viral life cycle. The capsid must maintain thermal and mechanical stability to protect the genome until it encounters a host cell. *In vitro* and *in vivo* studies have shown that the capsid stability depends on the serotype, the nature of the DNA (ssDNA or dsDNA), and the size of the viral genome. [260,270,280–283]. Additionally, the capsid remains stable for

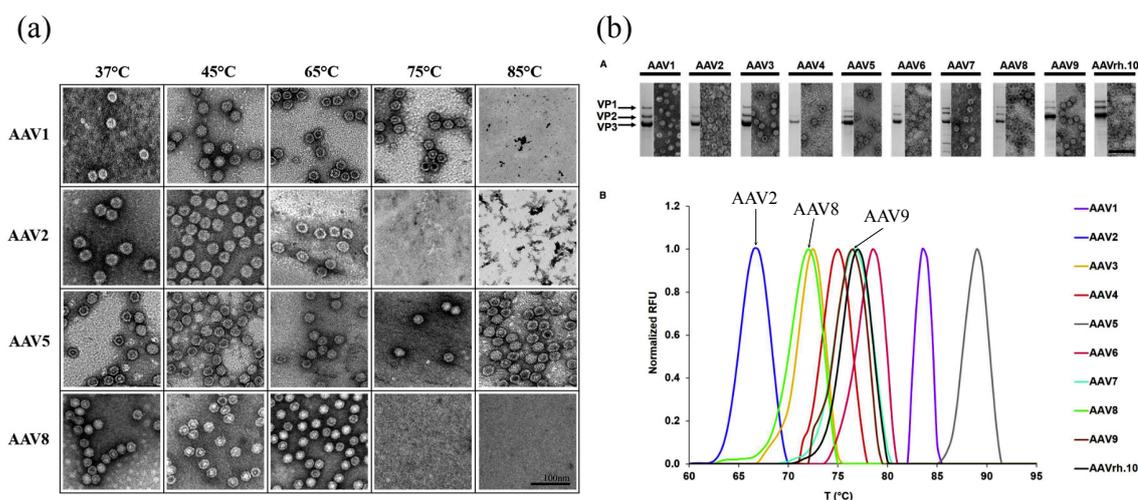


Figure 5.27: **Thermal stability of the AAV capsids** - (a) Negative-stained EM images for AAV1, AAV2, AAV5 and AAV8. AAV particles were heated at 37, 45, 65 and 75°C [281]. (b) Thermal stability of rAAV1-rAAV9 and rAAVrh.10. (A) Coomassie blue stained SDS-PAGE (left) and negative-stained EM (right). (B) Thermal profile for each serotype mentioned [282].

pH ranging from 5.5 to 8.5 [284].

Biochemical assays with antibodies recognising intact or denatured capsids showed that AAV2 is less thermostable than AAV8 [281]. Figure 5.27(a) shows images using negative-strain electron microscopy for which the amount of AAV8 capsids is greater than AAV2 capsid at 65 °C. Bennett *et al.*, [282] also studied the thermal stability of several serotypes of AAV. They showed that AA2 is less stable than AA1, AAV3-9 and AAVrh.10 [Fig. 5.27(b)]. Comparing the stability between AAV8 and AAV9, we see that AAV9 is more stable than AAV8. More interestingly, common AAV ancestors of AAV8 and AAV9 also displayed a gradual reduction of thermal stability [285].

Recently, the mechanical properties of AAV2 capsids have been characterised by nanoindentation experiment using the AFM [183]. The high-resolution AFM images identified the capsid orientation, and then the analysis was performed on the 2-fold symmetry axis capsid. [Fig. 5.28(a)]. Experimental analyses suggest that the encapsulated genome is not a determinant of the capsid stability since the resistance to a deformation between each type (ssAAV, scAAV or empty capsid) was negligible, and the stiffness value obtained was 1.5 N/m approximately. However, the quantitative statistical analysis has shown that the dsDNA containing capsid had greater resistance to deep deformation. Consequently, the experimental and the statistical analysis does not show the same behaviour [183]. Supporting the experimental results of Zeng *et al.*, Bennett *et al.*, [282] shown that the encapsulated genome is not a determinant of AAV capsid thermostability.

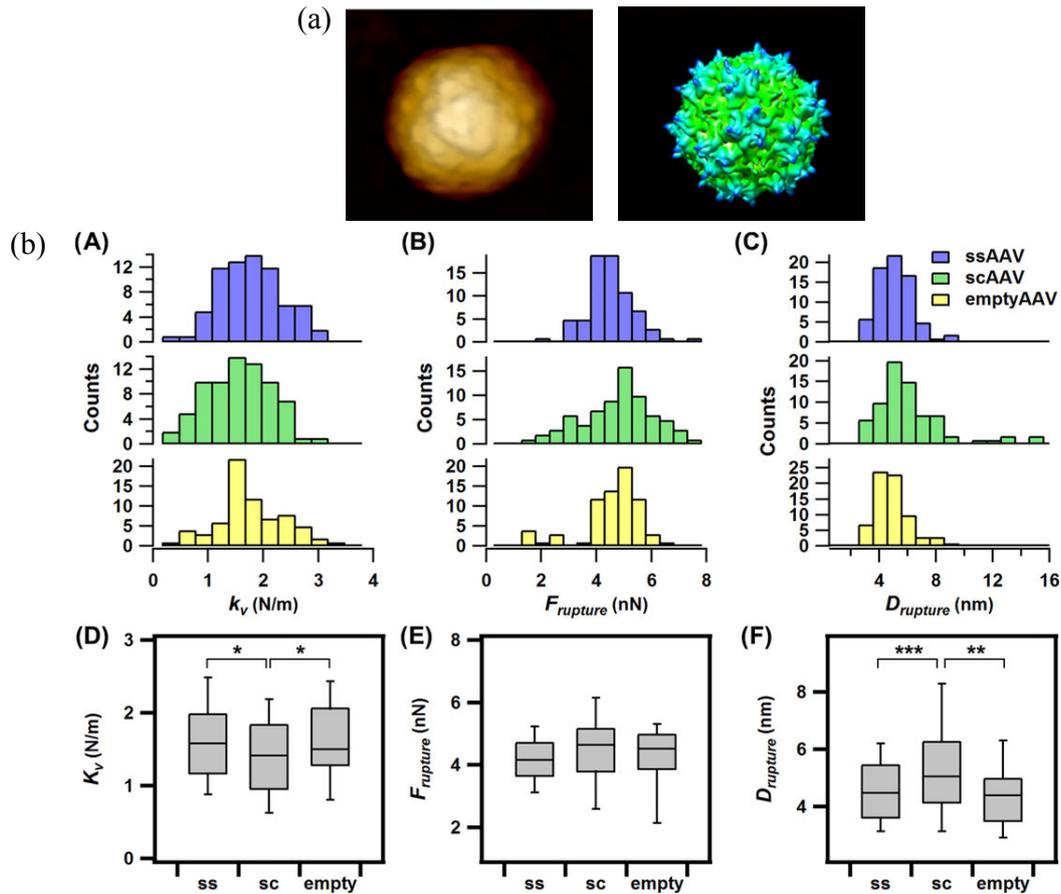


Figure 5.28: **Mechanical properties of AAV2 associated with 2-fold symmetry axis** - (a) AFM image showing the 2-fold symmetry axis of AAV2 capsid. (b) Capsid Stiffness, rupture force and their respective rupture diameter. Three cases were analysed, capsid containing ssDNA, scDNA or when the capsid is empty. Adapted from [183].

Conversely, Horowitz *et al.*, [260] show that the scAAV vectors required significantly higher temperatures to uncoat than when ssDNA is packaged. Then, the stability of the capsid depends on the nature of the DNA [Fig. 5.29(a)] as well as the size. For example, as the size of ssDNA increases, the ssAAV capsid thermal stability decreases [Fig. 5.29(b)].

5.7 Hepatitis B virus HBV

The hepatitis B virus is an enveloped DNA virus that belongs to the hepadnaviridae family [286], and their genome is a partially double-stranded circular DNA (rcDNA) where the viral polymerase is attached. The genetic material is inside an icosahedral capsid covered by a lipid membrane of cellular origin. Chronic infections caused by the

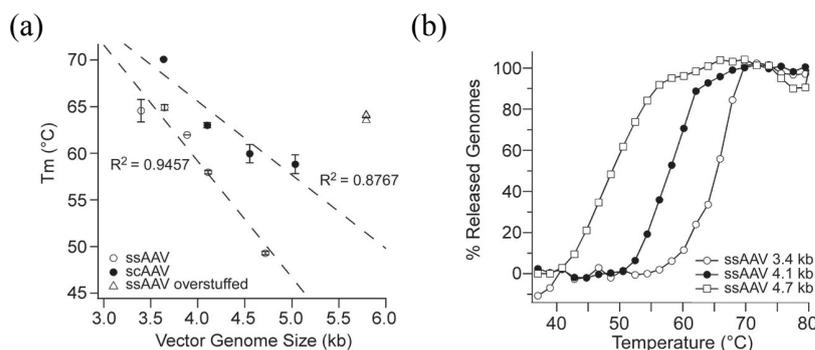


Figure 5.29: **AAV2 capsid stability as a function of genome length and self-complementarity** - (a) Comparison of ssAAV, scAAV, and oversized ssAAV vectors shows an inverse linear correlation of temperature T_m with vector genome size. (b) Percentage of released genomes for different vector genome lengths. The viral vectors were heated in 2° increments and held for 5 min at each temperature prior to acquiring fluorescent signal. Adapted from [260].

HBV virus are a significant public health problem worldwide and are the leading cause of liver cancer. Liver disease induced by HBV infections can be further aggravated by superinfection with Hepatitis D Virus HDV, a Satellite RNA Virus that depends on HBV to infect the host [287].

In particular, a major obstacle for the eradication of HBV is the persistence of the viral genome into the nucleus of the infected cell as a circular double-stranded DNA episome, called cccDNA (for circular covalently closed DNA). The cccDNA serves as the sole template for transcription of the viral pre-genomic RNA (pgRNA) and other subgenomic mRNAs, finally leading to the formation of new infectious particles [288, 289]. Thus, to achieve an HBV cure, it is necessary to silence the cccDNA transcriptionally or to be able to eliminate established HBV episomes or prevent their formation [290].

Several studies are currently conducted to understand how cccDNA is formed and maintained *in vitro* in human hepatocytes, and *in vivo* to achieve these aims. In particular, significant efforts have been made these last years to develop *in vivo* models of HBV infection consisting of immunodeficient mice engrafted with human hepatocytes or of immune-competent mice, in which the HBV genome has been introduced into murine hepatocytes via a viral vector such as those derived from Adenovirus or Adeno-Associated Virus [291, 292]. In this context, and as we explained in the previous section, the characterisation of the HBV capsids physical properties can play an essential role in obtaining effective treatment.

5.7.1 Viral cycle of HBV

Here, we will briefly describe the HBV viral cycle [293]. The viral cycle begins when the virions interact with the host cell through specific receptors found on the membrane surface [294–302], and the binding is probably irreversible [Fig. 5.30(1,2)]. The internalization mechanisms and their host factors remain unclear, where it has been proposed that this virus enters via endocytosis or by fusion with the plasma membrane [Fig. 5.30(3a,2b)]. If the entry is by endocytosis, the endocytic vesicle fuses with the virus lipid layer, and then the nucleocapsid is released into the cytoplasm [Fig. 5.30(3b)]. A recently published review discussed this problem and suggested that endosomal fusion seems to occur in late endosomes; however, the factors responsible for triggering endosomal fusion have not yet been identified [303]. Once the nucleocapsid is released, it is transported to the nucleus through microtubules [Fig. 5.30(5)] [304]. The nucleocapsid is trapped in a complex nuclear pore (nuclear basket) [305, 306] and then the rcDNA is released into the nucleoplasm [Fig. 5.30(6)] [307]. The liberated rcDNA is converted into covalently closed circular DNA cccDNA, the stable episomal transcriptional template for HBV messenger RNAs mRNAs [289].

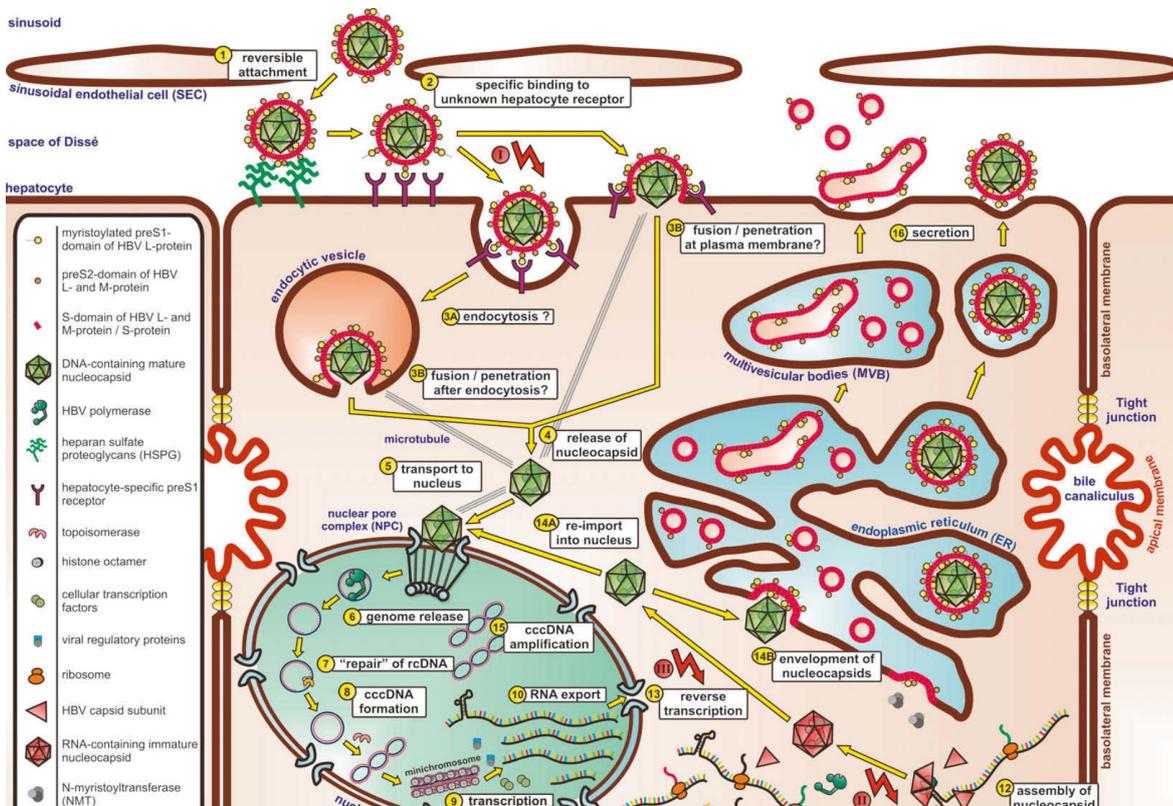


Figure 5.30: The replication cycle of hepatitis B virus - See details in the main text [308].

Once the rcDNA is released, the P protein anchored at the 5' end of the negative chain is removed. The deproteinisation reaction can be carried out by a host cell repair enzyme that removes complexes covalently or by a nucleolytic pathway mediated by endonucleases [Fig. 5.30(7)]. Next, both linear strands are joined mainly by host cellular DNA ligase and after an extension of the positive strand; thus, the cccDNA is formed. The cccDNA will then organise itself into a viral minichromosome [Fig. 5.30(8)]. HBV cccDNA can be epigenetically modified to regulate viral replication and the expression of viral genes. The cccDNA is used as a template for HBV replication and is then transcribed into HBV RNAs [Fig. 5.30(9)]. Subsequently, HBV RNAs are translocated into the cytoplasm [Fig. 5.30(10)].

The pre-core mRNA is first translated into a pre-core polypeptide. At the same time, a part of the pre-genomic RNA pgRNA is translated into core protein and P protein, and the other parts of pgRNA are encapsulated in the nucleocapsid and serve as templates for viral replication [Fig. 5.30(11)]. Next, the nucleocapsid is assembled [Fig. 5.30(12)]. The pgRNA is converted to rcDNA through a reverse transcription [Fig. 5.30(13)]. The nucleocapsids can shuttle back to the nucleus [Fig. 5.30(14a)] to maintain a relatively stable pool of cccDNA [Fig. 5.30(15)], and other nucleocapsids are enveloped by HBV envelope glycoproteins [Fig. 5.30(14b)]. Finally, the enveloped virion is secreted from the cell. [Fig. 5.30(16)]

5.7.2 HBV genome

The HBV contains a relaxed-circular DNA rcDNA genome of approximately 3.2 kb in length with a complete minus strand and an incomplete plus-strand [Fig. 5.31(a)] (review [309]). The viral genome encodes four parts of a reading frame that can be translated ORFs: C, P, S and X, which result in different RNAs with a common poly-A site.

The rcDNA is converted into covalently closed circular DNA cccDNA in infected cells, and cccDNA produces HBV RNAs of different lengths [Fig. 5.31(b)]: (i) 3.5 Kb RNA produces the protein product from C and P. (ii) 2.4 Kb RNA is translated into L-HBs. (iii) 2.1 Kb RNA synthesizes the other two surface antigens, M-HBs and S-HBs. (iv) 0.7 Kb RNA produces HBx. Four promoters regulate transcription for preS1, preS2, core, and X, and two enhancers, which is mediated by host RNA polymerase II machinery-dependent transcription [310]. The HBV RNAs coding the following proteins [Fig. 5.31(c)]:

- (i) The structural proteins, HBsAg, are large, medium, and small forms. The three

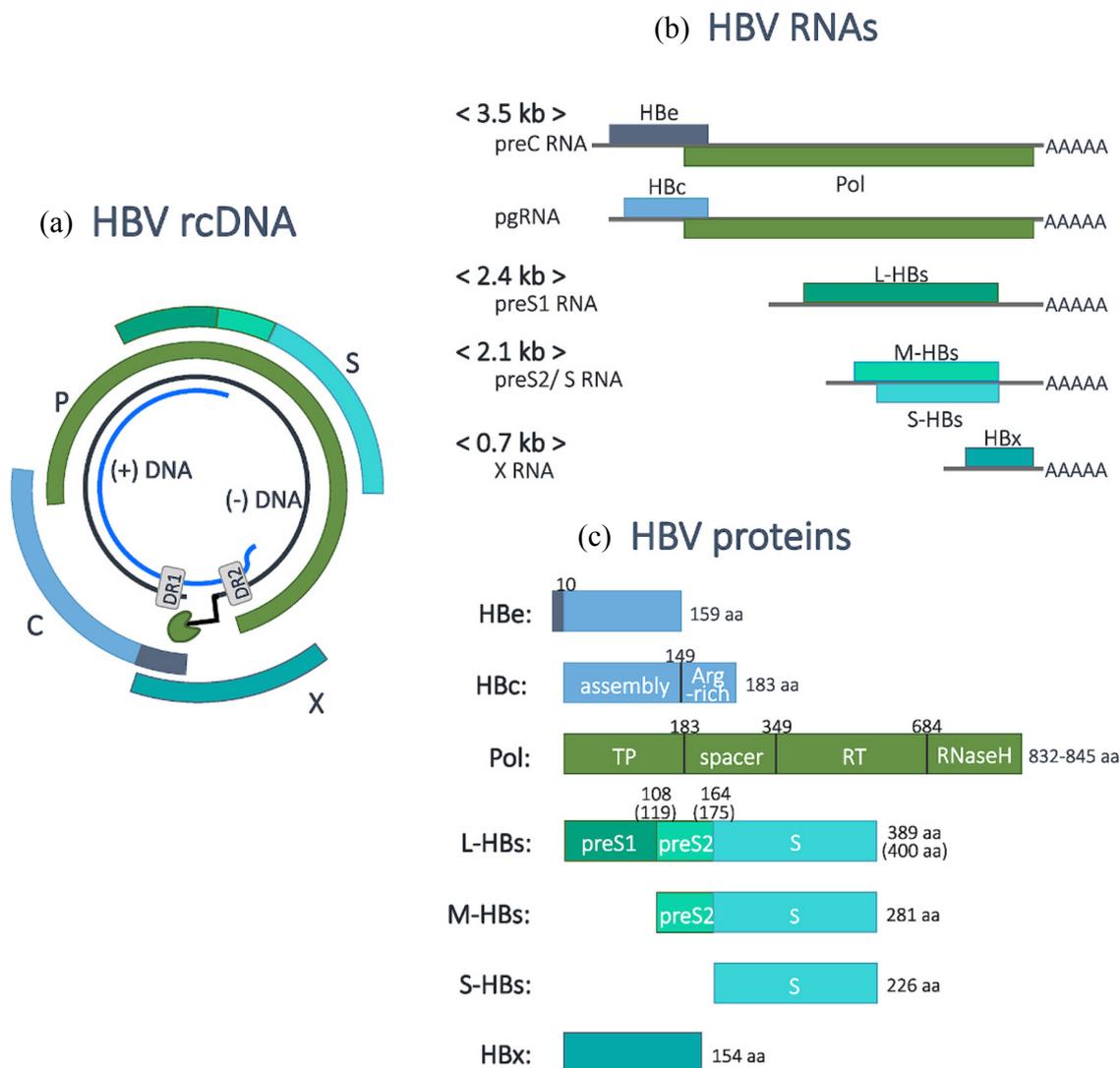


Figure 5.31: **Schematic representation of the structure of HBV genomic DNA, RNAs, and proteins** - (a) HBV genomic rcDNA and the encoded ORFs (C, P, S, X). (b) HBV RNAs (grey lines) produced by cccDNA transcription and the proteins (boxes) produced from the RNAs. The RNA lengths and the RNAs names are shown on the left, (c) HBV proteins and the domain structures. Amino acid numbers and the lengths are shown above the box and on the right, respectively [309].

HBs proteins share the common C-terminal S region. Also, M-HBs and L-HBs carry an extended region, called the preS2 region, at S region N-terminus. L-HBs further includes an extension preS1 region, which is essential for receptor binding during virus entry [295, 296, 301, 302], at the N-terminus of the preS2 and S regions.

(ii) The HBc or Cp antigen has an N-terminal assembly domain NTD is constituted by amino acid 1-149, and it is involved in the self-assembles into an icosahedral

capsid consisting of 90 or 120 dimers. The C-terminal functional domain is involved in packaging the 3.5 Kb viral pre-genomic RNA pgRNA associated with Pol [311].

(iii) HBe or external core antigen is produced by the translation of 3.5 Kb preC mRNA, having an extended 5' upstream region of the C gene and subsequent cleavage of the protein product at its C-terminus. The HBe is a hydrophobic region that forms a signal peptide that allows the protein to transit through the Golgi to the cellular surface. During the transport, the signal peptide and the basal tail are removed [312].

(iii) The polymerase is the most significant HBV protein, consisting of four domains: (1) the terminal protein (TP) domain, which is essential for binding to pgRNA and serves as a protein primer to initiate minus-strand DNA synthesis. (2) the spacer domain, whose function has not been well defined. (3) the reverse transcriptase (RT) domain has DNA elongation activity for reverse transcription and DNA-dependent DNA polymerisation. (4) the ribonuclease H (RNaseH) domain, which digests pgRNA after reverse transcription.

(iv) The transcriptional transactivator HBV X protein (HBx), which controls HBV transcription and transcription from cccDNA [313, 314]. It also plays a role in the development of HBV-related hepatocellular carcinoma.

5.7.3 Structure of HBV capsid

Dane *et al.*, [315] were the first visualising via an electron microscope the Hepatitis B Virus particles. At least three types of HBV capsids are observed in the serum of infected patients (review [309]). Virions, also called Dane particles, consist of a lipid membrane with three viral surface antigens HBs, large L-HBVs, middle M-HBs and small S-HBs surrounding a nucleocapsid composed of hepatitis B core protein HBc, viral polymerase (Pol), and viral genome DNA [Fig. 5.32(a)]. This virion presents spherical structures of 42 nm in diameter. Other smaller particles (22 nm in size) were identified in patient serum, which is much more abundant, including subviral particles (SVPs) that lack the nucleocapsid and are thus non-infectious [Fig. 5.32(b)]. Other non-infectious particles are produced during the infection process, including enveloped particles containing viral RNA instead of the viral genome and envelope-less particles (naked nucleocapsids) [Fig. 5.32(c)] [316].

Capsid assembly is initiated by rapid dimerisation of 2 monomer Cp through hydrophobic interactions forming a four-helix cluster. The Cp dimers are then trimerised in a slow reaction, followed by the association of these Cp hexamers generating the

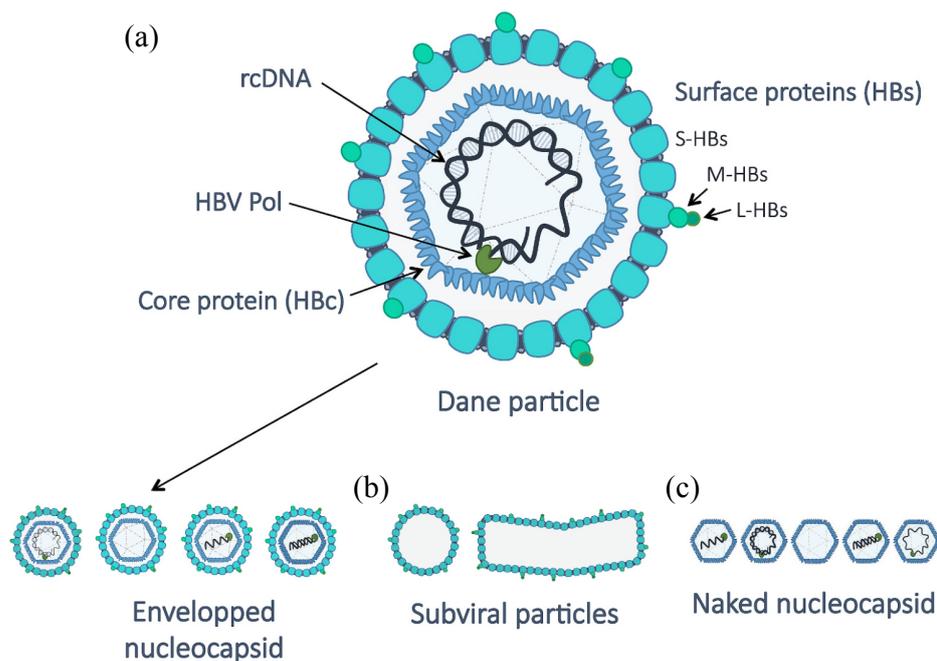


Figure 5.32: **Schematic representation of HBV particles** - (a) Infectious HBV virion (Dane particle). (b) Non-infectious HBV particles, including enveloped capsids containing immature DNA/RNA, subviral particles (sphere and filament), and naked nucleocapsids [309].

capsids [318–323]. The core proteins are organising into 90 or 120 dimers to form capsids with a $T = 3$ symmetry or in $T = 4$ symmetry capsid, respectively [Fig. 5.34]. The $T=4$ capsid is the major assembly product (95%) with a small amount $T=3$ capsid observed in infected human livers, recombinant *E. coli* capsids, and in vitro assembled recombinant HBcAg subunits [324–328]. The capsids appear to have a spherical shape. The capsid is 34 nm in size between two protrusions, and the inner layer is 24.4 nm for a capsid of $T=4$ symmetry [148].

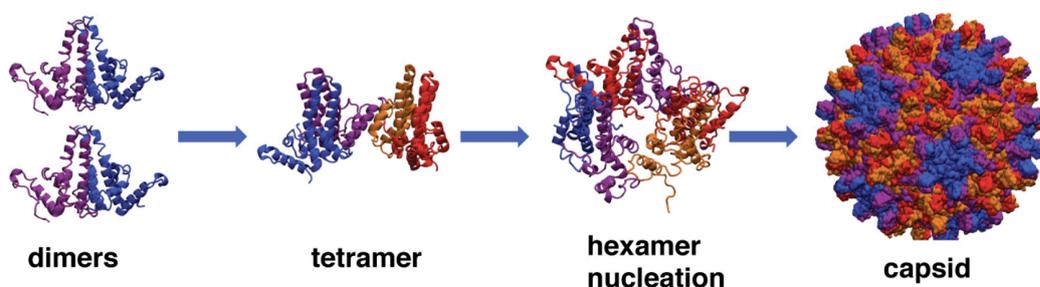


Figure 5.33: **Assembly of HBV capsid** - HBV capsid assembly process based on experimental data and mathematical modeling [317].

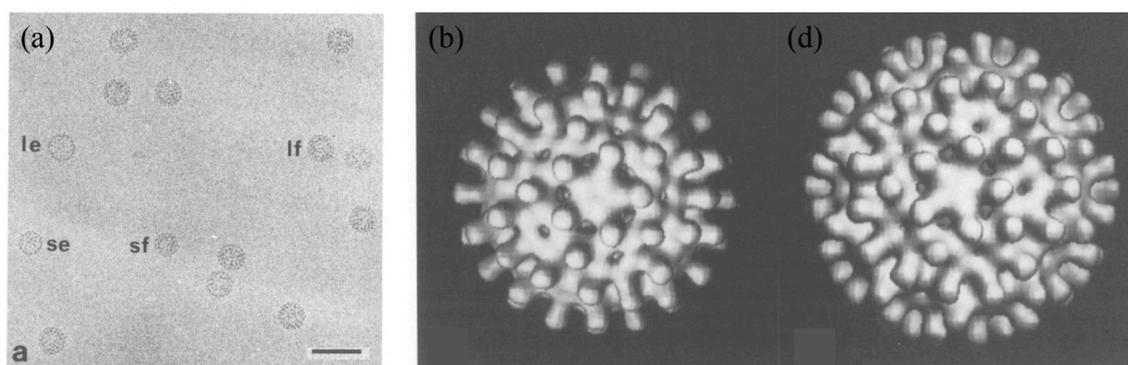


Figure 5.34: **Cryo-electron micrographs of HBV capsids** - (a) Electron cryomicrographs showing particles assembled from intact HBcAg protein. Different particles are visualised: large empty *le*, small empty *se*, large full *if* and small full *sf*. Scale bar 50 nm. (b) Surface representations of the computed three-dimensional maps of small empty (a) and large empty (b) core particles, showing the view along a 5-fold axis. The small particle shows 90 protruding spikes (dimers), arranged with dimer clustered $T=3$ icosahedral symmetry. The large particles shows 120 protruding spikes (dimers), arranged with dimer clustered $T = 4$ icosahedral symmetry. Adapted from [324].

5.7.4 HBV mechanical properties

Roos *et al.* [329], measured the mechanical response of the HBV capsid, for both $T=3$ and $T=4$ symmetry, via the atomic force microscopy AFM. Figure 5.35(a) shows the AFM images for both capsids. Most studies on the capsid deformation have revealed a remarkable linear elastic force-indentation response. However, HBV capsids reveals a continuous nonlinear behaviour at large indentations [see experimental data in figure 5.35(b)]. Due to the significant differences in the force curves in [329], the authors proposed a capsid orientation dependences for empty capsids, unlike the linear curves shown by Carrasco *et al.*, [156].

Also, Roos *et al.*, [329] performed simulations of capsid nanoindentations for two, three and five-fold symmetry axis, allowing the experimental data to be separated by matching with the simulation [Fig. 5.35(b)]. Finally, the results suggest that the Hepatitis B Virus capsids $T=4$ are stiffer and similar at the two and five-fold axis [Fig. 5.35(c,right)]. For $T=3$, the three and five-fold symmetry axis are stiffer and similar to the two-fold symmetry axis [Fig. 5.35(c,left)]. Simulations carried out by Arkhipov *et al.*, [176], belonging to the same Roos group, found reached the same conclusion.

Other previous studies did not show orientation dependence since it was not possible to identify the capsid orientation through the AFM images [175, 176]. Particularly,

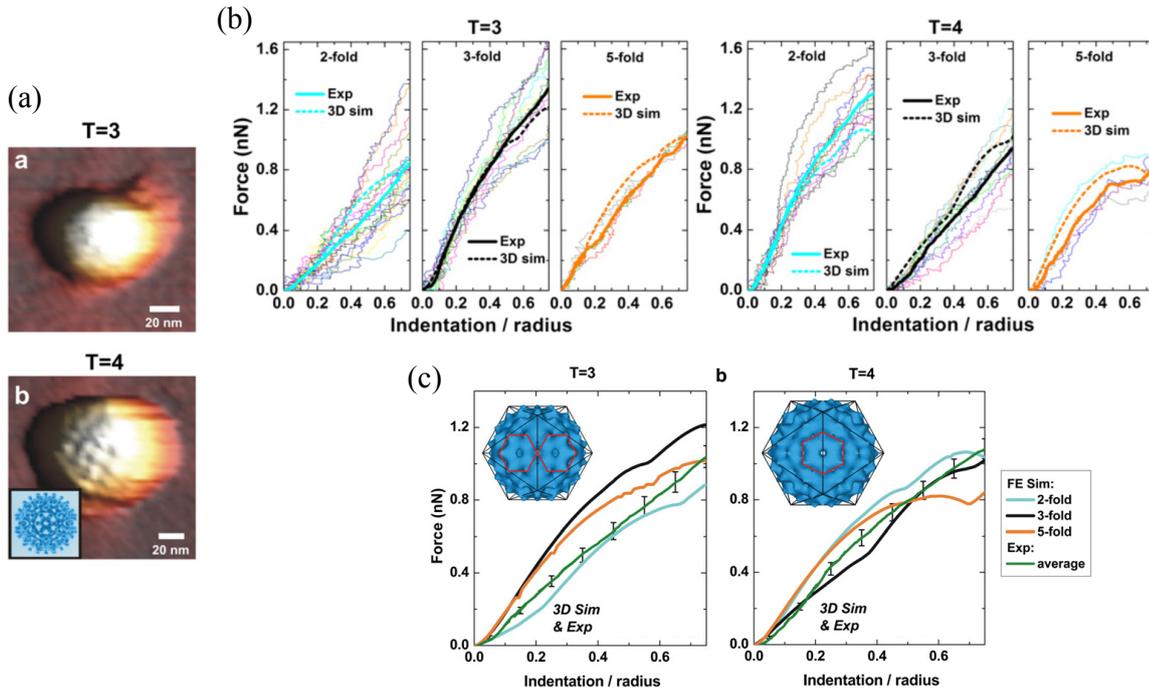


Figure 5.35: **AFM nanoindentation experiments for HBV capsid** - (a) AFM image of a T=3 (top) and T=4 (down) capsids. (b) Experimental curves grouped by apparent orientation for (left) T=3 and (right) T=4. Individual experimental curves are grouped according to the binning procedure based on the curves obtained from the 3D FE model. (c) Curves from the FE simulation separated along symmetry axes for T=3 (top) and T=4 (down) capsids. Also, the figure shows the average experimental force curve in green [329].

Utrecht *et al.*, [175], through their force curves, found a stiffness value of 0.1 N/m for T=3 and T=4.

5.8 Our work in this context

Viruses are nanosized, genome-filled protein containers with remarkable thermodynamic and mechanical properties from a physical point of view. Looking at the viral cycle, we can make the following hypothesis: Viral capsids need to be sufficiently stable to protect their genome in the extracellular environment but unstable enough to release their genomes into the cell nucleus. From this last statement, we can infer that the capsid mechanical properties play an essential role in the viral genome containing and ejection. Several recent works support this idea, pointing out the correlation between the mechanical properties of viruses and their ability to infect cells [330–332] but showing discrepancies in the value of measured stiffness. On the other hand, the

causal link between mechanics and infectivity remains unclear.

Next sections, we attempt to answer the following questions.

- Does the indentation position change the measured stiffness of the capsid? If the answer to this question is yes, can we define a protocol to perform a good estimation of this value?
- The capsid structure is not homogeneous; it has vertex, faces and edges. This structure depends on the protein number constituting the capsid. Are the mechanical properties depending on the capsid structure? For a specific viral capsid, could the measured stiffness be affected by the icosahedral geometry orientation?.
- Are the mechanical properties of the capsid relevant during the capsid uncoating process?
- HBV and AAV have approximately the same capsid size, but they have different nuclear entry. Transport receptors assist the transport of HBV while AAV crosses the nuclear membrane through a nuclear pore complex. Is it possible to describe this process through the mechanical properties of each capsid?
- ITRs genome of AAV play an essential role in gene therapy. Can the understanding of the ITR conformation increase the performance of gene therapy and the possible clinical applications?
- Is it possible to answer all the previous questions by using the AFM?

Chapter 6

AFM protocol to determine the viral capsid stiffness

Index

Summary	129
6.1 Capsid purification	130
6.2 Capsid deposition in mica	131
6.3 AFM imaging	134
6.4 Image analysis in a population of capsids	135
6.5 AFM image analysis for a single capsid	139
6.6 Methods to determine the capsid stiffness from z-F curves	144
6.6.1 Method 2 spring in series	144
6.6.2 $F(\delta)$ method	145
6.7 Nanoindentation on the surface	149
6.7.1 Force curve slope on a rigid surface gives us the cantilever stiffness	150
6.7.2 Indentations on the surface using a higher indentation force .	150
6.8 Nanoindentation on the capsid	152
6.8.1 One indentation point on the capsid centre	152
6.8.2 Indentation spatial profile over the capsid	154
6.9 Non-linear effects on force curves	157
6.10 Capsid average profile	161
6.11 Conclusion	161

Summary

This section will explain how the mechanical and morphological properties of capsids were obtained using AFM nanoindentation. The following sections show how the capsids are purified and the deposition protocol that allows visualising the capsids through the AFM images. Then, we show the protocol that allows the extraction of the morphological properties from those AFM images. Additionally, we will explain the protocol to do nanoindentation experiments through a single point in the capsid centre and an indentation profile. Finally, we analyse the force curves for the two protocols.

6.1 Capsid purification

The purifications of AAV capsids were carried out by Anna Salvetti's team at the International Research Center in Infectious Diseases (CIRI) at ENS de Lyon. The production of viral particles is based on the transfection of plasmids encoding the proteins (structural and regulatory) of viruses and the viral genome [Fig. 6.1].

Three types of plasmids are incorporated into human cells of the HEK293 line by ion shock provided by the Calcium Phosphate transfection method [Fig. 6.1(b,c)]: (i) The plasmid carrying the AAV-GFP genome encoding the transgene contains the sequence,

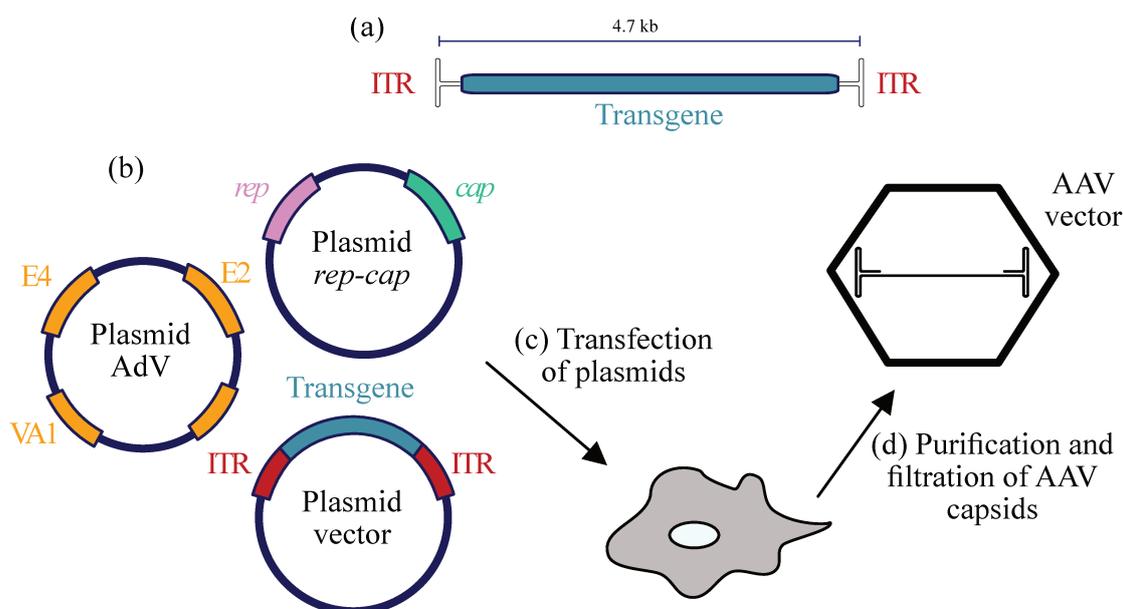


Figure 6.1: Diagram of AAV production process *in vitro* - See details in the main text.

which will be incorporated into the capsid as a viral genome [Fig. 6.1(a)]. (ii) The Rep-Cap plasmid encoding structural proteins (Rep) and regulatory proteins (Cap). (iii) The helper plasmid containing adenovirus genes encoding regulatory proteins must produce capsids by the transfected cell. After three days of incubation, the cells are suspended in buffer and then destroyed by freeze-thaw cycles. The lysate obtained is then purified by two successive density gradients of caesium chloride (CsCl). The density of correctly assembled AAV particles is well characterised. The CsCl gradient allows the recovery of functional capsids and removes cellular contaminants and malformed or empty capsids. Dialysis is performed to remove all traces of CsCl. At the end of these purification steps, the AAV particles are suspended in a PBS + buffer solution (phosphate-saline solution with divalent Mg^{2+} and Ca^{2+}). Their concentrations, measured by qPCR (Quantitative Polymerase Chain Reaction), range from 1010 vg/mL to 1013 vg/mL depending on the cell production efficiency and the more or less selective quality of the purification filtering [Fig. 6.1(d)]. The helper plasmids used for production were pDG, pDG8 and pDG9 from AAV2, AAV8 and AAV9, respectively.

