



# ETUDE STRUCTURALE PAR MICROSCOPIE ELECTRONIQUE ET CRISTALLOGRAPHIE AUX RAYONS X DE LA CAPSIDE DES ADENOVIRUS

Celine Fabry

## ► To cite this version:

Celine Fabry. ETUDE STRUCTURALE PAR MICROSCOPIE ELECTRONIQUE ET CRISTALLOGRAPHIE AUX RAYONS X DE LA CAPSIDE DES ADENOVIRUS. Biochimie [q-bio.BM]. Université Joseph-Fourier - Grenoble I, 2008. Français. NNT: . tel-00340464

HAL Id: tel-00340464

<https://theses.hal.science/tel-00340464>

Submitted on 20 Nov 2008

**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.

# THESE

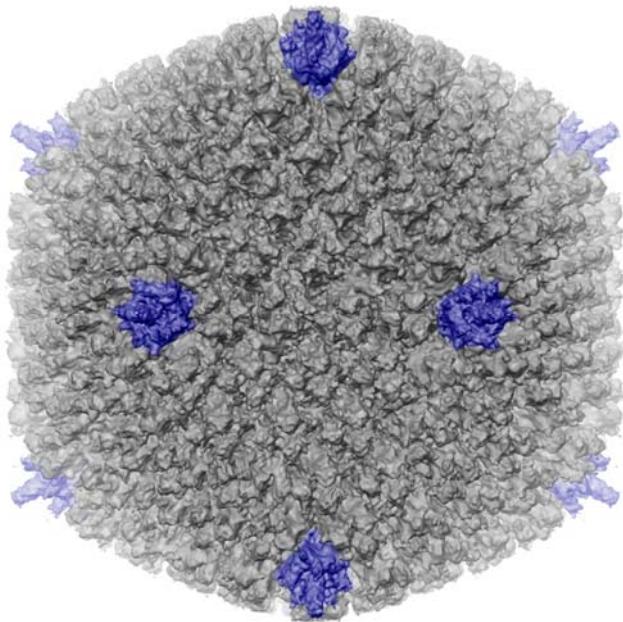
Pour l'obtention du diplôme de Docteur de l'Université Joseph Fourier - Grenoble I  
Spécialité : Biologie structurale et Nanobiologie

Présentée et soutenue publiquement par :

**Céline Fabry**

Le 24 Septembre 2008 à 14h

## ETUDE STRUCTURALE PAR MICROSCOPIE ELECTRONIQUE ET CRISTALLOGRAPHIE AUX RAYONS X DE LA CAPSIDE DES ADENOVIRUS



Devant le jury composé de :

Pr Pierre Boulanger	Rapporteur
Dr Daniel Thomas	Rapporteur
Dr Serge Perez	Président du Jury
Dr Daniel Levy	Examinateur
Dr Guy Schoehn	Directeur de thèse
Pr Rob Ruigrok	Co-directeur de thèse

Travaux effectués à l'UVHCI, Unit of Virus Host Cell Interactions.



## Résumé

Les Adénovirus sont des virus non enveloppés à symétrie icosaédrique. Découverts chez l'homme dans les années 50, ils constituent la famille des *Adenoviridae*. Ils existent une grande variété de souches pouvant infecter une grande variété d'hôtes allant de l'homme jusqu'au poisson. L'étude de l'Adénovirus a été stimulée par la découverte en 1962 par Trentin et collaborateurs de leur capacité à induire des tumeurs chez les bébés hamsters. Sur le plan structural, de nombreuses études en Microscopie Electronique ont été menées pour comprendre l'assemblage et la formation de la capsid. Depuis les années 90, avec la progression de l'informatique et des techniques liées à l'analyse d'images, la structure de la capsid des Adénovirus humains est de mieux en mieux comprise mais la localisation de certaines protéines mineures restaient encore incertaine ou inconnue. Nous avons, à travers l'utilisation de la Microscopie Electronique associée à l'analyse d'images, obtenu des reconstructions 3D pour plusieurs souches d'Adénovirus humains, canins et aviaires. L'obtention d'un modèle 3D à haute résolution pour l'Adénovirus humain de type 5 nous a permis de reconstituer un modèle quasi-atomique par la méthode de « fitting » et de déterminer avec plus de précisions la position de certaines protéines mineures comme la protéine IX, IIIa et VIII. L'étude d'une souche mutante d'Adénovirus de type 5 en complexe avec un Fab ainsi que d'une souche canine nous a permis de localiser précisément l'extrémité C terminale de la protéine IX. Enfin l'obtention d'une structure à haute résolution d'un virus immature dérivé de l'Adénovirus humain de type 2 a ouvert la voie sur l'étude et la compréhension de la maturation et de la phase précoce de son cycle cellulaire.

**Mots-clés:** Adénovirus, *Adenoviridae*, Microscopie Electronique, Reconstruction 3D, Fitting, Cristallographie, Virus à ADN double brin, symétries icosaédriques.

## Abstract

Adenoviruses are non-enveloped icosahedral viruses. They were found during the 50's and since constitute the huge *Adenoviridae* family. You can find a large variety of strains able to infect a large variety of hosts from humans to fishes. Fundamental studies on adenoviruses were largely stimulated since the discovery in 1962 by Trentin et al. of their ability to induce tumours in baby hamsters. From the structural point of view, many studies based on electron microscopy were done to understand how the adenovirus capsid is assembled. Since the last fifteen years, progress made in computer science and Electron Microscopy based techniques for 3D image analysis helped scientific community to get a better understanding of the adenovirus capsid structure but many minor proteins were uncertainly localised or totally unknown. Our work is based on Electron Microscopy and 3D image analysis. We obtained 3D models of different adenovirus strains like human, fowl and canine ones where two are below 10 Å resolution. A high resolution 3D reconstruction of Human Adenovirus type 5 helped us to reconstitute a quasi atomic model of the entire capsid and to determine the localisation of minor proteins like IX, IIIa and VIII. We also localised the C terminal part of protein IX from a 3D model of a mutant Human Adenovirus type 2 complexed with a Fab antibody and a canine Adenovirus. Finally, we were interested in the 3D structure of the non mature strain *ts1* derived from Human Adenovirus type 2. The “below10Å” 3D reconstruction obtained constitutes infamy into structural understanding of adenovirus maturation and early phase of infection.

**Keywords:** Adénovirus, *Adenoviridae*, Electron Microscopy, 3D reconstruction, Fitting, Crystallography, Double-stranded DNA viruses, icosahedral symmetries.



*Je remercie tout d'abord mes directeurs de thèse, le Dr. Guy Schoebn et le Pr. Rob Ruigrok avec qui j'ai beaucoup apprécié travailler durant mes quatre années de thèse. Je suis très reconnaissante de la patience dont ils ont fait preuve surtout pendant la dernière année. Ils ont accepté de me laisser plus de temps pour la rédaction de ce manuscrit et la préparation de ma soutenance. Cela m'a permis de mener à bien ma formation en Informatique. Je tiens à les remercier particulièrement pour cela. J'espère ne pas avoir trop traumatisé Guy en parlant à mon ordinateur lorsque nous partagions le même bureau.*

*Je remercie le Pr. Pierre Boulanger et le Dr. Daniel Thomas d'avoir accepté de lire ce manuscrit et d'être les rapporteurs de mon travail de thèse. J'aimerais aussi remercier tous les membres de mon jury, le Dr. Serge Perez, le Dr. Daniel Levy ainsi qu'à mes deux directeurs de thèse pour leur participation.*

*J'aimerais ensuite remercier toutes les personnes avec qui nous avons collaboré pour mener à bien ce projet de thèse sur l'étude de la structure des Adénovirus. A Lyon le Pr. Pierre Boulanger et le Dr. Manuel Rosa-Calatrava, à Montpellier le Dr. Eric Kremer, et à Zurich le Dr. Greber ainsi que leurs équipes respectives.*

*Dans notre laboratoire, j'aimerais tout particulièrement remercier mes amis encore au labo comme Francine Gérard, Majida El Bakkouri, Lucy Freeman, et ceux qui en sont partis : Thibault Géoui, Aurélie Albertini, Sophie Torres, Hatice Akarsu, Cédric Clapier, Jeanne Morinière, Carlos Fernando-Tornero pour tous les moments que nous avons passé ensemble. C'est vous tous qui m'avait fait aimé la vie au laboratoire.*

*J'aimerais maintenant remercier mes parents, Rose et Robert Fabry sans qui rien n'aurais pu arriver. Ils m'ont toujours soutenu dans mes choix quelque soit leur direction. Cela m'a toujours beaucoup touché.*

*La dernière personne à qui je dois beaucoup est la personne qui partage ma vie et qui l'a partagé durant toute ma thèse, Cyril Dian. Cyril m'a aussi toujours soutenu psychologiquement pendant l'écriture de cette thèse. Je tiens à exprimer toute mon affection et mon amour à Cyril pour sa présence et son soutien.*

*Malgré les difficultés que j'ai éprouvées pour rédiger ce manuscrit et soutenir ma thèse, je tiens à ce que vous sachiez que mon doctorat a été une expérience formidable pleine d'enrichissements aussi bien personnels que professionnels.*

*Si un petit génie me demandait de remonter le temps pour changer quelque chose à ma vie, je ne changerais rien à tout ca. Cette thèse a été pour moi une grande expérience de vie :-)*



## Table des Matières

<b>INTRODUCTION .....</b>	<b>1</b>
<b>I. La symétrie, un aspect fascinant de la nature.....</b>	<b>1</b>
I.1 Les opérateurs de symétrie élémentaires .....	2
I.2 Les Solides de Platon .....	5
I. 2.1 <i>Le cube</i> .....	6
I. 2.2 <i>Le tétraèdre</i> .....	7
I. 2.3 <i>L'octaèdre</i> .....	7
I. 2.4 <i>L'icosaèdre</i> .....	8
I. 2.5 <i>Le dodécaèdre</i> .....	9
I.3 La Symétrie dans les macromolécules biologiques et les virus .....	9
<b>II. La combinaison des données issues de la Microscopie Electronique et de la cristallographie aux rayons X, une méthode très adaptée à l'étude structurale de gros complexes macromoléculaires .....</b>	<b>14</b>
II.1 La Microscopie Electronique.....	14
II. 1.1 <i>Le principe de la Microscopie Electronique en transmission</i> .....	15
II. 1.2 <i>Le microscope électronique produit une image déformée de l'objet</i> .....	17
II. 1.3 <i>La préparation des échantillons biologiques et la prise d'images en MET</i> .....	20
II. 1.4 <i>La correction, l'analyse des images et le calcul d'une reconstruction 3D</i> .....	24
II.2 La Cristallographie aux rayons X .....	30
II.3 Obtention d'un modèle quasi-atomique par « fitting ».....	30
II. 3.1 <i>En corps rigide</i> .....	31
II. 3.2 <i>Les modes normaux et le « fitting » flexible</i> .....	33
<b>III. Objet du travail de thèse .....</b>	<b>34</b>
<b>STRUCTURE DE LA CAPSIDE DES ADENOVIRUS ET DES SOUS PARTICULES VIRALES .....</b>	<b>35</b>
<b>I. Introduction .....</b>	<b>35</b>
I.1 La famille des <i>Adenoviridae</i> .....	35
I.2 Un peu d'épidémiologie .....	37
I.3 Structure et organisation générale de la capsid.....	38
I. 3.1 <i>La symétrie de la capsid</i> .....	40
I. 3.2 <i>L'Hexon</i> .....	41
I. 3.3 <i>La Base du Penton</i> .....	42
I. 3.4 <i>La Fibre</i> .....	44
I. 3.5 <i>Les protéines structurales mineures</i> .....	45
I.4 Structure et Organisation du génome des Adénovirus .....	47
I. 4.1 <i>Structure du nucléoïde</i> .....	48
I. 4.2 <i>Unités transcriptionnelles et organisation du génome</i> .....	48
I.5 Le cycle viral.....	51
I. 5.1 <i>Entrée du virus dans la cellule et phase précoce d'infection</i> .....	53
I. 5.2 <i>RéPLICATION du génome viral</i> .....	54
I. 5.3 <i>Phase tardive d'infection et Lyse de la cellule infectée</i> .....	55
I. 5.4 <i>Conclusion sur le cycle virale de l'Adénovirus</i> .....	55
I.6 Structure du dodécaèdre et particularités du dodécaèdre de type 3 .....	56
I.7 Les Adénovirus Aviaires .....	58
I.8 Les Adénovirus Canins .....	59
I.9 Les Adénovirus, vecteurs pour la thérapie génique et anti-cancer.....	60
<b>II. Le modèle quasi atomique de la capsid de l'Adénovirus humain de type 5 .....</b>	<b>62</b>
II.1 Article 1 .....	62
II.2 Matériel et Méthodes complémentaires .....	73
II. 2.1 <i>Informations supplémentaires à l'article 1</i> .....	73
II. 2.2 <i>La microscopie électronique et l'analyse d'images</i> .....	80
II.3 Conclusion .....	88

<b>III. Etude structurale de la capsid de l'Adénovirus Aviaire de type 1 .....</b>	<b>89</b>
III.1 Matériel et Méthodes.....	89
<i>III. 1.1 Purification du virus CELO .....</i>	<i>89</i>
<i>III. 1.2 La microscopie électronique et l'analyse d'images .....</i>	<i>91</i>
<i>III. 1.3 Le fitting.....</i>	<i>92</i>
III.2 Résultats et discussion .....	92
III.3 Conclusion.....	98
<b>IV. Etude Structurale de la Capside de l'Adénovirus Canin de type 2.....</b>	<b>99</b>
IV.1 Article 2 .....	101
IV.2 Conclusion.....	114
<b>V. Etude structurale des Complexes Fab - Adénovirus Humain de type 5 pour la localisation de l'extrémité C terminale de la protéine IX.....</b>	<b>115</b>
V.1 Matériel et Méthodes.....	115
<i>V. 1.1 Les échantillons.....</i>	<i>115</i>
<i>V. 1.2 Préparation et Purification des Fab .....</i>	<i>117</i>
<i>V. 1.3 Préparation des complexes Virus – Fab.....</i>	<i>118</i>
<i>V. 1.4 CryoME .....</i>	<i>118</i>
V.2 Résultats et discussion.....	119
V.3 Conclusion .....	125
<b>VI. Etude structurale d'un mutant immature de l'Adénovirus Humain de type 2.....</b>	<b>126</b>
VI.1 Matériel et Méthodes.....	126
<i>VI. 1.1 La CryoME.....</i>	<i>126</i>
<i>VI. 1.2 Le « fitting » en corps rigide .....</i>	<i>127</i>
VI.2 Résultats et discussion .....	127
VI.3 Conclusion.....	132
<b>VII. Le modèle quasi atomique du dodécaèdre de l'Adénovirus Humain de type3.....</b>	<b>133</b>
VII.1 Article 3 .....	133
VII.2 Conclusion .....	145
<b>DISCUSSION GENERALE SUR LA STRUCTURE DE LA CAPSIDE DES ADENOVIRUS .....</b>	<b>146</b>
<b>CONCLUSION ET PERSPECTIVES.....</b>	<b>151</b>
<b>ANNEXES .....</b>	<b>153</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES .....</b>	<b>171</b>

## *Liste des Abréviations*

2D: **Bi**dimensionnel  
3D: **Tri**dimensionnel  
AA: **Acide Aminé**  
ADN: **Acide DésoxyriboNucléique**  
ADP: **Adenovirus Death Protein**  
AdV: **Adénovirus**  
AFSSA: **Agence Française de Sécurité Sanitaire des Aliments**  
AmMb: **Ammonium Molybdate**, molybdate d'ammonium  
ARN: **Acide RiboNucléique**  
at: **atmosphère**, unité de pression, 1 at = pression atmosphérique au niveau de la mer  
CAdV: **Canine Adenovirus**, Adénovirus canin  
CAR: **Coxsackie and Adenovirus Receptor**  
CC: **Coefficient de Corrélation**  
CELO: **Chicken Embryo Lethal Orphan**  
CryoME: **Cryo Microscopie Electronique**  
CTF: « **Contrast Transfer Function** » **Fonction de Transfert de Contraste**  
Da : **Dalton**, unité de masse atomique  
DBP : **DNA Binding Protein**, protéine de liaison à l'ADN  
dCMP : **désoxyCytidine MonoPhosphate**  
FAdV: **Fowl Adenovirus**, Adénovirus aviaire  
FEG : **Field Emission Gun**, Canon à Emission de champ  
FSC: **Fourier Shell Correlation**  
GON: **Group Of Nine Hexons**, Groupe des Neufs Hexons  
HAdV: **Human Adenovirus**, Adénovirus humain  
IBDV: **Infectious Bursal Disease Virus**  
Ig G: **Immunoglobuline de type G**  
ITR : **Inverted Terminal Repeat**, séquences inversées répétées  
LMH: **Leghorn Male Hepatoma**  
ME: **Microscopie Electronique**  
MET: **Microscopie Electronique en Transmission**  
MEB: **Microscope Electronique à Balayage**  
MEBT: **Microscope Electronique à Balayage en Transmission**  
nm : **nanomètre** =  $10^{-9}$  m  
pb: **paires de bases**  
PDB : **Protein Data Bank**, Base de données des structures de protéines  
PFT : **Polar Fourier Transform**, TF en coordonnées polaires  
PIF : **Purdue Image Format**, format d'image pour la ME développé à l'université de Purdue  
RMN: **Résonance Magnétique Nucléaire**  
SAB : **Sérum Albumine Bovine**  
SST: **Sodium silicotungstate**, Silicotungstate de sodium  
TBSV: **Tomato Bushny Stunt Virus**  
TF: **Transformée de Fourier**  
UA: **Unité Asymétrique**  
UrAc: **Uranyl Acetate**, Acétate d'uranylique  
VLP : **Viral Like Particles**  
WT : **Wild Type** ; de phénotype sauvage

**Code à 1 lettre et à 3 lettres des Acides Aminés :**

A	Ala	Alanine	M	Met	Méthionine
C	Cys	Cystéine	N	Asn	Asparagine
D	Asp	acide Aspartique	P	Pro	Proline
E	Glu	acide Glutamique	Q	Gln	Glutamine
F	Phe	Phénylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Sérine
H	His	Histidine	T	Thr	Thrénanine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophane
L	Leu	Leucine	Y	Tyr	Tyrosine

# INTRODUCTION

---

J'ai choisi de débuter mon manuscrit par une introduction générale sur les symétries ainsi que les méthodes biophysiques et informatiques que j'ai utilisées au cours de mes différents projets de recherche. J'ai décidé de ne pas traiter l'introduction biologique en rapport avec chacun de mes sujets dans cette partie par souci de structure et de clarté de l'exposé. Vous trouverez donc une introduction biologique au début des chapitres détaillants les projets dans lesquels j'ai été impliqué.

## *I. La symétrie, un aspect fascinant de la nature*

L'homme a toujours été fasciné par les assemblages symétriques, réguliers et répétitifs d'éléments présents dans la nature. La nature peut en effet donner naissance à des éléments de symétrie d'une troublante perfection. L'exemple le plus marquant est celui des minéraux. Les éléments chimiques des couches profondes de la croûte terrestre, sous l'action de la pression et de la température, s'assemblent pour former des cristaux minéraux d'une symétrie quasi parfaite. Le diamant, un cristal uniquement formé d'atomes de carbone illustre bien la fascination que l'homme peut éprouver. Il est aussi intéressant d'observer comment l'eau sous l'action d'une baisse de température peut former des assemblages très symétriques comme les flocons de neige (Figure 1a). Même au sein du monde vivant, la symétrie est très présente particulièrement dans le monde végétal mais aussi dans le monde animal (Figure 1). Certaines étoiles de mer, par exemple, présentent un axe de symétrie d'ordre 5 qui relie ses cinq bras (Figure 1).



*Figure 1: Nature et Symétrie*

(a) Photo d'une passiflore qui présente un axe de symétrie d'ordre 5 et 3 au niveau des étamines et de son pistil. (b) Photo d'un cristal de neige qui présente un axe de symétrie d'ordre 6 (Source : <http://www.snowcrystals.com/>). (c) Photo d'une étoile de mer qui présente un axe de symétrie d'ordre 5 en son centre (Source : <http://www.kluger.fr/productssimple3.html>).

Les symétries au sein d'un objet biologique présentent des avantages du point de vue de l'objet biologique lui-même ainsi que de celui de l'observateur. Du point de vue de l'observateur, cela permet une description mathématique précise de l'objet biologique que l'on cherche à étudier. Du point de vue biologique, la symétrie permet la mise en place de structures larges à partir d'une unité simple qui peut être répétée. Un exemple particulièrement frappant de l'utilité de la symétrie en biologie est celui de certains virus comme l'Adénovirus. Ce virus, d'assez grande taille (100 nm) ne contient que relativement peu d'informations génétiques. L'encapsidation de son génome nécessiterait une grande quantité de protéines différentes si ce virus n'était pas capable de fabriquer sa capsid à partir de seulement deux protéines différentes, assemblées en de nombreuses copies (cf. Partie 2). La symétrie est aussi visible dans l'art ou dans l'architecture à travers les créations de l'homme. De nombreuses œuvres d'art comme les tableaux de Victor Vasarely ainsi que des constructions illustres comme le Taj Mahal en Inde montrent comment l'homme à exploiter la symétrie.

J'ai choisi de centrer l'introduction générale de ma thèse sur la description mathématique de la symétrie. Une grande partie des données mathématiques est tirée du livre « Description de la symétrie : des groupes de symétrie aux structures fractales » de Jean Sivardié aux éditions EDP Sciences, 2004.

### I.1 Les opérateurs de symétrie élémentaires

La symétrie est un concept simple et complexe à la fois. Il existe une grande variété d'éléments de symétries, tous définis par des opérateurs mathématiques propres, mais je vais centrer mon exposé sur le type le plus simple : l'isométrie. La compréhension des isométries est importante pour l'étude des structures biologiques ainsi que les méthodes biophysiques qui permettent de déterminer la structure des objets biologiques comme les protéines ou les virus.

Une isométrie est une transformation qui ne modifie pas la taille de l'image par rapport à l'objet. On dit qu'elle conserve les longueurs. Toutes les isométries sont obtenues par combinaison de 3 opérateurs de symétrie élémentaires qui sont la translation, la rotation et l'inversion. Ce dernier opérateur (l'inversion) entraîne une modification de la chiralité de l'objet. Les molécules biologiques possédant une chiralité invariante, l'inversion est donc un opérateur de symétrie interdit dans le monde vivant. Je vais donc décrire dans cette partie les deux opérateurs restants de façon plus détaillée. Ceux-ci sont essentiels pour décrire les objets biologiques que j'ai étudié au cours de ma

thèse ainsi que pour le traitement des données obtenues en cristallographie aux rayons X, en ME ou encore à la méthode de « fitting » (cf. Chapitre II).

La translation. C'est l'opérateur de symétrie élémentaire le plus simple. Une translation  $\vec{t}$  est définie par un vecteur aux coordonnées (a, b, c). Cet opérateur permet le déplacement de l'objet d'un point M à un point M'. Le point M (x, y, z) est donc transformé en un point M' (x', y', z') où  $x' = x + a, y' = y + b, z' = z + c$ . Un système subit une translation  $\vec{t}$  si tous ces points subissent la même translation  $\vec{t}$ .

La rotation. La rotation est un opérateur de symétrie élémentaire majeur. Une symétrie de rotation simple se réalise autour d'un axe  $\Delta$  d'ordre N où N est un nombre entier. L'axe est défini par un vecteur unitaire  $\vec{u}$  qui donne son orientation dans l'espace ainsi qu'un angle  $\theta$  de rotation. Un axe de rotation d'ordre N est défini par N rotations d'un angle  $\theta = \frac{360^\circ}{N}$ . Pour un axe d'ordre  $N = 3$ , il faut appliquer 3 rotations successives de  $120^\circ$  chacune pour revenir à la position de départ (Figure 2).

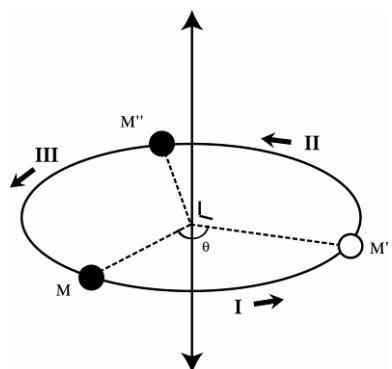


Figure 2 : Représentation schématique d'un axe de rotation d'ordre 3

La position M doit subir trois rotations (I, II et III) d'un angle  $\theta = \frac{360^\circ}{3} = 120^\circ$  pour revenir à sa même position initiale.

Mathématiquement, une rotation peut être décrite de plusieurs façons. En cristallographie, la représentation matricielle est privilégiée car les valeurs de la matrice sont exprimées en coordonnées fractionnaires, définies par rapport à la maille du cristal que l'on étudie. En ME, la représentation d'Euler (1707-1783) est majoritairement employée. En effet, l'objet étudié représente un objet isolé qui n'est pas sous forme cristalline. Ce que l'on cherche à déterminer est l'orientation selon laquelle cet objet a été projeté. Cette orientation ne peut donc pas être définie au sein d'un réseau comme en cristallographie. L'utilisation de la représentation d'Euler semble donc plus commode. Dans le premier cas, la transformation  $\alpha$  est alors décrite par une matrice  $3 \times 3$  orthogonale

$\Gamma(\alpha) = \begin{pmatrix} t_{xx} & t_{xy} & t_{xz} \\ t_{yx} & t_{yy} & t_{yz} \\ t_{zx} & t_{zy} & t_{zz} \end{pmatrix}$ . Cette représentation est particulièrement intéressante pour les isométries car

elle est *fidèle*. En effet, deux isométries distinctes sont représentées par deux matrices distinctes. Une matrice importante est la matrice identité représentant la transformation identité. Elle fait

correspondre un objet à lui-même et est définie par  $I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ . Dans le deuxième cas, Euler a

introduit un système de coordonnées qui permet de décomposer une rotation en un produit de trois rotations (Figure 3). Dans ce système de notation, on passe d'un référentiel  $Oxyz$  fixe à un référentiel lié au solide  $Ox'y'z'$ . La première rotation  $\Psi$  s'effectue autour de l'axe  $Oz$  et le repère passe en  $Ouvw$  (repère bleu sur la Figure 3). La deuxième rotation  $\theta$  s'effectue autour de l'axe  $Ou$  dans le nouveau repère et permet le passage au repère  $Ouwz'$  (en vert sur la Figure 3). Enfin une dernière rotation  $\varphi$  autour de l'axe  $z'$  permet le passage au repère final  $Ox'y'z'$  lié à l'objet (en rouge, vert sur la Figure 3).

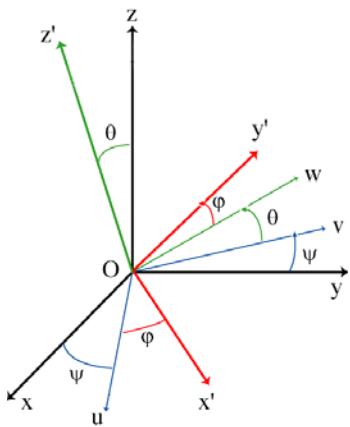


Figure 3 : Définition des angles d'Euler

La notation d'Euler est définie par trois rotations  $\Psi$ ,  $\theta$  et  $\varphi$  successives qui modifient le repère fixe dans lequel se situe l'objet à transformer. On passe d'un repère  $Oxyz$  à un repère  $Ox'y'z'$ .

Une isométrie peut être aussi être composée. En effet, au lieu d'une seule opération de symétrie, elle peut comporter une combinaison de plusieurs opérateurs élémentaires. On remarque, particulièrement en cristallographie, la combinaison d'une translation et d'une rotation qui donnent lieu à une translation hélicoïdale. Une rotation fait tout d'abord passer l'objet de la position  $M$  à  $M'$  puis une translation fait passer  $M'$  en  $M''$  (Figure 4). La compréhension des isométries est importante pour la suite de mon introduction. Je vais maintenant illustrer l'utilisation des symétries élémentaires à travers la description des solides de Platon.

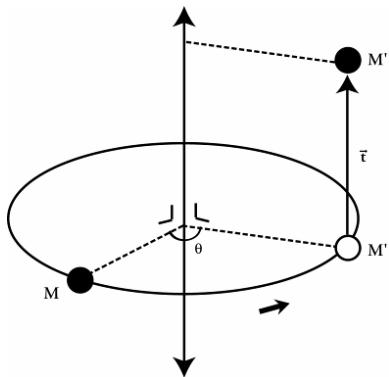


Figure 4 : Représentation schématique d'un axe hélicoïdal d'ordre 3

La position  $M$  subit tout d'abord une rotation d'un angle  $\theta = \frac{360^\circ}{3} = 120^\circ$  puis

la position  $M'$  subit une translation  $\vec{t}$ . L'objet  $M$  a donc subit une translation hélicoïdale. L'axe de rotation d'ordre devient alors un axe hélicoïdal.

## I.2 Les Solides de Platon

Les solides de Platon sont des références mathématiques qui me permettent de décrire la symétrie des particules biologiques que j'ai étudiées au cours de ma thèse. Les structures déterminées à partir d'images obtenues en ME sur lesquelles j'ai travaillé possèdent la géométrie de quatre sur cinq de ces solides. La compréhension de leur arrangement géométrique permet de mieux appréhender la structure des objets biologiques. Ces solides ont toujours fascinés les physiciens et ont de multiples utilisations. Un exemple ludique est leur utilisation sous forme de dés (Figure 5).



Figure 5 : Jeu de dés polyédriques à 4, 6, 8, 12 et 20 faces

Ces dés représentent, de gauche à droite un tétraèdre (4 faces), un cube (6 faces), un octaèdre (8 faces), un dodécaèdre (12 faces) et un icosaèdre (20 faces).

Les solides de Platon sont des polyèdres réguliers. Un polyèdre régulier est un volume limité par au moins 4 faces, toutes polygonales, régulières et identiques entre elles.

Théétète d'Athènes (-415, -395 ou -369), un mathématicien de la Grèce antique fit la première description mathématique de ces solides et prouva qu'il en existe cinq et seulement cinq : le tétraèdre, l'octaèdre, le cube, le dodécaèdre et l'icosaèdre. Les travaux de ce mathématicien sont connus par l'intermédiaire de Platon, célèbre philosophe de la Grèce antique (-427, -348). En effet, Platon reprit les travaux de Théétète dans le dialogue *Timée* datant de 360 avant JC et associa à chaque élément qui compose le monde (le feu, l'eau, la terre, l'air et l'univers) un des cinq solides. Par la suite, Euclide (-325, -265) reprit la démonstration de l'existence de seulement cinq polyèdres réguliers dans son œuvre *Les éléments*. Mathématiquement, ces 5 solides répondent au théorème de topologie énoncé par Euler en 1750. Si on considère un polyèdre convexe quelconque ayant S sommets, F faces et A arêtes, Euler démontra la relation suivante :

$$S + F - A = 2$$

Les solides de Platon vérifient le théorème d'Euler d'une façon particulière (Tableau 1) :

$$S \times z = F \times n = 2 \times A$$

Un peu de géométrie. Tous les solides de Platon sont inscriptibles dans une sphère et leurs faces sont des polygones réguliers isométriques. Toutes les faces sont identiques au sein d'un même solide. Chaque solide de Platon est aussi défini par un groupe de symétrie propre G (Tableau 1).

Donc pour chaque solide, on peut définir un ensemble de paramètres vérifiant le théorème de topologie d'Euler (Tableau 1). On remarque aussi qu'un axe de rotation d'ordre z passe par chacun des S sommets. Un axe d'ordre n est perpendiculaire et centré par rapport à chacune des F faces. Les solides de Platon possèdent tous un solide dual ou réciproque de même symétrie propre. Deux solides sont duals quand on peut les inscrire l'un dans l'autre. Les sommets du polyèdre dual passent par le centre de gravité ou un point quelconque des faces du solide initial. En accord avec le tableau 1, l'octaèdre et le cube sont duals ainsi que le dodécaèdre et l'icosaèdre. Le tétraèdre quant à lui est un cas un peu particulier puisqu'il est le dual de lui-même.

polyèdre	z	n	A	S	F	G
Tétraèdre	3	3	6	4	4	23
Octaèdre	4	3	12	6	8	432
Cube	3	4	12	8	6	432
Dodécaèdre	3	5	30	20	12	532
Icosaèdre	5	3	30	12	20	532

Tableau 1 : Caractéristiques des solides de Platon

Ce tableau décrit les différentes valeurs du théorème de topologie d'Euler pour chacun des cinq solides de Platon. z représente la coordinence ou valence de chaque sommet, qui correspond au nombre de faces ayant le même sommet. n correspond au nombre de cotés par face, A au nombre d'arêtes, S au nombre de sommets, F au nombre de faces et G au groupe de symétrie propre du solide.

Dans les parties suivantes, je vais détailler individuellement chaque solide de Platon en insistant particulièrement sur les solides qui m'ont servi de référence mathématique dans mes projets de recherche. L'icosaèdre et le dodécaèdre sont des volumes références pour l'étude de la géométrie de la capsid des adénovirus ainsi que des sous particules virales qui en sont dérivées. Le tétraèdre et l'octaèdre ont servi de référence pour l'étude de la structure de la protéase TET1 de *P. horikoshii*.

### I. 2.1 Le cube

Le cube ou hexaèdre régulier est le symbole de la Terre dans la philosophie de Platon. Composé de 6 faces carrées, le cube est le solide le plus difficile à mouvoir et par conséquent représente bien la Terre, l'élément le plus tenace. Son groupe de symétrie propre est 432 (Tableau 1).

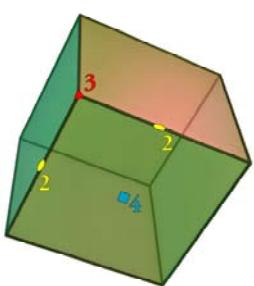


Figure 6 : Le cube

Le cube est représenté en perspective afin d'apprécier son volume en trois dimensions. L'axe d'ordre 3 présent à chaque sommet est coloré en rouge, l'axe 2 au milieu de chaque arête en jaune et l'axe d'ordre 4 au centre de chaque face est noté en bleu.

L'unité minimum nécessaire pour reconstituer un cube par symétrie est  $1/24^{\text{ème}}$  du volume total. En effet, on observe trois axes de rotation d'ordre  $n = 4$  au centre de chaque face,

perpendiculaires à deux faces opposées. Les quatre axes ternaires présents au niveau des sommets du cube sont disposés suivant les diagonales. Enfin un axe d'ordre 2 est présent au milieu de chaque arête et traverse deux arêtes opposées. L'octaèdre et le cube sont des solides duals puisqu'ils possèdent le même groupe de symétrie propre mais on peut aussi inscrire le tétraèdre, le dodécaèdre et l'icosaèdre au sein d'un cube. Cette propriété géométrique est particulièrement intéressante pour décrire la position des sommets des différents volumes inscrits dans un système de coordonnées cartésiennes (Litvin, 1990).

### I. 2.2 Le tétraèdre

Le tétraèdre est le symbole du Feu chez les grecs. Selon Platon, le Feu est pointu comme un poignard rappelant ainsi le tétraèdre. Ce volume est formé de quatre faces en forme de triangles équilatéraux égaux. Il est le seul solide de Platon à avoir un groupe de symétrie propre de 23 (Tableau 1). Le tétraèdre est donc le dual de lui-même. Il n'est inscriptible que dans un autre tétraèdre. Son groupe de symétrie implique un ordre 12 donc 1/12<sup>ème</sup> du volume est nécessaire et suffisant pour reconstruire le volume entier à l'aide de ses symétries propres. En effet, on observe quatre axes de rotation d'ordre 3 au sein d'un tétraèdre. Chaque axe 3 passe par un sommet et le centre de la face opposée, en accord avec les valeurs de n et z du tableau 1. Chacun des trois axes d'ordre 2 passe par le milieu de deux arêtes opposées. Ce solide est le modèle utilisé pour interpréter la géométrie de la forme à 12 sous-unités de la protéase TET1 de *P.horikoshii* en Annexe 1 de ma thèse. En effet, l'assemblage de la forme à 12 sous-unités de la protéine se réalise selon les mêmes symétries que celle d'un tétraèdre (cf. Partie 3).

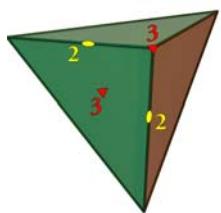


Figure 7 : Le tétraèdre

Le tétraèdre est représenté en perspective afin d'apprécier le volume en trois dimensions. L'axe d'ordre 3 présent à chaque sommet et au centre de chaque face est coloré en rouge, les axes 2 au milieu de chaque arête sont en jaune.

### I. 2.3 L'octaèdre

L'octaèdre est le symbole de l'Air dans la philosophie de Platon car les constituants de l'air sont si minuscules et si doux qu'on peut à peine les sentir. Ce solide, composé de 8 faces triangulaires, correspond à l'assemblage, par leur base, de deux pyramides à base carrée. Son groupe de symétrie propre est 432 comme le cube. Il est donc le solide dual du cube. L'unité minimum pour reformer un octaèdre par symétrie correspond donc au 1/24<sup>ème</sup> du volume total. On observe dans un

octaèdre un total de quatre axes d'ordre 3. Chaque axe coupe deux faces opposées en leur centre. Au niveau de chaque sommet, un axe d'ordre 4 traverse deux sommets opposés pour un total de trois axes d'ordre 4, en accord avec les valeurs de  $z$  et  $n$  de l'octaèdre (Tableau 1).

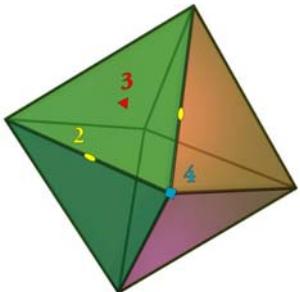


Figure 8 : L'octaèdre

L'octaèdre est représenté en perspective afin d'apprécier le volume en trois dimensions. L'axe d'ordre 3 présent au centre de chaque face est noté en rouge, les axes 2 au milieu de chaque arête sont en jaune et les axes d'ordre 4 sont représentés en bleu.

La forme à 24 sous unités ou 24-mer de la protéase TET1 de *P.horikoshii* possède le même groupe de symétrie propre que l'octaèdre. Ce solide nous a donc permis de définir les symétries présentes dans la protéine ainsi que de déterminer un modèle quasi-atomique complet de la protéine assemblée sous la forme d'un 24-mer (cf. Annexe 1).

#### I. 2.4 L'icosaèdre

Platon explique dans *Timée* que l'Eau s'échappe de la main comme si elle était constituée de petites boules faisant penser fortement à l'icosaèdre. Il attribua donc le symbole de l'Eau à l'icosaèdre. Ce solide est composé de 20 faces formées de triangles équilatéraux. Son groupe de symétrie propre est 532 (Tableau 1). Ce groupe possède un ordre de symétrie égal à 60, le plus haut. Seul ce solide permet d'assembler le plus grand nombre possible d'éléments équivalents sur une sphère. Il est composé de 6 axes d'ordre 5, 10 axes d'ordre 3 et 15 axes d'ordre 2. Le dodécaèdre, de même symétrie propre, est le dual de l'icosaèdre. C'est le seul solide qui possède une configuration d'axes de symétrie compatible avec l'icosaèdre (Tableau 1).

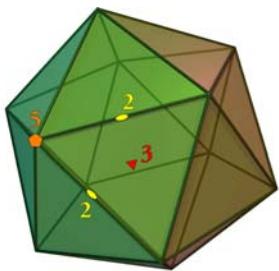


Figure 9 : L'icosaèdre

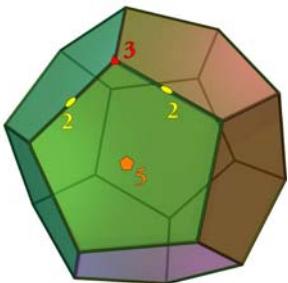
Un icosaèdre est représenté en transparent pour apprécier le volume en 3 dimensions. Un exemple des axes de symétrie qui le composent est noté en couleurs. L'axe d'ordre 2 est représenté en jaune, l'axe d'ordre 3 en rouge et l'axe 5 en orange.

Par symétrie 1/60<sup>ème</sup> du volume de départ est suffisant pour reconstituer le volume entier puisque chaque axe d'ordre 5 traverse 2 sommets opposés, les axes d'ordre 3 traversent perpendiculairement deux faces opposées en leur centre et enfin les axes d'ordre 2 se situent au milieu de chaque arête. L'icosaèdre est le plus complexe des solides de Platon. Malgré son apparente

complexité, il reste inscriptible dans un cube ce qui permet de décrire mathématiquement la symétrie icosaédrique par un jeu de 60 matrices. Il est utilisé comme modèle pour décrire la symétrie et la structure des virus icosaédriques en général, et en particulier celle de la capsid des Adénovirus (cf. I.3 et PARTIE 2).

### I. 2.5 Le dodécaèdre

Le dodécaèdre est le dernier des solides de Platon à avoir été découvert car ses faces ne sont ni de triangles ni des carrés mais des pentagones réguliers. Il représente l'Univers dans la philosophie de Platon et fût associé au cinquième élément appelé « éther » ou « esprit » en opposition avec les quatre autres éléments terrestres. L'éther est à l'origine un dieu primordial grec, personnifiant les parties supérieures du ciel.



**Figure 10 : Le Dodécaèdre**

Un dodécaèdre est représenté en transparent pour apprécier le volume en 3 dimensions. Un axe de rotation d'ordre 2 est présent au milieu de chaque arête, un axe d'ordre 3 au niveau de chaque sommet et un axe d'ordre 5 au milieu de chaque face.

Le dodécaèdre possède 12 faces pentagonales ainsi que 20 sommets. Il possède le même groupe de symétrie propre que l'icosaèdre qui est son solide dual (532) (Tableau 1). Il est donc inscriptible dans un icosaèdre mais aussi dans un cube. L'unité minimum nécessaire pour reconstituer le volume complet est, comme son dual,  $1/60^{\text{ème}}$  du volume total. Ce solide est composé de 6 axes de rotation d'ordre 5 qui passent par le centre de deux faces opposées. On observe 10 axes d'ordre 3 qui passent par 2 sommets opposés ainsi que 15 axes d'ordre 2 traversant le milieu de 2 arêtes opposées. La description des axes de symétrie du dodécaèdre est très similaire à celles de l'icosaèdre et constitue un jeu de 60 matrices.

Ce solide a servi de modèle pour l'étude de la structure 3D d'une sous particule virale dérivée de l'adénovirus (PARTIE 2). Cette particule est appelée dodécaèdre, en accord avec son modèle platonicien.

## I.3 La Symétrie dans les macromolécules biologiques et les virus

Le monde biologique, en termes de structure des macromolécules biologiques et en particulier des virus, offre une immense diversité de formes et de taille. La virosphère, éditée par le comité international de taxinomie des virus, montre bien la diversité de structures qu'il peut exister

au sein du monde viral (<http://www.ncbi.nlm.nih.gov/ICTVdb/Images/virosphere.htm>). Les protéines de capsidé, en présence ou non du génome viral, sont capables de s'assembler pour former des anneaux (Albertini et al., 2006) ou des structures hélicoïdales (Caspar & Klug, 1962) plus ou moins ordonnées comme beaucoup de nucléocapsides de virus à ARN négatif (Schoehn et al., 2004). La capsidé de certains virus peut aussi avoir la forme d'un icosaèdre. En effet, une grande partie des virus à ADN et ARN double brin possède une capsidé à symétrie icosaédrique. C'est le cas par exemple des Herpèsvirus et des Adénovirus (Baker et al., 1999). D'autres virus comme certains bactériophages montrent une combinaison de structures hélicoïdales et icosaédriques. En effet, le bactériophage T4 possède une capsidé à symétrie icosaédrique à laquelle est rattachée une queue à symétrie hélicoïdale (Leiman et al., 2004).

Dans la suite de mon exposé, je vais limiter ma description aux capsides virales à symétrie icosaédrique. Crick et Watson en 1956 montrèrent qu'une capsidé icosaédrique formée à partir d'un seul type de chaîne polypeptidique contient nécessairement 60 copies de cette chaîne, chacune réalisant des contacts équivalents en tout point de l'icosaèdre. Mais de nombreuses observations expérimentales, notamment par ME, ont montré que les virus contenaient en réalité beaucoup plus de 60 unités équivalentes. Caspar et Klug en 1962 ont donc proposé une explication à ce phénomène. Un icosaèdre peut accommoder un plus grand nombre de sous-unités si les contacts entre celles-ci ne sont pas tous identiques (Caspar & Klug, 1962). Les deux scientifiques ont ainsi introduit pour la 1<sup>ère</sup> fois la notion de « quasi-équivalence ». Cette notion est illustrée en [Figure 11](#) par la structure icosaédrique de la capsidé du TBSV et des Picornavirus.

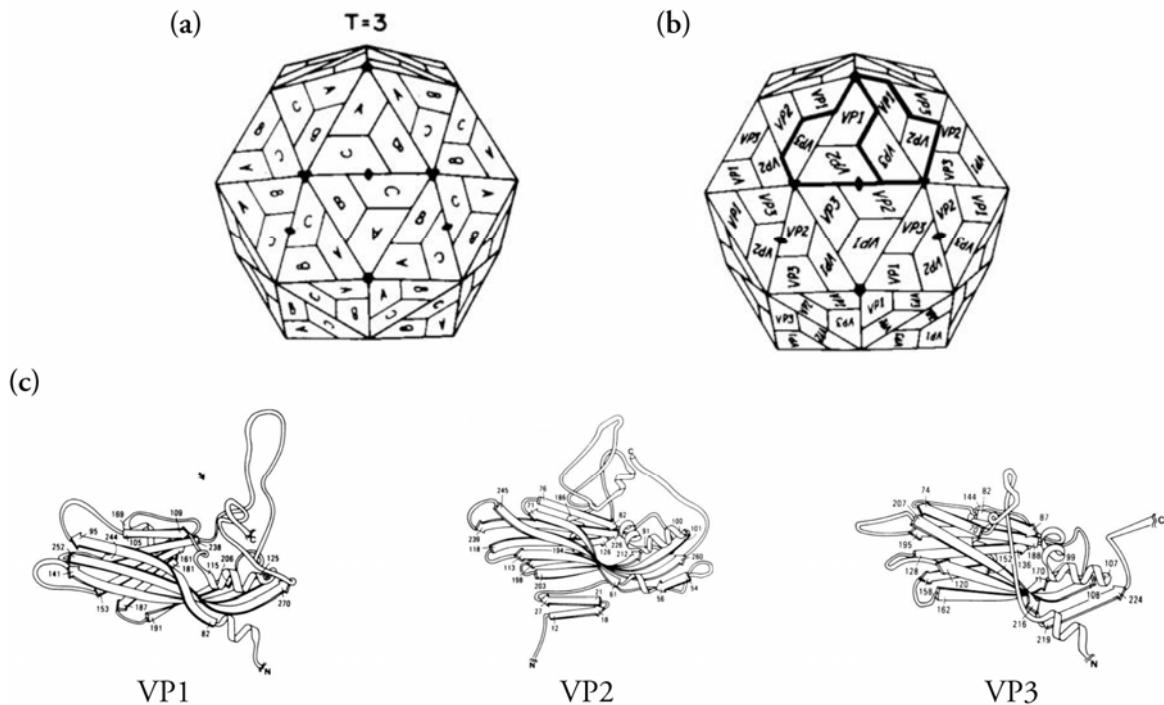


Figure 11: Exemple de quasi-équivalence chez les virus, le TBSV et les Picornavirus (Harrison, 1996; Harrison et al., 1978).

(a) Assemblage schématique de la capsidé du TBSV ( $T = 3$ ). Les capsomères ont tous la même structure mais sont présents dans trois environnements différents A, B et C. (b) Assemblage schématique de la capsidé des Picornavirus ( $T = 3$ ). Les sous unités formant l'unité asymétrique sont des protéines différentes VP1, VP2 et VP3. (c) Les trois protéines VP1, VP2 et VP3 constituant la capsidé des Picornavirus ont une structure différente.

Ils ont aussi proposé un modèle icosaédrique expliquant la structure de tous les virus. Le plus petit icosaèdre serait formé à partir d'une seule protéine qui formeraient des pentamères placés au niveau des 12 sommets de l'icosaèdre (5 chaînes x 12 sommets = 60 unités). Si l'on cherche à obtenir une capsid de plus grande taille, l'icosaèdre peut contenir des hexamères entre les différents pentamères, donnant ainsi naissance à un icosaèdre de taille de plus en plus importante. Les paramètres décrivant l'icosaèdre peuvent être énoncés mathématiquement. La grandeur mathématique qui le caractérise est le nombre de triangulation  $T$ . Cette valeur  $T$ , toujours entière, égale à  $T = h^2 + hk + k^2$ , décrit de façon simple le pavage de l'icosaèdre. Les paramètres  $h$  et  $k$  sont des nombres toujours entiers qui représentent le nombre de pas à faire au sein d'une face, dans deux directions, pour passer d'un sommet à l'autre (Figure 12). Si les pas dans la 2<sup>ème</sup> direction sont faits sur la gauche, on dit que la capsid est *levo* (*l*). Si les pas sont faits vers la droite, la capsid est *dextro* (*d*) (Figure 12). Le nombre de triangulation donne aussi le nombre théorique de chaînes polypeptidiques nécessaires pour former l'icosaèdre (= 60T) ainsi que le nombre d'environnements différents dans lesquels peuvent se trouver les capsomères.

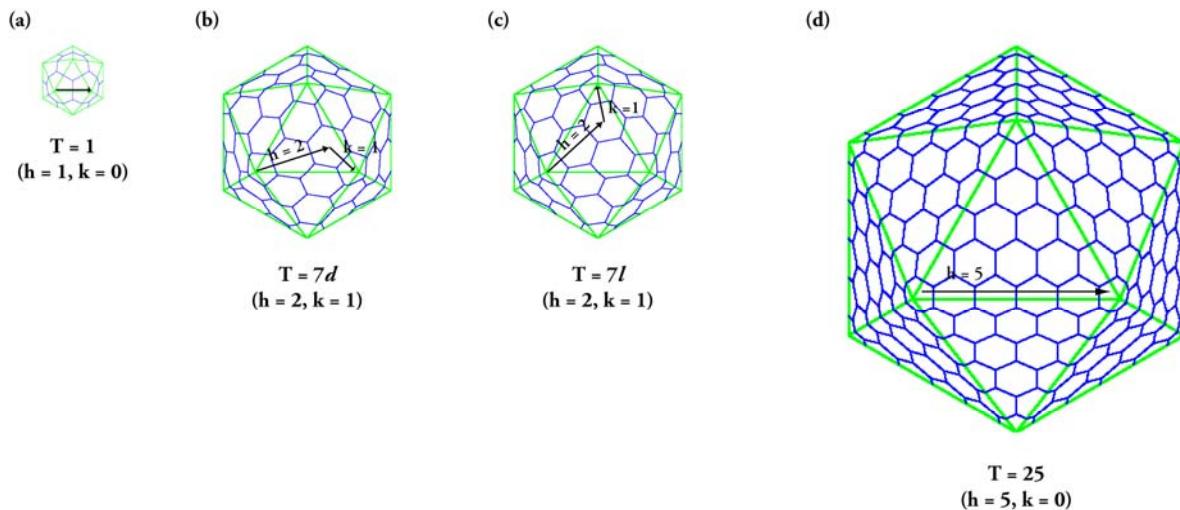


Figure 12 : Le pavage en hexamères d'un icosaèdre en fonction de son nombre de triangulation

(Source: The icosahedral server, [http://viperdb.scripps.edu/icos\\_server.php?icspage=icos\\_gal](http://viperdb.scripps.edu/icos_server.php?icspage=icos_gal))

(a) Dans un icosaèdre  $T = 1$ , il ne faut qu'un pas pour passer d'un axe 5 à l'autre. (b) Dans un icosaèdre  $T = 7d$ , il faut faire 2 pas dans une direction ( $h = 2$ ) puis 1 pas vers la droite pour passer d'un axe 5 à l'autre. (c) Dans un icosaèdre  $T = 7l$ , le pas dans la 2<sup>ème</sup> direction se réalise à gauche. (d) Dans un icosaèdre  $T = 25$ , il faut faire 5 pas selon une direction pour passer d'un axe 5 à l'autre.

La capsid d'un virus icosaédrique est aussi caractérisée par le paramètre  $S$  où  $S = 10 \times T + 2$ . Cette valeur représente le nombre de capsomères, pentamères et/ou hexamères, que l'on peut accoler par symétrie pour former une surface proche de la sphère. Certaines capsides virales suivent strictement la règle établie par Caspar et Klug en 1962 comme la capsid du TBSV (Baker et al., 1999). En effet, cette capsid possède un nombre de triangulation égal à 3 et les capsomères, tous identiques sont retrouvés dans trois environnements distincts.

Quantité d'autres virus brisent la règle de la quasi-équivalence établie par Caspar et Klug pour plusieurs raisons. Tout d'abord, chez certains virus, le nombre d'environnements différents est plus faible que le nombre de triangulation de la capsid. En effet, la capsid des papillomavirus possède une symétrie icosaédrique  $T = 7d$  mais contient seulement six environnements différents (Baker et al., 1999). Ensuite, la règle veut que la capsid soit formée à partir de pentamères et d'hexamères de la même protéine. Or la capsid des Adénovirus et du bactériophage PRD1 est composée de plusieurs protéines majeures par opposition au modèle établi. Dans ce cas, on dit que la capsid possède un pseudo nombre de triangulation  $T$ , égal à *pseudo*  $T = 25$  pour ces deux virus. Un icosaèdre  $T = 25$  contient  $60T = 1500$  chaînes polypeptidiques équivalentes. En accord avec le paramètre  $S$ , la capsid est formée de 252 capsomères ( $S = 10 \times T + 2 = 252$ ) correspondants à 12 pentamères, un à chaque sommet, ainsi que 240 trimères qui forment les faces de l'icosaèdre constituant ainsi un total de 780 chaînes protéiques majeures. Or l'adénovirus et le phage PRD1 ne suivent pas cette règle puisque les capsomères sont des trimères possédant une base hexagonale. Cette

particularité reproduit un pavage hexamèrique de la capsidé en créant un pseudo axe d'ordre 6 au sein des différents capsomères. Dans le cas de l'Adénovirus ou du bactériophage PRD1, on peut former la capsidé entière avec un nombre de chaînes plus faible par rapport au modèle établi par Caspar et Klug. Un autre exemple constitue celui de la capsidé des Polyomavirus comme le virus SV40. En effet, cette capsidé possède un nombre de triangulation  $T = 7d$  mais le pavage est réalisé par des pentamères uniquement. De plus la capsidé comme pour les papillomavirus ne contient que 6 environnements distincts. Un dernier exemple est celui du virus IBDV ( $T = 13l$ ) dont la structure a été résolue par cristallographie aux rayons X (Coulibaly et al., 2005). La structure atomique a montré que les capsomères sont en réalité des trimères qui forment des hexamères non quasi-équivalents suivant leur position dans la capsidé.

## II. La combinaison des données issues de la Microscopie Electronique et de la cristallographie aux rayons X, une méthode très adaptée à l'étude structurale de gros complexes macromoléculaires

Dans cette partie, je vais présenter les méthodes biophysiques et informatiques qui m'ont permis déterminer la structure 3D des particules biologiques que j'ai étudiées au cours de mes projets de recherche. La principale méthode porte le nom anglais de « fitting ». Ce mot n'a pas vraiment de traduction en français, c'est pourquoi je vais continuer à utiliser le terme « fitting » dans mon manuscrit. Le verbe anglais « to fit » signifie ajuster, faire correspondre. Cette méthode consiste, pour un complexe protéique, à faire correspondre une reconstruction 3D du complexe obtenue en ME avec les structures atomiques obtenues en cristallographie aux rayons X ou RMN des protéines qui composent le complexe. La première partie du travail consiste à obtenir une reconstruction 3D en ME. Je vais donc décrire le principe de la ME et du traitement informatique des images puis aborder brièvement la cristallographie aux rayons X pour finir par une description du principe du « fitting ».

### II.1 La Microscopie Electronique

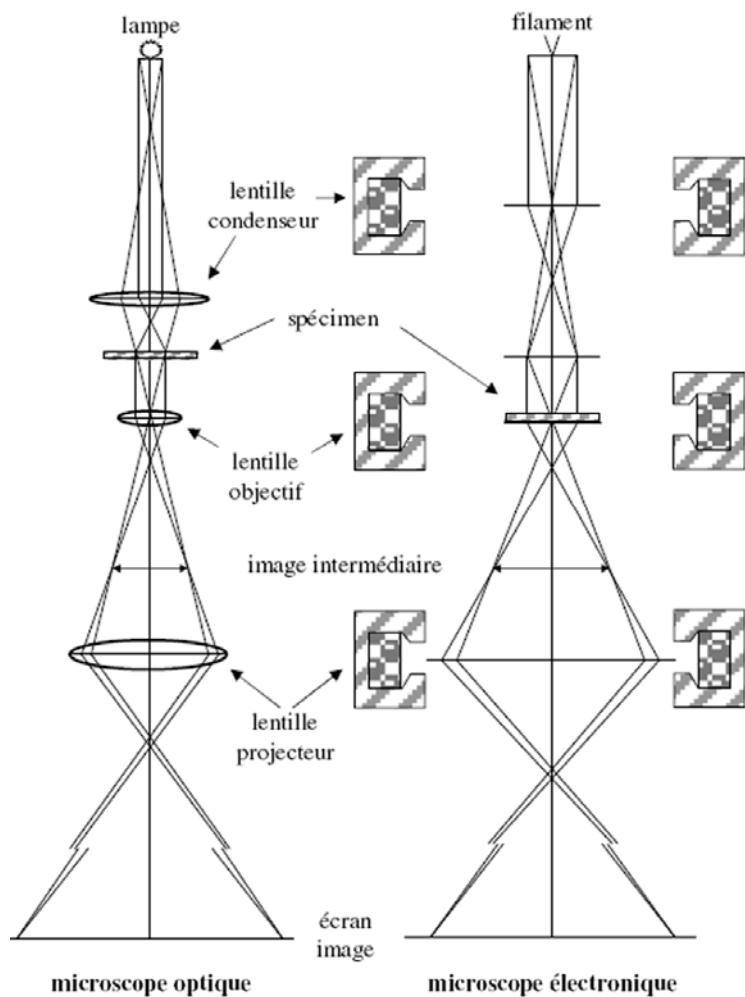
Les travaux de grands physiciens du début du 20<sup>ème</sup> siècle comme H. Poincaré, L. De Broglie ou H. Busch sur les électrons et les lentilles électromagnétiques furent les prémisses de ce qui devint au début des années 30 le microscope électronique. Max Knoll et son étudiant Ernst Ruska présentèrent en 1931 des images prises avec le 1<sup>er</sup> microscope électronique à deux lentilles. Les images montraient déjà des détails de l'ordre d'une dizaine de nanomètres. Ernst Ruska recevra le prix Nobel de physique en 1986 pour son invention. Par la suite, le développement de la ME se poursuivit, particulièrement après la seconde guerre mondiale.