6.2 Capsid deposition in mica

To study the capsids in a physiological environment, we modulated different chemical parameters of the buffer solutions. The goal here is to immobilise the capsids onto mica. The adhesion to the surface causes a deformation on the capsid; therefore, we must find an optimal condition not to deform the capsid too much. It remains glued enough so that it does not detach when the cantilever tip press on the capsid. Divalent cations play a bridge role between the negatively charged surface and the particle. The most common cations to adhere to viral particles are $MgCl_2$ and $NiCl_2$ [333,334], where $NiCl_2$ has been shown to provide stronger binding of the DNA in mica for measurements in liquid [335].

The cases analysed for AAV8 capsid are explained below:

1. Dried sample and PeakForce in air: the AAV solution was diluted in Tris 10 mM (ph=7.4) + (i) 5 mM $MgCl_2$, (ii) 1 mM $NiCl_2$ or (iii) 5 mM $NiCl_2$ [Fig. 6.2(b-d)]. Immediately after, 5 μ l of the solution was deposited onto mica disks previously cleaned using adhesive tape. After 10 min of incubation to favour adhesion, the sample was washed with 1 ml of ultrapure water and dried gently using nitrogen flow. AFM images show some smaller objects adhered to the surface, where the sample containing $NiCl_2$ 1 mM shows a higher amount of these objects [Fig. 6.2(b-d)]. The capsid height has been calculated (details

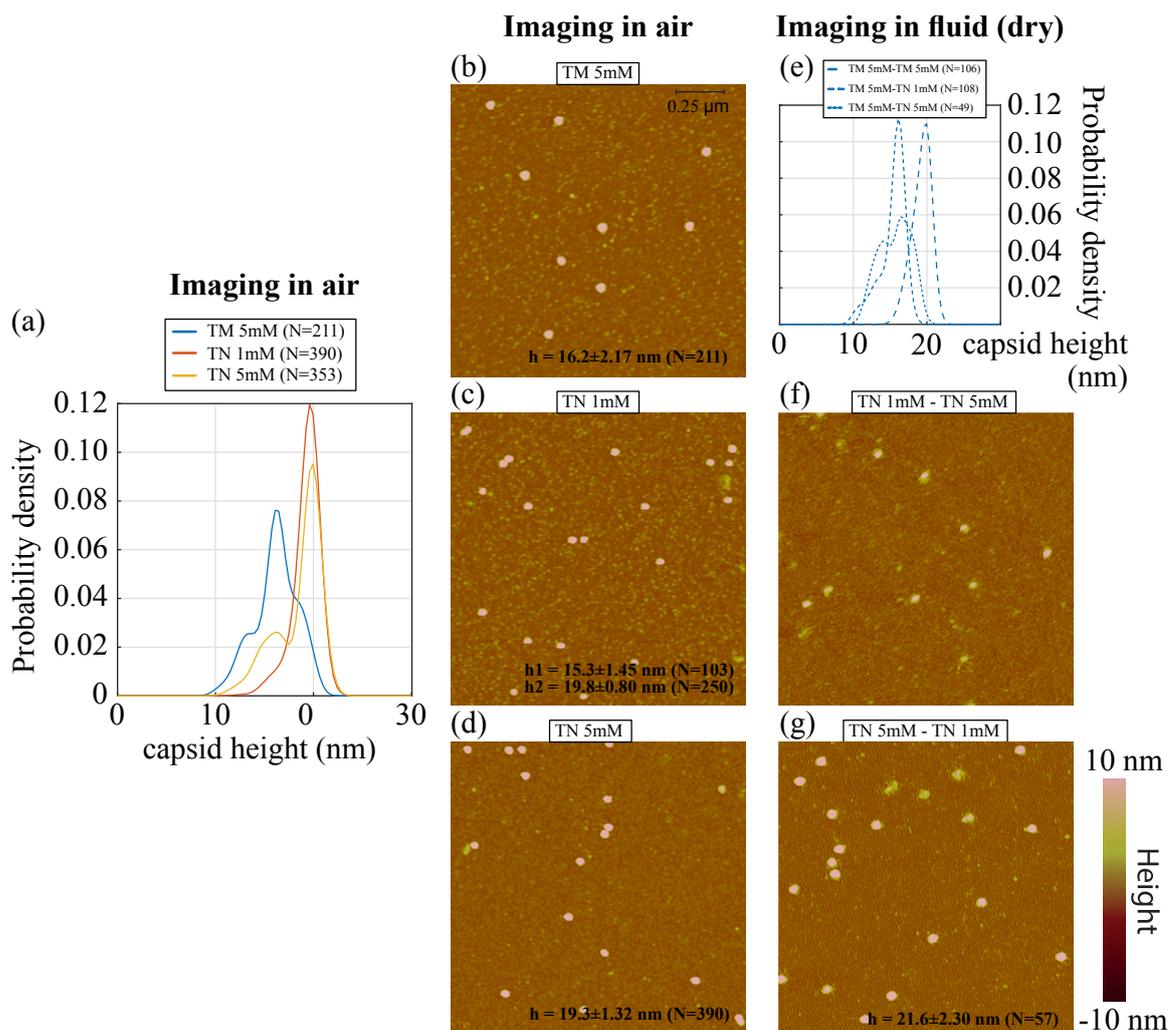


Figure 6.2: **Deposition conditions when the sample was dried** - AAV solution was diluted in Tris 10 mM (ph=7.4) + (i) 5 mM MgCl₂, (ii) 1 mM NiCl₂ or (iii) 5 mM NiCl₂. Immediately after, 5 μ l of solution was deposited onto mica disks previously cleaned using adhesive tape. After 10 min of incubation, the sample was washed with 1 ml of ultrapure water and dried gently using nitrogen flow. (a) The height of the capsids were determined from the AFM images of the conditions: 5 mM MgCl₂ (b), 1 mM NiCl₂ (c) and 5 mM NiCl₂ (d). (e,f,g) shows the probability density of the capsid height or AFM images when buffer was added to be able to analyze capsids in liquid.

on the analysis of the images in a later section). We noted that the magnesium generates a greater adhesion on the capsids generating a decrease in the apparent height [Fig. 6.2(a)]. The determined capsid height is similar to that reported in [280].

2. Dried sample and PeakForce in fluid: once the sample has been deposited in mica, according to protocol 1. We additionally added 30 μ l of 5 mM MgCl₂,

1 mM NiCl₂ or 5 mM NiCl₂ to each condition. Therefore, we will have nine samples to analyse [Fig. 6.2(e-g)]. When analysing viral capsids in a liquid medium, they will tend to behave in their natural state; therefore, it is expected that the height measurement will be close to the real capsid height. In case (i), we do not see variations if we add NiCl₂ in comparison with MgCl₂ that generated an increase in the apparent height. As we mentioned, nickel has a stronger interaction force. Therefore, although the capsid wants to return to its natural state, nickel does not allow it since the interaction force is stronger [Fig. 6.2(e)]. In particular, when the sample was prepared with 1 mM NiCl₂ in air [Fig. 6.2(c)], and then we added 5 mM NiCl₂, the capsids broke [Fig. 6.2(f)], consistent with recently mentioned. However, the opposite case did not generate a break in the capsids. The samples prepared with 5 mM NiCl₂ sufficiently adhered to the capsids in mica [Fig. 6.2(d)], but by adding 1 mM NiCl₂, that is, a less strong buffer, the capsids remained intact.

3. Not dried sample and PeakForce in fluid: the AAV solution was diluted in Tris 10 mM (ph=7.4) + **(i) 5 mM MgCl₂, (ii) 1 mM NiCl₂ or (iii) 5 mM NiCl₂**. Immediately after, 5 μ l of the solution was deposited onto mica disks previously cleaned using adhesive tape. After 20 min of incubation to favor adhesion, we added 30 μ l of **(i) 5 mM MgCl₂, (ii) 1 mM NiCl₂ or (iii) 5 mM NiCl₂** to the sample. As before, we have nine samples to analyse [Fig. 6.3]. Here we study the drying effect when the sample is prepared. We use the conditions measurements as if we had dried the sample, that is, a specific buffer to adhere the capsids in mica and another buffer (it can be the same or different) to obtain the images in liquid mode. In general, we see that the capsids height is closer to the expected value (\sim 25 nm); therefore, drying the sample generates a greater deformation caused by the mica adhesion. Unlike the TM5mM-TM5mM case (dry sample) [Fig. 6.2(e)], in this case no capsid adhered to the mica [Fig. 6.3(a)]. In the other two cases, TM 5mM-TN 1mM and TM 5mM-TN 5mM, either many more small particles were attached [Fig. 6.3(b)] or the capsids were broken [Fig. 6.3(ab)]. Although, the particles that managed to stick to the surface show a height that can be considered. On the other hand, when nickel was used to adhere to the capsids, generally good conditions were obtained for the capsids. The protocol selected was determined according to impurities adhered to the surface; therefore, TN1 mM-TN1 mM was chosen because it had fewer impurities than the other conditions.

Finally, the chosen deposition condition is: the AAV solution was diluted in Tris 10 mM (ph=7.4) + 1 mM NiCl₂, immediately after, 5 μ l of the solution was deposited onto mica disks previously cleaned using adhesive tape. After 20 min of incubation to

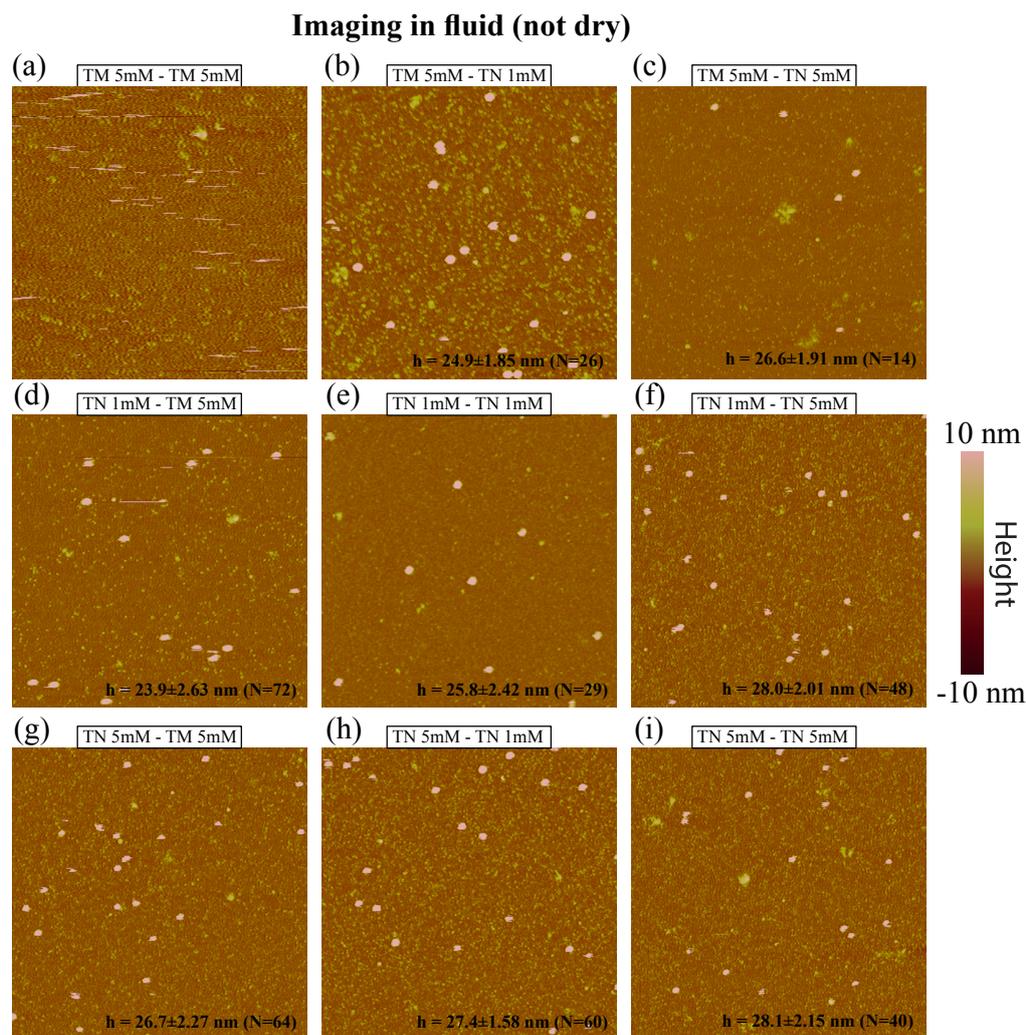


Figure 6.3: **Nondried sample deposition conditions** - AAV solution was diluted in Tris 10 mM (ph=7.4) + (i) 5 mM MgCl_2 , (ii) 1 mM NiCl_2 or (iii) 5 mM NiCl_2 . Immediately after, 5 μl of solution was deposited onto mica disks previously cleaned using adhesive tape. After 20 min of incubation to favor adhesion, we added 30 μl of (i) 5 mM MgCl_2 , (ii) 1 mM NiCl_2 or (iii) 5 mM NiCl_2 to the sample in order to visualize the capsid in liquid mode.

favour adhesion, we added 30 μl of $\text{TN}_{1\text{mM}}$ to the sample and another 70 μl of $\text{TN}_{1\text{mM}}$ in the AFM liquid imaging cell.

6.3 AFM imaging

Images were obtained using PeakForce mode in liquid. The cantilevers used, Bruker ScanAsyst-fluid-plus and Bruker SNL-10(A), have a triangular geometry with a nomi-

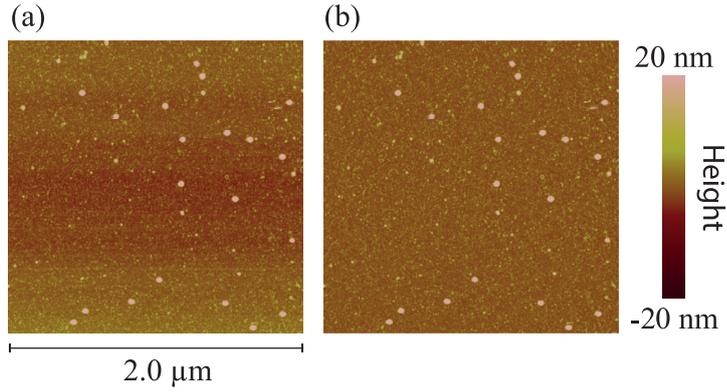


Figure 6.4: **AFM images of capsids** - (a) Original image. (b) Flattened image using NanoScope Analysis program.

nal radius of 2 nm. The nominal stiffnesses are 0.7 and 0.35 N/m, respectively; nevertheless, the cantilever spring constant k_c measured by thermal noise was around 1 and 0.4 N/m. The scanned images were squares of 2 μm side, with a resolution of 512x512 pixels and a scan frequency of 2 Hz [Fig. 6.4(a)]. The maximal force applied was 1000 pN for the three serotypes AAV2, AAV8 and AAV9. As in the previous chapter, the tilt and low noise frequency in the AFM image were removed using the NanoScope Analysis program [Fig. 6.4(b)].

6.4 Image analysis in a population of capsids

To obtain the capsid morphological properties, we applied a similar analysis image process as in section 3.3.1. The interpolation process and the applied spatial filter are the same as in DNA analysis [Fig. 6.5(b-c)]. In this case, the height threshold used was 6 nm; this value allows us to separate the capsid from the surface [Fig. 6.5(d)]. Then, we applied an area filter, where the minimum and maximum areas thresholds were 500 nm² and 2000 nm², respectively [Fig. 6.5(e)]. If there is no convolution effect between the cantilever tip and the capsid (ideal case), the area of the capsid 2D projection is approximately 650 nm² ($R_{capsid} = 12.5$ nm); therefore, the minimum area of a capsid should be around this value. The filters used until now do not select an isolated capsid; some capsids are still touching each other. To eliminate dimers or multimers of capsids, we used a parameter called the fractal number, defined as

$$F = P^2/4\pi A, \quad (6.1)$$

where P is the perimeter and A is the object area. We used $F < 0.8$ and $F > 1.2$ to removed ill-defined and agglomerated capsid, respectively [Fig. 6.5(f-g)]. Capsids that touch the edge of the image were also removed from the image. Then, we used a capsid filter; that is, we chose structures with a height greater than 10 nm. Thus, we make sure to choose only capsids. The coloured boxes represent the comparison between each step effect of the analysis with the previous image [Fig. 6.5]. Finally,

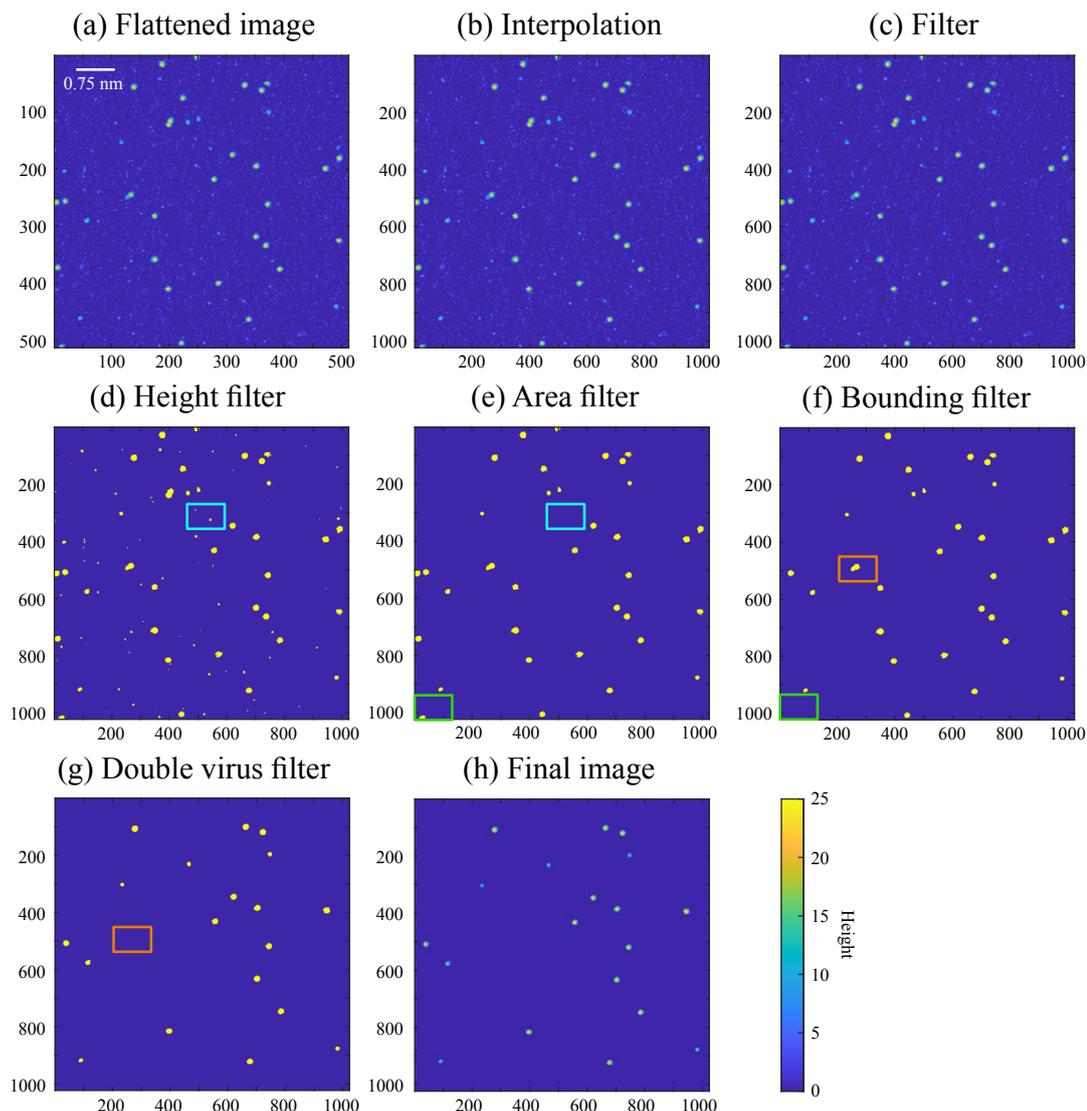


Figure 6.5: **Image analysis steps** - (a) Flattened image coming from NanoScope, (b) Interpolation and (c) noise filter performs an image [125]. The height filter (d) and area filter (e) remove all the small detection associated to noise and/or impurities of the image. (f) Bounding filter removes capsids touching the edge of the image, (g) Double capsid filter removes overlapped capsids. (h) shows the processed image. Here, all non-zero values represent the analysed capsid. The coloured boxes represent the comparison between each step effect of the analysis with the previous image.

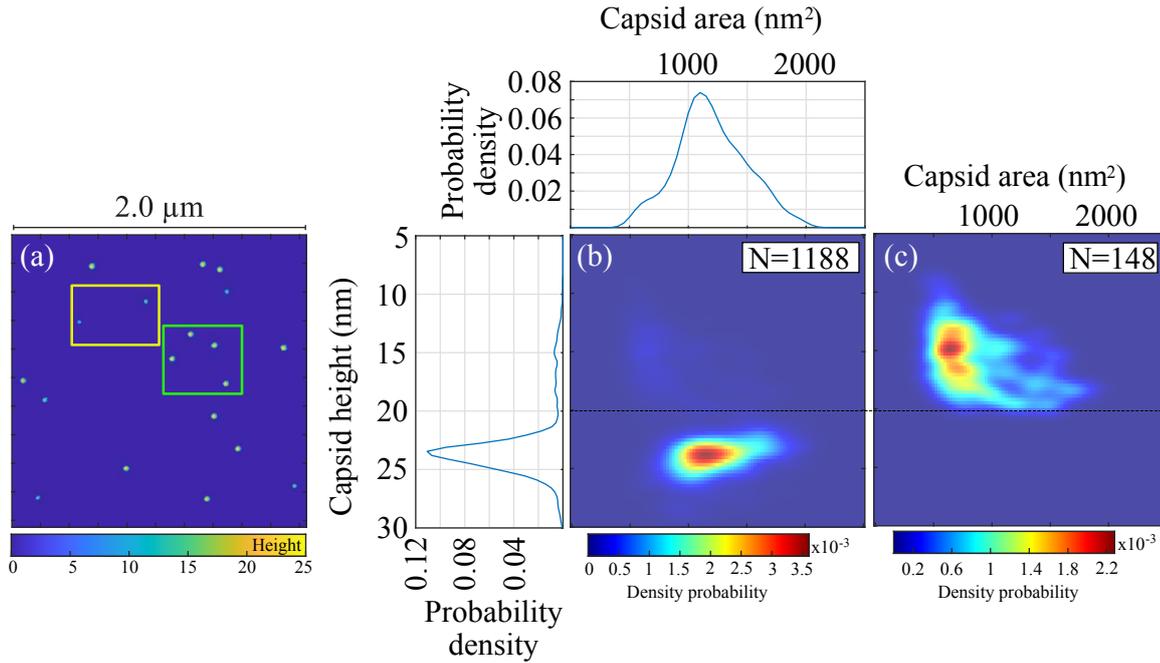


Figure 6.6: **Capsid height and area probability density for AAV8 capsid** - Correlation between the capsid viral height and area. (a) Edited image showing two capsid type. (b) Probability density map for 1188 capsids. From the profiles we note the more probable height is 23.9 ± 1.3 nm (capsids marked with a green square in (a)). The horizontal black line over 20 nm separates the capsids population in two groups with different areas. (c) Probability density of the capsids with a height smaller than 20 nm. We found 12% of the total amount of capsids form this second probability peak (capsids marked with a yellow square in (a))

we multiplied the treated and binarised image with the original image, resulting in an image with the original height values and only with well-defined isolated capsids [Fig. 6.5(h)]. Next, we calculated the different morphological properties as the height, area and volume of each capsid. To obtain the height capsid, we look at the maximum pixel value. The area was obtained from the addition of the number of pixels. The volume was obtained from all pixels value addition.

The analysis was performed over 48 images, with 1188 capsids selected according to the protocol explained above. The analysis was developed using AFM images obtained with the cantilever SNL-10. Two types of capsids are seen in the image 6.6(a), in which the smallest object has a small height. The capsids height probability density shows a peak, highly probable, for a height of 23.9 ± 1.3 nm (N=1040) [Fig. 6.6(b)]. For AAV9, it has been reported that the minimum diameter at the icosahedral 2-fold axis is 23 nm and 25 nm at the 3-fold. The maximum diameter is 29 nm at the tips of the protrusions surrounding the 3-fold axis and lies 3 nm from that axis [263]. The diameter between the tips of two protrusions for AAV2 and AAV8 are 29.4 and 30 nm,

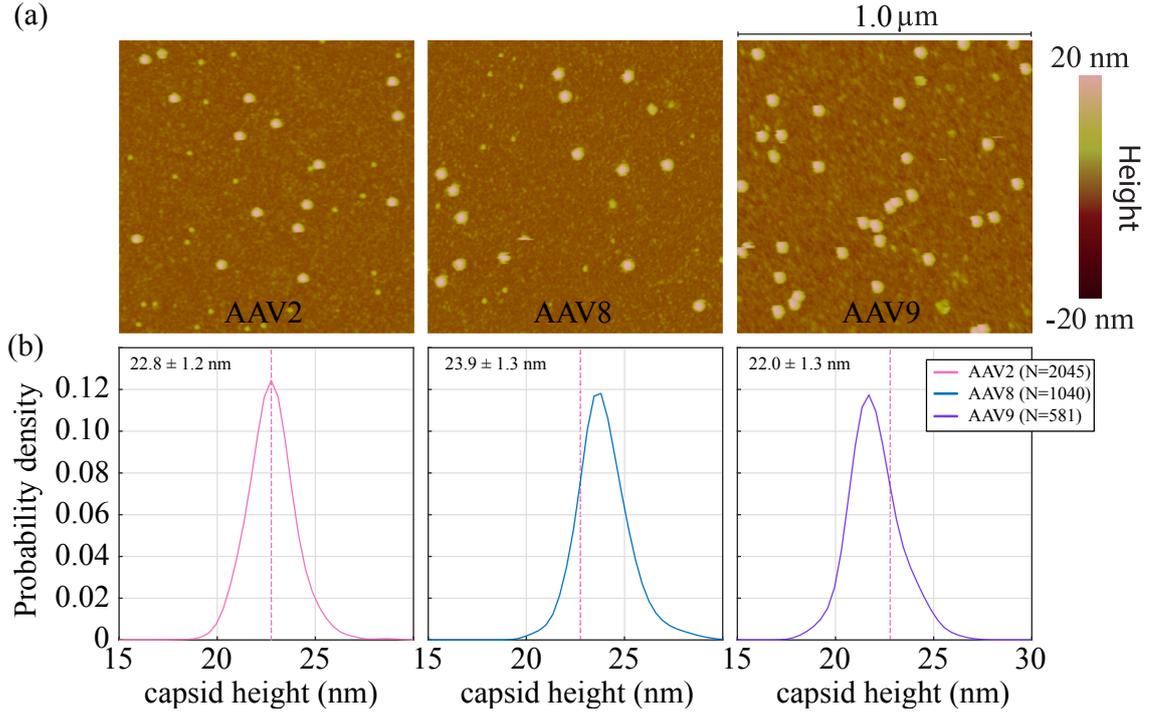


Figure 6.7: **Capsid height for AAV2, AAV8 and AAV9** - (a) AFM images for AAV2, AAV8 and AAV9 capsids. (b) Probability density of the capsid height, where the capsid AAV2, AAV8 and AAV9 have a height of 22.8 ± 1.2 nm, 23.9 ± 1.3 nm and 22.0 ± 1.3 nm, respectively.

respectively [148]. Our measure value suggests that the capsid adhesion to the surface generates a deformation on it, showing a height of 5 nm less than the expected value. Zeng *et al.*, [336] observed a 3-4 nm height decrease for Brome Mosaic Virus when the capsid was deposited on mica. The second peak, capsids with a height lower than 20 nm are empty capsids [Fig. 6.6(c)].

Subsequently, we compared the height capsid for AAV2, AAV8 and AAV9. AFM images show no visible differences, however analysing capsid height probability density, we noted the AAV8 capsid present the highest apparent height compared to AAV2 capsid, followed by AAV9 capsid. Zeng *et al.* 2017 [336] proposed that capsid reduced height measured in the AFM is due to the deformation and adhesion energy. Assuming the capsid like a sphere of radius r [Fig. 6.8(a)], the adhesion energy will deform the sphere a height Δh generating a circular contact zone of radius a [Fig. 6.8(b)]. The energy of the capsid will be the elastic energy E_{def} plus the contact adhesion energy E_{ad} , so we can write,

$$E_{total} = E_{def} + E_{ad}. \quad (6.2)$$

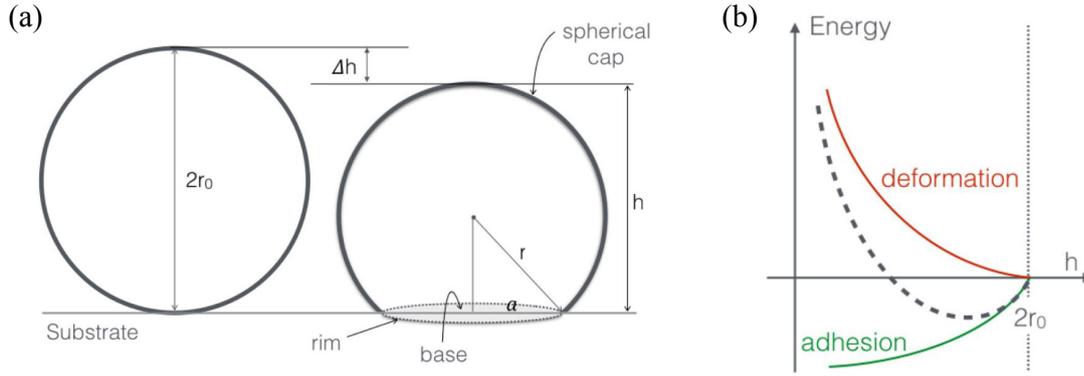


Figure 6.8: **Zeng model: capsid reduced height measured in the AFM is due by the deformation and adhesion energy** - (a) Scheme of a capsid without deformation on a substrate and another that has a decrease in height due to the adhesion and deformation energy (b).

Figure 6.8(b) shows the behaviour of each term and the total energy (segmented line) as a function of the deformed capsid height h . We used the same buffer to adhered the AAV2, AAV8 and AAV9 capsids to the mica surface, and we could suggest that they have similar adhesion energy. The total energy capsid is mainly due to the deformation energy, i.e. the reduced height should be given only by deformations energy comparing the serotypes. From the capsid height obtained from the AFM images [Fig. 6.7], and using the relation proposed by Zeng *et al.*, [336] for the elastic constant, we obtained the following relation:

$$k_v = \frac{E_{def}}{\Delta h^2}. \quad (6.3)$$

From this equation, we could infer that AAV8 capsid is stiffer than AAV2 and AAV9. We will determine these three capsid stiffness values in the later sections and see if this assumption is fulfilled.

6.5 AFM image analysis for a single capsid

Indentation experiments are generally performed on individual capsids. To extract the singular capsids morphological properties, we must perform a different image analysis from the one performed when there is a population of capsids. We will explain step by step how we developed the nanoindentation experiments and, in particular, the image analysis.

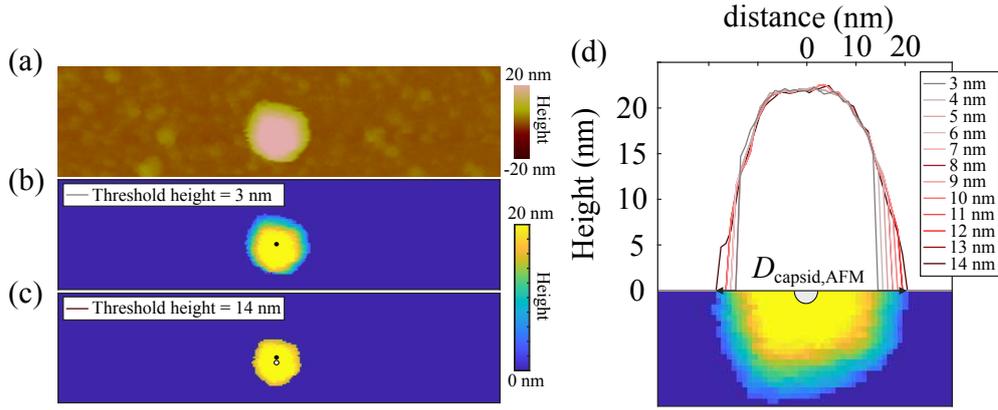


Figure 6.10: **Capsid centre from AFM images** - (a) Single capsid AFM image. (b) Edited AFM images using a height threshold of 3 nm. The black point represent the centre of mass position. (c) Idem as the case (a) but with a height threshold of 14 nm. The white point represent the centre of mass position. (d) Height profile over the capsid centre (white point in (c)) for different height threshold.

density that is not uniformly distributed. The 3 nm case has height values (light blue in the image) related to the capsid convolution effect and the cantilever tip generating a centre of mass position shift. Furthermore, the parameters used to obtain the image also play a role in the capsid shape. Figure 6.9 shows an image being scanned and a blue and red curve graph. These two curves represent a height profile for each vertical line of the image, where the red curve is the information when the cantilever goes from left to right, and the blue curve shows the information when the cantilever goes from right to left. We see that both profiles do not match when they pass over the capsid, effects generated by speed, number of image points or scanning force. Consequently, the centre of mass of the image was determined, excluding the convolution effect part. Next, we traced a horizontal line in the capsid centre to determine the capsid diameter [Fig. 6.10(d)]. We must mention that this threshold (14 nm) is used only to determine the correct center of the capsid.

To choose the right height threshold and then determine the $D_{capsid,AFM}$, we used a simple interpretation. Knowing the apparent capsid height and the geometric parameters of the cantilever tip, we could estimate the total diameter $D_{capsid,AFM}$ as can be seen in the diagram in figure 6.11(a). In large-scale AFM images, $2 \mu\text{m}$, the capsid height is determined from the highest point. According to the image parameters used, $1 \text{ px} = 3.8 \text{ nm}$, where 1 px contains 15 % of the information of the capsid ($R_{capsid} = 25 \text{ nm}$). In images with a single capsid, $1 \text{ px} = 1.2 \text{ nm}$, i.e. 1 px contains only 5 %. To have the same condition, we choose a radius of 3 px, equivalent to 15 % of the information of the capsid. The average heights for 62 capsids are shown in figure 6.11(b),

where the error bars represent the standard deviation of the data in the central region. The cumulative average height of all capsids is 22.5 nm, and the grey region represents the cumulative average height standard deviation [Fig. 6.11(b)]. Comparing the height capsid obtained from large-scale AFM images and AFM image with a single capsid, we noticed that the second case presents a reduction in the height measured. The height difference could be related to the applied force because the force is not the same in both cases.

In the ideal case, when the cantilever tips are intact ($R_{cantilever} = 2$ nm) and using a capsid height of 22.5 nm, the diameter $D_{capsid,AFM}$ determined should be approximately 37 nm [scheme in figure 6.11(b)]. Figure 6.11(c) shows that the $D_{capsid,AFM}$ values go up to 40 nm considering a height threshold of 3 nm; this case must happen when the tip does not have a radius of 2 nm. Bruker, the cantilever brand, mentions that the cantilever tip radius can go up to 12 nm. Therefore, we must take into account that the cantilever tip has not the same size every time. The minimum value of the height threshold such that the convoluted part between the cantilever tip and the capsid is not removed is 8.5 nm. To have a height threshold standard for all the images, we choose

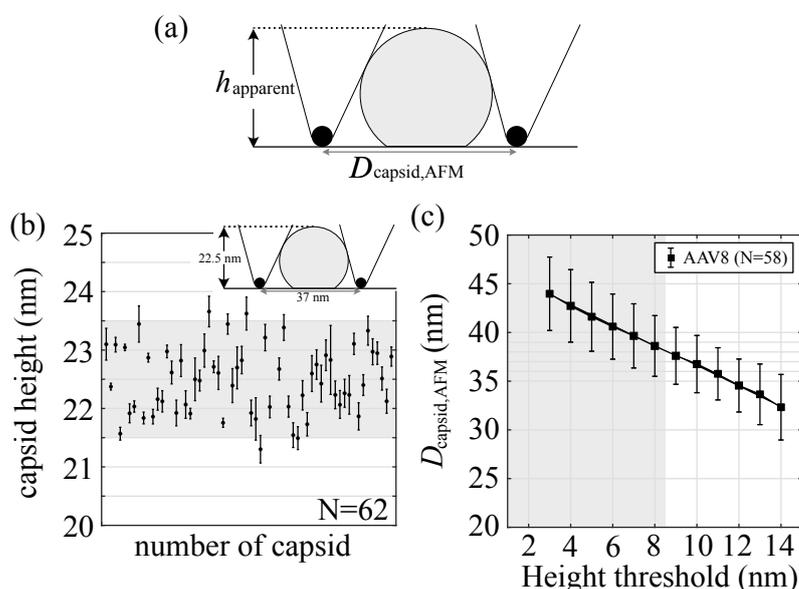


Figure 6.11: **Capsid apparent diameter determination** - (a) Schematic of a capsid with two cantilevers on each side. The distance between the centres of the two cantilever tips represents the capsid diameter ($D_{capsid,AFM}$) and $h_{apparent}$ is the capsid height, both are obtained from the AFM images. (b) The average height of 62 capsids. The capsid height was obtained considering a circular region with 3 nm in radius. In this case, 1 nm = 0.85 px. The error bars represent the standard deviation of data contains in this circular region. (c) $D_{capsid,AFM}$ as in function of the height threshold.

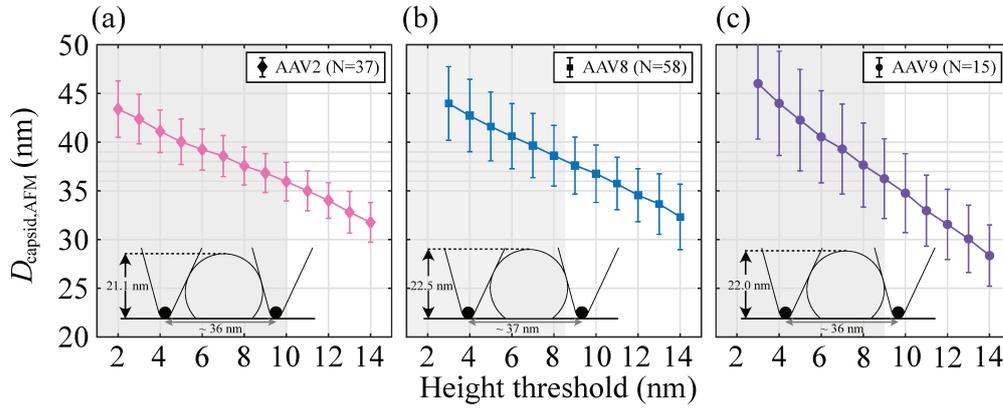


Figure 6.12: **Apparent diameter for AAV2, AAV8 and AAV9** - Apparent capsid diameter as a function of height threshold for (a) AA2, (b) AAV8 and (c) AAV9. The inset shows a schematic of the capsid with two cantilevers on each side. The capsid height shown is the height obtained from the AFM images, and the distance under the capsid is the width that must be found according to the configuration shown. The drawing represents the real scale between the capsid and the cantilever tip.

an intermediate value between 3 and 8.5 nm; that is, we choose a height threshold equal to 6 nm.

Additionally, this analysis has been performed for AAV2 and AAV9 capsids [Fig. 6.12]. The diameter $D_{\text{capsid,AFM}}$ as a function of the three serotype height threshold changes. This behaviour could be associated with how much the capsid was deformed in the mica. However, we will not address this possibility for the moment. According to the measured capsid apparent height, we see that the range of the height threshold for AA2 varies from 2 to 10 nm, so we choose 6 nm as an intermediate value [Fig. 6.12(a)]. For AAV9, the height threshold range varies from 3 to 9 nm; in the same way, the intermediate value chosen is 6 nm [Fig. 6.12(c)]. Finally, we have chosen 6 nm as a threshold height for the three types of capsids [Fig. 6.12].

Up to this point, we show the protocol for extracting the morphological properties of the capsids from the AFM images. Next, we will show how to extract the mechanical property of the capsid, precisely, the capsid stiffness from the nanoindentation curves.

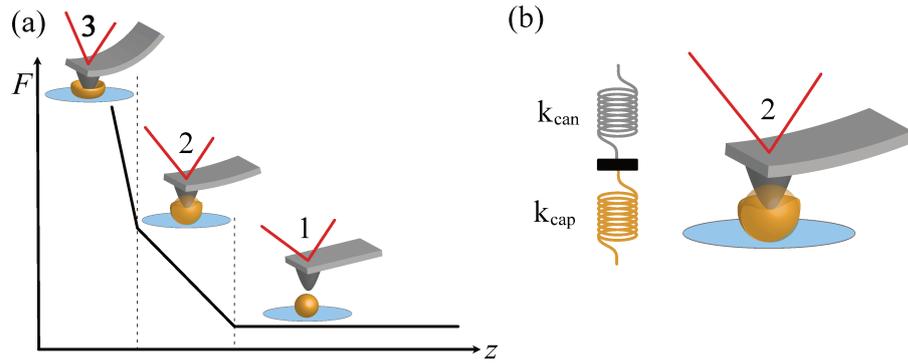


Figure 6.13: **Scheme of the typical z - F curves when a capsid is deformed** - (a) The z - F curves show three regimes: (1) the cantilever has not yet touched the capsid. (2) Both the capsid and the cantilever are deforming. (3) Deformations are given only by the cantilever deformation. (b) The capsid and cantilever set can be considered two springs in a series cantilever, assuming the capsid as an elastic shell.