En biologie, trois principaux types de microscope électronique sont couramment utilisés : le microscope électronique en transmission ou MET, le microscope électronique à balayage ou MEB et le microscope électronique à balayage en transmission ou MEBT. Le plus utilisé en biologie est le MET. Ce microscope génère un faisceau fixe d'électrons et permet de réaliser des études structurales à haute résolution sur des virus ou des complexes protéiques isolés. Il permet aussi d'étudier des coupes fines de cellules. Dans le cas du MEB, le faisceau d'électrons est une sonde mobile qui permet de balayer la surface d'un échantillon afin d'en obtenir une image ou de réaliser des mesures quantitatives sur la composition chimique de l'échantillon. Le MEBT quant à lui est une combinaison des deux premiers types de ME. Il permet d'obtenir une image classique en

transmission ainsi qu'une détermination précise de la masse de l'objet observé sous le faisceau d'électrons. Je vais focaliser mon exposé sur le MET. Il est possible, grâce à ce type de microscope et après traitement informatique des images, d'obtenir un modèle 3D d'objets biologiques comme la capsid de l'adénovirus ou la structure de la protéine TET1.

### *II. 1.1 Le principe de la Microscopie Electronique en transmission*

Le principe de fonctionnement d'un MET s'apparente à celui d'un microscope optique. Les électrons remplacent les photons. Le rayonnement traverse l'échantillon et une image se forme sur un écran phosphorescent au lieu de se former sur la rétine de l'observateur (Figure 13). Puisque le fonctionnement de ces deux microscopes est très similaire, pourquoi utilise t-on les électrons comme source de lumière à la place des photons ?



*Figure 13 : Principe comparé du microscope optique et du microscope électronique.*

Les deux microscopes sont en mode image. Dans le microscope électronique, les lentilles sont schématisées par les parties grises autour du faisceau.

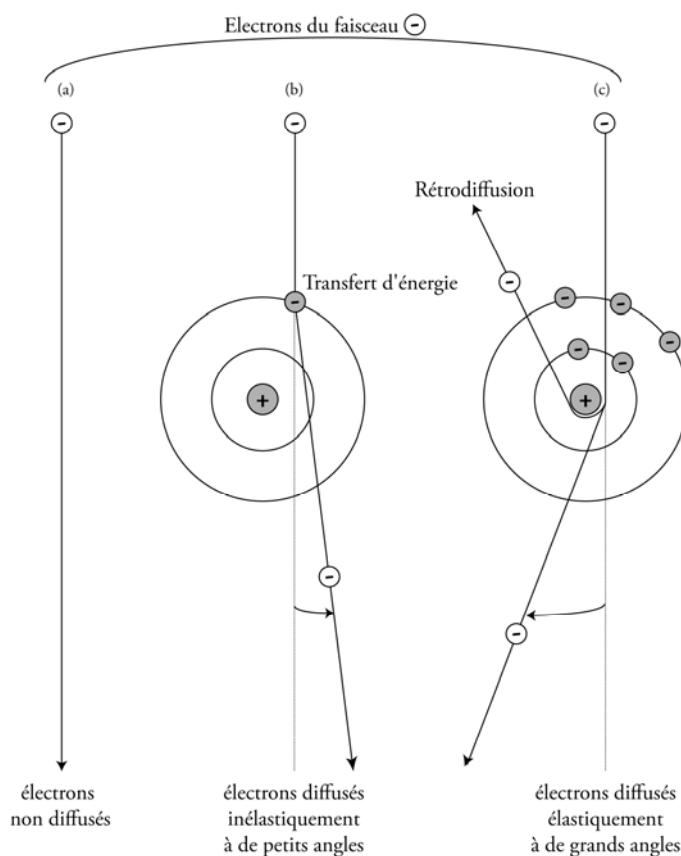
Le caractère ondulatoire de la lumière, que ce soit des photons ou des électrons, impose une limitation sur la taille des objets que l'on peut observer en microscopie. Ernst Abbe, physicien et industriel allemand, montra à la fin du 19<sup>ème</sup> siècle, que la distance entre deux points observés distinctement dans un microscope ne peut en théorie pas être inférieure à la moitié de la longueur d'onde utilisée pour l'observer. Ce critère théorique, confirmé par Rayleigh quelques années plus tard, pose la limite de l'utilisation des photons en microscopie. En effet, selon le critère d'Abbe, à une longueur d'onde de 500 nm, on ne peut pas observer des détails d'une taille inférieure à 250 nm. Si on veut observer des détails plus fins au sein d'un objet, il faut donc recourir à l'utilisation d'une autre onde électromagnétique possédant une longueur d'onde plus faible que les photons. Or les électrons possèdent une longueur d'onde  $\lambda$  inférieure à l'Angstrom, proportionnelle à la tension d'accélération  $V$  qui leur est appliquée conformément à la formule  $\lambda = \frac{1.23}{\sqrt{V + 10^{-6}V^2}}$  nm. Dans un MET couramment utilisé en biologie, les tensions d'accélération varient entre 100 et 300 kV ce qui produit des électrons possédant une longueur d'onde comprise entre  $3,7 \times 10^{-3}$  nm et  $1,97 \times 10^{-3}$  nm. Par exemple, pour une tension d'accélération de 100 kV, le plus petit détail théorique que l'on pourrait observer est de l'ordre de  $2 \times 10^{-3}$  nm.

Par analogie avec le microscope optique, Le MET est constitué d'un système de lentilles dans ce cas électromagnétiques qui permettent de condenser et focaliser les électrons émis par la source. La source peut être un simple filament de tungstène ou un cristal d'hexaborure de lanthane chauffé à très haute température (respectivement 3000°C et 1500°C). Plus récemment, des sources à émission de champ ont été développées. Ce type de source consiste à soumettre une pointe de tungstène (recouverte d'oxyde de zirconium) à une forte différence de potentiel électrique ce qui entraîne l'émission d'électrons par effet tunnel et délivre un faisceau d'électrons de meilleure qualité, plus brillant et plus cohérent que pour les deux premières sources.

Par la suite, le faisceau produit par la source est accéléré et se propage dans la colonne du MET pour atteindre l'échantillon présent au milieu de la lentille objectif ([Figure 13](#)). Le libre parcours des électrons dans l'air étant très faible, un vide poussé, d'au moins  $10^{-9}$  atmosphère, est maintenu au sein de la colonne. Le faisceau va tout d'abord être condensé par la lentille « condenseur » puis focalisé par la lentille « objectif » pour enfin être projeté sur un écran ou un film photographique par la lentille « projecteur » ([Figure 13](#)). L'image de l'échantillon peut être observée sur l'écran phosphorescent ou directement sur un film photographique après développement.

## II. 1.2 Le microscope électronique produit une image déformée de l'objet

Par analogie avec le microscope optique, c'est l'interaction de la lumière de la source avec l'échantillon qui est à l'origine de l'image. Dans un MET, les électrons du faisceau interagissent de deux façons différentes avec le spécimen et ces interactions sont à l'origine de la théorie de la formation de l'image. Dans un MET, l'image est une projection 2D d'un objet en 3D qui est l'échantillon. L'image obtenue n'est pas parfaite. Un MET est un système optique qui possède de nombreux défauts ce qui entraîne des déformations importantes dans l'image. Le contraste dans l'image ainsi qu'une partie de ces distorsions sont modélisés mathématiquement par la Fonction de Transfert de Contraste ou CTF (Contrast Transfer Function).



**Figure 14: Comportement des électrons du faisceau à la traversée d'un échantillon.**

(a) L'électron est transmis sans interagir avec l'échantillon. (b) Diffusion inélastique : l'électron du faisceau transfert de l'énergie au spécimen par excitation d'un électron d'un atome. L'électron incident subit une faible déviation liée à sa perte d'énergie. (c) Diffusion Elastique : l'électron du faisceau au voisinage d'un noyau est dévié sans perte d'énergie. Dans ce cas la déviation qu'il subit est forte.

Les électrons du faisceau sont représentés en blanc, ceux de l'échantillon en gris. Le noyau des atomes est représenté par un symbole + coloré en gris.

La formation de l'image : L'épaisseur de l'échantillon étant faible ( $e < 100$  nm), les électrons du faisceau le traversent sans être absorbé et statistiquement n'interagissent qu'une seule fois avec l'échantillon. Les électrons du faisceau vont essentiellement subir deux types de phénomènes de diffusion. Le premier cas est celui de la diffusion inélastique quand un électron du faisceau percute un électron de l'échantillon (Figure 14b). Dans ce cas, l'électron du faisceau donne une partie de son énergie. Sa trajectoire va être déviée à de petits angles (Figure 14b) et par perte d'énergie, sa longueur

d'onde va augmenter et sa vitesse réduire. L'électron de l'échantillon est excité. Si l'énergie qu'il reçoit est suffisante, il est éjecté de l'atome. Sinon l'électron subit une transition électronique puis retourne à son état fondamental par émission de rayons X. Ce transfert d'énergie est à l'origine de la destruction très rapide de l'échantillon soumis à un faisceau d'électrons (quelques secondes). Le deuxième phénomène de diffusion est appelé diffusion élastique. Lorsqu'un électron incident passe près du noyau d'un atome, celui-ci est attiré par la charge positive portée par le noyau (Figure 14c). La trajectoire de l'électron incident va ainsi être déviée à des grands angles. La déviation est proportionnelle au numéro atomique de l'atome qu'il rencontre. Plus l'atome est lourd, plus la déviation est forte. Il n'y a pas de transfert d'énergie au spécimen, ni de changement de longueur d'onde et de vitesse de l'électron incident. Dans certains cas, la déviation est tellement importante que l'électron incident peut être rétrodiffusé (Figure 14c).

L'interaction des électrons avec la matière forme une image mais on peut distinguer des détails dans l'image uniquement si celle-ci est contrastée. Le contraste dans une image de microscopie quelle soit optique ou électronique est de deux ordres : soit d'amplitude soit de phase. Dans le cas du MET, le contraste est toujours lié aux électrons diffusés élastiquement par l'échantillon. Le contraste d'amplitude est lié à l'élimination des électrons fortement déviés soit par les parois du microscope soit par le diaphragme d'objectif. La présence du diaphragme nous permet d'augmenter le contraste. Ceci entraîne la visualisation sur l'image de zones sombres (absence d'électrons) pour les parties de l'échantillon déviant fortement les électrons et des zones claires au niveau des zones déviant faiblement les électrons. Le marquage de l'échantillon par des sels de métaux lourds (technique de la coloration négative) exploite cette propriété pour augmenter le contraste dans l'image. Le contraste de phase quant à lui est lié à des phénomènes d'interférences entre les électrons diffusés entre eux et avec les électrons transmis. Ces interférences, constructives ou destructives entraînent des différences de phase à la sortie de l'échantillon. Ce contraste varie fortement avec la valeur de défocalisation de la lentille objectif. Dans une image réalisée en MET, le contraste de phase occupe une part plus importante que celui d'amplitude surtout pour les objets biologiques observés sans coloration.

Les aberrations du ME : L'image d'un objet obtenue en ME n'est pas parfaite. Elle subit de nombreuses déformations liées à la géométrie du ME en lui-même, à l'imperfection des lentilles utilisées, aux réglages réalisés par l'observateur ainsi qu'à la stabilité du microscope. Une partie de ces déformations est exprimée dans la CTF.

Le premier facteur limitant fortement la résolution d'un MET est l'aberration sphérique. Les rayons passant près du centre de la lentille ne sont pas focalisés au même endroit que ceux qui passent plus à l'extérieur. Tout point dans l'échantillon sera donc représenté par un disque appelé disque de moindre confusion. Cette aberration est quantifiée par une valeur d'aberration sphérique  $C_s$  exprimée en mm. Plus celle-ci est faible plus la résolution obtenue dans le MET est bonne. Ce paramètre majeur de limitation de la résolution est une composante de la CTF.

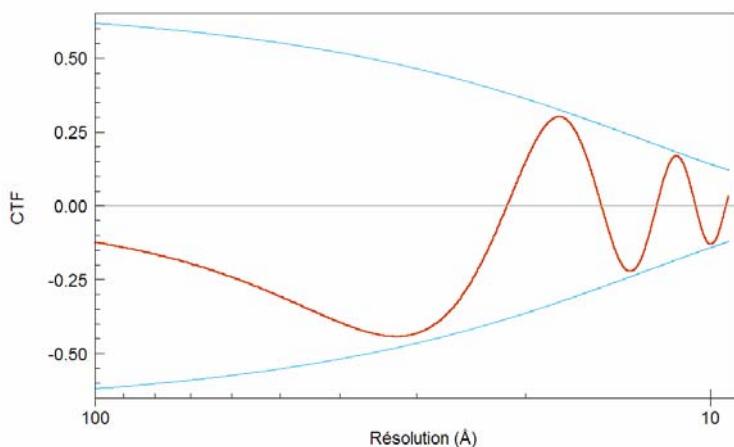
Le deuxième type de déformations est l'aberration chromatique. Suivant la cohérence de la source utilisée, les électrons du faisceau ne possèdent pas tous la même longueur d'onde et ne vont pas être tous focalisés au même endroit par la lentille. Cette distorsion n'est pas la principale cause de détérioration de la résolution. Son effet est non négligeable lorsque le spécimen observé possède une épaisseur importante.

D'autres phénomènes de distorsion s'ajoutent aux aberrations causées par les lentilles. L'image peut aussi contenir de l'astigmatisme et/ou de la dérive. Ces deux défauts sont accentués en présence de contamination magnétique qui vient perturber le fonctionnement des lentilles. L'astigmatisme ou aberration d'ellipticité, tout d'abord est causé par l'inhomogénéité du champ magnétique créé par rapport à l'axe optique. Cette asymétrie entraîne une variation de la position du plan focal suivant la direction des électrons incidents. Un point dans l'objet sera représenté par une ellipse dans l'image en présence d'astigmatisme. La correction de ce défaut est possible grâce à un correcteur d'astigmatisme au niveau du condenseur et de l'objectif. La correction de l'astigmatisme du condenseur est un réglage simple par opposition à celui de l'objectif. Malgré les possibilités de corrections, l'aberration d'ellipticité reste un facteur de limitation de la résolution dans un MET. Enfin l'image peut aussi contenir de la dérive. Ce défaut est perçu comme un effet de « flou » dans l'image et est causé par l'instabilité du MET et du porte-échantillon. En effet, l'enregistrement d'une image en MET nécessite un temps de pose de l'ordre de la seconde, des déplacements de quelques nm entraînent une détérioration importante de la qualité de l'image.

CTF: La CTF est une fonction qui décrit la variation du contraste d'amplitude et de phase dans une image de MET en fonction des fréquences spatiales  $v$  de l'image (Figure 15). La composante élastique de la CTF s'exprime mathématiquement par  $C(v) = A \cdot \sin[-\gamma(v)] - (1 - A) \cdot \cos[-\gamma(v)]$  où  $\gamma$  représente le déphasage,  $A$  correspond à la proportion du signal dû au contraste de phase par rapport au contraste d'amplitude. Le déphasage dépend de la fréquence spatiale  $v$  dans l'image et s'exprime mathématiquement de façon simple si on

$$\text{considère que l'image ne contient pas d'astigmatisme } \gamma(\nu) = \frac{2\pi}{\lambda} \left[ \frac{C_s \nu^4 \lambda^4}{4} - \frac{\Delta f \nu^2 \lambda^2}{2} \right] \text{ où } \Delta f$$

représente la valeur de la défocalisation,  $C_s$  la constante d'aberration sphérique du microscope électronique utilisé (Conway & Steven, 1999). Cette fonction est une sinusoïde amortie vers les hautes fréquences dont les paramètres dépendent des caractéristiques du microscope ( $C_s$ ), de la nature de la source, de la tension d'accélération utilisée ( $\lambda$ ) ainsi que de la défocalisation ( $\Delta f$ ). Au passage en zéro de la CTF, l'image ne contient pas d'informations (Figure 15). Puis l'amplitude de la CTF devient négative ce qui entraîne une inversion du contraste. Le caractère sinusoïdal de la CTF a un impact important au niveau de la qualité de l'image. Dans une gamme de fréquences où la CTF a une valeur négative, les zones de l'image visualisées plus claire quand le contraste est positif, sont maintenant visualisées foncées. Des basses fréquences au premier zéro de la CTF, l'information ne subit pas de modifications liées à la CTF. Au delà du premier zéro, l'image est altérée. Si on veut réaliser une reconstruction 3D à haute résolution, il est nécessaire de corriger les images pour les effets de la CTF. De plus, chaque image enregistrée en MET possède une CTF propre puisque que chaque image possède une valeur de défocalisation propre. Cette fonction permet aussi d'évaluer la qualité de l'image avant le traitement des données, notamment en jugeant de la résolution spatiale présente dans l'image.



*Figure 15: Représentation théorique de la CTF.*

La variation du contraste de phase est représentée en fonction de la résolution de l'image. L'information est atténue exponentiellement vers les hautes résolutions.

### II. 1.3 La préparation des échantillons biologiques et la prise d'images en MET

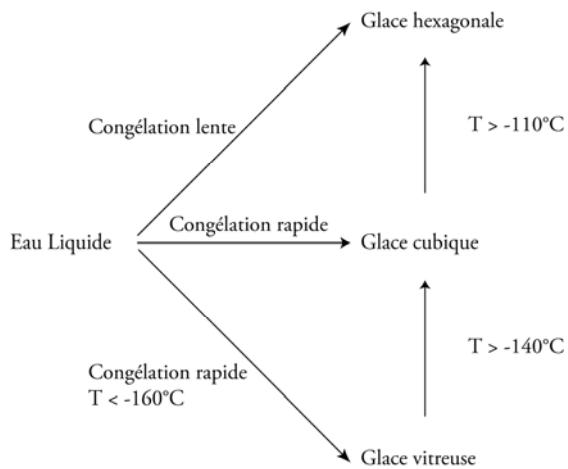
Les électrons constituent un rayonnement très destructeur pour les échantillons biologiques. La dose absorbée par l'échantillon soumis à un faisceau d'électrons correspond à celle qu'il pourrait absorber au centre d'un champignon atomique lors de l'explosion d'une bombe nucléaire. De plus, au sein de la colonne d'un microscope électronique règne un vide poussé. Si on veut observer un échantillon biologique en MET, celui-ci doit être préparé afin d'en préserver la structure. Divers

méthodes de préparation des échantillons biologiques ont été mises au point à partir de la 2<sup>ème</sup> guerre mondiale. Je ne vais détailler que les méthodes que j'ai utilisées pour étudier mes échantillons. Ces deux méthodes sont dominantes dans le monde de MET des échantillons biologiques. L'échantillon est toujours observé après dépôt sur une grille de cuivre. La prise d'images en MET se réalise sur des plans film ou à l'aide d'une caméra CCD en mode faible dose. En effet, pour obtenir une image de qualité en ME, il ne faut pas irradier la zone que l'on désire photographier. Le principe est toujours de réaliser les réglages du ME dans une zone adjacente de la zone à photographier afin que celle-ci ne subisse pas de dommages dus aux radiations. Le problème de la dose reçue par l'échantillon est récurrent en ME. Ainsi toutes les méthodes mises au point ont pour but de la minimiser.

*La coloration négative.* Hall en 1955 fût le premier à observer qu'une solution de sels de métaux lourds entraîne un effet de coloration négative. L'effet de coloration négative est caractérisé par l'observation de zones sombres au niveau des sels d'atomes lourds et des zones claires au niveau de l'échantillon ce qui produit un contraste à l'inverse du contraste classique où l'échantillon est visualisé plus sombre que le fond de l'image. Plus tard, Brenner et Horne montrèrent le même phénomène sur des capsides de bactériophages T2 et donnèrent le nom de coloration négative à cette technique (Brenner & Horne, 1959). Cette méthode est la plus simple à mettre en œuvre et permet l'observation rapide de l'échantillon avec un très fort contraste d'amplitude. Elle permet l'observation d'échantillons biologiques d'assez petite taille pouvant aller jusqu'à 50 kDa mais ne donne pas accès à la structure interne de la particule observée. Cette technique consiste à fixer l'échantillon par séchage d'une solution de sels de métaux lourds. L'échantillon va laisser une empreinte dans les sels de métaux lourds en séchant et cette empreinte va être observé dans le ME. Les métaux lourds sont des atomes possédant un noyau très chargé qui va dévier fortement les électrons du faisceau. Ces électrons vont ensuite être éliminés par le diaphragme d'objectif produisant ainsi un fort contraste d'amplitude dans l'image. Dans ce cas, l'échantillon est déposé sur une couche de carbone qui ensuite va venir recouvrir la grille en cuivre. Plusieurs solutions de colorants peuvent être utilisées comme le Silicotungstate de Sodium (SST), l'Acétate d'Uranylique (UrAc) ou le Molybdate d'Ammonium (AmMb). Le choix du colorant dépend de l'échantillon à fixer ainsi que du pH de la solution de colorant. Le SST est un colorant de choix qui produit un grain assez épais donnant un contraste important quand on observe un échantillon de petite taille (50 kDa) et peu concentré. Le molybdate d'ammonium, par contre, produit peu de contraste. Il est donc beaucoup plus utilisé pour les échantillons de grande taille comme les virus. Malgré sa rapidité d'exécution, cette méthode possède des inconvénients. Elle donne accès aux structures de surface

d'un échantillon mais pas aux structures internes comme l'intérieur d'un virus par exemple. De plus la fixation peut entraîner des artefacts structuraux. L'adsorption sur une surface, la présence de colorants et le séchage vont entraîner des déformations importantes de l'échantillon.

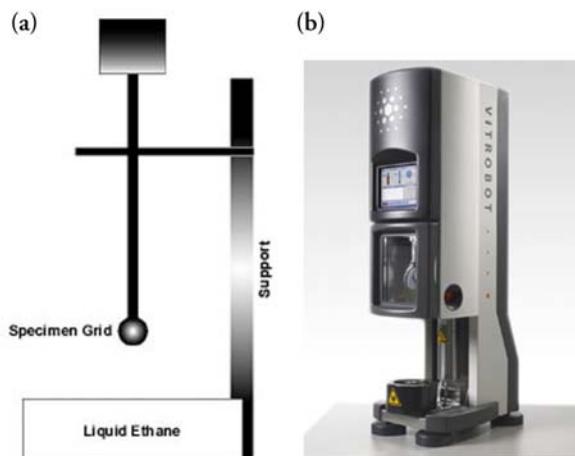
*La Méthode des échantillons « congelés-hydratés ».* Après l'utilisation de la coloration négative et d'autres méthodes de coloration que je n'ai pas détaillées, le challenge en ME fût d'observer les échantillons biologiques dans des conditions natives. Le principe est de piéger l'échantillon sans perturber les conditions de solvatation dans lequel celui-ci se trouve par cryo-fixation. La méthode des échantillons « congelés-hydratés » a été développée dans les années 80 (Adrian et al., 1984). Elle est lourde à mettre en œuvre et consiste à fixer l'échantillon dans une couche très mince de glace amorphe créée en refroidissant très rapidement la grille de cuivre portant l'échantillon. Seule la glace amorphe est transparente aux électrons. Les autres formes de glace obtenues par congélation plus lente ou à des températures plus élevées sont denses aux électrons et ne permettent pas l'observation de l'échantillon inclus (Figure 16).



**Figure 16: Diagramme des phases de la glace à  $P_{atm}$ .**

Le diagramme représente les transitions de phase entre les 3 principales formes de glace obtenues à pression atmosphérique. Il existe 3 formes prédominantes : les glaces cubiques et hexagonales sont opaques aux électrons alors que la glace vitreuse, nécessaire aux observations en MET, est translucide aux électrons. Adapté de Dubochet et al., 1988.

La cryo-fixation ne peut être réalisée que dans une solution d'éthane ou de propane liquéfiée par de l'azote liquide. Pour éviter les phénomènes de déformations de l'échantillon par adsorption, la couche de carbone possède des trous dans lesquels vient se placer l'échantillon. Celui-ci se trouve inclus dans une couche de glace amorphe au niveau des trous du carbone. La mise en œuvre de cette méthode nécessite un appareillage adapté à la congélation de l'échantillon, au maintien à température de l'échantillon (-180°C) ainsi qu'un MET refroidi à l'azote liquide, adapté à l'observation de tels échantillons. Tout d'abord, l'échantillon est congelé à l'aide d'un système de guillotine. La grille de cuivre est fixée au bout d'une pince, elle-même fixée à la guillotine, au dessus du bain d'éthane liquide (Figure 17).



**Figure 17 : Systèmes disponibles pour la congélation des grilles en cryoEM.**

(a) Schéma d'un système manuel de guillotine couramment utilisé pour la congélation des échantillons.

(b) Photo de la dernière génération du Vitrobot (FEI).

L'échantillon est déposé sur la grille puis l'excès de liquide est absorbé avec un papier filtre avant de plonger très rapidement la grille dans l'éthane liquide par abaissement du système de guillotine. Cette étape est déterminante pour la qualité de la congélation. Si l'épaisseur de glace est trop importante, la glace devient opaque aux électrons. Si elle est trop faible, l'échantillon se trouve agréger sur les bords de chaque trou car l'épaisseur de glace est trop faible au centre. Depuis 2002, des robots comme le Vitrobot (FEI) sont disponibles pour la congélation. L'échantillon est déposé sur la grille au sein d'une chambre sous atmosphère à humidité contrôlée ce qui empêche la grille de s'assécher et permet de contrôler assez finement l'épaisseur de glace au niveau de la grille. Après congélation, l'échantillon doit être maintenu à une température inférieure à -160 °C avant le transfert dans le MET afin d'éviter une transition de phase irréversible de la glace vitreuse (Figure 16). Pour cela, nous disposons d'un porte-échantillon spécial, refroidi à l'azote liquide ainsi qu'un système de transfert qui réduit la contamination de la grille par des dépôts d'eau atmosphérique. Le transfert de l'échantillon dans le microscope est réalisé le plus rapidement possible afin de limiter ces dépôts. Pour l'observation d'échantillons « congelés-hydratés », un MET adapté est nécessaire. En effet, le MET dispose d'un anti-contaminateur. Cette pièce, composée de deux lames de cuivre situées de part et d'autre de l'échantillon, est refroidie à l'azote liquide et permet de limiter la contamination de l'échantillon au sein du MET.

La cryo-fixation permet de diminuer fortement l'effet des dommages dus aux radiations. Ceux-ci sont réduits d'un facteur 2 à 5 lorsque l'échantillon est fixé dans la glace. Cette méthode, malgré sa mise en œuvre longue et délicate, permet d'observer l'échantillon dans un état très proche de sa structure native et permet de visualiser les structures internes comme l'organisation du génome au sein d'une capsidé virale (San Martin et al., 2001). Cette technique permet aussi d'analyser une

molécule dans différents environnements biochimiques ce qui permet l'observation de différents états fonctionnels. (Ranson et al., 2006).

Comme je l'ai déjà mentionné plus haut, la prise d'images en MET est réalisée en mode faible dose pour limiter les dommages dus aux radiations dans les images. Les images sont réalisées par paire focale sur des négatifs ou à l'aide d'une caméra CCD. Chaque zone est photographiée deux fois à deux valeurs différentes de défocalisation afin de compenser la perte d'informations liée à la CTF. Après développement, les négatifs sont scannés puis soumis à une méthode d'analyse d'images afin de reconstituer un modèle 3D de la particule observée dans le MET.

La numérisation des négatifs est réalisée à l'aide d'un scanner dédié à la digitalisation des négatifs provenant de la ME. En effet, celui-ci possède des spécifications particulières. Le pas utilisé pour la numérisation est très faible (jusqu'à 7  $\mu\text{m}$ ) afin de conserver les informations à haute résolution contenue dans les images. De plus, le scanner ne modifie pas ou peu la CTF des images. Les aberrations introduites par la numérisation sont les plus faibles possibles.

#### *II. 1.4 La correction, l'analyse des images et le calcul d'une reconstruction 3D*

Avec le développement rapide de l'informatique, l'analyse des images prises en MET a permis de reconstituer un modèle en 3 dimensions de la particule observée sous le faisceau d'électrons. Les techniques d'analyse d'images se sont développées à partir des années 70 avec les travaux de De Rosier et Klug (DeRosier & Klug, 1968). Selon la symétrie de la particule étudiée, les méthodes et les programmes diffèrent mais toutes s'appuient sur le principe que l'image d'un échantillon « congelé-hydraté » est la projection 2D de l'objet observé.

Le principe de reconstruction 3D des objets hélicoïdaux et icosaédriques est basé sur une approche dans l'espace de Fourier. En supposant que les projections 2D de l'objet couvrent l'ensemble des orientations possibles, par interpolation des différentes images dans l'espace de Fourier et Transformée de Fourier (TF) inverse, on peut reconstituer un modèle en 3D de l'objet de départ (DeRosier & Klug, 1968). Les premiers objets reconstruits possédaient une symétrie hélicoïdale (DeRosier & Moore, 1970). Par la suite, les méthodes ont été adaptées aux objets à symétrie icosaédrique (Crowther et al., 1970). De nombreuses équipes dans le monde ont ainsi développées des suites de programmes fonctionnant sous Unix puis Linux, permettant le traitement complet des images, de la sélection des particules jusqu'à la reconstruction 3D (Baker & Cheng, 1996; Frank et al., 1996; Fuller, 1987; Fuller et al., 1996; van Heel et al., 1996). Les particules isolées sont un cas un peu différent. Les méthodes de reconstruction ont été développées bien après

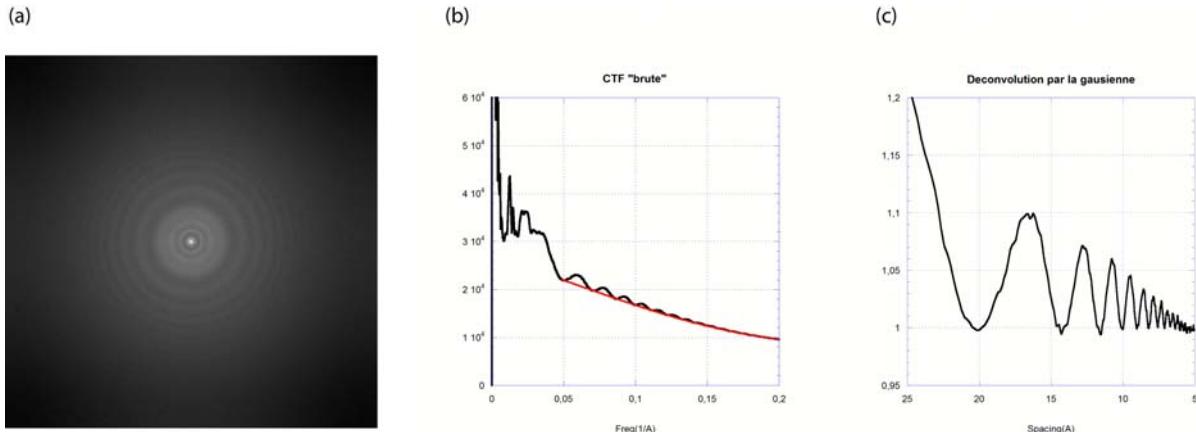
celles dédiées aux particules possédant une symétrie (Frank, 1981; Frank & van Heel, 1982; Frank et al., 1981; van Heel et al., 1996). Ces méthodes sont basées sur une reconstruction 3D dans l'espace réel, par épandage des images selon la 3<sup>ème</sup> dimension (Frank, 1981; Frank et al., 1996). Une partie du traitement informatique restant tout de même commune aux différentes méthodes, je vais décrire dans la suite de mon exposé le principe général de l'obtention d'une structure 3D en MET et les traitements informatiques communs.

Le travail d'analyse d'images commence toujours par un tri manuel des micrographies. Cette sélection est basée sur l'absence d'astigmatisme et de dérive dans l'image, aberrations qui détériorent énormément la qualité de la future reconstruction ainsi que sur la valeur de la défocalisation à laquelle l'image a été enregistrée. Seule une gamme restreinte de défocalisation est exploitable ( $1 < \Delta f < 3 \mu\text{m}$ ). En dessous de  $1 \mu\text{m}$ , le contraste est trop faible et l'orientation des particules n'est pas déterminable. Au dessus de  $3 \mu\text{m}$  de défocalisation, le premier zéro de la CTF se situe dans une zone de basses fréquences ce qui altère une trop grande partie de l'information pour obtenir une reconstruction 3D de qualité. De plus le nombre de zéros de la CTF va fortement augmenter. L'atténuation des hautes fréquences causée par la CTF sera très importante et va fortement diminuer la résolution de l'image. Les images en MET sont souvent enregistrées sur des négatifs. Il est donc nécessaire de numériser les micrographies à l'aide d'un scanner spécifique avant de procéder à un traitement. Ce scanner dispose d'un pas de numérisation très petit afin de ne pas détériorer la qualité des images (Zeiss Photoscan TD, pas de numérisation  $\geq 7 \mu\text{m}$ ). Les MET récents possèdent maintenant des caméras CCD permettant l'enregistrement numérique direct de la micrographie. Par la suite, les particules, pour chaque micrographie exploitable, doivent être sélectionnées et extraites de l'image de départ dans un nouveau fichier informatique. Ces nouveaux fichiers seront utilisés au cours du processus d'analyse d'images, sans jamais revenir à la micrographie de départ sauf au cours du travail de recentrage des particules (cf. Matériel et Méthodes). L'étape suivante, primordiale pour l'obtention d'une reconstruction 3D de qualité est la correction des aberrations liées à la CTF.

Correction de la CTF. Cette étape essentielle est à réaliser pour chaque micrographie. La méthode que nous avons utilisée pour corriger la CTF dans les images a été développé par James Conway et Alistair Steven (Conway & Steven, 1999). Elle consiste en deux traitements distincts réalisés sur les particules extraites après sélection. Mais avant de corriger les images, il faut avoir accès à la CTF de chaque image. Pour cela, le programme SUMPS permet de calculer le spectre de puissance moyen sur un jeu de particules données (Figure 18a) (Conway & Steven, 1999). Le spectre de puissance représente le carré de l'amplitude de la TF des particules. La représentation des

valeurs d'amplitude de la moyenne radiale du spectre de puissance en fonction des fréquences spatiales dans l'image permet de visualiser la CTF (Figure 18b).

La CTF est responsable de deux types de défauts qui peuvent être corrigés. Ces défauts sont visibles sur la Figure 19b. Le premier concerne les inversions de phase après chaque passage en zéro. On recherche à quelles fréquences spatiales la CTF passe en zéro pour réaliser une inversion de phase dans les zones où le contraste est inversé. Ce premier traitement permet de restaurer une partie de l'information dans l'image.



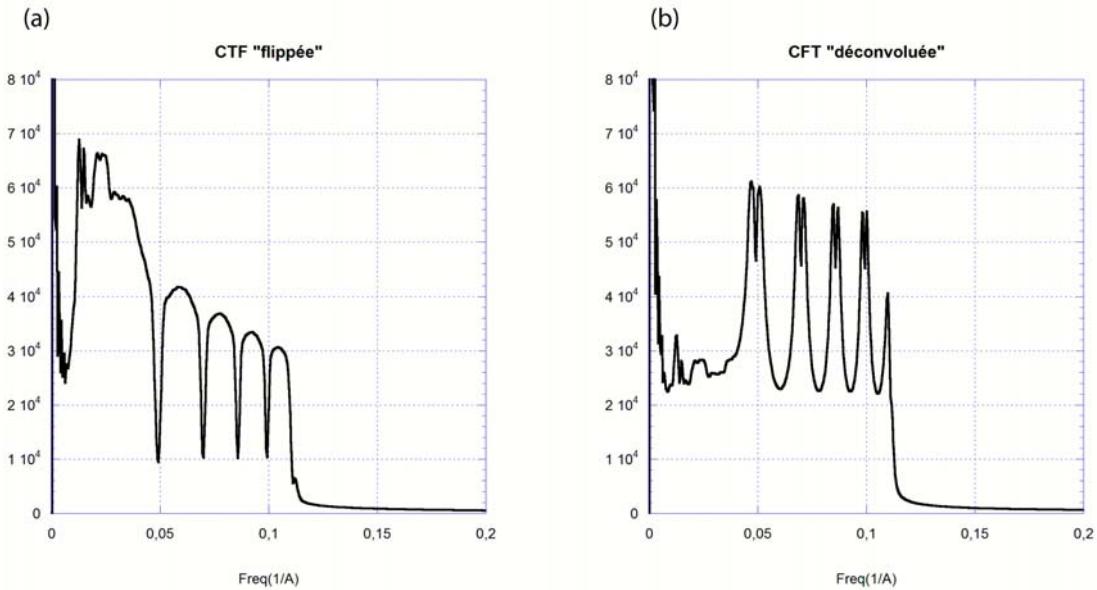
**Figure 18: Modélisation de la CTF à partir du spectre de puissance.**

(a) Spectre de puissance moyen de l'image. On remarque une alternance de cercles concentriques clairs et sombres. (b) Moyenne radiale de la CTF brute d'une image et en rouge la gaussienne calculée qui passe par les zéros de la CTF. Le premier zéro de la CTF est visible à  $0,05 = 1/20 \text{ \AA}$  (c) CTF après déconvolution par la gaussienne.

Les images traitées sont dites « flippées » (Figure 19a). Si on regarde la CTF des images « flippées », on ne remarque pas une énorme différence avec la CTF brute. En effet ce traitement agit sur la phase ce qui n'entraîne pas de modifications majeures au niveau des amplitudes de la CTF. L'atténuation des hautes fréquences est toujours présente (Figure 19a). Cet amortissement dans les hautes fréquences doit être corrigé, il s'agit là du deuxième défaut. Pour cela la CTF expérimentale propre à chaque image est modélisée comme en Figure 18b. Cette représentation est utilisée pour déterminer les paramètres  $m_1$ ,  $m_2$ ,  $m_3$  et  $m_4$  de la gaussienne responsable de l'atténuation qui se

traduit mathématiquement par :  $y = m_1 + m_2 \exp\left(\frac{-(m_0 - m_3)^2}{m_4^2}\right)$ . Cette gaussienne est déterminée au

niveau des zéros de la CTF (en rouge sur la Figure 18b) Une simple déconvolution permet ensuite de s'abolir des effets néfastes de la CTF (Figure 18c et Figure 19b). La déconvolution inclut aussi une inversion des phases comme dans le premier traitement et exclut les zéros.



**Figure 19: Modélisation des moyennes radiales de la CTF après correction.**

(a) Moyenne radiale de la CTF après inversion des phases : On ne voit que très peu de différences avec la moyenne radiale de la CTF « brute » (Figure 18b) car la correction a seulement lieu sur les phases sans toucher aux amplitudes. (b) Moyenne radiale de la CTF après déconvolution complète jusqu'à 9  $\text{\AA}$  de résolution, il n'y a plus d'information au-delà.

Les images traitées sont dites « déconvolues ». On peut donc voir les effets de déconvolution sur la Figure 19b. L'atténuation des hautes fréquences est corrigée mais celle-ci n'est pas parfaite. Ces corrections introduisent beaucoup de bruits dans les images (notamment au niveau des zéros), il va donc falloir moyenner plus d'images par la suite dans le processus de reconstruction 3D pour avoir un rapport signal sur bruit équivalent à celui des images brutes. La correction de la CTF n'est réalisée que jusqu'à une valeur de résolution donnée par l'expérimentateur. Au delà de la résolution choisie (1/9  $\text{\AA}$ ), on peut voir sur les figures 18b et 18c que l'image ne contient plus du tout d'information. Ces corrections permettent tout de même de restaurer une grande partie de l'information dans l'image. Les nouvelles images vont donc être utilisées séparément en deux jeux, un jeu d'images « flippées » l'autre d'images « déconvolues ». Les images « flippées » vont servir à rechercher les orientations et le centre des particules car elles sont moins bruitées alors que les images « déconvolues » vont être utilisées pour la reconstruction 3D (Figure 20).

Analyse des images. Le traitement des images est composé de deux étapes majeures, résumées dans la Figure 20. Quelque soit le type de symétrie de la particule, l'image observée sur la micrographie d'un échantillon « congelé-hydraté » est toujours la projection 2D de l'objet. L'enjeu dans un premier temps est de déterminer l'orientation initiale selon laquelle l'objet a été projeté. Le centre de la particule est aussi un paramètre important à déterminer et à affiner. Le référentiel dans lequel sont définies les orientations et les coordonnées des centres dépend des conventions

implémentées dans le programme utilisé. Je détaillerai ces conventions plus loin dans mon exposé (cf. Matériel et méthodes).

La première étape du travail informatique utilise le jeu de particules dites « flippées » car ce jeu d'images contient plus d'informations que la micrographie de départ, les zones de contraste inverse dans l'image étant restaurées par inversion de phase. De plus, ces images possèdent un rapport signal sur bruit beaucoup plus important que le jeu d'images « déconvolées ».

Dans la deuxième étape, les paramètres de centre et d'orientation précédemment définis pour chaque particule sont utilisés pour reconstituer le modèle en 3D de la particule étudiée ([Figure 20](#)). Le principe de reconstruction 3D est basé sur un processus itératif d'affinement. La recherche des orientations et le calcul d'une reconstruction 3D sont réalisés en boucle jusqu'à obtention d'une reconstruction finale à la résolution la plus haute possible. Pour cela, le modèle 3D néo calculé va être projeté selon un pas angulaire connu et défini par l'expérimentateur. Les reprojctions du modèle possèdent par définition une orientation connue qui est utilisée par comparaison avec les images de départ pour affiner les paramètres d'orientations ([Figure 20](#)). Je détaillerais l'obtention du modèle dans une partie de mon manuscrit (cf. II.2.2, Partie 2).

Malgré les progrès réalisés dans les trente dernières années, une reconstruction 3D en MET ne permet que rarement d'avoir des informations structurales à l'échelle atomique sur l'objet que l'on étudie. La ME, de part la longueur onde du rayonnement utilisé, devrait permettre d'atteindre une résolution atomique mais les aberrations et déformations que subit l'image sont encore trop importantes pour donner des données structurales à l'échelle atomique. Les récents développements de la ME concernant la correction des aberrations ainsi que des programmes de corrections et d'analyse d'images permettent maintenant d'atteindre de hautes résolutions. Il est possible, au-delà de 7 Å de résolution, d'identifier clairement la position des hélices  $\alpha$  comme récemment pour les modèles de la protéine GroEL, celui de la capsid de l'Adénovirus de type 5, obtenus à 6 Å de résolution ([Jiang & Ludtke, 2005; Saban et al., 2006](#)) ainsi que pour un des virus de la famille des *Reoviridae* à 3.88 Å de résolution ([Yu et al., 2008](#)).

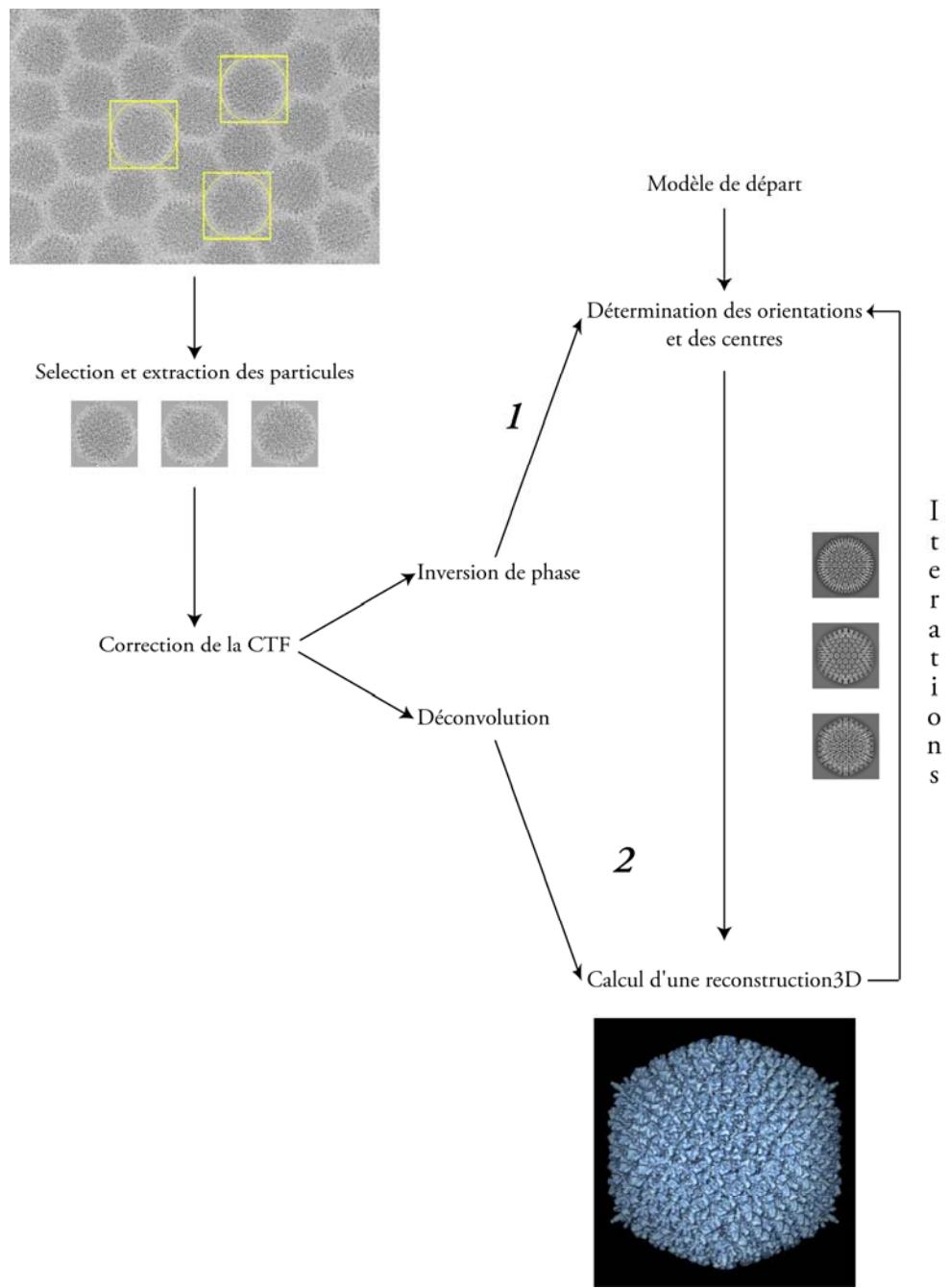


Figure 20: Principe général de l'analyse d'images en ME.

(a) Les particules sont sélectionnées et extraites à partir d'une micrographie. (b) La CTF est corrigée pour donner deux jeux d'images, « flippées » et « déconvolues ». (3) Etape 1: Les orientations de projection des particules sont déterminées par comparaison avec les reprojctions d'un modèle de départ à partir du jeu d'images « flippées ». (d) Etape 2: Le jeu d'images « déconvolues » est utilisé pour réaliser une reconstruction 3D en utilisant les orientations déterminées dans l'étape 1 puis le modèle obtenu sert à améliorer la recherche des orientations en calculant ses reprojctions selon un pas défini par l'expérimentateur. On s'abolit complètement du modèle de départ à partir de là.

## II.2 La Cristallographie aux rayons X

La radiocristallographie permet, en biologie, de déterminer la structure 3D à l'échelle atomique, de macromolécules comme des protéines ou des acides nucléiques à partir de cristaux. Déjà en 1952, des clichés de diffraction aux rayons X avaient permis à Watson et Crick de découvrir la structure de la double hélice d'ADN (Watson & Crick, 1953). La cristallographie aux rayons X devint particulièrement populaire dans les années 60 avec la découverte de la structure de la myoglobine et de l'hémoglobine par Max Perutz et John Kendrew (Kendrew et al., 1958; Perutz, 1964; Perutz et al., 1968). Ces découvertes leur ont valu le prix Nobel de Médecine (Watson et Crick) et de Chimie (Kendrew et Perutz) en 1962. Les techniques cristallographiques ont beaucoup progressées ces dernières années. Des nouveaux détecteurs de rayons X, couplés à l'utilisation du rayonnement synchrotron et les progrès fulgurants de l'informatique, permettent d'enregistrer rapidement des données de diffraction et de résoudre la structure atomique d'une protéine en un temps de plus en plus court. La principale limitation est l'obtention de cristaux. Les techniques de cristallogenèse ainsi que la robotisation permettent maintenant l'obtention de cristaux avec des quantités de matériel de plus en plus faibles mais la cristallogenèse reste souvent le facteur limitant dans le processus. De plus la cristallographie permet de résoudre la structure d'une molécule dans un cristal ce qui piège cette molécule dans un état conformationnel qui peut parfois être différent de celui en solution. Quand on s'intéresse à la structure d'un gros complexe multimérique, il est souvent plus simple de cristalliser les différentes sous-unités séparément. Malheureusement la structure d'une sous unité ne donne pas accès aux interactions ni à la structure globale du complexe protéique. La MET donne une structure basse résolution du complexe entier. Dans la perspective d'obtenir la structure atomique d'un gros complexe macromoléculaire, on peut, en utilisant la méthode de « fitting », tenter de reconstituer un modèle quasi-atomique à partir de sa reconstruction en MET et des structures atomiques de protéines qui composent ce complexe.

## II.3 Obtention d'un modèle quasi-atomique par « fitting »

La MET est une technique particulièrement adaptée pour obtenir la structure de gros complexes macromoléculaires. En effet, cette technique ne nécessite que peu de matériel biologique par rapport aux autres techniques de la biologie structurale (0,1 mg/ml pour la ME vs 10 mg/ml pour la cristallographie) mais la structure obtenue est à plus basse résolution par rapport à la

cristallographie qui restent à ce jour, avec la RMN, la seule technique de biologie structurale donnant accès à la structure atomique d'une macromolécule ; d'où le développement d'une méthode hybride de la biologie structurale : le « fitting ». Elle consiste à replacer, de manière suffisamment précise, à l'intérieur d'une reconstruction en MET les structures atomiques (RMN ou cristallographiques) des protéines qui lui correspondent pour reconstituer un modèle quasi-atomique.. Les premières méthodes mises au point calculent un recalage en corps rigide des structures atomiques (Figure 21). La structure atomique dans ce cas est considérée comme invariante et est déplacée sans déformation au sein de la reconstruction 3D issue de la ME. Très récemment, des méthodes de « fitting » flexible ont été développées. Les mouvements propres à la molécule sont calculés et la structure atomique est déformée en conséquence pour obtenir un meilleur résultat de « fitting ».

### *II. 3.1 En corps rigide*

Les premiers recalages datent de la deuxième moitié des années 90 (Roseman, 2000). Le placement des structures atomiques fût tout d'abord manuel à l'« œil » et réalisé surtout sur des capsides virales icosaédriques en complexe avec une partie de leur récepteur (Bella et al., 1998; Olson et al., 1993) ou en complexe avec un fragment d'anticorps (Hewat & Blaas, 1996; Stewart et al., 1997). Afin d'améliorer la qualité du placement manuel et la position des structures atomiques dans la reconstruction 3D, la communauté scientifique adapta des programmes dérivés de la cristallographie et de la RMN. Phoebe Stewart et ses collaborateurs obtinrent le premier recalage automatique ainsi qu'un modèle quasi-atomique en 1993 pour la capsid de l'Adénovirus de type 2 en utilisant le programme XPLOR, actuellement remplacé par CNS (Brunger et al., 1998). Par la suite, de nombreux programmes de « fitting » automatique ont été développés comme COLORES de Situs (Wriggers et al., 1999), COAN (Volkmann & Hanein, 1999), DOCKEM (Roseman, 2000), EMFIT (Rossmann, 2000), URO (Navaza et al., 2002) ou 3SOM (Ceulemans & Russell, 2004). Le package Situs développé par Willy Wriggers ainsi que URO développé par Jorge Navaza sont actuellement les deux programmes les plus distribués pour la communauté scientifique.

La problématique imposée dans la procédure de « fitting » automatisée est une recherche en corps rigide dans six dimensions. En effet, il faut déterminer la position de la molécule selon trois degrés de liberté en rotation (recherche rotationnelle) et trois degrés de liberté en translation (recherche translationnelle) selon les axes  $Ox$ ,  $Oy$  et  $Oz$  du repère dans lequel on se situe. Le choix

quant à l'utilisation de ces différents programmes doit prendre en compte plusieurs critères comme l'espace dans lequel se réalisent les calculs (réel ou réciproque). L'utilisation de l'espace réciproque accélère énormément le temps de calcul. Le 2<sup>ème</sup> critère dont il faut tenir compte est la prise en compte des symétries internes à la particule. En effet, pour une particule comme la capsidé des Adénovirus qui possède une symétrie d'ordre 60, il est nécessaire de fixer les symétries pendant la procédure de « fitting » pour éviter les superpositions d'atomes. Enfin, le dernier critère est le type de recherche effectuée par le programme. Certains logiciels comme COLORES ou EMFIT font une recherche exhaustive dans la totalité du repère défini par la reconstruction 3D. Il n'est donc pas nécessaire de réaliser un placement manuel approximatif avant de débuter la procédure. L'autre type de recherche implémentée dans URO ou EMFIT utilise la position définie par l'utilisateur lors du positionnement manuel. Le programme ne fait qu'affiner la position définie au départ par l'expérimentateur. Ce type de recherche réduit considérablement le temps de calcul mais nécessite un placement initial assez ajusté car URO ne tient pas compte des clashs entre capsomères.

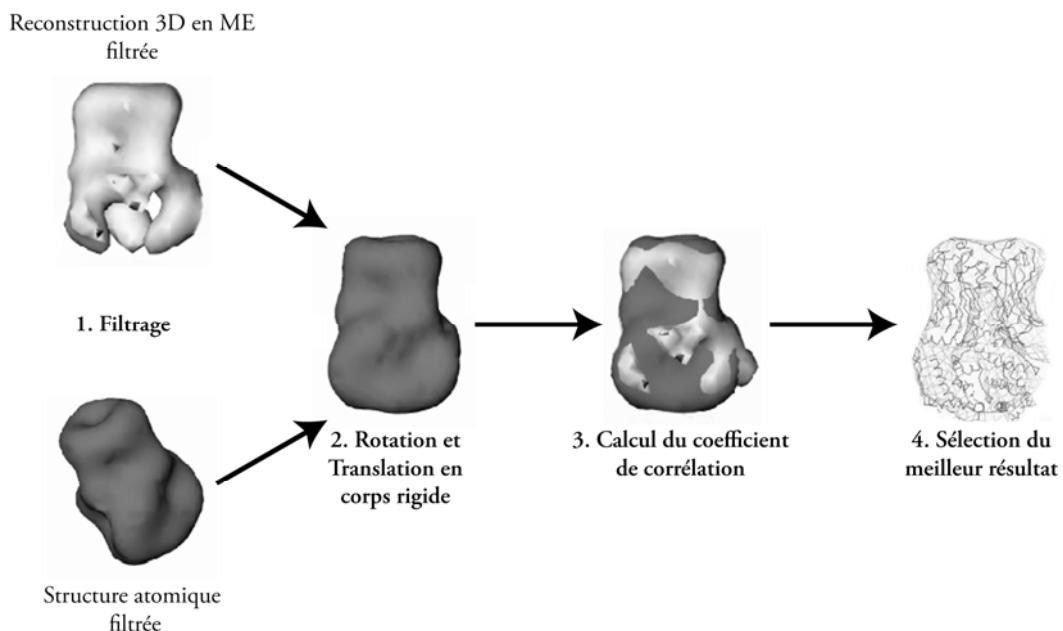


Figure 21: Représentation schématique du processus de "fitting" automatique.

- (a) La structure à recaler ainsi que la reconstruction 3D sont filtrées. Le type de filtre dépend du logiciel utilisé.
- (b) Le programme recherche les paramètres de translation et de rotation. (c) Calcul de critères de qualité qui est très souvent un coefficient de corrélation. (d) Sélection du meilleur résultat et création d'un nouveau fichier avec la position mise à jour de la structure atomique dans le référentiel de la reconstruction de MET.

Le principe de la procédure de « fitting », illustré en Figure 21, reste identique quelque soit le programme utilisé. Le logiciel a tout d'abord recourt à une série de filtrages sur la reconstruction 3D et la structure atomique afin de faciliter le placement. Ensuite la recherche de la bonne position est

réalisée selon les 6 degrés de liberté de l'espace. Enfin le programme calcule des critères de qualité afin de déterminer la meilleure solution et écrit un nouveau fichier contenant la structure atomique à la meilleure position déterminée pendant la procédure de « fitting ». L'utilisation de programmes automatiques de « fitting » a permis le calcul précis de critères de qualité importants pour évaluer le résultat final de « fitting ». Plusieurs critères de qualité sont disponibles, en fonction du programme utilisé, mais le principal reste le calcul d'un coefficient de corrélation CC. D'autres critères de qualité peuvent être calculés comme le facteur cristallographique  $R_f$  ou le critère Q ou écart quadratique entre la densité de la reconstruction 3D et celle du modèle atomique. Chaque programme possède une façon propre de calculer les critères de qualité. Dans ces conditions, il est souvent difficile de comparer des modèles quasi-atomiques obtenus avec deux programmes différents.

### *II. 3.2 Les modes normaux et le « fitting » flexible*

Au cours du processus de « fitting » en corps rigide, l'expérimentateur peut rencontrer certains problèmes. Si l'état conformationnel dans la structure atomique est différent de l'état conformationnel observé dans la reconstruction 3D, le recalage des données donne souvent un résultat médiocre. Il est nécessaire dans ce cas de « déformer » la structure atomique de la protéine pour que son positionnement dans le modèle de MET soit correct. Pour cela, des programmes sont capables de modifier et de déformer une structure atomique en calculant les modes normaux de la molécule. Ces modes sont basés sur l'énergie vibrationnelle de la molécule étudiée. A chaque mode normal correspond une conformation de la molécule. Le calcul des modes normaux à basse fréquence va permettre de trouver les mouvements de grande amplitude qui domine la dynamique de la protéine étudiée.

Les deux principaux programmes disponibles pour réaliser une déformation d'une structure sont NMFF (Normal Mode Flexible Fitting) développé par une équipe américaine du Scripps institute (Tama et al., 2004) et elNémo/NORMA développé par une équipe CNRS de Marseille (Suhre et al., 2006). NORMA est en fait une combinaison du programme de « fitting » en corps rigide URO avec un calcul des modes normaux NMA (Normal Mode Analysis). Si on ne désire pas utiliser le programme, le serveur web elNémo est disponible pour le calcul des modes normaux d'une molécule (Suhre & Sanejouand, 2004).

### **III. Objet du travail de thèse**

L'objectif de ma thèse était d'étudier la structure de la capsidé des Adénovirus. Dans un premier temps, le but était d'obtenir une reconstruction 3D en ME à haute résolution de la capsidé de l'Adénovirus humain de type 5. Puis dans un second temps de reconstituer un modèle quasi-atomique de la capsidé par la méthode de « fitting » de combinaison des données atomiques connues de protéines majeures et des données plus basse résolution obtenue en ME sur le virus entier. Ce travail m'a permis de proposer une localisation pour plusieurs protéines mineures au sein de la capsidé. Afin de confirmer leur localisation nous avons voulu étudier, avec les mêmes techniques de biophysique, la structure de la capsidé d'autres souches d'Adénovirus comme l'adénovirus aviaire de type 1, l'adénovirus canin de type 2 ou des souches modifiées de l'adénovirus humain de type 2 ou 5. En effet, ces différents virus ne possèdent pas le même réseau externe et interne de protéines mineures. Une publication en 2006 de la structure à haute résolution en MET de la capsidé de l'AdV-5 par Saban et collaborateurs ([Saban et al., 2006](#)) a révélé une localisation très précise des protéines de capsidé sur la base de résultats obtenus en ME et bioinformatique, sans apporter de preuves biochimiques. Une ambiguïté restait quant à la localisation de la protéine IIIa et de la protéine IX. J'ai donc réalisé un marquage par un Fab polyclonal spécifique d'un peptide accroché à l'extrémité carboxy-terminale de la protéine IX, combiné à une étude en MET. Le travail sur les différentes souches de virus ainsi que la combinaison avec les données cristallographiques nous a permis de confirmer la localisation de la protéine IX. Par la suite, mon expertise sur la méthode de « fitting » m'a amené à travailler sur deux projets annexes. Le premier consistait à étudier la structure d'une sous particule virale dérivée de l'Adénovirus humain de type 3 : Le dodécaèdre. J'ai donc réalisé le recalage de la structure cristallographique de la base du Penton de l'Adénovirus humain de type 2 dans les reconstructions 3D obtenues pour le dodécaèdre avec et sans fibre. Ces modèles nous ont permis de proposer une explication pour l'attachement de la fibre. Le dernier projet, en collaboration avec l'équipe de Bruno Franzetti à l'IBS, concernait l'étude d'une Aminopeptidase multimérique *TET1* extraite de l'organisme hyperthermophile *P.horikoshii*. In vitro, cette protéine est capable de former deux homomultimères à 12 sous unités (12-mer) et 24 sous unités (24-mer). Nos collaborateurs ayant résolu la structure atomique du 12mer, j'ai pu reconstituer un modèle quasi-atomique pour la forme à 24 sous unités (24-mer) (cf. Annexe 1).

# STRUCTURE DE LA CAPSIDE DES ADENOVIRUS ET DES SOUS PARTICULES VIRALES

---

Cette partie constitue le sujet principal de mon travail de recherche. Il s'intéresse à la structure de la capsid des adénovirus. Ces virus, découverts au milieu du 20<sup>ème</sup> siècle, ont été particulièrement étudiés dans la 2<sup>ème</sup> partie du siècle mais les connaissances biologiques et surtout structurales restent de nos jours assez limitées.

## *I. Introduction*

### I.1 La famille des *Adenoviridae*

Les Adénovirus (AdV) ont été isolés pour la première fois au début des années 50 à partir de prélèvements d'amygdales et de sécrétions respiratoires, respectivement, chez des enfants et des militaires atteints d'infections respiratoires aiguës (Hilleman & Werner, 1954; Rowe et al., 1953). Deux ans après sa découverte, l'agent pathogène responsable de ces infections est appelé Adénovirus à cause de sa présence dans les amygdales (« adeno » signifie glande). La découverte de Trentin et collaborateurs (Trentin et al., 1962) sur la capacité des Adénovirus à induire des tumeurs chez le hamster a fortement stimulée la recherche fondamentale par la suite.

La famille des *Adenoviridae* a été définie en 1976 (Norrby et al., 1976). A ce moment là, la famille n'était composée que de 2 genres distincts : les *Mastadenovirus* et les *Aviadenovirus*. Chaque genre est composé de plusieurs espèces annotées en fonction de l'hôte qu'il infecte et d'une lettre correspondant au sous groupe auquel il appartient (Figure 22). La classification la plus récente des Adénovirus a été établie sur la base d'identité de séquences de certains gènes ou parties du génome.

Le genre *Mastadenovirus* infecte une large gamme de mammifères incluant l'homme, le chien, le cheval, les ovins et les bovins. Les souches les plus étudiées à ce jour restent celles infectant l'homme en particulier les sérotypes 2 et 5 du sous-groupe C. Les *Mastadenovirus* sont les plus abondants et composent une multitude de sous-groupes. Le genre *Aviadenovirus* infectent uniquement des oiseaux. Il est composé de 3 sous-groupes. Le sérotype le plus étudié est le FAdV-1 également connu sous le nom de CELO (Chicken Embryo Lethal Orphan). Suite aux travaux de Benko et collaborateurs en 1998 la famille s'est enrichie d'un 3<sup>ème</sup> genre (Benko & Harrach, 1998).

Le fort pourcentage en bases A et T au sein de leur génome leur doit le nom de *Atadenovirus*. Ce genre comporte une variété assez importante d'hôtes similaires à ceux des *Mastadenovirus* et des *Aviadenovirus* mais se distinguent par l'organisation atypique de leur génome par rapport aux deux premiers genres, surtout dans la région E4 (Figure 29) (Davison et al., 2003). Le 4<sup>ème</sup> genre de la famille ne comporte que deux espèces, le FrAdV-1 et TAdV-3 isolés respectivement chez une grenouille et différents oiseaux (Davison & Harrach, 2002). Leur génome code pour une sialidase hypothétique d'où le nom de *Siadenovirus*. Enfin, le seul adénovirus connu pour infecter un poisson (esturgeon) est génétiquement le plus éloigné des 4 premiers genres de la famille et pourrait constituer le dernier genre connu à ce jour (Figure 22).

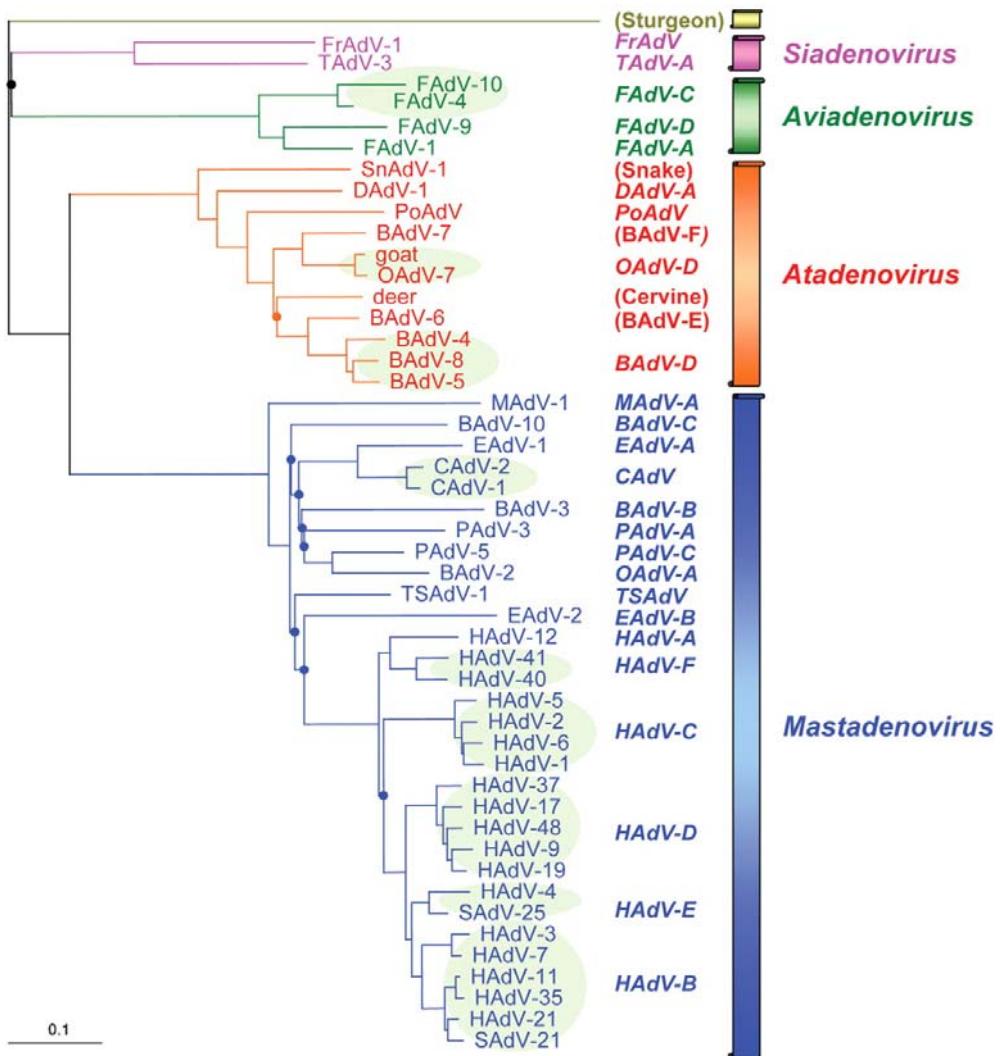


Figure 22: Arbre phylogénétique des adénovirus basé sur la distance entre les gènes codant pour la protéine structurale majeure : l'hexon (Davison et al., 2003).

Les membres des différents genres sont indiqués en couleurs différentes. Les souches virales appartenant à la même espèce sont entourées en vert. Le nom de chaque souche est indiqué en abrégé à la fin de chaque branche. L'espèce auxquels ils appartiennent est indiqué sur la droite en italique (pour les espèces reconnues par la nomenclature). B : bovin, C : canin, D : canard, E : équin, F : volaille, Fr : grenouille, H : humain, M : murin, O : ovin, P : porcin, Po : opossum, Sn : serpent, T : dinde, S : primates, TS : scadentien.

## I.2 Un peu d'épidémiologie

Maladies chez l'homme	Sous-groupes (sérotypes majeurs)
Maladie des voies respiratoires hautes	B, C
Maladie des voies respiratoires basses	B (3, 7, 21), E (4)
Maladie respiratoire aiguë	B (7, 21), E (4)
Kéatoconjonctivite épidémique	D (8, 19, 37)
Gastroentérite infantile	A (31), C(2), F (40, 41)
Fièvre Pharyngoconjonctivale	B (3, 7), E(4)
<hr/>	
Syndrome pertussique	C (5)
Conjonctivite hémorragique aiguë	B (11)
Cystite hémorragique aiguë	B (7, 11, 21, 35)
Maladie du patient immunodéprimé	A (31), B, C, D (29, 30, 37, 43, 45)
Maladies du système nerveux central	B (3, 7)
Maladies sexuellement transmissibles	C (2), D (19, 37)

**Tableau 2: Les maladies humaines associées aux Adénovirus.**

Maladies associées aux Adénovirus chez l'homme sain. Quand le sérotype n'est pas indiqué entre parenthèses, tous les sérotypes du groupe sont responsables de la maladie. Les 6 dernières lignes font référence à des maladies rares (Hierholzer, 1992).

Les Adénovirus ne sont responsables que de maladies assez bénignes. Seules les maladies causées par les 2 premiers genres *Mastadenovirus* et *Aviadenovirus* sont réellement documentées. Les maladies liées aux Adénovirus sont endémiques dans la population humaine. Elles entraînent le développement d'une immunité protectrice, spécifique du sérotype infectant mais on observe tout de même des infections épidémiques sporadiques. Chez les sujets immunodéprimés, même les souches relativement inoffensives d'Adénovirus peuvent entraîner des décès.

Ces virus, quelque soit le genre ou l'hôte, infectent essentiellement les cellules épithéliales des muqueuses respiratoires, gastro-entériques ou oculaires. Ils peuvent parfois atteindre le foie ou le rein. Chez l'homme, ils sont la cause d'infections respiratoires de type rhino-pharyngites, de gastroentérites ou d'infections oculaires comme certains types de conjonctivites bénignes qui peuvent prendre des formes plus sérieuses chez les enfants et les sujets immunodéprimés (Tableau 2). Les infections respiratoires, souvent limitées aux voies respiratoires hautes sont provoquées par trois sous groupes essentiellement : les sous-groupes B, C et E. Dans quelques cas, les groupes B et E peuvent provoquer une infection des voies respiratoires basses occasionnant des maladies respiratoires aiguës.

Le sérotype 5 est responsable d'un syndrome pertussique en association à une infection bactérienne. Le syndrome pertussique représente un état caractéristique de symptômes liés à une infection par la bactérie *Bordetella pertussis*, agent infectieux responsable de la coqueluche. Ce syndrome se caractérise par une forte toux et d'importantes diarrhées. Les Adénovirus semblent aussi impliquer dans quelques maladies rares ([Tableau 2](#)).

Ils existent 4 sérotypes distincts chez les Adénovirus aviaires, infectant uniquement les oiseaux.

Chez les bovins et les ovins, les Adénovirus sont la cause de maladies respiratoires bénignes. Chez le chien, il n'existe que deux sérotypes, le CAdV-1 et le CAdV-2. Le 1<sup>er</sup> est responsable d'une maladie hépatique aiguë appelée « Hépatite de Rubarth » souvent associé à des symptômes respiratoires, oculaires et entériques assez graves alors que le 2<sup>ème</sup> ne provoque qu'une infection respiratoire bénigne.

### I.3 Structure et organisation générale de la capsidé

Je vais vous présenter la structure et l'organisation générale de la capsidé des Adénovirus Humains du sous-groupe C : HAdV-2 et HAdV-5, telle qu'elle était connue en 2004. La structure et les particularités des souches de virus aviaires et canins ou d'autres souches humaines seront détaillées dans les parties I.6, I.7 et I.8.

Les Adénovirus sont des virus non enveloppés à symétrie icosaédrique. Leur génome est composé d'une molécule d'ADN linéaire double brin d'environ 35 000 paires de bases (pb), associée à plusieurs protéines mineures (cf. I.4). La capsidé possède un diamètre moyen de 85 nm et se compose de trois protéines structurales majeures (II, III et IV) ainsi que de nombreuses protéines mineures (IIIa, IX, VI, VIII). La première protéine majeure est l'Hexon (II). Elle forme les faces du virus. Les sommets de l'icosaèdre sont constitués par l'assemblage non covalent d'une protéine pentamérique : La Base du Penton (III) et d'une protéine trimérique qui se projette vers l'extérieur : La Fibre (IV).

Chaque face est constituée de 12 Hexons dont 9 forment le groupe des neuf Hexons ou GON (en gris clair et foncé sur la [Figure 23b](#)). Ceux-ci sont observés lors de la dissociation de la capsidé par simple dialyse ([Pereira & Wrigley, 1974](#)). L'observation de ces structures par MET a permis, quelques années plus tard, de définir la localisation et le rôle de la protéine IX, comme

ciment des GONs (Furciniti et al., 1989). De nombreuses autres protéines mineures viennent stabiliser la capsidé formée par les protéines majeures (Figure 23). Des reconstructions 3D en MET de la capsidé complète ont permis de compléter le placement des protéines mineures comme la protéine IIIa, aussi du coté externe de la capsidé. La communauté scientifique a donc admis, depuis 1993, la structure de la capsidé proposée par Phoebe Stewart obtenue à partir d'une reconstruction 3D en ME à 25 Å de résolution (Stewart et al., 1993) mais les preuves biochimiques et structurales restaient tout de même faibles. L'étude de la localisation précise de la protéine IX et IIIa fait l'objet d'une partie de mon travail de recherche. Les protéines IX et IIIa sont présentes sur la face externe du virus alors que les protéines VI et VIII ont été proposées sur la face interne mais leur localisation précise reste tout de même controversée.

Tous les Adénovirus ne possèdent pas les mêmes caractéristiques structurales en termes de composition en protéines majeures et mineures. En effet, les Adénovirus canins ou aviaires ne montrent pas le même réseau de protéines mineures. Les particularités de chaque Adénovirus dont j'ai étudié la structure seront détaillées dans la partie qui leur correspond (cf. I.7, I.8).

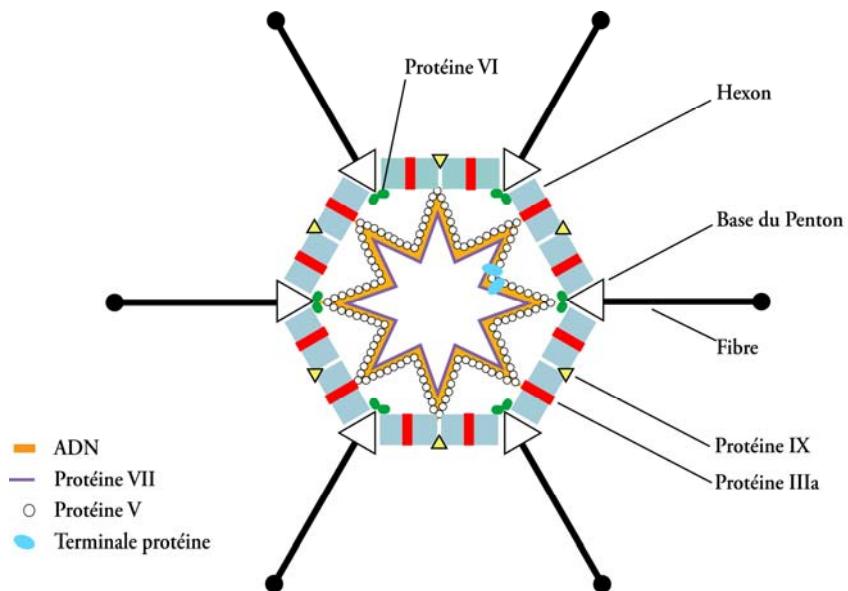


Figure 23: Vue schématique en coupe de la capsidé de l'Adénovirus de type 5 d'après Stewart et al., 1993.

Chaque type de protéine majeure (Hexon, Base du Penton, Fibre) ou mineure (IX, IIIa, V, VI, VII, tp) est coloré d'une couleur différente.

### I. 3.1 La symétrie de la capsid

La capsid des Adénovirus a une structure en icosaèdre, formé par les protéines majeures Hexon (II) et Base du Penton (III). Elle possède un nombre de triangulation *pseudo* T = 25 (Figure 12). Les facettes de l'icosaèdre sont formées par l'Hexon à l'état trimérique (Figure 23). La base du Penton sous forme de pentamère constitue les sommets. Cette capsid est composée de 20 faces et de 12 sommets en accord avec son modèle platonicien. Les axes de symétrie 5, 3 et 2 se situent respectivement au niveau des sommets, du centre des faces et du milieu des arêtes. La capsid possède un ordre de symétrie 60 et 60 matrices de rotation permettent donc de reconstituer la capsid complète si l'on connaît la position des constituants d'une unité asymétrique (UA). L'UA est la partie minimale nécessaire et suffisante pour reconstituer la capsid entière. Chez les Adénovirus, elle est constituée de 4 Hexons numérotés, par convention, de 1 à 4 et d'un monomère de Base du Penton (Figure 24). L'UA peut être choisie à deux endroits distincts (Burnett, 1985). Toujours encadrée par les 3 types d'axes de symétrie, l'UA peut contenir des Hexons appartenant au même GON (en bleu sur la Figure 24a) ou à deux GONs différents (en rose sur la Figure 24b). En effet, les Hexons 1 et 4 de la Figure 24 appartiennent au GON de la face adjacente. Dans ce cas, on considère que les Hexons appartiennent à la face adjacente. Ce qui conduit à la formation d'une UA à partir d'Hexons appartenant à deux faces distinctes.

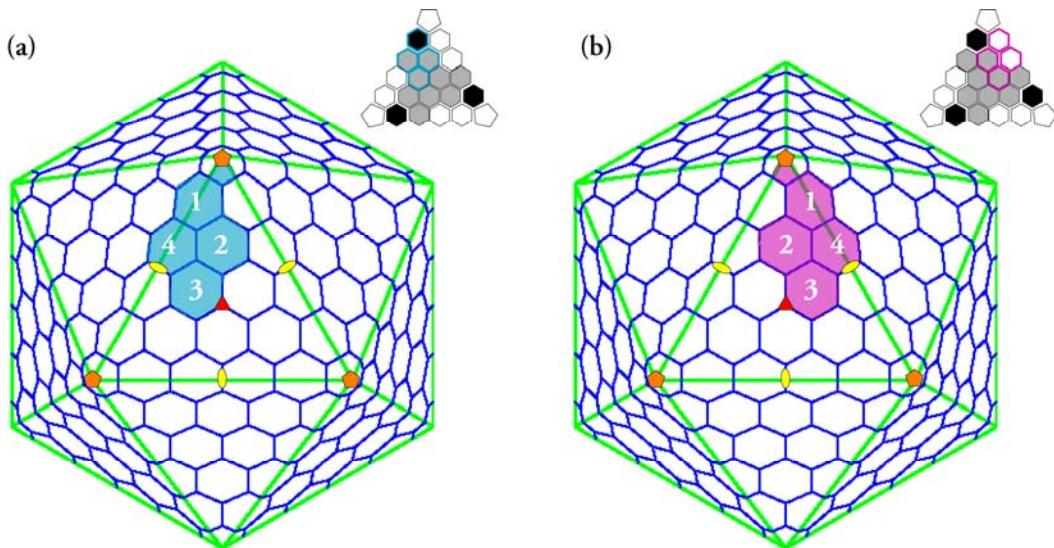


Figure 24: Choix de l'unité asymétrique avec les axes de symétrie (Burnett, 1985).

Les axes d'ordre 5, 3 et 2 sont respectivement représentés par un pentagone orange, un triangle rouge et une ellipse jaune. (a) La zone en bleu représente la 1<sup>ère</sup> position possible pour l'UA qui englobe uniquement des Hexons appartenant à la même face. (b) La zone en rose représente la 2<sup>ème</sup> position possible pour l'UA qui contient des Hexons appartenant à deux faces différentes. En encart, vues schématiques d'une face avec le GON en gris clair, les Hexons périmentaux de la face en noir, les Hexons des faces adjacentes en blanc et la Base du Penton représentée par un pentagone blanc.

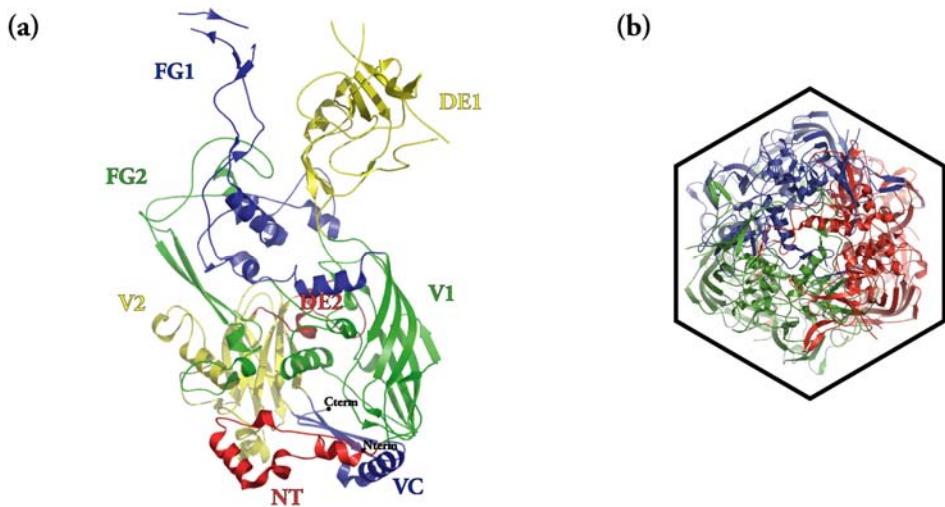
Cette capsidé possède un pseudo nombre de triangulation. Elle ne suit donc pas strictement la règle établie par Caspar et Klug concernant la quasi-équivalence. En effet, cette capsidé ne contient que 4 environnements différents au lieu de 25. Le nombre d'environnements correspond au nombre d'Hexons présents dans l'UA (Figure 24). En effet, l'Hexon numéro 1 est toujours en contact avec 5 autres Hexons et 1 Base de Penton. Il est appelé « péripentonal » car il est en contact direct avec la base du Penton. Les trois autres Hexons de l'UA (Hexons 2, 3 et 4 sur la Figure 24) sont toujours en contact avec 6 Hexons adjacents mais occupe des positions différentes par rapport aux axes de symétrie de la capsidé. En effet, l'Hexon numéro 2 se situe à la charnière entre 2 GONs de deux faces différentes, l'Hexon numéro 3 a une position centrale au niveau de l'axe d'ordre 3 de l'icosaèdre alors que l'Hexon 4 se situe au niveau des axes d'ordre 2. La présence des différents axes de symétrie impose donc 3 environnements distincts à ces trois Hexons. Le choix de la position de l'UA n'a pas d'influence sur le nombre d'environnements. J'ai donc choisi de fixer l'UA en position 1 (Figure 24a). Dans ce cas, l'UA appartient à une seule face ce qui facilite l'analyse des contacts entre protéines.

### I. 3.2 L'Hexon

Cette protéine, aussi appelée protéine II, est présente en 720 copies à la surface de la capsidé ce qui en fait la protéine la plus abondante des protéines structurales du virus. Chaque monomère d'environ 105 kDa s'assemble en trimère pour former 240 capsomères possédant une base hexagonale et un sommet triangulaire (Figure 25b). Cette protéine a été cristallisée et sa structure a été résolue tout d'abord chez l'Adénovirus de type 2 à 2,9 Å de résolution puis chez le type 5 à 2,5 Å (Burnett et al., 1985; Roberts et al., 1986; Rux & Burnett, 2000). Les parties N terminales de chaque monomère, partiellement ordonnées en hélices, sont responsables des interactions avec les deux monomères adjacents stabilisant ainsi le trimère.

Au niveau de sa base, l'Hexon possède une structure secondaire composée de deux feuillets  $\beta$  de 8 brins (V1 et V2 sur la Figure 25) et 18 hélices qui forment, dans le trimère, une zone très stable en tonneau  $\beta$  appelé « Viral Jellyroll ». Cette protéine se prolonge en surface par des domaines moins organisés (DE1, FG1 et FG2) composées de régions flexibles. Ces zones, trop flexibles pour être résolue en cristallographie, sont composées de sept boucles hypervariables (HRV1 à 7, Rux & Burnett, 2000) responsables de l'antigénicité importante de la capsidé. La structure de l'Hexon du CELO a été résolue récemment (Xu et al., 2007). Cette protéine, très stable à l'état trimérique

(résiste à une dénaturation par le SDS) ne s'oligomérise pas sans l'aide de la protéine 100K. La protéine 100K est une chaperonne codée par le virus qui aide à l'assemblage de la capsidé lors de la phase tardive d'infection.



**Figure 25: Structure atomique de l'Hexon de l'Adénovirus humain de type 5 (Rux & Burnett, 2000).**

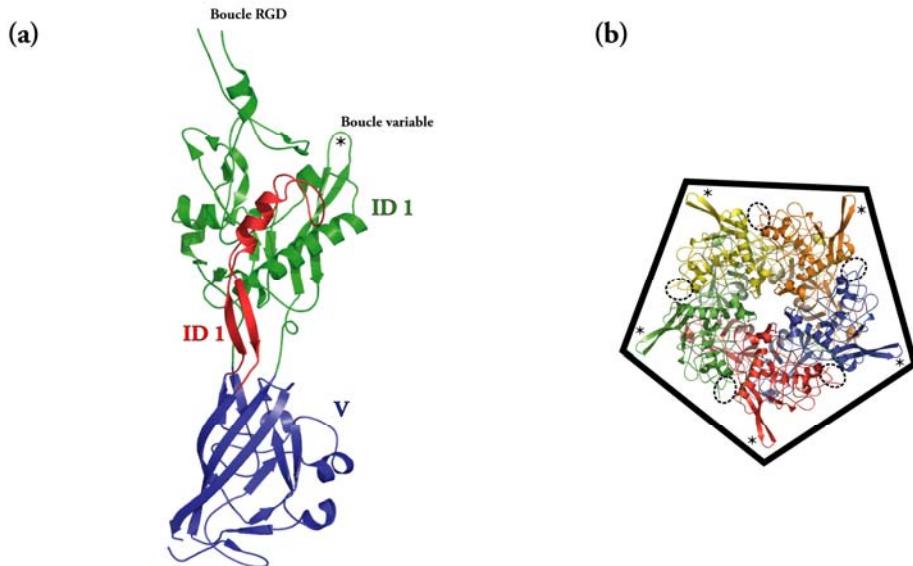
(a) Représentation en cartoon d'un monomère avec le nom des différentes régions. Les parties V1 et V2 constituent les zones en sandwich  $\beta$ . (b) Vue de dessus du trimère. Code PDB : 1P30.

### I. 3.3 La Base du Penton

La base du Penton ou protéine III est la 2<sup>ème</sup> protéine majeure de la capsidé des Adénovirus. Son monomère d'environ 60 kDa est présent en 60 copies à la surface du virus mais il est toujours présent à l'état pentamérique au niveau des 12 sommets de l'icosaèdre. Le pentamère possède une forme pentagonale (Figure 26b). Cette protéine réalise une association non covalente avec la fibre qui se projette à partir du sommet vers l'extérieur du virus. L'association non covalente de la Fibre avec la Base du Penton représente le Penton. Ce complexe est responsable de l'attachement et de l'entrée du virus dans la cellule.

La structure de la Base du Penton de l'Adénovirus de type 2 est connue depuis 2005. Celle-ci a été résolue à 3,5 Å de résolution à l'EMBL par le Dr. Chloé Zubieta et le Dr. Stephen Cusack (Figure 26) (Zubieta et al., 2005). Cette protéine présente une organisation générale en deux domaines distincts. La partie interne ou basale est constituée d'un feuillet  $\beta$  à 8 brins qui forme, au niveau du pentamère, un tonneau  $\beta$  à 5 feuillets constituant ainsi le « Viral Jellyroll » (V sur la Figure 26a). La partie externe est constituée en majorité d'hélices similaires aux régions DE et FG de l'Hexon. Deux zones de cette protéine présentent une très grande variabilité entre les différents sérotypes d'Adénovirus ainsi que chez d'autres espèces (boucle RGD et boucle variable, Figure 26).

Tout d'abord, on remarque la présence d'une grande boucle flexible et partiellement désordonnée dans la structure cristallographique. Cette boucle, de séquence et de longueur variable selon le sérototype, contient, chez les HAdV-5 et 2, un motif RGD (Arginine – Glycine – Acide Aspartique) conservé, responsable de l'interaction avec l'intégrine  $\alpha_v\beta_5$  lors de l'internalisation du virus.



**Figure 26: Structure atomique de la base du Penton de l'Adénovirus humain de type 2 (Zubieva et al., 2005).**

(a) Un monomère de la base du Penton est représenté en cartoon. Chaque région de la protéine est annotée d'une couleur différente. La zone V représente la partie de la protéine qui forme le tonneau  $\beta$  présent dans le pentamère. (b) Vue de dessus du pentamère. Chaque monomère est représenté d'une couleur différente. La boucle variable est annotée par une étoile et la boucle RGD est entourée en pointillés. Code PDB : 1X9P.

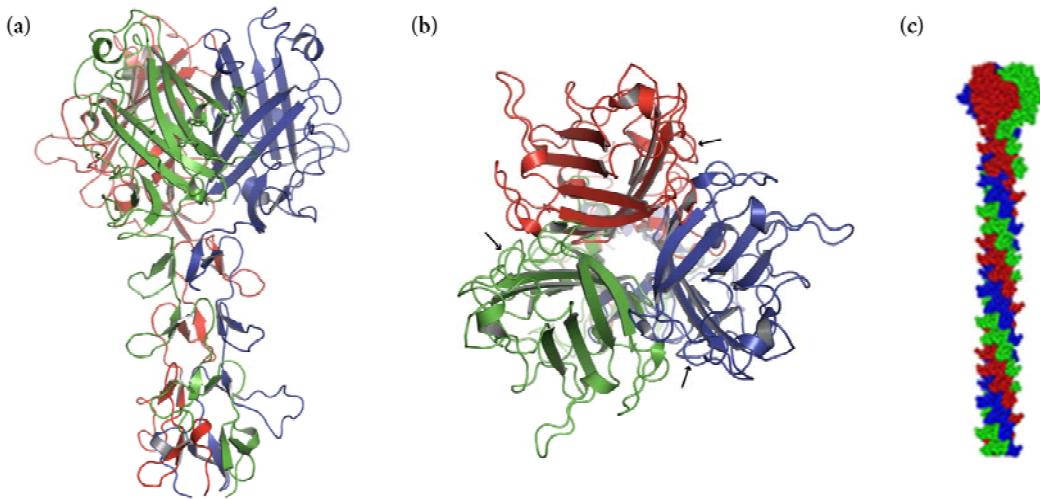
Une 2<sup>ème</sup> zone de longueur et de séquence variable est présente entre deux brins  $\beta$  (symbole \* sur la [Figure 26](#)). Cette boucle porte tout simplement le nom de « boucle variable ». La base du Penton de l'Adénovirus de type 2 est la seule base dont la structure est connue à ce jour mais des essais de cristallisation de la Base du Penton du CELO et du HAdV-3 sont en cours dans notre laboratoire. La base du Penton, chez certains sérotypes de virus, s'assemble en solution pour former une structure dodécaédrique de 12 bases : le dodécaèdre (cf. I.6)

### I. 3.4 La Fibre

La fibre, aussi appelée protéine IV, est la 3<sup>ème</sup> et dernière protéine majeure. La capsidé est constituée de 36 copies de cette protéine assemblée en trimères au niveau des 12 sommets de l'icosaèdre. La fibre, en liaison non covalente avec la Base du Penton, se projette vers l'extérieur pour réaliser l'interaction avec le récepteur primaire lors de la reconnaissance primaire de la cellule. La forme biologique en trimère est composée de 3 domaines principaux : une queue N terminale flexible, responsable de l'interaction avec la Base du Penton, une tige assez rigide de longueur variable en fonction du sérotype ainsi qu'une tête C terminale et globulaire qui permet l'interaction avec le récepteur primaire.

La tête de la Fibre de l'HAdV-5 fut la 1<sup>ère</sup> protéine à avoir été cristallisée en 1994 (Xia et al., 1994). La structure de nombreuses têtes est maintenant connue même chez d'autres espèces d'adénovirus comme le CELO ou le CAdV-2 (El Bakkouri et al., 2008; Guardado-Calvo et al., 2007; Seiradake et al., 2006). Toutes les têtes de fibre montrent un repliement homologue. Chaque monomère est constitué d'un sandwich  $\beta$  anti-parallèle formé de 2 feuillets  $\beta$  à 4 brins (Figure 27). Seulement chez les souches de virus qui reconnaissent CAR comme récepteur primaire, une zone particulièrement conservée, appelée boucle AB, est impliquée dans la liaison avec ce récepteur.

Le 2<sup>ème</sup> domaine de la protéine est la tige. En 1999, Van raaïj et collaborateurs ont résolu la structure de la tête de la fibre de l'HAdV-2 avec une partie de la tige (van Raaïj et al., 1999). La tige est repliée en motifs répétés de 15 acides aminés qui sont au nombre de 22 pour la fibre des HAdV-2 et HAdV-5 (Rugrok et al., 1990). Ces motifs ne sont pas identiques sur toute la longueur de la fibre mais leur répétition permet une variabilité évidente dans la longueur de la fibre. La tige est reliée à la tête par une région flexible nécessaire à sa fonction biologique. Le dernier domaine de la protéine, en N terminal, est la queue. Celle-ci contient un motif conservé (FNPVYPY) responsable de la liaison à une zone conservée de la Base du Penton qui forme un sillon entre deux monomères comme l'a montré la structure cristallographique de la Base du Penton de l'HAdV-2 en complexe avec le peptide N terminal de la fibre (Zubieta et al., 2005).



**Figure 27: La structure de la Fibre de l'Adénovirus de type 2.**

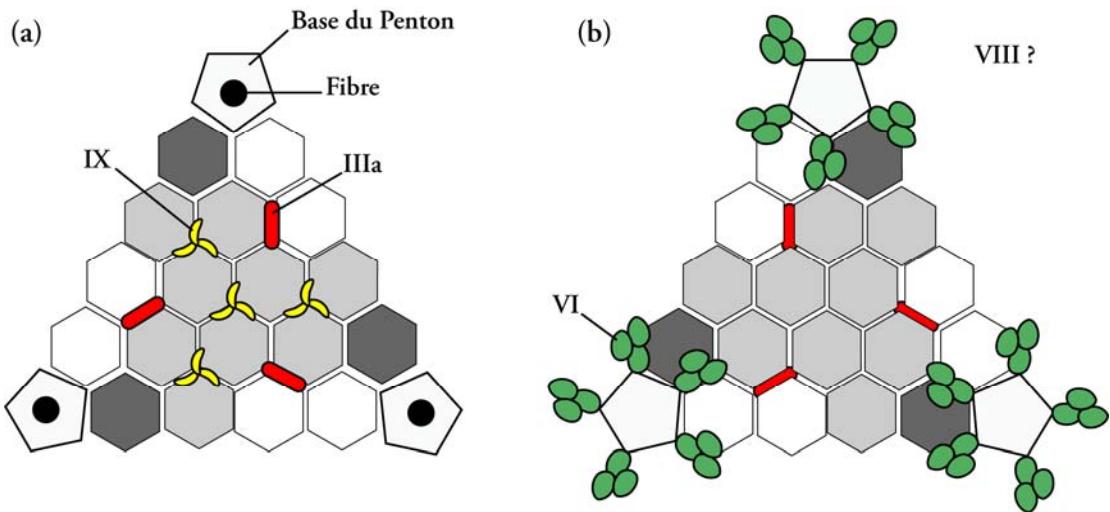
(a) Vue de coté de la tête et du début de la tige. (b) Vue de dessus de la tête. La boucle AB responsable de l'interaction avec le récepteur CAR est indiquée par une flèche. (c) Vue de coté du modèle de la Fibre complète présentée par [van Raaij et al., 1999](#).

### I. 3.5 Les protéines structurales mineures

Aucune structure atomique n'est disponible à ce jour pour les protéines mineures. Le séquençage du génome de l'HAdV-2 ainsi qu'une étude par marquage à la méthionine radioactive a permis de déterminer la stoechiométrie des protéines mineures de la capsid ([van Oostrum & Burnett, 1985](#)). Par la suite, des études structurales en ME ont établi la localisation de certaines protéines mineures mais leur position précise n'est pas connue en absence de structure atomique pour le virus entier. Je vais donc vous présenter la localisation et le rôle des protéines mineures des adénovirus humains du sous groupe C, admis par la communauté scientifique avant le début de ma thèse. Le travail que j'ai réalisé en 2005 a apporté des précisions structurales mais la reconstruction 3D d'un adénovirus humain obtenue à 6 Å de résolution par Saban et collaborateurs en 2006 a bouleversé le modèle établi ([Saban et al., 2006](#)). Je décrirai par la suite les particularités observées chez les adénovirus aviaires et canins concernant les protéines mineures dans les chapitres I.7 et I.8.

Protéine IIIa. Cette protéine est le produit C terminal de clivage d'un précurseur de 65 kDa par la protéase codée par le virus. La protéine maturée avoisinerait les 60 kDa mais sa taille exacte n'a pas été confirmée. Cette protéine semble être à l'état monomérique, en 60 copies dans la capsid ([van Oostrum & Burnett, 1985](#)). Dans le modèle de ME calculé en 1993, la seule zone pouvant accommoder une protéine de si grande taille était une région charnière à la surface de la capsid. La protéine IIIa a donc été attribuée à cette densité plus ou moins cylindrique à l'interface entre 2 GONs de 2 facettes adjacentes (En rouge sur la [Figure 28](#)). La protéine IIIa possèderait aussi une

partie interne de plus petite taille (Stewart et al., 1993). Cette protéine à l'interface entre deux GONs aurait donc un rôle stabilisateur maintenant deux faces entre elles.



**Figure 28: Localisation schématique des protéines de capsides chez les Adénovirus humains, admise depuis 1993.**  
 (a) Vue externe d'une face de la capside d'un Adénovirus humain du sous-groupe C. (b) Vue interne d'une face de la capside d'un Adénovirus humain du sous-groupe C. Adapté d'après les travaux de Stewart et al., 1993.

Protéine IX. Seuls les virus du genre *Mastadenovirus* codent cette protéine. Elle est maturée par la protéase virale à partir d'un précurseur d'environ 15 kDa. Au delà de son rôle structural au sein de la capside, cette protéine aurait d'autres fonctions au cours du cycle viral. En effet, la protéine IX est la seule protéine structurale produite au cours de la phase intermédiaire du cycle viral. Elle possèderait un rôle trans-activateur de la transcription des ARNm viraux (Rosa-Calatrava et al., 2001). D'après les alignements, les prédictions de structure secondaires et les travaux de mutagenèse dirigée, la protéine IX possèderait deux domaines structuraux distincts. Le domaine N terminal semble impliqué dans l'incorporation et le maintien de la protéine au sein de la capside alors que le domaine C terminal quant à lui est prédict en hélices alpha et semble responsable de son activité transcriptionnelle. Selon les études de stœchiométrie réalisées en 1985, elle serait présente en 240 copies dans la capside mature. Les travaux de Furciniti et collaborateurs en 1989 avait déjà permis de la localiser à l'état trimérique sur la partie externe des GONs (en jaune sur la Figure 28). Par la suite cette observation avait été confirmée par Stewart et al. (Furciniti et al., 1989; Stewart et al., 1993). Elle serait impliquée dans la stabilisation des GONs et responsable de la forte cohésion existante entre les Hexons d'une facette.

Protéine VI. La forme mature de la protéine VI est issue du clivage d'un précurseur d'environ 27 kDa au niveau de deux sites identifiés. Sa maturation libère un propeptide N terminal de 33 acides aminés ainsi qu'un peptide C terminal de 11 acides aminés, le pVIc. Stewart et collaborateurs ont proposés en 1993 une localisation périventrale interne de la protéine VI (en vert sur la Figure 28). La reconstruction obtenue montrait une densité globulaire formée de trois lobes en dessous des Pentons. Chaque lobe correspondrait à un trimère de protéines VI reliant le Penton au cœur formé par le génome (Stewart et al., 1993). Quelques années plus tard, la protéine VI ou seulement des parties de celle-ci ont été identifiés dans plusieurs processus biologiques. Le peptide pVIc est en fait un co-facteur de la protéase virale. La structure cristallographique de la protéase virale en complexe avec pVIc a montré que ce peptide forme un brin  $\beta$  responsable d'une augmentation très forte de l'activité catalytique de la protéine (300 fois) (Ding et al., 1996). Wodrich et collaborateurs ont montré en 2003 que la protéine VI est nécessaire au trafic intracellulaire des Hexons par la voie Importine  $\alpha$  - Importine  $\beta$  (Wodrich et al., 2003). Enfin Wiethoff et collaborateurs ont montré en 2005 que la partie N terminale de la forme mature de la protéine VI est impliquée dans la sortie du virus de l'endosome lors de la phase très précoce d'infection. En effet, les acides aminés 36 à 54 du précurseur de la protéine VI montre une activité lytique pour les membranes (Wiethoff et al., 2005).

Protéine VIII. C'est la protéine mineure la moins documentée. Sa forme mature est issue du clivage protéolytique d'un précurseur de 25 kDa environ. Dès 1975, la protéine VIII avait été placé sur la face interne de la capsid (Everitt et al., 1975). Sa stœchiométrie au sein de la capsid a été établie à 120 copies (van Oostrum & Burnett, 1985) ainsi que la découverte d'un virus thermolabile possédant une mutation ponctuelle dans la séquence de la protéine VIII suggère que cette protéine possède un rôle important de stabilisation de la capsid (Liu et al., 1985). Malheureusement la reconstruction 3D en ME obtenue en 1993 par Stewart et collaborateurs n'avait pas permis de localiser cette petite protéine.

#### I.4 Structure et Organisation du génome des Adénovirus

Les Adénovirus contiennent tous une molécule d'ADN double brin linéaire et non segmentée dont la longueur est variable suivant le genre. Le génome le plus court est celui des *Siadenovirus*, inférieur à 30 kpb. Les *Atadenovirus* et les *Mastadenovirus* possèdent un génome de longueur intermédiaire d'environ 33 kpb. Le dernier genre des *Aviadenovirus* possède le génome le plus long avec 45 kpb (Figure 28). L'ADN viral n'est jamais nu ni dans la capsid, ni dans la cellule

pendant le cycle d'infection. Il est compacté et protégé par des protéines virales et forme une structure appelé nucléoïde au sein de la capsidé.

#### *I. 4.1 Structure du nucléoïde*

Le nucléoïde est composé de l'ADN viral en complexe avec plusieurs protéines virales : les protéines V, VII, Tp et X (aussi appelée protéine  $\mu$ ).

Les extrémités 5' du génome ne sont jamais libres. Elles sont en permanence liées de façon covalente à la protéine Tp pour « Terminal protein » (Rekosh et al., 1977). Cette protéine d'environ 70 kDa est présente chez tous les Adénovirus (Figure 29). Elle possède un rôle important de protection du génome contre la dégradation et permet sa réPLICATION au cours du cycle viral.

L'ADN viral est entièrement recouvert par la protéine VII qui permet son compactage au sein de la capsidé (Figure 29). Cette petite protéine (17 kDa) est issue d'une maturation par protéolyse d'un précurseur de 20 kDa environ. Sa stoechiométrie est estimée à plus de 850 copies par capsidé (van Oostrum & Burnett, 1985). La structure 3D du nucléoïde n'est pas encore très bien comprise mais il est clair que le génome forme des unités répétitives faisant penser au nucléosome. Celles-ci sont composées d'ADN viral associé à la protéine VII et ne semblent pas posséder une taille fixe puisque une digestion du génome à la nucléase micrococcale donne des fragments de taille variable de 50 à 300 pb, tous liés à la protéine VII (Shenk, 1990).

La plus petite des protéines appartenant au nucléoïde est la protéine X ou  $\mu$ . Celle-ci ne fait que 5 kDa dans sa forme mature et semble impliquée dans le compactage du génome (Anderson et al., 1989).

La dernière protéine appartenant au nucléoïde est la protéine V. Elle n'est présente que chez les *Mastadénovirus* (Benko & Harrach, 2003). Cette protéine de 40 kDa environ serait impliquée dans le maintien du nucléoïde au niveau de la capsidé. En effet, la protéine V semble relier la protéine VII à la Base du Penton et la protéine VI (Figure 29).

#### *I. 4.2 Unités transcriptionnelles et organisation du génome*

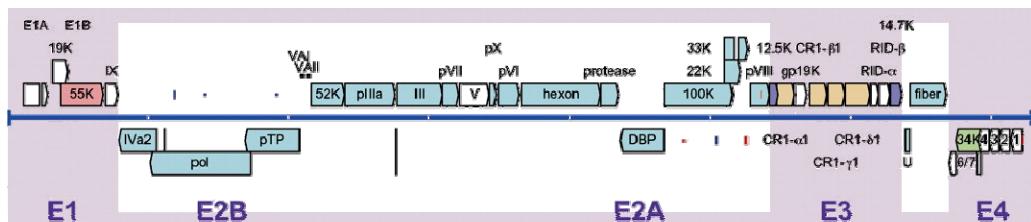
La description de l'organisation du génome des Adénovirus est possible depuis le séquençage complet du génome de plusieurs membres de chaque genre. Les premiers génomes à avoir été séquencés sont ceux de l'Adénovirus humain de type 2 en 1984 (Roberts et al., 1984) puis celui du type 5 en 1992 (Chroboczek et al., 1992). Le séquençage de la grande majorité des génomes a été

réalisé à partir du milieu des années 90. C'est le cas du FAdV-1 et du CAdV-2 en 1996 (Chiocca et al., 1996; Davison et al., 2003).

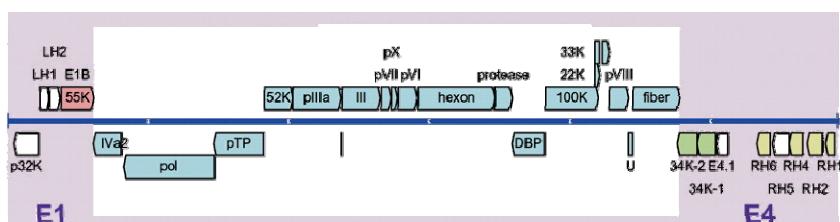
Malgré des différences marquées existantes entre les genres, le génome est toujours flanqué de deux origines de réplication au niveau de séquences inversées répétées (ITR) situées aux extrémités 5' et 3' du génome. Le génome des Adénovirus reste organisé en plusieurs parties transcriptionnellement distinctes, transcrtes dans les 2 sens et de façon différentielle au cours des phases précoce, intermédiaires et tardives du cycle cellulaire. Les gènes s'exprimant au cours de la même phase du cycle viral sont en général regroupés dans la même zone du génome sous forme d'unité transcriptionnelle. Chaque unité est désignée par une lettre E (Early) ou L (Late) suivant si celle-ci s'exprime, respectivement, pendant la phase précoce ou tardive du cycle cellulaire. Le génome possède en général plusieurs unités transcriptionnelles précoce (E1, E2, E3 et E4 chez les *Mastadenovirus*) et une unité tardive (L). La seule zone du génome où l'organisation est conservée à travers les différents genres d'Adénovirus est la partie centrale (encadré en blanc sur la [Figure 29](#)). Cette région code pour les précurseurs des protéines structurales (pII, pIII, pIV, pIIIa, pVI, pVIII) ainsi que pour les protéines nécessaires à la réplication du génome (pTp, polymérase et DBP) et à la maturation de la capsid (100K et protéase). L'ensemble de ces régions est transcrit en ARNm viraux par l'ARN polymérase de type 2. Seule la région VA (Viral ARN), présente uniquement chez les *Mastadenovirus*, est transcrte par l'ARN polymérase de type 3. Cette petite région code pour un ou deux ARN viraux (suivant le sérotype) dont la fonction cellulaire n'est pas encore très claire ([Figure 29](#)). Ceux-ci interviendraient tout de même dans la modulation de la réponse immunitaire de l'hôte ainsi que dans la transcription des ARNm viraux (Mathews & Shenk, 1991).

Le génome des *Mastadenovirus* (Adénovirus humains et canins) est organisé en cinq unités de transcription précoce dénotées E1A, E1B, E2, E3 et E4, en 2 unités intermédiaires IX et IVa2 et en une unité tardive L. Par convention, le génome est représenté avec la région E1 du coté gauche ([Figure 28](#)). Les régions E1A, E1B, IX, L et VA sont transcrtes de gauche à droite alors que les unités E2, E4 et IVa2 dans le sens opposé. La région E1 est la première région transcrte lorsque l'ADN viral entre dans le noyau de la cellule infectée. Les régions précoce E codent pour des protéines préparant la cellule à la réplication de l'ADN viral alors que l'unité tardive contient plutôt des gènes impliqués dans l'assemblage et la maturation de la capsid ([Figure 29](#)). Les particularités du génome des *Aviadenovirus* seront discutées en partie I.7 de cette introduction.

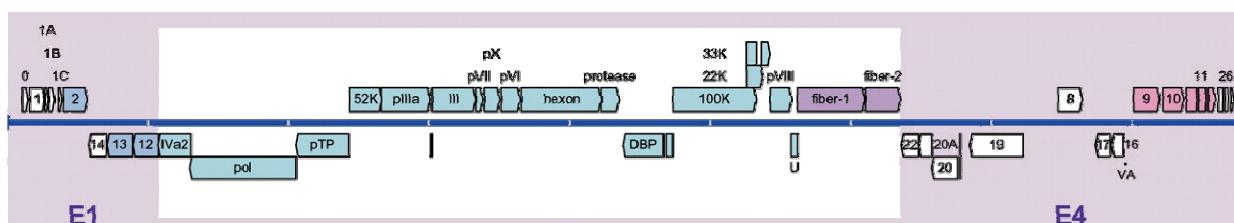
### Mastadenovirus: SAdV-25 (HAdV-E)



### Atadenovirus: OAdV-7 (OAdV-D)



### Aviadenovirus: FAdV-1 (FAdV-A)



### Siadenovirus: TAdV-3 (TAdV-A)

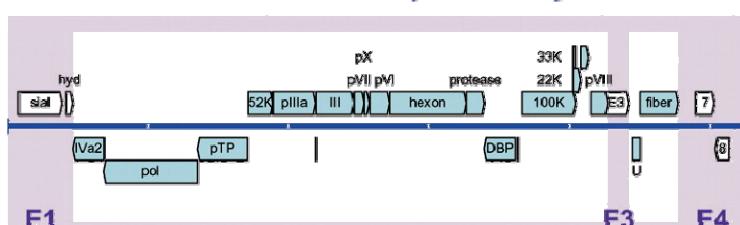


Figure 29: Organisation du génome des virus de la famille des Adenoviridae (Davison et al., 2003).

Le génome est représenté par une barre centrale de couleur bleue. Au sein de cette barre, l'espace entre deux traits blancs représente un intervalle de 5 kpb. Les cadres de lecture ouverts communs aux différents genres sont marquées en cyan. Les cadres de lecture ouverts ne possédant pas d'équivalents dans un autre genre sont notés en blanc. Les zones encadrées en blanc possèdent une organisation commune au sein de la famille des Adenoviridae.

## I.5 Le cycle viral

Seul le cycle cellulaire des HAdV-2 a été intensivement étudié. En effet ces deux virus sont modèles pour l'étude des Adénovirus. Ils sont produits assez facilement en laboratoire et de nombreux mutants ont pu être créés. Les données concernant le cycle cellulaire d'autres souches d'Adénovirus notamment les souches aviaires ou canines sont faibles mais le processus général d'infection reste conservé à travers la famille des *Adenoviridae*.

Les Adénovirus peuvent réaliser trois types de cycles cellulaires différents. Le premier est le cycle lytique lorsque le virus infecte une cellule permissive. Il aboutit à la production de nouvelles particules virales et à la lyse de la cellule infectée. Lorsque le virus pénètre dans une cellule non permissive, le déroulement complet du cycle lytique ne peut pas aboutir. Aucune particule virale n'est produite : c'est le cycle abortif. Enfin un dernier type de cycle cellulaire, le cycle transformant, peut être observé lorsque le virus infecte une cellule non permissive, plus rarement une cellule permissive. Dans ce cas, le virus immortalise la cellule infectée et déclenche le développement de tumeurs cancéreuses.

Dans la suite de mon exposé, je vais uniquement m'intéresser à la description du cycle lytique chez les adénovirus humains. Ce cycle d'infection dure entre 20 et 24h dans des cellules HeLa. Il est le seul entraînant la production de nouvelles particules virales actives. Schématiquement, ce cycle est organisé en deux grandes phases délimitées par la réplication du génome viral. La phase appelée précoce débute au moment où le virus interagit avec sa cellule hôte jusqu'à la réplication de l'ADN viral. L'événement de réplication constitue la phase qualifiée d'intermédiaire. Enfin la phase tardive commence à la réplication du génome et se poursuit jusqu'à la lyse de la cellule infectée. Dans la réalité, les transitions entre les différentes phases du cycle d'infection ne sont aussi marquées.

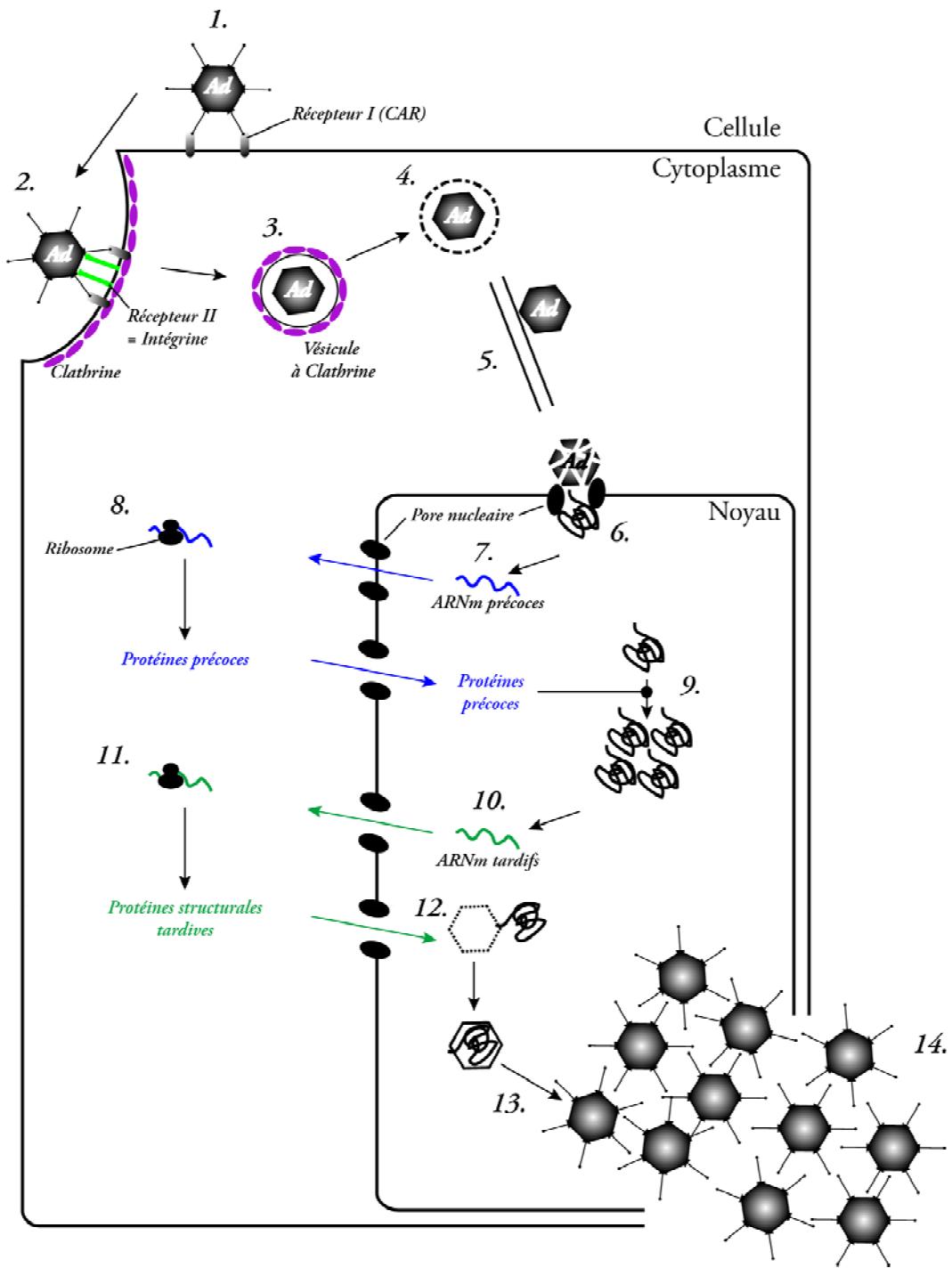


Figure 30: Le cycle viral de l'Adénovirus de type 2 et 5.