6.6 Methods to determine the capsid stiffness from z - F curves

A schematic of a typical z - F curve when a cantilever is deforming a capsid is shown in figure 6.13(a). Three regimes can be identified: (1) when the cantilever has not yet touched the capsid, then the force is zero. (2) The capsid and the cantilever are deforming simultaneously. (3) Deformations are given only by the cantilever. At this point, the capsid reaches its deformation limit. Therefore, the moment where we can analyse the capsid deformation is during regime 2. However, the calculation is not direct since the information curve contains both the capsid deformation and the cantilever.

To calculate the capsid stiffness, we used two methods: (i) the two spring in series method and (ii) the method that allows obtaining the stiffness from the force curves as a function of the deformation capsid.

6.6.1 Method 2 spring in series

The capsid and cantilever set can be considered two springs in series, assuming the capsid as an elastic shell [Fig. 6.13(b)]. The effective elastic constant k_{eff} of the capsid-cantilever system is calculated using the relation,

$$\frac{1}{k_{eff}} = \frac{1}{k_v} + \frac{1}{k_c}, \quad (6.4)$$

where the k_c and k_v are the cantilever and capsid stiffness, respectively. The capsid effective stiffness k_{eff} was calculated from the linear part of regime 2 [Fig. 6.13]. Solving, the k_v can be written as,

$$k_v = \frac{k_{eff}k_c}{k_{eff} - k_c}, \quad (6.5)$$

To refer to capsid stiffness using this method, we will use k_{v1} as the nomenclature.

6.6.2 $F(\delta)$ method

In the second method, the capsid stiffness k_{v2} was calculated from the horizontal difference between the reference curve and the z - F curves at a constant force, according to the following [Fig. 6.14],

$$F(z) = k_{eff}(z - z_0), \quad (6.6)$$

$$F_c(z) = k_c(z_c - z_{0c}), \quad (6.7)$$

where $F(z)$ and $F_c(z)$ are the mathematical representation of the force exerted on the capsid-cantilever system and the cantilever to deform it ($z - z_0$) and ($z_c - z_{0c}$), respectively. We used the same principle as method (1), the slope of the capsid-cantilever curve k_{eff} is defined as,

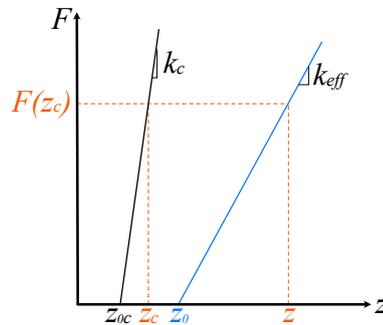


Figure 6.14: **Scheme showing two z - F curves** - Force curves as a function of the vertical displacement of the cantilever z . Blackline represents the cantilever deformation and the blue line represents the deformation of the capsid-cantilever system.

$$k_{eff} = \frac{k_c k_v}{k_c + k_v}. \quad (6.8)$$

Combining the equations 6.6, 6.7 and 6.8, and for a common force value, we obtained

$$F(z) = \frac{k_c k_v}{k_c + k_v} (z - z_0) = k_c (z_c - z_{0c}), \quad (6.9)$$

$$z_c = \frac{F(z)}{k_c} + z_{0c}. \quad (6.10)$$

From the figure 6.14, we defined a new δ as,

$$\delta = z - z_c, \quad (6.11)$$

and using the equation 6.10,

$$z = \delta + z_{0c} + \frac{F(z)}{k_c}. \quad (6.12)$$

We can redefine the variable $F(z)$ such as,

$$F(z) = \frac{k_c k_v}{k_c + k_v} \left(\delta + \frac{F(z)}{k_c} + z_{0c} - z_0 \right) \quad (6.13)$$

$$F(z) \left(1 - \frac{k_v}{k_c + k_v} \right) = \frac{k_c k_v}{k_c + k_v} \delta + \frac{k_c k_v}{k_c + k_v} (z_{0c} - z_0). \quad (6.14)$$

The last term $\frac{k_c k_v}{k_c + k_v} (z_{0c} - z_0)$ tends to zero if the contact points are aligned both at zero. Then,

$$F(z) = k_v \delta \quad (6.15)$$

However, if the contact points are not horizontally aligned, can we still extract the capsid stiffness? The answer is yes; we can extract it. If we do not remove the last term from the equation 6.14, $F(z)$ is defined as

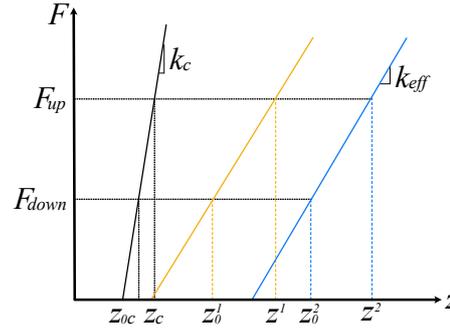


Figure 6.15: **Scheme showing two z - F curves 2** - Force curves as a function of the vertical displacement of the cantilever z . Blackline represents the cantilever deformation and the blue and yellow lines represent the deformation of two different capsid-cantilever systems.

$$\begin{aligned}
 F(z) &= k_v(\delta + z_{0c} - z_0) \\
 &= k_v[z - z_c + (z_{0c} - z_0)] \\
 &= k_v[z - z_0 - (z_c - z_{0c})].
 \end{aligned} \tag{6.16}$$

where $z - z_0$ is the difference of z in a given range of force for the capsid-cantilever curve, and the variable $z_c - z_{0c}$ is the same but for the cantilever curve. In this example, we have defined the force range $\Delta F = F_{up} - F_{down}$, where F_{down} is zero. However, this definition is true for nonzero F_{down} . Then, the difference of the forces is defined as follows [Fig. 6.15],

$$F_{up} - F_{down} = k_v[z - z_0 - (z_c - z_{0c})]. \tag{6.17}$$

We can generalise it to any other capsid-cantilever curve,

$$F_{up} - F_{down} = k_v[z^i - z_0^i - (z_c - z_{0c})]. \tag{6.18}$$

Finally, the capsid stiffness can be calculated from the following relation,

$$k_v = k(F(\delta)) = \frac{F_{up} - F_{down}}{z^i - z_0^i - (z_c - z_{0c})}, \tag{6.19}$$

$$k_{v2} = \frac{F_{up} - F_{down}}{z^i - z_0^i - (z_c - z_{0c})}. \tag{6.20}$$

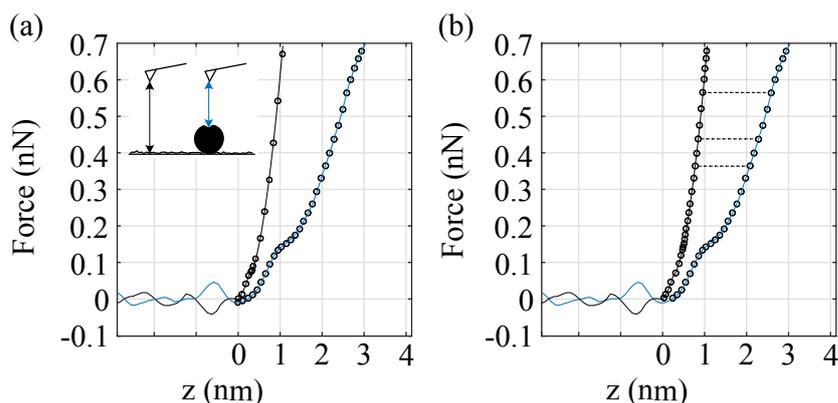


Figure 6.16: **Indentation curves interpolation** - (a) Force curves over the capsid (blue) and over the surface (black). There is a difference in the amount of point on each curve. (b) The interpolated force curve over the surface have the same amount of points than the curve over the capsid and with points in the same force values, allowing the subtraction of the curves.

The advantage of determining virus stiffness in this way is that it is unnecessary to determine the point of contact precisely. The contact point problem is one of the big problems that arise when determining the indentation distance on the virus.

It is necessary to do a new treatment on the curves to use this method, specifically on the curve with the cantilever information. As shown in our description, the curves must be subtracted at the same specific force; experimentally, we do not have this case. Figure 6.16(a) shows the reference curve (cantilever deformation) in black and the capsid-cantilever curve in blue. The circles represent the acquisition points. We noted that the curves do not have the same number of points; this is consistent because the cantilever travels different vertical distances, and for that route, the same number of points is set on these curves. Besides, it can be seen that the points are not equally spaced on the black curve; this is due to how the AFM software works. For example, the cantilever is descending vertically, and when it has reached a force of 0.2 nN, it has already set 95/100 points and realises that it only has 0.5 nm left, then it must fill that distance with the remaining points. Then, as the cantilever descends, it must calculate in each step how many points it has left; for this reason, the distance between points is not constant. To calculate the horizontal difference between the two curves, they must have the same number of points. To do this, we carry out an interpolation on the reference curve, fixing the same number of points as the capsid-cantilever curve [Fig. 6.16(b)]. This process was carried out for each capsid-cantilever curve since each curve has different behaviour.

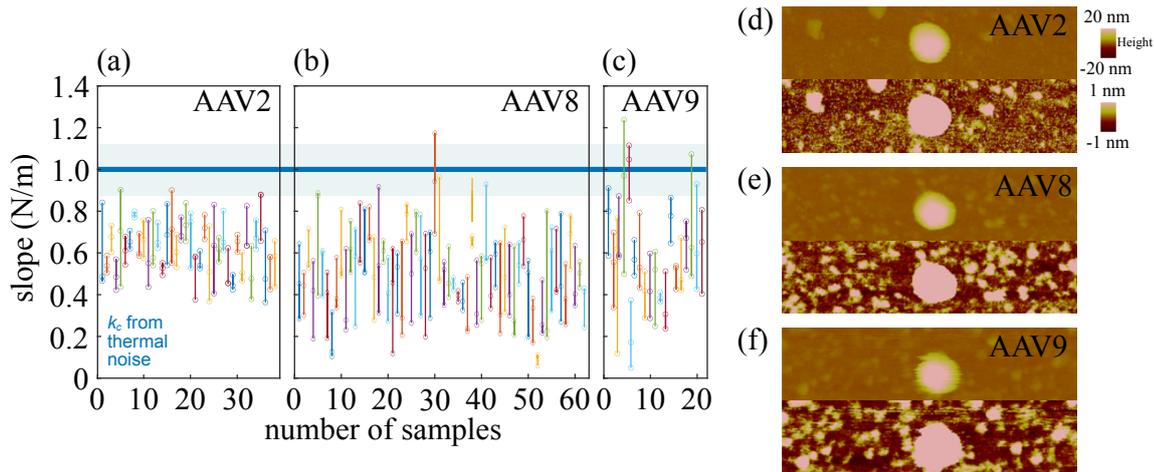


Figure 6.17: **Surface-Mica Analysis: contaminated layer** - The slope of the z - F curves for (a) AAV2, (b) AAV8 and (c) AAV9. Three repetitions were made in each indentation. Here, we averaged all the cantilever stiffness values obtained using thermal noise (blue line): this value is independent of the capsid type that has been indented. The light blue section contains the standard deviation of all stiffness values. (d,e,f) AFM images in two different scales for the three serotypes. We see impurities adhering to the surface at a scale that goes from -1.1 nm.

6.7 Nanoindentation on the surface

All z - F curves were obtained using Point and Shoot feature of Bruker acquisition software [Fig. 6.9(light blue box)], and they were obtained using the following parameters: the z -range was 100 nm, where each curve contains 1024 points. The ramp rate at which the piezo complete an extension-retraction cycle was 1 Hz approximately, and we used 700 pN of maximal force. Every time we indented a capsid, three indentations over the surface were obtained additionally. The indentation over the surface are required to calculate the stiffness capsid, as explained in the previous section. Besides, this reference curve is a control parameter in our protocol; the mica is rigid, then the z - F curve slope should give us the cantilever stiffness value since only the cantilever is deforming. Next, we will show the experimental results of indentations on a rigid (mica) surface to verify this assumption.

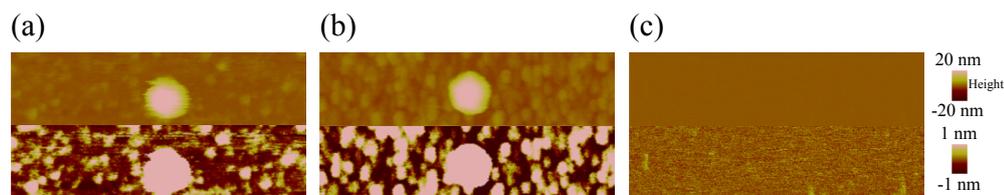


Figure 6.18: **Surface-Mica Analysis: contaminated layer** - Single AFM images for AAV9 capsid. The AFM images show a sample with two different capsid sample dilution (a,b) and another sample with not capsid solution (c).

6.7.1 Force curve slope on a rigid surface gives us the cantilever stiffness

Analysing the slope of the z - F curves on the surface, we noted that the slope values are smaller than the expected value (the cantilever stiffness value obtained from thermal noise) [Fig. 6.17(a-c)]. This difference could be related to impurities over the sample surface, as we can see in the AFM images on a smaller scale [Fig. 6.17(d-f)]. Particularly, we could argue that the limited dispersion in the slope values for AAV2 is related to a cleaner surface. The AFM images show fewer impurities for that capsid [Fig. 6.17(d)].

We have tested the effect of impurities by reducing the dilution of the sample containing the AAV9 capsids. As more capsids per unit area adhere to the surface, more impurities appear. [Fig. 6.18(b)]. We prepared a sample that only contains the buffer in the opposite case, where no impurity is visible [Fig. 6.18(c)]. These results suggest that the capsids solution causes contamination, possibly due to capsid proteins.

To verify that the cantilever deformation can be extracted from the z - F curves even if the impurities are present, we indented the surface using a higher force such we can find the regime when only the cantilever is deforming.

6.7.2 Indentations on the surface using a higher indentation force

To extract the stiffness of the cantilever from the z - F curves on surface, we used a maximum indentation force of 4 nN allowing to find the part when only the cantilever is deformed. In this case, we indented 58 random positions using the same cantilever can_1 , where the stiffness cantilever determined by thermal noise is 1.1 N/m.

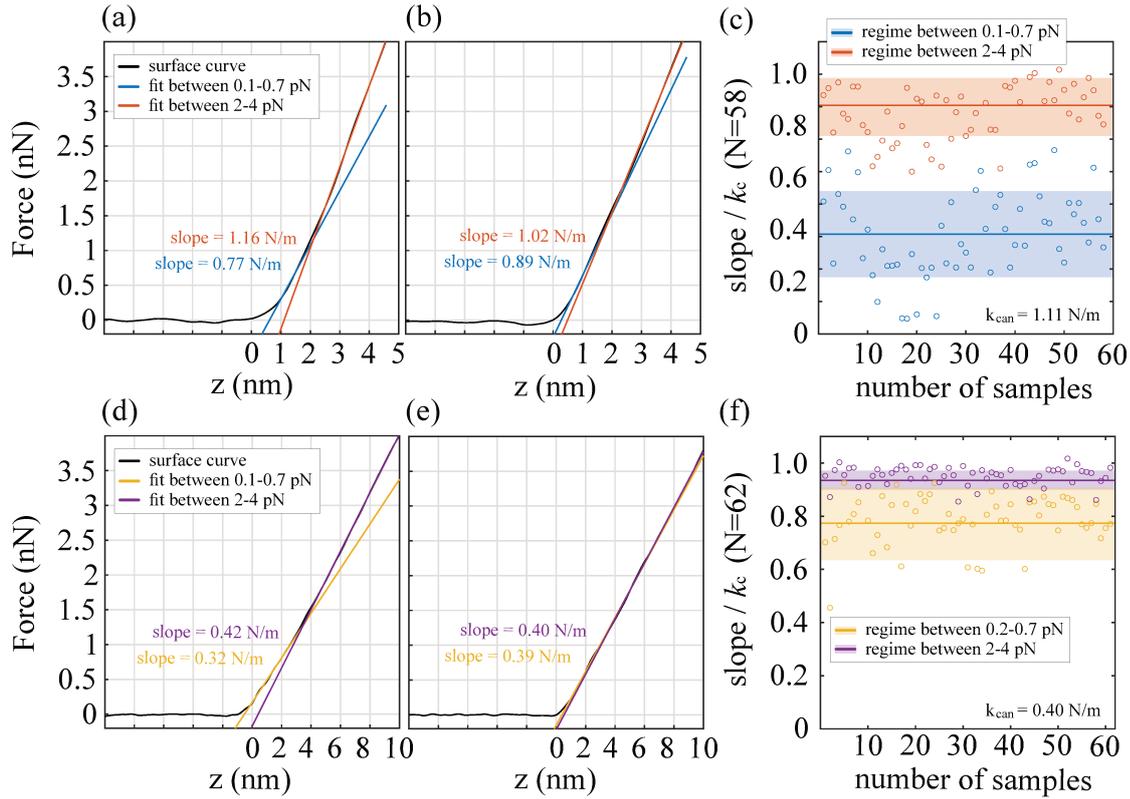


Figure 6.19: **Surface-Mica Analysis: z - F curves** - (a,b) Linear fit over the indentation curve using $k_c = 1.11$ N/m for two different fitting ranges. (c) Normalised effective stiffness (slope of z - F curve) of 58 capsids for the two fitting ranges. Figures (d,e) and (f) shows analogous results for cantilever $k_c = 0.40$ N/m.

The regime at small forces can be interpreted as an effect caused by the difference in stiffness between the cantilever and the surface; the contaminated layer is softer than the cantilever, so the contaminated layer will deform first. Also, we analysed the slopes with a softer cantilever can_2 : $k_c = 0.4$ N/m. The difference between the regime at small and large forces decreases [Fig. 6.19(d,e)], where the slopes set (yellow band) approaches to 1. In this case, we can say that both the cantilever and the contaminated layer have similar deformations. Consequently, the indentation curve does not have an abrupt transition.

Supposing the rigid surface (mica) is clean, the z - F curves should have only one regime since only the cantilever is deforming. In some particular cases, we had some curves with a linear behaviour [Fig. 6.19(b,e)], meaning the impurities layer is not homogeneous. Therefore, we can extract the cantilever stiffness value by choosing the best curve, the closer one to the value given by thermal noise. In conclusion, when the mica surface has some impurities adhered to, then the z - F curves will contain the cantilever deformation and information about these impurities. For this reason, this

effect should be taken into account for the estimation of the stiffness using the reference force curves on the surface.

6.8 Nanoindentation on the capsid

This section will show how to extract the AAV8 capsid stiffness. We performed two protocols, the first one with singular indentations on the capsid and the second one using a profile containing several horizontally aligned indentation points over the capsid.

6.8.1 One indentation point on the capsid centre

In this case, 58 capsids were indented on the capsid centre, then 58 indentations were realised on the surface. However, and according to the analysis carried out in the previous section, only the best curve on the surface was chosen [Fig. 6.20(a)]. The indentations were developed using the cantilever can_1 ($k_c = 1.0$ N/m). To obtain the z - F curves, the following parameters were used: the z range was 100 nm, where each curve contains 1024 points. The ramp rate at which the piezo complete an extension-retraction cycle was 1 Hz approximately, and we used 4 nN of maximal force [Fig. 6.20(b)].

We noted a large variability in the z - F curves [Fig. 6.20(b)]. During the indentation process, the capsid moves around due to thermal-piezo drift, generating that the point chosen to make the indentation, the centre of the capsid, is different from the actual point indentation. Hence, we propose that the curves with the greatest fluctuations or largest δ_{max} are indentation curves not in the capsid centre. Following this hypothesis, we chose some curves with small fluctuations, and they are compared with the surface curve [Fig. 6.20(c)]. Comparing both curves, we noticed that they behave in the same way for a force greater than 1.8 nN, meaning that only the cantilever is deforming. Therefore, the information on the capsid deformation is contained for a force less than 1.8 nN [gray box in figure 6.20(c)]. A similar analysis was performed using the cantilever can_2 ($k_c = 4.0$ N/m), where we also noted a large variability for the 61 indentation curves [Fig. 6.21(d)]. In this case, we have set the same no-deformation limit of the capsid: the slopes collapse around 1.8 nN [Fig. 6.21(e)].

We propose that the large range of the z - F curves [Fig. 6.20(b) and 6.21(a)] are due to the real indentation positions that is not the one given by the overlap of AFM images and Point and Shoot position. Another possible explanation is related to the

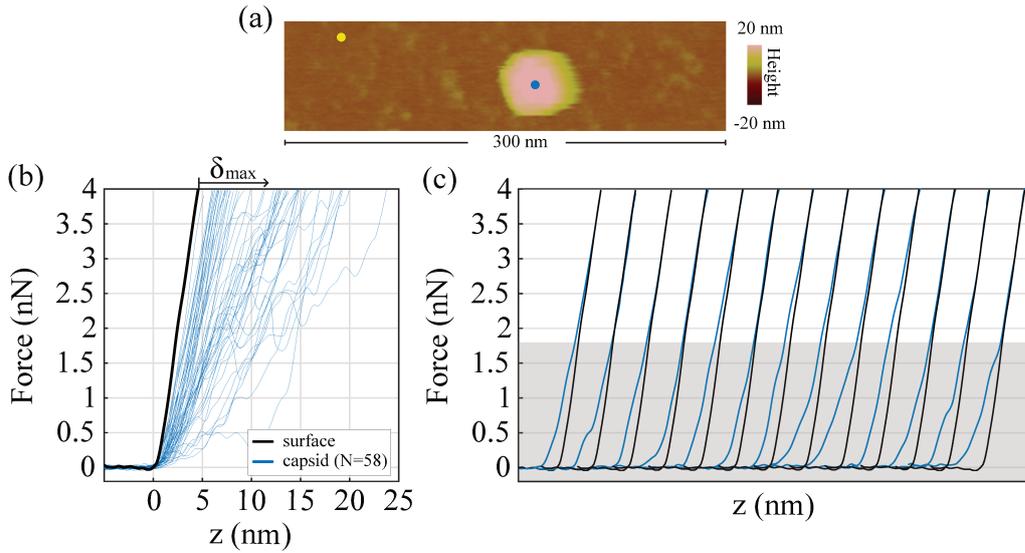


Figure 6.20: **Single capsid indentation curves for different positions using can_1** - (a) AFM image of a single capsid AAV8. (b) z - F curves for 58 capsids indented and 1 z - F curve on the surface. The indentation positions are shown in the AFM image. (c) Some z - F curves with smaller maximal indentation δ_{max} are shown. The grey region show the point where the capsid is deforming.

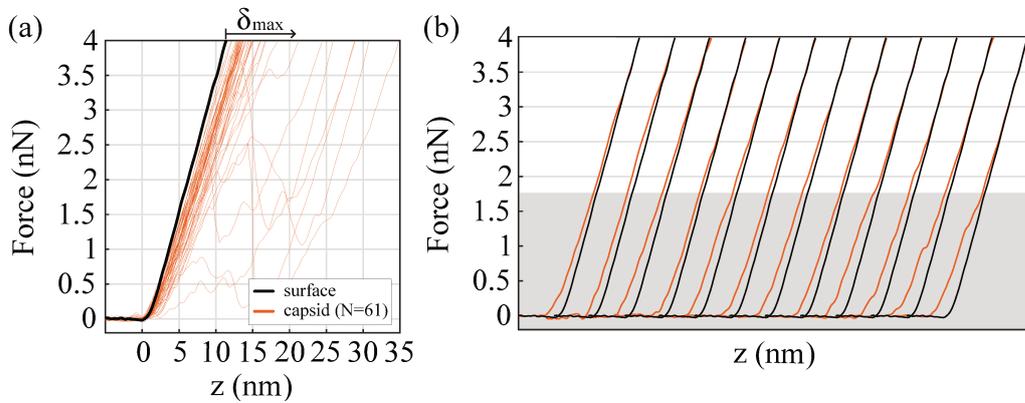


Figure 6.21: **Single capsid indentation curves for different positions using can_2** - (a) z - F curves for 61 capsids indented and 1 z - F curve on the surface. (b) Some z - F curves with smaller maximal indentation δ_{max} are shown. The grey region show the point where the capsid is deforming.

capsid orientation when it is deposited on the mica. Carrasco *et al.*, [156] shows that the capsid stiffness value is different when we are indented on a face, edge or vertex. To verify these assumptions, we made several horizontally aligned indentations, and then we choose the curve that best approximates the expected behaviour in the capsid centre, allowing us to exclude dependency related to the indentation position.

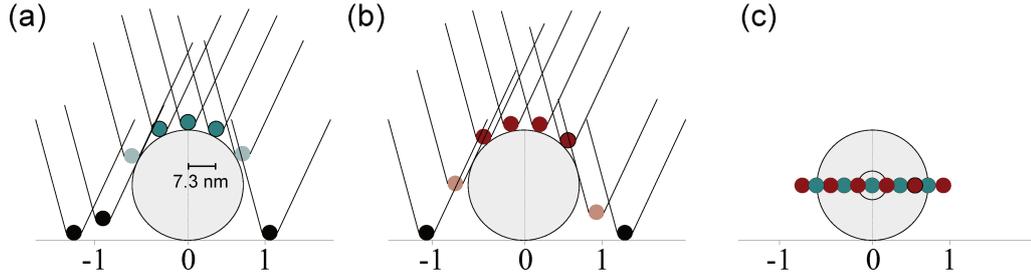


Figure 6.22: **Indentation positions scheme on the capsid** - (a) Indentation positions when a point is in the centre of the capsid.(b) Indentation positions when no points are touching the capsid centre. The burgundy and green tips represent the positions when the cantilever touches the capsid. (c) Superposition of the points that touch the capsid between case (a) and (b). The black circle represents the case where information on the capsid stiffness can always be extracted, that is, the points closest to the capsid centre.

6.8.2 Indentation spatial profile over the capsid

Unlike the singular points of indentation, we choose a lower force to make the indentation profile because we break the capsids at the border using 4 nN of maximal force: in this case, we used a maximal force of 0.7 nN. Furthermore, we know that the capsid is deforming in this range of forces [Fig. 6.20(c) and 6.21(b)], therefore, we can consider the entire curve to extract the capsid stiffness.

The protocol proposed here consists of 15 indentation points across the capsid. Then, we chose the curves close to the capsid centre. The indentation points are separated by a distance of 7.3 nm when we use the tip with a radius of 2 nm. According to this configuration, we have two limited situations of how the tip is positioned over the capsid. One position of the tip is located in the capsid centre, or there are two positions are located around the capsid centre [(a) and (b) in figure 6.22]. By superimposing several indentation profiles, the distance between the indentation points is reduced [Fig. 6.22(c)]. Therefore, we decide that the best case to obtained the capsid stiffness is taking the three central points of the profile [black circle of the capsid Fig. 6.22(c)] because we take into account the case (a) and (b). Assuming that we only study half of the indentation profile, only the central point and the point that follows it will be considered.

The z - F curves contain 1024 points and the ramp rate used is 1 Hz. Figure 6.23 shows the z - F curves using the cantilever can_1 and a pattern is identified respect to the reference curve (black line): from the left, the z - F curves present similar behaviour as the surface curve [Fig. 6.23(b)]. Next, the z - F curves slope decrease: we defined

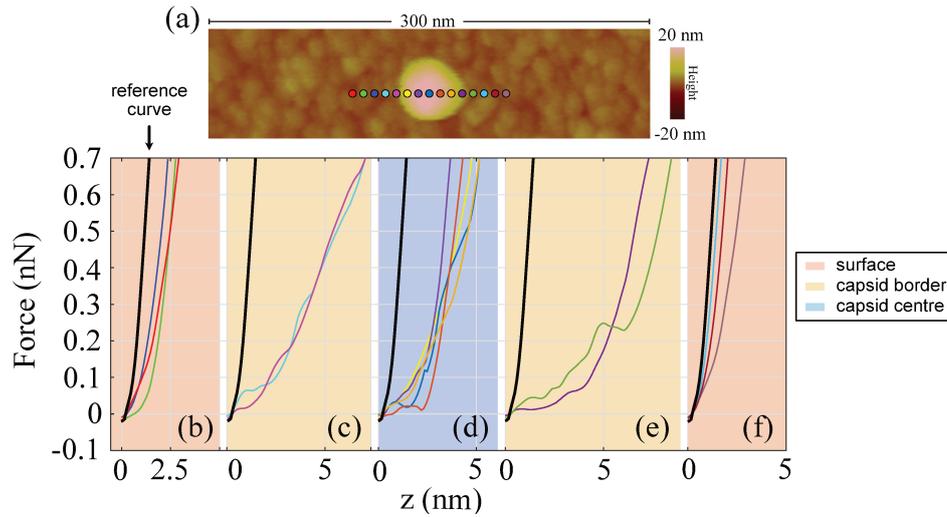


Figure 6.23: **The behaviour of the z - F curves depends on the indentation position -** (a) Single capsid AFM image. The circles represent the indentation position using NanoScope Point and Shoot feature. (b) Indentation curves separated by position with the colours showed in (a). There is a different behaviour and the curves can be sorted by their relative position.

those curves as curves on the capsid border [Fig. 6.23(c)]. The z - F curves slope starts to increase, meaning that the curves are on capsid top [Fig. 6.23(c)]. Finally, the z - F curves slope decrease again until it reaches the surface [Fig. 6.23(e,f)]. From the indentation profile, we identified that:

(i) The surface z - F curves show differences concerning the reference curve, that is, to the best curve chosen on the surface [Fig. 6.23(b,f)]. This is due to impurities adhering to the mica surface.

(ii) Although the curves at the surface or the curves at the capsid top are similar, they can be told apart according to the profile pattern. That is, between a curve on the surface and the capsid top, there is always a curve with a large maximum displacement [Fig. 6.23(c,e)].

(iii) Furthermore, the essential characteristic is that the z - F curves shape powerfully depend on the indentation position. The same overall behaviour is observed for AAV2 and AAV9 vector capsids.

The analysis was developed for many capsids, but we show just some of them in figure 6.24. As we see in that figure, we separated the surface, border and top capsid curves according to the orange, yellow and blue colours, respectively. The indentation positions on the AFM images are differentiated by colour as well. Curves on the capsid border have larger maximal displacement z_{max} in comparison with z - F curves

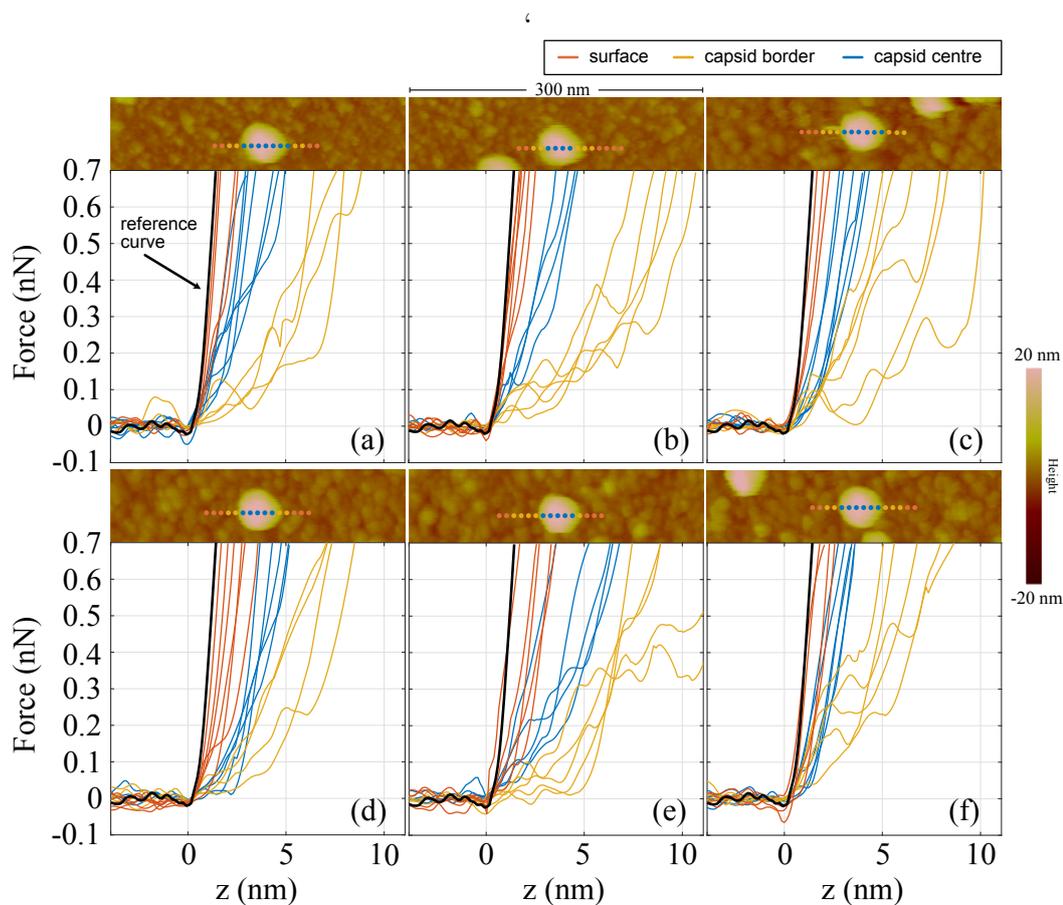


Figure 6.24: **Separating z - F curves according to indentation position** - Several examples of capsids that have been indented. The circles on the AFM image represent the indentation position using NanoScope Point and Shoot feature. The z - F curves are separated by position with the colours showed in the AFM image.

on the capsid top. Therefore, we can already emphasise that the different types of curves displayed in individual indentations [Figures 6.20(b) and 6.21(a)] are due to the indentations not being performed at the capsid top.

In conclusion, the indentation profile allows us to separate the effect of the capsid indentation position; however, we still have the effect of the capsid orientation. One way to identify the capsids orientation is through the AFM images; nevertheless, we could not obtain enough resolution to differentiate the pentamers or hexamers. Looking at the z - F curves for capsid centre (blue lines), it is possible to identify curves with different maximal displacement z that could be associated with the capsid orientation [blue curves in figure 6.24(b) and 6.24(e)]. We will discuss this topic later. Another behaviour that we noticed in the z - F curves, which are positioned mainly on the capsid centre, is a non-linear effect at small forces, which is discussed in the next section.

6.9 Non-linear effects on force curves

Visualising the z - F blue curves in figure 6.24], we noticed a non-linear regime that appears at small forces RsF (Regime at Small Force). To study this effect in details, we only show the curves passing mainly on the capsid centre for some capsids, as we can see in figure 6.25. A straight red-pointed line is drawn on each curve to visualise the non-linear effect better. The slope of the red line is given according to the range where the curve slope remains constant.

Comparing the RsF in the curves for different capsids, we observe no defined behaviour; in some cases, as we move away from the capsid centre, the RsF regime begins to grow [Fig. 6.25(a)]. In other cases, it happens as in the previous case, but just the central curve presents the largest RsF [Fig. 6.25(d)]. Also, it appears to be completely random. A possible cause of this effect is the spikes present on the capsid surface since the RsF does not have a defined shape, i.e. it does not constantly change as we move away from the capsid centre. It is important to mention that the all analysed range, the RsF and the linear part range, are within the deformation limit of the capsid shell. For the moment, we keep this assumption as speculation because we do not know how to test this idea.

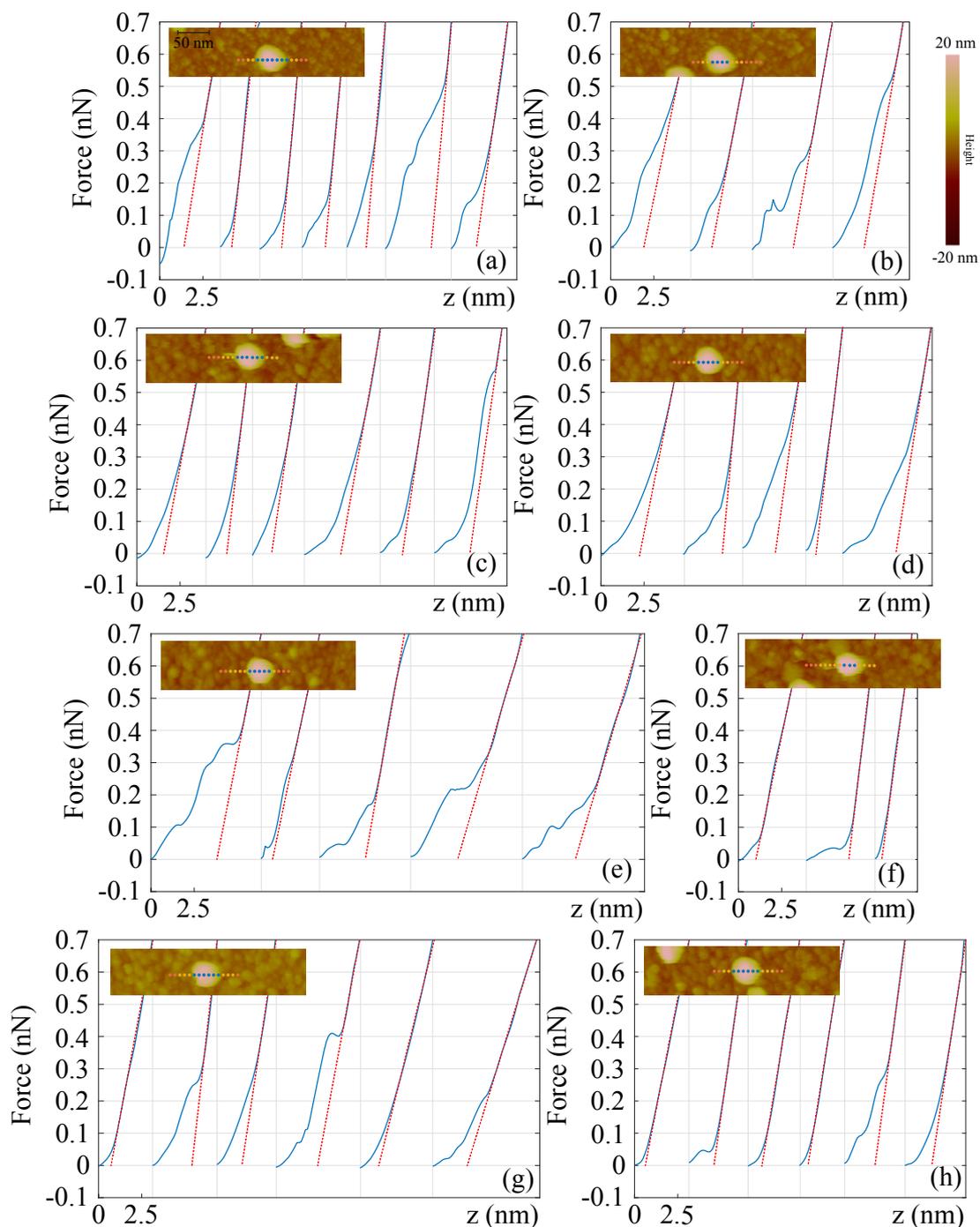


Figure 6.25: z - F curves on the AAV8 capsid top - Force curves for the central points of the indentation profile (blue lines). The indentation positions are shown in the AFM image where the points are separated by colors. The orange points represent the positions on the surface and the yellow and blue points represent the positions at the border and top of the capsid. A straight red line is drawn on each z - F curve from the maximum force, 700 pN, to the point where the curve changes its slope, showing the point where the curve stops being linear.

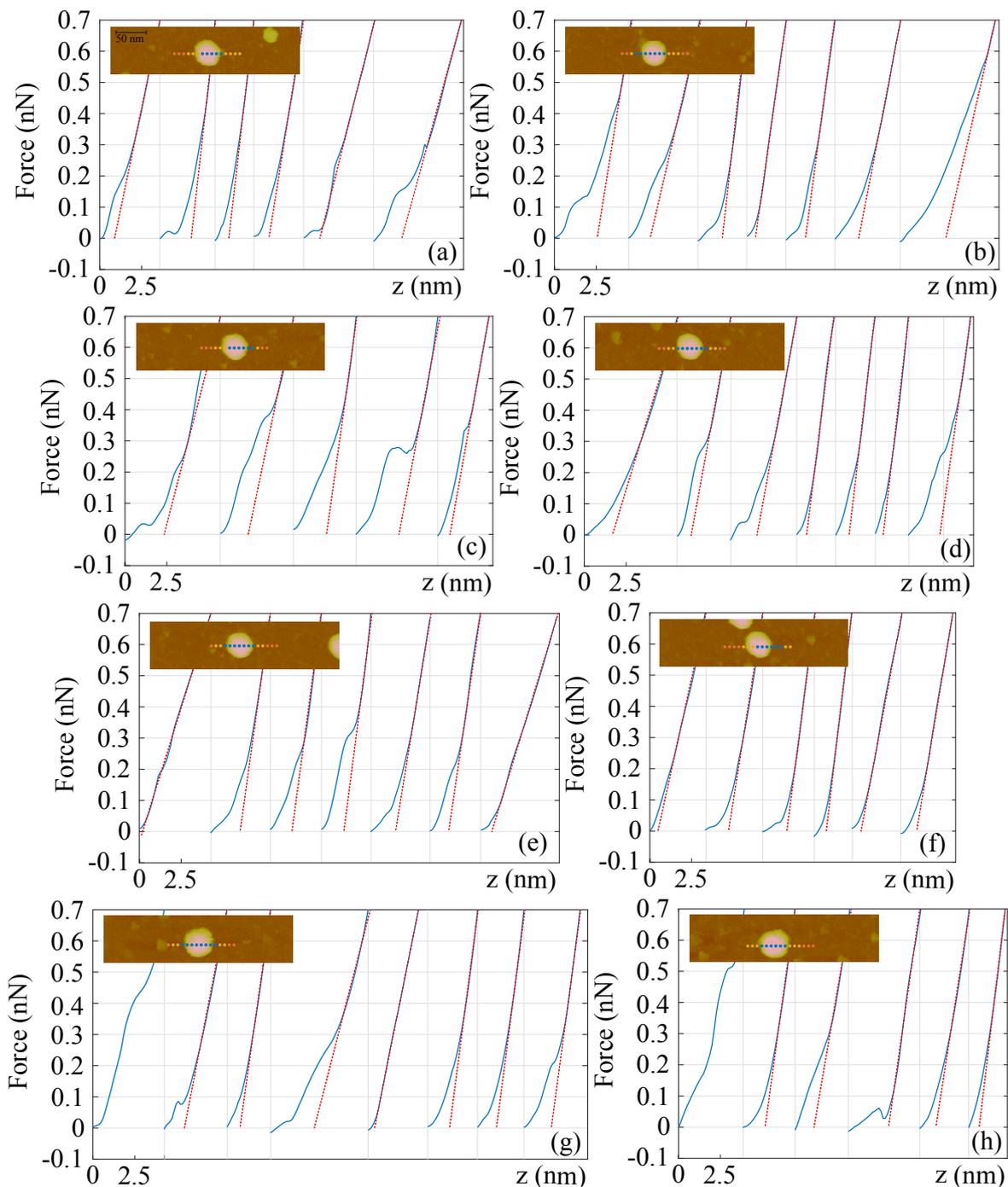


Figure 6.26: z - F curves on the AAV2 capsid top - Force curves for the central points of the indentation profile (blue lines). The indentation positions are shown in the AFM image where the points are separated by colors. The orange points represent the positions on the surface and the yellow and blue points represent the positions at the border and top of the capsid. A straight red line is drawn on each z - F curve from the maximum force, 700 pN, to the point where the curve changes its slope, showing the point where the curve stops being linear.

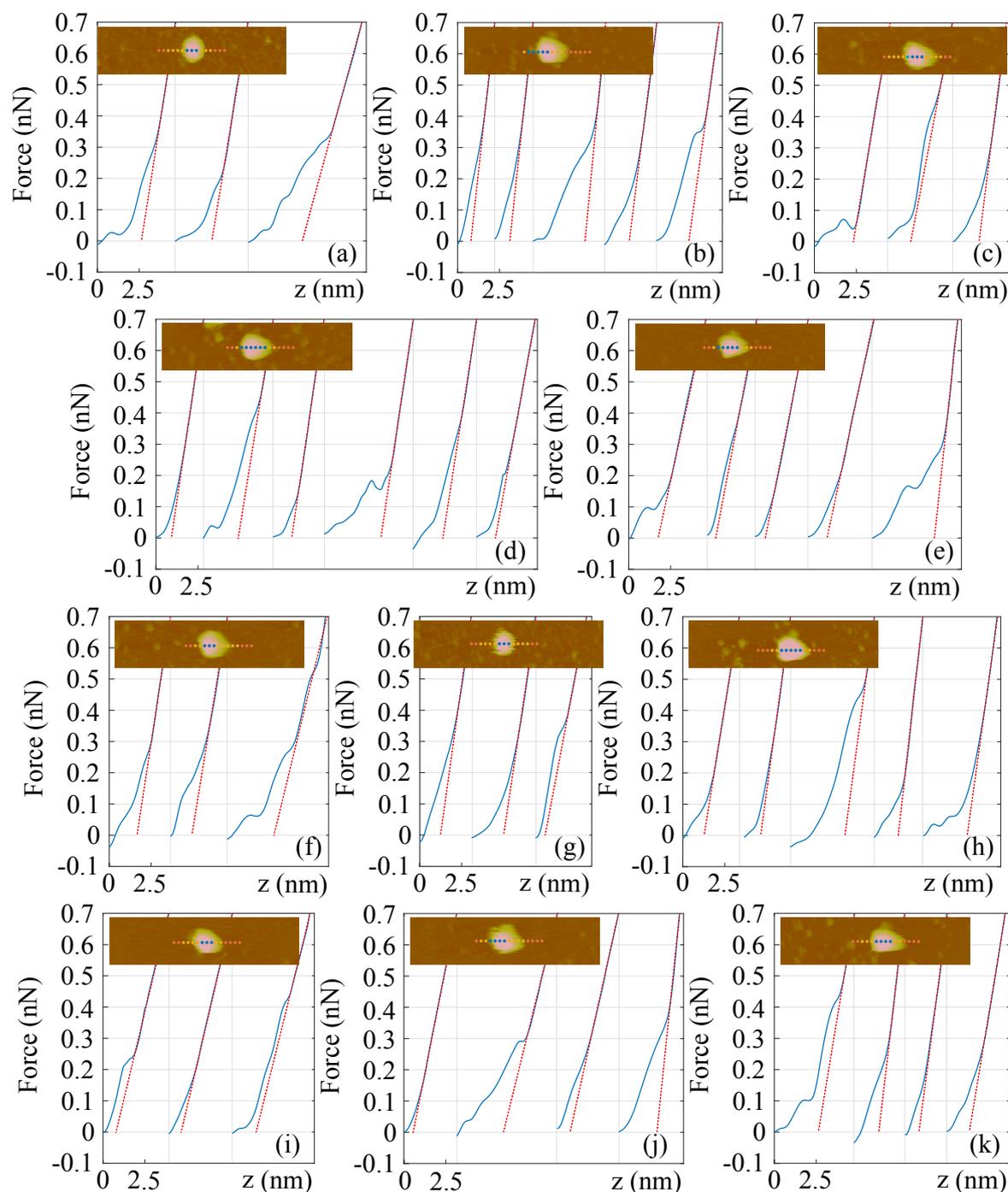


Figure 6.27: z - F curves on the AAV9 capsid top - Force curves for the central points of the indentation profile (blue lines). The indentation positions are shown in the AFM image where the points are separated by colors. The orange points represent the positions on the surface and the yellow and blue points represent the positions at the border and top of the capsid. A straight red line is drawn on each z - F curve from the maximum force, 700 pN, to the point where the curve changes its slope, showing the point where the curve stops being linear.

6.10 Capsid average profile

We have studied the behaviour indentation curves passing through the centre and border of the capsid to this point in the analysis. The remaining analysis is to obtain an average profile. For simplicity, we will use the indentation profile of the effective stiffness (capsid+cantilever deformation). In the results section, we will show the profile of the capsid stiffness.

As we see in figure 6.28(a), the profile is not perfectly centred at zero since the imaging and nanoindentation process are not simultaneous. First, the image is scanned, then we choose a possible capsid centre, and the indentation is followed. The seconds when the image has been scanned, and the indentation process can generate displacement of the capsid due to thermal drift [337–339]. Consequently, the positions of each indentation point are aligned to the capsid centre in the AFM images. The following steps are carried out to obtain a profile centred: we choose the two lowest points of the profile around the central point (local maximum) [grey points in the figure 6.28(b)]. Then, we calculated the half distance $d/2$ between the two points. All the points are subtracted $d/2$, leaving the profile aligned at zero [Fig. 6.28(c)]. Finally, we divided the profile positions to the capsid radius obtained from the AFM image [Fig. 6.28(d)].

Figure 6.29(a) shows all the overlapping of 58 profiles, and the general profile shape is still maintained. The average profile is obtained setting $r/R_{virus,AFM}$ ranges with a step of 0.2. Then, we obtained the median, equal to 50 % of the data (black square), and the upper and lower limits represent 75 % and 25 % of the data, respectively [Fig. 6.29(b)].

6.11 Conclusion

We have established an indentation protocol to obtain the stiffness of a viral capsid as a function of the indentation position. We observed that there is a large variety of z - F curves when singular indentations are made over the capsid. This effect is most likely because the capsid is not deformed at the centre. Also, we show that the slope of the z - F curves are smaller at the border than at the centre. These results suggest a dependence of effective stiffness as a function of position. Furthermore, the z - F curves show a non-linear behaviour for small forces that could be related to the spike on the capsid surface.

We will then use this protocol to analyse the stiffness of icosahedral capsids and

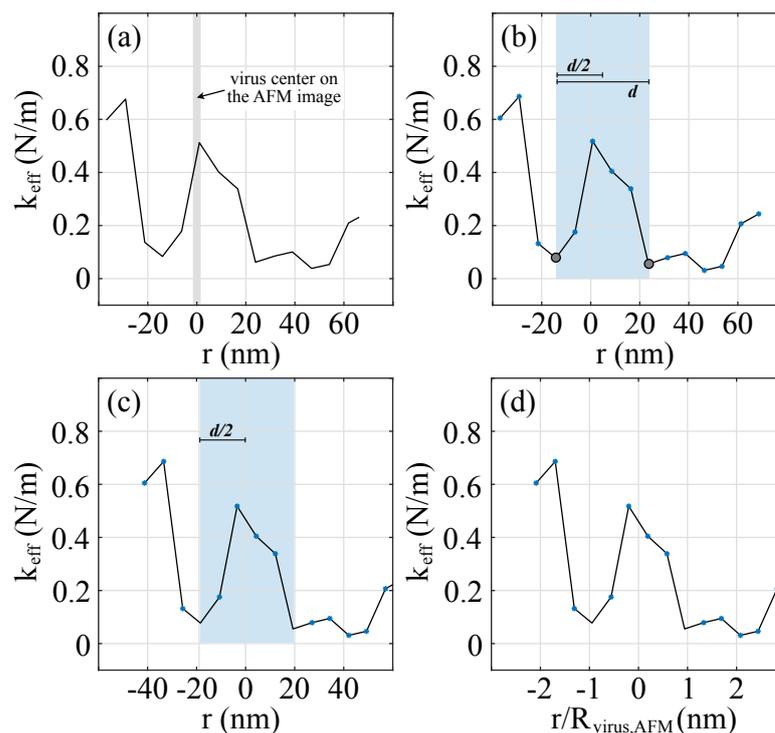


Figure 6.28: **Effective stiffness profiles alignment** - (a) Original profile, the grey vertical line represent the capsid centre in the AFM image. (b) We choose the first minimums values around the center and defined the distance d between these points. (c) All the points are subtracted $d/2$, redefining the center of the profile. (d) Finally, the profile is normalised by $R_{virus,AFM}$.

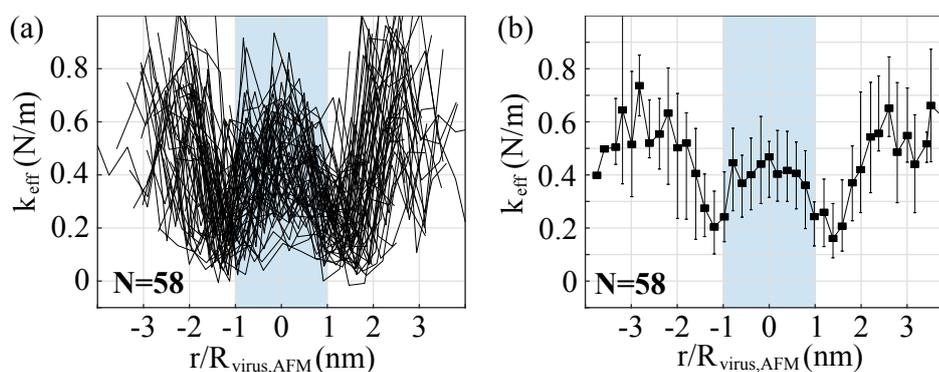


Figure 6.29: **Effective stiffness average profile** - (a) Figure show the aligned k_m profiles for 58 capsids, the light blue zone represent the indentations over the capsid. (b) Average profile of the effective stiffness k_{eff} . The black squares represented median of the data and the upper-lower error bars are the 75% and 25%.of the points, respectively.

study the possible dependence of stiffness on the geometry and orientation of the capsids.

Chapter 7

Stiffness of icosahedral capsids

Index

Summary	163
7.1 AAV8 capsid	164
7.1.1 Experimental Stiffness profile	164
7.1.2 Stiffness profiles according to the values of central points	166
7.1.3 Simulations of indentation numerical on viral capsid by Lucas Menou and Martin Castelnovo	168
7.1.4 Cantilever spring constant effect	173
7.1.5 Cantilever tip shape effect	176
7.2 AAV2 and AAV9 capsids	179
7.3 HBV capsid	182
7.3.1 Morphology of HBV capsid	182
7.3.2 Stiffness profile according to the values of central points	183
7.3.3 Influence of capsid facet in the stiffness profile	186
7.4 Conclusions	188

Summary

This section will study the stiffness heterogeneity using an indentation profile for two different small capsids: Adeno Associated Virus (AAV, T=1) and Hepatitis B Virus (HBV, T=4). First, we will analyse the effect of the stiffness and size cantilever using

the indentation profile for the AAV8 capsid. Then the stiffness of the capsids of AAV2, AAV9 are determined. Although both virus capsids have an icosahedral shape and similar size, HBV and AAV have 240 and 60 capsids proteins, respectively; consequently, AAV capsid has a more facet aspect than HBV capsid. Therefore, we are interested in studying the effect of faceting on the indentation profile. Finally, the HBV capsid will be analysed and compared to the AAV capsid.

The nanoindentation project on capsids was developed in collaboration with the PhD student Lucas Menou and CNRS researcher Martin Castelnovo (Physics Laboratory, ENS of Lyon). Lucas made simulations of AFM indentations on viral capsids, changing cantilever and capsid geometry. The simulations were performed in order to mimic the experimental conditions. Also, a geometrical model allowing to describe the stiffness profile was developed due to the collaboration. More details in [340].

7.1 AAV8 capsid

7.1.1 Experimental Stiffness profile

According to the nanoindentation protocol, we have two methods to determine the stiffness capsid k_{v1} (two spring constant in series method) and k_{v2} (stiffness capsid from the $\delta-F$ curves). First, we analysed the profiles using the can_1 ($k_c = 1.0$ N/m and

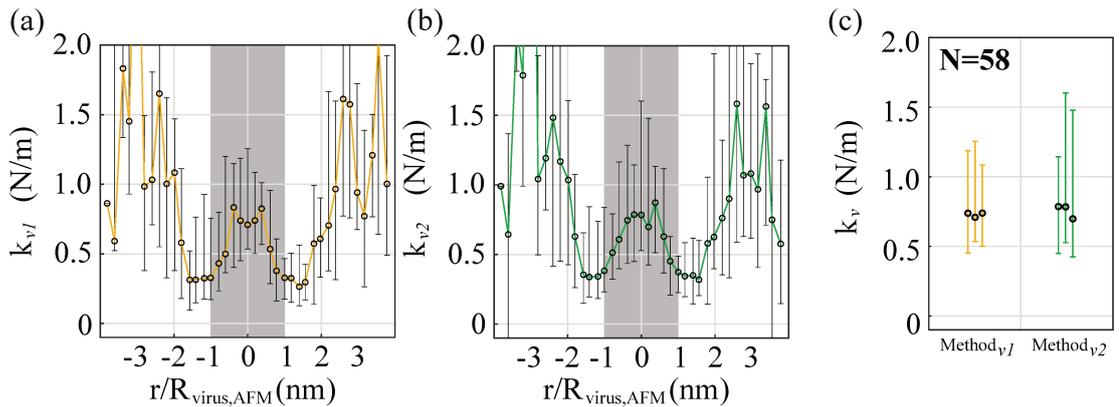


Figure 7.1: **Stiffness profiles for AAV8 using the best reference curve-** For various indentations on the surface, the best $z-F$ curve is chosen to determine the stiffness of the capsid. (a) Stiffness profile using the method k_{v1} (two spring constant in series method) and (b) k_{v2} (stiffness capsid from the $\delta-F$ curves). (c) Comparison of the three central points of the profiles for the mentioned cases.

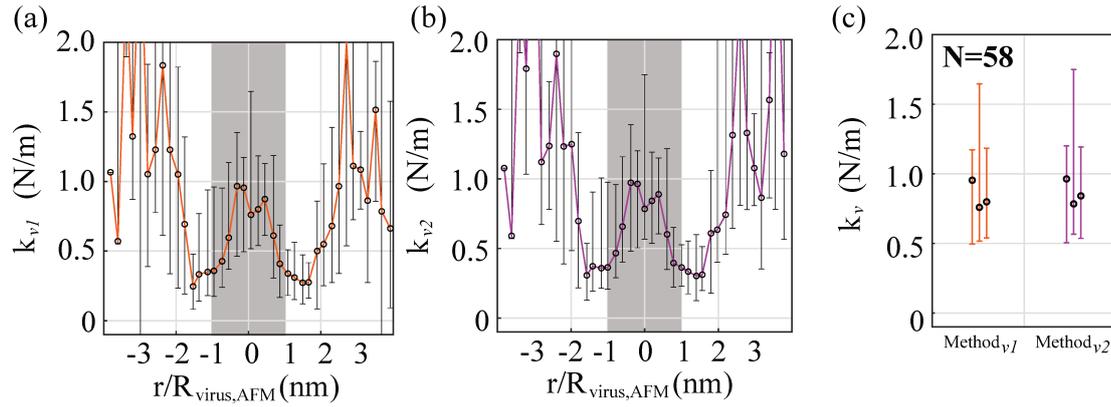


Figure 7.2: **Stiffness profiles for AAV8 using created curve with slope value obtained from thermal noise-** A linear curve was created as a reference curve. The slope of this curve was determined from thermal noise. (a) Stiffness profile using the method k_{v1} and (b) k_{v2} . (c) Comparison of the three central points of the profiles for the mentioned cases. The black circle represent the median of the data and the upper-lower error bars are the 75% and 25% of the points, respectively.

$R_c = 2$ nm) when we choose the best reference curve [Fig. 7.1]. The stiffness profiles show that there are no significant differences using method 1 [Fig. 7.1(a)] or method 2 [Fig. 7.1(b)]. As the three central points mostly represent the capsid deformation, we compared their median values [Fig. 7.1(c)]: not differences are visualised as well. Hence, the k_v obtained is not affected by the method used.

We have done the same analysis, but in this case, we create a new reference curve in which the slope value was obtained from the thermal noise method [Fig. 7.2]. Similar to the previous case, the methods to determine the capsid stiffness show no differences in the complete profile [Fig. 7.2(a,b)] and the three central points [Fig. 7.2(c)]. Also, we noted that the stiffness profiles using either the best reference curve or the created curve are similar, suggesting that the best reference curve gives the real cantilever information.

All the analysis performed below was performed using the method that allows determining the stiffness capsid from the δ - F curves, k_{v2} , as well as the actual reference curve (the best reference curve) since it contains actual information about the indentation.

7.1.2 Stiffness profiles according to the values of central points

As we have seen so far, the stiffness profile shows a dependence as a function of the indentation position, where it is more rigid in the centre than at the capsid border [Fig. 7.3(a)]. The advantage of studying stiffness using this profile is that it allows us to distinguish the value closest to the capsid centre more accurately, that is, the stiffness value where the vertical indentation is completely radial. As we mentioned before, the points that best represent the indentation in the capsid centre are the three central points of the profile (see figure 6.22).

Figure 7.3(b,c) shows the probability density function (pdf) of the three centre points where we noted two well differentiable peaks, one centred at 0.5 and the other centred at 1.8 N/m, approximately [Fig. 7.3(b)]. Additionally, we have analysed the stiffness values using the cumulative distribution function (cdf) [Fig. 7.3(c)]. This type

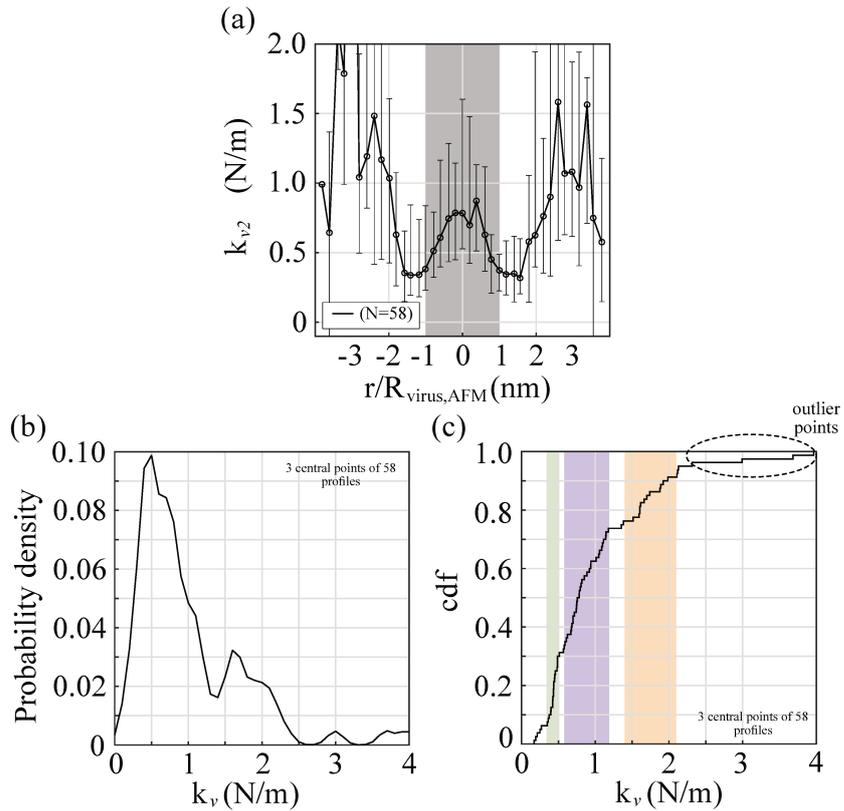


Figure 7.3: **AAV8: three central points of the indentation profile-** (a) Stiffness profile. The black circle represent the median of the data and the upper-lower error bars are the 75% and 25% of the points, respectively. The points contained in the colored region are indentation points on the capsid top. (b) Probability density and (c) the cumulative distribution function of the three central points, where 58 capsids have been indented.

of graph allows us to identify groups of a data set by changes in the curve slope, where the slope decreases and then increases again, seen as a valley. We have identified three regions (green, purple, and damask colour), an extra one compared to the probability density plot. The central values analysis shows that the 58 indented capsids present different stiffness values. This result is expected since we have not differentiated the capsids according to the orientation of the capsid. Therefore, we think that these different peaks or groups are associated with this effect.

We think it would be interesting to study the capsid orientation according to the stiffness profile type and not only according to the central value. To do that, we firstly identified three profiles that did not show a dependence of stiffness concerning the

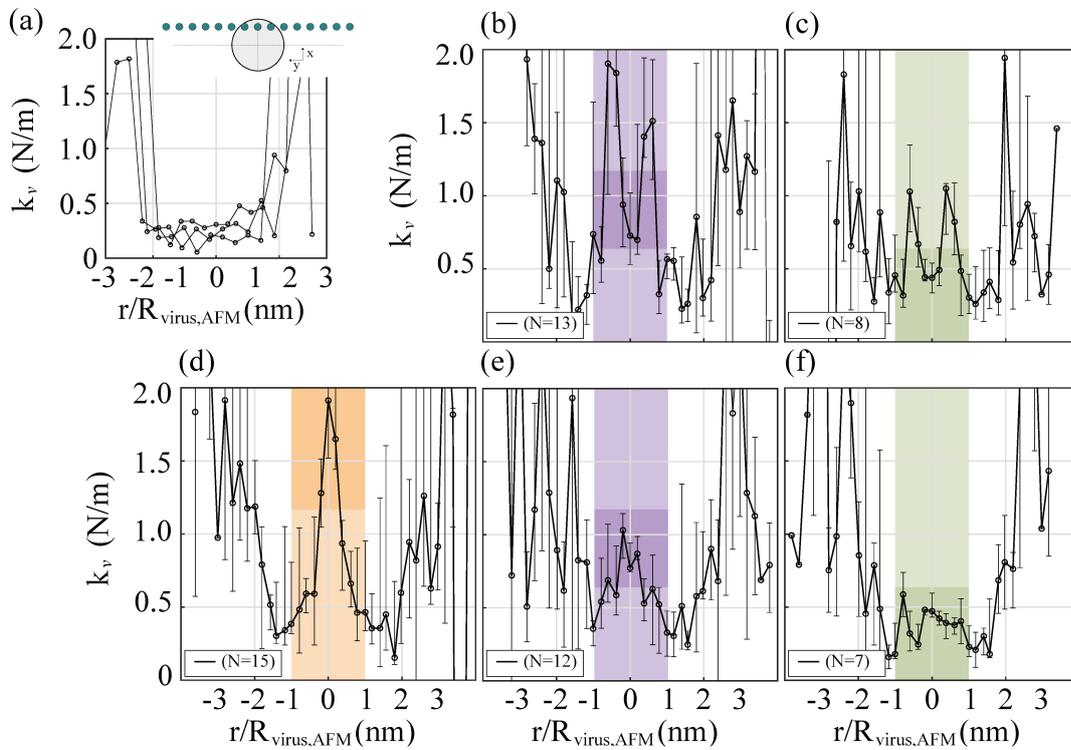


Figure 7.4: **AAV8: stiffness profile according to the value of the central positions-** (a) Indentation profile that did not pass vertically through the centre of the capsid. In this case, it goes through the capsid border, making all the values constant (inset show a scheme of this situation). The remaining profiles show certain behaviours: (b,c) the central values are smaller than at the border, showing a valley in the centre. The other type shows larger values in the centre (d,e,f). The rectangles colour show profiles according to a certain range of stiffness values for the central values, where the green, purple and damask colours contain values ranging from 0-0.7 N/m, 0.7-1.2 N/m and 1.2-2 N/m, respectively. The black circle represent the median of the data, and the upper-lower error bars are the 75% and 25% of the points, respectively.

position: this type of profile did not pass vertically through the centre of the capsid. Figure 7.4(a) shows a scheme where the central indentation has well alignment over the x-axis but not at the y-axis. On the other hand, the profiles that show dependence with the positions present varied behaviours; in some cases, the central points values are greater than the points next to these three central points [Fig. 7.4(d,e)] or are smaller than the points next to these three central points [Fig. 7.4(b,c)]. Although the pdf and cdf of the central values are good tools to know that there are groups related to the capsid orientation, they cannot separate the profiles that show a valley or mountain behaviour. The mountain and valley behaviour were separated by hand due to an inability to detect each profile type automatically.

Looking only at the central points, we noticed that the one group has values close to 0.4 N/m [Fig. 7.4(c,f)] and the other has central values around 0.75 N/m [Fig. 7.4(b,e)]. The points next to these three central points decrease or increase values are related to the different trajectories of the cantilever passing over the capsid. The third group does not present two behaviours: the central value is close to 1.8 N/m, and it falls fast as it moves away from the centre [Fig. 7.4(d)]. The different types of stiffness profiles clearly show that each trajectory has an associated capsid orientation. The remaining analysis is to associate a stiffness profile with a specific orientation, a discussion that will take place in the next subsection.

7.1.3 Simulations of indentation numerical on viral capsid by Lucas Menou and Martin Castelnovo

As we mentioned, this work has been developed collaboratively with Lucas Menou and Martin Castelnovo. They allowed us to complement our experimental results with numerical indentation simulations of a thin shell. As a result, a geometrical model was developed to describe the behaviour of the indentation curves.

7.1.3.1 Geometrical model

The curves with large indentations (smaller slopes) are related to the indentation on the capsid border, and as we get closer to the centre of the capsid, the slope of the force curve becomes smaller. An interpretation of this behaviour is related to the vertical deformation generated by the cantilever tip. Figure 7.5(a) shows a scheme of two Δ - F curves with different slopes: the blue and green line represents the indentation position at the centre and border of the capsid, respectively.

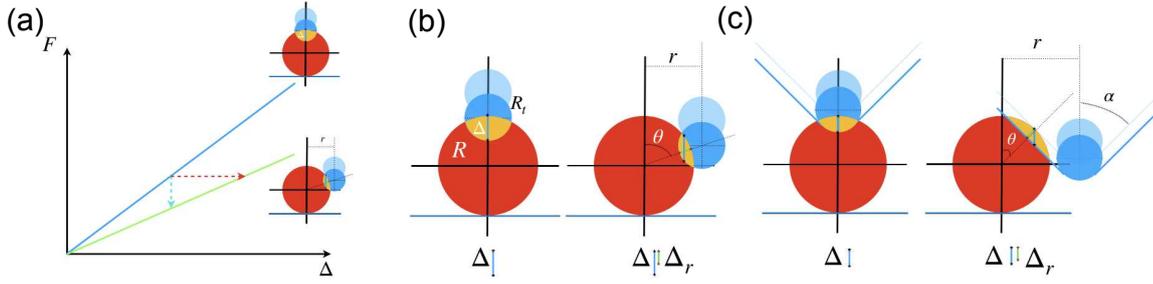


Figure 7.5: **Scheme of effective indentation for different positions and cantilevers-**
 (a) Two different indentation curves with different slopes. Indentation near to the center of the capsids (blue line) shows a higher stiffness than indentation close to the border (green line) (b) Deformation details for vertical (left) and border (right) indentations with a tip of radius R_t . While vertical indentation deformation Δ is completely measured by the cantilever, indentation close to the border generates an additional deformation Δ_r in the radial direction which is not fully transmitted to the cantilever (c) Bigger conical tip of angle α with spherical tip. In this case the conical part of the tip could have contact with the border of the capsids even if the spherical tip is not in contact.

In a sphere-sphere indentation, the indentation depth Δ that the tip generates on the capsid is mainly produced by vertical indentation at the central position. However, the real indentation depth Δ_r decreases as we move away from the centre. We can also interpret this behaviour through the intersection (yellow section in figure 7.5(b)) between the viral particle and the spherical tip of the cantilever since the area section is decreasing too. From a straightforward geometrical computation, we can determine the real indentation depth Δ_r ,

$$\Delta_r = \frac{1 + \frac{\bar{\Delta}}{R-\bar{\Delta}} - \sqrt{1 + \left(\frac{\bar{\Delta}}{R-\bar{\Delta}}\right)^2 \left(\frac{1}{\cos^2 \theta} - 1\right)}}{1 + \frac{\bar{\Delta}}{R-\bar{\Delta}}}, \quad (7.1)$$

where R is the capsid radius and the variable $\bar{\Delta}$ is defined as $\bar{\Delta} = \frac{\Delta \cos \theta}{1 + R_t/R}$. The angle θ is the angle between the initial axis of symmetry and the position r of the tip (see figure 7.5(b)), which can be written as:

$$\cos \theta = \sqrt{1 - \sin^2 \theta} = \sqrt{1 - \left(\frac{r}{R_t + R}\right)^2}, \quad (7.2)$$

where R_t is the tip radius. Assuming small indentation $\Delta \ll R$, the real indentation depth can be reduced to,

$$\Delta_r = \Delta \cos \theta = \Delta \sqrt{1 - \left(\frac{r}{R_t + R}\right)^2}. \quad (7.3)$$

The previous estimation assumes that the capsid does not move laterally upon vertical indentation. In actual nanoindentation experiments, the capsid adhesion with the substrate is not strong enough, so we noticed that the capsid was deformed and slightly shifted sideways. As we mentioned earlier, the indentation process begins with obtaining an image to locate the capsid position, then we indented and finally, we obtained another image to verify that the capsid is intact. We saw that the capsid had escaped from the imposed indentation constraint by sliding sideways on certain occasions. We think that this occurs in the weak adherence regime. As a consequence, it is essential to take horizontal displacement into account in the geometric model.

The corrected indentation depth can be rewritten assuming that in addition to the true indentation along the symmetry axis of the deformation, there is a horizontal shift Δ_h . Then, the equation 7.3 to the lowest order in Δ is written as,

$$\Delta_r = \Delta \cos \theta = \Delta \sqrt{1 - \left(\frac{r + \Delta_h}{R_t + R}\right)^2}. \quad (7.4)$$

For the capsid to hold in a fixed position, a horizontal component of the indentation force $F_{ind,h} = k_{top} \Delta_r \frac{r + \Delta_h}{R_t + R}$ must be balanced with a horizontal restoring force F_h defined as $F_h \sim k_{adh} \Delta_h$. From these two relations, we can estimate the horizontal shift as:

$$\Delta_h \sim \Delta \frac{k_{top}}{k_{adh}} \frac{r}{R_t + R} \quad (7.5)$$

From equation 7.5, we can deduce that the horizontal displacement increases with the lateral position of the tip r and decreases with the adhesion efficiency k_{adh} . For this reason, it is crucial to find an optimal deposition condition of viral capsids to avoid horizontal displacements due to vertical indentation caused by the tip. Now, we can redefine the indentation depth Δ_r considering the horizontal shift Δ_h ,

$$\Delta_r = \Delta \sqrt{1 - \left(\frac{r}{R_t + R}\right)^2 (1 + A)^2}. \quad (7.6)$$

We have rewritten the horizontal shift as $\Delta_h = Ar$, where A introduces a rescaling that depends on specific adhesion parameters and geometrical ones. It is essential to mention that the equation 7.6 is valid for a spherical tip.

We introduce a new parameter for a conical tip, the aperture angle α of the cone [Fig. 7.5(c)]. Similarly to the spherical case, the indentation depth also depends on the lateral position of the tip: however, this proof is valid for $r \leq R$. For tip position $r > (R_t + R) \cos \alpha$ or when $\theta = \pi/2 - \alpha$, the capsid is deformed by the conical part of the indenter. In this case, the indentation depth Δ_r is identical wherever the lateral position r since the curvature that plays in the indentation is zero, and then the overlap with the capsid remains constant with r . By geometrical analysis, the Δ_r is written as:

$$\Delta_r = \Delta \sin \alpha. \quad (7.7)$$

We have analytically shown that the indentation position on capsids and the tip shape play an important role in extracting the indentation depth Δ_r , which seems to be due to a purely geometric effect. The decrease in the indentation depth when the tip is off-centred also generates a decrease in the mechanical response to balance this deformation. In addition, there is a complementary horizontal response that generates capsid displacements that depend on both the indenter position and the adhesion force. It is necessary to mention that all these factors must be considered to reach an agreement between the analytical, experimental, and numerical results. The latter will be analysed below.

7.1.3.2 Simulation of AFM indentations

In the thin shell model, in particular, in the thin shells like viral capsids, the deformation of the capsid shell is modelled as the contribution of the out-of-plane bending and in-plane stretching/compression energy. The bending energy contributes to the curvature of the capsid shell and the stretching energy to the in-plane deformation. In this theory, a linear elastic response is expected when the indentation is in the order of the shell thickness [124]. Figure 7.6(a) shows the elastic energy of two indentation positions, where the maximal indentation depth is set to $\delta_{max} = 0.3R$ in the radial direction. Comparisons between the energy curve for indentations in the centre and the border shows the same behaviour as the experimental data. Therefore, the simulations reproduced the indentation experiments. To calculate the effective stiffness of the capsid, a quadratic fit of the form $E_{elas} = \frac{k}{2}\Delta^2$ was used in the energy curve. In order to reproduce the experimental results, the simulations were carried out using a conical tip (see more details about the simulations in [340]).

Figure 7.6(b) shows stiffness profiles for different trajectories by which the shell ori-

entation was imposed, such that a pentamer or a hexamer is facing up. The hexamer is delimited by three equidistant pentamers, corresponding to a triple-axis of symmetry of the icosahedron. The different trajectories carried out are pentamer-pentamer, pentamer-hexamer, hexamer-pentamer and hexamer-hexamer, where the pentamers and hexamers are represented by numbers 5 and 6, respectively. Note that the shape of the profile is a function of the mentioned trajectories. When pentamers are in the centre, the measured effective stiffness is more significant than hexamers located in the centre. We deduced a behaviour for the remaining orientation from these trajectories: orientation 2 or an edge on the capsid. When we have the hexamer-pentamer trajectory, the tip passes through (i) the centre of the hexamer, (ii) an edge, (iii) a hexamer near to the capsid border, and (iv) through a pentamer on the capsid border [Fig. 7.6(b)]. The effective stiffness starts with a value close to 1.6 N/m, followed by an increase to 1.8 N/m, and finally decreasing again to 1.6 N/m. When the stiffness increased to 1.8 N/m, the indentation was over an edge. From here, we can suggest that the face is softer than an edge; more generally, a vertex is more rigid than an edge, and the edge is more rigid than a face.

Comparing the simulations and our experimental data, we can associate a specific type of profile with a capsid orientation. The profile of the figure 7.4(d) is associated with a 5-fold symmetry axis since this profile also has the highest stiffness value; this is also consistent with that a pentamer is more rigid than a hexamer [187]. The profiles

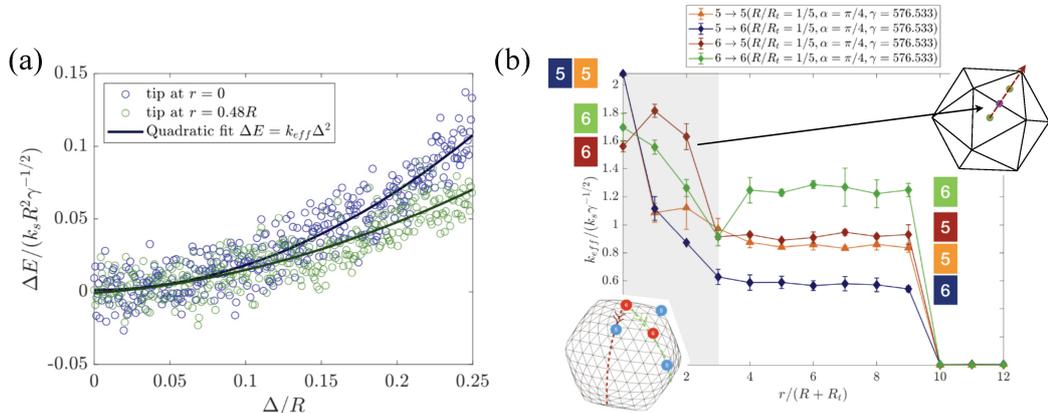


Figure 7.6: **Numerical indentation of a capsid with triangular number $T=1$** - (a) Quadratic fits of the elastic energy as a function of the indentation depth for a numerical simulation of a capsid modelled as a thin shell. Note that for indentation out of the centre, the capsid stiffness is smaller (green circles). (b) Extracted numerical stiffness as a function of the normalised indentation position, presented as a mean value of 4 repetitions with errors bars constructed as the root mean square (*rms*). The numerical stiffness was simulated for different trajectories where the numbers 5 and 6 represent the pentamers and hexamers.

with central values between the range 0.7-1.2 N/m and 0-0.5 N/m are associated with a 2-fold and 3-fold symmetry axis, respectively.

All this analysis allowed us to interpret our experimental results with the simulations and the geometric model, in particular, to identify the orientation of the capsid without using the AFM image. In addition, the indentation profile opens the opportunity to analyse geometric effects related to the position of the indenter. Next, we will study the stiffness profile by changing certain factors such as the elastic constant or the shape of the cantilever.

7.1.4 Cantilever spring constant effect

It is recommended to use a cantilever with an elastic constant similar to the capsid to perform nanoindentation experiments on viral capsids, preferably smaller to avoid damage to the capsid [13]. However, this can be a challenge if the capsid stiffness can vary from 0.6 to 1.4, depending on the orientation [156]. Supposing the cantilever is very soft, then a considerable force must be used to deform the most rigid orientation, but at the same time, that force can break the capsid with a softer orientation. Until now, we have used a cantilever can_1 with stiffness equal to 1 N/m (measurement obtained with thermal noise), and we determined the stiffness value of the capsid between 0.5 and 2 N/m. This section will analyse the indentation profile using a softer cantilever can_2 (0.4 N/m), which means it is softer than the capsid stiffness. It is expected that no differences are found since the capsid deformation is not related to how much the cantilever deforms.

As with can_1 , some force curves were analysed to determine the point where the non-linear regime RsF disappears [Fig. 7.7]. The regime RsF does not present a fixed pattern in the profile curves; it may be larger for the central curves [Fig. 7.7(b,c,e)] or it can increase as we move away from the central curves [Fig. 7.7(f)]. However, the regime ends at approximately 300 pN, not at 400 pN using can_1 . As we can see, this regime does not disappear using a softer cantilever, although it could disappear if we tried an even softer cantilever. In this case, we have to use a greater maximum deformation force than the one we used in our experiments to achieve deformation of the capsid.