1. Le virus s'attache au récepteur primaire via la tête de la fibre. 2. Le virus est reconnu par son récepteur secondaire, une intégrine, via la boucle RGD de la base du Penton. 3. Le virus est internalisé dans une vésicule à clathrine. 4. Cette vésicule fusionne avec un lysosome. Le virus se désassemble partiellement et s'échappe de l'endolysosome par acidification du pH. 5. la capside partiellement désassemblée voyage via les microtubules jusqu'au nucléopore. 6. Le génome viral entre dans le noyau via le nucléopore 7. Les ARNm précoce sont transcrits. 8. puis traduits en protéines précoce. 9. Certaines protéines précoce retournent dans le noyau et déclenchent la réplication du génome. 10. Les ARNm tardifs sont produits. 11. Ces ARNm sont traduits en protéines tardives dans le cytoplasme. 12. Les protéines tardives essentiellement structurelles retournent dans le noyau et forment des procapsides qui vont intégrer les génomes néo-synthétisés. 13. Les procapsides sont ensuite maturées par protéolyse spécifique et vont s'accumuler dans le noyau. 14. Le virus est libéré par lyse de la cellule infectée.

### *I. 5.1 Entrée du virus dans la cellule et phase précoce d'infection*

Le premier événement de la phase précoce est l'attachement du virus à sa cellule hôte. La tête de la fibre est responsable de la liaison du virus à la membrane cellulaire par l'intermédiaire d'un récepteur membranaire primaire (1. sur la [Figure 30](#)). Le type de récepteur primaire dépend du sérotype de virus infectant la cellule. En effet, le premier récepteur identifié fût CAR (Coxsackie and Adenovirus Receptor) ([Bergelson et al., 1997](#)). Celui-ci est utilisé par les sérotypes humains du sous-groupe A, C et F ([Figure 22](#)). CAR est présent dans de nombreux tissus épithéliaux donnant ainsi un tropisme important aux Adénovirus capables de s'y attacher. Le sous-groupe D utilise l'acide sialique comme récepteur primaire ([Arnberg et al., 2000](#)). Ces sucres sont présents en partie terminale des chaînes carbohydrates des glycoprotéines membranaires de la cellule. Les virus du sous-groupe B comme l'HAdV-3 se lie au récepteur membranaire CD46 ([Defer et al., 1990; Gaggar et al., 2003](#)). Ce virus peut aussi utiliser les récepteurs membranaires CD80/CD86 comme récepteur primaire ([Short et al., 2004](#)).

Après attachement de la fibre, la Base du Penton reconnaît une intégrine via sa boucle RGD (2. sur la [Figure 30](#)). Les intégrines reconnues par les Adénovirus humains sont principalement de deux types :  $\alpha_v\beta_3$  et  $\alpha_v\beta_5$  ([Wickham et al., 1993](#)). Ce récepteur membranaire est constitué de deux types de chaînes,  $\alpha$  et  $\beta$  formant un hétérodimère. La boucle RGD du Penton interagit en fait avec la sous-unité  $\alpha$  des intégrines.

L'interaction avec les intégrines déclenche l'entrée du virus par endocytose dans une vésicule à clathrines (3. sur la [Figure 30](#)). La simple reconnaissance des intégrines cellulaires peut aussi permettre au virus de pénétrer dans la cellule via un processus de macropinocytose ([Meier & Greber, 2004](#)). Immédiatement après son entrée, le virus se retrouve enfermé dans une vésicule endosomale. Pour injecter son ADN dans le noyau cellulaire, ce virus doit s'en échapper et voyager jusqu'au pore nucléaire. L'acidification du pH dans l'endosome par fusion avec une vésicule lysosomale est responsable de l'entrée du virus dans le cytoplasme. La baisse du pH déclenche la déstabilisation et le désassemblage partiel de la capsid et plus particulièrement la libération de la protéine VI ([Greber et al., 1993](#)). La protéase virale, présente dans la capsid mature aurait ici un rôle important dans la dégradation de certaines protéines mineures notamment de la protéine VI ([Greber et al., 1996](#)). Cette protéine après dégradation et en présence d'un pH acide, semble avoir une activité lytique sur

les membranes ce qui entraînerait la destruction de la vésicule (4. sur la [Figure 30](#)) et la libération du virus dans le cytoplasme ([Wiethoff et al., 2005](#)).

Dans le cytoplasme, le virus subit ensuite un déshabillage progressif et un transport via les microtubules jusqu'au pore nucléaire où l'ADN viral entre dans le noyau ([Greber et al., 1997](#); [Greber et al., 1993](#); [Leopold et al., 2000](#); [Mabit et al., 2002](#)).

L'entrée du génome est marquée par l'expression des gènes précoces. (7. et 8. sur la [Figure 30](#)) L'expression de ces gènes permet de préparer la cellule à la réplication virale, de protéger la cellule infectée contre les mécanismes de défense immunitaire et de synthétiser les protéines virales nécessaires à la réplication de l'ADN viral. Les premières protéines exprimées sont les protéines E1A. Ces protéines précoces avec l'aide des produits des régions E1B et E4 font entrer la cellule en phase S (phase du cycle de division cellulaire normal durant laquelle la réplication de l'ADN cellulaire a lieu). L'expression des régions E3 et VA permettent au virus de lutter contre certains mécanismes immunitaires. Enfin les protéines E1A et E1B activent l'expression de toutes les régions précoces permettant ainsi la production des protéines virales nécessaires à la réplication (produits des régions E2 du génome viral). La fin de la phase précoce est marquée par une accumulation de protéines précoces E2.

### *I. 5.2 Réplication du génome viral*

Le début de la réplication du génome viral marque le début de la phase tardive entre 8 à 10h post-infection dans des cellules HéLa (9. sur la [Figure 30](#)).

Les régions E2A et E2B, exprimées au cours de la phase précoce, codent pour des protéines indispensables à la réplication virale comme la polymérase virale ADN-dépendante, la DBP (DNA Binding Protein) et le précurseur pTp de la protéine terminale.

La réplication débute toujours au niveau des ITR présentes aux extrémités du génome. La polymérase virale utilise le précurseur pTp comme amorce pour la réplication. L'initiation de la réplication débute par la formation d'une liaison covalente entre un résidu sérine de la pTp et un nucléotide désoxycytidine monophosphate (dCMP). La polymérase recopie ensuite le brin matrice d'ADN viral pendant la phase d'elongation. Durant cette phase, la protéine DBP se fixe sur l'ADN simple brin formant ainsi une chaîne de protéines qui entraîne le déroulement de la molécule d'ADN ([Dekker et al., 1997](#); [Monaghan et al., 1994](#)). La polymérase peut ainsi facilement recopier le brin matrice d'un bout à l'autre après son détachement de la pTp.

L'expression des régions IVa2 et IX se produit pendant une phase qualifiée d'intermédiaire qui débute après la réplication de l'ADN viral mais avant l'expression des protéines tardives. Ces protéines, outre leur rôle dans l'assemblage de la capsidé semblent avoir de nombreuses fonctions au cours du cycle viral notamment dans l'activation des gènes tardifs (Rosa-Calatrava et al., 2001; Tribouley et al., 1994; Zhang et al., 2001).

#### *I. 5.3 Phase tardive d'infection et Lyse de la cellule infectée*

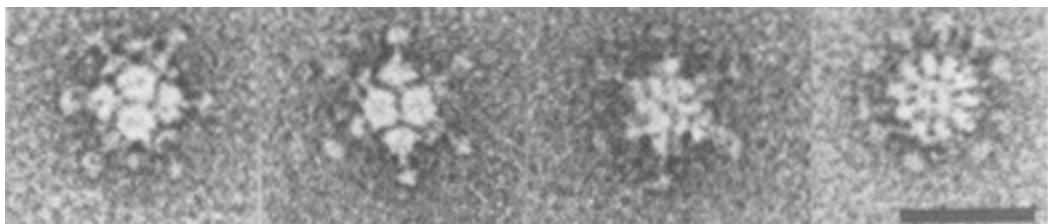
La réplication de l'ADN viral marque donc le début de la phase tardive (9. sur la Figure 30). Durant cette phase, les régions tardives du génome viral sont exprimées (10. sur la Figure 30). Les protéines de capsidé exceptée la protéine IX sont produites et importées dans le noyau ainsi que cinq autres protéines nécessaires à l'assemblage et à la maturation de la capsidé : les protéines 100K, 52K, 33K, 22K et la protéase virale (Figure 29 et 11. sur la Figure 30). Les protéines de capsidé s'assemblent dans le noyau, avec l'aide de la chaperonne virale 100K, pour former des procapsides dans lesquelles le génome viral néo-synthétisé va être incorporé (implication de la protéine IVa2 et 52K, (Perez-Romero et al., 2006; Zhang et al., 2001) (12. sur la Figure 30). Les capsides vont ensuite subir une maturation par dégradation protéolytique par la protéase virale et s'accumuler dans le noyau de la cellule infectée (13. sur la Figure 30). Enfin la protéine ADP (Adenovirus Death Protein), seule protéine de la région E4 à être exprimée pendant la phase tardive, va déclencher la lyse de la cellule infectée ce qui libère les nouveaux virions dans le milieu extracellulaire (14. sur la Figure 30). Les nouveaux virions fonctionnels libérés dans l'espace intercellulaire peuvent maintenant infecter de nouvelles cellules.

#### *I. 5.4 Conclusion sur le cycle virale de l'Adénovirus*

Le cycle viral des Adénovirus de type 2 et de type 5 est le plus connu. Malgré le grand nombre d'études réalisés sur le cycle cellulaire des Adénovirus humains, on ne sait que très peu de choses sur les mécanismes cellulaires mis en jeu. On ne connaît pas non plus l'impact de ces virus sur la modulation de la réponse immunitaire de l'hôte, ni son rôle dans l'induction de potentiel cancer. De nombreuses études doivent être encore menées pour percer les mystères du cycle cellulaire des Adénovirus à la fois chez l'Homme mais aussi chez les mammifères.

## I.6 Structure du dodécaèdre et particularités du dodécaèdre de type 3

Au cours de la phase tardive, les protéines structurales de la capsid sont produites en très grandes quantités et s'accumulent dans le cytoplasme de la cellule infectée. Chez certains sérotypes, il arrive que certaines de ces protéines s'auto-assemblent. C'est le cas de la Base du Penton et de la Fibre qui interagissent pour former une sous-particule virale appelé Dodécaèdre. In vivo, ces structures ont été observées pour la première fois à partir de cellules infectées par les sérotypes 3 puis 4, 7, 11 et 9/15 (Neurath & Rubin, 1968; Norrby, 1966; Norrby, 1968; Norrby et al., 1967; Norrby & Wadell, 1967). Les sérotypes 1, 2, 5, 6, 12, 16, 40 et 41 ne sont pas connus pour former des dodécaèdres in vivo mais la structure de la Base du Penton du sérotype 2 a tout de même été résolue sous forme dodécaédrique (Zubieta et al., 2005). La Base du Penton semble donc capable de former des dodécaèdres, au moins in vitro, chez certains sérotypes. Cette particule d'environ 250 Å de diamètre est formée par l'assemblage non covalent de 12 Pentons (Base du Penton + Fibre) et porte le nom de dodécaèdre en référence au solide de Platon qui lui correspond (cf. Partie 1, I.2.5). Les 12 Pentons sont reliés par la présence d'axe de symétrie d'ordres 5, 3 et 2.



**Figure 31: Images en coloration négative du dodécaèdre de type 3 (Schoehn et al., 1996).**

De gauche à droite : Dodécaèdre de type 3 observés en coloration négative selon l'axe d'ordre 2, 2, 3 et 5. La barre d'échelle représente 500 nm.

Au cours de ma thèse, je me suis intéressée au dodécaèdre de l'Adénovirus de type 3. Etant le premier à avoir été observé, c'est le plus étudié des dodécaèdres. Cette structure se forme préférentiellement après protéolyse d'au moins 40 acides aminés en N terminal. Schoehn et collaborateurs en 1996 ont obtenu la première reconstruction 3D en ME du dodécaèdre de l'HAdV-3 à 20 Å de résolution (Schoehn et al., 1996). Ce modèle avait permis de décrire la cavité interne du dodécaèdre ainsi que de proposer un modèle pour l'attachement de la fibre à la Base du Penton en absence de données cristallographiques (disponibles seulement depuis 2005).

L'Adénovirus humain de type 3 appartient au sous-groupe B (Figure 22). Il possède une fibre beaucoup plus courte et rigide que les sérotypes 2 et 5 (Wu et al., 2003). Celle-ci possède 6 répétitions et ne mesure qu'environ 60 Å de long. L'HAdV-3 reconnaît CD46 et CD80/86 comme

récepteurs primaires (Cf. Introduction). La sous particule virale semble se comporter différemment de la particule virale dont elle est issue. En effet, le dodécaèdre aurait en plus la capacité de lier les héparanes sulfates à la surface de la cellule par l'intermédiaire de la Base du Penton (Vives et al., 2004). Les héparanes sulfate sont des sucres présents au niveau des glycoprotéines de surface de la cellule.

La description d'une cavité interne au dodécaèdre ainsi que de voies alternatives d'entrée dans la cellule a entraîné le développement et l'étude du dodécaèdre en tant que vecteur de thérapie à la fin des années 90 (Fender et al., 1997).

## I.7 Les Adénovirus Aviaires

Les Adénovirus aviaires ont été identifiés comme agent infectieux en 1957 (Yates & Fry, 1957). Ils constituent le genre des *Aviadenovirus*. Il existe 4 sérotypes distincts : le FAdV-1 ou CELO, FAdV-4, FAdV-9 et FAdV-10. Peu de temps après leur découverte, les travaux de Sarma et collaborateurs ont montré que CELO induisait des tumeurs chez les bébés hamsters (Sarma et al., 1965). Cette observation a suscité l'intérêt de la communauté scientifique et a fait de CELO le plus étudié des Adénovirus aviaires.

Malgré la longueur importante de leur génome, les adénovirus aviaires ne contiennent pas de région E3. Ces virus codent, à la place, pour une quantité de protéines précoces, spécifiques à leur genre, appartenant aux régions E1 et E4 (ORF en blanc sur la Figure 29) (Davison et al., 2003). Les régions contenant ces gènes sont donc situées aux extrémités du génome. La région E4 est particulièrement étendue et code entre autres, comme chez les *Mastadenovirus*, pour un petit ARN VA transcrit par l'ARN polymérase de type III dont la fonction cellulaire est encore inconnue chez les virus aviaires.

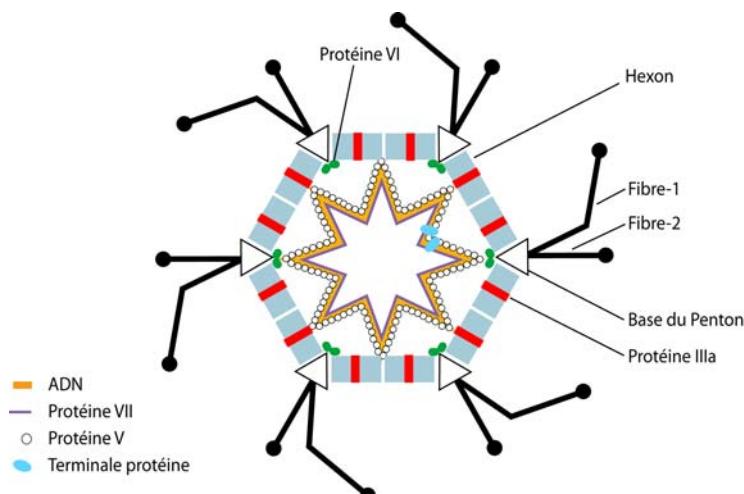


Figure 32: Représentation schématique en coupe de la capsid du CELO

La région centrale du génome, exprimée au cours de la phase tardive d'infection, est particulièrement intéressante en termes de structure de la capsid. En effet, cette région ne contient pas de gène codant pour le précurseur de la protéine IX mais contient deux gènes distincts de fibre : fibre-1 et fibre-2 (Figure 29). La capsid ne contiendrait donc aucune protéine IX en surface contrairement aux Adénovirus humains. Les Hexons ne forment pas de GONs lors du désassemblage de la capsid. Le génome code pour les protéines mineures structurales IIIa, VI et VIII mais en absence de données structurales, leur localisation reste encore inconnue. On suppose tout de même

que leur positionnement est identique à celui des Adénovirus humains (Figure 32). L'étude structurale de cette capsid, par comparaison avec les adénovirus humains pourrait permettre la localisation plus précise de la protéine IX et IIIa. Le Penton est formé de la Base du Penton lié aux deux types de fibre. En effet, des observations en ME montrent que la fibre-1 (longue) et la fibre-2 (courte) sont présentes toutes les deux à chacun des sommets de la capsid (cf. III.2). La fibre-1 est homologue à la fibre des Adénovirus humains alors que la fibre-2 ne possède que peu de similarités de séquences avec d'autres fibres courtes d'Adénovirus. Ces caractéristiques ont de conséquences importantes sur le tropisme et le cycle cellulaire du virus. Les récepteurs primaires et secondaires n'ont pas encore été identifiés. La fibre-1 semble tout de même capable d'interagir avec CAR mais la Base du Penton ne possédant pas de séquence RGD ne peut pas interagir avec une intégrine cellulaire pour l'internalisation du virus. Ceci affecte le tropisme du virus. En effet celui-ci est capable d'infecter des cellules humaines par l'intermédiaire de la fibre-1 mais le mécanisme d'internalisation n'a pas encore été élucidé (Tan et al., 2001). La structure de la tête de la fibre-1 a été résolue en 2007 par une équipe espagnole et celle de la fibre-2 a été déterminée dans notre laboratoire (Guardado-Calvo et al., 2007, El Bakkouri et al., 2008). Récemment la structure de l'Hexon du CELO a aussi été résolue par Xu et collaborateurs à 4 Å de résolution (Xu et al., 2007). J'ai donc pu utiliser cette structure à la fin de ma thèse pour essayer de reconstituer un modèle quasi-atomique de la capsid du CELO par « fitting » (cf. IV.2).

## I.8 Les Adénovirus Canins

Ils existent deux sérotypes d'adénovirus canins : le CAdV-1 et le CAdV-2. Ces virus appartiennent au genre des *Mastadenovirus*. L'organisation de leur génome et de leur capsid ne diffère pas de celle des Adénovirus humains. J'ai étudié au cours de ma thèse le sérotype le moins virulent, le CAdV-2. Il existe très peu de différences entre les Adénovirus canins et humains mais il semble tout même important de s'y intéresser quand on étudie la structure de la capsid des Adénovirus pour une raison majeure. L'étude de ce virus possède un fort intérêt en termes de vecteurs pour la thérapie (cf. partie suivante). La tête de la fibre du CAV-2 a la capacité d'interagir avec CAR puisque celle-ci a été cristallisée en complexe avec le domaine d'interaction de CAR avec les fibres d'Adénovirus (Seiradake et al., 2006). Ce virus semblerait donc posséder un tropisme pour les cellules humaines par l'intermédiaire de CAR sans entraîner de réactions immunitaires, les structures de l'Hexon humain et canin étant assez différentes. Il reste tout de même dans ce cas, des

problèmes d’hemaglutination importante. La capsid du CAdV-2 et a fortiori, celle du sérotype 1 montre le même réseau de protéines que chez les Adénovirus humains mais les séquences de ces protéines (surtout celle de la protéine IX) montrent des différences importantes qui ont certainement des conséquences sur la structure de la capsid que je discuterai en partie IV.2.

### I.9 Les Adénovirus, vecteurs pour la thérapie génique et anti-cancer

Les Adénovirus ont été largement caractérisés puis manipulés en laboratoire depuis leur découverte dans les années 50. L’essentiel des informations de cette partie est tirée d’un revue publiée en 2000 par W.C. Russell ([Russell, 2000](#)). Les Adénovirus délivrent leur génome directement dans le noyau dans un large panel de tissus et se répliquent de façon importante même dans des cellules quiescentes ce qui fait de ce virus un candidat de choix pour la thérapie génique et anti-cancer.

La thérapie génique se rapporte aux traitements des maladies via l’utilisation de gènes thérapeutiques sous forme d’ARN ou d’ADN dans le cas des Adénovirus. Le ou les gènes thérapeutiques sont apportés par un vecteur, principalement, de deux manières différentes : si l’on peut prélever le tissu malade à traiter, on procède dans ce cas à un traitement dit *ex vivo* en culture puis le tissu traité est réinjecté chez le patient par voie systémique ou locale. Dans l’autre cas, le vecteur est injecté au patient et doit apporter le ou les gènes thérapeutiques directement au niveau des cellules à traiter. C’est dans cette approche dite *in vivo* que les Adénovirus ont beaucoup été utilisés et développés. Quatre types de vecteurs viraux sont actuellement utilisées : les Adénovirus, les virus associés aux Adénovirus (AAV), les Lentivirus, les Herpèsvirus. Leur utilisation chez l’homme entraîne de nombreux problèmes comme le ciblage efficace du tissu à traiter mais aussi des réponses immunitaires importantes et difficiles à contrôler. Les premiers Adénovirus utilisés possédaient une délétion des régions E1 et E3 du génome permettant ainsi l’insertion d’un ou plusieurs gènes thérapeutiques sur environ 6,5 Kpb. L’élimination de la région E1 évite que le vecteur ne se réplique dans son hôte. En effet, il est nécessaire de limiter le plus possible la réplication virale lors d’un traitement par thérapie génique car celle-ci est difficile à contrôler et peut entraîner des réponses immunitaires fortes et non souhaitées chez les patients. Par la suite, afin de diminuer les réponses immunitaires, les générations de vecteurs adénoviraux étaient partiellement voir totalement délestés de la région E2 de leur génome. Malgré cela, des réponses immunitaires fortes ont été constatées jusqu’à un incident tragique qui se produisit lors d’un essai de thérapie génique sur l’homme. Un jeune homme atteint d’une déficience enzymatique décéda des suites d’une injection intra hépatique

d'Adénovirus. Cet incident freina fortement l'engouement de la communauté scientifique quant à l'utilisation des Adénovirus comme vecteur de thérapie chez l'homme mais mena au développement d'une dernière génération de vecteurs dits « gutless » ne possédant plus que les séquences inversées répétées et les séquences d'encapsidation. Ces souches nécessitent alors l'aide d'une souche adénovirale dite « helper » pour se répliquer. De problèmes ont encore été rencontrés avec ces virus notamment la présence de contamination par les Adénovirus helper ainsi que la faible stabilité des souches gutless. La mort du jeune homme des suites d'une forte activation immunitaire dénote aussi d'un problème de ciblage lié au tropisme des Adénovirus. En effet, ces virus possèdent un tropisme important pour tous les tissus de type épithélium. Il est difficile de traiter certains tissus comme le tissu nerveux. Afin de traiter durablement un patient à l'aide d'un gène thérapeutique, il est nécessaire que celui-ci s'intègre au génome du patient. Ce n'est pas le cas des Adénovirus où le transgène ne peut pas s'intégrer au génome. Ces différentes raisons font que l'Adénovirus n'est plus un candidat de choix pour la thérapie face à certains lentivirus ou AAV. L'induction d'une forte réponse immunitaire et l'absence d'intégration du transgène peuvent tout de même présenter des avantages en terme de thérapie. En effet, si on cherche à lutter contre un cancer, il est intéressant de stimuler les défenses immunitaires du patient. Au cours de ces dernières années, de nombreux efforts ont été fait pour le développement des Adénovirus comme vecteur pour la thérapie anti-cancer. En effet, ces virus sont utilisés pour l'expression de gènes suppresseurs de tumeurs comme le gène p53 induisant ainsi l'apoptose des cellules tumorales. De plus, il a été observé que certaines souches d'Adénovirus possédant une mutation au niveau du gène E1 55K auraient un effet oncolytique. Des Adénovirus exprimant des gènes suicides en combinaison avec des vecteurs exprimant des cytokines ainsi qu'un traitement par chimiothérapie ou radiothérapie ont donné des résultats prometteurs en termes de réduction des tumeurs. Tous ces traitements, aussi bien la thérapie anti-cancer que la thérapie génique, nécessitent une modification du ciblage et du tropisme cellulaire du virus pour une meilleure efficacité. Malgré les nombreuses études fondamentales réalisées depuis la découverte des Adénovirus, peu de connaissances sur sa structure ainsi que sur son cycle de réplication sont disponibles. Le développement et la mise au point d'un vecteur passeront par une connaissance approfondie de la structure de la capsid ainsi que du rôle des protéines qui la compose.

## II. *Le modèle quasi atomique de la capsid de l'Adénovirus humain de type 5*

Cette partie de ma thèse est consacrée à l'étude structurale en MET de la capsid de l'Adénovirus humain de type 5. La dernière reconstruction 3D obtenue à 20 Å de résolution en MET a été publiée en 1997 (Stewart et al., 1997). Ces travaux ont permis d'établir un schéma général de la structure de la capsid mais depuis cette date, aucune reconstruction 3D à plus haute résolution n'a été publiée malgré les grands progrès réalisés en informatique et le développement des MET. Nous nous sommes donc intéressés à la structure de la capsid afin d'apporter des connaissances plus précises sur la structure de la capsid du HAdV-5. J'ai choisi dans ce manuscrit de ne pas apporter de précisions supplémentaires à l'article publié dans le Journal EMBO mais de réaliser un complément de Matériel et Méthodes sur l'analyse d'images et la reconstruction 3D des objets icosaédriques en MET.

### II.1 Article 1

#### “A quasi-atomic model of human adenovirus type 5 capsid”

Auteurs: C.M.S. Fabry, M. Rosa-Calatrava, J.F. Conway, C. Zubieta, S. Cusack, R.W.H. Ruigrok & G. Schoehn.

Publié en 2005 dans *Embo Journal*, Vol.24, pages 1645-1654.

Les coordonnées du modèle quasi-atomique ont été déposées dans la PDB sous les codes d'accès 2BVI et 2BLD pour respectivement, la position des 4 Hexons de l'UA et de la Base du Penton, avec les 60 matrices de rotation nécessaires à la reconstitution du modèle quasi-atomique. Les modèles 3D de microscopie électronique ont été déposés dans la base de données EMDep sous les codes d'accès 1111, 1112 et 1113 pour, respectivement, les échantillons muft, sift et WT.

# A quasi-atomic model of human adenovirus type 5 capsid

Céline MS Fabry<sup>1,2</sup>, Manuel Rosa-Calatrava<sup>3</sup>, James F Conway<sup>4</sup>, Chloé Zubieta<sup>2</sup>, Stephen Cusack<sup>2</sup>, Rob WH Ruigrok<sup>1,2</sup> and Guy Schoehn<sup>1,2,\*</sup>

<sup>1</sup>Laboratoire de Virologie Moléculaire et Structurale, FRE 2854 CNRS-Université Joseph Fourier, Grenoble Cedex, France, <sup>2</sup>EMBL Grenoble Outstation, Grenoble Cedex, France, <sup>3</sup>Laboratoire de Virologie et Pathogénèse Virale, Faculté de Médecine et Institut Fédératif de Recherche RTH Laennec, Lyon, France and <sup>4</sup>Institut de Biologie Structurale, Grenoble Cedex, France

Adenoviruses infect a wide range of vertebrates including humans. Their icosahedral capsids are composed of three major proteins: the trimeric hexon forms the facets and the penton, a noncovalent complex of the pentameric penton base and trimeric fibre proteins, is located at the 12 capsid vertices. Several proteins (IIIa, VI, VIII and IX) stabilise the capsid. We have obtained a 10 Å resolution map of the human adenovirus 5 by image analysis from cryo-electron micrographs (cryoEMs). This map, in combination with the X-ray structures of the penton base and hexon, was used to build a quasi-atomic model of the arrangement of the two major capsid components and to analyse the hexon–hexon and hexon–penton interactions. The secondary proteins, notably VIII, were located by comparing cryoEM maps of native and pIX deletion mutant virions. Minor proteins IX and IIIa are located on the outside of the capsid, whereas protein VIII is organised with a  $T=2$  lattice on the inner face of the capsid. The capsid organisation is compared with the known X-ray structure of bacteriophage PRD1.

The EMBO Journal (2005) 24, 1645–1654. doi:10.1038/sj.emboj.7600653; Published online 21 April 2005

Subject Categories: structural biology; microbiology & pathogens

Keywords: cryo-electron microscopy; minor capsid protein; structure

## Introduction

Adenoviruses are double-stranded DNA viruses that are found in all vertebrates (Davison *et al.*, 2003). Over 50 human serotypes cause a variety of rather mild diseases such as colds, eye infections and diarrhoea (Horwitz, 2001). Adenoviruses have an icosahedral capsid with a pseudo  $T$  number of 25 (Stewart *et al.*, 1991) and a total protein mass of around 125 MDa (van Oostrum and Burnett, 1985). The major capsid components are the hexon, a trimeric protein

with a hexagonal shape at its base, and the penton that is a noncovalent complex between the pentameric penton base and the trimeric fibre protein. In all, 240 hexons form the 20 facets of the icosahedron, whereas the pentons form and project from the 12 vertices. The fibre binds to host cell receptors with its C-terminal knob domain, whereas protein loops that extend from the penton base and that contain an Arg–Gly–Asp (RGD) sequence (except for the enteric adenovirus 40 and 41; Albinsson and Kidd, 1999) are required to bind to a secondary cell receptor (integrin  $\alpha v \beta 3/5$ ) to trigger endocytosis of the virus (Greber *et al.*, 1993; Wickham *et al.*, 1993). The capsid also contains minor components, proteins IIIa, VI, VIII and IX, that glue the hexons and pentons together at specific positions in the capsid.

Adenoviruses are also widely used as vectors for gene transfer and in anticancer protocols (Russell, 2000). However, their use as therapeutic agents for humans is problematic due to the massive immune response against the capsid proteins, due to the limited period of expression of trans-genes owing to nonintegration of viral DNA into the host genome and due to the difficulty in targeting specific cell types (Russell, 2000; Horwitz, 2001). Several of these problems may be overcome by modifying one or several of the viral capsid proteins, but for this a detailed knowledge of the capsid structure must be available.

Adenoviruses produce para-crystalline inclusions in the nuclei of infected cells (Boulanger *et al.*, 1970), but it has not yet been possible to grow crystals of purified virus suitable for high-resolution structure determination. The atomic structure of the hexon of human adenovirus type 5 (Ad5) (Rux and Burnett, 2000) is available, as is that of the human adenovirus type 2 (Ad2) penton base (Zubieta *et al.*, 2005). Thus, the shape and fold of these capsomers are known, as well as the interaction of the monomers in the hexon trimer and penton base pentamer. The penton base structure also gives information regarding the contacts between the bases in the penton base dodecahedron, a subviral particle consisting of 12 penton bases (Schoehn *et al.*, 1996; Fender *et al.*, 1997; Zubieta *et al.*, 2005), and a view of how the trimeric fibre interacts with the penton base (Zubieta *et al.*, 2005). Information about the architecture of the intact Ad capsid has come from three-dimensional (3D) reconstructions of virus particles imaged by cryo-electron microscopy (cryoEM). The first reconstruction, at 35 Å resolution, showed the complex organisation of the capsid with the overall positions of the major capsomers and the stabilising role of some of the minor capsid proteins (Stewart *et al.*, 1991). Subsequently, a 25 Å reconstruction of the capsid allowed the crystal structure of the hexon to be placed into the EM map, giving additional information on the capsid organisation (Stewart *et al.*, 1993). However, the atomic details on the capsomer interfaces at the different positions in the triangular facets and the interactions with the minor proteins were not revealed.

\*Corresponding author. EMBL, 6 rue Jules Horowitz, BP 181 38042 Grenoble, France. Tel: +33 4 76 20 70 96; Fax: +33 4 76 20 71 99; E-mail: schoehn@embl-grenoble.fr

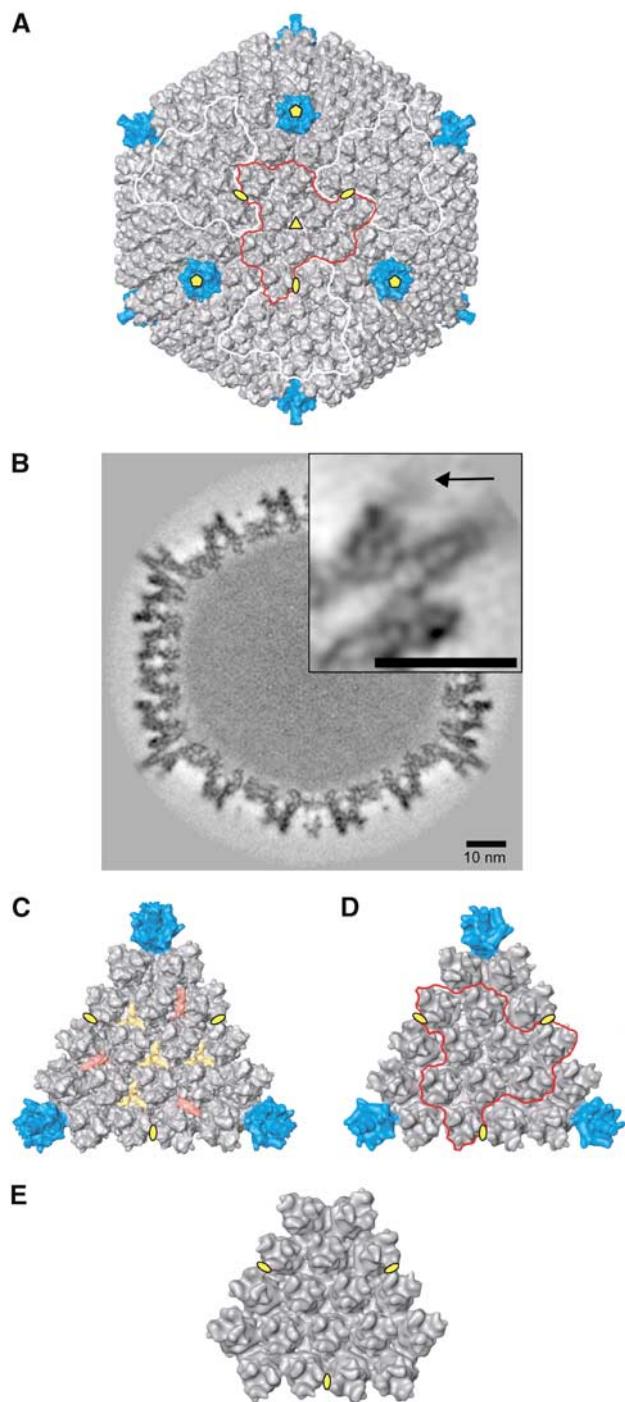
Received: 2 February 2005; accepted: 30 March 2005; published online: 21 April 2005

Here we describe the structure of the entire human Ad5 virion determined at a resolution of 10 Å by cryoEM. We also determined the structures of a mutant virus lacking proteins IIIa and IX, and the same virus after several cycles of freezing/thawing, which led to the expulsion of the pentons and the viral DNA. The difference maps between these three structures allow the densities corresponding to some of the minor proteins and the penton to be clearly delimited. We have generated a quasi-atomic model of the entire capsid by fitting the atomic structures of the hexon (Ad5) and the penton base (Ad2) into the 10 Å cryoEM map. Exposed protein loops disordered in the crystal structures of hexon and penton base are clearly visible in the cryoEM map, as are the loops involved in contacts between the capsomers. It is thus possible to describe, at the atomic level, these different contacts that nicely illustrate the quasi-equivalence principle: hexon loops interact differently with each other according to the position of the hexon in the capsid. We also present an image of the contacts between the fibre and the penton base as well as new details of the minor proteins that stabilise the internal face of the capsid. Finally, we compare the 10 Å structure of Ad with the recently reported atomic structure of the bacteriophage PRD1 capsid (Abrescia *et al*, 2004).

## Results and discussion

### EM models for wild-type (wt) Ad5 and Ad5 pIX° mutant viruses

We used two Ad5 viruses: a GFP expressing Ad5 with a wt capsid (wt Ad5) and an Ad5 pIX° mutant virus with a deletion in the pIX coding sequence. The wt Ad5 was prepared for EM after a single cycle of freezing/thawing (necessary for transport between the laboratory of production and the site of the EM) that did not have any apparent influence on the structure of the virus. The Ad5 pIX° mutant virus was also prepared after a single freezing/thawing step (*sift*) or after multiple cycles of freezing/thawing (*muft*). Image reconstruction from cryoEM images for the three samples gave models with resolutions of about 10 Å for wt Ad5 (Fourier shell correlation criterion of 0.3, see Supplementary data), 16 Å for *sift* pIX° virus and 22 Å for *muft* pIX° virus. CryoEM images showed that the interior of the *muft* pIX° particles was much less dark than that of wt Ad5, indicating that these viruses had lost their DNA (not shown). Figure 1A shows a surface rendering of the EM density for wt Ad5 with the pentons (penton base plus fibre) coloured in blue. A slice of the density map cutting through the middle of the virus particle and four pentons shows the densities of the major and minor capsid proteins and the particle interior filled with DNA (Figure 1B). However, no detailed information on the DNA organisation is visible. The size of the capsid between the tips of the hexons on opposite facets is 890 Å. The fibre is only visible up to the third repeat in the shaft at which point it is flexible, thus smearing out the density (Ruigrok *et al*, 1990; Stewart *et al*, 1991; Wu *et al*, 2003). As the entire map contains so much information, for clarity, Figures 1C-E show only a single facet of wt Ad5, *sift* pIX° and *muft* pIX° viruses. When Ad is dissociated *in vitro* under mild conditions, one of the structures that remains intact is the 'Groups Of Nine hexons' or GON (Laver *et al*, 1969; Furciniti *et al*, 1989). GONs are indicated on four adjacent facets in Figure 1A and on the facet in Figure 1D. The capsid is composed of the



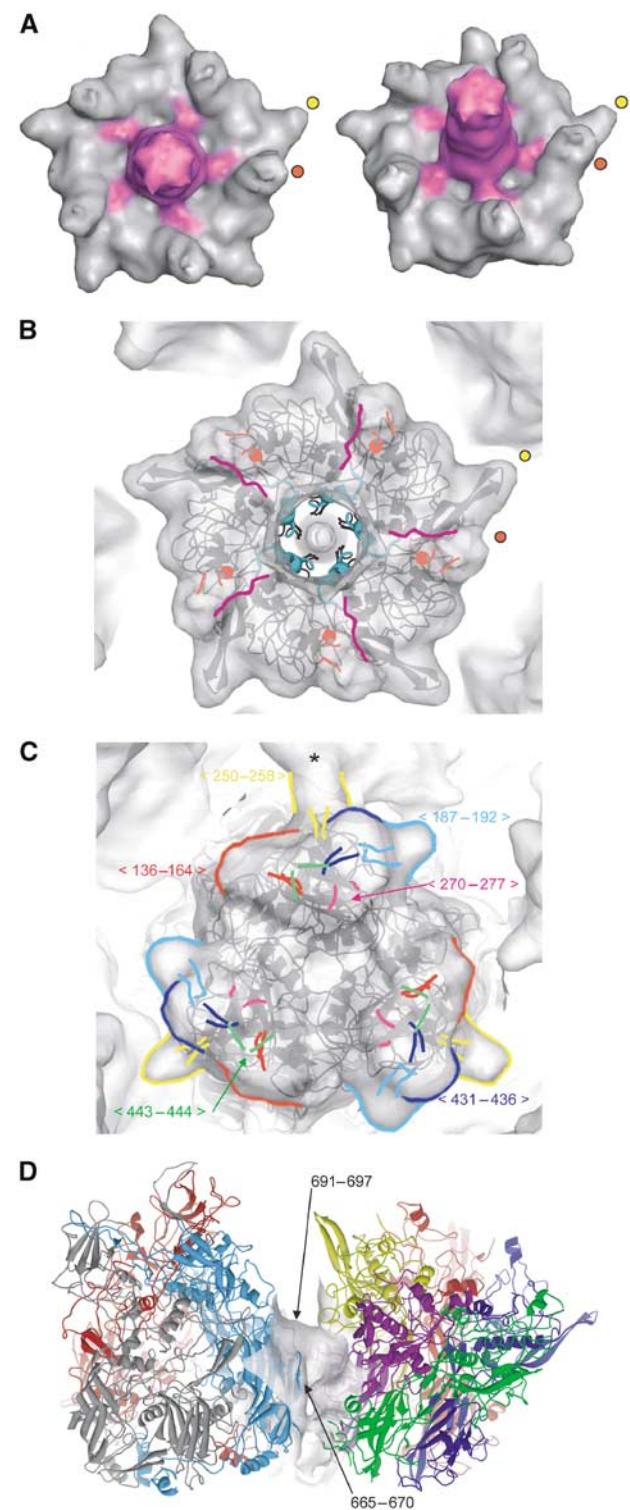
**Figure 1** CryoEM reconstructions of adenovirus. (A) Isosurface representation of the 3D reconstruction of human Ad5 by cryoEM at 10 Å resolution viewed down a three-fold axis. The pentons are coloured in blue and the different symmetry axes are indicated in yellow (pentagons for five-fold, a triangle for three-fold and ellipses for two-fold). The GONs are also outlined on four adjacent facets (in red and white). (B) Central section through the density of the Ad5 seen in a 222 orientation. The protein and DNA densities are dark grey. The upper right inset is a magnification of the section through one penton and one protuberance. The arrow indicates the distal limit of the protuberance extension. Scale bars represent 10 nm. (C) Isosurface representation of one facet of the wt Ad5 capsid in the same orientation as that in (B). Red rods and yellow triskelions represent, respectively, monomers of minor protein IIIa and trimers of minor protein IX. (D) Isosurface representation of one facet of the pIX° sift Ad5 capsid at 16 Å resolution. One GON is outlined in red. (E) Isosurface representation of one facet of the pIX° muft Ad5 capsid at 22 Å.

pentons, the hexons around the penton (peripentonal hexons) and the GONs, plus minor capsid proteins. The structure of *sift* pIX° virus can be super-imposed on that of the wt virus and the difference density reveals two features: four triskelion-like (yellow) and three rod-like (red) structures are visible on each facet exterior (Figure 1C). The triangular complexes correspond to four trimers of protein IX as previously suggested by Furciniti *et al* (1989), who performed image analysis on GONs. The red molecules correspond to three distinct monomers of protein IIIa (Stewart *et al*, 1993). As pointed out before (Furciniti *et al*, 1989; Stewart *et al*, 1993), IX links the hexons within a GON, whereas IIIa stabilises the contacts between neighbouring GONs. The EM reconstructions show that in the *sift* pIX° sample both IX and IIIa are absent. Whereas the absence of IX was expected, we find that IIIa, although present in freshly purified pIX° virus (see Materials and methods), is released from the capsid after freezing/thawing. The most likely explanation for this is that, in the absence of IX, the GONs are less rigid, allowing expulsion of IIIa upon freezing. The structure of *muft* pIX° (Figure 1E) is the same as that of *sift* Ad5 pIX°, despite the fact that it also lacks all its pentons and viral DNA, although the peripentonal hexons remain in place.

#### Fitting the atomic structure of the penton base into the EM model

The isolated EM density for the penton is shown in Figure 2A. The visible part of the fibre and its connection with the penton base are shown in pink. The two protuberances of the penton base are indicated by red and yellow dots. The atomic structure of the Ad2 penton base was fitted into the EM envelope of wt Ad5 (Figure 2B). The penton base sequences from Ad2 and Ad5 are 98.6% identical, so it can be assumed that they have the same structure. The atomic structure shows only one protuberance at the top of the base (the variable loop, Zubieta *et al*, 2005) made up by two antiparallel  $\beta$ -strands, residues 142–169 (yellow dots). The second extension visible in the EM model (red dots) corresponds to residues 296–372, the loop that contains the RGD sequence (RGD loop). The part of the RGD loop shown in

Figure 2A has the same density as the rest of the penton. However, in the magnified view of the density section in Figure 1B, it can be seen that diffuse density extends further out. Such weak density, probably resulting from flexibility, has been reported before in the Ad3 dodecahedron cryoEM structure (Schoehn *et al*, 1996) and in previous reconstructions of the virus (Chiu *et al*, 1999). Similarly, in the penton base crystal structure, most of the RGD loop is missing due to flexibility (Zubieta *et al*, 2005). Only 18 amino-acid residues



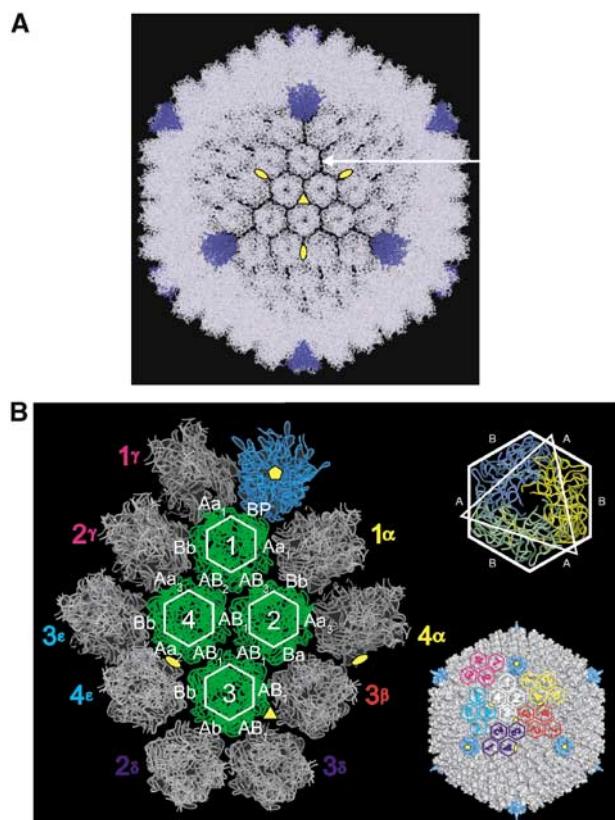
**Figure 2** Fitting the atomic structure in the EM densities of pentons and hexons. (A) Isosurface view of two isolated pentons from the wt Ad5 capsid cryoEM map. The right penton is tilted by about 20° compared to the left one. The fibre and the connection between the fibre and the penton base are coloured in pink. The two protuberances are marked by yellow and red dots (yellow: variable loop; red: RGD loop). (B) Fitting of the penton base crystal structure in complex with amino acids 11–20 from the N-terminal part of the fibre into the cryoEM envelope at 10 Å resolution. The figure is labelled with yellow and red dots as in (A). The two strands of the RGD loop seen in the X-ray structure are coloured in orange. The fibre peptide is in pink and the part of the penton base interacting with the fibre is in cyan. (C) Fitting of the hexon crystal structure into hexon 2 (according to Figure 3B) of the cryoEM envelope at 10 Å resolution. Densities for missing loops in the crystal structure that are present in the EM map are labelled in different colours. \* indicates the contact between Hexon 2 and protein IIIa through the yellow loop (250–258). (D) Detailed view of the peripentonal hexon–penton base atomic interactions. Hexon (grey, red, blue) and penton (in five colours) are in ribbons, whereas the cryoEM density of the connection is in grey. Interactions occur through a hexon B face (in blue). Arrows indicate two out of the three interacting hexon loops.

are observed, which rise to about 20 Å above the level where the fibre attaches to the base. In the EM reconstruction the RGD loop is about 30 Å long with the same density as the penton base. Depending on the secondary structure in the rest of this loop (the crystal structure gives evidence for at least one helical region; Zubieta *et al*, 2005), the total length of the flexible RGD loop is predicted to be between 60 and 70 Å for Ad5. The smeared-out low density visible for the penton base (Figure 1B, arrow) extends to at least 50 Å, indicating that the RGD loop is not totally disordered.

The crystal structure of the penton base without fibre peptide, when fitted into the EM map (not shown), does not account for the connection between the fibre and the penton base (coloured pink in the EM density in Figure 2A). However, fitting the penton base structure containing amino acids 11–20 from the N-terminus of the fibre protein (Zubieta *et al*, 2005) shows that these amino acids fall inside and fill up this connection (Figure 2B). Owing to the imposition of icosahedral symmetry during the image-processing procedure, the densities of the five connections were equal even though the fibre is a trimer. In reality, three connections should have a density 5/3 times higher and two should be absent. This five-fold averaging only affects the connections because the conformational change that occurs upon fibre binding is thought to be cooperative with all penton base monomers having the same structure even if only three tails are bound to them (Zubieta *et al*, 2005). The three tails of the individual trimeric fibres can probably associate randomly with the five binding sites on the penton base. In Figure 1B, it can be seen that the lowest part of the fibre shaft is hollow. The structure of the fibre shaft at the level of the regular fibre repeats is dense and rigid (van Raaij *et al*, 1999). The first regular shaft repeat starts at amino acid 47 (Chroboczek *et al*, 1995). The hollow structure at the base of the fibre, probably composed of the proline-rich region 21–46, appears to constitute a rigid link between the first 20 residues that are fixed onto the penton base and the first repeat in the shaft.

#### Fitting the atomic structure of the hexon into the Ad5 EM model

As the GON contains a three-fold axis, there are only three unique hexons in a GON (hexons 2, 3 and 4 in Figure 3B). The fourth structurally unique hexon in the capsid is the peripentonal hexon (hexon 1 in Figure 3B; Burnett, 1985). The crystal structure of the Ad5 hexon (Rux and Burnett, 2000) was fitted independently into the four unique hexon densities that make up the asymmetric unit (AU), with the assumption that all hexons have essentially the same structure. As for the penton, the hexon crystal structure fits very well into the EM map, with the addition that a number of loops at the top of the hexons and disordered in the crystal structure are resolved (Figure 2C). The good fit of the crystal structure to the EM density, particularly near the base of the hexon (not shown), indicates that the N- and C-termini of the hexon (of which only the four N-terminal and five C-terminal residues are not ordered in the crystal; Rux and Burnett, 2000) are unlikely to be available for extensive hexon–hexon and hexon–penton contacts, as seen in the capsid structure of bacteriophage PRD1 (Abrescia *et al*, 2004).



**Figure 3** Quasi-atomic model of the Ad5 capsid. (A) View down a three-fold axis (similar to Figure 1A) of the quasi-atomic model of the Ad5 capsid made by fitting the hexon (light blue) and penton base (dark blue) crystal structures into the 10 Å resolution EM map. Three-fold (triangle) and two-fold (ellipses) symmetries axes are indicated in yellow. The arrow indicates one position of protein IIIa. (B) Detailed view of the different capsomer contacts. The penton base is in blue. The four hexons coming from one AU are coloured in green with a white hexagon. The hexons at the outside of the AU are in grey and are labelled uniquely. The positions of six AUs (including only the hexons) are also indicated in the complete virus on the right lower part of the figure. The top right shows that the hexon is a trimeric protein with a hexagonal base. Hexagon sides formed exclusively from one monomer are denoted as B; those formed from two monomers are denoted as A. At the left, the different contacts between the capsomers are labelled in white with the following convention: the labels are in capitals when the interaction side belongs to the same AU, and in small letters when coming from another AU; Aa1, Aa2 and Aa3 represent, respectively, contacts between peripentonal hexons, hexons connected across a two-fold symmetry axis, and hexon–hexon connected by protein IIIa. All the atomic contacts are listed in Table I.

After fitting all the unique positions of the major constituents of the capsid, we built a quasi-atomic model of the capsid using its 532 symmetry with the pentons in dark blue and the hexons in light blue (Figure 3A). In the quasi-atomic model, it is apparent that the spacing between the hexons depends on their position on the facet. The separation is particularly wide at the position of protein IIIa (see below, Figure 1C and arrow in Figure 3A).

#### Hexon–hexon and hexon–penton contacts in the quasi-atomic model

Hexons are trimeric complexes with a pseudo-hexagonal base, the six faces of which are of two types depending on

whether the face is made up of only one monomer or of two monomers (respectively B and A in Figure 3B). There are eight different hexon–hexon contacts (Table I), depending on whether these are contacts between peripentonal hexons or hexons in the GON, unique positions in the GON or whether minor proteins are present between the hexons (Figure 3B). When the face comes from a hexon or penton base within the AU it is labelled with a capital A, B or P, and when it comes from a hexon outside the AU it is labelled with 'a' or 'b' (Figure 3B), as originally introduced by Burnett (1985). For the different contacts, all possible interactions at a distance  $\leq 5\text{ \AA}$  between the surfaces of the hexons and the penton are given in Table I. All contacts concern the lower double jellyroll region of the hexon (eight-stranded antiparallel  $\beta$  sheets, regions V1, V2, see Figure 4A). The nine different hexons in the GON are all making four contacts (contacts

AB<sub>1</sub>) with an angle of about 10°. This is also the case between hexons 1–4 and 1–2 in the same AU. Deviations from this angle are located at the IIIa interface, where the angle is about 30° (no hexon–hexon contact), between peripentonal hexons (30° and two contacts) and between hexons through the two-fold axis (12° and four contacts). The structure of the GON is further stabilised through the presence of protein IX at these AB<sub>1</sub> contacts. It is of course possible that in the viral capsid the hexon loops are slightly re-arranged compared to their conformation in the crystal structure. This, combined with the limited resolution of the EM map, implies that residue-to-residue contacts cannot be given to better than a few residues. Although the loops are not specifically enriched in charged or hydrophobic residues, in the majority of cases two interacting loops contain residues with opposite charges (see Table I).

**Table I** Hexon–hexon, hexon–penton base and hexon–minor protein contacts in the capsid

*Contacts between hexons from two different AUs*

Aa <sub>1</sub> Between peripentonal hexons		Aa <sub>2</sub> Two-fold axis		Aa <sub>3</sub> Presence of IIIa		Bb		Ba ou Ab	
A	a	A	a	A	a	B,b(3)	b,B(4)	B,b(2)	A,a(3)
57–62	57–62	65–70	725–727*	—	—	665–670	349–356	349–356	65–70*
<b>TDRSQR</b>	<b>TDRSQR</b>	<b>LRFIPV</b>	<b>DSS</b>			SIPSRN	GQASQLNA	GQASQLNA	LRFIPV
65–70	725–727*	186–193	186–193			691–697	340–344	665–670	725–727
<b>LRFIPV</b>	<b>DSS</b>	<b>GVEGQTPK</b>	<b>GVEGQTPK</b>			GSGYDPY	STGNM	SIPSRN	DSS
		725–727*	65–70				70–80		
		<b>DSS</b>	<b>LRFIPV</b>				<b>VDREDTAYSYK</b>		
		734–741*	734–741*						
		<b>DRLLTPNE</b>	<b>DRLLTPNE</b>						

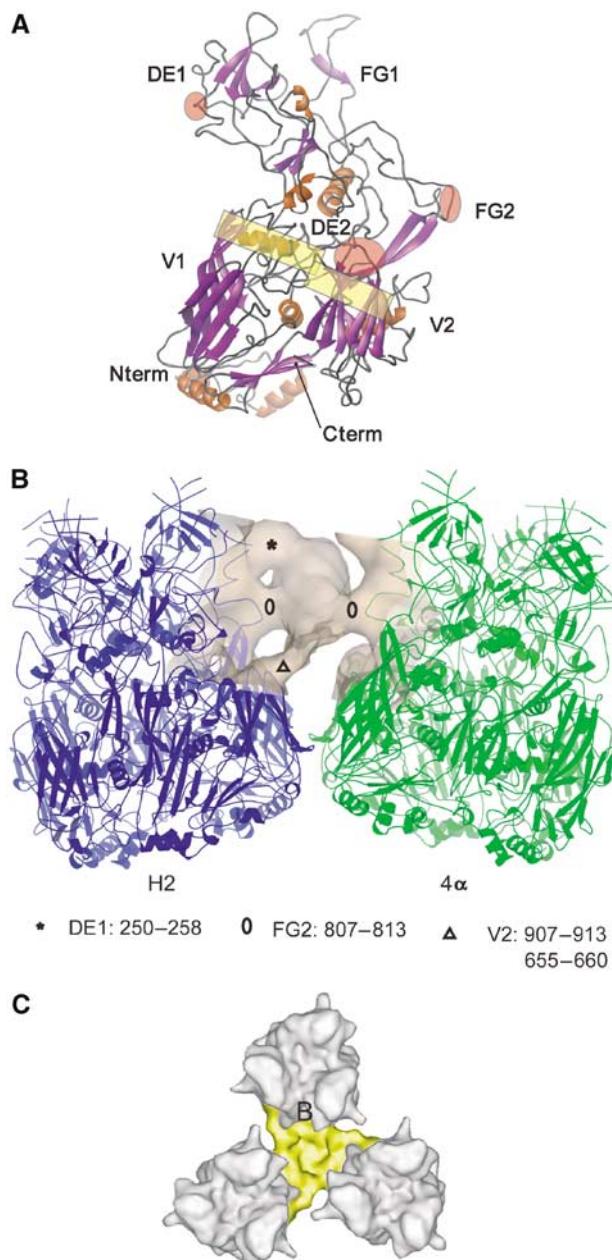
*Contacts between hexons in the same AU*

AB <sub>1</sub> Presence of IX		AB <sub>2</sub>		AB <sub>3</sub>		Contacts between hexon and penton		
A	B	A(1)	B(4)	A(2)	B(1)	B	P	
65–70*	70–76	725–727	665–670	725–727	665–670	346–354		450–455
<b>LRFIPV</b>	<b>VDREDTA</b>	<b>DSS</b>	<b>SIPSRN</b>	<b>DSS</b>	<b>SIPSRN</b>	VLAGQASQL		<b>RSTSQI</b>
87–92*	349–356	734–741	340–344			70–80		
<b>VGDNRV</b>	<b>GQASQLNA</b>	<b>DRLLTPNE</b>	<b>STGNM</b>			<b>VDREDTAYSYK</b>		
725–727	665–670					665–670		
<b>DSS</b>	<b>SIPSRN</b>					SIPSRN		<b>NDYSPGE</b>
734–741	340–344					691–697		82–86
<b>DRLLTPNE</b>	<b>STGNM</b>					<b>GSGYDPY</b>		<b>LNYQN</b>

*Residues of hexon that are in contact with IIIa (Aa<sub>3</sub> interface)*

A(2)	a(4)
250–258	807–813
<b>VKQQNGKLE</b>	<b>DTKYKDY</b>
807–813*	
<b>DTKYKDY</b>	
655–660	
IPANAT	
907–913*	
<b>DPMDEPT</b>	

Only contact regions (numbers that give the positions of the amino acid in the sequence) that are separated by less than 5 Å are listed. Hexons have two types of sides, B and A, depending on the fact whether they are made up of one or two monomers, respectively. Capitals are used for the hexons that make part of the AU and lower case for hexons of another AU. There are three types of Aa contact depending on their position in the capsid (i.e. between peripentonal hexon, when IIIa is present and between hexons on the two-fold axis). '\*' indicates that the contacts are made by two different monomers in an A-side of the corresponding hexon. The amino acids involved in the interaction are listed and the charged amino acids are in bold. When a zigzag line is present, two loops from one capsomer are interacting with one loop from another capsomer.



**Figure 4** Connections between hexon and proteins IIIa/IX at the atomic level. (A) Ribbon view of one monomer of the hexon. The secondary structure elements are marked, as well as some of the interacting regions with proteins IIIa (red ellipses) and IX (yellow rectangles). The N and C termini are also marked. (B) Atomic interactions between protein IIIa (EM density) and its two neighbouring hexons H2 and  $4\alpha$  (ribbons). The asterisk, triangle and circles mark the interactions of IIIa with the hexons and the corresponding structural elements are indicated at the bottom of this figure and in part A. (C) Detailed top view of the hexon–protein IX interaction on the three-fold of the virus. The protein IX monomer wraps only around the hexon B face.

The peripentonal hexons interact with the penton through a B-face (Figures 2D and 3B). The penton base–hexon interaction is also just below the middle of the two structures and involves loops of the jellyroll motifs of both proteins (Figure 2D; Table I). Only one hexon loop implicated in hexon–hexon contacts (loop 665–670) is also involved in the hexon–penton base contact; the two others are different. Two of the penton base loops, 99–105 and 450–455, are also

involved in inter-penton base contacts in the penton base dodecahedron where penton bases interact directly (Zubieta *et al*, 2005).

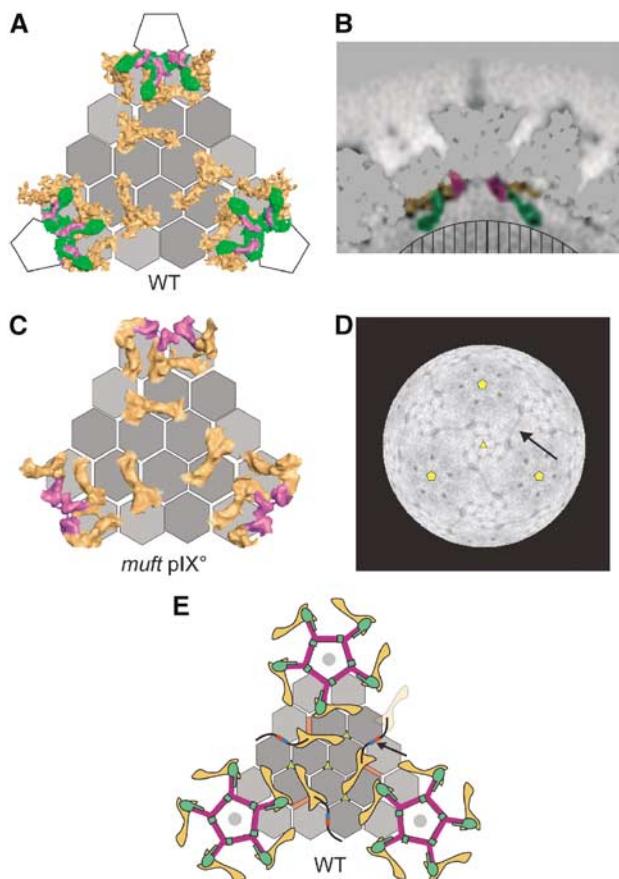
The interactions that a hexon makes with its neighbours depend on its position in the facet and this illustrates the principle of quasi-equivalence. For example, loop 725–727 interacts either with loop 65–70 (when the hexons are across the two-fold axis) or with loop 665–670 (hexons around the three-fold axis).

#### Position and contacts of minor proteins IIIa and IX on the outside of the capsid

Protein IIIa forms a rod with a length of about 60 Å and a maximum width of about 30 Å and links two neighbouring GONs at an  $\text{Aa}_3$  hexon–hexon contact (Figures 1C and 3B and Table I). It lies at an angle of 45° to the capsid surface and touches the side of two hexons near their middle at the level of structural elements FG2 (807–813, Figures 4A and B). As seen in Figure 4B, the top of protein IIIa is attached to one hexon by a loop that is missing in the X-ray structure (250–258, element DE1) and the bottom of protein IIIa is connected to two elements (655–660; 907–913) from the V2 domain of the same hexon (Rux and Burnett, 2000; Figures 4A and B and Table I). This position leads to a separation in space of the two neighbouring hexons that have no densities that approach within the 5 Å cutoff distance that was used to determine subunit contacts (see arrow in Figure 3A). This implies that at this position capsid integrity is maintained only through hexon–IIIa–hexon interactions.

As the density of the internal minor proteins in the *sift* Ad5 pIX° reconstruction does not differ from those in wt virus (not shown but similar to Figure 5A), we conclude that IIIa is positioned only on the outside of the capsid. In a previous model of the Ad2 capsid structure, IIIa was proposed to traverse the width of the capsid with a rod-like density (corresponding to what we see) near the top of the hexon and another domain near the bottom of the hexon (Stewart *et al*, 1993). However, we assign the latter density on the inside of the capsid to a distinct minor capsid component, most likely protein VIII (see below and Figure 5). The reason for this different interpretation is probably due to the fact that the resolution of our model is better (10 Å versus 25 Å) and also because the atomic model for the hexon that was used in 1993 was corrected in 2000 (Rux and Burnett, 2000). At the resolution of our 3D reconstruction, the volume for IIIa corresponds only to about half the volume of a protein of 63 kDa. Most of the C-terminal 200 amino acids of IIIa are predicted to be disordered (FOLDINDEX; Uversky *et al*, 2000) and may occupy an extended conformation in the virus that is too diffuse to be resolved at the present resolution or may deviate from icosahedral symmetry. There is a nonattributed thin density running through the two-fold axis at the inside of the particle that starts just under the protein IIIa (see Discussion and Figure 5D). It is not possible to see if this feature is present or absent in the *sift* capsid that lacked IIIa, because it is only visible at higher resolution.

Protein IX is positioned at the  $\text{AB}_1$  hexon contacts (Figures 1C and 3B and Table I). This trimeric molecule has three arms with a length of about 40 Å (from the triskelion axis to the end of one arm) and a maximum diameter of 20 Å. The protein only wraps around the B face of the hexon, leaving the opposite A face free (Figure 4C). The interactions between



**Figure 5** Structural views of the inside of Ad5 capsid. (A) View of the inside of one facet of the wt Ad5. Hexons and penton base are represented diagrammatically with the GON in dark grey. The densities of the minor proteins coming from difference imaging of the cryoEM map with the wt 10 Å filtered quasi-atomic model exhibit three kinds of extra densities: a hammer-like orange shape (120 copies in the capsid), which probably corresponds to protein VIII (see text) and magenta and green densities that have not been attributed. The interactions between protein VIII and the four hexons are the same around the five-fold and around the three-fold symmetry axis. (B) Central slice through the density of one penton showing the internal organisation of the minor proteins under the five-fold symmetry axis. This view is generated from the difference map between the 10 Å filtered quasi-atomic model and the wt Ad5 cryoEM map. The missing density (coming from the subtraction of the quasi-atomic model) is in grey. The remaining different densities under the five-fold are coloured according to Figure 5A. The green density extends up to 40 Å under the hexon. Parallel lines indicate schematically the DNA. (C) Same view as in (A), but using the difference image from the *muft pIX°* adenovirus mutant reconstruction and the quasi-atomic model filtered to 18 Å. The colour codes are the same as in (A) and (B). (D) Spherical section view (one voxel thick) at a radius of 330 Å down the three-fold axis of the wt Ad5 cryoEM reconstruction that shows a thin connection (10 Å in diameter) over the two-fold symmetry axis (arrow). This connection is visible between radius 325 and 335 in Å (i.e. in four slices). In the map, the extremities of this connection start just below the big part of the orange hammer-like density, which is also vertically beneath the bottom of protein IIIa (see (E)). (E) Schematic distribution of the major and minor proteins in one facet of the Ad particle seen from the inside of the virus particle. Hexons are in grey (GON in dark grey) with a hexagonal shape; pentons are in white with a grey disk representing the fibre. Protein IIIa is in red, protein IX in yellow, VIII in orange. Un-attributed densities around the five-fold symmetry axis are colour coded according to (A)–(C). The connections across the two-fold axis (coming from VIII or IIIa) are in red and blue.

IX and the hexon B face are too numerous to be listed in Table I, but they are clearly mediated via the top of jelly-roll V1 and V2 of the hexon (yellow rectangles in Figure 4A). As IX is relatively thin and sits on the top of the hexon-hexon interacting region, most of the normal interactions between the hexons are not perturbed. The fact that the hexon-hexon contacts can occur independently of the presence of IX probably explains why capsids can form in the absence of this protein although the virus is less stable at elevated temperatures (Colby and Shenk, 1981).

#### Positions of minor capsid proteins on the inside of the capsid

Adenoviruses have several minor proteins on the inner surface of the capsid. Polypeptide VI is thought to form a ring underneath the peripentonal hexons (Stewart *et al*, 1993). The position of protein VIII has not yet been assigned (Stewart *et al*, 1993), but was recently proposed as a possible size-determining or ‘tape-measure’ protein similar to P30 of PRD1 (Abrescia *et al*, 2004; Rux and Burnett, 2004). Figure 5A and B show the EM density map of the inside of wt Ad5 capsid after subtraction of the crystallographic densities of hexon (virtually the entire molecule) and penton base (excluding residues 1–52 since their structure is not known) filtered to about 10 Å resolution (IIIa and IX have also been removed to clarify the figure). Figure 5C shows the inside of the *muft pIX°* map after subtraction of the hexon density filtered to 18 Å (this virus sample had already lost its pentons and DNA and probably all other proteins only attached to pentons and DNA). The difference map of the *sift pIX°* mutant is similar to Figure 5A, except for being at lower resolution (not shown).

The most conspicuous feature in Figure 5A and C (and also in the difference map of *sift pIX°*) is the orange hammer-like structure organised in a  $T=2$  lattice (120 copies), with a volume corresponding to a protein of about 20 kDa (this is probably an overestimate due to the many fine details in the structure that could not be taken into account). Protein VIII is the only capsid protein likely to be associated with both DNA and hexons, and which is present with this stoichiometry. Stewart and Burnett (1995) estimated the presence of about  $127 \pm 3$  copies, assuming a molecular weight of processed VIII of 15.4 kDa (c.f. full-length precursor pVIII Ac2-227, 24.6 kDa). In previous, lower-resolution EM maps of Ad (Stewart *et al*, 1993), the hammer-like density was thought to be part of protein IIIa. However, due to the fact that the density is present in the *muft pIX°* capsid (which does not contain protein IIIa) and particularly its presence in a  $T=2$  lattice, this assignment is excluded. Note that the exact state of protein VIII in mature viral capsids is not totally established, since it is uncertain which potential protease sites are actually cleaved and which fragments actually remain in the virus. A recent analysis suggests that whereas empty viral capsids contain full-length pVIII, mature viral particles contain pVIII-derived 7.6 and 12 kDa peptides with estimated copy numbers of 160 and 130 per particle (Vellekamp *et al*, 2001). Other analyses reveal the likely presence of only the so-called propeptide domain of pVIII (Ac2-111, MW 12 kDa, protease cleavage at LAGG/F) in the virion (Blanche *et al*, 2001; Liu *et al*, 2003). These results are generally consistent with the volume and stoichiometry of the hammer-like EM density we observe.

The  $T=2$  lattice implies two nonequivalent positions in the capsid for protein VIII (Figure 5A). The first is around the five-fold axis where five copies of protein VIII generate a ring that connects the peripentonal hexons and links them to the GON. These molecules remain in place after the removal of the penton. The second localisation occurs around the three-fold axis, where three molecules of VIII form a second ring that most likely stabilises the GONs as well as makes links between different GONs, complementing, respectively, the binding function of proteins IX and IIIa on the external side of the capsid. Independent of their location around the three- or the five-fold axis, each protein VIII appears, at the resolution of the map, to interact in exactly the same manner with four hexons. Looking carefully at spherical slices of the difference map at a radius between 325 and 335 Å, another feature with a diameter of 10 Å is apparent. This makes a connection across the two-fold axis between the rings of protein VIII around the three-fold axis, linking the GONs of two facets (arrow in Figure 5D). It is not possible to be sure of the origin of this linker at the current resolution, but it does suggest that unfolded regions of minor capsid proteins (possibly VIII or IIIa) are indeed involved in forming extended arms, perhaps in a fashion reminiscent of P30 of PRD1 (Abrescia *et al*, 2004; Rux and Burnett, 2004).

The second major feature of the difference map is the magenta coloured ring of density around the five-fold axis in Figure 5A–C. In the wt Ad5 map, two additional green blobs of density connected by a green arm are also present (Figure 5A and B), but not when the pentons and the DNA are removed (Figure 5C). The largest green blob is attached to protein VIII and to the magenta density, and appears to be a hollow sphere. The smaller green blob and the arm are only attached to the magenta ring. We do not have any positive way of identifying the protein nature of these densities that may consist of the amino-terminal 52 amino acids of the penton base, protein VI, other viral proteins or even viral DNA. However, the magenta ring cannot consist of the N-terminus of the penton base because it remains there when the penton is expelled. Stewart *et al* (1993) also observed density underneath the peripentonal hexons and attributed this to protein VI. Protein VI has nonspecific DNA-binding activity and also is an activator of the Ad protease (Russell and Precious, 1982). It is predicted to have 44% disordered residues (FOLDINDEX, residues 1–32; 65–126; 170–177). However, neither the green, the magenta, nor the green plus magenta regions together represent enough density for the 342 copies (7.6 MDa) of protein VI estimated to be present in the virion (van Oostrum and Burnett, 1985; Stewart and Burnett, 1995) unless a large portion of the protein is disordered. Whatever the identity of these densities, the largest green blob points radially inwards towards the middle of the virus particle (up to 40 Å under the base of the hexon protein, Figure 5B), away from the capsid and in contact with the DNA at exactly the same position as the trans-membrane protein P16 of PRD1 that binds the capsid to the viral DNA on the inner side of the viral lipid membrane (Abrescia *et al*, 2004).

#### Organisation of the DNA in the capsid

There have been a number of models proposing the organisation of the DNA in human adenoviruses (Newcomb *et al*, 1984). From our results, it is apparent, as from the previous

work of Stewart *et al* (1991, 1993), that the DNA is not organised with any icosahedral symmetry, because, if it were, dsDNA strands would have been visible around the symmetry axes. Other viruses, such as Herpes virus and bacteriophages lambda and T4, have closely packed dsDNA that is readily visible in cryoEM images (Lepault *et al*, 1987; Booy *et al*, 1991). The absence of any ordered density directly on the micrographs or in the 3-D model indicates that there is neither icosahedral symmetry nor liquid crystalline organisation of the DNA of Ad5. Nevertheless, it is still possible to deduce some information about the DNA packing in the virus from the EM map. Only two types of connections seem to be present between the capsid proteins and the DNA (Figure 5B and not shown). The first one occurs through the spherical blob of green density that points to the inside of the particle. The arms of protein VIII, independent of their position, make the second type of connection.

#### Comparison between adenovirus and PRD1 capsid structures

Hypotheses on an evolutionary relationship between adenoviruses and bacteriophage PRD1 are based on the similarity in overall organisation of their protein capsids and the structures of the hexon (P3 in PRD1) and penton base (P31). However, for both hexons and pentons, whereas the bottom parts of the proteins have the same fold, the upper elaborations are not present in PRD1 (Benson *et al*, 1999, 2004; Rux and Burnett, 2000; Abrescia *et al*, 2004; Zubietta *et al*, 2005). The Ad capsid is heavily stabilised by a network of connections made by interactions with at least four types of minor protein both on the inside and the outside of the capsid (Figure 5E). On the outside of the Ad capsid, proteins IIIa and IX interact mainly with the upper part of the hexons, above the jelly-roll domain. There cannot be any equivalent proteins or interactions in the PRD1 structure because this top part of P3 is absent.

In PRD1, the contacts between P3–P3 and P3–P31 are stabilised by specific interactions of their respective N- and C-terminal peptides, resulting in specific conformations of these termini at specific positions in the AU. In Ad, all hexons are connected with each other by protein VIII which, therefore, has a role functionally equivalent to the N and C termini of P3 (PRD1 hexon). Furthermore, it seems that for PRD1, the P3–P31 connection is stronger than the P3–P3 connection, since when PRD1 empty particles or the sus1 mutant (a mutant that does not package DNA) are treated with SDS, P31 plus peripentonal P3 are released (Martin *et al*, 2001). For Ad empty capsids (the *muft* structure in this work and also natural empty capsids; Vellekamp *et al*, 2001), only the penton is ejected with the peripentonal hexons remaining in place, probably because of the stabilisation of the contact between the GON and the peripentonal hexons by protein VIII.

The only minor capsid protein of PRD1 that is organised with icosahedral symmetry is the ‘tape-measure protein’ P30. Interestingly, two PRD1 P30 proteins meet over the two-fold axis just like the weak density seen in Ad (arrows in Figure 5D and E). Another apparent similarity is the presence of a protein element attached to the peripentonal hexons (the ‘green density’ for Ad, P16 for PRD1), which extends inwards probably to interact with the viral DNA (Figure 5B). However, the major difference between the two viruses remains the

presence of the lipid membrane in PRD1, needed for the infection mechanism, but which is absent in adenoviruses.

## Materials and methods

### Viruses

E1/E3-deleted Ad5 (the E1 region was replaced by the gene for GFP) that is wt for the capsid (wt Ad5) was obtained from the Gene Vector Production Network (<http://www.gvnp.org/>). For the recombinant Ad5 pIX° virus, viral genomes were constructed as infectious plasmids by homologous recombination in *Escherichia coli*, as described previously (Rosa-Calatrava *et al*, 2001). Ad5 pIX° contained an additional deletion that removed the entire coding sequence for minor capsid protein IX. Growing, purification, titration and storage were as described previously (Rosa-Calatrava *et al*, 2001). As checked by SDS-PAGE and immunoblotting using polyclonal antibody against the virus capsid (gift from P Boulanger), penton base, fibre and IIIa proteins were present both in infected cell extracts and after CsCl purification, whereas IX was absent (see Supplementary data and not shown).

### Electron cryomicroscopy

Frozen hydrated specimens of wt virus were prepared on quantifoil grids as described previously (Schoehn *et al*, 1997) and with quantifoil grids covered with a thin layer of continuous carbon for Ad5 pIX° in order to increase the virus concentration (Schoehn *et al*, 1997). Specimens were imaged with less than 10 electrons/Å<sup>2</sup> at a temperature close to -186°C by using a Gatan 686 cryo-holder in a LaB<sub>6</sub> Philips CM200 electron microscope operating at 200 kV. Tobacco mosaic virus (TMV) was used to determine the absolute magnification of the microscope, which was found to be 28 600±100 times.

### Image analysis

Images were screened by optical diffraction, and the best were digitised on a Z/I Imaging scanner (Photoscan TD) with a pixel size of 7 µm that corresponds to a sample pixel size of 2.45 Å (see calibration above). The particles were selected interactively using x3d and boxed into 427×427 pixels with a circular mask applied of radius 209 pixels plus four fade pixels (Conway and Steven, 1999). The starting model for the image analysis procedure was calculated from images of negatively stained (2% ammonium molybdate) wt Ad5 using the classical common lines orientation search (Crowther *et al*, 1970) and performing single-particle reconstructions. The best single-particle reconstruction (showing clear  $T=25$  triangulation) was then used as the starting model in a model-based approach using a modified version of the Polar Fourier Transform (PFT) algorithm (Baker and Cheng, 1996; Conway and Steven, 1999). From a total of 250 images of negatively stained particles, 120 were used in a final reconstruction at 25 Å resolution (Fourier Shell Correlation 0.5 criterion; Yang *et al*, 2003). This negative stain model was used to start analysis of the cryoEM images. Determination of particle origin and orientation were again performed using the model-based PFT programs after images had been corrected for contrast transfer function effects (CTFMIX, Conway and Steven, 1999).

## References

- Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martin C, Burnett RM, Stuart DI, Bamford DH, Bamford JK (2004) Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**: 68–74
- Albinsson B, Kidd AH (1999) Adenovirus type 41 lacks an RGD alpha(v)-integrin binding motif on the penton base and undergoes delayed uptake in A549 cells. *Virus Res* **64**: 125–136
- Baker TS, Cheng RH (1996) A model-based approach for determining orientations of biological macromolecules imaged by cryo-electron microscopy. *J Struct Biol* **116**: 120–130
- Benson SD, Bamford JK, Bamford DH, Burnett RM (1999) Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98**: 825–833

For wt Ad5, we selected 6131 images of particles from nine focal pair micrographs plus three single negatives. After completing PFT refinement, one supplementary cycle was carried out using a new version of PFT: PFT2 (<http://www.niams.nih.gov/rcc/labbranch/lsbr/software>). Of the total selected particles, two-thirds (4100) were chosen for the final reconstruction according to a figure of merit calculated from a combination of the PFT coefficients ccPRJ and ccCMP. The reconstruction includes data to a spatial frequency limit of (9 Å<sup>-1</sup>) and the resolution was estimated at (10.3 Å<sup>-1</sup>), where the Fourier shell correlation between independent half data-set density maps crossed below 0.3 (see Supplementary data).

For the *sift* Ad5 pIX° virus, we selected 1202 images of particles from 10 focal pair micrographs plus two single negatives, of which 1062 were included in the final reconstruction, yielding a resolution of 15.5 Å. For the *muft* Ad5 pIX° virus, we selected 820 images of particles from five focal pair micrographs. For the final reconstruction we included 660 particles, obtaining a resolution of 22 Å.

The contour level of the isosurface views was chosen using three filtered X-ray structures of the hexon (10, 16 and 22 Å) so that the surface contours the right molecular weight of the protein with an average protein density of 0.84 Da/Å<sup>3</sup>.

### Fitting the atomic structures of hexon and penton base into the EM map and contacts between capsomers

The atomic structures of the Ad5 hexon and that of the Ad2 penton base were fitted into the EM model for wt Ad5. The atomic structures were placed manually into the four hexons that make up the AU and into one penton base using the program 'O' (Jones *et al*, 1991). The optimisation procedure was carried out using SITUS (Wriggers and Birmanns, 2001) for the penton base and using URO (Navaza *et al*, 2002) for the four hexons. The quality of the final fit is given by a correlation coefficient of 94.4% and crystallographic *R* factor of 26.7%. From the optimised positions of the four hexons in the AU plus the penton, the quasi-atomic model of the whole capsid was built using 532 symmetry.

### Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

## Acknowledgements

We thank the Production and Control department of Genethon which is supported by the Association Française contre les Myopathies (AFM), for providing us with the Ad5GFP sample. We also thank Dr David Belnap for providing pft3dr2 software and Jorge Navaza for help with the URO software. Figures were made using AMIRA (TGS Europe, Merignac, France), Povscript (Avatar Software AB) and Chimera (Pettersen *et al*, 2004). JFC was supported in part by an Action Thématique et Incitative sur Programme (ATIP) from the CNRS. The EM map has been deposited in the EM database associated with the Macromolecular Structure Database. The accession numbers are, respectively, 1I11, 1I12 and 1I13 for *muft*, *sift* and wild-type Ad. PDB accession number for the quasi-atomic model is 2bld.

- Benson SD, Bamford JK, Bamford DH, Burnett RM (2004) Does common architecture reveal a viral lineage spanning all three domains of life? *Mol Cells* **16**: 673–685
- Blanche F, Monegier B, Faucher D, Duchesne M, Audhuy F, Barbot A, Bouvier S, Daude G, Dubois H, Guillemin T, Maton L (2001) Polypeptide composition of an adenovirus type 5 used in cancer gene therapy. *J Chromatogr A* **921**: 39–48
- Booy FP, Newcomb WW, Trus BL, Brown JC, Baker TS, Steven AC (1991) Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus. *Cell* **64**: 1007–1015
- Boulanger PA, Torpier G, Biserte G (1970) Investigation on intranuclear paracrystalline inclusions induced by adenovirus 5 in KB cells. *J Gen Virol* **6**: 329–332

- Burnett RM (1985) The structure of the adenovirus capsid. II. The packing symmetry of hexon and its implications for viral architecture. *J Mol Biol* **185**: 125–143
- Chiu CY, Mathias P, Nemerow GR, Stewart PL (1999) Structure of adenovirus complexed with its internalization receptor, alphavbeta5 integrin. *J Virol* **73**: 6759–6768
- Chroboczek J, Ruigrok RW, Cusack S (1995) Adenovirus fiber. *Curr Top Microbiol Immunol* **199** (Part 1): 163–200
- Colby WW, Shenk T (1981) Adenovirus type 5 virions can be assembled *in vivo* in the absence of detectable polypeptide IX. *J Virol* **39**: 977–980
- Conway JF, Steven AC (1999) Methods for reconstructing density maps of 'single' particles from cryoelectron micrographs to subnanometer resolution. *J Struct Biol* **128**: 106–118
- Crowther RA, Amos LA, Finch JT, De Rosier DJ, Klug A (1970) Three dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. *Nature* **226**: 421–425
- Davison AJ, Benko M, Harrach B (2003) Genetic content and evolution of adenoviruses. *J Gen Virol* **84**: 2895–2908
- Fender P, Ruigrok RW, Gout E, Buffet S, Chroboczek J (1997) Adenovirus dodecahedron, a new vector for human gene transfer. *Nat Biotechnol* **15**: 52–56
- Furciniti PS, van Oostrum J, Burnett RM (1989) Adenovirus polypeptide IX revealed as capsid cement by difference images from electron microscopy and crystallography. *EMBO J* **8**: 3563–3570
- Greber UF, Willets M, Webster P, Helenius A (1993) Stepwise dismantling of adenovirus 2 during entry into cells. *Cell* **75**: 477–486
- Horwitz MS (2001) Adenoviruses. In *Fields' Virology*, Knipe DM, Howley PM (eds), pp 2301–2326. Philadelphia, USA: Lippincott Williams and Wilkins
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47** (Part 2): 110–119
- Laver WG, Wrigley NG, Pereira HG (1969) Removal of pentons from particles of adenovirus type 2. *Virology* **39**: 599–604
- Lepault J, Dubochet J, Baschong W, Kellenberger E (1987) Organization of double-stranded DNA in bacteriophages: a study by cryo-electron microscopy of vitrified samples. *EMBO J* **6**: 1507–1512
- Liu YH, Vellekamp G, Chen G, Mirza UA, Wylie D, Twarowska B, Tang JT, Porter FW, Wang S, Nagabhushan TL, Pramanik BN (2003) Proteomic study of recombinant adenovirus 5 encoding human p53 by matrix-assisted laser desorption/ionization mass spectrometry in combination with database search. *Int J Mass Spectrom* **226**: 55–69
- Martin CS, Burnett RM, de Haas F, Heinkel R, Rutten T, Fuller SD, Butcher SJ, Bamford DH (2001) Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions. *Structure (Camb)* **9**: 917–930
- Navaza J, Lepault J, Rey FA, Alvarez-Rua C, Borge J (2002) On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr D Biol Crystallogr* **58**: 1820–1825
- Newcomb WW, Boring JW, Brown JC (1984) Ion etching of human adenovirus 2: structure of the core. *J Virol* **51**: 52–56
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605–1612
- Rosa-Calatrava M, Grave L, Puvion-Dutilleul F, Chatton B, Kedinger C (2001) Functional analysis of adenovirus protein IX identifies domains involved in capsid stability, transcriptional activity, and nuclear reorganization. *J Virol* **75**: 7131–7141
- Ruigrok RW, Barge A, Albiges-Rizo C, Dayan S (1990) Structure of adenovirus fibre. II. Morphology of single fibres. *J Mol Biol* **215**: 589–596
- Russell WC (2000) Update on adenovirus and its vectors. *J Gen Virol* **81**: 2573–2604
- Russell WC, Precious B (1982) Nucleic acid-binding properties of adenovirus structural polypeptides. *J Gen Virol* **63** (Part 1): 69–79
- Rux JJ, Burnett RM (2000) Type-specific epitope locations revealed by X-ray crystallographic study of adenovirus type 5 hexon. *Mol Ther* **1**: 18–30
- Rux JJ, Burnett RM (2004) Adenovirus structure. *Hum Gene Ther* **15**: 1167–1176
- Schoehn G, Fender P, Chroboczek J, Hewat EA (1996) Adenovirus 3 penton dodecahedron exhibits structural changes of the base on fibre binding. *EMBO J* **15**: 6841–6846
- Schoehn G, Moss SR, Nuttall PA, Hewat EA (1997) Structure of Broadhaven virus by cryoelectron microscopy: correlation of structural and antigenic properties of Broadhaven virus and bluetongue virus outer capsid proteins. *Virology* **235**: 191–200
- Stewart PL, Burnett RM (1995) Adenovirus structure by X-ray crystallography and electron microscopy. *Curr Top Microbiol Immunol* **199** (Part 1): 25–38
- Stewart PL, Burnett RM, Cyrklaff M, Fuller SD (1991) Image reconstruction reveals the complex molecular organization of adenovirus. *Cell* **67**: 145–154
- Stewart PL, Fuller SD, Burnett RM (1993) Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy. *EMBO J* **12**: 2589–2599
- Uversky VN, Gillespie JR, Fink AL (2000) Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* **41**: 415–427
- van Oostrum J, Burnett RM (1985) Molecular composition of the adenovirus type 2 virion. *J Virol* **56**: 439–448
- van Raaij MJ, Mitraki A, Lavigne G, Cusack S (1999) A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**: 935–938
- Vellekamp G, Porter FW, Sutjipto S, Cutler C, Bondoc L, Liu YH, Wylie D, Cannon-Carlson S, Tang JT, Frei A, Voloch M, Zhuang S (2001) Empty capsids in column-purified recombinant adenovirus preparations. *Hum Gene Ther* **12**: 1923–1936
- Wickham TJ, Mathias P, Cheresh DA, Nemerow GR (1993) Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell* **73**: 309–319
- Wriggers W, Birmanns S (2001) Using situs for flexible and rigid-body fitting of multiresolution single-molecule data. *J Struct Biol* **133**: 193–202
- Wu E, Pache L, Von Seggern DJ, Mullen TM, Mikyas Y, Stewart PL, Nemerow GR (2003) Flexibility of the adenovirus fiber is required for efficient receptor interaction. *J Virol* **77**: 7225–7235
- Yang S, Yu X, Galkin VE, Egelman EH (2003) Issues of resolution and polymorphism in single-particle reconstruction. *J Struct Biol* **144**: 162–171
- Zubieta C, Schoehn G, Chroboczek J, Cusack S (2005) The structure of the human adenovirus 2 penton. *Mol Cell* **17**: 121–135

## II.2 Matériel et Méthodes complémentaires

Nous disposons de deux échantillons d'Adénovirus de type 5 recombinants. Ces échantillons sont conservées à -80°C dans un tampon Tris 10 mM pH 8,5 contenant 1 M Sucrose, 1 mM MgCl<sub>2</sub>, 150 mM NaCl et 0,005 % Tween80. Le premier échantillon, noté WT, est sauvage pour la capsid. Un aliquote de 50 µl est décongelé puis imagé par cryoME après retrait du sucrose par centrifugation. Cet échantillon n'a subi qu'un seul cycle de congélation-décongélation. Le deuxième échantillon est un virus recombinant noté pIX°, délété pour le gène de la protéine IX. Celui-ci est séparé en deux lots : le premier, noté sift pIX°, est imagé par cryoME directement après décongélation. Il ne subit donc qu'un seul cycle de congélation-décongélation comme l'échantillon WT. Le deuxième lot, noté muft pIX°, subit plusieurs cycles de congélation-décongélation. La concentration du virus WT est de l'ordre de 10<sup>12</sup> particules/ml. Celle du virus mutant pIX° est 1000 fois plus faible, d'environ 10<sup>9</sup> particules/ml.

### *II. 2.1 Informations supplémentaires à l'article 1*

Informations supplémentaires publiées en complément de l'article 1

**Supplementary information to: A quasi-atomic model of human Adenovirus type 5 capsid**

By: Fabry, Rosa-Calatrava, Conway, Zubieta, Cusack, Ruigrok and Schoehn.

**Methods**

**Fitting the atomic structures of hexon and penton base into the EM map.**

The atomic structures of the Ad5 hexon (residues 5-946, full length of hexon is 951 residues, Rux and Burnett, 2000; PDB number 1RUX) and that of the Ad2 penton base (residues 52-567, 1-48 are missing, 49-51 disordered, 299-372 disordered; Zubieta et al. 2005, PDB number 1X9P; structure of penton base plus a 20 amino acid N-terminal peptide, PDB number 1X9T) were fitted into the EM model for wt Ad5.

*Penton base.* The atomic structure the penton base was placed manually into one penton base using the program “O” (Jones et al, 1991). Using SPIDER (Frank et al, 1996), the atomic model of the penton base was filtered to a resolution of 11 Å in order to create a binary mask (value of 1 for everything inside the penton base and 0 for everything outside). This mask was then enlarged by a factor of 1.6 in order to be sure that any loops that are visible in the EM model of the penton base but invisible in the atomic model (due to disorder) are included for the positioning process. The EM map was the multiplied by the 11 Å binary mask. The mask contained the density of the penton base but also some density from neighbouring molecules (such as neighbouring hexons or minor proteins). The position of the atomic model of the penton base was then optimised into the isolated penton from the EM map with SITUS (Wriggers and Birmanns 2001). The 11 other pentons were generated using icosahedral symmetry and the quality of the fit was then checked manually using “O”.

*Hexons.* The atomic structures of the hexon was placed manually into the four hexons that make up the Asymmetric Unit (AU) using the program “O” (Jones et al, 1991). URO (Navaza et al, 2002) was used to generate a preliminary quasi-atomic model of the entire hexon shell of the virus

using the 4 manually positioned atomic structures of hexon in the AU. For this first model, a mask was made using SPIDER in order to exclude as much density as possible that did not belong to the hexon trimers. This mask was then used to restart URO and to find the final position of the four hexons in the asymmetric unit. The quality of the final fit is given by a correlation coefficient of 94.4 % and crystallographic Rfactor of 26.7 %. We also checked the quality of the fit visually and with the softwares mapman and mama (Upsala Software Factory). For the 240 hexons, an averaged value of 0.9 % of the atoms from the X-ray structure falls outside the EM map (hexon 1 : 0.83%, hexon 2: 0.86 % hexon 3 : 0.92; hexon 4: 0.96 %). These difference are not significant enough to say that one hexon fit is better then the other. In the case of the penton base this value is 1.82% which means that this fit is of slightly lower quality. The scale of the EM map was also tested using URO and this agreed with the value obtained using TMV as a magnification standard (i.e. 2.45 Å/pixel). From the optimised positions of the 4 hexons in the asymmetric unit plus the penton, the quasi-atomic model of the whole capsid was built using 532 icosahedral symmetry.

### **Generation of EM difference maps**

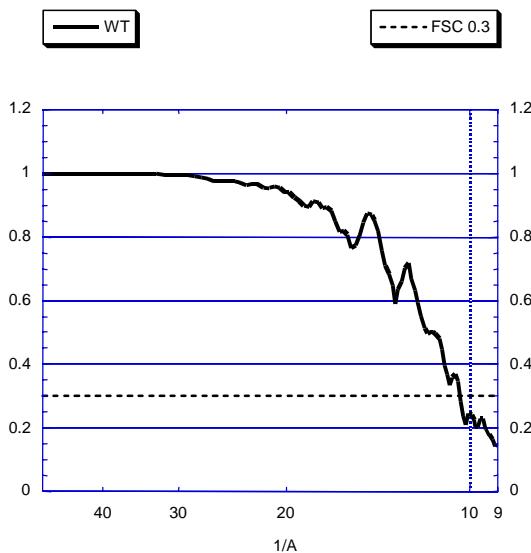
The difference maps were generated by using the quasi atomic model of the capsid. This model was Fourier filtered to the same resolution as the EM maps. We then created a binary mask (value of 0 for everything inside the capsid and 1 for everything outside) that included the right molecular weight of the capsid. Then, the different EM maps were multiplied by their corresponding binary masks to generate the difference maps.

### **Contacts between capsomers.**

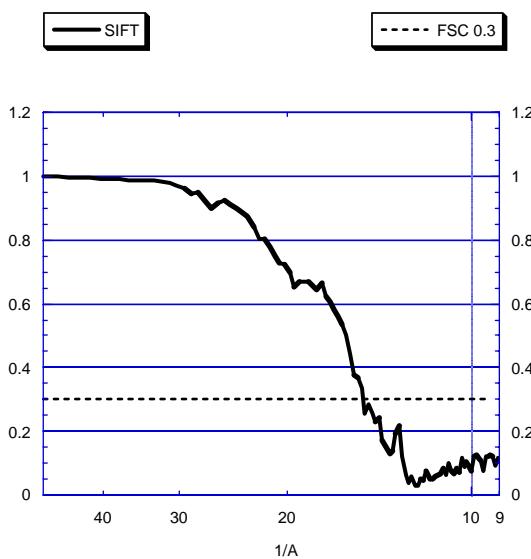
Protein contacts between capsomers were defined as any pair with atoms closer than 5 Å. This analysis was done by visual inspection of the EM density and the quasi-atomic model using “O” and verified using Contact v5.0 from the CCP4 suite.

**Determination of the resolution for the EM reconstructions of wt Ad5 (A), *sift* Ad5 pIX° (B) and *muft* Ad5 pIX° (C).**

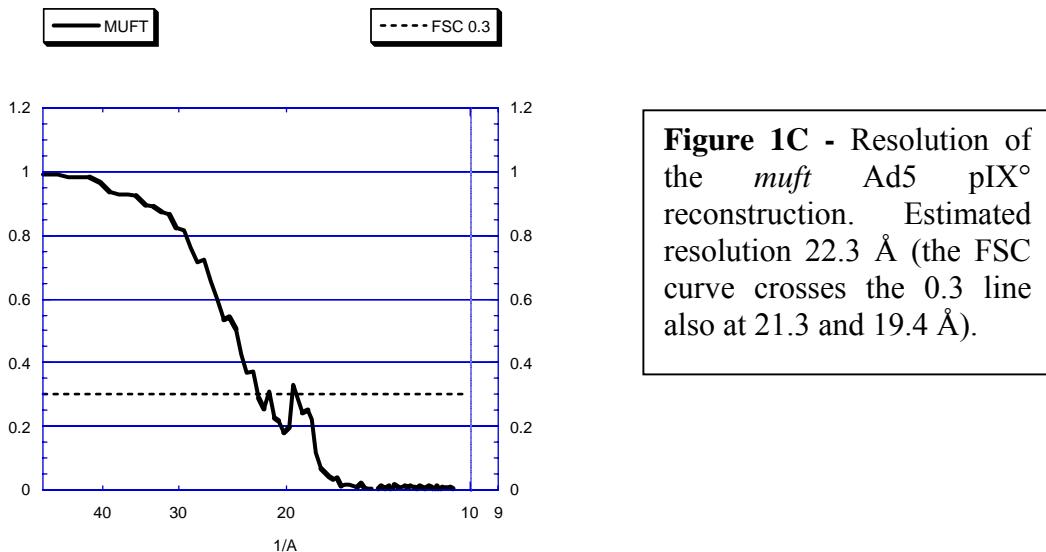
The resolution of the EM reconstructions structures was determined by Fourier Shell Correlation using a threshold value of 0.3 (Yang et al., 2003). The images used for the reconstructions were split into two halves in order to calculate two independent reconstructions. The Fourier shell correlations for these two reconstructions are indicated in Figure 1.



**Figure 1A** - The resolution of the wild type Ad5 EM reconstruction was estimated by Fourier shell correlation (FSC). The horizontal dashed line represents a conservative estimate where the power in the EM map falls to a FSC of 0.3. The estimated value for the resolution is 10.3 Å.



**Figure 1B** - Resolution of the *sift* Ad5 pIX° reconstruction. Estimated resolution: 15.0 Å.



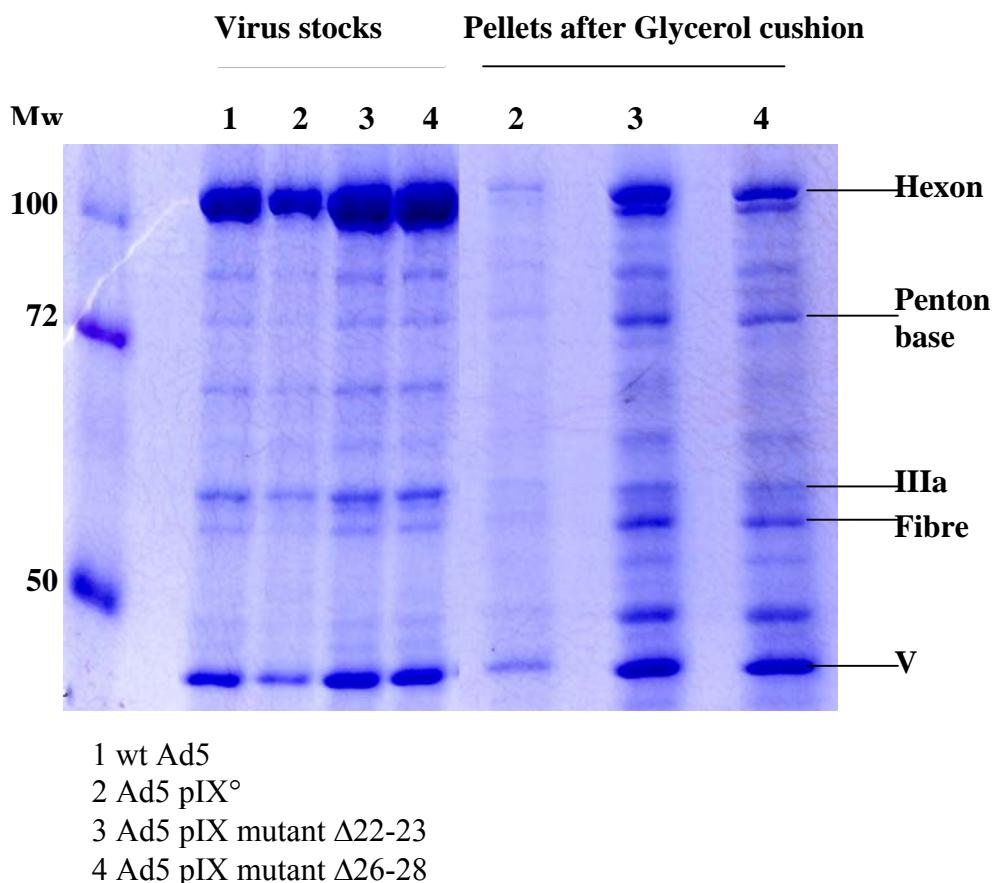
**Figure 1C** - Resolution of the *muft* Ad5 pIX° reconstruction. Estimated resolution 22.3 Å (the FSC curve crosses the 0.3 line also at 21.3 and 19.4 Å).

**Biochemical evidence that most of protein IIIa in Ad5 pIX° virus and in Ad5 viruses mutated in the gene for pIX, is dissociated from virus after freezing/thawing.**

The *sift* pIX mutant virus neither contained pIX, which was expected since the gene for this protein had been deleted, but also protein IIIa. We analysed stocks of wt, pIX° and other mutant Ad5 viruses that had short deletions in the gene for pIX (Ad5 pIX mutant Δ22-23 and mutant Δ26-28, described in Rosa-Calatrava et al., 2001). We analysed the same viruses on SDS-PAGE after pelleting them through a 20% glycerol cushion in order to remove viral proteins that have dissociated from the capsid (Figure 2). All viruses had been frozen/thawed a single time (so comparable with the *sift* pIX° virus sample used for the cryo-EM reconstruction). In all virus stocks the band for protein IIIa was much stronger than that for the fibre, in agreement with the fact that there are 60 IIIa molecules per virus particle compared with only 36 fibre monomers per virion. After the centrifugation step, only a small amount of the pIX° mutant virus was found in the pellet but much more of the other two mutant viruses. However, in the pellets of all mutant viruses, the ratio between the densities for fibre and protein IIIa had been inverted with a strong band for fibre and only a weak band for IIIa. This indicates that these viruses had lost a significant amount of their protein IIIa, probably through the freezing/thawing treatment. We were not able to obtain virus

preparations that had lost all protein IIIa. This may imply that some IIIa was retained in the pIX° virus that was analysed by cryo-EM. However, if the occupancy of IIIa was significantly lower than that of the other virus proteins, the cut-off values for the EM density would have resulted in elimination of the density for IIIa. Penton base protein and protein IIIa are also the first proteins that dissociate from wt Ad5 when incubated at elevated temperatures (Rexroad et al., 2003).

**Figure 2;** SDS Page analysis of wt and mutant Ad5 viruses, stock solutions and after pelleting the virus through a glycerol cushion. Numbers on the left indicate molecular weight markers whereas some of the viral proteins are indicated on the right.



## References

- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. and Leith, A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol*, **116**, 190-199.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, **47 (Pt 2)**, 110-119.
- Navaza, J., Lepault, J., Rey, F.A., Alvarez-Rua, C. and Borge, J. (2002) On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr D Biol Crystallogr*, **58**, 1820-1825.
- Rexroad, J.; Wiethoff, C.M., Green, A.P., Kierstead, T.D., Scott, M.O. and Middaugh, C.R. (2003). Structural stability of adenovirus type 5. *J Pharm Sci* **92**, 665-678.
- Rosa-Calatrava, M., Grave, L., Puvion-Dutilleul, F., Chatton, B. and Kedinger, C. (2001). Functional analysis of adenovirus protein IX identifies domains involved in capsid stability, transcriptional activity and nuclear reorganisation. *J Virol* **75**, 7131-7141.
- Rux, J.J. and Burnett, R.M. (2000) Type-specific epitope locations revealed by X-ray crystallographic study of adenovirus type 5 hexon. *Mol Ther*, **1**, 18-30.
- Wriggers, W. and Birmanns, S. (2001) Using situs for flexible and rigid-body fitting of multiresolution single-molecule data. *J Struct Biol*, **133**, 193-202.
- Yang, S., Yu, X., Galkin, V.E. and Egelman, E.H. (2003). Issues of resolution and polymorphism in single-particle reconstruction. *J. Struct. Biol.* **144**, 162-171.
- Zubieta, C., Schoehn, G., Chroboczek, J. and Cusack, S. (2005) The structure of the human adenovirus 2 penton. *Mol Cell*, **17**, 121-135.

## II. 2.2 La microscopie électronique et l'analyse d'images

Nous avons procédé à une étude en MET par cryoEM. Les négatifs enregistrés après sélection ont été scannés. Le traitement et l'analyse peuvent ainsi débuter avec la sélection des particules. Pour les 3 lots de virus imagés en CryoME, les particules sont extraites manuellement à l'aide du logiciel **X3D** (Conway et al., 1993) dans une sélection carrée de 427 pixels de coté. Une sélection circulaire d'un rayon de 209 pixels, entourée d'un masque doux de 4 pixels est appliquée à chaque particule. Les particules sont extraites dans des fichiers de type PIF (Purdue Image Format). Les spectres de puissance sont calculés avec le programme **SUMPS** (Conway & Steven, 1999) puis les CTF sont modélisées à partir des valeurs des spectres de puissance dans le logiciel **Kaleidagraph** (Synergy software, <http://www.synergy.com>). Les courbes calculées nous donnent la position des zéros de la CTF. La défocalisation est estimée pour chaque image à l'aide du programme **CTFZERO** (Conway & Steven, 1999) en fixant la part de contraste d'amplitude dans les images à 7% pour la cryoME. Enfin la valeur de défocalisation est utilisée dans **CTFMIX** (Conway & Steven, 1999) pour appliquer une inversion de phase jusqu'à 9 Å. La valeur de la défocalisation ainsi que les paramètres de la gaussienne déterminée dans Kaleidagraph sont utilisés pour déconvoluer totalement les images jusqu'à 9 Å de résolution (Figure 19). L'étape suivante consiste à déterminer les centres et les orientations des particules afin de réaliser une reconstruction 3D. L'orientation d'un objet icosaédrique est par convention définie selon les angles d'Euler depuis 1968 (Figure 33) (Klug & Finch, 1968). La détermination de l'orientation des particules est réalisée uniquement dans l'unité asymétrique. Le processus de l'analyse d'images est basé sur le théorème de la projection centrale (Bragg, 1929). Ce théorème démontre que la TF de la projection d'un objet 3D est une section centrale de la TF 3D de l'objet. En connaissant l'orientation de la projection, on peut déterminer l'orientation de la section dans la TF 3D et lorsque l'espace inverse est peuplé de suffisamment de sections, une TF inverse permet de calculer un modèle 3D de l'objet étudié (Figure 35).

La détermination des orientations des particules est réalisée par comparaison des images « flippées » avec les reprojctions d'un modèle 3D. Dans cette approche, un premier modèle 3D de la capsid de l'Adénovirus doit être calculé. Celui-ci est calculé par la méthode des lignes communes sur des particules indépendantes observées en coloration négative (2% AmMb). Cela permet d'orienter les particules à symétrie icosaédrique sans modèle de départ. Les particules sont sélectionnées de la même manière que précédemment. L'orientation de 10 particules indépendantes

est déterminée à l'aide du programme EMICOFV (Fuller et al., 1996). Pour chaque particule, une reconstruction 3D est calculée. Le modèle montrant clairement un arrangement en  $T= 25$  des capsomères est conservé comme modèle de départ pour la suite de l'analyse.

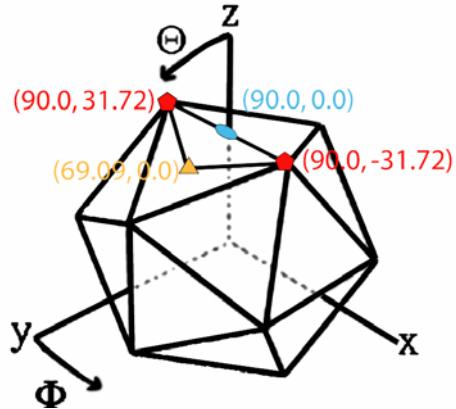
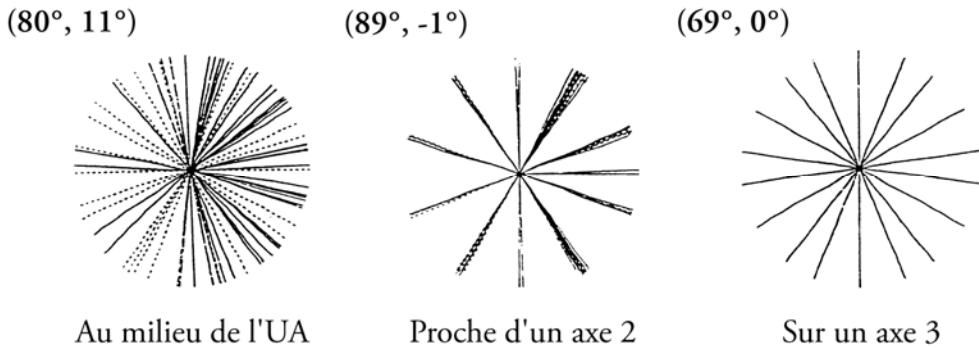


Figure 33: Convention d'orientation d'un icosaèdre d'après Klug & Finch, 1968.

Les axes X, Y et Z passent tous par des axes 2 de l'icosaèdre. Les rotations  $\Theta$  et  $\Phi$  se font respectivement autour des axes X et Z. La troisième rotation  $\Psi$  est appliquée suivant le nouvel axe Z. Le pentagone (en rouge), le triangle (en orange) et l'ellipse (en bleu) représentent respectivement les axes 5, 3 et 2. Ces axes encadrent l'unité asymétrique. Les valeurs entre parenthèses correspondent aux angles  $\Theta$  et  $\Phi$  pour chaque axe de symétrie. Pour une unité asymétrique,  $69,09^\circ < \Theta < 90^\circ$  ;  $-31,72^\circ < \Phi < 31,72^\circ$  ;  $0^\circ < \Psi < 180^\circ$ .

La méthode des lignes communes. Cette méthode, inventée par Crowther en 1971, est basée sur le théorème de la section centrale ainsi que sur les symétries icosaédriques (Crowther, 1971). Puisque une image en MET est une projection d'un objet 3D, et la TF de cette projection est une section centrale de la TF 3D de l'objet, l'intersection de deux sections centrales différentes est une droite qui coupe l'espace de Fourier en son centre. Cette droite forme ce que l'on appelle une ligne commune. Les amplitudes et les phases de la TF le long de cette ligne sont en effet identiques entre les deux sections centrales. Pour un objet icosaédrique, il existe deux types de lignes communes : les lignes communes intra-particulaires et inter-particulaires. Les lignes intra-particulaires proviennent de la présence de symétries icosaédriques sur une seule particule alors que les lignes inter-particulaires proviennent de la présence des symétries sur des particules différentes. L'application d'une symétrie et de son inverse sur une particule génère une paire de lignes communes intra-particulaires. La symétrie icosaédrique génère un total de 37 paires dont deux paires par axe d'ordre 5 (6), 1 paire par axe d'ordre 3 (10) et 1 paire par axe d'ordre 2 (15).

La technique de Crowther repose sur la similitude des phases le long des lignes communes intra-particulaires. On connaît la répartition des lignes communes intra-particulaires pour toutes les orientations de l'icosaèdre donc on peut les comparer aux lignes communes calculées expérimentalement. Dans la pratique, on recherche l'orientation pour laquelle les résidus de phase le long des lignes communes sont les plus faibles. Cette méthode marche assez bien pour les particules orientées au centre de l'unité asymétrique mais fonctionne très mal pour les particules proches des axes de symétrie où les lignes communes sont dégénérées (Figure 34) (Fuller et al., 1996).

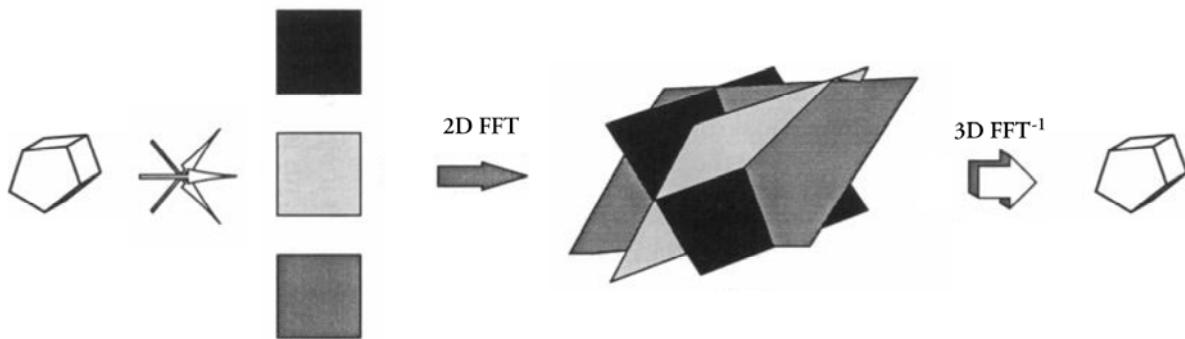


**Figure 34: Distribution des lignes communes dans une projection pour différentes orientations de l'icosaèdre.**

L'orientation des angles  $\Theta$  et  $\Phi$  est indiquée entre parenthèse. Pour chaque orientation, les 37 paires de lignes communes intra-particulaires sont représentées. De gauche à droite, plus on se rapproche d'un axe de symétrie, plus les lignes communes sont dégénérées ou superposées. Les lignes représentées en pointillés correspondent aux lignes communes partiellement superposées des 15 axes d'ordre 2.

Pour obtenir des résultats satisfaisants avec la méthode des lignes communes, les particules dont on recherche l'orientation doivent être bien centrées.

Calcul d'une reconstruction 3D. A partir de chacune des particules et de leur orientation déterminée dans l'étape précédente, on calcule une reconstruction 3D dans le programme **EM3DR**. Ce programme est aussi utilisé chaque fois que l'on veut réaliser une reconstruction 3D d'un objet icosaédrique lors des étapes suivantes du traitement des images. Le principe de reconstruction 3D est aussi basé sur le théorème de la projection centrale (Figure 35).



**Figure 35: Principe de Reconstruction 3D dans l'espace de Fourier basé sur le théorème de la projection centrale d'après Chiu et al., 1993.**

L'orientation des images est déterminée, la TF 2D est calculé pour chaque image. Les sections centrales leur correspondant sont combinées et une TF 3D inverse est calculée pour reconstituer un modèle 3D de l'objet de départ.