Comparing the stiffness profiles for both cantilevers, we see a slight difference in the median profile [Fig. 7.8 (a)], however it does not mean that the capsid stiffness values obtained above are different in this case. The general profile shape will depend on the number of profiles that show a valley or profiles with mountain behaviour. The

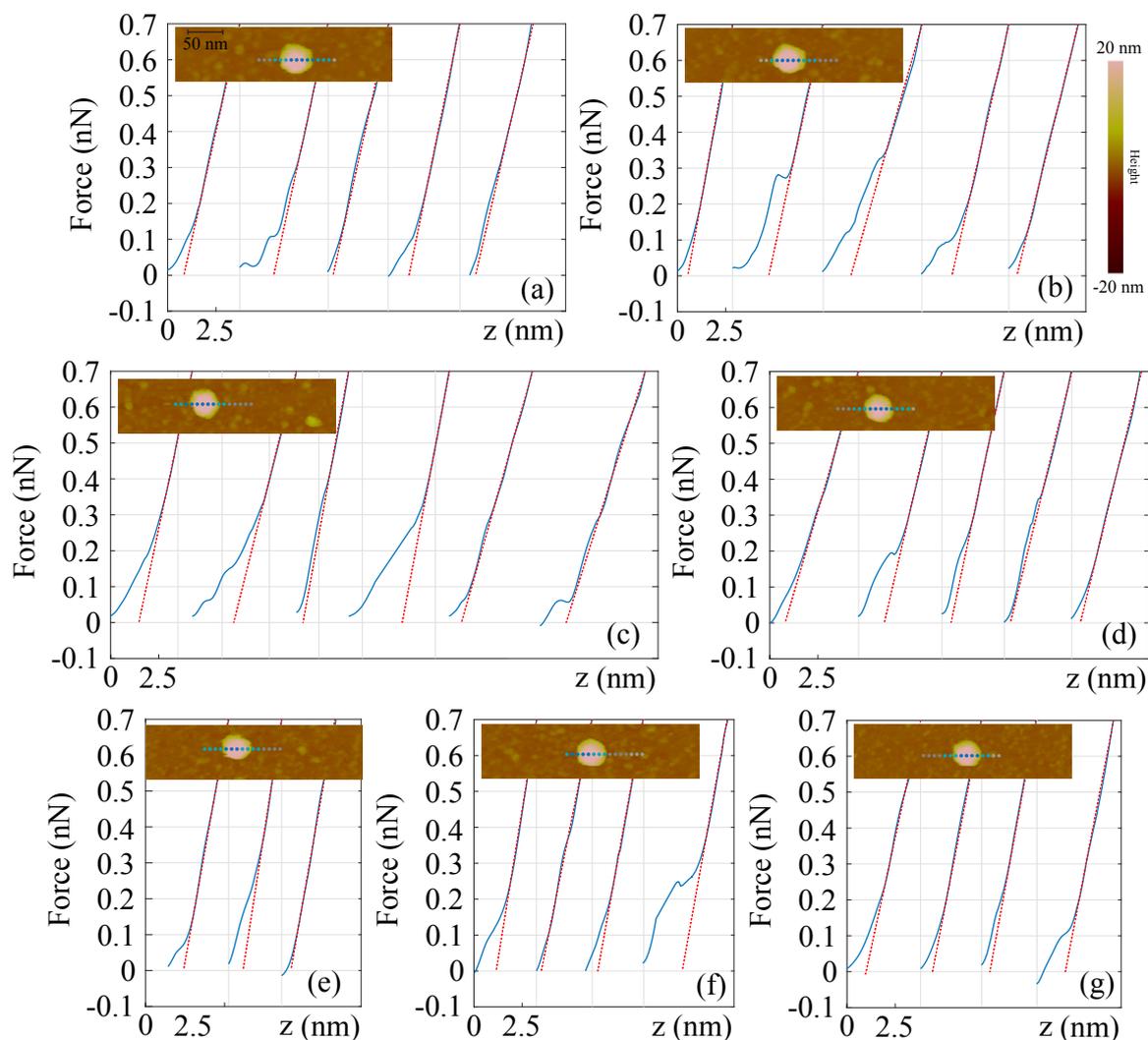


Figure 7.7: z - F curves on the AAV8 capsid top - softer cantilever- Force curves for the central points of the indentation profile (blue lines). The indentation positions are shown in the AFM image where the points are separated by colors. The grey points represent the positions on the surface and the light blue and blue points represent the positions at the border and top of the capsid. A straight red line is drawn on each z - F curve from the maximum force, 700 pN, to the point where the curve changes its slope, showing the point where the curve stops being linear.

probability density of the three central points is similar to can_1 , mainly with the peak centred at 0.5, the second peak disappears here [Fig. 7.8(b)]. To compare the stiffness groups between can_1 and can_2 , the coloured sections of can_1 have been replicated in the cdf of can_2 [Fig. 7.8(c)]. The two groups (green and purple sections) identified before are not completely clear; the curve changes slightly in slope. Nevertheless, according to the stiffness groups, the profiles are well differentiable, as we will see now.

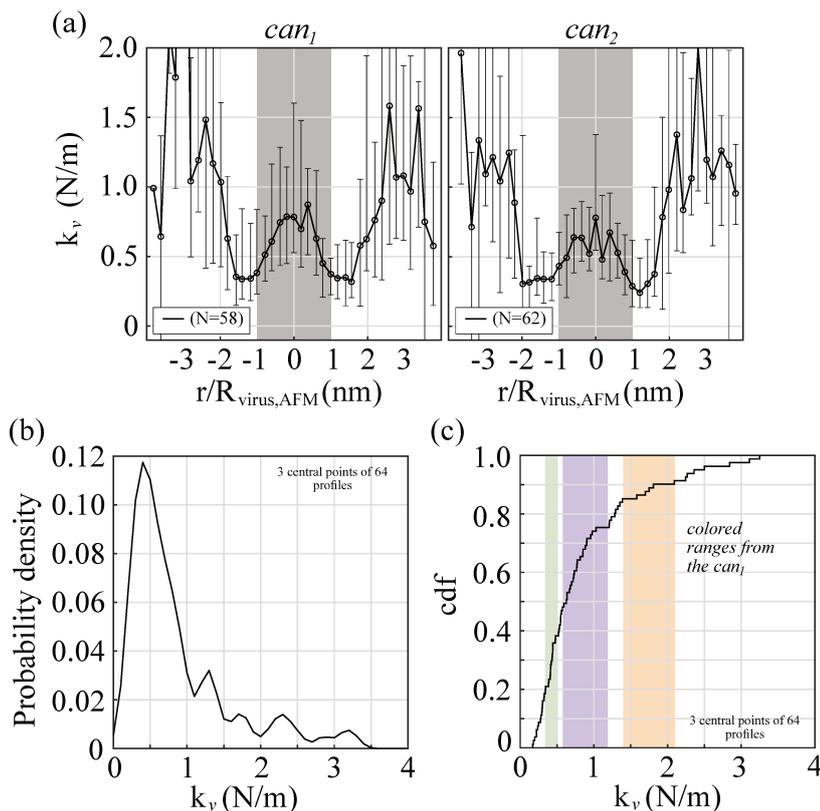


Figure 7.8: **AAV8: three central points of the indentation profile using can_2** - (a) Stiffness profile for can_1 and can_2 . The black circle represent the median of the data, and the upper-lower error bars are the 75% and 25% of the points, respectively. The points contained in the coloured region are indentation points on the capsid top. (b) Probability density and (c) the three central points cumulative distribution function, where 62 capsids have been indented. The positions of the coloured sections are those chosen for can_1 ; thus, we can compare the new behaviour of cdf with that obtained with can_1 .

As expected, we found the same types of profiles and the central stiffness values as with the can_1 [Fig. 7.7]. From the 62 indentation profiles, we have excluded nine profiles that did not pass vertically through the centre of the capsid. We should note that the analysis through an indentation profile allows us to address both axes through which the capsid can move to extract a more precise value of the stiffness capsid. Analysing in detail the profiles for both cantilevers, we noted that the valley-shaped profiles have a slight difference, specifically the ends of the valley, figure 7.4(b,c) and 7.9(b,c). The difference is due to each group profiles number; however, the values found are contained in the error bars.

The results show that regardless of the stiffness of the cantilever, the capsid stiffness not change. However, it must be taken into account when the capsid is deformed. As we saw, the force range for a cantilever was 400-700 pN, while for the softer cantilever,

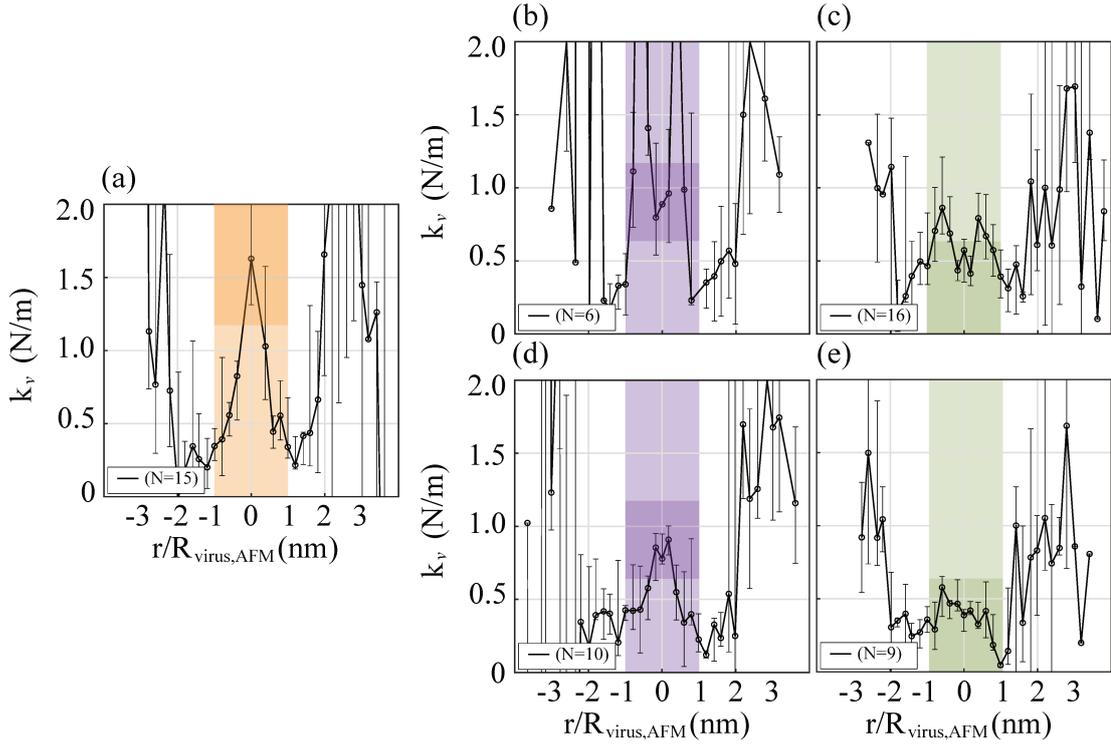


Figure 7.9: **AAV8: stiffness profile according to the value of the central positions using can₂**- The profiles show certain behaviours: (b,c) the central values are smaller than at the border, showing a valley in the centre. The other type shows larger values in the centre (a,d,e). The coloured zones represent the central point of the profile and the darker rectangle show the profiles according to a certain range of stiffness values, where the green, purple and damask colours contain values ranging from 0-0.7 N/m, 0.7-1.2 N/m and 1.2-2 N/m, respectively. The black circle represent the median of the data, and the upper-lower error bars are the 75% and 25% of the points, respectively.

the range was 300-700 pN.

7.1.5 Cantilever tip shape effect

The previous analyses were performed using a cantilever tip smaller than the capsid ($R_{can} \sim 2$ nm and $R_{cap} \sim 13$ nm). Hence, we could mention that the deformations that have been made on the capsid are local deformations. We developed indentation experiments with a bigger cantilever tip than the capsid size to understand the effect of local and global deformations. We used two conical cantilevers with the same stiffness value, but they have tip radius of 2 and 20 nm, can₁ and can_{1a} respectively. Although the can_{1a} also has a conical geometry, the tip is much larger than the capsid $R_{cap} \sim 13$ nm. Therefore, the contact will be of the sphere-sphere type.

The indentation profile shows a displacement in all stiffness values [Fig. 7.10(a)]. Consequently, the centre values have also been displaced. Figure 7.10(b) shows the probability density for both cantilevers, where the black and grey lines are the data for can_{1a} and can_1 , respectively. In this case, the subgroups have been redefined where we have marked several ranges with grey sections in figure 7.10(c).

Following the protocol, we identified the same profiles, but the central points differed from the central points with the smallest tip [Fig. 7.11]. The 5-fold symmetry axis profile increased to 25% at the profile centre [Fig. 7.11(a)]. Aznar *et al.*, [341] analysed the effect of the cantilever tip size for a capsid with a triangular value equal to $T=7$. The authors mainly showed that indentations on the 5-fold symmetry axis show no stiffness variations unless the cantilever tip is large enough to deform more than one capsomer: in our case, the cantilever tip can deform more than one capsomer [Fig.

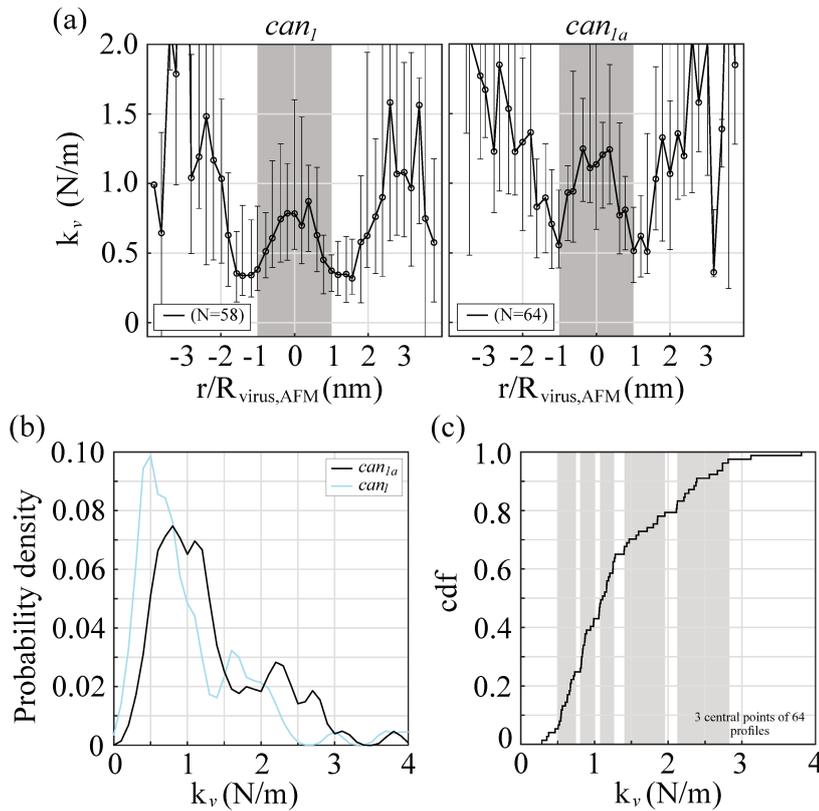


Figure 7.10: **AAV8: three central points of the indentation profile using can_2** - (a) Stiffness profile for can_1 and can_{1a} . The black circle represent the median of the data, and the upper-lower error bars are the 75% and 25% of the points, respectively. The points contained in the coloured region are indentation points on the capsid top. (b) Probability density of the three central points for both cantilevers and (c) the cumulative distribution function of the three central points, where 64 capsids have been indented. The grey sections show the subgroups of the stiffness values for can_{1a} .

7.11(a)]. The profiles associated with 2-fold symmetry axis, valley and mountains shape, show an increase in stiffness values [Fig. 7.11(b,d)]. In contrast, the valley-shaped profiles associated with a 3-fold symmetry axis do not show variations between the small and large cantilever tip [Fig. 7.11(c)], although there are variations for the mountain-shaped profile [Fig. 7.11(d)].

Finally, we have shown experimentally that the cantilever tip size can influence the value of the stiffness obtained in indentation experiments. As mentioned by Aznar *et al.* [341], the stiffness value increases continuously with the tip radius up to converge to a constant value for tips that tend to behave as flat. Hence, the measure will depend on the area (or capsomeres) that the tip can form.

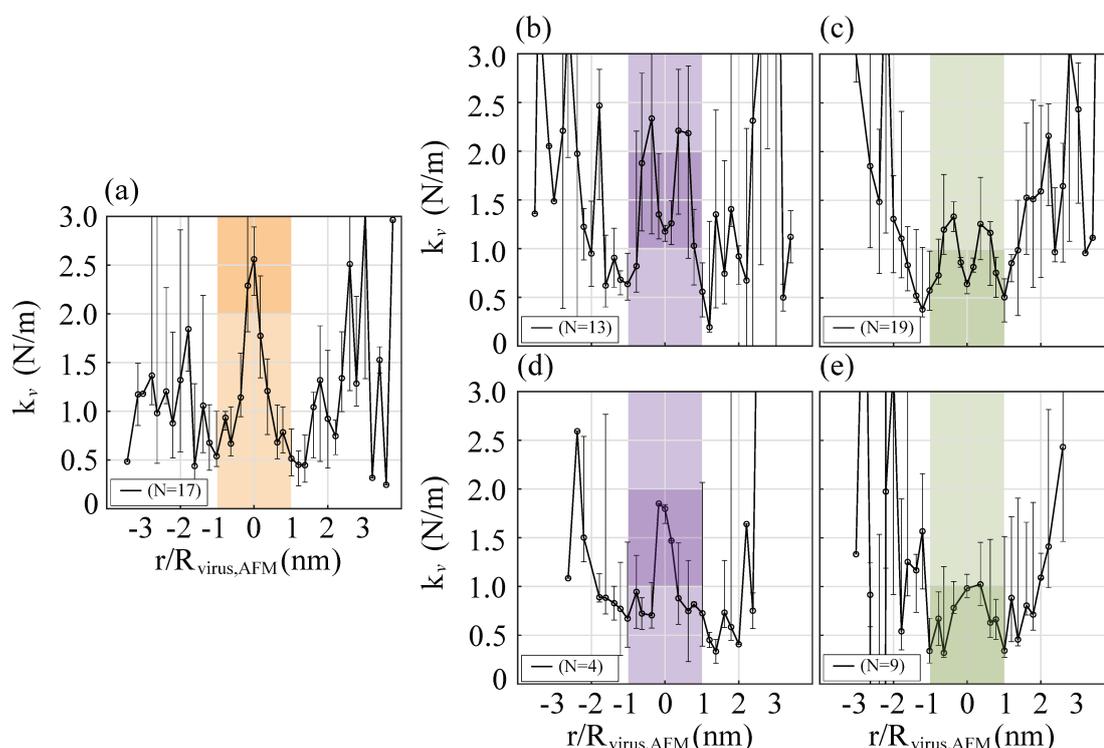


Figure 7.11: **AAV8: stiffness profile according to the value of the central positions using can_{1a}** - The profiles show certain behaviours: (b,c) the central values are smaller than at the border, showing a valley in the centre. The other type shows larger values in the centre (a,d,e) The coloured zones represent the central point of the profile and the darker rectangle show the profiles according to a certain range of stiffness values, where the green, purple and damask colours contain values ranging from 0-1 N/m, 1-2 N/m and 2-3 N/m, respectively. The black circle represent the median of the data, and the upper-lower error bars are the 75% and 25% of the points, respectively.

7.2 AAV2 and AAV9 capsids

Once the protocol for measuring capsid stiffness has been established, this thesis project focuses on more applied questions. AAV virus serotypes have different receptors for cell binding and tissue tropism. After the capsid crosses the plasma barrier of the cell, the capsid crosses a second nuclear membrane and then release the genome. Each stage of the viral cycle is crucial to generate efficient transduction. However, these processes are not fully understood. Hence, we think that mechanical properties (capsid stiffness) can play an essential role in describing the viral capsid behaviour during the viral cycle. This section will analyse the AAV2 and AAV8 capsids stiffness, although discussing the effect of mechanical properties on the viral cycle will be addressed in the destabilisation section. Here, we will only focus on determining the stiffness of the capsid.

In the first instance, we compared the stiffness profiles, and we see that the central

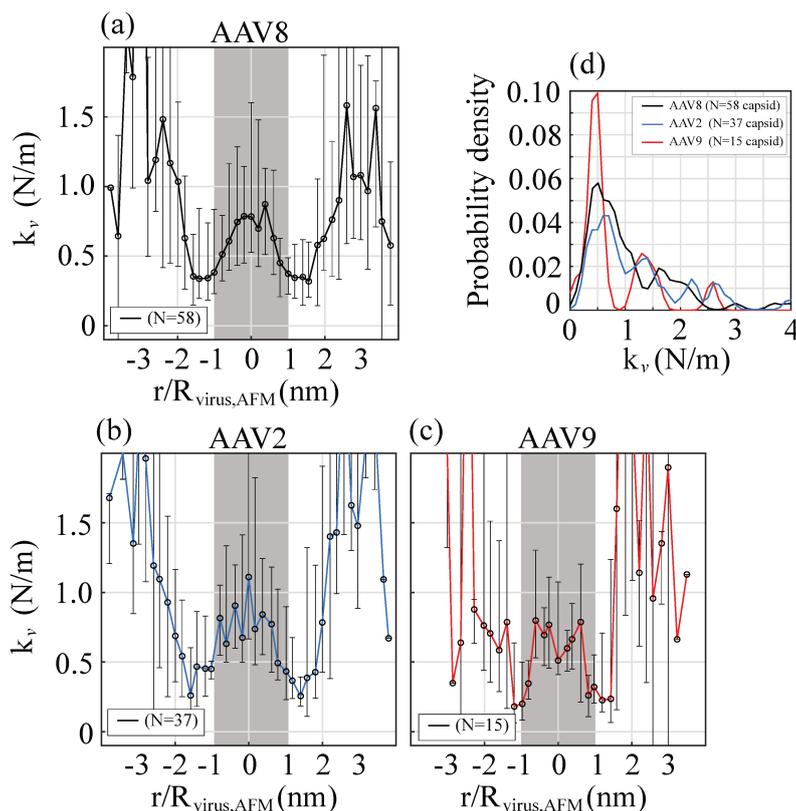


Figure 7.12: **Stiffness profiles for AAV2, AAV8 and AAV9 capsids** - Stiffness profile for (a) AAV8, (b) AAV2 and (c) AAV9. The black circle represent the median of the data and the upper-lower error bars are the 75% and 25% of the points, respectively. The points contained in the colored region are indentation points on the capsid top. (d) Probability density of the three central points for both the conditions mentioned.

values are higher for AAV2, followed by AAV8 and finally AAV9 as the softest one [Fig. 7.12(a,b,c)]. Unlike AAV8, which we indented 58 capsids, only 37 and 15 capsids were measured for AAV2 and AAV9, respectively. The probability density of the three central points shows the same behaviour as the profile. The subgroups shift to smaller values following the order: AAV2-AAV8-AAV9 [Fig. 7.12(d)].

Figure 7.13 show all the profiles types for AAV2, AAV8 and AAV9. Because we only indented 15 capsids for AAV9, we do not have enough data to identify different types of profiles. Comparing only the profiles with a valley shape for AAV2 and AAV8, we see no great differences with the capsids associated with the 3-fold symmetry axis [Fig. 7.13(d)]. On the contrary, the 2-fold symmetry axis capsids presented differences where the central values follow the order already mentioned [Fig. 7.13(b)]. For the 5-fold symmetry axis, we see that AAV8 is softer than AAV2 [Fig. 7.13 (a)]. The AAV9 capsid is more complicated since we are not completely sure that this behaviour is associated with a 5-fold or 2-fold symmetry axis since we do not have enough data to compare. However, following the idea that AAV9 appears to be softer than the other two capsids, this profile should be associated with orientation 5.

Up to this point in the analysis, we show the AAV2 is more rigid than AAV8 and AAV9, and the capsid stiffness depends on the position where it is indented, which seems to be due to geometric components. It is possible to identify the capsid orientation using an indentation profile. We found that the capsid stiffness does not change using a softer cantilever to deform the capsid, although it is essential to know the range of capsid deformation. Also, we analyse the effect of the cantilever tip size, where we found that the stiffness increases for larger tips. The following section will study Hepatitis B virus HBV capsids to determine if we find this same dependence with indentation position. Additionally, the HBV capsid is similar in size to the AAV capsid, which will allow us to study whether the possible differences in mechanical properties are relevant during the nuclear entry process and release of the viral genome.

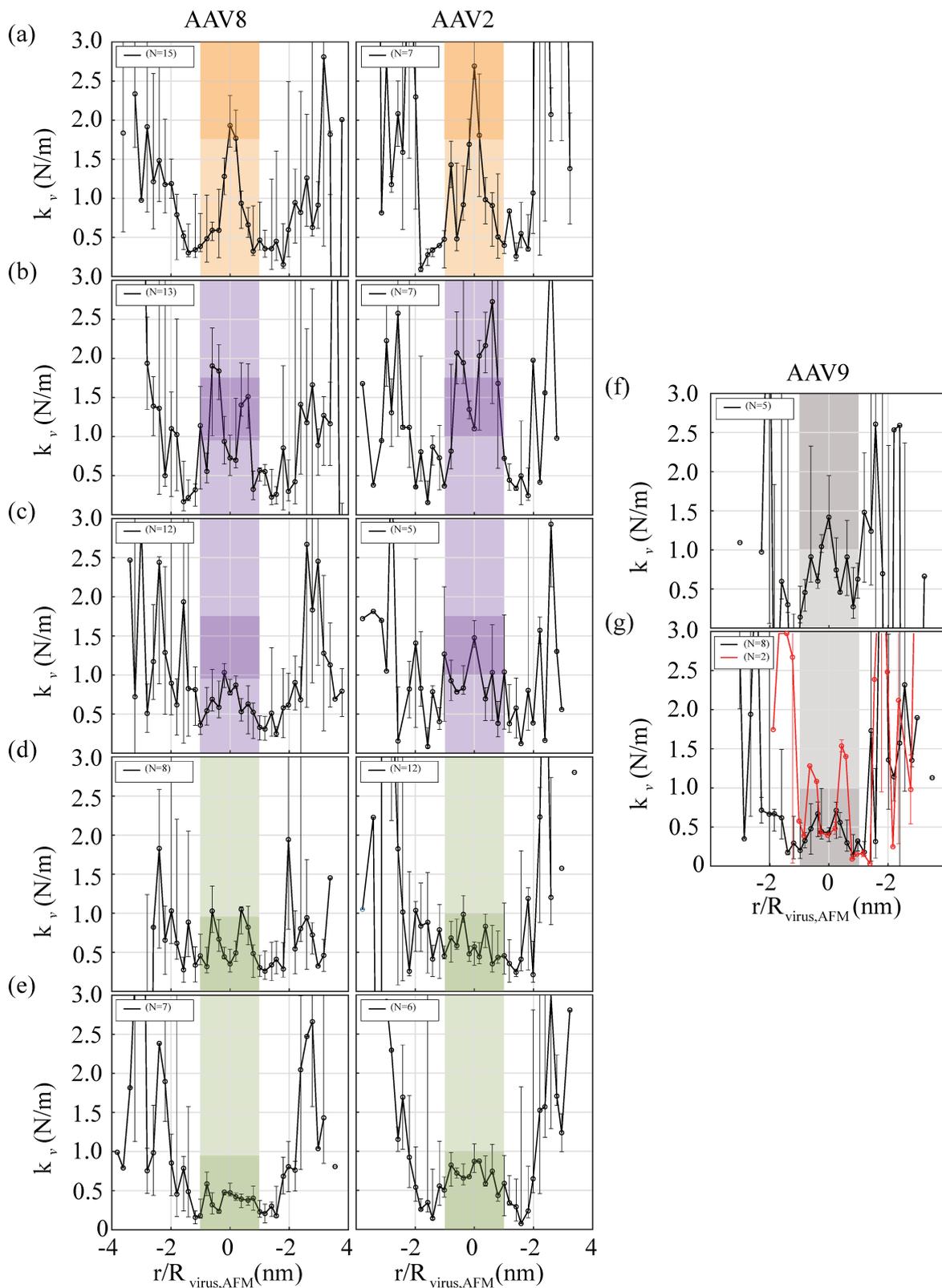


Figure 7.13: **Stiffness profiles for AAV2, AAV8 and AAV9 capsids-** (a) Stiffness profile of AAV2 (a), AAV8 (b) and AAV9 (c). The black circle represent the median of the data and the upper-lower error bars are the 75% and 25% of the points, respectively. The points contained in the gray region are indentation points on the capsid.

7.3 HBV capsid

Before determining the capsid stiffness, we will show some AFM images containing the capsid and their respective capsid height probability density. As we mentioned before, the height parameters are important to verify the capsid conditions. then, we applied the same protocol as AAV to determine the stiffness profile capsid.

The study of the HBV capsid mechanical properties is a work carried out in collaboration with Lauriane Lecoq at the Institut de Biologie et Chimie des Protéines (IBCP) in Lyon. Lauriane has developed a method to assemble HBV capsids from the Cp183 proteins inside bacteria. No viral pre-genomic RNA is present during self-assembly inside the bacteria, so instead, bacterial RNA is packaged during the Cp183 self-assembly process. For the detailed protocol see for example Wang *et al.*, 2019, Chevreuil *et al.*, 2020 and Lecoq *et al.*, 2020 [342–344].

7.3.1 Morphology of HBV capsid

Images were obtained using PeakForce mode in liquid. The cantilevers used, Bruker ScanAsyst-fluid-plus, have a triangular geometry with a nominal radius of 2 nm. The nominal stiffness is 0.7; nevertheless, the cantilever spring constant k_c measured by thermal noise was around 1 N/m. For larger images, the scanned images are squares of 2 μm side, with a resolution of 512x512 pixels and a scan frequency of 3 Hz [Fig. 6.4(a)]. The maximal force applied was 1000 pN. For single capsid AFM images, we first scanned a square image with 2 μm size, with a resolution of 256x256 pixels at a scan rate of 2 Hz. The maximum force applied was set to 200 pN; we do not break the capsid with this force. The image was limited to 300 nm in size, with the capsid on the image centre. Figure 6.9 shows a screenshot of the Nanoscope software where we see how the capsid is being scanned. All the parameters used here are the same as the AFM imaging of AAV.

The HBV capsid size between protrusion is 34 nm, and the protrusion length is 2 nm, approximately [148, 345]. Then, the height we consider here for HBV is 32 nm. Comparing the height obtained from the AFM images between the two methods: (i) from a capsids population [Fig. 7.14(a)] or (ii) from single capsids [Fig. 7.14(b)], a small difference is identified. The height capsid probability density shows that the average value of 24.5 ± 1.7 nm from (i) is smaller than the height from (ii) 25.4 ± 2.3 . Also, we noticed that the reduced height is ~ 7 nm, and it is almost three times the height reduction founded for AAV. This could be because the capsid stiffness of HBV

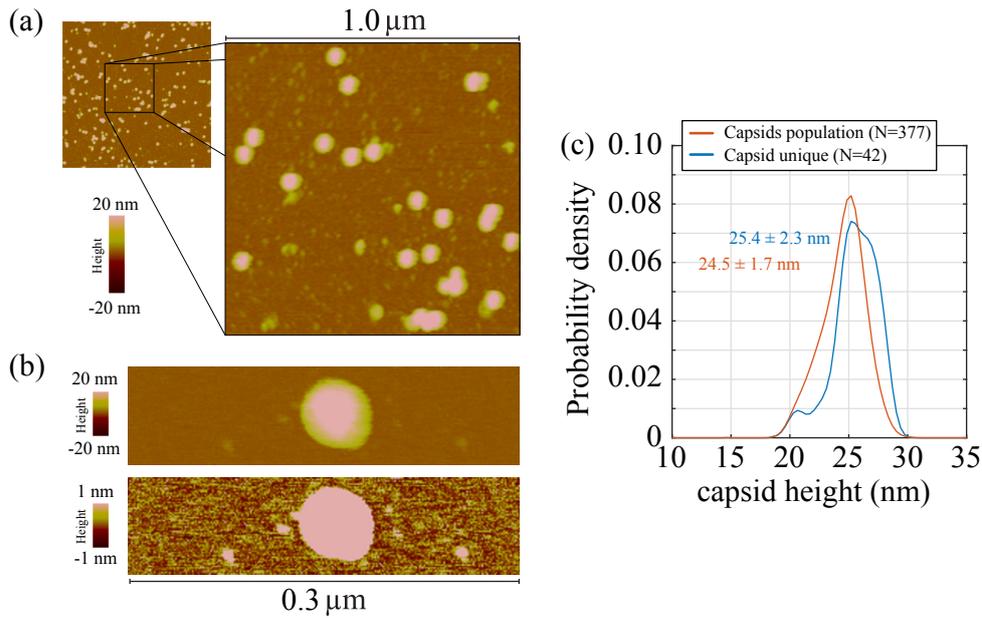


Figure 7.14: **HBV capsid height** - (a) AFM image of several HBV capsids. (b) AFM image of a single HBV capsid with two different z-scale. (c) Probability density of the capsid height determined by using the population analysis (red line, section 6.4) and using isolated capsid analysis (blue line, section 6.5).

and AAV2 (two-fold axis) are 0.1 N/m 1.5 N/m, respectively [175, 195]. Therefore, HBV capsid is expected to be much softer and more affected by adhesion force.

7.3.2 Stiffness profile according to the values of central points

Performing the same analysis that we performed for AAV capsid (section 6.9), we plotted the central curves for several indentation profiles [Fig. 7.15]. The visualisation of these curves allows us to identify if the capsid has a linear response. Unlike AAV force curves, HBV capsid has a RsF less pronounced at small forces. This could be related to the HBV capsid protrusions are homogeneously distributed on the surface and with a constant size of 2 nm. It would be interesting to characterise these protrusions mechanical properties and establish a role with biological processes. To avoid the regime at small forces RsF, we used the curve from 300 pN until 700 pN.

The stiffness profile in figure 7.16(a) shown a position-dependent stiffness as in AAVs capsid, although the stiffness values on the capsid are smaller. Figure 7.16(b) shows the probability density of the three central points of the stiffness profile in which we identified two main peaks. The first peak is centered at 0.16 N/m approximately and a second peak at 0.35 N/m. The three-point cumulative distribution function

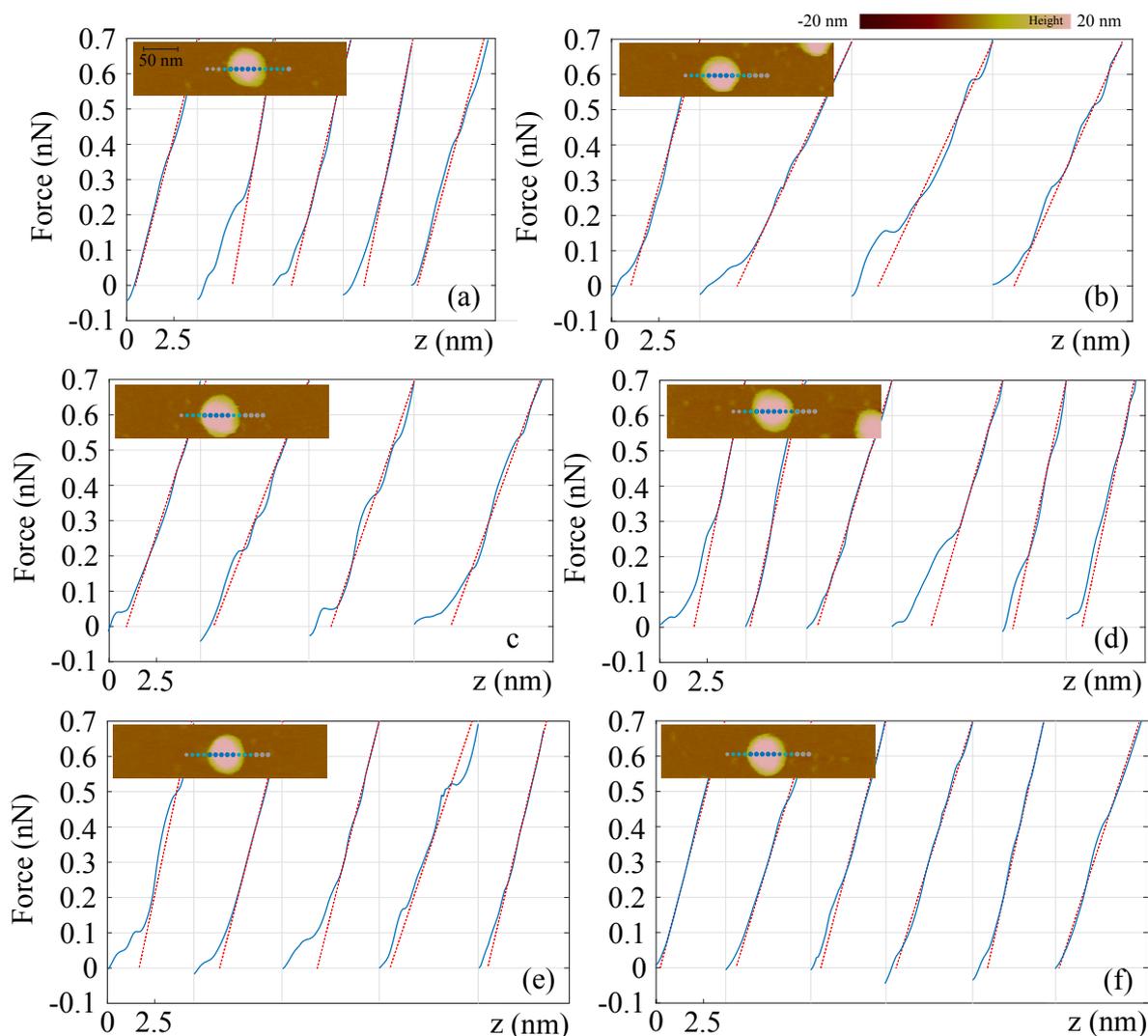


Figure 7.15: z - F curves on the HBV capsid top - Force curves for the central points of the indentation profile (blue lines). The indentation positions are shown in the AFM image where the points are separated by colors. The grey points represent the positions on the surface and the light blue and blue points represent the positions at the border and top of the capsid. A straight red line is drawn on each z - F curve from the maximum force, 700 pN, to the point where the curve changes its slope, showing the point where the curve stops being linear.

shows a group between 0-0.16 N/m that we do not identify through the probability density height [Fig. 7.16(c)].

Looking at the profiles in figure 7.17, we found different profiles. Figure 7.17(a,b) shows a profiles with a single main peak while 7.17(c,d,e) have two centered peaks. Comparing our results with those obtained by [175] (capsid stiffness ~ 0.1 N/m), we see that our experimental values are higher at the centre of the profile. The other

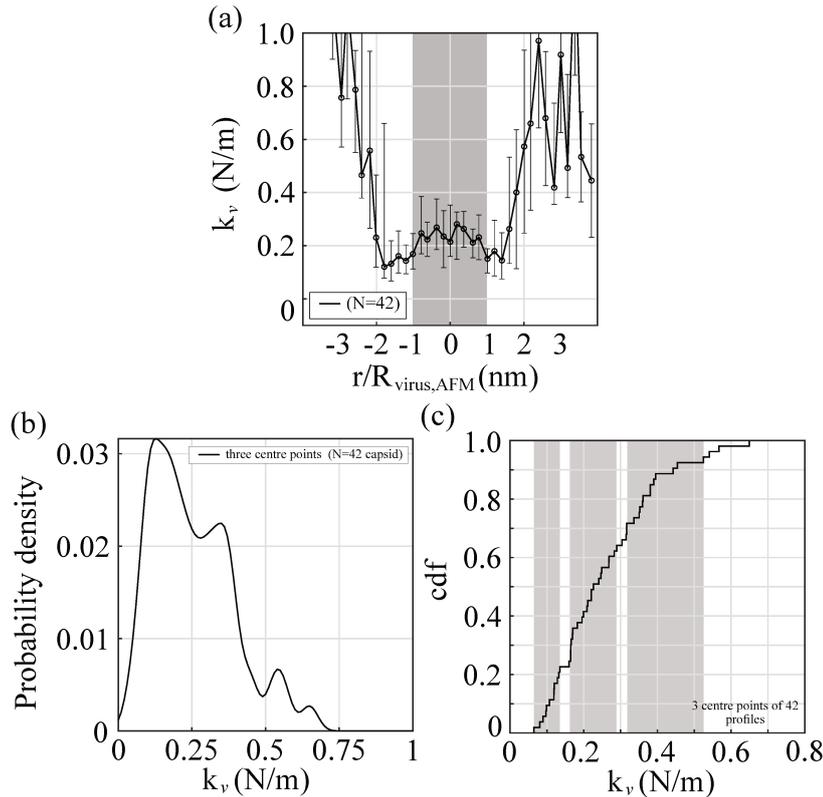


Figure 7.16: **HBV: stiffness profile and the centre points of the profile** - (a) Stiffness profile for HBV. The black circle represent the median of the data and the upper-lower error bars are the 75% and 25% of the points, respectively. The points contained in the coloured region are indentation points on the capsid top. (b) Probability density of the three central points. (c) The cumulative distribution function of the three central points, where 62 capsids have been indented. The grey sections show the subgroups of the stiffness values.

difference is that they are considered stiffness values of the centre and border of the capsid and values of all orientations, giving them a stiffness value smaller than ours. Although we can differentiate the capsid orientation with the indentation profile, in this case, we do not have enough data to associate a specific orientation. Also, association with the simulations performed by Lucas Menou and Martin Castelnovo [Fig. 7.17(f)] is complicate because in the simulation the ratio between the capsid radius R_v and the cantilever tip radius R_t is $R_t/R_v = 1/5$, while in our experiments $R_t/R_v = 1/8$. The bigger tip can result in a loss of resolution in the simulation, not allowing the differentiation between pentamers and hexamers.