Dans ce programme, l'espace est décrit en coordonnées cylindriques où l'axe z est situé sur l'axe 5 de l'icosaèdre. La mise en place de ce repère impose la présence d'un axe de symétrie 5 selon z ainsi que de tous les axes d'ordre 2 du plan x, y. Une TF de chaque particule est calculée puis celle-ci est utilisée pour peupler l'espace de Fourier dans l'orientation de la reprojection correspondante

(Figure 35). L'échantillonnage de l'espace réciproque en coordonnées cylindriques permet la décomposition de la TF( $S_r, S_z, S\phi$ ) totale en une somme de fonctions  $G_n(S_r, S_z)$ , indépendantes de l'angle  $\phi$ . Par la suite, le calcul des fonctions  $G_n$  est réalisé par inversion d'une matrice (Figure 36).

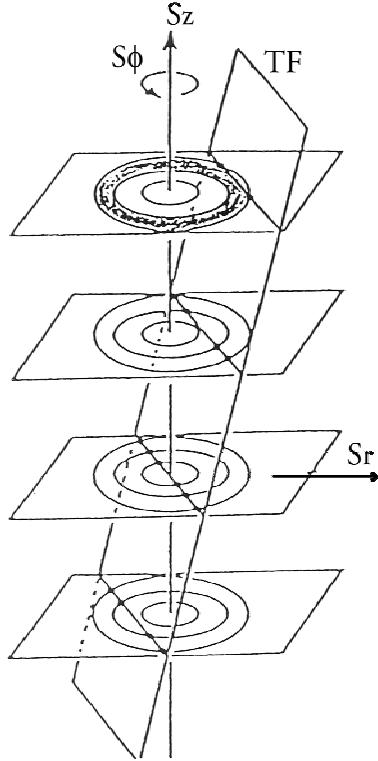


Figure 36: Passage à la TF en coordonnées cylindriques ou Fourier-bessel.

En coordonnées cylindriques, la TF ne dépend plus de l'angle  $\phi$  mais correspond à la somme des fonctions  $G_n$  comme le montre la formule de la TF :

$$TF(S_r, S_\phi, S_z) = \sum_n G_n(S_r, S_z) \exp^{in(S_\phi + \frac{\pi}{2})}$$

La valeur de l'inverse des valeurs propres de cette matrice renseigne sur le peuplement de l'espace de Fourier par les TF des particules de départ (cf. qualité du modèle 3D obtenu). Enfin, une TF inverse en coordonnées cylindriques permet de calculer un modèle 3D à partir des images expérimentales. La reconstruction sert ensuite de référence pour l'alignement d'un jeu de particules observées en coloration négative.

Affinement des orientations grâce à la comparaison avec les reprojctions d'un modèle. Cette étape est réalisée avec le programme **PFT** (Polar Fourier Transform). PFT permet de déterminer l'orientation d'une particule par comparaison avec les reprojctions d'un modèle. Ce programme crée une base de données des reprojctions d'un modèle défini selon un pas en degrés donné par l'utilisateur puis affine l'orientation et le centre des particules expérimentales dans l'espace de Fourier par comparaison avec la base de données (Figure 37).

La base de données des reprojctions est réalisée après passage en coordonnées polaires car le programme travaille dans l'espace de Fourier. Cet échantillonnage permet de limiter les calculs à l'épaisseur de la capsidé mais également à une certaine gamme de résolution (Figure 38).

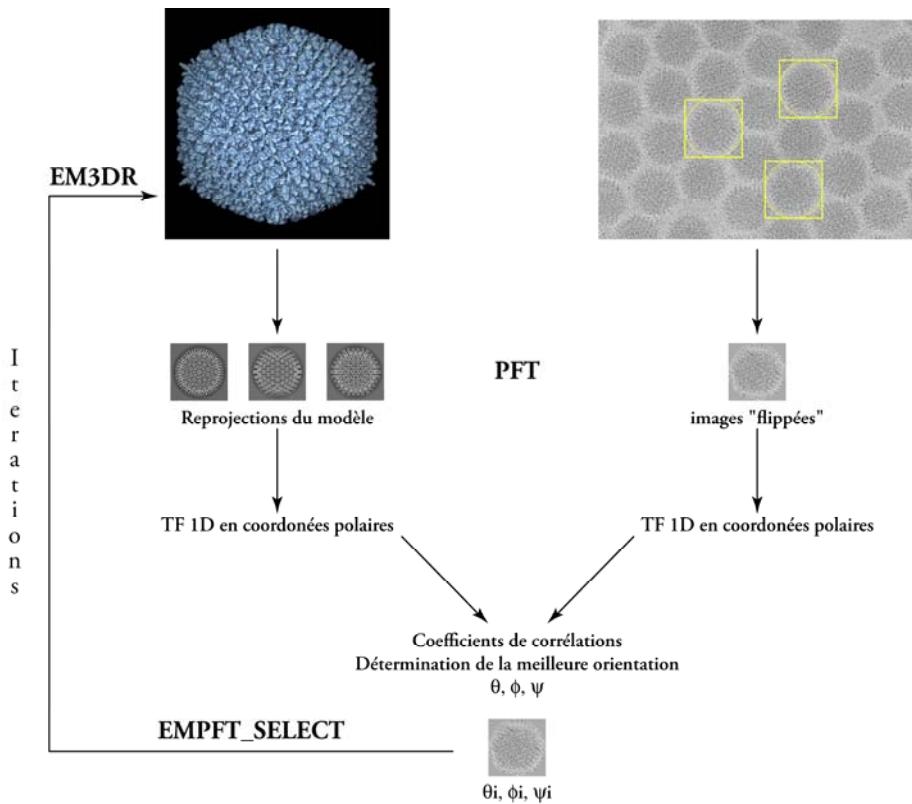


Figure 37: Principe de la détermination de l'orientation des particules avec PFT.

Le modèle est reprojeté selon un pas angulaire défini pour créer une base de données. La base de données et les particules à orienter sont échantillonnées en coordonnées polaires dans l'espace de Fourier puis sont comparées. Pour chaque particule, la meilleure orientation est sélectionnée et stockée dans un fichier texte. Un tri sur la base de la valeur des CC est réalisé avec EMPFT\_SELECT puis une reconstruction 3D est calculée avec EM3DR.

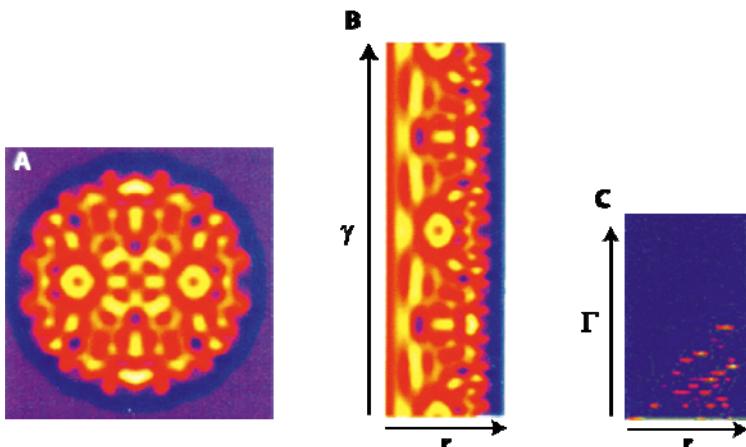


Figure 38: Passage à la TF en coordonnées polaires d'après Fuller et al., 1996.

(A) Projection du modèle 3D selon un angle donné. (B) Echantillonage radial de la reprojeciton. (C) Transformée de Fourier en coordonnées polaires de la reprojeciton radiale obtenue en B.

Le principe d'utilisation de ce programme est une recherche large des orientations et des centres dans les premiers cycles d'affinement puis la recherche est réduite uniquement aux orientations en fixant les centres des particules. Le pas de reprojeciton est dans un premier temps plus grossier (de l'ordre de  $2^\circ$ ) puis il est réduit au fur et à mesure de l'amélioration de la qualité de la reprojeciton (jusqu'à  $0,2^\circ$  environ suivant la mémoire vive disponible). La procédure de recherche des orientations peut être arrêtée quand les orientations des particules ne varient plus au cours des cycles d'affinement. La comparaison de la TF en coordonnées polaires des images avec les

reprojections du modèle donnent trois coefficients de corrélation différents : PFTCC, PRJCC et CMPCC.

Le premier appelé PFTCC décrit la qualité des angles  $\theta$  et  $\phi$ , le deuxième noté PRJCC correspond aux valeurs de centre x, y ainsi qu'à l'angle psi. Le dernier CMPCC prend en compte toutes ces valeurs. Pour chaque particule, l'orientation déterminée ainsi que les valeurs des CC correspondantes sont stockées dans un fichier texte.

Les particules sont ensuite soumises à une sélection selon une valeur seuil de CC définie par l'utilisateur à l'aide du programme EMPFT\_SELECT. Généralement, nous utilisons soit un seuil de CC égal à la moyenne des CC ou un seuil un peu plus drastique équivalent à la moyenne à laquelle on ajoute la valeur de la déviation standard des CC. Les orientations des particules sélectionnées, définies avec les angles d'Euler ([Figure 33](#)), sont retranscrites dans un fichier texte de sortie puis utilisées comme entrée par le programme de reconstruction EM3DR.

Cela a permis d'obtenir une reconstruction 3D de la capsid du HAdV-5 à partir de 120 particules en coloration négative à 25 Å de résolution. Cette reconstruction sert ensuite de modèle pour la recherche des centres et des orientations des particules extraites des images de CryoME (article 1).

*Utilisation des programmes PFT2 et EM3DR2 pour les images de CryoME.* Depuis 2005, nous avons introduit dans le processus d'analyse d'images deux nouveaux programmes PFT2 et EM3DR2. Ces deux programmes sont une évolution des versions antérieures de PFT et EM3DR. Ces deux programmes nécessitent un temps de calcul beaucoup plus important mais le résultat obtenu est de meilleure qualité pour plusieurs raisons. Nous les utilisons de la même manière que PFT et EM3DR ([Figure 37](#)). PFT2 détermine l'orientation de particules prises proches de la focalisation de façon plus précise que son prédecesseur. De plus, sur le même système informatique, on peut définir un pas angulaire plus fin qu'avec PFT (fonction de la capacité en mémoire vive de l'ordinateur utilisé). Ceci a une importance significative lorsqu'on cherche à orienter de gros objets comme la capsid des Adénovirus. En effet, plus l'objet à orienter est grand plus le pas de reprojection doit être petit si on désire minimiser l'erreur ([Tableau 3](#)). De plus, le programme calcule plusieurs orientations pour chaque particule et les évaluent à l'aide des mêmes critères de corrélation que dans PFT (PFTCC, PRJCC et CMPCC). Une figure de mérite calculée à partir de ces 3 critères permet de garder la meilleure orientation. Le véritable critère de sélection est donc ici la valeur de la figure de mérite.

Ceci permet de ne pas éliminer des particules que le programme PFT orientait de façon incorrecte. La sélection des particules retenues est réalisée à l'aide d'un programme BSOFT appelé BPARTSEL. On ne conserve pour la reconstruction qu'entre 50 et 70 % des particules totales. Les particules retenues servent à calculer une reconstruction 3D à l'aide d'EM3DR2. L'utilisation conjointe de ces programmes permet souvent d'améliorer la résolution de la reconstruction finale de 1 ou 2 Å.

Qualité du modèle 3D obtenu. La qualité du modèle final dépend de plusieurs facteurs. Cela dépend tout d'abord de la qualité des images de départ ainsi que de la qualité de leur CTF. Le pas de reprojctions lors de la recherche des orientations avec PFT a une grande influence sur la résolution de la reconstruction finale (Tableau 3). Le peuplement correct de l'espace de Fourier par des sections centrales balayant la totalité de l'unité asymétrique est aussi très important. En effet, il représente le nombre d'orientations différentes dont nous disposons pour la reconstruction. Ce critère est évalué par la valeur de l'inverse des valeurs propres dans le programme EM3DR. Quand cette valeur est égale à 1, le nombre d'orientations dont nous disposons est juste suffisant pour permettre le calcul d'une reconstruction. Quand cette valeur est inférieure à 1, nous disposons de plus d'informations que nécessaire donc l'information va être moyennée et le résultat de la reconstruction sera de meilleure qualité. Il est aussi possible d'estimer la résolution de la reconstruction par la méthode de corrélation des coques de Fourier, traduction de Fourier Shell Correlation (FSC) (Harauz & van Heel, 1986) issue de la méthode de corrélation annulaire de Fourier (Saxton & Baumeister, 1982). Le principe de cette méthode est de séparer aléatoirement les particules utilisées pour la reconstruction en deux lots et de calculer deux reconstructions 3D à partir de ces deux lots. Le programme BRESOLVE du package BSOFT est ensuite utilisé pour comparer dans l'espace de Fourier les valeurs des phases dans des coques pour les 2 modèles à chaque valeur de fréquences spatiales ou résolutions. La valeur des FSC diminue des basses vers les hautes fréquences. On définit la limite de résolution au passage d'une valeur seuil de FSC correspondant à une valeur de fréquence spatiale. Plusieurs critères seuil existent. Le premier critère fixe la limite de résolution à 50% de corrélation entre les 2 reconstructions 3D (FSC = 0,5). Cette valeur de FSC correspond à un rapport signal sur bruit de 1. En 2003 il a été montré que la comparaison des 2 volumes introduit un biais qui est responsable d'une sous évaluation de la résolution réelle de la reconstruction à FSC = 0,5 (Yang et al., 2003). En effet, seule la moitié des données a servi à calculer les reconstructions qui sont donc de moins bonne qualité. La limite de résolution est donc étendue à la valeur de FSC à 30 % de

corrélation (FSC = 0,3). Le dernier critère, noté  $3\sigma$ , est celui donné quand la courbe de FSC coupe la courbe correspondant à 3 écarts types de plus que la courbe de corrélation de 2 volumes ne contenant que du bruit (Orlova et al., 1997). Dans mon laboratoire, nous utilisons le critère FSC = 0,3 introduit par Yang et al. pour la détermination de la résolution de toutes nos reconstructions 3D.

		Pas de reprojection en degrés				
		2,0°	1,0°	0,5°	0,25°	0,1°
D (Å)	250	8,7	4,4	2,2	1,1	0,43
	500	17,5	8,7	4,4	2,2	0,87
	1000	34,9	17,5	8,7	4,4	1,75
	2000	69,8	34,9	17,5	8,7	3,49

*Tableau 3: Erreur sur le positionnement d'un point en fonction du pas angulaire de reprojection.*

Ces valeurs sont calculées d'après Conway & Steven, 1999. Les valeurs du tableau sont en Å. Pour une particule de 1000 Å de diamètre, l'erreur sur le positionnement d'un point avec un pas de 0,5 ° est 8.7 Å ce qui va déterminer la résolution maximale ne pouvant pas être dépassée.

### II.3 Conclusion

Nous avons obtenu trois reconstructions 3D pour les 3 échantillons WT, sift et muft dont nous disposions à, respectivement, 10 Å, 16 Å et 22 Å de résolution. A la date de publication de nos travaux, la capsidé entière d'un Adénovirus n'avait jamais été reconstruite et analysée à une résolution aussi haute. Le modèle à haute résolution de la capsidé du HAdV-5 nous a permis de recalier les structures atomiques de protéines majeures, Hexon du HAdV-5 et Base du Penton du HAdV-2 afin de reconstituer un modèle quasi-atomique. Ce modèle, constitué uniquement de protéines majeures, nous a tout d'abord permis de décrire, de façon précise, les zones de contacts entre toutes les protéines majeures de la capsidé. Par la suite, nous avons calculé des cartes de différence entre les reconstructions 3D des échantillons WT, sift, muft et le modèle quasi-atomique filtré à la même résolution que les reconstructions. La soustraction de la densité des protéines majeures ainsi que la comparaison de l'échantillon WT aux sift et muft nous a permis de décrire la position de certaines protéines mineures comme la protéine IX, IIIa, VI et VIII et de mettre à jour le schéma admis depuis 1993 concernant la position des protéines de capsidé (Stewart et al., 1993). Nous avons notamment pu préciser la structure et la localisation de la protéine VI et donner une localisation précise de la protéine VIII qui n'avait pas pu être déterminé auparavant. Enfin le modèle quasi-atomique en combinaison avec la carte de différence calculé avec la capsidé WT a montré les zones des protéines majeures (Hexon et Base du Penton) en contact avec les protéines mineures. Tous les contacts protéiques au sein de la capsidé que ce soit entre deux protéines majeures ou entre une protéine majeure et une protéine mineure, ont pu être évalués avec une précision de l'ordre de quelques acides aminés. Nos travaux sur la structure de la capsidé de l'Adénovirus de type 5 par combinaison entre la ME et le recalage des structures atomiques des protéines majeures ont eu un fort impact sur la connaissance fondamentale de la structure de ce virus. La description précise des contacts protéiques permet de déterminer correctement les zones de la capsidé susceptibles d'accepter des modifications en vue d'un reciblage du virus. En absence de données biochimiques précises combinées à des études structurales, des doutes persistent sur la position des protéines mineures dans la capsidé des Adénovirus en particulier pour les protéines IX et IIIa.

### ***III. Etude structurale de la capsid de l'Adénovirus Aviaire de type 1***

Cette partie est consacrée à l'étude structurale en MET de la capsid de l'Adénovirus aviaire de type 1 ou virus CELO. Comme je l'ai déjà mentionné en introduction, aucune séquence similaire à celle de la protéine IX des Adénovirus humains n'a été retrouvée dans le génome du virus CELO. On suppose donc que la capsid de ce virus ne contient aucune protéine IX. Nous nous sommes donc intéresser à la structure de sa capsid afin de comprendre comment une capsid d'Adénovirus peut s'assembler et se maintenir en absence de protéine IX à sa surface. Nous voulions vérifier si une protéine de virus CELO encore inconnue pourrait remplacer la protéine IX des *Mastadenovirus*. De plus, à chaque sommet de la capsid, on retrouve deux fibres différentes. Nous voulions étudier l'interaction de deux fibres avec une seule Base du Penton afin de comprendre comment une protéine pentamérique comme la base du Penton peut accommoder deux trimères de fibre sachant que chaque monomère de la base du Penton a la capacité de lier un monomère (Zubieta et al., 2005).

#### **III.1 Matériel et Méthodes**

##### ***III. 1.1 Purification du virus CELO***

Nous disposons de deux échantillons différents de virus CELO. Le premier provient d'une collaboration avec le Dr. Matthew Cotten (Axxima Pharmaceuticals, Allemagne). Le virus CELO, après purification sur gradient de Chlorure de Césium (CsCl), nous a été envoyé à -80°C dans une solution de conservation contenant 40% de glycérol. La présence du glycérol et de CsCl dans le tampon de congélation du virus ne permet pas de réaliser directement une étude en CryoME. Cet échantillon doit subir une étape de purification afin d'éliminer le glycérol. Le deuxième lot de virus provient d'une collaboration avec le Dr Patrick Langlois de l'Agence Française de Sécurité Sanitaire des Aliments à Ploufragan (AFSSA). Ce lot est produit sur une lignée cellulaire LMH d'hépatocarcinomes de poulet (Leghorn Male Hepatoma) jusqu'à lyse complète des cellules. Le virus est ensuite récolté dans le milieu de culture et envoyé à 4°C. Dans ce cas, il est nécessaire de purifier le virus des contaminants du milieu de culture avant des études en CryoME. J'ai donc mis au point un protocole de purification du virus à partir du milieu de culture.

Précipitation au Sulfate d'Ammonium. Cette étape n'est réalisée que sur le deuxième échantillon de CELO car elle permet d'éliminer une grande partie des contaminants du milieu de culture cellulaire dont la Sérum Albumine Bovine (SAB). Une précipitation au sulfate d'ammonium, adaptée des travaux de Schagen et collaborateurs (Schagen et al., 2000), est réalisée sur le milieu de culture. Le milieu est incubé de façon croissante jusqu'à 40 % de saturation en sulfate d'ammonium, 2h30 sous agitation à température ambiante. Dans ce cas, la protéine SAB, contaminant majeur du milieu de culture, reste soluble alors que le virus précipite. Le précipité contenant le virus est en suite culoté par centrifugation à température ambiante pendant 15 min à 1600 g. Le culot est recouvert par 1/40<sup>ème</sup> du volume de départ en PBS 1x, et stockée à 4° C. De cette manière, le précipité se resuspend lentement ce qui évite d'endommager le virus.

Gradient de Chlorure de Césium. Cette étape est réalisée pour les 2 types d'échantillons pour éliminer le glycérol ou terminer la purification du CELO après la précipitation. Le surnageant de l'étape précédente ou la solution de virus du Dr. Cotten est déposé au dessus d'un gradient de densité en CsCl formé de deux couches : une couche inférieure à une concentration de 4 M CsCl et une couche supérieure à 2,2 M dans un tampon 10 mM tris pH 7,5. Le gradient est ultracentrifugé à 210000 g pendant 12h dans une ultracentrifugeuse Beckman-Coulter munie d'un rotor SW41 ou SW55 en fonction du volume à centrifuguer. Cette méthode est adaptée des travaux de Kanegae et collaborateurs (Kanegae et al., 1994). Après centrifugation, le virus intact se situe à l'interface entre les 2 couches de CsCl (Figure 39a). Le virus étant dense, il forme donc une bande concentrée opalescente juste au dessus de la deuxième couche. Il est récolté par ponction à l'aide d'une seringue au niveau de cette bande. Le gradient contient plusieurs autres bandes contaminantes qui migrent à une densité plus faible dans le gradient.

Désalage et Concentration de la Solution Virale. La dernière étape de la purification constitue le désalage et la concentration de la solution virale. Il est nécessaire de désaler l'échantillon afin d'éliminer la forte concentration de CsCl présente qui ne permet pas de réaliser une étude en CryoME car elle entraîne une perte complète de contraste en CryoME. J'ai utilisé plusieurs méthodes de désalage sans réel succès. J'ai tout d'abord réalisé une dialyse classique dans un tampon tris 10 mM pH 7,5 mais cette technique trop rapide provoque un éclatement des particules virales. J'ai par la suite réalisé une dialyse progressive dans un bain de 600 ml d'un tampon 10 mM Tris pH 7,5 contenant 3 M NaCl pour compenser la dilution du CsCl. Dans ce bain est ajouté environ 3 l de

tampon Tris 10 mM pH 7,5 à une vitesse de 2 ml/min, à l'aide d'une pompe péristaltique. Lors de l'ajout, la solution de dialyse est homogénéisée par agitation magnétique. Cette méthode permet de diluer progressivement le CsCl ainsi que le NaCl sans entraîner de choc osmotique pour le virus. Cette dialyse se termine par deux bains de 1 l de tampon Tris pH 7,5 150 mM NaCl. Cette étape de désalage est vraiment déterminante pour la qualité finale de l'échantillon. En effet, le virus CELO semble beaucoup moins stable au cours de la dialyse qu'un Adénovirus humain de type 5. Il m'a été difficile même avec la dialyse progressive d'obtenir un échantillon intact en fin de purification. La dialyse progressive a l'inconvénient majeur de diluer fortement l'échantillon. Il est donc nécessaire de concentrer le virus avant d'observer le virus en CryoME. J'ai donc essayé plusieurs méthodes de concentrations. La première, à l'aide d'un concentrateur, n'a pas donné de bons résultats. Le virus même à faible vitesse est détruit au cours de la concentration. Par la suite, j'ai utilisé une cassette d'absorption Vivapore5 (Vivascience) munie d'une membrane en cellulose régénérée (c/o = 30kDa). Cette méthode, beaucoup plus longue, ne détruit pas le virus mais une grande quantité d'échantillon est perdue au cours de cette étape. Je discuterai de la purification complète de ce virus ainsi que de sa stabilité dans la partie Résultats et Discussion. Après concentration, la solution virale est conservée à température ambiante plutôt qu'à 4°C pour prévenir de l'agrégation du virus.

### *III. 1.2 La microscopie électronique et l'analyse d'images*

La solution concentrée de virus est tout d'abord observée en coloration négative dans 2% AmMb afin de vérifier l'intégrité des capsides avant congélation. Les contrôles sont réalisés sur le MET microscope JEOL 1210-EX de l'EMBL à un grossissement nominal de 40 000 fois. Ce contrôle permet également d'évaluer sa concentration.

La concentration en virus étant trop faible pour permettre la congélation sur une grille à trous classique, le virus est préparé pour la cryoME sur une grille Quantifoil recouverte d'une couche de carbone, de la même façon que le HAdV-5 pIX° (cf. II.2.2).

La prise d'images est réalisée de la même manière que pour tous les échantillons d'HAdV-5, sur un MET Philips CM200 à un grossissement calibré de 28 600 fois. Les négatifs sélectionnés ont été numérisés à un pas de 14 µm ce qui correspond à une taille réel de pixel de 5,1 Å. La sélection manuelle des particules est réalisée dans X3D dans une boîte de 243 pixels de côté. La correction de CTF est réalisée jusqu'à 20 Å de résolution. Les images flippées et déconvolées sont utilisées dans PFT et EM3DR pour calculer une reconstruction 3D de la capsid du CELO en utilisant la

reconstruction 3D de l’HAdV-5 à 10 Å de résolution comme modèle de départ. Nous avons ensuite continué l’analyse des images avec PFT2 et EM3DR2 en espérant améliorer la qualité de la reconstruction finale.

### *III. 1.3 Le fitting*

J’ai utilisé le programme URO (Navaza et al., 2002) afin de reconstituer un modèle quasi-atomique de la capsid du CELO en recalant la structure de l’Hexon du HAdV-5 dans la reconstruction 3D que nous avions obtenu. Pour cela, j’ai manuellement replacé les 4 Hexons de l’UA dans la reconstruction avec « O » puis j’ai utilisé URO en combinaison avec les matrices de rotation icosaédriques pour améliorer mon placement manuel (Jones et al., 1991). La recherche de la position optimale dans URO a été réalisé jusqu’à une résolution de 30 Å. En 2007, la structure de l’Hexon du CELO a été publiée à 4 Å de résolution (Xu et al., 2007). J’ai donc essayé à la fin de ma thèse de recaler cette structure au sein de la reconstruction 3D que nous avions obtenu, de la même manière qu’avec la structure de l’Hexon du HAdV-5.

## **III.2 Résultats et discussion**

Purification du virus. La purification de ce virus a été une étape assez complexe à mettre au point. Je disposais de très peu d’échantillons au départ et le rendement de purification est faible. L’échantillon fourni par le Dr Cotten, qui a permis de réaliser la reconstruction 3D, représente 600 µl de virus purifié à une concentration inférieure à  $10^9$  particules/ml au départ dans 40% (v/v) glycérol. Le virus fourni par le Dr. Langlois est récupéré après la lyse dans le milieu de culture cellulaire correspondant à 4 boites de 25 cm<sup>2</sup> (V= 75 ml). Dans les deux cas, Il a été difficile, après désalage et concentration, d’obtenir la concentration en virus nécessaire pour réaliser une étude satisfaisante en CryoEM (au moins  $10^{10}$  particules/ml). De plus, le virus CELO semble, in vitro, beaucoup moins stable que les Adénovirus humain de type 2 et 5 habituellement manipulés en laboratoire. Après chaque étape de purification, nous avons visualisé l’échantillon en MET par la méthode de coloration négative afin de déterminer son état (intact, partiellement désassemblé ou cassé). Dans ces contrôles, on considère que l’échantillon est en cours de désassemblage lorsque le colorant (AmMb) commence à pénétrer à l’intérieur des virus. Dès lors, on ne peut pas l’utiliser pour réaliser une étude en CryoME.

Les deux premières étapes de la purification (précipitation et gradient de CsCl) n'ont présenté aucun problème particulier pour la stabilité de l'échantillon viral. Le virus ponctionné au niveau du gradient était intact (Figure 39 b et c). L'échantillon fourni par le Dr. Cotten donne un bien meilleur résultat à l'issu du gradient, l'échantillon étant déjà purifié (Figure 39c). L'autre échantillon contient encore de nombreux contaminants visibles dans le fond de l'image de la Figure 39b.

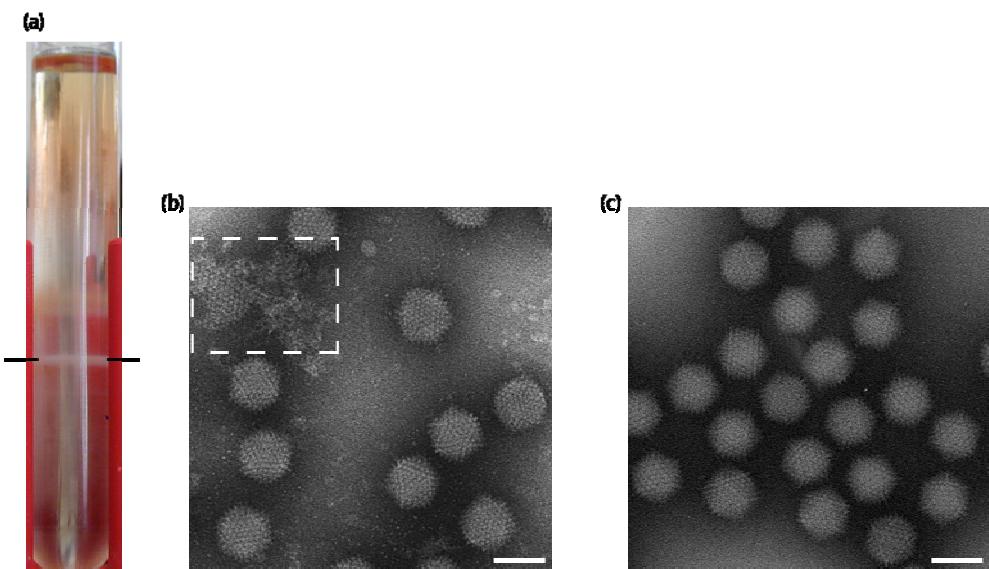


Figure 39: Purification du CELO sur gradient de CsCl.

(a) Photo du gradient de CsCl pour l'échantillon présent dans le milieu de culture (Dr. Langlois). Le trait noir indique la bande de virus intact. (b) et (c) Images en MET dans 2% AmMb de (b) bande de virus intact ponctionnée sur le gradient montrée en (a), (c) bande de virus intact récupérée après centrifugation de l'échantillon déjà purifié (Dr. Cotten). La barre d'échelle représente 100 nm. La zone encadrée en pointillés blancs montrent l'état d'une capsid virale abimée et cassée.

La difficulté majeure de la purification de ce virus se situe au niveau des étapes de désalage et de concentration. Le désalage a été réalisé par différentes méthodes de dialyse mais aucune n'a donné de résultats satisfaisants. Les contrôles en MET après dialyse classique ont toujours montré un échantillon détérioré où la majorité des virus sont désassemblés comme le montre la zone encadrée en pointillés blancs sur la Figure 39b. La dialyse progressive a donné de meilleurs résultats mais une partie des capsides virales sont tout de même cassées (Figure 39b). Par manque d'échantillon de départ, je n'ai pas pu tester les méthodes de désalage par retard comme les colonnes de type PD10 (Biorad) ou Microbiospin6 (Biorad). Ces méthodes utilisent une colonne constituée d'une résine ou d'un gel possédant une taille d'exclusion très faible. Le virus est exclu et passe dans le volume mort alors que les molécules d'une taille inférieure à la taille d'exclusion (glycérol, sels) sont retardées. J'ai pu constater ultérieurement leur efficacité sur un échantillon d'Adénovirus humain de type 5 (cf.

V.1.1). La dialyse entraîne la déstabilisation du virus alors que ces méthodes permettent de désaler un échantillon sans le détériorer.

En ce qui concerne la concentration, seule la méthode par absorption a permis d'obtenir un échantillon intact. Toutes les méthodes par centrifugation se sont avérées infructueuses. Dans 100% des cas, le virus était complètement désassemblé, de façon similaire à la zone entourée en [Figure 39b](#).

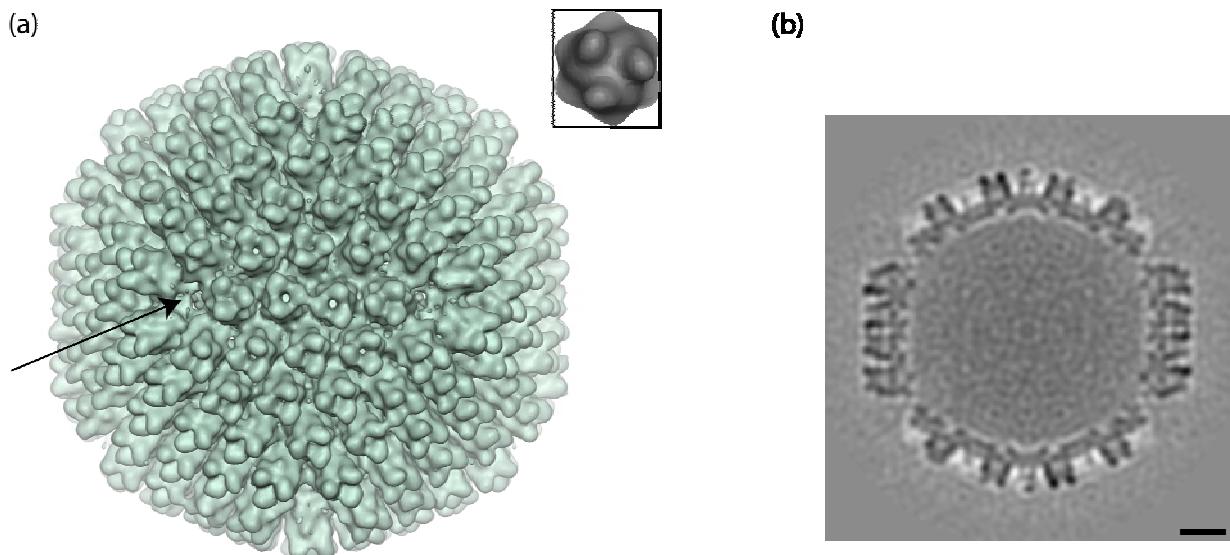
Le stockage de ce virus après concentration s'est également avéré délicat. Nous avons remarqué, lors des contrôles en coloration négative après concentration, que le virus CELO a tendance à agréger à 4°C. Les échantillons en fin de purification sont donc toujours conservés à température ambiante ce qui prévient de l'agrégation.

La purification d'un virus est beaucoup plus délicate que celle d'une protéine. Dans certains cas, l'agrégation et la dénaturation d'une protéine sont des phénomènes réversibles. Lorsqu'un Adénovirus commence à se désassembler, ce phénomène est un processus irréversible et l'échantillon devient inutilisable pour toutes études structurales. Il faut donc être particulièrement prudent au cours de la purification.

Modèle 3D de la capsid du FAdV-1 ou CELO. Nous avons étudié en CryoME un échantillon purifié de CELO à partir d'une aliquote fourni par le Dr. Cotten et enregistré un jeu de 25 paires focales. Le contrôle en MET par coloration négative montrait une grande majorité de capsides intactes comme en [Figure 39c](#). La concentration en virus étant très faible (inférieure à celle de l'Adénovirus de type 5 mutant pIX°), nous n'avons pu sélectionner qu'un total de 633 particules à partir de 14 négatifs scannés qui contenait le plus grand nombre de particules. A partir de ces images, nous avons obtenu un modèle 3D de la capsid du CELO à 30 Å de résolution (FSC 0,5) avec 426 de ces particules. Aucune reconstruction 3D n'a jamais été publiée sur le virus CELO. Le modèle 3D que nous avons obtenu constitue donc la première étude structurale réalisée à moyenne résolution sur la capsid complète de ce virus.

Le modèle est malheureusement incomplet puisqu'on remarque l'absence de Pentons au niveau des sommets de la capsid (Flèche sur la [Figure 40a](#)). Il ne nous a donc pas permis d'étudier l'interaction des deux fibres ni les protéines liées sous la Base du Penton. Le contrôle effectué en coloration négative pourtant des particules intactes puisque le colorant (AmMb) n'avait pas pénétré à l'intérieur des particules virales ([Figure 39c](#)). Malheureusement, même lors d'un contrôle en MET, nous ne pouvons pas déterminer de façon précise la présence de Pentons. En effet, le

colorant ne pénètre dans les capsides que lorsqu'une partie de l'ADN viral est sorti. A ce stade de désassemblage, les Pentons sont déjà détachés de la capsid. Seule une reconstruction 3D va réellement montrer l'état de l'échantillon que l'on a étudié en MET.



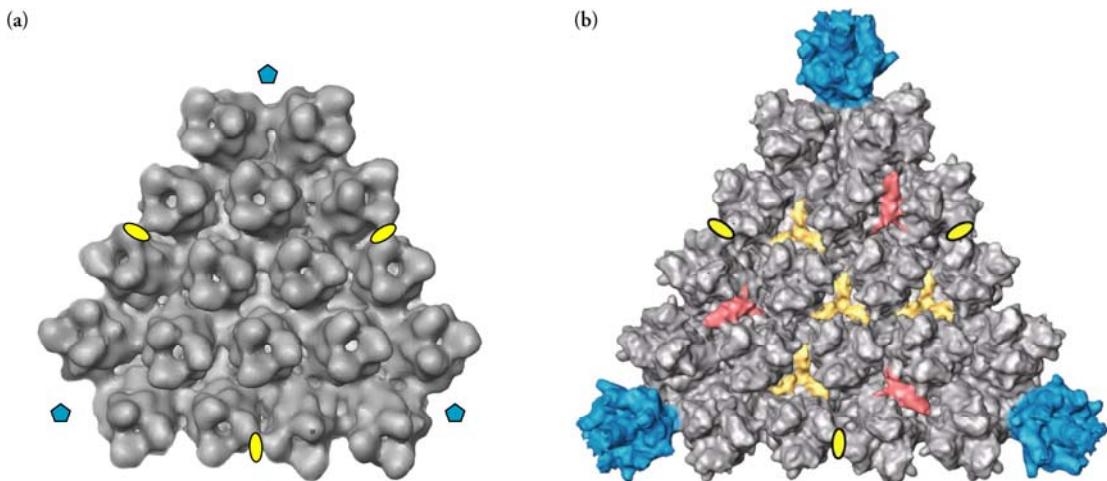
**Figure 40: Reconstruction 3D du CELO.**

(a) Reconstruction 3D finale de la capsid du CELO vue selon l'axe 2 de l'icosaèdre. La flèche montre l'absence de Pentons dans la reconstruction obtenue ; Encart : Structure cristallographique de l'Hexon du virus CELO filtrée à 30 Å de résolution avec SPIDER (Frank et al., 1996). (b) Vue transversale en coupe de la Reconstruction 3D obtenue pour le CELO.

On remarque sur cette reconstruction une forme beaucoup plus arrondie de l'Hexon avec des sillons plus creusés entre les monomères, par rapport à l'Hexon de l'Adénovirus de type 5 (cf. Article 1) ainsi que la présence de trous au niveau du centre de certains Hexons (au niveau de l'axe 2 de l'icosaèdre, Figure 41). Ces trous sont des artefacts apparus au cours du processus de reconstruction 3D. Un filtrage à 30 Å de résolution de la structure cristallographique d'un Hexon ne montre pas de trou de densité au centre (encart en Figure 40a). Le désassemblage partiel des Pentons ainsi que la sortie de l'ADN viral pourrait avoir un effet global sur la structure de la capsid, même au niveau du squelette formé par les Hexons. Au cours de la reconstruction 3D, on ne réalise pas de tri entre les particules intactes et celles en cours de désassemblage. La reconstruction 3D est calculée à partir d'une moyenne dans l'espace de Fourier de l'ensemble des particules sélectionnées ce qui introduit une hétérogénéité importante. Nous avons essayé sans succès de séparer les deux populations de virus (avec et sans Pentons). Il n'y a probablement pas uniquement deux populations mais un ensemble de capsides contenant de 1 à 12 Pentons par virus. Ceci est illustré par le fait que si l'on choisit un niveau de contour moins restrictif pour l'isosurface du modèle 3D, on commence à voir apparaître

de la densité résiduelle au niveau des sommets et les trous au centre des Hexons commencent à se combler, information provenant certainement des capsides intactes. De plus, l'histogramme de la répartition des orientations des particules sur l'UA montrent une répartition correcte mais certaines orientations en particulier au niveau du milieu de l'UA ( $\theta = 80^\circ$ ,  $\phi = 12^\circ$  et  $\theta = 85^\circ$ ,  $\phi = -12^\circ$ ) sont absentes (cf. Annexe 2). L'absence de certaines orientations pourrait être une des causes expliquant la qualité moyenne de la reconstruction.

La reconstruction 3D du CELO ne semble pas contenir de densités pouvant correspondre aux protéines IX ou IIIa des *Mastadenovirus* à l'extérieur de la capsid. Le désassemblage partiel des capsides dans notre échantillon pourrait être à l'origine de l'absence de ces protéines mais il est difficile de déterminer les causes de l'absence des protéines IX et IIIa dans la capsid du virus CELO en absence de reconstruction 3D complète d'un échantillon intact.



**Figure 41: Comparaison du modèle 3D de la capsid du CELO avec celui de la capsid du HAdV-5.**  
 (a) Vue selon l'axe 3 de l'icosaèdre d'une face de la reconstruction 3D du CELO (réalisé avec CHIMERA, Pettersen et al., 2004). (b) Vue selon l'axe 3 de l'icosaèdre d'une face de la reconstruction 3D du HAdV-5 (cf. Article 1).

« Fitting » en corps rigide. Malgré la qualité moyenne de la reconstruction que nous avons obtenue, nous avons tout de même tenté de reconstituer un modèle quasi atomique de la capsid du CELO. L'absence de Penton sur la reconstruction 3D ne m'a pas permis de recaler les structures cristallographiques de la Base du Penton mais j'ai tout de même pu réaliser le recalage de la structure cristallographique de l'Hexon du HAdV-5 à l'aide du programme URO pour les 4 hexons de l'UA (même position que pour HAdV-5). Le dernier cycle a donné des critères de qualité assez faibles qui dénotent d'un positionnement assez médiocre des 4 Hexons. Ceci a pour conséquence la reconstitution d'un modèle quasi atomique avec des erreurs de positionnement et des clashs entre atomes au niveau de certaines UA. Nous n'avons donc pas pu reconstituer, à partir de l'Hexon du

HAdV-5, un modèle quasi atomique satisfaisant. La structure cristallographique de l'Hexon du virus CELO est disponible depuis le printemps 2007 (Xu et al., 2007). J'ai donc pu comparer cette structure à celle de l'Hexon du HAdV-5. Si on regarde séparément les structures de l'Hexon du CELO et du HAdV-5, ceux-ci sont très similaires. On ne remarque pas de différences majeures au niveau de l'organisation générale de la structure (Figure 42 a et b). La structure cristallographique de l'Hexon du virus CELO contient moins de structures secondaires. Ceci est dû à la résolution de cette structure (4 Å). Par contre, en réalisant une superposition à l'aide du logiciel CCP4MG (Potterton et al., 2004), on remarque une différence majeure entre ces deux structures (Flèche sur la Figure 42c) au niveau d'une boucle en surface dans l'Hexon du HAdV-5 (AA 198 à 218) (RMSD sur 855 C<sub>α</sub> = 1,09 Å).

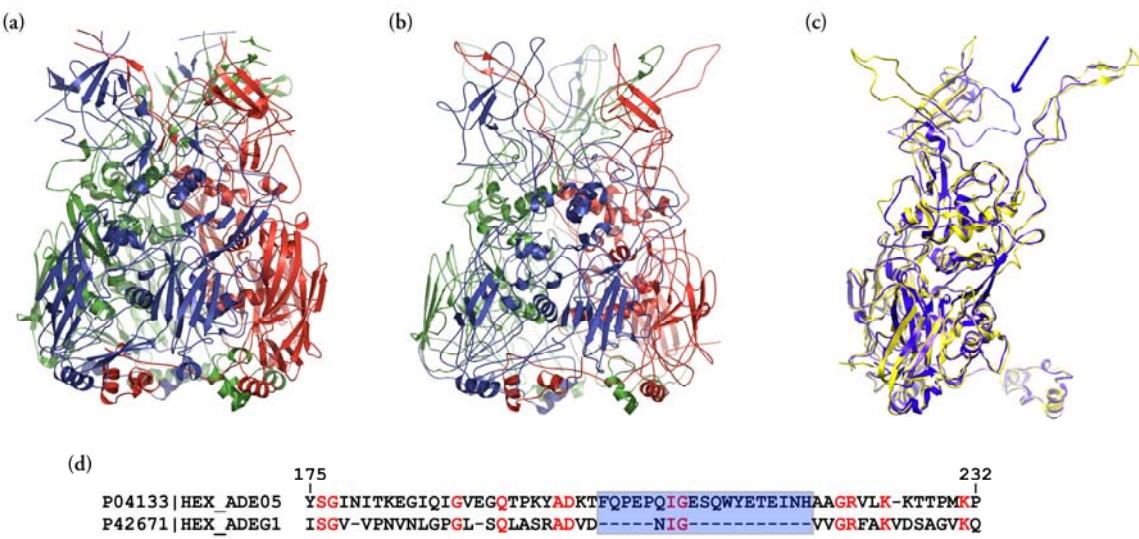


Figure 42: Comparaison des structures cristallographiques de l'Hexon du HAdV-5 et du virus CELO.

(a) Structure cristallographique de l'Hexon du HAdV-5 représentée en « Ribbon », PDB : 1P30. (b) Structure cristallographique de l'Hexon du CELO représentée en « Ribbon », PDB : 2INY. (d) Superposition d'un monomère d'Hexon du CELO en jaune et du HAdV-5 en bleu. (d) Alignement des séquences de l'Hexon du HAdV-5 (PO4133) et du CELO (P42671).

Cette zone, responsable de la liaison avec le monomère adjacent au sein du trimère d'Hexon est absente chez le virus CELO (flèche sur la [Figure 42c](#) et zone encadrée en bleu sur la [Figure 42d](#)). Ceci explique pourquoi les sillons entre les monomères sont beaucoup plus creusés dans ce virus et l'apparition de trous dans la reconstruction 3D. La présence de cette boucle peut expliquer dans un premier temps le mauvais positionnement des Hexons lors de la procédure de « fitting ». Les boucles en surface influencent beaucoup la position des Hexons à la fin de la procédure de recalage. En effet, cette boucle ne correspondant à aucune densité dans la reconstruction 3D impose un positionnement radialement plus bas des Hexons. A la suite de la publication de l'Hexon du CELO,

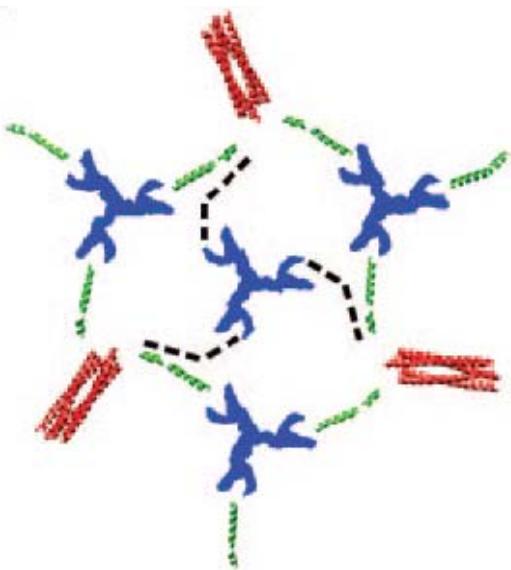
j'ai donc essayé de reconstituer un modèle quasi atomique à partir de cette structure. Pour cela, j'ai procédé de la même manière qu'avec l'Hexon du HAdV-5 mais le positionnement était toujours moyen (CC = 0,58, R<sub>f</sub> = 0,579, Q = 0,347). En effet, Il semblerait que la présence de trous artéfactuels de densité au centre de certains Hexons dans la reconstruction 3D empêche un recalage correct des structures cristallographiques. C'est donc avant tout la qualité de la reconstruction 3D que nous avons obtenu qui semble être à l'origine des problèmes rencontrés au cours du processus de recalage des structures atomiques de l'Hexon et notre tentative de reconstitution d'un modèle quasi atomique de qualité.

### III.3 Conclusion

Nous avons purifié et imaginé un échantillon de virus CELO en CryoME. Même si la purification s'est avérée plus complexe que pour un Adénovirus humain, nous avons tout de même obtenu à l'issue de l'analyse des images une reconstruction 3D de la capsid du virus CELO à environ 30 Å de résolution. Cette reconstruction ne possède pas de Pentons au niveau de sommets de la capsid à cause d'un désassemblage précoce de la capsid au cours de la purification mais le modèle 3D obtenu reste la première visualisation à moyenne résolution de la capsid d'un Adénovirus aviaire. Nous n'avons pas pu conclure sur l'interaction des deux fibres au niveau de la Base du Penton sur la localisation des protéines structurales mineures IX et IIIa. Notre tentative de calculer un modèle quasi atomique de la capsid n'a pas donné un résultat aussi satisfaisant que nous le pensions à cause de la qualité de la reconstruction 3D ainsi que de l'échantillon de départ. N'ayant pas la possibilité de produire le virus au laboratoire pour des raisons de sécurité, je n'ai pas pu renouveler la purification du virus mais mon expérience acquise sur la manipulation du CELO ainsi que sur les méthodes de concentration et de désalage du virus (cf. V) permettront de mener à bien la purification du virus. Ce projet m'a permis d'acquérir de l'expérience dans la manipulation de virus fragile. Avec la publication de la structure atomique de l'Hexon du CELO, une reconstruction 3D de qualité à moyenne résolution (> 20 Å de résolution) permettrait de reconstituer un modèle quasi atomique et d'analyser la localisation des protéines mineures structurales chez le virus CELO.

#### IV. Etude Structurale de la Capside de l'Adénovirus Canin de type 2

Suite à la publication de l'article de Saban et al., 2006 qui remet en question la localisation établie des protéines mineures comme la protéine IX ou la protéine IIIa chez les Adénovirus humains, nous avons voulu apporter des précisions concernant la localisation des extrémités N et C terminales de la protéine IX. En effet, la reconstruction obtenue par Saban et al. A 6 Å de résolution a permis une analyse très précise de la capsid du HAdV-5 qui a révélé une organisation différente de protéines mineures de surface de ce qui avait été précédemment établi. En effet, les auteurs ont montré que les densités situées en surface de la capsid des HAdV-5 sont uniquement dues à la présence de la protéine IX comme le montre la [Figure 43](#). Le C terminal de la protéine IX formerait, selon leur modèle, les densités précédemment attribuées à la protéine IIIa.



*Figure 43: Modèle publié par Saban et al. en 2006 sur la structure de la protéine IX au sein de la capsid du HAdV-5.*  
En bleu sont représentées les densités en forme de triskelion précédemment attribuées au N Terminal de la protéine IX. Les extrémités C Terminales de la protéine IX sont représentées en rouge alors que la zone charnière entre les deux extrémités de la protéine IX est représentée en vert.

Parmi la centaine de sérotypes identifiés, les Adénovirus humains ne représentent que la moitié d'entre eux. Or ce sont les seuls sérotypes à avoir fait l'objet d'études structurales avancées. C'est pourquoi nous nous sommes intéressés aux sérotypes non humains comme l'Adénovirus canin de type2 ou CAdV-2. Ce virus fait partie du genre *Mastadenovirus* et ne comporte donc pas, au niveau de sa capsid, de différences majeures d'organisation par rapport à un adénovirus humain. Par contre, si l'on s'intéresse plus en détails à la séquence des protéines qui le composent et en particulier à celle de la protéine IX, on remarquera que la protéine IX du CAdV-2 est plus courte d'environ 25% que celle du HAdV-5 ([Figure 44](#)). La zone charnière riche en alanine, présente au milieu de la séquence de la pIX du HAdV-5 (AA 60 à 90 en jaune, [Figure 44](#)) est beaucoup plus courte et n'est

pas composée des mêmes acides aminés. Ceci entraîne probablement des différences structurales assez importantes chez CAdV-2. Nous nous sommes donc intéressé à la structure de sa capsid par CryoME afin de localiser les différentes parties de la protéine IX ainsi que de comprendre les conséquences des différences observées au niveau de sa structure primaire.

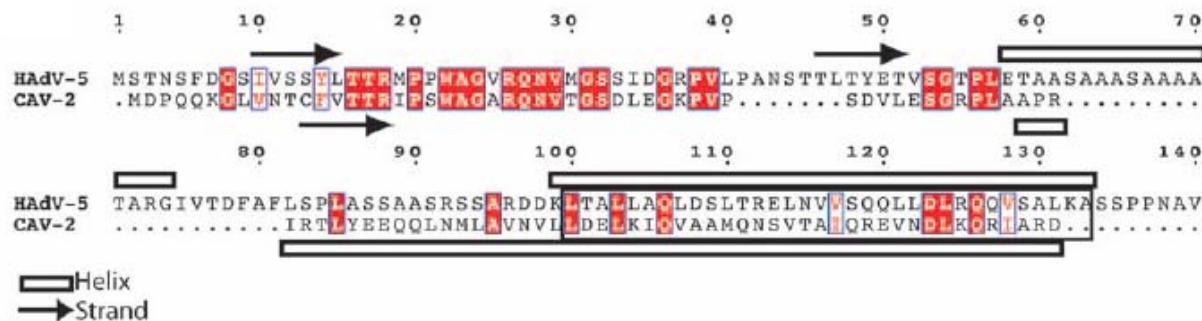


Figure 44: Alignement de séquences de la protéine IX du CAdV-2 et du HAdV-5.

L'alignement est réalisé avec le programme CLUSTAL

#### IV.1 Article 2

##### “Three-dimensional structure of canine adenovirus serotype 2 capsid”

Auteurs: G. Schoehn, M. El Bakkouri, C.M.S. Fabry, L.F. Estrozi, L. Le, D.T. Curiel, O. Billet, A.V. Kajava, R.W.H. Ruigrok & E.J. Kremer.

Publié en 2008 dans *Journal of Virology*, Vol. 82, pages 3329-3341.

## Three-Dimensional Structure of Canine Adenovirus Serotype 2 Capsid<sup>▽</sup>

Guy Schoehn,<sup>1,2\*</sup> Majida El Bakkouri,<sup>1</sup> Céline M. S. Fabry,<sup>1</sup> Oliver Billet,<sup>3,4</sup> Leandro F. Estrozi,<sup>1</sup> Long Le,<sup>5</sup> David T. Curiel,<sup>5</sup> Andrey V. Kajava,<sup>6</sup> Rob W. H. Ruigrok,<sup>1</sup> and Eric J. Kremer<sup>3,4\*</sup>

*Unit for Virus Host Cell Interaction, UMR 5233, Université Joseph Fourier, EMBL, CNRS, Grenoble, France<sup>1</sup>; Institut de Biologie Structurale Jean-Pierre Ebel, UMR5075 CEA-CNRS-UJF, Grenoble, France<sup>2</sup>; Institut de Génétique Moléculaire de Montpellier, CNRS 5535, Montpellier, France<sup>3</sup>; Universités Montpellier I and II, Montpellier, France<sup>4</sup>; Division of Human Gene Therapy, Departments of Medicine, Obstetrics and Gynecology, Pathology, and Surgery, and the Gene Therapy Center, The University of Alabama at Birmingham, Birmingham, Alabama 35294<sup>5</sup>; and Centre de Recherche de Biochimie Macromoléculaire, UMR 5237 CNRS-Universités Montpellier I and II, Montpellier, France<sup>6</sup>*

Received 6 November 2007/Accepted 28 December 2007

**T**here are more than 100 known adenovirus (AdV) serotypes, including 50 human serotypes. Because AdV-induced disease is relatively species specific, vectors derived from nonhuman serotypes may have wider clinical potential based, in part, on the lack of ubiquitous memory immunity. Whereas a few of the human serotype capsids have been studied at the structural level, none of the nonhuman serotypes has been analyzed. The basis laid by the analysis of human AdV (hAdV) has allowed us to determine and compare the three-dimensional structure of the capsid of canine serotype 2 (CAV-2) to that of hAdV serotype 5 (hAdV-5). We show that CAV-2 capsid has a smoother structure than the human serotypes. Many of the external loops found in the hAdV-5 penton base and the hexon, against which the antibody response is directed, are shorter or absent in CAV-2. On the other hand, the CAV-2 fiber appears to be more complex, with two bends in the shaft. An interesting difference between the human and canine viruses is that the C-terminal part of protein IX is in a different position, making an antenna sticking out of the CAV-2 capsid. The comparison between the two viruses allows the identification of sites that should be easy to modify on the CAV-2 capsid for altering tissue tropism or other biological activities.

*Adenoviridae* are nonenveloped icosahedral viruses with 26- to 44-kb double-stranded DNA genomes (26). They infect a variety of vertebrates including mammals, fish, birds, and reptiles (12). The more than 50 human adenovirus (hAdV) serotypes are the most extensively studied because of their ability to induce respiratory, ocular, and enteric infections in immunocompetent individuals. hAdV infections are a significant cause of morbidity and mortality in newborns and immunocompromised individuals (28). AdVs are also used for gene therapy trials and anticancer treatments (6) although the widespread immunity against human serotypes severely compromises their use as vectors. Because AdV-induced morbidity is relatively species specific, vectors derived from nonhuman AdVs may be more clinically useful than those derived from human serotypes, based, in part, on the potential lack of ubiquitous memory immunity (cellular and humoral) (5, 11, 18, 30, 35, 36).

One of the best characterized nonhuman AdVs is the canine serotype 2 (cAdV-2, commonly referred to as CAV-2) (29). We previously showed that CAV-2 vectors could lead to preferential transduction of neurons *in vivo* (49), long-term (>1 yr) expression in the brain (50), and a lower cellular immune response (27, 49). The majority of donors in a random cohort harbored no or only low levels of anti-CAV-2 neutralizing antibodies (30, 37), and less

than half of this cohort harbored anti-CAV-2 memory T cells (38). When the donors harbored an anti-CAV-2 cellular response, it was on the average 10-fold lower than the response to hAdV serotype 5 (hAdV-5). Finally, in contrast to hAdV-5-based vectors, CAV-2 vectors poorly transduce or mature human dendritic cells or activate the complement cascade (37, 40), both of which play pivotal roles in orchestrating and bridging innate, adaptive, and memory immune responses. These findings suggest that CAV-2 vectors may have specific clinical advantages, especially in terms of safety (39).