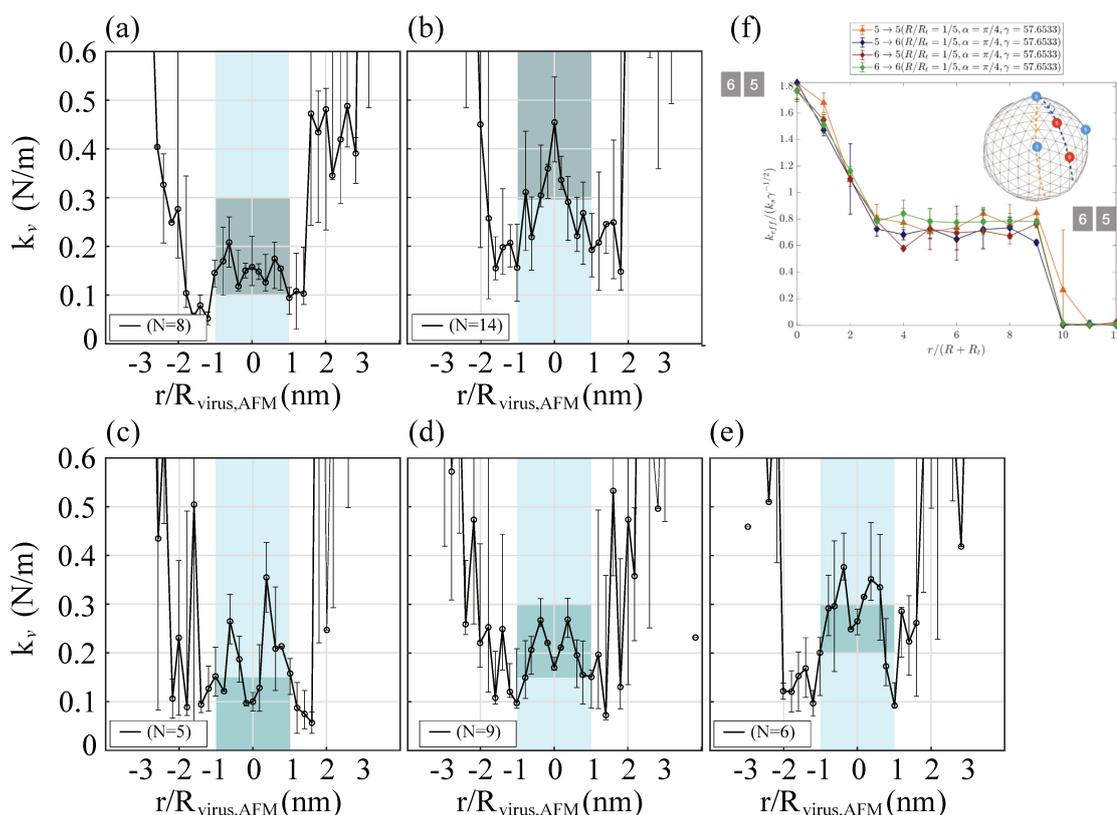


Figure 7.17: **AAV8: stiffness profile according to the value of the central positions**
- The profiles show certain behaviours: (a,b) The stiffness value at the centre are larger than at the border. (c,d,e) The stiffness values are smaller than at the border, showing a valley in the centre. The coloured zones represent the profile central point, and the darker rectangles show the profiles according to a certain range of stiffness values. The black circle represent the median of the data, and the upper-lower error bars are the 75% and 25% of the points, respectively. (f) The stiffness profiles corresponding to particular paths along the shell surface is shown for a spherical shell. Four trajectories have been considered: pentamer-pentamer, pentamer-hexamer, hexamer-hexamer and hexamer-pentamer and the profiles are similar.

7.3.3 Influence of capsid facet in the stiffness profile

We could obtain a stiffness profile with a more precisely stiffness capsid from the analysis protocol. Comparing both stiffness profiles directly, we noted that AAV stiffness values are bigger than HBV values [Fig. 7.18(a)]. Additionally, the stiffness value at the centre is approximately two times the border value for both capsids. Also, we noted that there are bigger fluctuations in the stiffness value for AAV than for HBV [error bars in figure 7.18(a)]. Figures 7.18(b,c) show the stiffness as a function of the normalised position obtained by numerical simulations. In particular, figure 7.18(b) shows a simulated profile for a faceted capsid (c_1) and figure 7.18(c) shows a profile

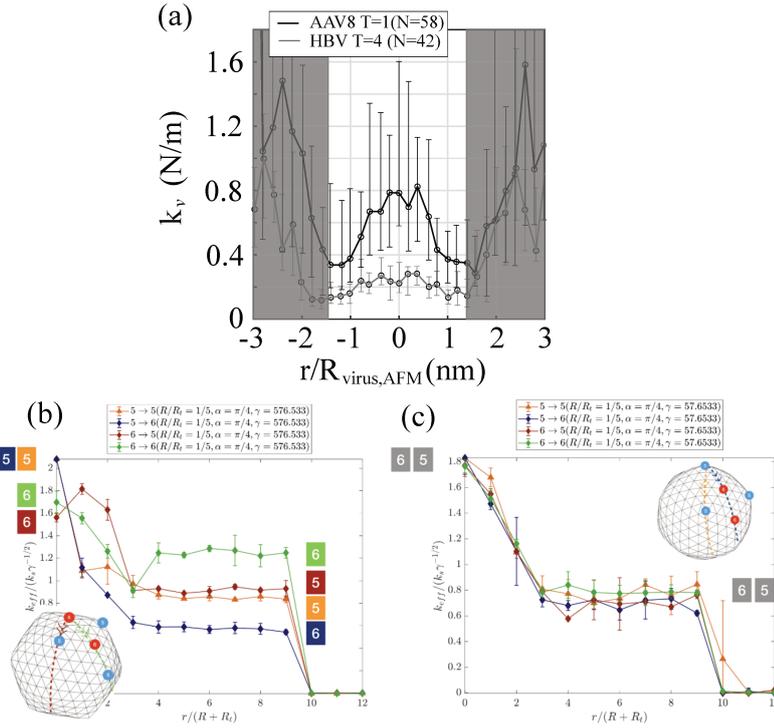


Figure 7.18: **Comparing AAV and HBV capsids profiles** - (a) Stiffness profile of AAV and HBV capsids. (b,c) The stiffness profiles corresponding to particular paths along the shell surface are shown for spherical shell with $\gamma \sim 58$ (b) or faceted shells with $\gamma \sim 580$ (c). Four trajectories have been considered: pentamer-pentamer, pentamer-hexamer, hexamer-hexamer and hexamer-pentamer and the profiles are similar.

for a capsid less faceted (c_2), with an aspect more spherical. From the simulation, we know that a more faceted capsid c_1 exhibit a bigger difference between the orientations followed by the indenter [Fig. 7.18(b)]. On the contrary, c_2 have the same profiles for all different trajectories [Fig. 7.18(c)]. This effect could explain the fluctuations of the stiffness in our experimental results. We have bigger error bars for AAV, the most faceted capsid T=1, and smaller error bars for HBV capsids, less faceted capsids. To summarise, we found that less faceted capsids, that is, capsids that have a spherical appearance, are more difficult to differentiate their orientation. The sensitivity to detect orientation is a function of the R_t/R ratio, specifically related to the geometric limitation; regardless of the indentation position, a large enough point will deform vertices, edges, and faces simultaneously.

7.4 Conclusions

To calculate the capsid stiffness from the force curves, we used two methods: (i) Hooke's law for two springs in series k_{v1} , and (ii) we isolated the capsid deformation from the indentation curve by subtracting a reference curve who have the cantilever deformation k_{v2} . We found negligible differences comparing the approximated theoretical model (i) and the stiffness calculated from the $\delta - F$ curves (ii). The force curves show that for a maximum indentation force, the capsid deforms more at the border. A simple geometrical interpretation can explain this effect: the indentation that the tip generates on the capsid is mainly produced by vertical indentation at the central position; however, the actual indentation decreases as we move away from the centre. Also, we can associate this effect through the intersection area between the viral capsid and the tip section deforming the capsid. This intersection area decreases as it moves away from the centre too. Consequently, the experimental stiffness profiles show that the capsid is stiffer in the centre than at the border. Additionally, we have determined the stiffness of some AAV serotypes, and we show that the AAV2 capsid is stiffer than AAV8, and AAV8 is stiffer than the AAV9 capsid.

The position-dependence was corroborated with numerical simulations carried out by Lucas Menou and Martin Castelnovo, where they analysed different trajectories as the experimental indentation protocol. Our analysis method allowed us to obtain the stiffness in the capsid centre. We were also able to separate the stiffness profile based on the orientation of the capsid. We classify each effective orientation stiffness, with the 5-fold symmetry axis being the most rigid, followed by 2-fold and 3-fold symmetry axis being the softest. Furthermore, our analysis method is robust to changes in the stiffness of the cantilever. Analysis performed with a larger tip showed that the measured capsid stiffness increased. The contact area could explain this change in the stiffness value that the cantilever tip deforms. That is, the larger the cantilever tip, the greater the deformation area. Consequently, a greater stiffness is obtained because all the capsid structure is deformed at the same time.

Comparing the stiffness profiles between the capsids of AAV8 (60 capsid proteins) and HBV (240 capsid proteins), we found that more faceted capsids turn out to be more homogeneous when facing orientation changes. This can be seen in the smaller fluctuations of the stiffness values in the profiles. The simulation results show that for less faceted capsids, there is no difference between the orientations. However, our experimental measures show a difference in the profiles according to the capsids orientation; nevertheless, these differences are smaller than the more faceted capsid (AAV8). The difference between the simulation and the experimental work suggests

a dependence on the tip size, as seen in Aznar *et al.*, [341]. As a perspective, an indentation could be performed over the HBV capsid, varying the tip size. We expect then to observe a collapse of the profiles.

Chapter 8

Thermal stability of AAV and HBV capsids

Index

Summary	190
8.1 Deposition conditions and AFM imaging	191
8.2 Genome uncoating for HBV capsid	192
8.3 Genome uncoating of AAV capsids	194
8.4 Conclusions	196

Summary

This section is focused on understanding the genome release process by heating the capsid and establishing a relationship with the mechanical properties (obtained in the previous section). Typically two physical parameters in the literature are related to nanomechanics measurements and may play a role in the genome release process. One of them is their ability to deform the capsid (stiffness), and the maximum force that the capsid can resist before breaking. The last parameter is more complicated to measure through nanoindentation experiments. The force to break capsid is identified as a discontinuity in the z - F curves. However, it is not easy to indent right in the centre due to the AFM piezo stage thermal drift. It is also impossible to use the indentation profile because the capsid first breaks at the border before reaching the centre.

Furthermore, our interpretation will focus on the capsid stiffness to understand the

capsid stability, that is, its ability to uncoat its genome. To mimic uncoating in vitro, we heat the capsid and then use AFM imaging to see if the capsid morphology changes and if DNA leaks out of the capsid. The capsid analysed are HBV and three AAV serotypes: AAV2, AAV8 and AAV9.

8.1 Deposition conditions and AFM imaging

The stock solution of viral capsids was initially diluted in Tris 10mM (pH 7.4), (i) then, the sample was incubated at certain temperatures for a given time. For HBV, the sample was incubated at 70, 75 and 80°C for 15 min. Another condition has been used: the sample was incubated for 15, 20, 25 and 30 min at 80°C. For AAV2, the sample was incubated at 50, 60 and 75°C for 15 min. After that, the reaction was quenched for a few seconds on ice. As for the protocol developed in [280], (ii) we added MgCl₂ 5mM to favour the DNA adhesion on the mica surface. (iii) The sample was deposited onto a freshly cleaved mica surface for 5 min before rinsing with 1 ml of ultrapure water and gently drying by nitrogen flow.

The images have been obtained using Tapping Mode in air. The cantilever used, TESP-300, has a nominal stiffness of 40 N/m with a resonant frequency of ~ 300 kHz. The images obtained have 3 μm in size, with a resolution of 512×512 and a scan frequency of 2 Hz.

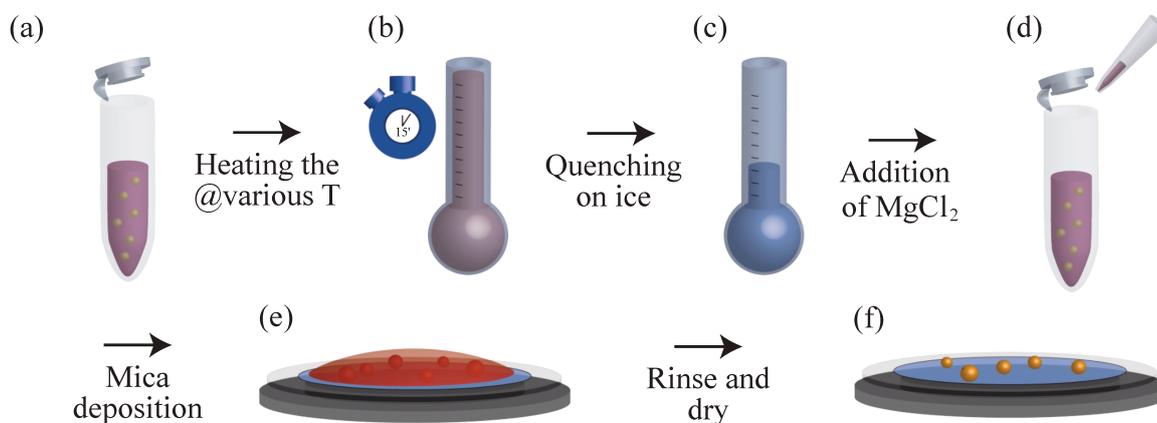


Figure 8.1: **Stability Protocol** - Illustration of the protocol of stability for sample preparation and deposition. Adapted from [280].

8.2 Genome uncoating for HBV capsid

When the temperature is varied from 70°C to 80°C (15 min of incubation) [Fig. 8.2(a)], no significant differences can be seen in the AFM images. The capsid height probability density shows the same results, where the average height (around 12 nm) is the same for the mentioned conditions [Fig. 8.2(b)]. The HBV capsid exhibits great stability to temperature effects. As we showed in the previous section, the HBV capsid is much softer than AAV capsids. The softer stiffness of the HBV capsid may explain the large reduction in height when imaging in air. The height of the capsid in fluid gives us a more real value, although it is also affected by the softer stiffness of HBV capsid [Fig. 7.14]. Finally, these results show the high thermal stability of HBV capsid, related to

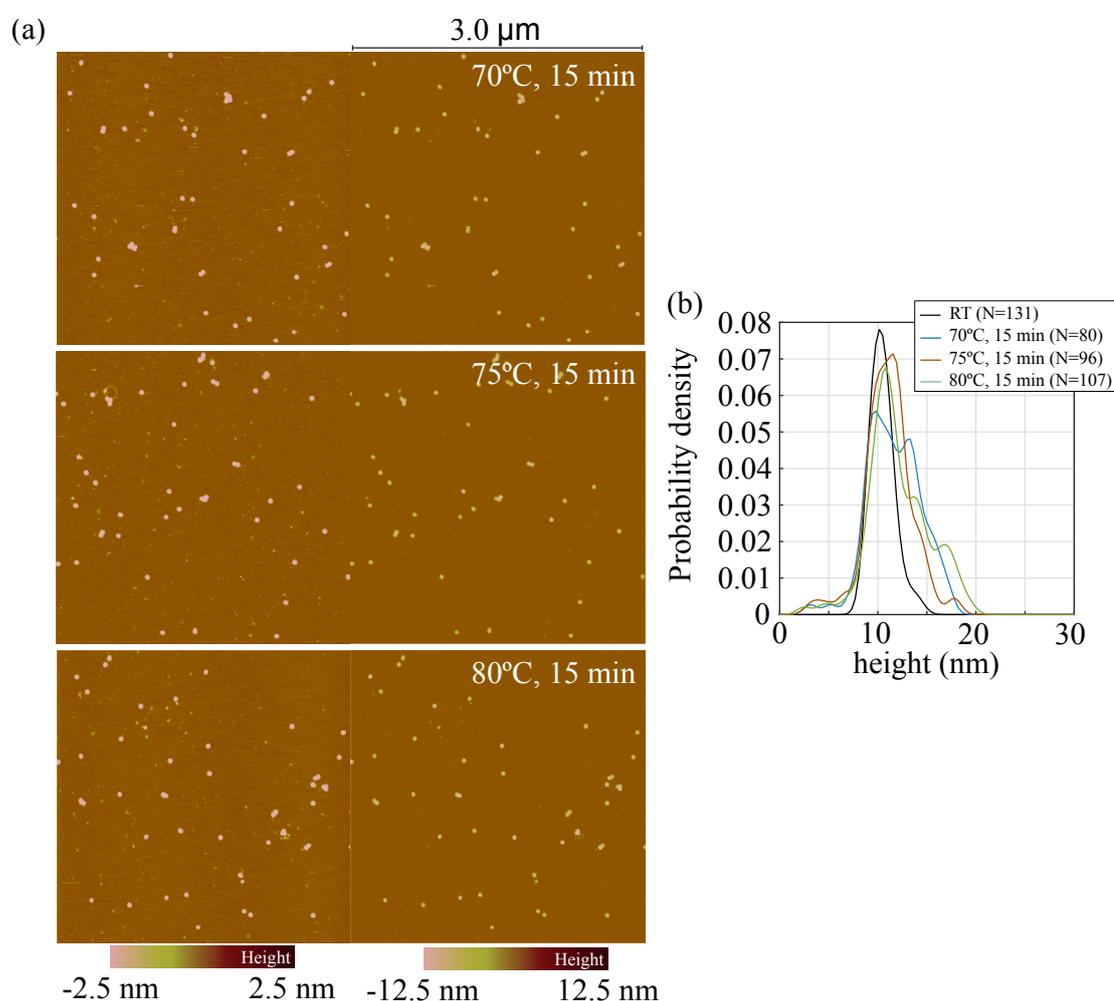


Figure 8.2: **AFM imaging of heat-induced HBV capsids destabilisation 1** - (a) AFM images of HBV capsids for different heating conditions: 70, 75 and 80°C during 15 min of incubation. (b) Probability density of the capsid height for the three conditions and Room Temperature as a comparison.

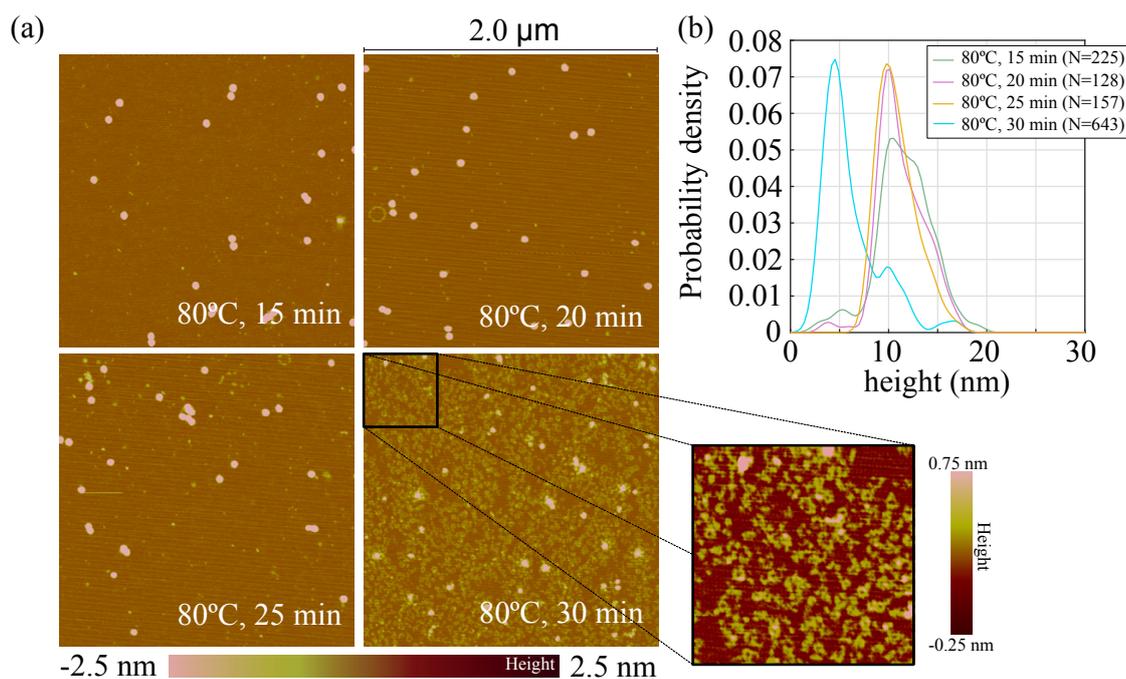


Figure 8.3: **AFM imaging of heat-induced HBV capsids destabilisation 2** - (a) AFM images of HBV capsids for different heating conditions: 80°C during 15, 20, 25 and 30 min of incubation. (b) Probability density of the capsid height for the four first conditions.

the large capacity to deform the capsid.

To further destabilise the capsid, we have varied the incubation time at 80°C: the sample was incubated for 15, 20, 25 and 30 min. AFM images show that the capsid eventually underwent ruptures when the sample was exposed for 30 min [Fig. 8.3(a)]. Furthermore, it can be seen that the height of the capsid decreases slightly as the incubation time increases [Fig. 8.3(b)]. The capsid height of the sample heated for 30 min decreased by almost half compared to the other conditions, indicating that it has been destabilised. The AFM image of this condition shows small structures that might be capsomers bound to mica.

The great stability of the HBV capsid could be associated with the capsid stiffness being softer. In fact, It is in the softest capsids category among several viral capsids that have been characterised using indentation experiments [149]. Also, under thermal fluctuations, HBV capsid can experience large scale deformations without capsid failure [149, 176, 329].

8.3 Genome uncoating of AAV capsids

We will interpret the study about thermal stability of AAV capsids obtained in the group before. Julien Bernaud, a PhD student of Cendrine Moskalenko, studied the AAV8 and AAV9 capsids stability when the environment varies, specifically when the capsid is subjected to different temperatures. Consequently, the single-stranded DNA starts to be ejected from the capsid [280]. As we have determined the stiffness of the AAV2 capsid, we have also studied its thermal stability, allowing us to interpret the thermal stability and the mechanical properties of the three serotypes.

Julien showed that AAV8 capsids begin to release DNA after a temperature of 60°C, and at that temperature, there are no signs of breakdown as seen in figure 8.4(a). As the temperature increases, the DNA is completely expelled or is still attached to the capsid. At 75°C, capsids were completely destabilised, and all the DNA has been

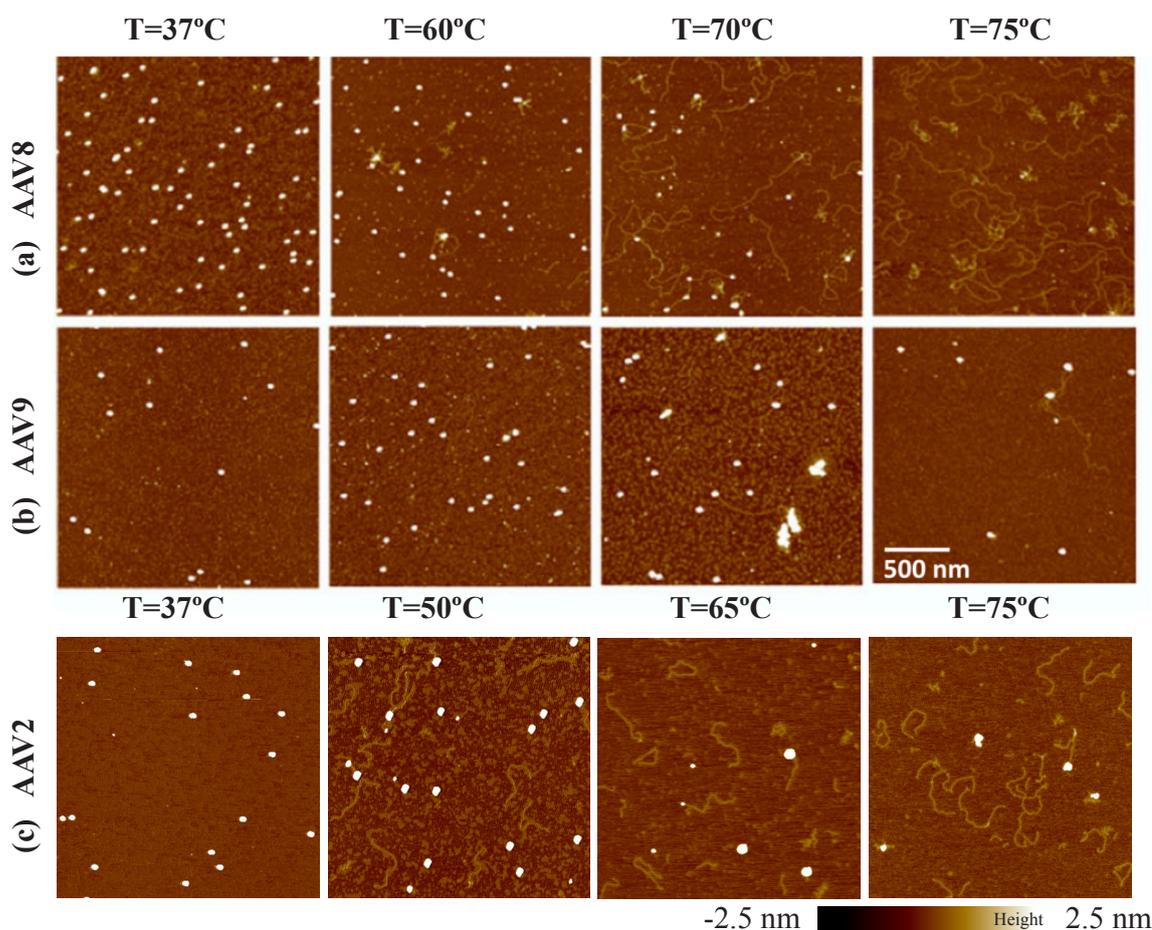


Figure 8.4: Comparison of thermal stability between AAV2, AAV8 and AAV9 - AFM images of (a) AAV8, (b) AAV9 and (c) AAV2 capsids when subjected to different temperatures. AFM images of (a,b) were obtained and adapted from [280].

released [Fig. 8.4(b)]. The protocol was also applied to the AAV9 [Fig. 8.4(b)], showing a greater resistance to destabilisation. At the maximum temperature tested, 75°C, intact capsids are still present, and the ratio between intact capsids and released DNA is dominant for capsids. AFM images for AAV2 capsids [Fig. 8.4(c)] shows that at room temperature, 37°C, the capsids are stable. We noticed that the AAV2 capsid destabilised earlier than AAV8 and AAV9.

To study the capsid evolution as a function of temperature, we analyse the correlation between the maximum height of the capsid and the equivalent diameter ($D_{eq} = 2\sqrt{A/\pi}$, where A is the area) [Fig. 8.5]. Focusing on the height projection of the 2D maps, we can see that as the temperature increases, the capsid height decreases, except

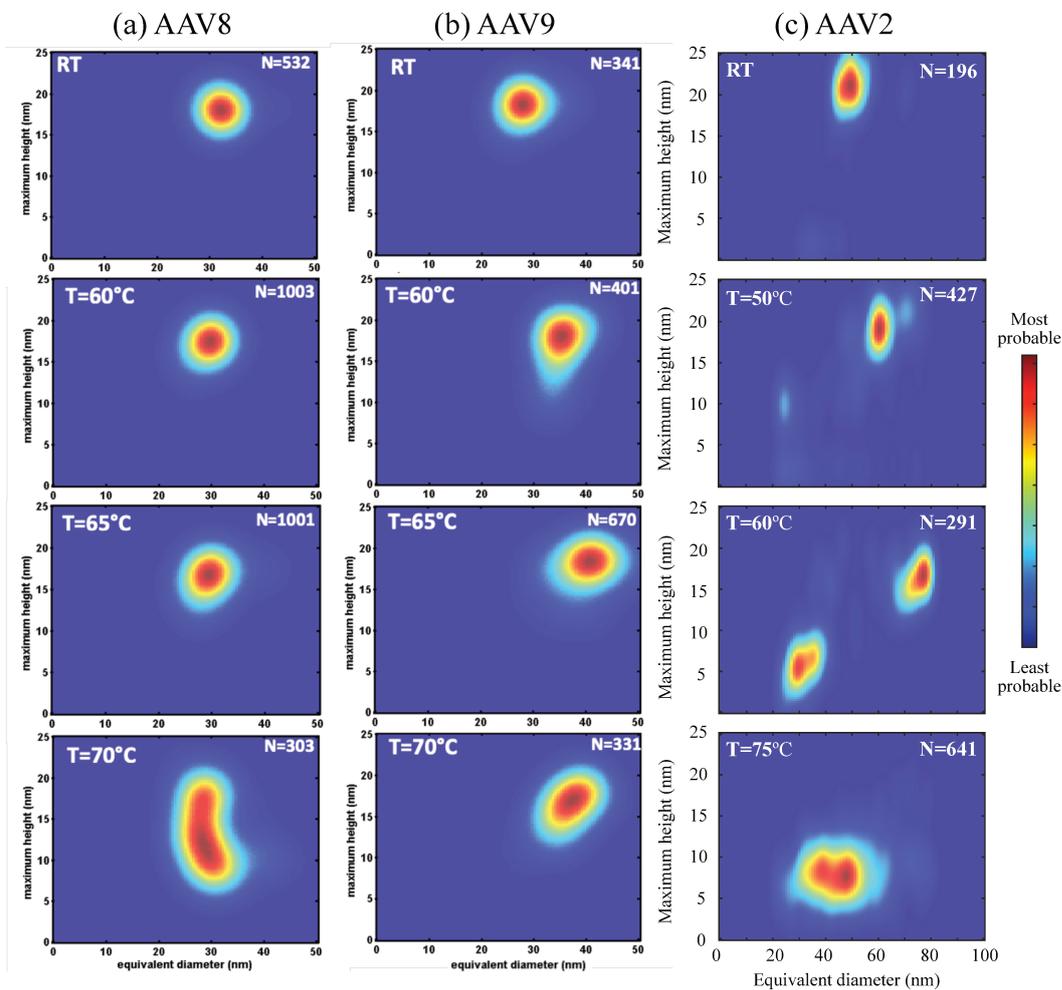


Figure 8.5: Correlation between capsid maximum height and the equivalent diameter as a function of the temperature - 2D probability density of height and equivalent diameter for (a) AAV8, (b) AAV9 and (c) AAV2 capsids for different temperatures. Data in (a,b) were obtained from [193].

for AAV9. It remains almost constant, suggesting that this capsid does not have an intermediate state of DNA release. On the contrary, AAV8 and AAV2 have an intermediate state since the capsid height decreases before the capsid completely disassembles at temperatures higher than 80°C. For AAV2 at 60°C, the same probability of finding an intact capsid and capsid with DNA partially released and at 75°C only remains the capsid with the DNA releasing.

From the mechanical properties analysis, we found that the capsid stiffness of AAVs is: $k_{AAV9} < k_{AAV8} < k_{AAV2}$. Comparing the results from the capsid stability and mechanical properties, we see that the more stable capsid AAV9, is at the same time, the softer capsid. In general, the viral capsids must travel among various environments; thus, the capsids have evolved to protect the viral genome against several physical constraints, which can be thermal, chemical or mechanical. As we mentioned, AAV9 is the serotype that has been associated with all tissue tropisms [191]. Perhaps their tropism could relate to their capacity to resist different environments. Thus, to better understand virus biology, it seems essential to increase our knowledge of viruses physical properties and optimise their use as viral vectors.

8.4 Conclusions

We found that softer capsids can withstand temperature changes without uncoating the capsid. The HBV capsid has an average stiffness of 0.3 N/m withstands temperatures of 80°C (30 min of incubation). In contrast, the capsid of AAV2, being the stiffest among AAV2, AAV8, and AAV9, broke at the lowest temperature 50°C. These results suggest that AAV2 is less stable than AAV8, followed by AAV9. Through Differential Scanning Fluorimetry (DSF), Bennet *et al.*, [282] measured the protein unfolding by monitoring changes in fluorescence as a function of temperature, and they found that AAV2 capsid is less stable than AAV8 and AAV9, the same tendency as our experiments. In conclusion, we say that softer capsids are more resistant to thermal destabilisation.

Chapter 9

ITR-DNA visualisation on AFM images

This chapter has been removed from the online version.

Conclusions and Perspectives

During this thesis, I had the opportunity to work on three biological systems: (i) Mechanical properties and thermal stability of AAV and HBV capsids. (ii) ITRs characterisation: these structures are characteristic of the AAV genome. (iii) R-loops objects characterisation; these objects are generated during the transcription process. These projects follow different subareas; however, we have established a common point in this thesis, using the physics principles to understand these processes better. The tool that allowed us to address these diverse issues was the Atomic Force Microscope AFM.

The AFM allows obtaining images of surfaces, which is an excellent tool for visualising and characterising different biological objects at the morphology level. Additionally, the AFM allows measuring mechanical properties of these objects, allowing establishing a relationship between the release process of the viral genome of AAV and HBV viruses and the mechanical properties of the objects involved in this process. We had the opportunity to work on different capsids, thanks to Anna Salvetti (AAV) and Lauriane Lecoq (HBV).

We developed a nanoindentation protocol to measure the mechanical properties and obtain the stiffness of the viral capsid (in the centre of the capsid) with high precision. The nanoindentation profile also shows the capsid stiffness depends on the position; it is almost twice in the centre than at the border of the capsid. We determined the stiffness of the AAV8 capsid by changing both the spring constant and the size of the cantilevered tip to verify our results. We found that the stiffness of the capsid is independent of the constant elastic cantilever but not of the tip size. The stiffness of the capsid using a larger tip increased stiffness, depending on the ratio between the tip size and capsid size. Small tips will deform the capsid locally and be sensitive to the icosahedral geometry and facets of the capsid. Bigger tips will deform the capsid as a complete object, so the more capsomeres that deform, the greater the stiffness. The stiffness must converge to a constant value for sufficiently large tips.

Although we could not identify the capsid orientation through the AFM images,

our protocol allowed us to differentiate orientation according to the stiffness profile behaviour. To verify these results, Lucas Menou and Martin Castelnovo developed AFM nanoindentation numerical simulations to compare both profiles. We analysed different experimental and numerical trajectories and found that the stiffness in a pentamer is more significant than in a hexamer: in terms of capsid orientation, $5' > 2' > 3'$ symmetry axis. Also, the stiffness profile falls with the indentation distance to the centre. We associated a profile to each orientation for AAV8 and AAV2. For the AAV9 capsid, we did not have enough data to differentiate a profile with each orientation. Hence, more experiments are required to classify the indentation profiles according to the orientation of this capsid. We found that AAV9 is much softer than AAV8 at the average profile level, followed by AAV2. Despite the level of the central points of the profile, AAV2 is more rigid than AAV8.

Analysing the stiffness profiles, we noted that the more faceted capsids (AAV, 60 capsid proteins) are more inhomogeneous when facing orientation changes. On the contrary, the capsid HBV, which has 240 capsid proteins, have stiffness value fluctuations smaller than AAV. The simulation results show that for the less faceted capsids, there is no difference between the orientations. However, in this case, the tip was larger than the capsid size compared to our experiments. As a perspective, indentations could be made on the HBV capsid, varying the tip size. We expect that the profiles collapse for big enough tips. For HBV, the stiffness at the centre of the capsid is approximately 0.3 N/m. Greater than the value found in [175] which is around 0.1 N/m. The difference between the values could be explained if we omit the position dependence over the capsid.

An interesting effect was visualised in the force curves; a non-linear behaviour was exhibited at small forces. We showed that this effect is not constant at the profile level. In some profiles, the non-linear regime was smaller in the centre or near the capsid centre, and sometimes it was random. We suggest that this effect could be related to the protrusions on the capsid surface. When we analysed the force curves of HBV, we noticed that the non-linear force regime was less pronounced. This is consistent because the protrusions of the HBV capsid are distributed in a more homogeneous way over the capsid surface. It would be interesting to study the non-linear force regime in detail since the capsid protrusions play an essential role during the viral cycle binding process.

Additionally, we studied the thermal stability of AAV and HBV capsids. In general, we found that the softer capsids are more resistant to thermal fluctuations. By comparing the capsids of AAV, we determine that AAV9 is the most resistant. This is interesting because this serotype has been associated with all tissue tropisms, and per-

haps its ability to express in various cellular tissues is related to its ability to resist long deformations of the capsid. In our laboratory biophysics team, Fabien Montel works on the transport of biomolecules in a confined environment. Specifically, he observed the passage of individual capsids through artificial membranes with calibrated pores of different diameters. Preliminary experiments with AAV capsids showed that the only serotype they could visualise, that is, that was able to pass through the artificial pore, is the AAV9 capsid. Therefore, the stiffness of the capsid could play a fundamental role in the nuclear entry process.

This result makes us wonder if the HBV capsid is small enough to pass directly through complex nuclear pores since it is smaller than the pores themselves and assuming that the softer stiffness supports transport. So why does this virus need transport receptors to cross the nuclear membrane? Why does it remain stuck in the pore? Besides, why does the viral genome release occur in the pore?

The mechanical properties study is not the only way to understand the replication processes of viruses. For decades, researchers have tried to understand the virus processes to provide specific cures. Advances have reached such a point that viruses are currently being used to introduce genetic material into target cells using non-viral or viral vehicles. The efficiency of genes transfer depends on several factors. In particular, Laurianne Laugel and Magalie Penaud-Budloo are interested in increasing the gene transfer efficiency of recombinant AAV vectors and understanding the inverted terminal repeats role (ITR) in this process.

In a first approach to understanding the structure of the ITR, and together with Marianne Laugel and Magalie Penaud-Budloo, we managed to develop a protocol that allows its visualisation through AFM images. We show that the ITRs of rAAV DNA can generate different conformations. Besides, we noted that the ITRs could adhere to the capsids, where sometimes one capsid was attached to one DNA end, and at other times two capsids attached to each DNA end. Hence, we suggest that this interaction could depend on the type of conformation of the ITR. From a general perspective, more experimental data are needed to establish a complete interpretation. Another interaction we identified was between a capsid and another part of the DNA that is not necessarily at the DNA end. When there was no capsid attached in that position, secondary structures were identified: structures with a height more significant than the average DNA height. We suggest that the protomers generate these secondary structures in the AAV genome since they coincide with the secondary structure positions. If this is the case, it would be interesting to study this effect since the protomer is a regulatory element of the viral vector genome, determining the transgene level of expression and in which cells it will be expressed.

During the transcription process, some secondary structures can also be generated, specifically structures called R-loops. In collaboration with Vincent Vannosthuysse, we have studied the R-loops objects resulting from the *in vitro* transcription of the AIRN gene. These R-loops objects disappear when RNase H is applied to digest DNA/RNA hybrids. To study and quantitatively R-loops objects, I developed an algorithm in Matlab that allowed us to determine the morphological and mechanical properties of DNA containing R-loops objects. We identify three types of conformations that we call blob, spur and loops objects. These R-loop objects differ mainly by the amount of DNA contained, where the spur object contains more DNA than the blobs and loops objects. The R-loop object initial generation position depends on the object type, although the final position variations depend on the object volume type.

When the *airn* gene was mutated, the R-loop objects conformation were affected, although these mutations did not affect the amount of DNA-RNA hybrids. Also, we found that a new R-loop objects formation position but with a lower frequency of appearance. To know if this type of R-loop object is identified in other genes, a similar analysis was carried out on the *sum3* gene. AFM images did not show R-loop objects, suggesting that the extruded ssDNA does not form detectable R-loop objects. In conclusion, we propose that the non-template stranded sequence determines the three-dimensional architecture of R-loops *in vitro*. Analysis of the DNA mechanical properties for the *airn* gene showed that the R-loop objects globally change the DNA molecule persistence. When the R-loops objects were not detected in the AFM images (*sum3* gene), the DNA persistence did not undergo structural changes. Perhaps these structural differences are related to the pathological or physiological behaviours of the R-loops. Therefore, from a biological perspective, it would be interesting to see whether the production of R-loops at the *airn* or *sum3* genes *in vivo* also lead to different consequences in terms of cell fate.

Bibliography

- [1] Gerd Binnig, Calvin F Quate, and Ch Gerber. Atomic force microscope. Physical review letters, 56(9):930, 1986.
- [2] Gerd Binnig, Heinrich Rohrer, Ch Gerber, and Edmund Weibel. Surface studies by scanning tunneling microscopy. Physical review letters, 49(1):57, 1982.
- [3] E Clayton Teague, Fredric E Scire, Saul M Baker, and Stephen W Jensen. Three-dimensional stylus profilometry. Wear, 83(1):1–12, 1982.
- [4] Pedro J de Pablo and Mauricio G Mateu. Mechanical properties of viruses. structure and physics of viruses, pages 519–551, 2013.
- [5] Brunero Cappella and Giovanni Dietler. Force-distance curves by atomic force microscopy. Surface science reports, 34(1-3):1–104, 1999.
- [6] Yves F Dufrêne, David Martínez-Martín, Izhar Medalsy, David Alsteens, and Daniel J Müller. Multiparametric imaging of biological systems by force-distance curve-based afm. Nature methods, 10(9):847, 2013.
- [7] P.C. Braga and D. Ricci. Atomic Force Microscopy: Biomedical Methods and Applications. Methods in molecular biology Atomic force microscopy. Humana Press, 2004.
- [8] Gerald Göring, Philipp-Immanuel Dietrich, Matthias Blaicher, Swati Sharma, Jan G Korvink, Thomas Schimmel, Christian Koos, and Hendrik Hölscher. Tailored probes for atomic force microscopy fabricated by two-photon polymerization. Applied Physics Letters, 109(6):063101, 2016.
- [9] Pedro J de Pablo. Atomic force microscopy of viruses. Structure and physics of viruses, pages 247–271, 2013.
- [10] John E Sader, James WM Chon, and Paul Mulvaney. Calibration of rectangular atomic force microscope cantilevers. Review of scientific instruments, 70(10):3967–3969, 1999.

- [11] Peter Eaton and Paul West. Atomic force microscopy. Oxford university press, 2010.
- [12] Josep Canet-Ferrer, Eugenio Coronado, Alicia Forment-Aliaga, and Elena Pinilla-Cienfuegos. Correction of the tip convolution effects in the imaging of nanostructures studied through scanning force microscopy. Nanotechnology, 25(39):395703, 2014.
- [13] Wouter H Roos. How to perform a nanoindentation experiment on a virus. In Single Molecule Analysis, pages 251–264. Springer, 2011.
- [14] Fred Griffith. The significance of pneumococcal types. Epidemiology & Infection, 27(2):113–159, 1928.
- [15] Oswald T Avery, Colin M MacLeod, and Maclyn McCarty. 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. The Journal of experimental medicine, 79(2):137–158, 1944.
- [16] Alfred D Hershey and Martha Chase. Independent functions of viral protein and nucleic acid in growth of bacteriophage. Journal of general physiology, 36(1):39–56, 1952.
- [17] F Miescher. Letter i; to wilhelm his; tübingen, february 26th, 1869. Die Histochemischen und Physiologischen Arbeiten von Friedrich Miescher—Aus dem sissenschaft—lichen Briefwechsel von F. Miescher, 1:33–8, 1869.
- [18] Friedrich Miescher-Rüsch. Ueber die chemische Zusammensetzung der Eiterzellen. 1871.
- [19] James D Watson and Francis HC Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature, 171(4356):737–738, 1953.
- [20] Bruce K Duncan and Jeffrey H Miller. Mutagenic deamination of cytosine residues in dna. Nature, 287(5782):560–561, 1980.
- [21] Ronnie Pedersen, Alexandria N Marchi, Jacob Majikes, Jessica A Nash, Nicole A Estrich, David S Courson, Carol K Hall, Stephen L Craig, and Thomas H LaBean. Properties of dna. In Handbook of nanomaterials properties, pages 1125–1157. Springer, 2014.
- [22] Rosalind E Franklin and Raymond George Gosling. The structure of sodium thymonucleate fibres. i. the influence of water content. Acta Crystallographica, 6(8-9):673–677, 1953.