However, there exist certain enigmas concerning various aspects of CAV-2 biology. For example, the *in vitro* paradigm for AdV tropism is predominately based on studies using hAdV-2 or hAdV-5 in epithelial cells. Initially, the fiber knob binds to the coxsackievirus AdV receptor (CAR) (4, 52). Then, a conserved Arg-Gly-Asp (RGD) motif in the penton base binds to integrins, triggering endocytosis (24). Although CAV-2 uses CAR to infect cells (48), the capsid does not contain an identifiable integrin-interacting motif, yet it traffics in epithelia cells with similar kinetics (8). Furthermore, although both hAdV-5 and CAV-2 use CAR, CAV-2 preferentially transduces neurons when injected into the brain while hAdV-5 preferentially transduces astrocytes (1, 32). CAV-2 is also capable of an impressive level of axonal transport in neurons that can be up to 100-fold greater than that of hAdV-5 (49). In addition, in contrast to previous studies with hAdV-5 (16, 17), we have been unable to complex CAV-2 virions with divalent cations or lipoplexes (unpublished data).

To resolve these and other CAV-2 enigmas, we need to better understand the capsid structure. Based on the data from different hAdVs, the AdV capsid structure appears to be rel-

\* Corresponding author. Mailing address for Guy Schoehn: Unit for Virus Host Cell Interaction, 6 rue Jules Horowitz, BP181, 38042 Grenoble, France. Phone: 33 4 7620 9423. Fax: 33 4 7620 7199. E-mail: schoehn@embl.fr. Mailing address for Eric J. Kremer: Institut de Génétique Moléculaire de Montpellier, CNRS 5535, 1919 Route de Mende, 34293 Montpellier, France. Phone: 33 4 6761 3672. Fax: 33 4 6704 0231. E-mail: eric.kremer@igmm.cnrs.fr.

▽ Published ahead of print on 23 January 2008.

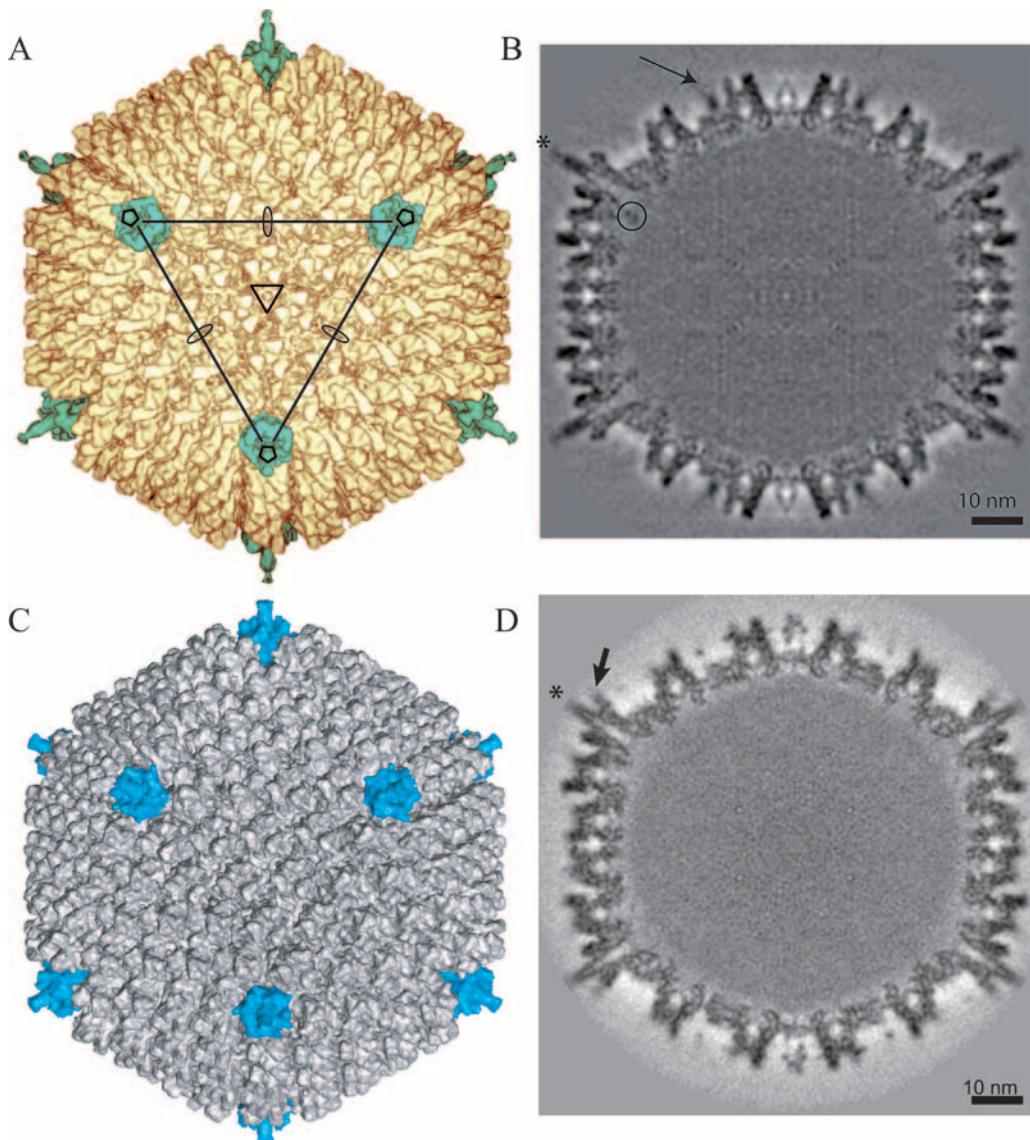


FIG. 1. 3D reconstruction of CAV-2 at 12-Å resolution and comparison with the previously determined structure of hAdV-5 (MSD accession number 1111) at 10-Å resolution. (A) Isosurface representation of CAV-2 as viewed down the threefold axis. The pentons are shown in petrol blue, and the rest of the capsid is shown in yellow. The icosahedral axes are indicated by an ellipse (twofold), a triangle (threefold), and a pentagon (fivefold). A continuous line connecting the three pentons pointing in your direction delimits a facet. (B) A central slice through the CAV-2 EM reconstruction showing that the capsid density is well defined. One of the fibers sticking out radially from the capsid is indicated by a star. A circle surrounds the nonattributed extra density present under the penton base. The arrow indicates protein IX density. Scale bar, 10 nm. (C) Isosurface representation of hAdV-5 view down the threefold axis. The pentons are shown in blue, and the rest of the capsid is shown in gray. (D) A central slice through the hAdV-5 EM reconstruction showing that the capsid density is well defined. The fibers sticking out radially from the capsid are indicated by a star. The RGD loop is indicated by an arrow. Scale bar, 10 nm.

atively well conserved. The capsid contains 7 out of the 12 polypeptides present in the virion. The major structural components are the trimeric hexons, 240 of which form the 20 facets of the icosahedron (51). At each of the 12 vertices are a pentameric penton base and an externally projecting trimeric fiber. The atomic structures of these three individual components are known for hAdV-2 or hAdV-5 (44, 54, 58).

The hexon trimer has a pseudohexagonal base with three towers extending upwards. The triangular top is rotated  $\sim 10^\circ$  counterclockwise with respect to the hexagonal base. The loops extending from the irregularly shaped towers (44) have not

been solved by X-ray crystallography, although they have been visualized and modeled in recent cryo-electron microscopy (cryo-EM) structures (15, 45).

In contrast to the hexon, the penton base is wide at the top (58). It has an overall tulip shape, with each of its five petals having a right-handed twist around the fivefold axis. The top of the base is almost flat with the exception of two protuberances. One of these extensions, which are variable in size depending on the serotype, contains the above-mentioned integrin-interacting motif. The other extension is a hypervariable loop that also varies in size depending on the serotype (20).

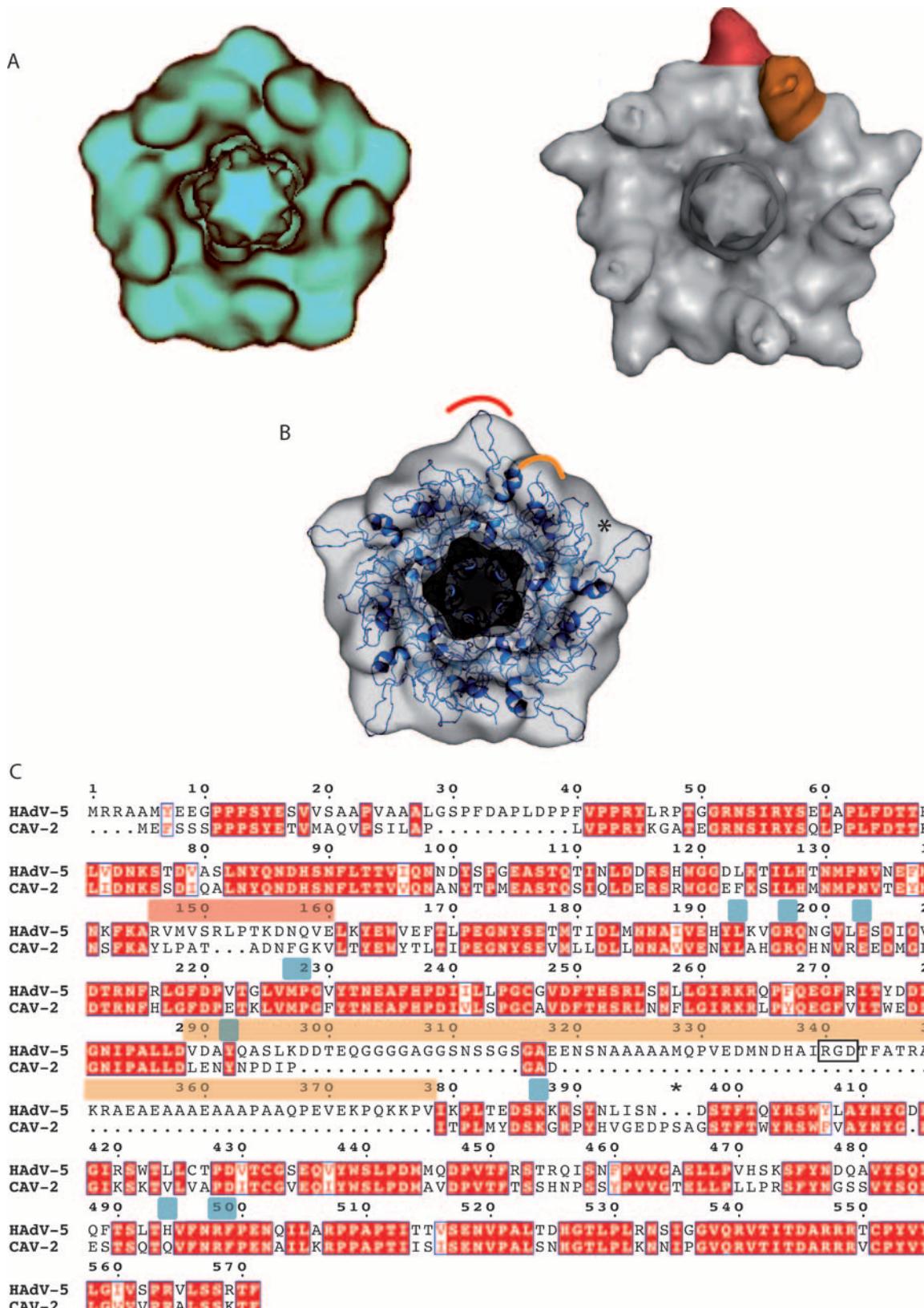


FIG. 2. The CAV-2 penton base. (A) Comparison between the computationally isolated penton from the CAV-2 capsid in blue (left) and the hAdV-5 capsid in gray (right). One of the five hypervariable regions is highlighted in red, and one protuberance harboring the RGD sequence is shown in orange in the hAdV-5 penton. (B) Fitting of the hAdV-2 penton base X-ray structure in blue (PDB accession number 1X9T) into the 12-Å CAV-2 cryo-EM envelope. The red and orange parts correspond to the regions indicated in panel A. The star indicates a region where the

The trimeric AdV fiber contains a highly conserved N-terminal sequence that binds to the base, a thin shaft of variable length, and a globular knob at the distal end. The length, the flexibility of the fiber shaft, and the sequence of the knob domain also vary between serotypes (9). Four of the 23 repeats of the hAdV-2 fiber shaft were crystallized, and the resolved structure demonstrated a novel triple  $\beta$ -spiral (54). Notably though, the entire fiber was not crystallized, probably due to its flexible structure. The number of repeats can vary from a minimum of five for hAdV-3 to at least 46 for bovine AdV-3 (BAV-3) (43).

Apart from these major proteins, the capsid is stabilized by hexon-associated proteins (51). Using high-resolution cryo-EM three-dimensional (3D) reconstructions combined with secondary structure predictions, a new location of the minor structural proteins has been proposed. Protein IX forms a connected complex of four trimers in the middle of each facet (45). Protein IIIa is located inside the capsid underneath the penton base, connected to its N-terminal part (45). A total of 120 copies of protein VIII are located on the inner face of the capsid, and each makes the link between four hexon trimers around the three- and fivefold axes (15).

Here, we have determined the 3D structure of the CAV-2 capsid and compared it to the known X-ray and EM structures of hAdV-5 or hAdV-2 by fitting the different atomic resolution structures into the EM density. Globally, we found that CAV-2 capsid has a smoother structure than the human serotypes. Many of the external loops found in hAdV-2 and hAdV-5 (hAdV-2/5) are either shorter or absent in CAV-2. On the other hand, the CAV-2 fiber appears to be more complex, with two bends in the shaft. The knowledge of the structure will allow the identification of sites where functional loops may be added to CAV-2 to potentially modify tissue tropism.

#### MATERIALS AND METHODS

**CAV-2 vectors and recombinant fiber.** CAVGFP and CAVpIX-GFP were prepared as previously described (30, 31). Briefly, CAVGFP is a CAV-2 vector with a deletion of the E1 region and with a green fluorescent protein (GFP) expression cassette. The capsid contains no modifications. CAVpIX-GFP is a vector containing GFP fused to the C-terminal end of protein IX. All vectors were purified by double banding on CsCl gradients, and CsCl was removed using PD-10 columns (Pharmacia). The vectors were stored in phosphate-buffered saline containing 10% glycerol. Stocks were  $2 \times 10^{12}$  physical particles/ml with at least 1 infectious particle/5 physical particles. The approximately 1.7-kb CAV-2 fiber coding region was cloned in pBlueBac4.5/V5-His-TOPO (Invitrogen). Recombinant baculovirus was generated, isolated, amplified, and purified following the manufacturer's protocol. The His-tagged fiber was purified using standard protocols.

**Negative staining.** Four microliters of the recombinant CAV-2 fiber sample ( $\sim 0.1$  mg/ml) was loaded in the mica-carbon interface as described previously (46). The sample was stained using 2% sodium silico tungstate, pH 7.5, and air dried. Images were taken under low-dose conditions in an EX1200-II JEOL electron microscope working at 100 kV and with a nominal magnification of 40,000. The images were scanned on a Z/I Imaging scanner (Photoscan TD) with a pixel size of 14  $\mu$ m (3.5  $\text{\AA}$  per pixel at the sample level). A total of 800 fibers

were selected using X3d (10), cut into 200- by 200-pixel squares, transferred into the SPIDER program (19), and subjected to classification and averaging.

**Cryo-EM.** For cryo-EM, the glycerol was removed, and the sample was concentrated to  $\sim 1$  mg/ml. Four microliters of sample was loaded onto a Quantifoil R2/1 holey grid (Quantifoil Micro Tools GmbH, Germany), blotted for 1 to 2 s to remove the excess liquid, and then rapidly plunged into liquid ethane cooled using liquid nitrogen. The frozen grid was transferred into an LaB6 CM200 Philips electron microscope using a GATAN 626 cryoholder. The images were taken under low-dose conditions at 200 kV (less than  $10 \text{ e}^-/\text{\AA}^2$ ) and with a nominal magnification of 27,500. Each area was imaged twice using two different defocus values, starting with the image closest to focus. The negatives were developed in full-strength D19 developer for 12 min.

**Image analysis.** Negatives were screened for astigmatism and drift by optical diffraction, and only those showing information up to 12  $\text{\AA}$  were digitized using the Photoscan TD with a pixel size of 7  $\mu$ m (2.54  $\text{\AA}$  per pixel at the specimen level as calibrated using tobacco mosaic virus). The particles were selected interactively using X3d (10) and boxed into 447- by 447-pixel squares with a circular mask applied of radius 219 pixels plus four fade pixels. The images were corrected for the contrast transfer function effect, as described previously, using the program CTFFIND (10). A total of 5,350 particles from 21 negatives were used in the image analysis process. Determination of particle origin and orientation were performed with the model-based polar Fourier transform (PFT) programs using the previously determined 3D structure of hAdV-5 (3, 15). After PFT refinement was completed, a supplementary cycle was carried out using a new version of PFT, PFT2 (D. M. Belnap, J. B. Heymann and J. F. Conway, unpublished results). The final reconstruction was obtained using the PFT2 and EM3DR2 programs with a total of 2,780 particles (about half of the total). The resolution of the final map was estimated to 12  $\text{\AA}$  by Fourier shell correlation (53) calculated between independent half-data set maps and by applying a correlation limit of 0.3 (not shown). The same process was essentially used for CAVpIX-GFP, but we used Quantifoil grids covered by a thin layer of continuous carbon to increase the visible virus concentration. Only 800 particles were used in this reconstruction, which limited the resolution to 25  $\text{\AA}$ . Map visualization was performed using the programs WEB (19) and PYMOL (14), and the contour level was chosen to include an average protein density of 0.84 Da/ $\text{\AA}^3$ . The EM maps were uploaded into the EM database associated with the Macromolecular Structure Database (see below).

**Fitting of the atomic structures of hexon and penton base into the EM density.** The atomic structure of the hAdV-5 hexon and that of the hAdV-2 penton base were placed manually into the corresponding CAV-2 EM densities using the program PYMOL (14). These manual fits were then optimized using SITUS (55).

**Molecular modeling of the 3D structure of CAV-2 fiber.** To model the 3D structure of the CAV-2 fiber, we used the crystal structure of the homotrimeric C-terminal CAV-2 fiber head (residues 361 to 542) (47) and two known crystal structures of homologous proteins: reovirus attachment protein  $\sigma 1$  (7) and the C-terminal part of the hAdV-2 fiber containing four repeats of the shaft domain followed by the head domain (54). The hAdV-2 structure allowed us to model the last four repetitive elements of the CAV-2 fiber shaft and the region where this shaft is linked to the head. The N-terminal repetitive structure of the CAV-2 fiber shaft was also modeled by using the  $\beta$ -spiral fragment of the known hAdV-2 structure as a template and in accordance with the repeat alignment shown on Fig. 3. The two longer repeats (see Fig. 3C, repeats 4 and 10) of hAdV-2 having 7- and 6-residue insertions in the middle of the repeats were modeled by using the crystal structure of reovirus attachment protein  $\sigma 1$  (7). The  $\sigma 1$  protein has one such long repeat containing a similar 5-residue insertion. The model was built by using the Insight II program (13).

**Comparison with hAdV-5.** All comparisons between the CAV-2 and the hAdV-5 EM structures were done using the hAdV-5 3D reconstruction previously determined in the laboratory and freely available in the EM database associated with the Macromolecular Structure Database (<http://www.ebi.ac.uk/msd-srv/emsearch/index.html>) under accession number 1111.

X-ray structure did not fill the EM density. (C) Sequence alignment of the hAdV-5 and CAV-2 penton bases visualized with ESPript (23). The RGD and hypervariable loops are indicated by bars above the sequence in the same colors as used in panels A and B. The RGD sequence is boxed with a black rectangle. Conserved residues are in white text and highlighted in red; partially conserved residues are in red text. The star highlights three extra amino acids present in CAV-2 compared to hAdV-5 that could fill up the density indicated by the star in panel B. The amino acids involved in fiber binding for hAdV-5 are indicated by blue bars above the sequence.

**Sequence alignments and structure prediction.** All sequence alignments were carried out using CLUSTAL (http://www.ebi.ac.uk/clustalw/). Secondary structure predictions were made using PREDATOR (http://bioweb.pasteur.fr/seqanal/interfaces/predator-simple.html) and coil-coiled predictions were made using http://www.russell.embl-heidelberg.de/cgi-bin/coils-svr.pl.

**Protein structure accession numbers.** EM density maps have been deposited in the Macromolecular Structure Database at the European Bioinformatics Institute under accession numbers 1462 and 1463 for CAV-2 and CAVpIX-GFP, respectively.

## RESULTS AND DISCUSSION

**EM model of CAV-2.** We determined the 3D structure of CAV-2 capsid by cryo-EM at 12-Å resolution. The dimension and the thickness of the CAV-2 capsid shell are similar to those of hAdV-5 (Fig. 1). The icosahedral shape is well defined and exhibits the pseudo T=25 triangulation number imposed by the trimeric hexon protein (Fig. 1A and B). There are 12 triangular hexons per facet and one penton on each vertex. The hexons are packed in an orientation similar to that of hAdV-5, including the slight tilt at each of the 20 edges of the virion (30° versus 10° for the other hexons belonging to the same facets). The penton base also has roughly the same shape in both capsids. The CAV-2 fiber is flexible, which explains why only part of it was visible in the 3D reconstruction (Fig. 1B, star). Finally, densities are present in the middle of the facet between the hexons. This density, attributed to protein IX, exhibits a triskelion-like shape, which is similar to that of hAdV-5 (Fig. 1; see also Fig. 5).

**The CAV-2 penton base is smooth.** Sequence alignment between the hAdV-5 and CAV-2 penton bases shows that 310 of the 571 amino acids are identical and that 14 amino acids are similar between the two species. However, the CAV-2 penton base sequence is ~20% shorter than the hAdV-5 sequence (477 versus 571) (Fig. 2C). Nonetheless, the overall shape of the two pentagons is similar (Fig. 2A). A slice through the density (Fig. 1B) shows a cone-like geometry that fits perfectly into the vertices surrounded by the hexons.

The CAV-2 penton base isosurface representation is smoother than that of hAdV-2/5 (Fig. 2A). By comparing the EM structures (Fig. 2A), we identified two protuberances that are smaller in the CAV-2 base (Fig. 2A, right panel, red and orange areas). The atomic structure of the hAdV-2 penton base (99% identical to that of hAdV-5) was fitted into the CAV-2 penton base EM density in order to identify the residues that are responsible for the smooth appearance of the CAV-2 capsid (Fig. 2B). The first difference is located in the hypervariable loop (Fig. 2A, red area, and C, sequence indicated by a red bar), which is 3 amino acids shorter in the CAV-2 base. On average, the amino acids composing this loop are also smaller in size (PATADNFG) (Fig. 2C). The second difference (Fig. 2, orange area) lies on the protuberances located on the top of the base that contains the RGD sequence in most hAdV (residues 289 to 379 for hAdV-5) (Fig. 2C). Due to its flexibility, the entire RGD loop structure is not visible either in the atomic structure of the hAdV-2 penton base or in the hAdV-5 cryo-EM structure. This protuberance is also present in the CAV-2 base but is much less pronounced because it is only 12 amino acids long. The CAV-2 loop does not contain an RGD sequence. Despite the small size of the protuberance, it is clearly visible in the cryo-EM map. This rein-

forces the idea that the N-terminal part of the fiber interacts with the base of the RGD loop (20) and contributes to the size of the protuberance.

Next to the orange loop in the CAV-2 penton base there are three extra amino acids after the RGD loop, indicated by a star in the fit shown in Fig. 2B and in the sequence in panel C. This corresponds to an empty region in terms of both the fit and sequence (Fig. 2B and C, stars). This could easily explain the smoother pentagonal shape of the CAV-2 base (slightly larger on the middle of the edge and smaller on the pentagon vertex).

According to our sequence alignment (Fig. 2C), the N-terminal part of the CAV-2 penton base is also 15 amino acids shorter than that of hAdV-5. The currently unresolved N-terminal part of the penton base (amino acids 1 to 49) contains two highly conserved PPXY (where X is any residue) motifs implicated in interactions with host cell WW domain-containing ubiquitin ligases (21, 22). In the mature virion, this region is probably located under the penton base, where it interacts with other viral capsid proteins (20, 45). The PPXY sequences are also present in CAV-2, suggesting that these sequences are critical for some phase in the AdV life cycle (22). The 15-amino-acid difference is too small to be interpreted in the EM map because it is in contact with other proteins on the inside of the virus capsid.

On the CAV-2 penton base, 9 out of the 10 residues involved in fiber tail-penton base interaction are conserved compared to hAdV-5 (Fig. 2C, indicated by blue bars). The only difference is that histidine 494 of hAdV-5 is a glutamine in CAV-2. The fiber sequence that interacts with the base is also conserved: amino acids 13 to 19 follow the conserved sequence seen in hAdV (58). The binding of the fiber to the penton base should therefore be very similar to that described for hAdV-2 (58).

**The CAV-2 fiber contains two bends.** The CAV-2 fiber is 542 amino acids long: the first 41 amino acids make up the tail, the central 320 amino acids comprise the shaft, and the last 182 amino acids form the knob (Fig. 3C). In hAdV-5 each of the ~15-amino-acid repeats is 13 Å long (42) and induces a rotation of ~50° along the fiber axis (54). Based on sequence analysis, Rasmussen et al. (41) predicted that the CAV-2 shaft had 18 repeats with 18 amino acids per repeat. Our sequence analysis suggests that there are 18.5 repeats (Fig. 3C). The number of amino acids/repeat varies: seven repeats have the quintessential 15-amino-acid motif, the two repeats with 16 amino acids have the “extra” amino acid on the outside of the repeat, the three repeats with 17 amino acids have the extra amino acids inside the repeat, one repeat has 18 amino acids (extra internal), and two repeats have 19 amino acids (extra external). Two of the three longest repeats (21 to 22 amino acids) contain six or seven extra internal amino acids. All hAdV fiber sequences contain an irregular, long repeat near the N-terminal attachment site on the penton base, which has been linked to a possible bend at this position (42). However, this irregular repeat tends to be the third repeat and not the fourth, as is the case for the CAV-2 fiber (9). The much longer fibers of BAV-3 and the avian chicken embryo lethal orphan virus have additional bends, also related to irregular repeats (25, 43). Wu et al. (57) showed that such a 6- to 7-amino-acid internal insert can introduce a bend in the hAdV-5 shaft that may allow greater access of the penton base’s integrin-interacting motifs to the cell surface (56).

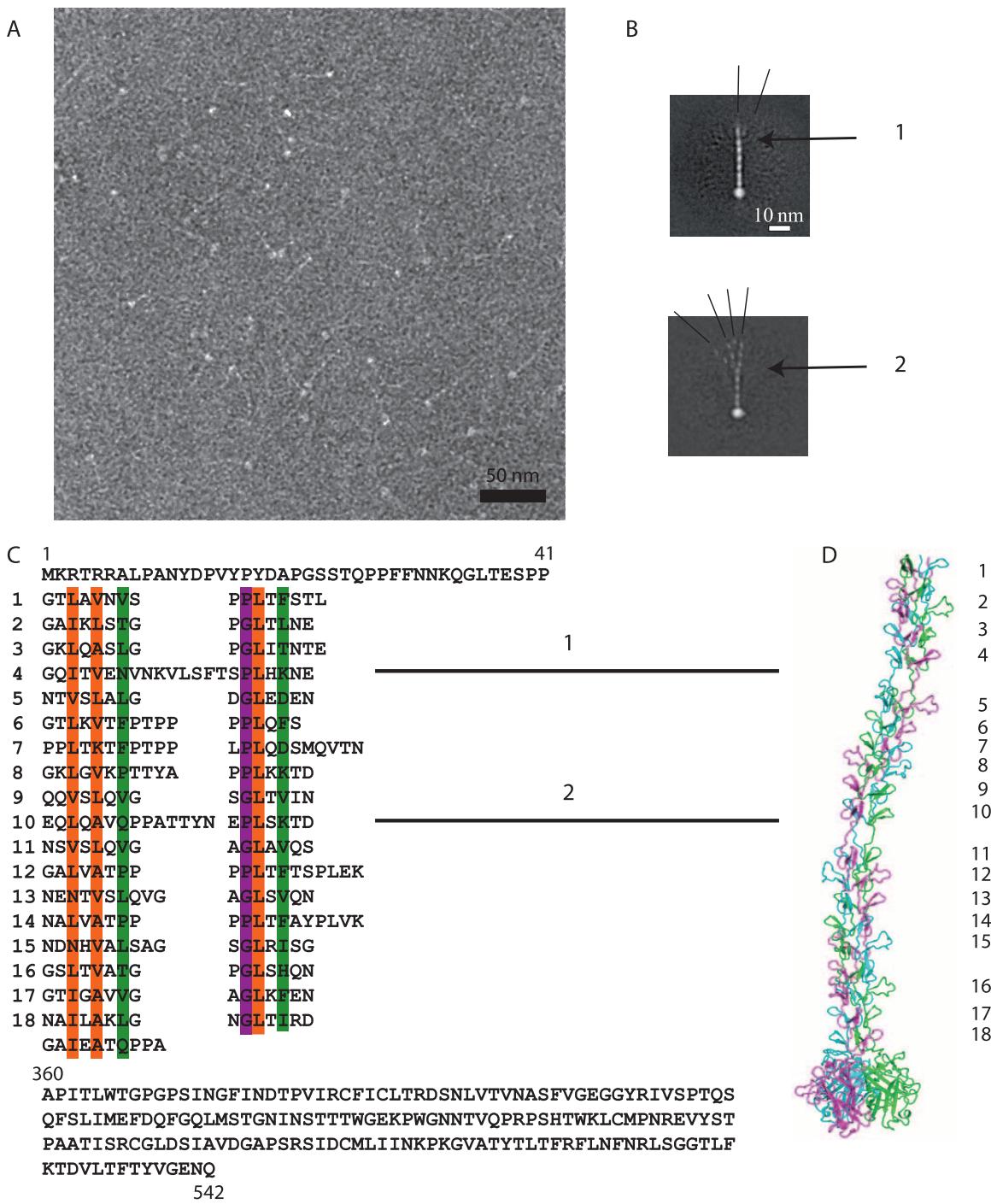
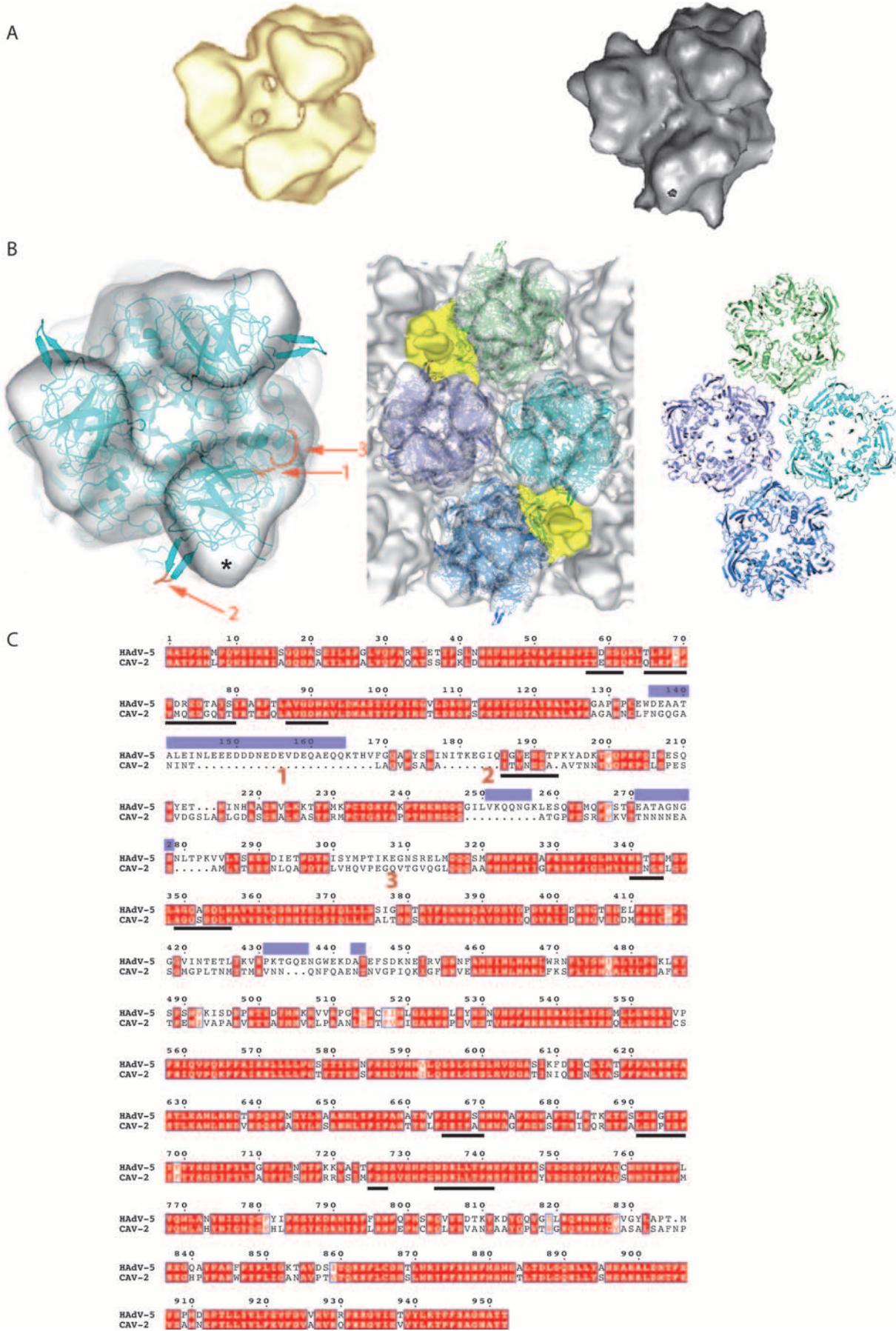


FIG. 3. The CAV-2 fiber. (A) Electron micrograph of negatively stained CAV-2 fibers. The thin shafts and their globular knobs are clearly visible. Scale bar, 50 nm. (B) Averaged image of fibers selected from micrographs like that shown in panel A. The fiber is generally straight but can exhibit two kinks as indicated by the numbers 1 and 2. The different lines show the angle that the N-terminal part makes with the rest of the shaft. (C) Sequence of the CAV-2 fiber shaft showing the repeat motifs that correspond to the location of the bends. The top part of the sequence is the N-terminal tail of the fiber, and the lower part is the C-terminal knob. The repeat residues involved in the hydrophobic core are in orange, the ones forming the peripheral hydrophobic patches are in green, and the conserved glycine or proline residues are highlighted in purple according to van Raaij et al. (54). (D) Hypothetical model of the CAV-2 fiber showing the predicted bends at repeats 4 and 10. The repeats are indicated with the same numbers as in the sequence shown in panel C.

As mentioned previously, the CAV-2 fiber was too flexible to be reconstructed by cryo-EM on the entire virion. To complete our structural analysis, we imaged recombinant CAV-2 fibers by negative staining (Fig. 3A). After image averaging (the fiber

head and the apical part of the different fibers are superposed but their N-terminal parts show the different positions they can adopt), we also found at least two bends in the shaft (Fig. 3B). The location of the hinges correlates well with the 21- and



22-amino acid-long repeats with six to seven internal amino acids (Fig. 3B, repeats 4 and 10).

To test the agreement of the EM data with the (i) suggested repeat alignment, (ii) dimensions of the predicted atomic structure, and (iii) locations and properties of the kink regions, we modeled the 3D structure of the fiber (Fig. 3D). Recently, the crystal structure of the homotrimeric C-terminal CAV-2 fiber head (residues 361 to 542) was determined (47). We also took advantage of two other known crystal structures of homologous proteins corresponding to different parts of the fiber to model its overall structure. The first structure that we used as a template was the C-terminal part of the hAdV-2 fiber containing four repeats of the shaft domain followed by the receptor-binding head domain (54). This structure allowed us to model the last four repetitive elements of the CAV-2 fiber shaft and the region where this shaft is linked to the head domains. Most of the CAV-2 repeats are similar to those in hAdV-2, and, therefore, the crystal structure of the hAdV-2 repeats was used to model the remaining part of the CAV-2 fiber. The exceptions were repeats 4 and 10 (Fig. 3C) having 7- and 6-residue insertions between two conserved  $\beta$ -strand motifs. These repeats were modeled by using the crystal structure of the reovirus attachment protein  $\sigma$ 1 (7). In the structure of  $\sigma$ 1, one repeat contains a similar 5-residue insertion. This insertion interrupts the progression of the regular  $\beta$ -spiral and introduces a bend into the  $\sigma$ 1 trimer. The structure of the insertion is loose, with slightly different main chain conformations in each monomer extended along the fiber axis. This structure was used to model the two longer repeats of CAV-2.

If our model is correct, the CAV-2 trimeric fiber should be about 330 Å long. The  $\beta$ -spiral is 280 Å long with two flexible bends and a C-terminal globular head domain contributing 50 Å (Fig. 3D). This model, with respect to the number and location of the bends and the length, is in a good agreement with our EM data.

**The CAV-2 hexon.** The general dimensions and twist of the homotrimeric CAV-2 hexon are very similar to those of hAdV-5 (Fig. 4A). Globally, the CAV-2 hexon sequence is shorter than that of hAdV-5 (Fig. 4C) (908 versus 952 amino acids) with all the differences located in the N-terminal half. The fit into the EM density of the CAV-2 hexon was carried out using the hAdV-5 X-ray structure (Protein Data Bank [PDB] 1P30) (Fig. 4B). Because the loops that are absent in the hAdV-5 X-ray structure (Fig. 4C, blue bars above the sequence) often correspond to a deletion in the CAV-2 sequence, the fit was good, and only one loop stays outside the density (Fig. 4B, left). This loop (amino acids 194 to 202 in

hAdV-5 versus 153 to 161 in CAV-2) was visible in the atomic structure of the hAdV-5 hexon but not in that of the hAdV-2 hexon. The sequence just before this loop is absent in the CAV-2 hexon (Fig. 4B and C, region 2), which could explain the absence of density in the EM structure. The structural rearrangement caused by this difference could fill the empty part in the CAV-2 EM hexon structure (Fig. 4B, asterisk). Further, there is a 24-amino-acid gap after amino acid 144 (Fig. 4B and C, region 1). This region is located on the top of the hAdV-5 hexon and was not ordered in the X-ray structure. This may explain the smoother shape of the CAV-2 hexon compared to the EM density of the hAdV-5 hexon (Fig. 4A, right). The last structural difference, also located on the top of the hexon, is likely caused by the significant sequence divergence around amino acid 310 (Fig. 4B and C, region 3 in the hAdV-5 sequence). Interestingly, the protruding structures in the hAdV hexons, which contain most of the epitopes recognized by anti-hexon antibodies (44), are absent in CAV-2.

As mentioned above, we have been unable to complex CAV-2 vectors with divalent cations or lipoplexes, suggesting a more neutral charge of the external CAV-2 capsid. Because the capsid charge is dominated by the hexon charges, the data presented here support this prediction. For example, region 1 in Fig. 4C is composed mainly of negatively charged residues in hAdV-5, and, more generally, all the hAdV-5 loops that are missing in the CAV-2 capsid are rich in charged residues.

We also performed a multiple fitting on the capsid using four hexons surrounding the twofold axis of the virus. The regions involved in the interhexon contacts are the same as those previously determined for hAdV-5 (15) (Fig. 4C, bold lines). Furthermore, in most cases, the amino acids are conserved between the two species. The four hexons present around the twofold axis in CAV-2 interact using the same five loops that are used in hAdV-5: loops 65 to 70, 292 to 296, 614 to 619, 674 to 676, and 683 to 690 in CAV-2 corresponding to loops 65 to 70, 340 to 344, 665 to 670, 725 to 727, and 734 to 741 in hAdV-5 (Fig. 4B and C, bold lines). Only the interaction region 2 (Fig. 4B) is absent in CAV-2.

**Minor proteins.** (i) **Protein IX.** A major difference in the 3D structure of hAdV and CAV-2 is located on the top of the triskelion structure attributed to the N-terminal domain of protein IX, where a rod-like density is present in CAV-2. This density is absent in the hAdV-5 reconstruction (Fig. 1B, arrow, and C and 5C). The sequence of protein IX is 25% shorter in CAV-2 than in hAdV-5 (105 versus 140 amino acids) (Fig. 5E). By comparing the 3D reconstructions of CAV-2 and hAdV, we should be able to locate protein IX within the capsid.

FIG. 4. The CAV-2 hexon capsomer. (A) Comparison of computer-isolated hexons from CAV-2 (yellow) and hAdV-5 (gray) imaged in isodensity surface view. (B) Fitting of hAdV-5 hexon X-ray structure (PDB 1P30) into the CAV-2 cryo-EM envelope. Either one hexon alone (left) or the four hexons surrounding a twofold axis (middle) have been fitted using SITUS and imaged at the same time in the EM density. On the right, only the bottom wall of the pseudo-atomic model of the four hexons is shown. In the hexon alone, three regions are indicated on the X-ray structure; they correspond to large differences in the sequence of CAV-2 compared to that of hAdV-5 (see below). The red loops and regions 1 to 3 in the single hexon on the left are discussed in the text and correspond to sequence elements 1 to 3 indicated in panel C. The star indicates a region in the EM density that is not filled by the density derived from the crystal structure, as discussed in the text. (C) Sequence alignment between the CAV-2 and hAdV-5 hexon bases. Conserved residues are in white text and highlighted in red; partially conserved residues are in red text. The loops not resolved in the X-ray structure are indicated by blue bars above the sequences. The amino acid sequences involved in the contacts between hexons as described by Fabry et al. (15) are underlined with a bold line. They are conserved between CAV-2 and hAdV-5. Regions 1, 2, and 3 from panel B are indicated.

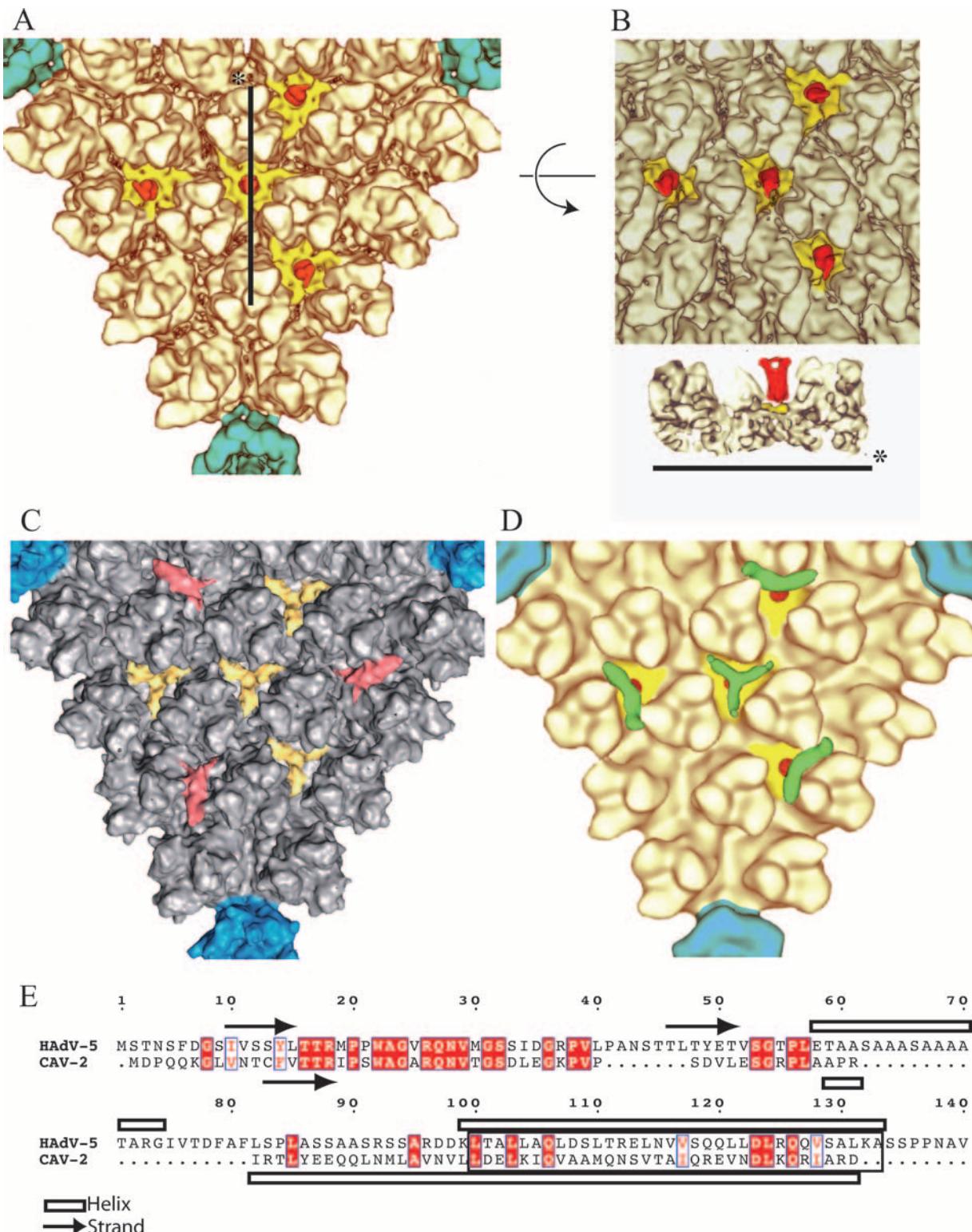


FIG. 5. The minor protein IX. (A) Detailed view of one facet of CAV-2 seen down the threefold axis of the virion. The pentons are located on the vertices of the triangle and are shown in petrol blue, and the hexons are shown in pale yellow as in Fig. 1. Additional densities present as four groups of trimeric triskelions assigned to be protein IX are yellow (N-terminal part) and red (C-terminal part). (B) The top part represents the center of the facet described as in panel A but slightly tilted to see the extension of the C terminal part of protein IX from its N terminal part. The lower part is a slice through the capsid along the line shown in panel A. The color codes are the same as in panel A. (C) Detailed view of one facet of hAdV-5 seen down the threefold axis of the virion. The pentons are located on the vertices of the triangle and are in blue, and the hexons are gray. Protein IX is shown in yellow, and the cylindrical density that we initially assigned to protein IIIa (15) but that was recently reassigned to the C-terminal part of protein IX (45) is in red. (D) Detailed view of the threefold axis of the CAVpIX-GFP reconstruction. The

In 2005, Fabry et al. (15) published the 3D structure of the wild-type capsid of hAdV-5 as well as the 3D structure of a mutant virus with a deletion of protein IX. In the mutant structure the yellow and red densities shown in Fig. 5B were not present. At the time of that study the position represented by the red density was thought to be that of protein IIIa; the authors thus assumed that the lack of the red densities in the protein IX-deleted virus was due to a diminished stability of the capsid, i.e., the absence of protein IX could have resulted in the loss of protein IIIa (15). Recently, Saban et al. (45), using a 6-Å resolution map in which the  $\alpha$ -helices in the capsid were visible plus secondary structure predictions, reassigned the red density to a four-helical bundle built from the C-terminal parts of four copies of protein IX. The triskelion structure was considered to be made up from the N-terminal parts of three copies of protein IX. The density for protein IIIa was reassigned to a highly helical density present underneath the penton base (45). The reassignment of the position of the C-terminal domain of protein IX was confirmed by Marsh et al., who showed that a protein IX mutant virus with GFP added at the C terminus had an additional density above the reassigned position for the C terminus (33), i.e., the position of the red domains shown in Fig. 5C.

In the CAV-2 capsid (Fig. 5A and B), the triskelion N-terminal part of protein IX is present at the same position in the capsid as in hAdV-5 (Fig. 5C). The notable difference between the two reconstructions is that in CAV-2 a cylindrical density is present on the top of the triskelion, whereas the red density observed in the hAdV-5 capsid (Fig. 5C) was absent. The red density on top of the triskelion in CAV-2 (Fig. 5A and B) has more or less the same dimensions as the red rods in hAdV-5. It is therefore likely that the cylindrical density on top of the triskelion in the CAV-2 structure is also the C-terminal domain of protein IX. This interpretation is reinforced by comparing the sequences of the two proteins. Protein IX is a two-domain protein consisting of an N-terminal part that contains some  $\beta$ -strands and a C-terminal,  $\alpha$ -helical part, both parts being connected by a linker, as shown by sequence alignment and secondary structure prediction (Fig. 5E). Saban et al. (45) did not detect any helices in the triskelion structure, confirming the assignment to the N-terminal domain of protein IX. This part of protein IX interacts with the hexon protein and should therefore be conserved. The N termini of the CAV-2 and the hAdV-5 sequences are the most similar between the two proteins (40% identity for the 52 first amino acids). However, the linker sequence (amino acids 40 to 81 in hAdV-5) is almost absent in CAV-2 (amino acids 39 to 53), and therefore the two domains must be close together in CAV-2. In contrast, in hAdV-5 the C and N termini could be quite far apart from each other, as suggested by Saban et al. (45). Although the CAV-2 C-terminal part of protein IX is slightly smaller and different from that of hAdV-5, this region

is still predicted to be helical and organized as a coil-coiled domain (Fig. 5E). Therefore, we suggest that three C-terminal domains of protein IX make up the rod on top of the triskelion structure in the CAV-2 capsid. To test this hypothesis we performed cryo-EM and calculated a 3D reconstruction of a CAV-2 virion harboring a C-terminal pIX-GFP fusion protein (31). While the quality of the reconstruction (25-Å resolution) is lower than that of the native CAV-2 capsid, it is good enough to recognize the triskelion, the cylindrical red density, and, at the top of that, a small extra density (Fig. 5D, green areas). This latter density is smaller than the size of GFP, probably because it is flexible. The flexible linker between the C-terminal part of protein IX and GFP (30, 31) would allow movement, and therefore only the average position is visible in the 3D reconstruction. We believe that our combined data comparing the reconstructions of CAV-2 and the capsid harboring the protein IX-GFP fusion confirm the localization of the C-terminal domain of CAV-2 protein IX. Our results also confirm the assignment of Saban et al. (45) of the red rods in the hAdV structure shown in Fig. 5C as the C-terminal domain of human virus protein IX.

**(ii) The interior of the capsid. (a) Protein IIIa.** As mentioned above, based on a 6-Å EM reconstruction and secondary structure predictions (45), the hAdV protein IIIa was reassigned to a helix-rich density lying under the penton base. This density is also present on the inside of the CAV-2 capsid (not shown). It exhibits the same topology as the human protein IIIa: there is an arm bound to the bottom part of the penton base, and, more toward the inside of the capsid, a globular region which resembles an empty sphere is visible and of same size as in hAdV-5. The CAV-2 protein IIIa is slightly smaller than the hAdV-5 protein (567 amino acids versus 585), but except for the 15 first amino acids, the N-terminal part predicted to be  $\alpha$ -helical is very conserved. The C-terminal part is notably different and smaller in CAV-2. However, at the present resolution we are unable to see such differences.

**(b) Protein VIII.** A total of 120 copies of protein VIII have been counted on the inner part of the hAdV-5 capsid (15). The same number of proteins can be counted in CAV-2, and each interacts with four hexons (not shown). The density corresponding to protein VIII seems to be slightly smaller and less continuous in CAV-2 than in hAdV-5. The sequences of these proteins are very similar (more than 66% similarity or identity), and therefore the differences observed in the density are probably due to a lack of resolution.

Finally, there is also an extra density just under the CAV-2 penton base (Fig. 1B, circle) that is absent in the hAdV-5 structure, but at this stage it remains unassigned.

**Conclusion.** This work shows the first 3D structure of a nonhuman AdV. There are well over 100 AdV serotypes currently identified, with less than half of them human. We predict that this number of nonhuman AdVs is only the tip of the

penton and the hexons are as described in panel A. Protein IX is in yellow (N-terminal part) and red (C-terminal part), and the extra density assigned to GFP is in green. (E) Sequence alignment of hAdV-5 and CAV-2 protein IX. Conserved residues are in white text and highlighted in red; partially conserved residues are in red text. The arrows and open bars represent predicted  $\beta$ -strands and  $\alpha$ -helices, respectively, as given by the secondary structure prediction software (<http://bioweb.pasteur.fr/seqanal/interfaces/predator-simple.html>). The predicted coil-coiled is boxed into a rectangle.

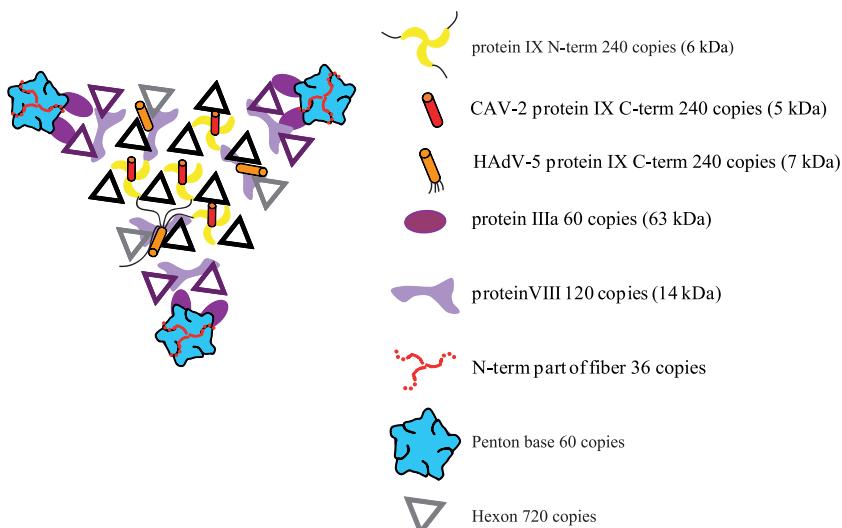


FIG. 6. Schematic view of one facet of the CAV-2 capsid seen from the outside of the capsid. The figure shows a summary of all structural data on the protein localization in the AdV capsid. The hexons highlighted in black belong to the group of nine whereas the violet ones are the peripentonal hexons (26). The different proteins occupy the same positions in the capsid (hAdV and the CAV-2) except for the C-terminal part of protein IX that can either be on the top of the N terminus of protein IX in the CAV-2 capsid or be near the twofold axis in the hAdV capsid. For the hAdV protein IX, a thin black line indicates the position of the linker between the N- and the C-terminal parts of protein IX as suggested by Saban et al. (45).

iceberg. We are just beginning to understand and uncover the potential of many of these viruses, in particular, of the non-human serotypes. Our study highlights the structural differences and similarities between hAdV-2/5 and CAV-2 and definitively localizes the N- and C-terminal parts of protein IX in this CAV-2 capsid. The molecular organization of the AdV capsid as it is known today is shown schematically in Fig. 6. The CAV-2 penton base and hexon have fewer and/or smaller loops around the conserved bodies of these capsid proteins. In the hAdV-2/5 penton base these loops are involved in internalization via the integrin-interacting domain (34), possibly binding coagulation factors (2), and are targeted by the humoral immune response following infection (37). The biological significance of many of the structural differences is likely to be complex. Yet we can now begin to rationally understand and resolve some of the CAV-2 enigmas. The capsid structure will allow us to rescan the external CAV-2 capsid for integrin-interacting motifs that could possibly induce internalization. Because of its C-terminal domain extending outside of the capsid wall, the CAV-2 protein IX may have an additional role during capsid disassembly at the nuclear pore by binding the anterograde motor kinesin (U. Greber, unpublished data). The smoother shape of the CAV-2 hexons and penton bases due to the lack of highly antigenic loops present in human serotypes explains the lack of cross-reacting human neutralizing antibodies (37).

Biologically active loops can be added onto the CAV-2 penton base and hexon at the sites where such loops are present in the human viruses. In particular, it will be interesting to add an RGD loop to the CAV-2 penton base at the same site where this is found in the human virus. Apart from adding loops to pentons and hexons, the specific position of the C-terminal domain of CAV-2 protein IX is an ideal site to position additional protein domains for specific interactions with host cells.

Finally, we may also be able to better understand and assay CAV-2 tropism by manipulating the rigidity of the flexible shaft of the fiber.

#### ACKNOWLEDGMENTS

We thank Pascal Fender and Harry Wodrich for discussions and Emmanuelle Neumann for help with the electron microscope. We thank Vasiliki Kalatzis for critical reading of the manuscript. We thank the other members of our laboratories for constructive comments, suggestions, and help during the course of this study. We thank the IFR 122 recombinant protein platform for the production of the CAV-2 fiber.

E.J.K. is an INSERM fellow. G.S. was supported in part by a Jeunes Chercheurs grant from the Agence Nationale pour la Recherche. D.T.C. was funded by NIH award 5R01CA111569. This work was supported, in part, by the Association Française contre les Myopathies (AFM), Vaincre les Maladies Lysosomales, and the Fondation pour le Recherche Médicale. O.B. was an AFM postdoctoral researcher, M.E.B. is a Ph.D. student financed by the Rhône Alpes region, and C.M.S.F. is a Ph.D. student financed through a MENRT fellowship from the French government.

The authors have no conflicting financial interests.