- [23] Maurice Hugh Frederick Wilkins, Alex R STOKES, and Herbert R Wilson. Molecular structure of nucleic acids: molecular structure of deoxyribose nucleic acids. Nature, 171(4356):738–740, 1953.
- [24] Silvia Hormeno, Borja Ibarra, José L Carrascosa, José M Valpuesta, Fernando Moreno-Herrero, and J Ricardo Arias-Gonzalez. Mechanical properties of high-gc content dna with a-type base-stacking. Biophysical journal, 100(8):1996–2005, 2011.
- [25] Andrew H-J Wang, Gary J Quigley, Francis J Kolpak, James L Crawford, Jacques H Van Boom, Gijs van der Marel, and Alexander Rich. Molecular structure of a left-handed double helical dna fragment at atomic resolution. Nature, 282(5740):680–686, 1979.
- [26] Andrew H-J Wang, Satoshi Fujii, Jacques H van Boom, Gijs A van der Marel, Stan AA van Boeckel, and Alexander Rich. Molecular structure of r (gcg) d (tatacgc): a dna–rna hybrid helix joined to double helical dna. Nature, 299(5884):601–604, 1982.
- [27] Martin Egli, Nassim Usman, SG Zhang, and Alexander Rich. Crystal structure of an okazaki fragment at 2-Å resolution. Proceedings of the National Academy of Sciences, 89(2):534–538, 1992.
- [28] Steven B Zimmerman and Barbara H Pfeiffer. A rna–dna hybrid that can adopt two conformations: an x-ray diffraction study of poly (ra)·poly (dt) in concentrated solution or in fibers. Proceedings of the National Academy of Sciences, 78(1):78–82, 1981.
- [29] Edith Hantz, Valery Larue, Patrick Ladam, Laurence Le Moyec, Catherine Gouyette, and Tam Huynh Dinh. Solution conformation of an rna–dna hybrid duplex containing a pyrimidine rna strand and a purine dna strand. International journal of biological macromolecules, 28(4):273–284, 2001.
- [30] Agnes Noy, Alberto Pérez, Manuel Márquez, F Javier Luque, and Modesto Orozco. Structure, recognition properties, and flexibility of the dna-rna hybrid. Journal of the American Chemical Society, 127(13):4910–4920, 2005.
- [31] Jeffrey I Gyi, Andrew N Lane, Graeme L Conn, and Tom Brown. Solution structures of dna-rna hybrids with purine-rich and pyrimidine-rich strands: Comparison with the homologous dna and rna duplexes. Biochemistry, 37(1):73–80, 1998.
- [32] Andrew N LANE, Susanne EBEL, and Tom BROWN. Nmr assignments and solution conformation of the dna· rna hybrid duplex d (gtgaactt)· r (aaguucac). European journal of biochemistry, 215(2):297–306, 1993.

- [33] Thomas E Cheatham and Peter A Kollman. Molecular dynamics simulations highlight the structural differences among dna: Dna, rna: Rna, and dna: Rna hybrid duplexes. Journal of the American Chemical Society, 119(21):4805–4825, 1997.
- [34] Yong Xiong and Muttaiya Sundaralingam. Crystal structure of a dna· rna hybrid duplex with a polypurine rna r (gaagaagag) and a complementary polypyrimidine dna d (ctcttcttc). Nucleic acids research, 28(10):2171–2176, 2000.
- [35] Chen Zhang, Hang Fu, Yajun Yang, Erchi Zhou, Zhijie Tan, Huijuan You, and Xinghua Zhang. The mechanical properties of rna-dna hybrid duplex stretched by magnetic tweezers. Biophysical journal, 116(2):196–204, 2019.
- [36] S Arnott, MHF Wilkins, W Fuller, and R Langridge. Molecular and crystal structures of double-helical rna: Iii. an 11-fold molecular model and comparison of the agreement between the observed and calculated three-dimensional diffraction data for 10-and 11-fold models. Journal of molecular biology, 27(3):535–548, 1967.
- [37] S Arnott, F Hutchinson, M Spencer, MHF Wilkins, W Fuller, and R Langridge. X-ray diffraction studies of double helical ribonucleic acid. Nature, 211(5046):227–232, 1966.
- [38] Richard E Dickerson, Horace R Drew, Benjamin N Conner, Richard M Wing, Albert V Fratini, and Mary L Kopka. The anatomy of a-, b-, and z-dna. Science, 216(4545):475–485, 1982.
- [39] Struther Arnott. Historical article: Dna polymorphism and the early history of the double helix. Trends in biochemical sciences, 31(6):349–354, 2006.
- [40] Junhua Zhao, Albino Bacolla, Guliang Wang, and Karen M Vasquez. Non-b dna structure-induced genetic instability and evolution. Cellular and molecular life sciences, 67(1):43–62, 2010.
- [41] Albino Bacolla and Robert D Wells. Non-b dna conformations, genomic rearrangements, and human disease. Journal of Biological Chemistry, 279(46):47411–47414, 2004.
- [42] Albino Bacolla, Adam Jaworski, Jacquelynn E Larson, John P Jakupciak, Nadia Chuzhanova, Shaun S Abeysinghe, Catherine D O’Connell, David N Cooper, and Robert D Wells. Breakpoints of gross deletions coincide with non-b dna conformations. Proceedings of the National Academy of Sciences, 101(39):14162–14167, 2004.

- [43] Guliang Wang and Karen M Vasquez. Naturally occurring h-dna-forming sequences are mutagenic in mammalian cells. Proceedings of the National Academy of Sciences, 101(37):13448–13453, 2004.
- [44] Guliang Wang and Karen M Vasquez. Non-b dna structure-induced genetic instability. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 598(1-2):103–119, 2006.
- [45] Boris P Belotserkovskii, Sergei M Mirkin, and Philip C Hanawalt. Dna sequences that interfere with transcription: implications for genome function and stability. Chemical reviews, 113(11):8620–8637, 2013.
- [46] Shozeb M Haider, Gary N Parkinson, and Stephen Neidle. Structure of a g-quadruplex–ligand complex. Journal of molecular biology, 326(1):117–125, 2003.
- [47] B. Alberts. Molecular Biology of the Cell. W.W. Norton, 2017.
- [48] Andrés Aguilera and Tatiana García-Muse. R loops: from transcription byproducts to threats to genome stability. Molecular cell, 46(2):115–124, 2012.
- [49] Martina GL Perego, Michela Taiana, Nereo Bresolin, Giacomo P Comi, and Stefania Corti. R-loops in motor neuron diseases. Molecular neurobiology, 56(4):2579–2589, 2019.
- [50] Patricia Richard and James L Manley. R loops and links to human disease. Journal of molecular biology, 429(21):3168–3180, 2017.
- [51] Matthias Groh and Natalia Gromak. Out of balance: R-loops in human disease. PLoS Genet, 10(9):e1004630, 2014.
- [52] Aaron R Haeusler, Christopher J Donnelly, and Jeffrey D Rothstein. The expanding biology of the c9orf72 nucleotide repeat expansion in neurodegenerative disease. Nature Reviews Neuroscience, 17(6):383–395, 2016.
- [53] Monika HM Schmidt and Christopher E Pearson. Disease-associated repeat instability and mismatch repair. DNA repair, 38:117–126, 2016.
- [54] Youssef A Hegazy, Chrishan M Fernando, and Elizabeth J Tran. The balancing act of r-loop biology: The good, the bad, and the ugly. Journal of Biological Chemistry, 295(4):905–913, 2020.
- [55] Deepankar Roy, Zheng Zhang, Zhengfei Lu, Chih-Lin Hsieh, and Michael R Lieber. Competition between the rna transcript and the nontemplate dna strand during r-loop formation in vitro: a nick can serve as a strong r-loop initiation site. Molecular and cellular biology, 30(1):146–159, 2010.

- [56] Deepankar Roy and Michael R Lieber. G clustering is important for the initiation of transcription-induced r-loops in vitro, whereas high g density without clustering is sufficient thereafter. Molecular and cellular biology, 29(11):3124–3133, 2009.
- [57] Deepankar Roy, Kefei Yu, and Michael R Lieber. Mechanism of r-loop formation at immunoglobulin class switch sequences. Molecular and cellular biology, 28(1):50–60, 2008.
- [58] Paul A Ginno, Paul L Lott, Holly C Christensen, Ian Korf, and Frédéric Chédin. R-loop formation is a distinctive characteristic of unmethylated human cpG island promoters. Molecular cell, 45(6):814–825, 2012.
- [59] Vladimir A Kuznetsov, Vladyslav Bondarenko, Thidathip Wongsurawat, Surya P Yenamandra, and Piroon Jenjaroenpun. Toward predictive r-loop computational biology: genome-scale prediction of r-loops reveals their association with complex promoter structures, g-quadruplexes and transcriptionally active enhancers. Nucleic acids research, 46(15):7566–7585, 2018.
- [60] Konstantina Skourti-Stathaki, Nicholas J Proudfoot, and Natalia Gromak. Human senataxin resolves rna/dna hybrids formed at transcriptional pause sites to promote xrn2-dependent termination. Molecular cell, 42(6):794–805, 2011.
- [61] Ryan Patrick Mackay, Qinqin Xu, and Paul M Weinberger. R-loop physiology and pathology: A brief review. DNA and Cell Biology, 39(11):1914–1925, 2020.
- [62] Robert Stolz, Shaheen Sulthana, Stella R Hartono, Maika Malig, Craig J Benham, and Frederic Chedin. Interplay between dna sequence and negative superhelicity drives r-loop structures. Proceedings of the National Academy of Sciences, 116(13):6260–6269, 2019.
- [63] Wesley I Sundquist and Aaron Klug. Telomeric dna dimerizes by formation of guanine tetrads between hairpin loops. nature, 342(6251):825–829, 1989.
- [64] Lamia Wahba, Lorenzo Costantino, Frederick J Tan, Anjali Zimmer, and Douglas Koshland. S1-drip-seq identifies high expression and polyA tracts as major contributors to r-loop formation. Genes & development, 30(11):1327–1338, 2016.
- [65] Hannah E Mischo, Belén Gómez-González, Pawel Grzechnik, Ana G Rondón, Wu Wei, Lars Steinmetz, Andrés Aguilera, and Nick J Proudfoot. Yeast sen1 helicase protects the genome from transcription-associated instability. Molecular cell, 41(1):21–32, 2011.

- [66] Lamia Wahba, Jeremy D Amon, Douglas Koshland, and Milena Vuica-Ross. Rnase h and multiple rna biogenesis factors cooperate to prevent rna: Dna hybrids from generating genome instability. Molecular cell, 44(6):978–988, 2011.
- [67] Marc Drolet, Pauline Phoenix, Rolf Menzel, Eric Massé, Leroy F Liu, and Robert J Crouch. Overexpression of rnase h partially complements the growth defect of an escherichia coli delta topa mutant: R-loop formation is a major problem in the absence of dna topoisomerase i. Proceedings of the National Academy of Sciences, 92(8):3526–3530, 1995.
- [68] Xialu Li and James L Manley. Inactivation of the sr protein splicing factor asf/sf2 results in genomic instability. Cell, 122(3):365–378, 2005.
- [69] Tatiana García-Muse and Andrés Aguilera. R loops: from physiological to pathological roles. Cell, 179(3):604–618, 2019.
- [70] Aziz El Hage, Sarah L French, Ann L Beyer, and David Tollervey. Loss of topoisomerase i leads to r-loop-mediated transcriptional blocks during ribosomal rna synthesis. Genes & development, 24(14):1546–1558, 2010.
- [71] Sandie Tuduri, Laure Crabbé, Chiara Conti, Hélène Tourrière, Heidi Holtgreve-Grez, Anna Jauch, Véronique Pantesco, John De Vos, Aubin Thomas, Charles Theillet, et al. Topoisomerase i suppresses genomic instability by preventing interference between replication and transcription. Nature cell biology, 11(11):1315–1324, 2009.
- [72] José M Santos-Pereira and Andrés Aguilera. R loops: new modulators of genome dynamics and function. Nature Reviews Genetics, 16(10):583–597, 2015.
- [73] Thomas G Fazzio. Regulation of chromatin structure and cell fate by r-loops. Transcription, 7(4):121–126, 2016.
- [74] Poshen B Chen, Hsiuyi V Chen, Diwash Acharya, Oliver J Rando, and Thomas G Fazzio. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nature structural & molecular biology, 22(12):999, 2015.
- [75] Christopher Grunseich, Isabel X Wang, Jason A Watts, Joshua T Burdick, Robert D Guber, Zhengwei Zhu, Alan Bruzel, Tyler Lanman, Kelian Chen, Alice B Schindler, et al. Senataxin mutation reveals how r-loops promote transcription by blocking dna methylation at gene promoters. Molecular cell, 69(3):426–437, 2018.

- [76] Konstantina Skourti-Stathaki and Nicholas J Proudfoot. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes & development*, 28(13):1384–1396, 2014.
- [77] Dorothy Yanling Zhao, Gerald Gish, Ulrich Braunschweig, Yue Li, Zuyao Ni, Frank W Schmitges, Guoqing Zhong, Ke Liu, Weiguo Li, Jason Moffat, et al. Snn and symmetric arginine dimethylation of rna polymerase ii c-terminal domain control termination. *Nature*, 529(7584):48–53, 2016.
- [78] Wei-Ting Lu, Ben R Hawley, George L Skalka, Robert A Baldock, Ewan M Smith, Aldo S Bader, Michal Malewicz, Felicity Z Watts, Ania Wilczynska, and Martin Bushell. Drosha drives the formation of dna: Rna hybrids around dna break sites to facilitate dna repair. *Nature communications*, 9(1):1–13, 2018.
- [79] Corina Ohle, Rafael Tesorero, Géza Schermann, Nikolay Dobrev, Irmgard Sinning, and Tamás Fischer. Transient rna-dna hybrids are required for efficient double-strand break repair. *Cell*, 167(4):1001–1013, 2016.
- [80] Takaaki Yasuhara, Reona Kato, Yoshihiko Hagiwara, Bunsyo Shiotani, Motohiro Yamauchi, Shinichiro Nakada, Atsushi Shibata, and Kiyoshi Miyagawa. Human rad52 promotes xpg-mediated r-loop processing to initiate transcription-associated homologous recombination repair. *Cell*, 175(2):558–570, 2018.
- [81] Yaqun Teng, Tribhuwan Yadav, Meihan Duan, Jun Tan, Yufei Xiang, Boya Gao, Jianquan Xu, Zhuobin Liang, Yang Liu, Satoshi Nakajima, et al. Ros-induced r loops trigger a transcription-coupled but brca1/2-independent homologous recombination pathway through csb. *Nature communications*, 9(1):1–12, 2018.
- [82] Lorenzo Costantino and Douglas Koshland. The yin and yang of r-loop biology. *Current opinion in cell biology*, 34:39–45, 2015.
- [83] Shir Toubiana and Sara Selig. Dna: Rna hybrids at telomeres—when it is better to be out of the (r) loop. *The FEBS journal*, 285(14):2552–2566, 2018.
- [84] Olga M Mazina, Havva Keskin, Kritika Hanamshet, Francesca Storici, and Alexander V Mazin. Rad52 inverse strand exchange drives rna-templated dna double-strand break repair. *Molecular cell*, 67(1):19–29, 2017.
- [85] Havva Keskin, Ying Shen, Fei Huang, Mikir Patel, Taehwan Yang, Katie Ashley, Alexander V Mazin, and Francesca Storici. Transcript-rna-templated dna recombination and repair. *Nature*, 515(7527):436–439, 2014.
- [86] Mary Muers. The perils of transcription. *Nature Reviews Genetics*, 12(3):156–156, 2011.

- [87] Hallie Wimberly, Chandan Shee, PC Thornton, Priya Sivaramakrishnan, Susan M Rosenberg, and PJ Hastings. R-loops and nicks initiate dna breakage and genome instability in non-growing *escherichia coli*. Nature communications, 4(1):1–11, 2013.
- [88] A Beletskii and Ashok S Bhagwat. Transcription-induced mutations: increase in c to t mutations in the nontranscribed strand during transcription in *escherichia coli*. Proceedings of the National Academy of Sciences, 93(24):13919–13924, 1996.
- [89] Sarah K Dickerson, Eleonora Market, Eva Besmer, and F Nina Papavasiliou. Aid mediates hypermutation by deaminating single stranded dna. The Journal of experimental medicine, 197(10):1291–1296, 2003.
- [90] Christopher S Nabel, Emily K Schutsky, and Rahul M Kohli. Molecular targeting of mutagenic aid and apobec deaminases. Cell Cycle, 13(2):171–172, 2014.
- [91] Wenjian Gan, Zhishuang Guan, Jie Liu, Ting Gui, Keng Shen, James L Manley, and Xialu Li. R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes & development, 25(19):2041–2056, 2011.
- [92] Tokio Kogoma. Stable dna replication: interplay between dna replication, homologous recombination, and transcription. Microbiology and Molecular Biology Reviews, 61(2):212–238, 1997.
- [93] Julie Sollier, Caroline Townsend Stork, María L García-Rubio, Renee D Paulsen, Andrés Aguilera, and Karlene A Cimprich. Transcription-coupled nucleotide excision repair factors promote r-loop-induced genome instability. Molecular cell, 56(6):777–785, 2014.
- [94] Stephan Hamperl, Michael J Bocek, Joshua C Saldivar, Tomek Swigut, and Karlene A Cimprich. Transcription-replication conflict orientation modulates r-loop levels and activates distinct dna damage responses. Cell, 170(4):774–786, 2017.
- [95] Kevin S Lang, Ashley N Hall, Christopher N Merrikh, Mark Ragheb, Hannah Tabakh, Alex J Pollock, Joshua J Woodward, Julia E Dreifus, and Houra Merrikh. Replication-transcription conflicts generate r-loops that orchestrate bacterial stress survival and pathogenesis. Cell, 170(4):787–799, 2017.
- [96] María García-Rubio, Paula Aguilera, Juan Lafuente-Barquero, José F Ruiz, Marie-Noelle Simon, Vincent Geli, Ana G Rondón, and Andrés Aguilera. Yra1-bound rna–dna hybrids cause orientation-independent transcription–replication collisions and telomere instability. Genes & development, 32(13-14):965–977, 2018.

- [97] Brian S Plosky. The good and bad of rna: Dna hybrids in double-strand break repair. Molecular cell, 64(4):643–644, 2016.
- [98] Amandine Bonnet, Ana R Grosso, Abdessamad Elkaoutari, Emeline Coleno, Adrien Presle, Sreerama C Sridhara, Guilhem Janbon, Vincent Géli, Sérgio F de Almeida, and Benoit Palancade. Introns protect eukaryotic genomes from transcription-associated genetic instability. Molecular cell, 67(4):608–621, 2017.
- [99] Raquel Boque-Sastre, Marta Soler, Cristina Oliveira-Mateos, Anna Portela, Catia Moutinho, Sergi Sayols, Alberto Villanueva, Manel Esteller, and Sonia Guil. Head-to-head antisense transcription and r-loop formation promotes transcriptional activation. Proceedings of the National Academy of Sciences, 112(18):5785–5790, 2015.
- [100] Dmitry Klinov, Benjamin Dwir, Eli Kapon, Natalia Borovok, Tatiana Molotsky, and Alexander Kotlyar. High-resolution atomic force microscopy of duplex and triplex dna molecules. Nanotechnology, 18(22):225102, 2007.
- [101] Yuri L Lyubchenko and Luda S Shlyakhtenko. Afm for analysis of structure and dynamics of dna and protein–dna complexes. Methods, 47(3):206–213, 2009.
- [102] Pablo Ares, Maria Eugenia Fuentes-Perez, Elías Herrero-Galán, José M Valpuesta, Adriana Gil, Julio Gomez-Herrero, and Fernando Moreno-Herrero. High resolution atomic force microscopy of double-stranded rna. Nanoscale, 8(23):11818–11826, 2016.
- [103] Carlos Bustamante, James Vesenka, Chun Lin Tang, William Rees, Martin Guthold, and Rebecca Keller. Circular dna molecules imaged in air by scanning force microscopy. Biochemistry, 31(1):22–26, 1992.
- [104] T Thundat, DP Allison, RJ Warmack, GM Brown, KB Jacobson, JJ Schrick, and TL Ferrell. Atomic force microscopy of dna on mica and chemically modified mica. Scanning microscopy, 6(4):2, 1992.
- [105] Helen G Hansma, DE Laney, Magdalena Bezanilla, Robert L Sinsheimer, and Paul K Hansma. Applications for atomic force microscopy of dna. Biophysical journal, 68(5):1672–1677, 1995.
- [106] David Pastré, Olivier Piétrement, Stéphane Fusil, Fabrice Landousy, Josette Jeusset, Marie-Odile David, Loïc Hamon, Eric Le Cam, and Alain Zozime. Adsorption of dna to mica mediated by divalent counterions: a theoretical and experimental study. Biophysical journal, 85(4):2507–2518, 2003.

- [107] JS Ellis, HG Abdelhady, S Allen, MC Davies, CJ Roberts, SJB Tendler, and PM Williams. Direct atomic force microscopy observations of monovalent ion induced binding of dna to mica. Journal of microscopy, 215(3):297–301, 2004.
- [108] Colin L Freeman, Lea Dieudonne, Matthew J Collins, and Karina K Sand. Survival of environmental dna in natural environments: Surface charge and topography of minerals as driver for dna storage. bioRxiv, 2020.
- [109] Elizabeth E Fry, Susan M Lea, Terry Jackson, John WI Newman, Fiona M Ellard, Wendy E Blakemore, Robin Abu-Ghazaleh, Alan Samuel, Andrew MQ King, and David I Stuart. The structure and function of a foot-and-mouth disease virus–oligosaccharide receptor complex. The EMBO journal, 18(3):543–554, 1999.
- [110] Adam Siddiqui-Jain, Cory L Grand, David J Bearss, and Laurence H Hurley. Direct evidence for a g-quadruplex in a promoter region and its targeting with a small molecule to repress c-myc transcription. Proceedings of the National Academy of Sciences, 99(18):11593–11598, 2002.
- [111] Julian L Huppert and Shankar Balasubramanian. G-quadruplexes in promoters throughout the human genome. Nucleic acids research, 35(2):406–413, 2007.
- [112] Yong Qin and Laurence H Hurley. Structures, folding patterns, and functions of intramolecular dna g-quadruplexes found in eukaryotic promoter regions. Biochimie, 90(8):1149–1171, 2008.
- [113] Kelly J Neaves, Julian L Huppert, Robert M Henderson, and J Michael Edmondson. Direct visualization of g-quadruplexes in dna using atomic force microscopy. Nucleic acids research, 37(18):6269–6275, 2009.
- [114] Caroline Bardin and Jean Louis Leroy. The formation pathway of tetramolecular g-quadruplexes. Nucleic acids research, 36(2):477–488, 2008.
- [115] Luda S Shlyakhtenko, Alexander Y Lushnikov, Ming Li, Lela Lackey, Reuben S Harris, and Yuri L Lyubchenko. Atomic force microscopy studies provide direct evidence for dimerization of the hiv restriction factor apobec3g. Journal of Biological Chemistry, 286(5):3387–3395, 2011.
- [116] Luda S Shlyakhtenko, Alexander Y Lushnikov, Atsushi Miyagi, and Yuri L Lyubchenko. Specificity of binding of single-stranded dna-binding protein to its target. Biochemistry, 51(7):1500–1509, 2012.
- [117] Yuri L Lyubchenko and Luda S Shlyakhtenko. Imaging of dna and protein- dna complexes with atomic force microscopy. Critical Reviews™ in Eukaryotic Gene Expression, 26(1), 2016.

- [118] Claudio Rivetti, Chip Walker, and Carlos Bustamante. Polymer chain statistics and conformational analysis of dna molecules with bends or sections of different flexibility. Journal of molecular biology, 280(1):41–59, 1998.
- [119] C Bustamante, JF Marko, ED Siggia, and S Smith. Entropic elasticity of lambda-phage dna. Proc. Natl. Acad. Sci USA, 88:10009, 1991.
- [120] Claudio Rivetti, Martin Guthold, and Carlos Bustamante. Scanning force microscopy of dna deposited onto mica: Equilibration versus kinetic trapping studied by statistical polymer chain analysis. Journal of molecular biology, 264(5):919–932, 1996.
- [121] Guillaume Lamour, Julius B Kirkegaard, Hongbin Li, Tuomas PJ Knowles, and Jörg Gsponer. Easyworm: an open-source software tool to determine the mechanical properties of worm-like chains. Source code for biology and medicine, 9(1):1–6, 2014.
- [122] Julien Moukhtar, Emeline Fontaine, Cendrine Faivre-Moskalenko, and Alain Arnéodo. Probing persistence in dna curvature properties with atomic force microscopy. Physical review letters, 98(17):178101, 2007.
- [123] Patrick R Heenan and Thomas T Perkins. Imaging dna equilibrated onto mica in liquid using biochemically relevant deposition conditions. ACS nano, 13(4):4220–4229, 2019.
- [124] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. Theory of elasticity (in russian). 1965.
- [125] Graham Treece. The bitonic filter: linear filtering in an edge-preserving morphological framework, 2016.
- [126] Fernando Moreno-Herrero, J Colchero, and AM Baro. Dna height in scanning force microscopy. Ultramicroscopy, 96(2):167–174, 2003.
- [127] Nobuyuki Otsu. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1):62–66, 1979.
- [128] Maika Malig and Frederic Chedin. Characterization of r-loop structures using single-molecule r-loop footprinting and sequencing. In RNA-Chromatin Interactions, pages 209–228. Springer, 2020.
- [129] Divakaran Murugesapillai, Serge Bouaziz, L James Maher, Nathan E Israeloff, Craig E Cameron, and Mark C Williams. Accurate nanoscale flexibility measurement of dna and dna–protein complexes by atomic force microscopy in liquid. Nanoscale, 9(31):11327–11337, 2017.

- [130] Yeraldinne Carrasco-Salas, Amélie Malapert, Shaheen Sulthana, Bastien Molcrette, Léa Chazot-Franguiadakis, Pascal Bernard, Frédéric Chédin, Cendrine Faivre-Moskalenko, and Vincent Vanoosthuyse. The extruded non-template strand determines the architecture of r-loops. Nucleic acids research, 47(13):6783–6795, 2019.
- [131] Stella R Hartono, Amélie Malapert, Pénélope Legros, Pascal Bernard, Frédéric Chédin, and Vincent Vanoosthuyse. The affinity of the s9. 6 antibody for double-stranded rnas impacts the accurate mapping of r-loops in fission yeast. Journal of molecular biology, 430(3):272–284, 2018.
- [132] Wang-Shick Ryu. Discovery and classification. Molecular Virology of Human Pathogenic Viruses, page 3, 2017.
- [133] Wang-Shick Ryu. Virus life cycle. Molecular Virology of Human Pathogenic Viruses, page 31, 2017.
- [134] Freya Molenberghs, Johannes J Bogers, and Winnok H De Vos. Confined no more: viral mechanisms of nuclear entry and egress. The International Journal of Biochemistry & Cell Biology, page 105875, 2020.
- [135] Michael Kann, Beate Sodeik, Angelika Vlachou, Wolfram H Gerlich, and Ari Helenius. Phosphorylation-dependent binding of hepatitis b virus core particles to the nuclear pore complex. Journal of Cell Biology, 145(1):45–55, 1999.
- [136] Sarah Cohen and Nelly Pante. Pushing the envelope: microinjection of minute virus of mice into xenopus oocytes causes damage to the nuclear envelope. Journal of general virology, 86(12):3243–3252, 2005.
- [137] Sarah Cohen, Ali R Behzad, Jeffrey B Carroll, and Nelly Pante. Parvoviral nuclear import: bypassing the host nuclear-transport machinery. Journal of general virology, 87(11):3209–3213, 2006.
- [138] Sarah C Nicolson and R Jude Samulski. Recombinant adeno-associated virus utilizes host cell nuclear import machinery to enter the nucleus. Journal of virology, 88(8):4132–4144, 2014.
- [139] Wu Xiao, Kenneth H Warrington, Patrick Hearing, Jeffrey Hughes, and Nicholas Muzyczka. Adenovirus-facilitated nuclear translocation of adeno-associated virus type 2. Journal of virology, 76(22):11505–11517, 2002.
- [140] Joseph M Kelich, Jiong Ma, Biao Dong, Qizhao Wang, Mario Chin, Connor M Magura, Weidong Xiao, and Weidong Yang. Super-resolution imaging of nuclear

- import of adeno-associated virus in live cells. Molecular Therapy-Methods & Clinical Development, 2:15047, 2015.
- [141] TS Baker, NH Olson, and SD Fuller. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiology and Molecular Biology Reviews, 63(4):862–922, 1999.
- [142] Francis HC Crick and James D Watson. Structure of small viruses. Nature, 177(4506):473–475, 1956.
- [143] Martin Castelnovo. Viral self-assembly pathway and mechanical stress relaxation. Physical Review E, 95(5):052405, 2017.
- [144] Donald LD Caspar and Aaron Klug. Physical principles in the construction of regular viruses. In Cold Spring Harbor symposia on quantitative biology, volume 27, pages 1–24. Cold Spring Harbor Laboratory Press, 1962.
- [145] Jack Lidmar, Leonid Mirny, and David R Nelson. Virus shapes and buckling transitions in spherical shells. Physical Review E, 68(5):051910, 2003.
- [146] Jianfei Hua, Alexis Huet, Carlos A Lopez, Katerina Toropova, Welkin H Pope, Robert L Duda, Roger W Hendrix, and James F Conway. Capsids and genomes of jumbo-sized bacteriophages reveal the evolutionary reach of the hk97 fold. MBio, 8(5), 2017.
- [147] Mark N Hamblin, Jie Xuan, Daniel Maynes, H Dennis Tolley, David M Belnap, Adam T Woolley, Milton L Lee, and Aaron R Hawkins. Selective trapping and concentration of nanoparticles and viruses in dual-height nanofluidic channels. Lab on a Chip, 10(2):173–178, 2010.
- [148] Mauricio Carrillo-Tripp, Craig M Shepherd, Ian A Borelli, Sangita Venkataraman, Gabriel Lander, Padmaja Natarajan, John E Johnson, Charles L Brooks III, and Vijay S Reddy. Viperdb2: an enhanced and web api enabled relational database for structural virology. Nucleic acids research, 37(suppl_1):D436–D442, 2009.
- [149] Mauricio G Mateu. Mechanical properties of viruses analyzed by atomic force microscopy: a virological perspective. Virus research, 168(1-2):1–22, 2012.
- [150] Milagros Castellanos, Rebeca Pérez, Pablo JP Carrillo, Pedro J de Pablo, and Mauricio G Mateu. Mechanical disassembly of single virus particles reveals kinetic intermediates predicted by theory. Biophysical journal, 102(11):2615–2624, 2012.
- [151] Pedro J de Pablo. Introduction to atomic force microscopy. In Single Molecule Analysis, pages 197–212. Springer, 2011.

- [152] JL Cuellar, F Meinhoewel, M Hoehne, and E Donath. Size and mechanical stability of norovirus capsids depend on pH: a nanoindentation study. Journal of general virology, 91(10):2449–2456, 2010.
- [153] Pedro J de Pablo, Mercedes Pérez, Carolina Carrasco, and José L Carrascosa. Direct visualization of single virus restoration after damage in real time. Journal of biological physics, 44(2):225–235, 2018.
- [154] IL Ivanovska, PJ De Pablo, B Ibarra, G Sgalari, FC MacKintosh, JL Carrascosa, CF Schmidt, and GJL Wuite. Bacteriophage capsids: tough nanoshells with complex elastic properties. Proceedings of the National Academy of Sciences, 101(20):7600–7605, 2004.
- [155] JP Michel, IL Ivanovska, MM Gibbons, WS Klug, CM Knobler, GJL Wuite, and CF Schmidt. Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. Proceedings of the National Academy of Sciences, 103(16):6184–6189, 2006.
- [156] C Carrasco, A Carreira, IAT Schaap, PA Serena, J Gomez-Herrero, MG Mateu, and PJ De Pablo. Dna-mediated anisotropic mechanical reinforcement of a virus. Proceedings of the National Academy of Sciences, 103(37):13706–13711, 2006.
- [157] Milagros Castellanos, Rebeca Pérez, Carolina Carrasco, Mercedes Hernando-Pérez, Julio Gómez-Herrero, Pedro J de Pablo, and Mauricio G Mateu. Mechanical elasticity as a physical signature of conformational dynamics in a virus particle. Proceedings of the National Academy of Sciences, 109(30):12028–12033, 2012.
- [158] Carolina Carrasco, Milagros Castellanos, Pedro J de Pablo, and Mauricio G Mateu. Manipulation of the mechanical properties of a virus by protein engineering. Proceedings of the National Academy of Sciences, 105(11):4150–4155, 2008.
- [159] Mavis Agbandje-McKenna, Antonio L Llamas-Saiz, Feng Wang, Peter Tattersall, and Michael G Rossmann. Functional implications of the structure of the murine parvovirus, minute virus of mice. Structure, 6(11):1369–1381, 1998.
- [160] Maija Vihinen-Ranta, Dai Wang, Wendy S Weichert, and Colin R Parrish. The vp1 n-terminal sequence of canine parvovirus affects nuclear transport of capsids and efficient cell infection. Journal of virology, 76(4):1884–1891, 2002.
- [161] Juan Reguera, Aura Carreira, Laura Riobos, José María Almendral, and Mauricio G Mateu. Role of interfacial amino acid residues in assembly, stability, and conformation of a spherical virus capsid. Proceedings of the National Academy of Sciences, 101(9):2724–2729, 2004.

- [162] Svenja Bleker, Florian Sonntag, and Jürgen A Kleinschmidt. Mutational analysis of narrow pores at the fivefold symmetry axes of adeno-associated virus type 2 capsids reveals a dual role in genome packaging and activation of phospholipase a2 activity. *Journal of virology*, 79(4):2528–2540, 2005.
- [163] Glen A Farr and Peter Tattersall. A conserved leucine that constricts the pore through the capsid fivefold cylinder plays a central role in parvoviral infection. *Virology*, 323(2):243–256, 2004.
- [164] Pavel Plevka, Susan Hafenstein, Lei Li, Anthony D’Abramo, Susan F Cotmore, Michael G Rossmann, and Peter Tattersall. Structure of a packaging-defective mutant of minute virus of mice indicates that the genome is packaged via a pore at a 5-fold axis. *Journal of virology*, 85(10):4822–4827, 2011.
- [165] Susan F Cotmore and Peter Tattersall. Encapsidation of minute virus of mice dna: aspects of the translocation mechanism revealed by the structure of partially packaged genomes. *Virology*, 336(1):100–112, 2005.
- [166] Jason A King, Ralf Dubielzig, Dirk Grimm, and Jürgen A Kleinschmidt. Dna helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *The EMBO journal*, 20(12):3282–3291, 2001.
- [167] Jennifer Timpe, Joyce Bevington, John Casper, John D Dignam, and James P Trempe. Mechanisms of adeno-associated virus genome encapsidation. *Current gene therapy*, 5(3):273–284, 2005.
- [168] Mariska GM van Rosmalen, Glen R Nemerow, Gijs JL Wuite, and Wouter H Roos. A single point mutation in precursor protein vi doubles the mechanical strength of human adenovirus. *Journal of biological physics*, 44(2):119–132, 2018.
- [169] Joost Snijder, Vijay S Reddy, Eric R May, Wouter H Roos, Glen R Nemerow, and Gijs JL Wuite. Integrin and defensin modulate the mechanical properties of adenovirus. *Journal of virology*, 87(5):2756–2766, 2013.
- [170] Ana J Pérez-Berná, Alvaro Ortega-Esteban, Rosa Menéndez-Conejero, Dennis C Winkler, Margarita Menéndez, Alasdair C Steven, S Jane Flint, Pedro J De Pablo, and Carmen San Martín. The role of capsid maturation on adenovirus priming for sequential uncoating. *Journal of Biological Chemistry*, 287(37):31582–31595, 2012.
- [171] Alvaro Ortega-Esteban, Gabriela N Condezo, Ana J Pérez-Berná, Miguel Chillón, S Jane Flint, David Reguera, Carmen San Martín, and Pedro J De Pablo. Mechanics of viral chromatin reveals the pressurization of human adenovirus. *ACS nano*, 9(11):10826–10833, 2015.

- [172] Marian Baclayon, Glen K Shoemaker, Charlotte Uetrecht, Sue E Crawford, Mary K Estes, BV Venkataram Prasad, Albert JR Heck, Gijs JL Wuite, and Wouter H Roos. Prestress strengthens the shell of norwalk virus nanoparticles. Nano letters, 11(11):4865–4869, 2011.
- [173] Wouter H Roos, Kerstin Radtke, Edward Kniesmeijer, Hylkje Geertsema, Beate Sodeik, and Gijs JL Wuite. Scaffold expulsion and genome packaging trigger stabilization of herpes simplex virus capsids. Proceedings of the National Academy of Sciences, 106(24):9673–9678, 2009.
- [174] Ivan Liashkovich, Wali Hafezi, Joachim E Kühn, Hans Oberleithner, Armin Kramer, and Victor Shahin. Exceptional mechanical and structural stability of hsv-1 unveiled with fluid atomic force microscopy. Journal of cell science, 121(14):2287–2292, 2008.
- [175] Charlotte Uetrecht, Cees Versluis, Norman R Watts, Wouter H Roos, Gijs JL Wuite, Paul T Wingfield, Alasdair C Steven, and Albert JR Heck. High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis b virus capsids. Proceedings of the National Academy of Sciences, 105(27):9216–9220, 2008.
- [176] Anton Arkhipov, Wouter H Roos, Gijs JL Wuite, and Klaus Schulten. Elucidating the mechanism behind irreversible deformation of viral capsids. Biophysical journal, 97(7):2061–2069, 2009.
- [177] Mercedes Hernando-Pérez, Roberto Miranda, María Aznar, José L Carrascosa, Iwan AT Schaap, David Reguera, and Pedro J de Pablo. Direct measurement of phage phi29 stiffness provides evidence of internal pressure. Small, 8(15):2366–2370, 2012.
- [178] J Snijder, C Uetrecht, RJ Rose, R Sanchez-Eugenía, Gerardo Anibal Marti, J Agirre, DMA Guérin, GJL Wuite, AJR Heck, and WH Roos. Probing the biophysical interplay between a viral genome and its capsid. Nature chemistry, 5(6):502–509, 2013.
- [179] Irena Ivanovska, Gijs Wuite, Bengt Jönsson, and Alex Evilevitch. Internal dna pressure modifies stability of wt phage. Proceedings of the National Academy of Sciences, 104(23):9603–9608, 2007.
- [180] Alex Evilevitch, Wouter H Roos, Irena L Ivanovska, Meerim Jeembaeva, Bengt Jönsson, and Gijs JL Wuite. Effects of salts on internal dna pressure and mechanical properties of phage capsids. Journal of molecular biology, 405(1):18–23, 2011.

- [181] Natalia Martín-González, Mercedes Hernando-Pérez, Gabriela N Condezo, Marta Pérez-Illana, Antonio Šiber, David Reguera, Philomena Ostapchuk, Patrick Hearing, Carmen San Martín, and Pedro J de Pablo. Adenovirus major core protein condenses dna in clusters and bundles, modulating genome release and capsid internal pressure. Nucleic acids research, 47(17):9231–9242, 2019.
- [182] Mariska GM van Rosmalen, Chenglei Li, Adam Zlotnick, Gijs JL Wuite, and Wouter H Roos. Effect of dsdna on the assembly pathway and mechanical strength of sv40 vp1 virus-like particles. Biophysical journal, 115(9):1656–1665, 2018.
- [183] Cheng Zeng, Sven Moller-Tank, Aravind Asokan, and Bogdan Dragnea. Probing the link among genomic cargo, contact mechanics, and nanoindentation in recombinant adeno-associated virus 2. The Journal of Physical Chemistry B, 121(8):1843–1853, 2017.
- [184] A Ortega-Esteban, Ana J Pérez-Berná, R Menéndez-Conejero, SJ Flint, C San Martín, and Pedro José de Pablo. Monitoring dynamics of human adenovirus disassembly induced by mechanical fatigue. Scientific reports, 3(1):1–7, 2013.
- [185] Aida Llauro, Benjamin Schwarz, Ranjit Koliyatt, Pedro J de Pablo, and Trevor Douglas. Tuning viral capsid nanoparticle stability with symmetrical morphogenesis. ACS nano, 10(9):8465–8473, 2016.
- [186] Mercedes Hernando-Pérez, Elena Pascual, María Aznar, Alina Ionel, José R Castón, Antoni Luque, José L Carrascosa, David Reguera, and Pedro J de Pablo. The interplay between mechanics and stability of viral cages. Nanoscale, 6(5):2702–2709, 2014.
- [187] Roya Zandi and David Reguera. Mechanical properties of viral capsids. Physical Review E, 72(2):021917, 2005.
- [188] Florence Tama and Charles L Brooks III. Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. Journal of molecular biology, 345(2):299–314, 2005.
- [189] Florence Tama and Charles L Brooks III. The mechanism and pathway of ph induced swelling in cowpea chlorotic mottle virus. Journal of molecular biology, 318(3):733–747, 2002.
- [190] BM Hespenheide, DJ Jacobs, and MF Thorpe. Structural rigidity in the capsid assembly of cowpea chlorotic mottle virus. Journal of Physics: Condensed Matter, 16(44):S5055, 2004.

- [191] Chengwen Li and R Jude Samulski. Engineering adeno-associated virus vectors for gene therapy. Nature Reviews Genetics, pages 1–18, 2020.
- [192] Sirika Pillay and Jan E Carette. Host determinants of adeno-associated viral vector entry. Current opinion in virology, 24:124–131, 2017.
- [193] Julien Bernaud. Propriétés physiques de capsides virales étudiées à l'échelle du virus unique par microscopie à force atomique : exemples du rétrovirus VIH-1 et du parvovirus AAV. PhD thesis, 2015. Thèse de doctorat dirigée par Moskalenko, Cendrine Physique Lyon, École normale supérieure 2015.
- [194] Mario Mietzsch, Judit J Péntzes, and Mavis Agbandje-McKenna. Twenty-five years of structural parvovirology. Viruses, 11(4):362, 2019.
- [195] Ran Zhang, Lin Cao, Mengtian Cui, Zixian Sun, Mingxu Hu, Rouxuan Zhang, William Stuart, Xiaochu Zhao, Zirui Yang, Xueming Li, et al. Adeno-associated virus 2 bound to its cellular receptor aavr. Nature microbiology, 4(4):675–682, 2019.
- [196] Qing Xie, John M Spear, Alex J Noble, Duncan R Sousa, Nancy L Meyer, Omar Davulcu, Fuming Zhang, Robert J Linhardt, Scott M Stagg, and Michael S Chapman. The 2.8 Å electron microscopy structure of adeno-associated virus-dj bound by a heparinoid pentasaccharide. Molecular Therapy-Methods & Clinical Development, 5:1–12, 2017.
- [197] Lin-Ya Huang, Ami Patel, Robert Ng, Edward Blake Miller, Sujata Halder, Robert McKenna, Aravind Asokan, and Mavis Agbandje-McKenna. Characterization of the adeno-associated virus 1 and 6 sialic acid binding site. Journal of virology, 90(11):5219–5230, 2016.
- [198] Sandra Afione, Michael A DiMattia, Sujata Halder, Giovanni Di Pasquale, Mavis Agbandje-McKenna, and John A Chiorini. Identification and mutagenesis of the adeno-associated virus 5 sialic acid binding region. Journal of virology, 89(3):1660–1672, 2015.
- [199] Thomas F Lerch and Michael S Chapman. Identification of the heparin binding site on adeno-associated virus serotype 3b (aav-3b). Virology, 423(1):6–13, 2012.
- [200] Jason O'Donnell, Kenneth A Taylor, and Michael S Chapman. Adeno-associated virus-2 and its primary cellular receptor—cryo-em structure of a heparin complex. Virology, 385(2):434–443, 2009.
- [201] Hazel C Levy, Valorie D Bowman, Lakshmanan Govindasamy, Robert McKenna, Kevin Nash, Kenneth Warrington, Weijun Chen, Nicholas Muzyczka, Xiaodong

- Yan, Timothy S Baker, et al. Heparin binding induces conformational changes in adeno-associated virus serotype 2. Journal of structural biology, 165(3):146–156, 2009.
- [202] Susan Hafenstein, Laura M Palermo, Victor A Kostyuchenko, Chuan Xiao, Marc C Morais, Christian DS Nelson, Valorie D Bowman, Anthony J Battisti, Paul R Chipman, Colin R Parrish, et al. Asymmetric binding of transferrin receptor to parvovirus capsids. Proceedings of the National Academy of Sciences, 104(16):6585–6589, 2007.
- [203] Alberto López-Bueno, Mari-Paz Rubio, Nathan Bryant, Robert McKenna, Mavis Agbandje-McKenna, and José M Almendral. Host-selected amino acid changes at the sialic acid binding pocket of the parvovirus capsid modulate cell binding affinity and determine virulence. Journal of virology, 80(3):1563–1573, 2006.
- [204] Zhijian Wu, Aravind Asokan, and R Jude Samulski. Adeno-associated virus serotypes: vector toolkit for human gene therapy. Molecular therapy, 14(3):316–327, 2006.
- [205] W Ding, L Zhang, Z Yan, and JF Engelhardt. Intracellular trafficking of adeno-associated viral vectors. Gene therapy, 12(11):873–880, 2005.
- [206] Maxim Salganik, Fikret Aydemir, Hyun-Joo Nam, Robert McKenna, Mavis Agbandje-McKenna, and Nicholas Muzyczka. Adeno-associated virus capsid proteins may play a role in transcription and second-strand synthesis of recombinant genomes. Journal of virology, 88(2):1071–1079, 2014.
- [207] Florian Sonntag, Svenja Bleker, Barbara Leuchs, Roger Fischer, and Jürgen A Kleinschmidt. Adeno-associated virus type 2 capsids with externalized vp1/vp2 trafficking domains are generated prior to passage through the cytoplasm and are maintained until uncoating occurs in the nucleus. Journal of virology, 80(22):11040–11054, 2006.
- [208] Krishna J Fisher, Guang-Ping Gao, Matthew D Weitzman, Ronald DeMatteo, John F Burda, and James M Wilson. Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. Journal of virology, 70(1):520–532, 1996.
- [209] Forrest K Ferrari, Thaddeus Samulski, Thomas Shenk, and Richard Jude Samulski. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. Journal of virology, 70(5):3227–3234, 1996.

- [210] Dan Wang, Phillip WL Tai, and Guangping Gao. Adeno-associated virus vector as a platform for gene therapy delivery. Nature reviews Drug discovery, 18(5):358–378, 2019.
- [211] Richard A McPherson, Leonard J Rosenthal, and James A Rose. Human cytomegalovirus completely helps adeno-associated virus replication. Virology, 147(1):217–222, 1985.
- [212] Robert ML Buller, John E Janik, Edwin D Sebring, and James A Rose. Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication. Journal of virology, 40(1):241–247, 1981.
- [213] Friedrich W Weindler and REGINE Heilbronn. A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. Journal of virology, 65(5):2476–2483, 1991.
- [214] LONG-SHENG Chang, YANG Shi, and THOMAS Shenk. Adeno-associated virus p5 promoter contains an adenovirus e1a-inducible element and a binding site for the major late transcription factor. Journal of Virology, 63(8):3479–3488, 1989.
- [215] Matthew D Weitzman and R Michael Linden. Adeno-associated virus biology. In Adeno-Associated Virus, pages 1–23. Springer, 2012.
- [216] Clare E Thomas, Theresa A Storm, Zan Huang, and Mark A Kay. Rapid uncoating of vector genomes is the key to efficient liver transduction with pseudotyped adeno-associated virus vectors. Journal of virology, 78(6):3110–3122, 2004.
- [217] I Sipo, H Fechner, S Pinkert, L Suckau, X Wang, S Weger, and W Poller. Differential internalization and nuclear uncoating of self-complementary adeno-associated virus pseudotype vectors as determinants of cardiac cell transduction. Gene therapy, 14(18):1319–1329, 2007.
- [218] Kenneth I Berns. Parvovirus replication. Microbiology and Molecular Biology Reviews, 54(3):316–329, 1990.
- [219] Kevin D Foust, Emily Nurre, Chrystal L Montgomery, Anna Hernandez, Curtis M Chan, and Brian K Kaspar. Intravascular aav9 preferentially targets neonatal neurons and adult astrocytes. Nature biotechnology, 27(1):59–65, 2009.
- [220] Lauren M Drouin and Mavis Agbandje-McKenna. Adeno-associated virus structural biology as a tool in vector development. Future virology, 8(12):1183–1199, 2013.

- [221] Eric Zinn and Luk H Vandenberghe. Adeno-associated virus: fit to serve. Current opinion in virology, 8:90–97, 2014.
- [222] Melissa A Kotterman and David V Schaffer. Engineering adeno-associated viruses for clinical gene therapy. Nature Reviews Genetics, 15(7):445–451, 2014.
- [223] Guangping Gao, Luk H Vandenberghe, Mauricio R Alvira, You Lu, Roberto Calcedo, Xiangyang Zhou, and James M Wilson. Clades of adeno-associated viruses are widely disseminated in human tissues. Journal of virology, 78(12):6381–6388, 2004.
- [224] Sebastian Bass-Stringer, Bianca C Bernardo, Clive N May, Colleen J Thomas, Kate L Weeks, and Julie R McMullen. Adeno-associated virus gene therapy: translational progress and future prospects in the treatment of heart failure. Heart, Lung and Circulation, 27(11):1285–1300, 2018.
- [225] Melissa A Kotterman, Thomas W Chalberg, and David V Schaffer. Viral vectors for gene therapy: translational and clinical outlook. Annual review of biomedical engineering, 17:63–89, 2015.
- [226] Mark A Kay. State-of-the-art gene-based therapies: the road ahead. Nature Reviews Genetics, 12(5):316–328, 2011.
- [227] Hildegard Büning, Anke Huber, Liang Zhang, Nadja Meumann, and Ulrich Hacker. Engineering the aav capsid to optimize vector–host-interactions. Current opinion in pharmacology, 24:94–104, 2015.
- [228] Benjamin E Deverman, Bernard M Ravina, Krystof S Bankiewicz, Steven M Paul, and Dinah WY Sah. Gene therapy for neurological disorders: progress and prospects. Nature reviews Drug discovery, 17(9):641–659, 2018.
- [229] Jian-Yun Dong, Pei-Dong Fan, and Raymond A Frizzell. Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. Human gene therapy, 7(17):2101–2112, 1996.
- [230] JUAN Li, RICHARD JUDE Samulski, and XIAO Xiao. Role for highly regulated rep gene expression in adeno-associated virus vector production. Journal of virology, 71(7):5236–5243, 1997.
- [231] Nor Chejanovsky and Barrie J Carter. Replication of a human parvovirus non-sense mutant in mammalian cells containing an inducible amber suppressor. Virology, 171(1):239–247, 1989.

- [232] DS Im and NICHOLAS Muzyczka. Partial purification of adeno-associated virus rep78, rep52, and rep40 and their biochemical characterization. Journal of Virology, 66(2):1119–1128, 1992.
- [233] Ralf Dubielzig, Jason A King, Stefan Weger, Andrea Kern, and Jürgen A Kleinschmidt. Adeno-associated virus type 2 protein interactions: formation of pre-encapsidation complexes. Journal of virology, 73(11):8989–8998, 1999.
- [234] Douglas E Smith, Sander J Tans, Steven B Smith, Shelley Grimes, Dwight L Anderson, and Carlos Bustamante. The bacteriophage φ 29 portal motor can package dna against a large internal force. Nature, 413(6857):748–752, 2001.
- [235] Susan F Cotmore, Mavis Agbandje-McKenna, Marta Canuti, John A Chiorini, Anna-Maria Eis-Hubinger, Joseph Hughes, Mario Mietzsch, Sejal Modha, Mylène Ogliastro, Judit J Péntzes, et al. Ictv virus taxonomy profile: Parvoviridae. The Journal of general virology, 100(3):367, 2019.
- [236] Florian Sonntag, Kristin Schmidt, and Jürgen A Kleinschmidt. A viral assembly factor promotes aav2 capsid formation in the nucleolus. Proceedings of the National Academy of Sciences, 107(22):10220–10225, 2010.
- [237] F Sonntag, K Köther, K Schmidt, M Weghofer, C Raupp, K Nieto, A Kuck, B Gerlach, B Böttcher, OJ Müller, et al. The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. Journal of virology, 85(23):12686–12697, 2011.
- [238] JA Rose, JV Maizel, JK Inman, and AJ Shatkin. Structural proteins of adenovirus-associated viruses. Journal of virology, 8(5):766–770, 1971.
- [239] RM Buller and JA Rose. Characterization of adenovirus-associated virus-induced polypeptides in kb cells. Journal of virology, 25(1):331–338, 1978.
- [240] Lauriel F Earley, John M Powers, Kei Adachi, Joshua T Baumgart, Nancy L Meyer, Qing Xie, Michael S Chapman, and Hiroyuki Nakai. Adeno-associated virus (aav) assembly-activating protein is not an essential requirement for capsid assembly of aav serotypes 4, 5, and 11. Journal of virology, 91(3), 2017.
- [241] Stefanie Grosse, Magalie Penaud-Budloo, Anne-Kathrin Herrmann, Kathleen Börner, Julia Fakhiri, Vibor Laketa, Chiara Krämer, Ellen Wiedtke, Manuel Gunkel, Lucie Ménard, et al. Relevance of assembly-activating protein for adeno-associated virus vector production and capsid protein stability in mammalian and insect cells. Journal of virology, 91(20), 2017.

- [242] Anna C Maurer, Simon Pacouret, Ana Karla Cepeda Diaz, Jessica Blake, Eva Andres-Mateos, and Luk H Vandenberghe. The assembly-activating protein promotes stability and interactions between aav's viral proteins to nucleate capsid assembly. Cell reports, 23(6):1817–1830, 2018.
- [243] Anne-Kathrin Herrmann, Stefanie Große, Kathleen Börner, Chiara Krämer, Ellen Wiedtke, Manuel Gunkel, and Dirk Grimm. Impact of the assembly-activating protein on molecular evolution of synthetic adeno-associated virus capsids. Human Gene Therapy, 30(1):21–35, 2019.
- [244] E Lusby, KH Fife, and KI Berns. Nucleotide sequence of the inverted terminal repetition in adeno-associated virus dna. Journal of virology, 34(2):402–409, 1980.
- [245] Dong-Soo Im and Nicholas Muzyczka. The aav origin binding protein rep68 is an atp-dependent site-specific endonuclease with dna helicase activity. Cell, 61(3):447–457, 1990.
- [246] DOUGLAS M McCARTY, Daniel J Pereira, Irene Zolotukhin, Xiaohuai Zhou, John H Ryan, and Nicholas Muzyczka. Identification of linear dna sequences that specifically bind the adeno-associated virus rep protein. Journal of Virology, 68(8):4988–4997, 1994.
- [247] RO Snyder, DS Im, TIEHUA Ni, XIAO Xiao, RJ Samulski, and N Muzyczka. Features of the adeno-associated virus origin involved in substrate recognition by the viral rep protein. Journal of virology, 67(10):6096–6104, 1993.
- [248] Richard O Snyder, Richard Jude Samulski, and Nicholas Muzyczka. In vitro resolution of covalently joined aav chromosome ends. Cell, 60(1):105–113, 1990.
- [249] Manuel AFV Gonçalves. Adeno-associated virus: from defective virus to effective vector. Virology journal, 2(1):1–17, 2005.
- [250] Rebecca P Haberman, Thomas J McCown, and Richard Jude Samulski. Novel transcriptional regulatory signals in the adeno-associated virus terminal repeat a/d junction element. Journal of Virology, 74(18):8732–8739, 2000.
- [251] R J Samulski, Long-Sheng Chang, and T Shen. A recombinant plasmid from which an infectious adeno-associated virus genome can be excised in vitro and its use to study viral replication. Journal of virology, 61(10):3096–3101, 1987.
- [252] Susan K McLaughlin, P Collis, PL Hermonat, and N Muzyczka. Adeno-associated virus general transduction vectors: analysis of proviral structures. Journal of virology, 62(6):1963–1973, 1988.

- [253] MARK A Labow and KENNETH I Berns. The adeno-associated virus rep gene inhibits replication of an adeno-associated virus/simian virus 40 hybrid genome in cos-7 cells. Journal of virology, 62(5):1705–1712, 1988.
- [254] Lucia Haronikova, Jan Coufal, Iva Kejnovska, Eva B Jagelska, Miroslav Fojta, Petra Dvořáková, Petr Muller, Borivoj Vojtesek, and Vaclav Brazda. Ifi16 preferentially binds to dna with quadruplex structure and enhances dna quadruplex formation. PLoS One, 11(6):e0157156, 2016.
- [255] Seamus R Morrone, Tao Wang, Leeza M Constantoulakis, Richard M Hooy, Michael J Delannoy, and Jungsan Sohn. Cooperative assembly of ifi16 filaments on dsdna provides insights into host defense strategy. Proceedings of the National Academy of Sciences, 111(1):E62–E71, 2014.
- [256] Megan H Orzalli, Sara E Conwell, Christian Berrios, James A DeCaprio, and David M Knipe. Nuclear interferon-inducible protein 16 promotes silencing of herpesviral and transfected dna. Proceedings of the National Academy of Sciences, 110(47):E4492–E4501, 2013.
- [257] Anna-Maria Herzner, Cristina Amparo Hagmann, Marion Goldeck, Steven Wolter, Kirsten Kübler, Sabine Wittmann, Thomas Gramberg, Liudmila Andreeva, Karl-Peter Hopfner, Christina Mertens, et al. Sequence-specific activation of the dna sensor cgas by y-form dna structures as found in primary hiv-1 cdna. Nature immunology, 16(10):1025–1033, 2015.
- [258] Xavier Lahaye, Takeshi Satoh, Matteo Gentili, Silvia Cerboni, Cécile Conrad, Ilse Hurbain, Ahmed El Marjou, Christine Lacabaratz, Jean-Daniel Lelièvre, and Nicolas Manel. The capsids of hiv-1 and hiv-2 determine immune detection of the viral cdna by the innate sensor cgas in dendritic cells. Immunity, 39(6):1132–1142, 2013.
- [259] DM McCarty, Paul E Monahan, and Richard J Samulski. Self-complementary recombinant adeno-associated virus (scaav) vectors promote efficient transduction independently of dna synthesis. Gene therapy, 8(16):1248–1254, 2001.
- [260] Eric D Horowitz, K Shefaet Rahman, Brian D Bower, David J Dismuke, Michael R Falvo, Jack D Griffith, Stephen C Harvey, and Aravind Asokan. Biophysical and ultrastructural characterization of adeno-associated virus capsid uncoating and genome release. Journal of virology, 87(6):2994–3002, 2013.
- [261] Mavis Agbandje-McKenna and Jürgen Kleinschmidt. Aav capsid structure and cell interactions. In Adeno-Associated Virus, pages 47–92. Springer, 2012.

- [262] Brittney L Gurda, Christina Raupp, Ruth Popa-Wagner, Matthias Naumer, Norman H Olson, Robert Ng, Robert McKenna, Timothy S Baker, Jürgen A Kleinschmidt, and Mavis Agbandje-McKenna. Mapping a neutralizing epitope onto the capsid of adeno-associated virus serotype 8. Journal of virology, 86(15):7739–7751, 2012.
- [263] Michael A DiMattia, Hyun-Joo Nam, Kim Van Vliet, Matthew Mitchell, Antonette Bennett, Brittney L Gurda, Robert McKenna, Norman H Olson, Robert S Sinkovits, Mark Potter, et al. Structural insight into the unique properties of adeno-associated virus serotype 9. Journal of virology, 86(12):6947–6958, 2012.
- [264] Brittney L Gurda, Michael A DiMattia, Edward B Miller, Antonette Bennett, Robert McKenna, Wendy S Weichert, Christian D Nelson, Wei-jun Chen, Nicholas Muzyczka, Norman H Olson, et al. Capsid antibodies to different adeno-associated virus serotypes bind common regions. Journal of virology, 87(16):9111–9124, 2013.
- [265] Hyun-Joo Nam, Brittney L Gurda, Robert McKenna, Mark Potter, Barry Byrne, Maxim Salganik, Nicholas Muzyczka, and Mavis Agbandje-McKenna. Structural studies of adeno-associated virus serotype 8 capsid transitions associated with endosomal trafficking. Journal of virology, 85(22):11791–11799, 2011.
- [266] Hyun-Joo Nam, Michael Douglas Lane, Eric Padron, Brittney Gurda, Robert McKenna, Erik Kohlbrenner, George Aslanidi, Barry Byrne, Nicholas Muzyczka, Sergei Zolotukhin, et al. Structure of adeno-associated virus serotype 8, a gene therapy vector. Journal of virology, 81(22):12260–12271, 2007.
- [267] Robert Ng, Lakshmanan Govindasamy, Brittney L Gurda, Robert McKenna, Olga G Kozyreva, R Jude Samulski, Kristin N Parent, Timothy S Baker, and Mavis Agbandje-McKenna. Structural characterization of the dual glycan binding adeno-associated virus serotype 6. Journal of virology, 84(24):12945–12957, 2010.
- [268] Lakshmanan Govindasamy, Eric Padron, Robert McKenna, Nicholas Muzyczka, Nikola Kaludov, John A Chiorini, and Mavis Agbandje-McKenna. Structurally mapping the diverse phenotype of adeno-associated virus serotype 4. Journal of virology, 80(23):11556–11570, 2006.
- [269] Qing Xie, Weishu Bu, Smita Bhatia, Joan Hare, Thayumanasamy Somasundaram, Arezki Azzi, and Michael S Chapman. The atomic structure of adeno-associated virus (aav-2), a vector for human gene therapy. Proceedings of the National Academy of Sciences, 99(16):10405–10410, 2002.

- [270] Lauren M Drouin, Bridget Lins, Maria Janssen, Antonette Bennett, Paul Chipman, Robert McKenna, Weijun Chen, Nicholas Muzyczka, Giovanni Cardone, Timothy S Baker, et al. Cryo-electron microscopy reconstruction and stability studies of the wild type and the r432a variant of adeno-associated virus type 2 reveal that capsid structural stability is a major factor in genome packaging. Journal of virology, 90(19):8542–8551, 2016.
- [271] Yong Zi Tan, Sriram Aiyer, Mario Mietzsch, Joshua A Hull, Robert McKenna, Joshua Grieger, R Jude Samulski, Timothy S Baker, Mavis Agbandje-McKenna, and Dmitry Lyumkis. Sub-2 Å ewald curvature corrected structure of an aav2 capsid variant. Nature communications, 9(1):1–11, 2018.
- [272] Matthew Burg, Claire Rosebrough, Lauren M Drouin, Antonette Bennett, Mario Mietzsch, Paul Chipman, Robert McKenna, Duncan Sousa, Mark Potter, Barry Byrne, et al. Atomic structure of a rationally engineered gene delivery vector, aav2. 5. Journal of Structural Biology, 203(3):236–241, 2018.
- [273] Thomas F Lerch, Qing Xie, and Michael S Chapman. The structure of adeno-associated virus serotype 3b (aav-3b): insights into receptor binding and immune evasion. Virology, 403(1):26–36, 2010.
- [274] Thomas F Lerch, Jason K O’Donnell, Nancy L Meyer, Qing Xie, Kenneth A Taylor, Scott M Stagg, and Michael S Chapman. Structure of aav-dj, a retargeted gene therapy vector: cryo-electron microscopy at 4.5 Å resolution. Structure, 20(8):1310–1320, 2012.
- [275] Lakshmanan Govindasamy, Michael A DiMattia, Brittney L Gurda, Sujata Halder, Robert McKenna, John A Chiorini, Nicholas Muzyczka, Sergei Zolotukhin, and Mavis Agbandje-McKenna. Structural insights into adeno-associated virus serotype 5. Journal of virology, 87(20):11187–11199, 2013.
- [276] Qing Xie, Thomas F Lerch, Nancy L Meyer, and Michael S Chapman. Structure–function analysis of receptor-binding in adeno-associated virus serotype 6 (aav-6). Virology, 420(1):10–19, 2011.
- [277] Sujata Halder, Kim Van Vliet, J Kennon Smith, Thao Thi Phuong Duong, Robert McKenna, James M Wilson, and Mavis Agbandje-McKenna. Structure of neurotropic adeno-associated virus aavrh. 8. Journal of structural biology, 192(1):21–36, 2015.
- [278] Kyle Mikals, Hyun-Joo Nam, Kim Van Vliet, Luk H Vandenberghe, Lauren E Mays, Robert McKenna, James M Wilson, and Mavis Agbandje-McKenna. The

- structure of aavrh32. 33, a novel gene delivery vector. Journal of structural biology, 186(2):308–317, 2014.
- [279] Balasubramanian Venkatakrishnan, Joseph Yarbrough, John Domsic, Antonette Bennett, Brian Bothner, Olga G Kozyreva, R Jude Samulski, Nicholas Muzyczka, Robert McKenna, and Mavis Agbandje-McKenna. Structure and dynamics of adeno-associated virus serotype 1 vp1-unique n-terminal domain and its role in capsid trafficking. Journal of virology, 87(9):4974–4984, 2013.
- [280] Julien Bernaud, Axel Rossi, Anny Fis, Lara Gardette, Ludovic Aillot, Hildegard Büning, Martin Castelnovo, Anna Salvetti, and Cendrine Faivre-Moskalenko. Characterization of aav vector particle stability at the single-capsid level. Journal of biological physics, 44(2):181–194, 2018.
- [281] Vamseedhar Rayaprolu, Shannon Kruse, Ravi Kant, Balasubramanian Venkatakrishnan, Navid Movahed, Dewey Brooke, Bridget Lins, Antonette Bennett, Timothy Potter, Robert McKenna, et al. Comparative analysis of adeno-associated virus capsid stability and dynamics. Journal of virology, 87(24):13150–13160, 2013.
- [282] Antonette Bennett, Saajan Patel, Mario Mietzsch, Ariana Jose, Bridget Lins-Austin, C Yu Jennifer, Brian Bothner, Robert McKenna, and Mavis Agbandje-McKenna. Thermal stability as a determinant of aav serotype identity. Molecular Therapy-Methods & Clinical Development, 6:171–182, 2017.
- [283] Yinxia Xu, Ping Guo, Junping Zhang, Matthew Chrzanowski, Helen Chew, Jenni A Firrman, Nianli Sang, Yong Diao, and Weidong Xiao. Effects of thermally induced configuration changes on raav genome’s enzymatic accessibility. Molecular Therapy-Methods & Clinical Development, 2020.
- [284] Richard O Snyder and Joyce Francis. Adeno-associated viral vectors for clinical gene transfer studies. Current gene therapy, 5(3):311–321, 2005.
- [285] Eric Zinn, Simon Pacouret, Vadim Khaychuk, Heikki T Turunen, Livia S Carvalho, Eva Andres-Mateos, Samiksha Shah, Rajani Shelke, Anna C Maurer, Eva Plovie, et al. In silico reconstruction of the viral evolutionary lineage yields a potent gene therapy vector. Cell reports, 12(6):1056–1068, 2015.
- [286] Neil Rajoriya, Christophe Combet, Fabien Zoulim, and Harry LA Janssen. How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis b. time for an individualised approach? Journal of hepatology, 67(6):1281–1297, 2017.

- [287] Dulce Alfaiate, Paul Dény, and David Durantel. Hepatitis delta virus: From biological and medical aspects to current and investigational therapeutic options. Antiviral research, 122:112–129, 2015.
- [288] Ju-Tao Guo and Haitao Guo. Metabolism and function of hepatitis b virus cccdna: Implications for the development of cccdna-targeting antiviral therapeutics. Antiviral research, 122:91–100, 2015.
- [289] Michael Nassal. Hbv cccdna: viral persistence reservoir and key obstacle for a cure of chronic hepatitis b. Gut, 64(12):1972–1984, 2015.
- [290] Mirjam B Zeisel, Julie Lucifora, William S Mason, Camille Sureau, Jürgen Beck, Massimo Levrero, Michael Kann, Percy A Knolle, Monsef Benkirane, David Durantel, et al. Towards an hbv cure: state-of-the-art and unresolved questions—report of the anrs workshop on hbv cure. Gut, 64(8):1314–1326, 2015.
- [291] Kazuaki Chayama, C Nelson Hayes, Nobuhiko Hiraga, Hiromi Abe, Masataka Tsuge, and Michio Imamura. Animal model for study of human hepatitis viruses. Journal of gastroenterology and hepatology, 26(1):13–18, 2011.
- [292] Claudia Dembek and Ulrike Protzer. Mouse models for therapeutic vaccination against hepatitis b virus. Medical microbiology and immunology, 204(1):95–102, 2015.
- [293] Jie Wang, Hongxin Huang, Yongzhen Liu, Ran Chen, Ying Yan, Shu Shi, Jingyuan Xi, Jun Zou, Guangxin Yu, Xiaoyu Feng, et al. Hbv genome and life cycle. Hepatitis B Virus Infection, pages 17–37, 2020.
- [294] Andreas Schulze, Philippe Gripon, and Stephan Urban. Hepatitis b virus infection initiates with a large surface protein–dependent binding to heparan sulfate proteoglycans. Hepatology, 46(6):1759–1768, 2007.
- [295] Dieter Glebe, Stephan Urban, Eva V Knoop, Nilgün Çağ, Peter Krass, Stefanie Grün, Aiste Bulavaite, Kestutis Sasnauskas, and Wolfram H Gerlich. Mapping of the hepatitis b virus attachment site by use of infection-inhibiting pres1 lipopeptides and tupaia hepatocytes. Gastroenterology, 129(1):234–245, 2005.
- [296] Yann Le Duff, Matthieu Blanchet, and Camille Sureau. The pre-s1 and antigenic loop infectivity determinants of the hepatitis b virus envelope proteins are functionally independent. Journal of virology, 83(23):12443–12451, 2009.
- [297] Azeneth Barrera, Bernadette Guerra, Lena Notvall, and Robert E Lanford. Mapping of the hepatitis b virus pre-s1 domain involved in receptor recognition. Journal of virology, 79(15):9786–9798, 2005.

- [298] Corinna M Leistner, Stefanie Gruen-Bernhard, and Dieter Glebe. Role of glycosaminoglycans for binding and infection of hepatitis b virus. Cellular microbiology, 10(1):122–133, 2008.
- [299] Eloi R Verrier, Che C Colpitts, Charlotte Bach, Laura Heydmann, Amélie Weiss, Mickaël Renaud, Sarah C Durand, François Habersetzer, David Durantel, Georges Abou-Jaoudé, et al. A targeted functional rna interference screen uncovers glypican 5 as an entry factor for hepatitis b and d viruses. Hepatology, 63(1):35–48, 2016.
- [300] Philippe Gripon, Isabelle Cannie, and Stephan Urban. Efficient inhibition of hepatitis b virus infection by acylated peptides derived from the large viral surface protein. Journal of virology, 79(3):1613–1622, 2005.
- [301] Yi Ni, Florian A Lempp, Stefan Mehrle, Shirin Nkongolo, Christina Kaufman, Maria Fälth, Jan Stindt, Christian Königer, Michael Nassal, Ralf Kubitz, et al. Hepatitis b and d viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. Gastroenterology, 146(4):1070–1083, 2014.
- [302] Huan Yan, Guocai Zhong, Guangwei Xu, Wenhui He, Zhiyi Jing, Zhenchao Gao, Yi Huang, Yonghe Qi, Bo Peng, Haimin Wang, et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis b and d virus. elife, 1:e00049, 2012.
- [303] Charline Herrscher, Philippe Roingeard, and Emmanuelle Blanchard. Hepatitis b virus entry into cells. Cells, 9(6):1486, 2020.
- [304] Birgit Rabe, Dieter Glebe, and Michael Kann. Lipid-mediated introduction of hepatitis b virus capsids into nonsusceptible cells allows highly efficient replication and facilitates the study of early infection events. Journal of virology, 80(11):5465–5473, 2006.
- [305] Hung-Cheng Li, Er-Yi Huang, Pei-Yi Su, Szu-Yao Wu, Ching-Chun Yang, Young-Sun Lin, Wen-Chang Chang, and Chiaho Shih. Nuclear export and import of human hepatitis b virus capsid protein and particles. PLoS Pathog, 6(10):e1001162, 2010.
- [306] André Schmitz, Alexandra Schwarz, Michael Foss, Lixin Zhou, Birgit Rabe, Julia Hoellenriegel, Miriam Stoeber, Nelly Panté, and Michael Kann. Nucleoporin 153 arrests the nuclear import of hepatitis b virus capsids in the nuclear basket. PLoS Pathog, 6(1):e1000741, 2010.

- [307] Birgit Rabe, Mildred Delaleau, Andreas Bischof, Michael Foss, Irina Sominskaya, Paul Pumpens, Christian Cazenave, Michel Castroviejo, and Michael Kann. Nuclear entry of hepatitis b virus capsids involves disintegration to protein dimers followed by nuclear reassociation to capsids. PLoS Pathog, 5(8):e1000563, 2009.
- [308] Stephan Urban, Andreas Schulze, Maura Dandri, and Joerg Petersen. The replication cycle of hepatitis b virus. Journal of hepatology, 52(2):282–284, 2010.
- [309] Senko Tsukuda and Koichi Watashi. Hepatitis b virus biology and life cycle. Antiviral Research, page 104925, 2020.
- [310] LESLIE B Rall, DN Standring, O Laub, and WJ Rutter. Transcription of hepatitis b virus by rna polymerase ii. Molecular and cellular biology, 3(10):1766–1773, 1983.
- [311] Juergen Beck and Michael Nassal. Hepatitis b virus replication. World journal of gastroenterology: WJG, 13(1):48, 2007.
- [312] Volker Bruss. Envelopment of the hepatitis b virus nucleocapsid. Virus research, 106(2):199–209, 2004.
- [313] Stephanie Schädler and Eberhard Hildt. Hbv life cycle: entry and morphogenesis. Viruses, 1(2):185–209, 2009.
- [314] Julie Lucifora, Silke Arzberger, David Durantel, Laura Belloni, Michel Strubin, Massimo Levrero, Fabien Zoulim, Olivier Hantz, and Ulrike Protzer. Hepatitis b virus x protein is essential to initiate and maintain virus replication after infection. Journal of hepatology, 55(5):996–1003, 2011.
- [315] DS Dane, CH Cameron, and Moya Briggs. Virus-like particles in serum of patients with australia-antigen-associated hepatitis. The lancet, 295(7649):695–698, 1970.
- [316] Jianming Hu and Kuancheng Liu. Complete and incomplete hepatitis b virus particles: formation, function, and application. Viruses, 9(3):56, 2017.
- [317] Anna Pavlova, Leda Bassit, Bryan Cox, Maksym Korablyov, Chris Chipot, Kiran Verma, Olivia O Russell, Raymond F Schinazi, and James C Gumbart. Mechanism of action of hbv capsid assembly modulators predicted from binding to early assembly intermediates. bioRxiv, 2020.
- [318] Lisa Selzer and Adam Zlotnick. Assembly and release of hepatitis b virus. Cold Spring Harbor perspectives in medicine, 5(12):a021394, 2015.

- [319] Adam Zlotnick, Balasubramanian Venkatakrishnan, Zhenning Tan, Eric Lewellyn, William Turner, and Samson Francis. Core protein: a pleiotropic keystone in the hbv lifecycle. Antiviral research, 121:82–93, 2015.
- [320] Adam Zlotnick, Ira Palmer, Joshua D Kaufman, Stephen J Stahl, Alasdair C Steven, and Paul T Wingfield. Separation and crystallization of $t=3$ and $t=4$ icosahedral complexes of the hepatitis b virus core protein. Acta Crystallographica Section D: Biological Crystallography, 55(3):717–720, 1999.
- [321] Charlotte Uetrecht, Ioana M Barbu, Glen K Shoemaker, Esther Van Duijn, and Albert JR Heck. Interrogating viral capsid assembly with ion mobility–mass spectrometry. Nature chemistry, 3(2):126, 2011.
- [322] Corinne A Lutomski, Nicholas A Lykтей, Elizabeth E Pierson, Zhongchao Zhao, Adam Zlotnick, and Martin F Jarrold. Multiple pathways in capsid assembly. Journal of the American Chemical Society, 140(17):5784–5790, 2018.
- [323] Dale A Shepherd, Kris Holmes, David J Rowlands, Nicola J Stonehouse, and Alison E Ashcroft. Using ion mobility spectrometry–mass spectrometry to decipher the conformational and assembly characteristics of the hepatitis b capsid protein. Biophysical journal, 105(5):1258–1267, 2013.
- [324] RA Crowther, NA Kiselev, B Böttcher, JA Berriman, GP Borisova, V Ose, and P Pumpens. Three-dimensional structure of hepatitis b virus core particles determined by electron cryomicroscopy. Cell, 77(6):943–950, 1994.
- [325] Linda M Stannard and Margaret Hodgkiss. Morphological irregularities in dane particle cores. Journal of General Virology, 45(2):509–514, 1979.
- [326] John M Kenney, Carl-Henrik von Bonsdorff, Michael Nassal, and Stephen D Fuller. Evolutionary conservation in the hepatitis b virus core structure: comparison of human and duck cores. Structure, 3(10):1009–1019, 1995.
- [327] Paul T Wingfield, Stephen J Stahl, Robert W Williams, and Alasdair C Steven. Hepatitis core antigen produced in escherichia coli: subunit composition, conformation analysis, and in vitro capsid assembly. Biochemistry, 34(15):4919–4932, 1995.
- [328] A Zlotnick, N Cheng, JF Conway, FP Booy, AC Steven, SJ Stahl, and PT Wingfield. Dimorphism of hepatitis b virus capsids is strongly influenced by the c-terminus of the capsid protein. Biochemistry, 35(23):7412–7421, 1996.
- [329] WH Roos, MM Gibbons, A Arkhipov, C Uetrecht, NR Watts, PT Wingfield, AC Steven, AJR Heck, K Schulten, WS Klug, et al. Squeezing protein shells:

- how continuum elastic models, molecular dynamics simulations, and experiments coalesce at the nanoscale. Biophysical journal, 99(4):1175–1181, 2010.
- [330] Hong-Bo Pang, Liron Hevroni, Nitzan Kol, Debra M Eckert, Marianna Tsvitov, Michael S Kay, and Itay Rouso. Virion stiffness regulates immature hiv-1 entry. Retrovirology, 10(1):1–11, 2013.
- [331] Sai Li, Christian Sieben, Kai Ludwig, Chris T Höfer, Salvatore Chiantia, Andreas Herrmann, Frederic Eghiaian, and Iwan AT Schaap. ph-controlled two-step uncoating of influenza virus. Biophysical journal, 106(7):1447–1456, 2014.
- [332] María Medrano, Alejandro Valbuena, Alicia Rodríguez-Huete, and Mauricio G Mateu. Structural determinants of mechanical resistance against breakage of a virus-based protein nanoparticle at a resolution of single amino acids. Nanoscale, 11(19):9369–9383, 2019.
- [333] Nitzan Kol, Micha Gladnikoff, David Barlam, Roni Z Shneck, Alan Rein, and Itay Rouso. Mechanical properties of murine leukemia virus particles: effect of maturation. Biophysical journal, 91(2):767–774, 2006.
- [334] Ferry Kienberger, Rong Zhu, Rosita Moser, Dieter Blaas, and Peter Hinterdorfer. Monitoring rna release from human rhinovirus by dynamic force microscopy. Journal of virology, 78(7):3203–3209, 2004.
- [335] Helen G Hansma and Daniel E Laney. Dna binding to mica correlates with cationic radius: assay by atomic force microscopy. Biophysical journal, 70(4):1933–1939, 1996.
- [336] Cheng Zeng, Mercedes Hernando-Pérez, Bogdan Dragnea, Xiang Ma, Paul Van Der Schoot, and Roya Zandi. Contact mechanics of a small icosahedral virus. Physical Review Letters, 119(3):038102, 2017.
- [337] M D’Acunto and O Salvetti. Pattern recognition methods for thermal drift correction in atomic force microscopy imaging. Pattern Recognition and Image Analysis, 21(1):9, 2011.
- [338] Philipp Rahe, Ralf Bechstein, and Angelika Kühnle. Vertical and lateral drift corrections of scanning probe microscopy images. Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena, 28(3):C4E31–C4E38, 2010.
- [339] Iaroslav Gaponenko, Philippe Tückmantel, Benedikt Ziegler, Guillaume Rapin, Manisha Chhikara, and Patrycja Paruch. Computer vision distortion correction of scanning probe microscopy images. Scientific reports, 7(1):1–8, 2017.

- [340] Lucas Menou. Elasticité de capsides virales et défauts topologiques. PhD thesis, 2020. Thèse de doctorat dirigée par Castelnovo, Martin Physique Lyon 2020.
- [341] María Aznar, Sergi Roca-Bonet, and David Reguera. Viral nanomechanics with a virtual atomic force microscope. Journal of Physics: Condensed Matter, 30(26):264001, 2018.
- [342] Lauriane Lecoq, Marie-Laure Fogeron, Beat H Meier, Michael Nassal, and Anja Böckmann. Solid-state nmr for studying the structure and dynamics of viral assemblies. Viruses, 12(10):1069, 2020.
- [343] Shishan Wang, Marie-Laure Fogeron, Maarten Schledorn, Marie Dujardin, Susanne Penzel, Dara Burdette, Jan Martin Berke, Michael Nassal, Lauriane Lecoq, Beat H Meier, et al. Combining cell-free protein synthesis and nmr into a tool to study capsid assembly modulation. Frontiers in molecular biosciences, 6:67, 2019.
- [344] Maelenn Chevreuil, Lauriane Lecoq, Shishan Wang, Laetitia Gargowitsch, Naima Nhiri, Eric Jacquet, Thomas Zinn, Sonia Fieulaine, Stéphane Bressanelli, and Guillaume Tresset. Nonsymmetrical dynamics of the hbv capsid assembly and disassembly evidenced by their transient species. The Journal of Physical Chemistry B, 124(45):9987–9995, 2020.
- [345] Xuekui Yu, Lei Jin, Jonathan Jih, Chiaho Shih, and Z Hong Zhou. 3.5 Å cryoem structure of hepatitis b virus core assembled from full-length core protein. PloS one, 8(9):e69729, 2013.
- [346] Katherine C Chen, John J Tyson, Muriel Lederman, Ernest R Stout, and Robert C Bates. A kinetic hairpin transfer model for parvoviral dna replication. Journal of molecular biology, 208(2):283–296, 1989.