#### REFERENCES

1. Akli, S., C. Caillaud, E. Vigne, L. D. Stratford-Perricaudet, L. Poenaru, M. Perricaudet, A. Kahn, and M. R. Peschanski. 1993. Transfer of a foreign gene into the brain using adenovirus vectors. *Nat. Genet.* **3**:224–228.
2. Baker, A. H., J. H. McVey, S. N. Waddington, N. C. Di Paolo, and D. M. Shayakhmetov. 2007. The influence of blood on *in vivo* adenovirus bio-distribution and transduction. *Mol. Ther.* **15**:1410–1416.
3. Baker, T. S., and R. H. Cheng. 1996. A model-based approach for determining orientations of biological macromolecules imaged by cryo-electron microscopy. *J. Struct. Biol.* **116**:120–130.
4. Bergelson, J. M., J. A. Cunningham, G. Drogue, E. A. Kurt-Jones, A. Krithivas, J. S. Hong, M. S. Horwitz, R. L. Crowell, and R. W. Finberg. 1997. Isolation of a common receptor for coxsackie B viruses and adenoviruses 2 and 5. *Science* **275**:1320–1323.
5. Both, G. 2002. Xenogenic adenoviruses, p. 447–479. *In* D. Curiel and J. Douglas (ed.), *Adenoviral vectors for gene therapy*. Academic Press, San Diego, CA.
6. Campos, S. K., and M. A. Barry. 2007. Current advances and future chal-

- lenges in adenoviral vector biology and targeting. *Curr. Gene. Ther.* **7**:189–204.
- Chappell, J. D., A. E. Prota, T. S. Dermody, and T. Stehle. 2002. Crystal structure of reovirus attachment protein signal reveals evolutionary relationship to adenovirus fibre. *EMBO J.* **21**:1–11.
  - Chillon, M., and E. Kremer. 2001. Trafficking and propagation of canine adenovirus vectors lacking a known integrin-interacting motif. *Hum. Gene Ther.* **12**:1815–1823.
  - Chroboczek, J., R. W. Ruigrok, and S. Cusack. 1995. Adenovirus fiber. *Curr. Top. Microbiol. Immunol.* **199**:163–200.
  - Conway, J. F., and A. C. Steven. 1999. Methods for reconstructing density maps of “single” particles from cryoelectron micrographs to subnanometer resolution. *J. Struct. Biol.* **128**:106–118.
  - D'Ambrosio, E., N. Del Grosso, A. Chicca, and M. Midulla. 1982. Neutralizing antibodies against 33 human adenoviruses in normal children in Rome. *J. Hyg.* **89**:155–161.
  - Davison, A. J., M. Benko, and B. Harrach. 2003. Genetic content and evolution of adenoviruses. *J. Gen. Virol.* **84**:2895–2908.
  - Dayringer, H. E., A. Tramontano, S. R. Sprang, and R. J. Fletterick. 1986. Interactive program for visualization and modeling of protein, nucleic acids and small molecules. *J. Mol. Graph.* **4**:82–87.
  - DeLano, M. C., and Y. Cao. 2002. High b-value diffusion imaging. *Neuroimaging Clin. N. Am.* **12**:21–34.
  - Fabry, C. M., M. Rosa-Calatrava, J. F. Conway, C. Zubietta, S. Cusack, R. W. Ruigrok, and G. Schoehn. 2005. A quasi-atomic model of human adenovirus type 5 capsid. *EMBO J.* **24**:1645–1654.
  - Fasbender, A., J. H. Lee, R. W. Walters, T. O. Moninger, J. Zabner, and M. J. Welsh. 1998. Incorporation of adenovirus in calcium phosphate precipitates enhances gene transfer to airway epithelia in vitro and in vivo. *J. Clin. Investig.* **102**:184–193.
  - Fasbender, A., J. Zabner, M. Chillon, T. O. Moninger, A. P. Puga, B. L. Davidson, and M. J. Welsh. 1997. Complexes of adenovirus with polycationic polymers and cationic lipids increase the efficiency of gene transfer in vitro and in vivo. *J. Biol. Chem.* **272**:6479–6489.
  - Flomberg, P., V. Piaskowski, R. L. Truitt, and J. T. Casper. 1995. Characterization of human proliferative T cell responses to adenovirus. *J. Infect. Dis.* **171**:1090–1096.
  - Frank, J., M. Radermacher, P. Penczek, J. Zhu, Y. Li, M. Ladjadj, and A. Leith. 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**:190–199.
  - Fuschiotti, P., G. Schoehn, P. Fender, C. M. Fabry, E. A. Hewat, J. Chroboczek, R. W. Ruigrok, and J. F. Conway. 2006. Structure of the dodecahedral penton particle from human adenovirus type 3. *J. Mol. Biol.* **356**:510–520.
  - Galinier, R., E. Gout, H. Lortat-Jacob, J. Wood, and J. Chroboczek. 2002. Adenovirus protein involved in virus internalization recruits ubiquitin-protein ligases. *Biochemistry* **41**:14299–14305.
  - Garcel, A., E. Gout, J. Timmins, J. Chroboczek, and P. Fender. 2006. Protein transduction into human cells by adenovirus dodecahedron using WW domains as universal adaptors. *J. Gene. Med.* **8**:524–531.
  - Gouet, P., X. Robert, and E. Courcelle. 2003. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**:3320–3323.
  - Greber, U. F. 2002. Signalling in viral entry. *Cell Mol. Life Sci.* **59**:608–626.
  - Hess, M., A. Cuzange, R. W. Ruigrok, J. Chroboczek, and B. Jacrot. 1995. The avian adenovirus penton: two fibres and one base. *J. Mol. Biol.* **252**:379–385.
  - Horwitz, M. 2001. Adenoviruses, p. 2149–2171. In D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus (ed.), *Fields virology*, 4th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
  - Keriel, A., C. Rene, C. Galer, J. Zabner, and E. J. Kremer. 2006. Canine adenovirus vectors for lung-directed gene transfer: efficacy, immune response, and duration of transgene expression using helper-dependent vectors. *J. Virol.* **80**:1487–1496.
  - Kojaoghlianian, T., P. Flomberg, and M. S. Horwitz. 2003. The impact of adenovirus infection on the immunocompromised host. *Rev. Med. Virol.* **13**:155–171.
  - Kremer, E. J. 2004. CAR chasing: canine adenovirus vectors—all bite and no bark? *J. Gene. Med.* **6**(Suppl. 1):S139–S151.
  - Kremer, E. J., S. Boutin, M. Chillon, and O. Danos. 2000. Canine adenovirus vectors: an alternative for adenovirus-mediated gene transfer. *J. Virol.* **74**:505–512.
  - Le, L. P., J. Li, V. V. Ternovoi, G. P. Siegal, and D. T. Curiel. 2005. Fluorescently tagged canine adenovirus via modification with protein IX-enhanced green fluorescent protein. *J. Gen. Virol.* **86**:3201–3208.
  - Le Gal La Salle, G., J. J. Robert, S. Berrard, V. Ridoux, L. D. Stratford-Perricaudet, M. Perricaudet, and J. Mallet. 1993. An adenovirus vector for gene transfer into neurons and glia in the brain. *Science* **259**:988–990.
  - Marsh, M. P., S. K. Campos, M. L. Baker, C. Y. Chen, W. Chiu, and M. A. Barry. 2006. Cryoelectron microscopy of protein IX-modified adenoviruses suggests a new position for the C terminus of protein IX. *J. Virol.* **80**:11881–11886.
  - Medina-Kauwe, L. K. 2003. Endocytosis of adenovirus and adenovirus capsid proteins. *Adv. Drug Deliv. Rev.* **55**:1485–1496.
  - Olive, M., L. Eisenlohr, N. Flomberg, S. Hsu, and P. Flomberg. 2002. The adenovirus capsid protein hexon contains a highly conserved human CD4<sup>+</sup> T-cell epitope. *Hum. Gene. Ther.* **13**:1167–1178.
  - Paillard, F. 1997. Advantages of nonhuman adenoviruses versus human adenoviruses. *Hum. Gene. Ther.* **8**:2007–2009.
  - Perreau, M., M. C. Guerin, C. Drouet, and E. J. Kremer. 2007. Interactions between human plasma components and a xenogenic adenovirus vector: reduced immunogenicity during gene transfer. *Mol. Ther.* **15**:1998–2007.
  - Perreau, M., and E. J. Kremer. 2005. Frequency, proliferation, and activation of human memory T cells induced by a nonhuman adenovirus. *J. Virol.* **79**:14595–14605.
  - Perreau, M., and E. J. Kremer. 2006. The conundrum between immunological memory to adenovirus and their use as vectors in clinical gene therapy. *Mol. Biotechnol.* **34**:247–256.
  - Perreau, M., F. Mennechet, N. Serratrice, J. N. Glasgow, D. T. Curiel, H. Wodrich, and E. J. Kremer. 2007. Contrasting effects of human, canine, and hybrid adenovirus vectors on the phenotypical and functional maturation of human dendritic cells: implications for clinical efficacy. *J. Virol.* **81**:3272–3284.
  - Rasmussen, U. B., Y. Schlesinger, A. Pavirani, and M. Mehtali. 1995. Sequence analysis of the canine adenovirus 2 fiber-encoding gene. *Gene* **159**:279–280.
  - Ruigrok, R. W., A. Barge, C. Albiges-Rizo, and S. Dayan. 1990. Structure of adenovirus fibre. II. Morphology of single fibres. *J. Mol. Biol.* **215**:589–596.
  - Ruigrok, R. W. H., A. Barge, S. K. Mittal, and B. Jacrot. 1994. The fibre of bovine adenovirus is very long, but bent. *J. Gen. Virol.* **75**:2069–2073.
  - Rux, J. J., P. R. Kuser, and R. M. Burnett. 2003. Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution X-ray crystallographic, molecular modeling, and sequence-based methods. *J. Virol.* **77**:9553–9566.
  - Saban, S. D., M. Silvestry, G. R. Nemerow, and P. L. Stewart. 2006. Visualization of  $\alpha$ -helices in a 6-angstrom resolution cryoelectron microscopy structure of adenovirus allows refinement of capsid protein assignments. *J. Virol.* **80**:12049–12059.
  - Schoehn, G., F. M. Vellieux, M. Asuncion Dura, V. Receveur-Brechot, C. M. Fabry, R. W. Ruigrok, C. Ebel, A. Roussel, and B. Franzetti. 2006. An archaeal peptidase assembles into two different quaternary structures: a tetrahedron and a giant octahedron. *J. Biol. Chem.* **281**:36327–36337.
  - Seiradake, E., H. Lortat-Jacob, O. Billet, E. J. Kremer, and S. Cusack. 2006. Structural and mutational analysis of human Ad37 and canine adenovirus 2 fiber heads in complex with the D1 domain of coxsackie and adenovirus receptor. *J. Biol. Chem.* **281**:33704–33716.
  - Soudais, C., S. Boutin, S. S. Hong, M. Chillon, O. Danos, J. M. Bergelson, P. Boulanger, and E. J. Kremer. 2000. Canine adenovirus type 2 attachment and internalization: coxsackievirus-adenovirus receptor, alternative receptors, and an RGD-independent pathway. *J. Virol.* **74**:10639–10649.
  - Soudais, C., C. Laplace-Builhe, K. Kissa, and E. J. Kremer. 2001. Preferential transduction of neurons by canine adenovirus vectors and their efficient retrograde transport in vivo. *FASEB J.* **15**:2283–2285.
  - Soudais, C., N. Skander, and E. J. Kremer. 2004. Long-term in vivo transduction of neurons throughout the rat central nervous system using novel helper-dependent CAV-2 vectors. *FASEB J.* **18**:391–393.
  - Stewart, P. L., R. M. Burnett, M. Cyrlaff, and S. D. Fuller. 1991. Image reconstruction reveals the complex molecular organization of adenovirus. *Cell* **67**:145–154.
  - Tomko, R., R. Xu, and L. Philipson. 1997. HCAR and MCAR: the human and mouse cellular receptors for subgroup C adenoviruses and group B coxsackieviruses. *Proc. Natl. Acad. Sci. USA* **94**:3352–3356.
  - Van Heel, M. 1987. Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* **21**:111–123.
  - van Raaij, M. J., A. Mitraki, G. Lavigne, and S. Cusack. 1999. A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**:935–938.
  - Wrighers, W., and S. Birnmanns. 2001. Using situs for flexible and rigid-body fitting of multiresolution single-molecule data. *J. Struct. Biol.* **133**:193–202.
  - Wu, E., and G. R. Nemerow. 2004. Virus yoga: the role of flexibility in virus host cell recognition. *Trends Microbiol.* **12**:162–169.
  - Wu, E., L. Pache, D. J. Von Seggern, T. M. Mullen, Y. Mikyas, P. L. Stewart, and G. R. Nemerow. 2003. Flexibility of the adenovirus fiber is required for efficient receptor interaction. *J. Virol.* **77**:7225–7235.
  - Zubietta, C., G. Schoehn, J. Chroboczek, and S. Cusack. 2005. The structure of the human adenovirus 2 penton. *Mol. Cell* **17**:121–135.

## IV.2 Conclusion

Dans cette partie, nous nous sommes intéressés à la structure de la capsid du CAdV-2 par CryoME et « fitting » des structures cristallographiques des protéines majeures de la capsid. Nous avons obtenu, pour la première fois une reconstruction 3D de la capsid d'un Adénovirus non humain à 12 Å de résolution. C'est aussi la première reconstruction 3D d'un Adénovirus canin. Notre étude montre les différences et les similitudes qui peuvent exister entre la structure de la capsid des Adénovirus humains et canins et localise de façon convaincante les extrémités N et C terminales de la protéine IX chez CAdV-2. La capsid du virus canin semble plus lisse que celle des virus humains. De nombreuses boucles externes et flexibles présentes dans la structure de l'Hexon et de la Base du Penton des virus humains sont plus courtes voir absentes de la structure des protéines majeures de capsid chez CAdV-2. Ceci donne naissance à une capsid faisant penser à la reconstruction 3D du virus CELO et explique la faible antigénicité du CAdV-2 présente chez l'homme. La forte antigénicité des virus humains réside en la présence de séquences hypervariables à la surface des Hexons. La comparaison des modèles 3D obtenu pour le CAV et le CAVpIX-GFP nous a permis de localiser de façon précise la protéine IX. On observe notamment une densité cylindrique au dessus des triskelions formés par la protéine IX. Cette densité cylindrique, jamais observée auparavant, a été attribuée aux extrémités C terminales de la protéine IX car elle fait fortement penser à la densité qui avait été précédemment attribuée à la protéine IIIa chez les virus humains. Sa présence s'explique justement par la présence d'une zone charnière au milieu de la séquence de la protéine IX, plus courte chez les virus canins que chez les virus humains. La dernière partie de notre étude s'intéresse à la modélisation de la fibre du CAdV-2. Nous avons pu, à partir d'images prises en coloration négative de la fibre exprimée de façon recombinante, déterminer que cette protéine possède au moins deux zones flexibles. A partir de ces données, nous avons élaboré un modèle atomique de la fibre en combinant la structure de la tête de la fibre du CAdV-2 avec la structure d'une partie de la tige résolue chez HAdV-2 ainsi qu'avec une partie de la protéine d'attachement σ1 des Réovirus. L'absence de boucles en surface au niveau de la Base du Penton et de l'Hexon ainsi que la forme atypique des extrémités C terminales de la protéine IX constituent des sites potentiels de modifications de la capsid en vue d'un reciblage efficace du virus dans le cadre du développement d'un vecteur optimisé pour la thérapie génique.

## V. Etude structurale des Complexes Fab - Adénovirus Humain de type 5 pour la localisation de l'extrémité C terminale de la protéine IX

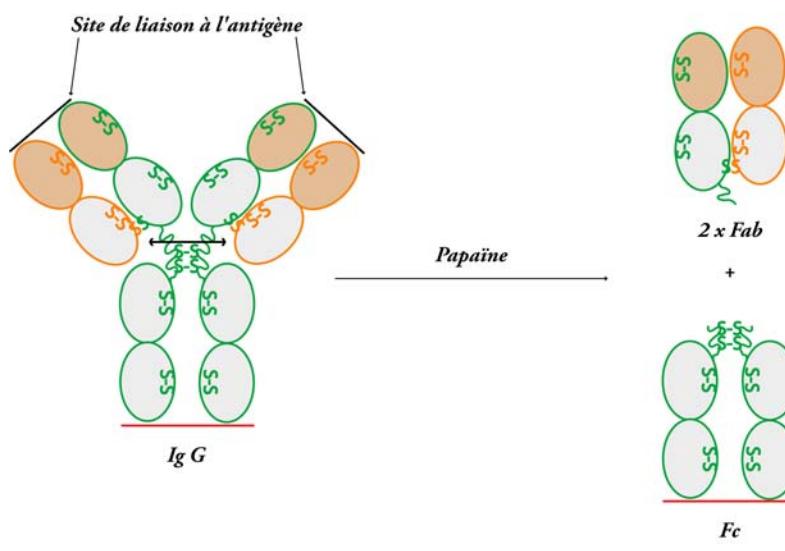
L'étude structurale réalisée chez l'Adénovirus canin de type 2 semble localiser de façon convaincante les différentes parties de la protéine IX mais rien ne nous prouve qu'il en est de même pour les virus humains. La localisation de l'extrémité N terminale de la protéine n'a pas été controversée au cours des études structurales précédemment menées. L'ensemble de la communauté scientifique s'accorde pour placer cette extrémité en contact avec la base des Hexons. C'est pourquoi cette partie de ma thèse est uniquement consacrée à la visualisation de l'extrémité C-terminale de la protéine IX au niveau de la capsid des Adénovirus humains. De récents travaux publiés par Saban et collaborateurs semblent montrer, chez les virus humains, que la densité précédemment identifiée comme la protéine IIIa sur les reconstructions 3D en CryoME serait peut être un complexe réalisé par quatre extrémités C terminale de protéines IX (Saban et al., 2006). Au vue de ces données, nous avons donc tenté d'apporter une preuve structurale à la localisation de l'extrémité C terminale de la protéine IX dans la capsid des Adénovirus humains. Pour cela, nous avons réalisé un marquage anticorps et procédé à une étude structurale en CryoME sur le complexe virus entier - anticorps.

### **V.1 Matériel et Méthodes**

#### *V. 1.1 Les échantillons*

Grâce à une collaboration avec le Pr. Pierre Boulanger et le Dr. Manuel Rosa-Calatrava (IFR Laennec, Lyon), nous disposons tout d'abord d'un échantillon de HAdV-5 mutant à une concentration d'environ  $8 \times 10^{12}$  particules/ml. Ce virus noté HAdV-5\_SY12 présente une modification au niveau de la protéine IX. En effet, un peptide de 12 acides aminés noté SY12 (TAYSSYMKGGKF) est fusionné à l'extrémité C terminale de la protéine IX sans peptide de liaison. Après expression et purification, la solution virale est conservée à - 80°C dans un tampon PBS 1x contenant 10% de glycérol. Nous disposons, de plus, d'un lot d'anticorps polyclonaux de lapin dirigés contre le peptide SY12. Ces anticorps ont été récoltés dans le sérum d'un lapin après 87 jours post-injection et purifiés sur une colonne d'affinité contre le peptide SY12 lui-même. Après 87 jours post-infection, le sérum est essentiellement constitué d'Immunoglobulines de type G (Ig G).

La connaissance du type d'anticorps ainsi que de l'espèce de l'animal duquel ils proviennent est importante pour la mise au point d'un protocole de purification adapté. Leur spécificité pour le peptide SY12 a été vérifiée par Western Blot. La solution d'anticorps à une concentration de 0,8 mg/ml est conservée à -20 °C dans un tampon PBS 1x contenant 0,1% p/v SAB ainsi que du thimerosal. Le thimerosal est une molécule cyclique utilisée comme agent de conservation des anticorps. Sa forte absorbance dans l'UV nous oblige à utiliser la méthode de Bradford pour la mesure des concentrations protéiques.



**Figure 45: Principe de coupure d'un anticorps à la papaine.**

Le clivage enzymatique à la papaine permet de séparer l'anticorps en 2 fragments Fab et 1 fragment Fc.

La double flèche noire représente le site de coupure de la papaine. Les domaines entourés en vert appartiennent à la chaîne lourde, ceux en orange constituent la chaîne légère. Les domaines marqués en gris correspondent aux domaines constants et ceux en marron clair aux domaines variables des chaînes lourdes et légères. Le trait rouge représente la zone de liaison à la protéine A au niveau de la partie Fc.

Un anticorps est une molécule de l'immunité acquise ou adaptative, utilisée pour détecter et neutraliser les agents pathogènes. Les anticorps sont secrétés par une cellule dérivée des lymphocytes B, les plasmocytes après exposition à un antigène quelconque. Ces molécules sont des glycoprotéines de la superfamille des immunoglobulines formées de 4 chaînes polypeptidiques (150 kDa environ): 2 chaînes lourdes (50 kDa environ chacune) et 2 chaînes légères (25 kDa environ chacune) qui sont reliées entre elles par un nombre variable de ponts disulfures intra et inter-chaînes (S-S sur la Figure 45) assurant ainsi une rigidité à la molécule. Un anticorps possède trois extrémités, deux Fab responsables de la liaison avec l'antigène ainsi qu'une extrémité Fc constante qui relie les deux Fab (Figure 45). Les deux extrémités capables d'interagir avec l'antigène étant reliées, l'attachement d'un anticorps à son antigène entraîne souvent une agrégation de l'antigène par pontage via les molécules d'anticorps. Pour réaliser une étude structurale en MET, il est nécessaire d'empêcher ce pontage. Pour cela, une des solutions est d'individualiser les domaines Fab par digestion enzymatique à la papaine (Figure 45). Les Fab obtenus sont, par la suite, purifiés et séparés du reste des fragments

(anticorps non digérés, fragments Fc et SAB) avant de réaliser un complexe avec le virus. Je vais donc détailler dans les parties suivantes le protocole de digestion des anticorps et de purification des fragments Fab puis la réalisation d'un complexe entre le virus entier et les fragments Fab purifiés. Enfin je présenterai les résultats que nous avons obtenu à l'issu de l'analyse d'images en MET concernant la localisation de l'extrémité C terminale de la protéine IX chez le HAdV-5.

### *V. 1.2 Préparation et Purification des Fab*

Tous les tubes et concentrateurs utilisés pour la manipulation des anticorps sont préalablement incubés une nuit à 4° C avec une solution de passivation (1% Lait écrémé dans du PBS 1X) afin d'éviter l'interaction non spécifique des anticorps avec le plastique des tubes. Le principe de purification des Ig G de lapin est basé sur la propriété de la protéine A à fixer les extrémités Fc des Ig G avec une bonne affinité (entre 8 et 15 mg d'Ig G par ml de gel de protéine A).

*Colonne d'affinité à la protéine A.* Cette étape est réalisée à l'aide du Nab Protein A plus spin kit (Pierce), conformément au protocole. Un volume de 700 µl d'anticorps purifié est incubé 1h à température ambiante, sous agitation, avec 150 µl de résine protéine A. La résine est lavée avec 3 volumes colonne de PBS 1x puis l'anticorps est élué de la résine avec 3 volumes colonnes de tampon d'élution (fourni dans le kit). L'acidité du tampon d'élution est neutralisée à l'aide d'une solution de 1 M Tris pH 8. La solution d'anticorps est ensuite concentrée à l'aide d'un concentrateur vivaspin2 (Vivascience) possédant une membrane Hydrosart (c/o = 30 kDa). Cette étape permet d'éliminer une grande partie de la SAB et du thimerosal qui contamine fortement l'échantillon. Lors de cette première étape de purification, l'anticorps est retenu sur la résine alors que la SAB et le thimérosal passent directement dans le tampon de lavage de la résine.

*Digestion à la papaine.* Cette digestion est réalisée par de la papaine immobilisée sur des billes d'agarose (Pierce) avec un rapport E:S de 10:1 en quantité conformément au protocole fourni. L'anticorps est incubé avec le gel de papaine sous agitation à 37 °C pendant 12h. L'arrêt de la réaction est réalisé sur glace avec 1 volume de tampon de lavage 10 mM tris pH 7,5 puis le gel est rincé avec 1 volume équivalent de tampon de lavage.

*Colonne d'affinité à la protéine A.* Cette étape est aussi réalisée à l'aide du Nab Protein A plus spin kit conformément au protocole. La résine est incubée 2h à température ambiante sous agitation avec la solution d'anticorps. Les fragments Fab sont récupérés au lavage de la résine et les fragments Fc ainsi que les anticorps partiellement ou non digérés sont élués conformément au protocole. La

solution récupérée après lavage contenant les fragments Fab est enfin concentrée à l'aide d'un concentrateur Vivaspin2 Hydrosart ( $c/o = 30$  kDa) et la concentration protéique est mesurée avec la méthode de Bradford.

#### *V. 1.3 Préparation des complexes Virus – Fab*

Une aliquote de 50  $\mu$ l de virus HAdV-5\_SY12 est décongelé puis le glycérol est éliminé par centrifugation sur une unité Microbiospin6 (Biorad) conformément au protocole fourni. Le virus est ensuite incubé pendant 2h à température ambiante, sans agitation, avec la solution de Fab précédemment préparée à un ratio molaire de protéine IX:Fab de 1:10. La présence d'anticorps à la surface du virus HAdV-5\_SY12 est vérifiée par observation en MET par la méthode de coloration négative dans 2% AmMb.

#### *V. 1.4 CryoME*

Nous avons étudié un échantillon de virus HAdV-5\_SY12 seul comme contrôle puis en complexe avec le Fab purifié anti-SY12. Dans les deux cas, les images ont été réalisées par le Dr. Guy Schoehn sur le Microscope JEOL 2010F de l'IBS à un grossissement nominal de 30 000 fois. Les images sélectionnées ont été scannées avec un pas de 7  $\mu$ m sur le scanner ZEISS de l'IBS. Cela correspond à une taille calibrée de pixel de 2,3  $\text{\AA}$ . Pour le traitement des images, les particules ont été sélectionnées manuellement dans X3D à l'aide d'une sélection carrée de 447 pixels de coté. Comme pour les traitements des images du HAdV-5 WT, la recherche des orientations et des centres a été réalisée avec PFT puis PFT2 à partir de la reconstruction 3D à 10  $\text{\AA}$  de résolution du HAdV-5. La reconstruction 3D est calculée avec EM3DR puis EM3DR2. La CTF des images a été inversées jusqu'à 15  $\text{\AA}$  et déconvolées à 8  $\text{\AA}$  pour l'échantillon contrôle. Pour l'échantillon en complexe avec le Fab, les images ont été inversées jusqu'à 15  $\text{\AA}$  de résolution pour la recherche des orientations et déconvolées à 8  $\text{\AA}$  pour réaliser une reconstruction 3D.

## V.2 Résultats et discussion

Purification du Fab à partir d'un lot d'anticorps polyclonaux anti-SY12 de lapin. J'ai mis au point la purification des Fab à partir des lots d'anticorps fournis par nos collaborateurs de Lyon. L'échantillon contient une concentration en SAB aussi forte que celle de l'anticorps (1 mg/ml) comme le montre l'étoile sur la [Figure 46a](#). Malheureusement cela ne permet pas de réaliser une étude en MET. La présence de contaminants en trop grande concentration entraîne beaucoup de bruits dans le fond de l'image et donne une reconstruction 3D finale très bruitée. Il faut donc en premier lieu éliminer la SAB. Pour cela, j'ai réalisé une première étape de purification par affinité sur résine de protéine A. L'anticorps fixé sur la résine est élué indépendamment des contaminants ([Figure 46b](#), Avt). Après cette première étape, il reste encore de la SAB dans l'échantillon mais faiblement par rapport à la concentration en anticorps ce qui ne gène pas pour réaliser une étude en MET. Par la suite, l'anticorps subit une digestion à la papaïne. La coupure des chaînes lourdes a libéré deux fragments d'environ 25 kDa chacun. On peut voir le résultat de la digestion sur la [Figure 46 \(Après Papaïne\)](#). Deux bandes de taille différentes apparaissent entre 25 et 30 kDa alors que la bande de la chaîne lourde à 50 kDa a presque entièrement disparue. La présence d'une bande surnuméraire après digestion à la papaïne d'une taille inférieure à 15 kDa correspond la digestion en deux parties du fragment Fc ([Figure 46](#)). En effet lorsque la digestion à la papaïne est trop longue, il se produit un clivage supplémentaire à environ la moitié du Fc. Ce clivage entraîne l'apparition d'un demi-fragment Fc qui se lie à la protéine A et de la 2<sup>ème</sup> partie qui a perdu sa capacité d'attachement avec la protéine A. La coupure du fragment Fc en deux parties explique pourquoi l'intensité de la bande du Fc est plus faible que celle du Fab sur la piste R, Après Papaïne de la [Figure 46b](#).

La purification des fragments Fab après coupure à la papaïne est réalisée par une 2<sup>ème</sup> étape d'affinité pour la protéine A. Cette étape a permis de séparer les fragments Fab des Fc comme on peut le voir sur les pistes Fab et Fc de la [Figure 46](#). Les bandes correspondantes aux parties d'anticorps ne sont pas nettes sur le gel. Ces bandes sont assez étendues, laissent des « trainées » ou traces sur le gel certainement à cause de la glycosylation des chaînes lourdes et légères d'anticorps. La glycosylation d'une protéine entraîne la présence de plusieurs formes de la protéine en solution avec des poids moléculaires différents.

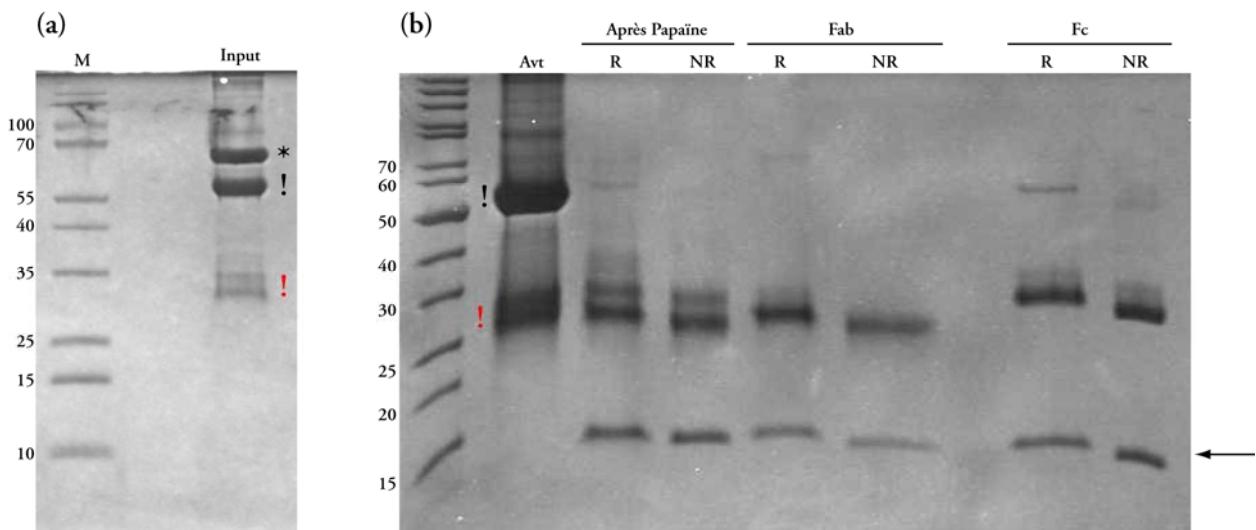


Figure 46: Gels 12% Tris-Tricine en conditions dénaturantes de contrôle en cours de purification.

M = Marqueur de poids moléculaire, le poids correspondant à chaque bande est marqué sur le côté gauche du gel, les valeurs de masse moléculaires sont exprimées en kDa. (a) Anticorps anti-SY12 : \* = SAB ; ! = chaînes lourdes d'Ig G ; ! = chaînes légères d'Ig G. (b) Avt = Anticorps à la sortie de la 1<sup>ère</sup> colonne de protéine A. Après papaïne : Echantillon après la digestion à la papaïne. Après digestion à la papaïne, Fab : Fraction non retenue après une 2<sup>eme</sup> étape de purification sur protéine A. Fc : Fraction élue après une 2<sup>eme</sup> étape de purification sur protéine A. La flèche indique la bande surnuméraire après digestion à la papaïne. NR = conditions non réductrices, R = conditions réductrices.

Ceci a pour conséquence la présence, sur un gel dénaturant, d'une série de bandes très proches au niveau de la bande majoritaire laissant ainsi un effet de « trainées » sur le gel. La 2<sup>ème</sup> étape de purification ne permet malheureusement pas d'éliminer une partie du Fc puisque celui-ci est clivé en deux parties dont une qui n'a pas d'affinité pour la protéine A. A la fin de la purification l'échantillon de Fab obtenu reste suffisamment pur pour ne pas gêner l'observation et la formation des complexes Fab avec le virus entier HAdV-5\_SY12. Afin de vérifier si l'anticorps n'est pas dénaturé, j'ai déposé en parallèle le même échantillon en conditions réductrices et non réductrices.

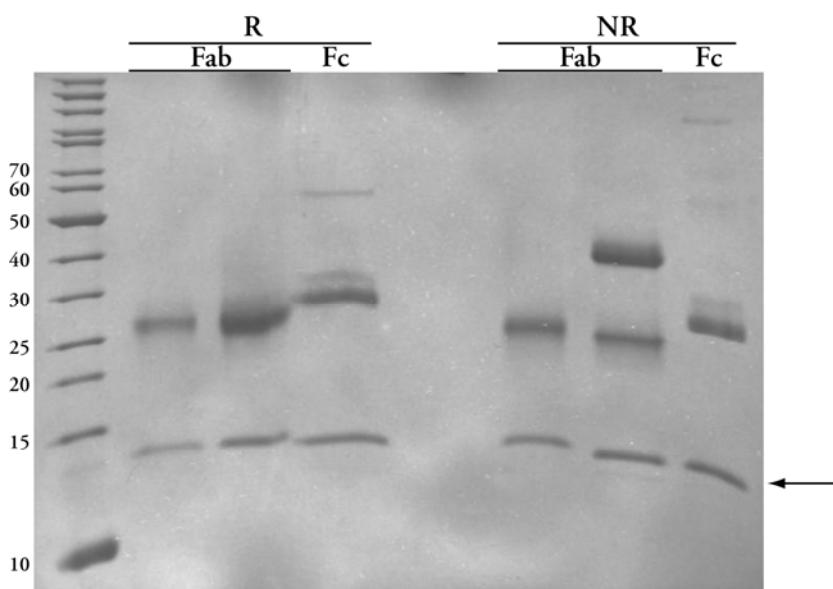
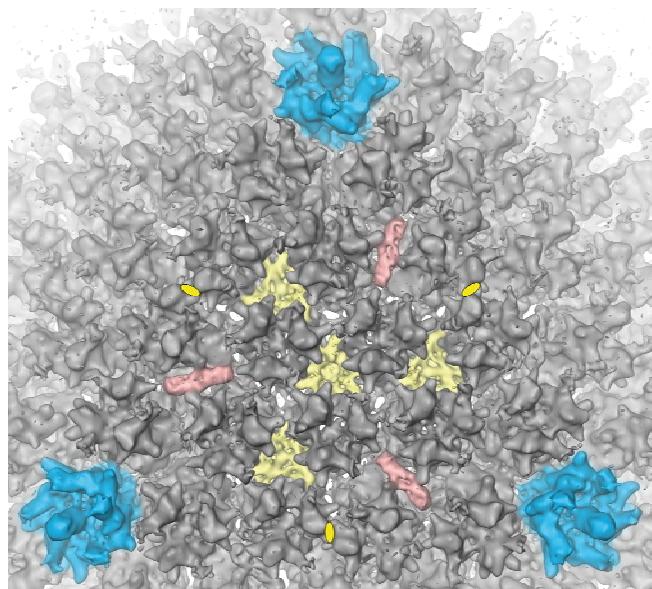


Figure 47: Gel 12% Tris-Tricine en conditions dénaturantes de contrôle en fin de purification.

R correspond à l'échantillon purifié déposé en présence d'un agent réducteur le  $\beta$ - mercaptoéthanol alors que les 3 pistes suivantes ont été déposées sur le gel en absence de cet agent réducteur. Fab : Fraction non retenue sur la protéine A. Fc : Fraction élue de la protéine A. La flèche indique la bande surnuméraire observée après digestion à la papaïne. NR = conditions non réductrices, R = conditions réductrices.

En conditions réductrices, les ponts disulfures sont réduits et les différentes chaînes d'anticorps sont individualisées. En conditions non réductrices, les différentes chaînes sont dénaturées mais restent attachées par les ponts disulfures qui les relient. On peut voir que le Fab existe sous sa forme non réduite seulement après la concentration finale de l'échantillon. En effet, le tampon de digestion pour la papaine contient 20 mM cystéine réduite. La présence de la cystéine entraîne la réduction de l'anticorps. En solution, il n'existe plus que la forme oxydée de la molécule comme on peut le voir sur la [Figure 47](#). Après la concentration, la cystéine initialement présente passe à travers la membrane du concentrateur et est ainsi séparée des fragments Fab. Après concentration, on peut voir apparaître sur le gel une bande, entre 40 et 50 kDa, correspondant au Fab oxydé.

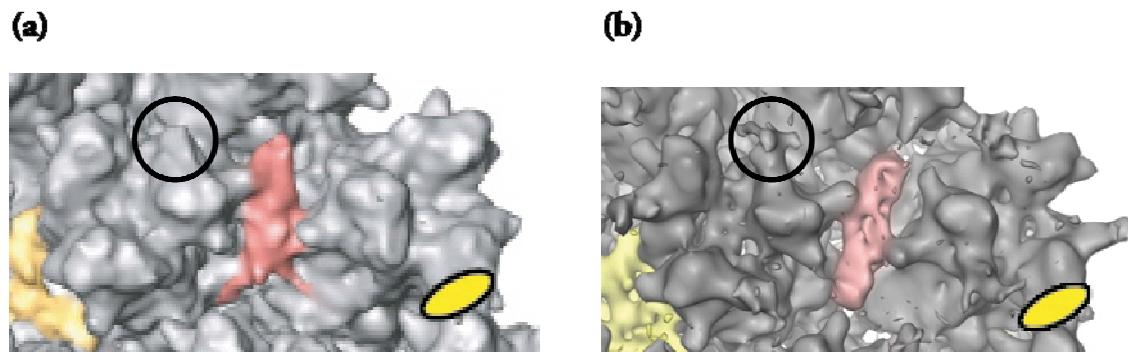
Le modèle 3D du virus HAdV-5 SY12. Le modèle a été calculé à partir de 18 micrographies sur lesquelles ont pu être sélectionnées un total de 4905 particules dont 60% (2943) ont servi à calculer une reconstruction 3D finale d'environ 13 Å de résolution à  $FSC_{0,5}$  (cf. FSC en Annexe 4). Le modèle 3D calculé semble complet ([Figure 48](#)). Toutes les protéines majeures sont présentes au sein de la capsid. Les Pentons sous forme entière sont observés au niveau des sommets et les Hexons forment bien le reste de la capsid ([Figure 48](#)). Ceci renseigne sur la qualité de l'échantillon viral utilisé pour réaliser la CryoME. Cet échantillon semble avoir bien supporté la congélation et était intact au moment de sa préparation pour la ME. En termes de protéines mineures externes, ce virus mutant semble posséder toutes les caractéristiques structurales d'un HAdV-5 WT. La position des protéines mineures externes semble conservée. On observe bien, à sa surface, les densités précédemment attribuées aux protéines IX et IIIA ([Figure 48](#)).



*Figure 48: Vue de dessus d'une face de la Reconstruction 3D obtenue pour le HAdV-5 SY12.*  
Les Pentons sont colorés en Bleu. On voit les densités précédemment attribuées aux protéines IX en jaune et IIIa en rose. Les axes de symétrie d'ordre 2 sont indiqués par des ellipses jaunes.

On note tout de même une différence au niveau de la reconstruction calculée par rapport à celle obtenue à haute résolution du HAdV-5 WT. En effet, les boucles flexibles situées au sommet des Hexons sont beaucoup plus détaillées dans la nouvelle reconstruction (Zone entourée sur la Figure 49a et b). Un niveau de détail important est aussi visible pour le Penton où l'on remarque clairement les zones de contact entre la Base et la Fibre. L'utilisation d'un microscope FEG dans cette étude montre clairement la qualité des images ainsi que l'importance de l'utilisation d'une source plus cohérente. Pour une résolution plus faible, l'utilisation d'un microscope FEG nous a permis de visualiser plus de détails que dans la reconstruction à 9 Å que nous avions obtenue pour le HAdV-5 WT.

De plus, on remarque du côté externe de la capsid une différence majeure au niveau de la densité précédemment attribuée à la protéine IIIa. Cette densité semble légèrement plus longue et plus étoffée au niveau des extrémités. Elle ne possède pas la même inclinaison que celle du virus WT (Figure 49). Il est difficile de quantifier ce changement d'orientation mais celui-ci est tout de même notable sur une vue de dessus de la capsid (Figure 49).



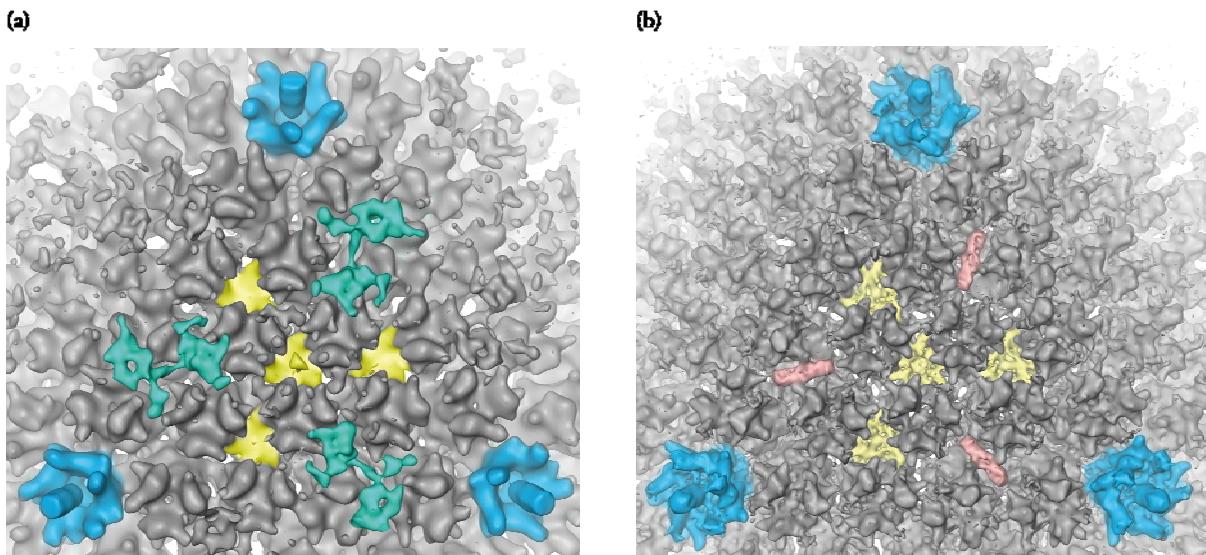
**Figure 49: Comparaison des densités attribuées à la protéine IIIa entre le virus WT et SY12.**  
 (a) Densité attribuée à la protéine IIIa dans la reconstruction du HAdV-5 WT en rose (cf. Article 1). (b) Même densité dans la reconstruction du HAdV-5\_SY12, en rose. L'ellipse jaune marque la position d'un axe icosaédrique d'ordre 2. Le cercle montre une boucle flexible, du côté externe de l'Hexon. On remarque aussi sur les deux figures la position de l'extrémité N terminale de la protéine IX en jaune.

Chez le virus WT, la densité est orientée de façon verticale et réalise un contact avec la partie basse des Hexons du côté gauche et du côté droit. La protéine est assez allongée et possède des extrémités internes et externes plus fines que sa partie intermédiaire. Chez le virus mutant HAdV-5\_SY12, la protéine est orientée de façon plus horizontale. Ceci a pour conséquence l'absence de contacts avec la partie basse des deux Hexons qui l'encadrent. En effet, la zone la plus interne de contact avec les Hexons observée dans la capsid du virus WT n'existe plus chez ce virus mutant. La densité forme maintenant un cylindre où les extrémités sont de même largeur que la zone intermédiaire.

Nous n'avons pas détecté de modifications même mineures de la densité précédemment attribuée à la protéine IX (en jaune sur la [Figure 48](#)). La seule modification externe de la capsidé observée se situe au niveau de la densité précédemment attribuée à la protéine IIIa. Cette observation vient confirmer ce qui a été observée par Saban et collaborateurs en 2006 ([Saban et al., 2006](#)). Les auteurs ont attribué la densité de la protéine IIIa à l'extrémité C terminale de la protéine IX. En effet, les auteurs arrivent à replacer dans cette densité jusqu'à quatre hélices alpha correspondant aux extrémités C-terminale de quatre protéines IX. Les données que nous avons obtenues semblent confirmer cette observation. La seule modification protéique existante entre le virus WT et la souche SY12 est l'addition d'un peptide de 12 acides aminés attachés directement à l'extrémité C terminale de la protéine IX. La présence de ce peptide viendrait donc modifier les contacts que la protéine IX peut réaliser avec les Hexons adjacents. Les 2 extrémités de la densité rose ([Figure 49](#)) sont modifiées dans notre reconstruction ce qui laisse à penser qu'au moins un peptide de 12 acides aminés est présent au niveau de chaque extrémité. On peut donc dire que la densité est composée d'au moins deux extrémités C terminale positionnées en tête bêche. Cela confirme que la région en jaune ([Figure 48](#)) ne constituerait que les extrémités N terminales de la protéine. Cette modification de localisation de la protéine IX explique aussi l'absence totale de densités externes que nous avions observées au niveau de la capsidé du virus HAdV-5 où le gène de la protéine IX avait été supprimé. Cette hypothèse semble plus simple que celle proposée en 2005 ([Article 1](#)) où la protéine IIIa se serait détachée après déstabilisation de la capsidé en absence de protéine IX. Nous avons voulu néanmoins réaliser un marquage anticorps au niveau du peptide SY12 afin de confirmer les résultats obtenus.

*Le modèle 3D du virus HAdV-5 SY12 en complexe avec un Fab anti-SY12.* 11 micrographies de bonne qualité (sans astigmatisme, ni dérive) ont permis de réaliser la sélection manuelle de 2548 particules dont 949 ont permis de calculer une reconstruction 3D à 22 Å de résolution ( $FSC_{0,5}$ ) du virus HAdV-5\_SY12 en complexe avec le fragment de Fab anti-SY12. J'ai volontairement choisi de me placer à  $FSC_{0,5}$  car la courbe ne permet de déterminer de façon précise la résolution à  $FSC_{0,3}$  ([Annexe 4](#)). Avec 949 particules et l'utilisation d'un microscope FEG, nous pensions obtenir une reconstruction 3D à une résolution un peu plus importante. La présence des fragments Fab attachés à la surface du virus est responsable d'une augmentation du bruit autour des capsides dans les images. Ce phénomène accroît la difficulté à orienter les particules correctement à l'aide du

programme PFT. Nous avons donc utilisé assez rapidement PFT2 pour orienter les particules ce qui nous a permis d'obtenir un modèle 3D à moyenne résolution. Le modèle 3D calculé pour le virus en complexe avec le Fab semble complet. Toutes les protéines majeures sont présentes dans la capsid. Les Pentons sous forme complète sont observés au niveau des sommets et les Hexons forment bien le reste de la capsid (Figure 50). La forme globale de la capsid est conservée par rapport au contrôle sans marquage anticorps. La partie N terminale de la protéine IX est bien présente au niveau du GON (en jaune sur la Figure 50). Elle forme de la même manière que chez les autres virus humains que nous avons étudiés un triskelion stabilisant les Hexons au sein d'une face.



**Figure 50: Comparaison des reconstructions 3D obtenues pour le HAdV-5\_SY12 en présence et en absence d'anticorps.**  
 (a) Vue de dessus d'une face de la reconstruction obtenue pour le virus SY12 en complexe avec un Fab anti-SY12. La densité où se fixe le Fab est colorée en vert. Si le niveau de contour choisi est le même qu'en b, la densité verte est peu visible. (b) Vue de dessus d'une face de la reconstruction du même virus en absence de marquage par anticorps. Sur les 2 figures, les extrémités N-terminales des protéines IX sont colorées en jaune et les Pentons en bleu.

On peut tout de même remarquer une différence majeure au niveau de la surface du virus. La densité rose observée sur la reconstruction contrôle a disparu. Il est maintenant nécessaire de jouer sur le niveau de contour si l'on veut observer une densité dans cette zone (en vert sur la Figure 50). Lorsqu'on fixe le niveau de contour à une valeur faisant apparaître les densités moins fortes, on observe maintenant une région allongée assez fine qui se prolonge au niveau de ses deux extrémités par une zone large de densité assez diffuse (En vert sur la Figure 50). Au niveau des extrémités, les zones de densité diffuse sembleraient correspondre à l'attachement du Fab au niveau du peptide SY12. La présence des fragments Fab au niveau de l'extrémité C terminale de la protéine IX déstabilisent complètement la densité observée au niveau du virus contrôle. La présence de densités correspondant au Fab au niveau des deux extrémités confirme la présence de peptides SY12 de

chaque coté et donc le placement en tête bêche d'au moins deux extrémités C terminales de protéines IX dans cette zone mais il nous est difficile de conclure définitivement sur le nombre d'extrémités C terminales qui composent la densité précédemment attribuée à la protéine IIIa. La fixation des Fab au niveau du peptide dans cette zone semble entraîner la déstabilisation complète des hélices alpha présentes au niveau de l'extrémité C terminale de la protéine IX. La fixation des Fab provoque aussi une flexibilité importante dans cette région de la capsidne nous permettant pas de visualiser plus précisément l'anticorps à ce niveau. La forme que nous observons pour l'extrémité C terminale de la protéine IX attachée au Fab est le résultat de la moyenne réalisée lors du processus d'analyse d'images et de reconstruction 3D entre des capsides dépourvues d'anticorps et des capsides où l'anticorps est attaché. Ceci a pour conséquence l'apparition d'une densité plus faible possédant en partie la forme et l'orientation de la densité de départ observée dans le virus contrôle. Nous aurions aimé réaliser une étude plus approfondie du contact entre le Fab et le modèle quasi-atomique de la capsidne entière par « fitting » de la structure atomique d'un Fab mais nous avons obtenu des résultats convaincants quant à la localisation de l'extrémité C terminale de la protéine IX chez les Adénovirus Humains.

### V.3 Conclusion

Après les résultats obtenus concernant la localisation de l'extrémité C terminale de la protéine IX chez l'Adénovirus Canin de type 2, nous avons apporté ici une preuve structurale de sa localisation chez les Adénovirus humains. Nous avons donc isolé et purifié un fragment Fab reconnaissant le peptide SY12 présent à l'extrémité C terminale de la protéine IX chez une souche modifiée de HAdV-5. Nous avons ensuite étudié par CryoME et analyse d'images l'échantillon viral en absence et en présence de Fab à sa surface. Cette technique nous a permis de reconstituer un modèle 3D de la capsidne en absence et en présence du fragment Fab à sa surface. L'analyse de ces deux modèles permet de conclure définitivement quant à la localisation d'au moins deux extrémités C terminales de protéines IX positionnées en tête bêche au niveau des densités qui avait été précédemment attribuée à la protéine IIIa. La présence des extrémités C terminales de la protéine IX au niveau de la densité attribuée à la protéine IIIa avait été introduite par Saban et collaborateurs en 2006. Notre étude structurale conforte l'hypothèse avancée par Saban et collaborateurs où la densité attribuée à la protéine IIIa ne serait donc qu'une partie de la protéine IX mais celle-ci ne nous permet pas de conclure définitivement sur le nombre de protéines IX qui composent cette densité.

## ***VI. Etude structurale d'un mutant immature de l'Adénovirus Humain de type 2***

Durant ma thèse, j'ai essentiellement travaillé sur des souches d'Adénovirus matures afin de comprendre l'assemblage des protéines de capsid chez le virus actif. Nous voulions par la suite nous intéresser à des souches virales immatures ou partiellement maturées afin d'étudier la maturation des Adénovirus. La seule souche connue d'Adénovirus humain bloquée à un stade intermédiaire de maturation est une souche appelée *ts1*. Cette souche est très peu étudiée et aucune caractérisation biochimique ni structurale n'a encore été réalisée et publiée. Nous avons mené une étude structurale en MET en collaboration avec le Dr Urs F. Greber (Zurich, Suisse) qui a réalisé avec son équipe une étude préliminaire sur ce virus. La souche *ts1* est dérivée du HAdV-2. Elle n'est pas biologiquement active. En effet, ce virus possède une mutation thermosensible au sein du gène codant pour la protéase qui conduit à son inactivation. D'après les travaux préliminaires réalisés par l'équipe du Dr. Greber, l'inactivation de la protéase donne lieu à la production de capsides immatures qui semblent contenir la protéase sous une forme inactive. Cette souche d'Adénovirus semble capable de pénétrer la cellule mais échoue à sortir de l'endosome lors de la phase précoce d'infection. A travers l'étude structurale en MET de la souche *ts1*, nous voulions déterminer les zones de la capsid qui subissent une maturation par protéolyse ainsi qu'identifier la localisation potentielle de la protéase virale au sein de la capsid des Adénovirus. Dans cette étude, je me suis plus particulièrement intéressée au processus de reconstruction 3D et de reconstitution d'un modèle quasi-atomique.

### **VI.1 Matériel et Méthodes**

Un échantillon purifié de virus mutant *ts1* nous a été fourni par le Dr. U.F. Greber. La solution concentrée de virus est conservée à – 80°C dans un tampon 20 mM Tris ph 7,5, contenant 10% de glycérol et 100 mM NaCl. L'excès de glycérol a été retiré par centrifugation à l'aide d'un Microbospin 6 (Biorad).

#### ***VI. 1.1 La CryoME***

Cet échantillon a été étudié par CryoME sur le microscope JEOL FEG de l'IBS à un grossissement nominal de 30000 par le Dr Guy Schoehn. La préparation de l'échantillon et l'analyse d'images ont été réalisées conformément au protocole établi pour la souche WT du HAdV-5 (Cf. II). Des images sélectionnées pour leur qualité (absence d'astigmatisme et de dérive) ont été scannées avec un pas de 7 µm et ont permis d'extraire un total de 2916 particules. Après extraction, les

particules sont placées dans une boîte de 447 pixels par 447 pixels ce qui correspond à une taille nominale de pixels de 2,33 Å. La CTF a été corrigée jusqu'à 15 Å pour le jeu d'images « flippées » et à 7 Å pour le jeu d'images « déconvolées ». Nous avons utilisé pour la recherche des orientations et des centres, PFT/EM3DR dans un premier temps puis PFT2/EM3DR2 afin d'améliorer la résolution finale de la reconstruction 3D obtenue.

### VI. 1.2 Le « fitting » en corps rigide

Le recalage des structures atomiques de la Base du Penton et de l'Hexon a été réalisé avec URO. Les structures atomiques utilisées sont celle de la Base du Penton en complexe avec le peptide N terminal de la fibre (Code PDB : 1X9T) et celle de l'Hexon du HAdV-5 (Code PDB : 1P30). Un premier placement manuel a été réalisé à partir du modèle quasi-atomique du HAdV-5. Par la suite, la procédure de fitting a permis d'ajuster l'échelle de la carte de ME. L'échelle a été déterminée à 1% de moins ce qui correspond à une taille de pixel de 2,30 Å. Les critères de qualité du « fitting » sont calculés après masquage des extra-densités liées aux protéines mineures ou aux boucles absentes des structures cristallographiques des protéines majeures de capsid.

## VI.2 Résultats et discussion

La CryoME A partir de 878 particules parmi un total de 2916, nous avons pu reconstituer un modèle 3D à 12 Å de résolution ( $FSC_{0,3}$ ) de la souche *ts1* dérivée du HAdV-2 (Figure 51a). Le graphique correspondant des FSC se trouve en annexe 5.

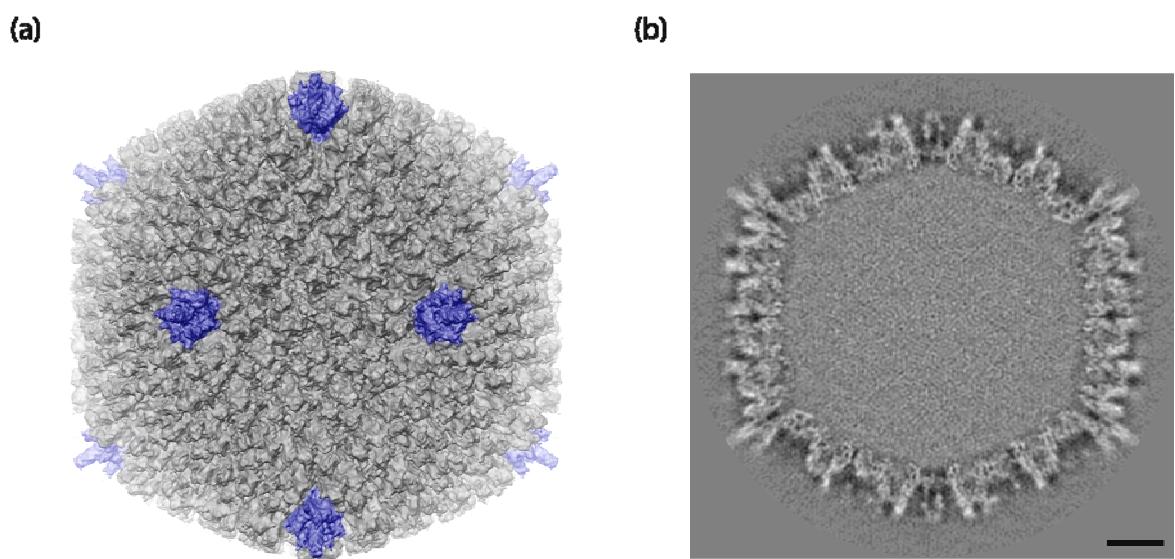
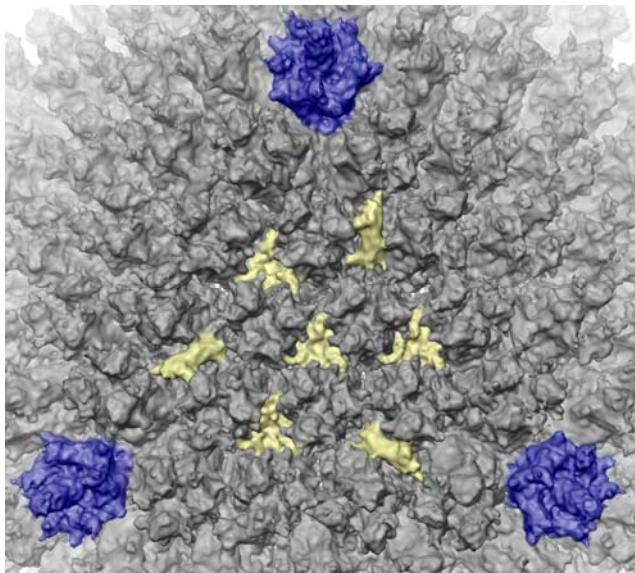


Figure 51: Modèle 3D en CryoME à 12 Å de la souche mutante du HAdV-2 : le *ts1*.

(a) Reconstruction 3D finale de la capsid du mutant *ts1* vue selon l'axe 2 de l'icosaèdre. (b) Vue transversale en coupe de la densité (réalisé avec SPIDER/JWEB). La barre d'échelle sur la figure b représente 100 nm.

On peut voir sur l'image en isosurface du virus ainsi que sur une coupe transversale que la densité est très détaillée (Figure 51). Un tel niveau de détails rend l'observation et l'analyse du modèle particulièrement difficile surtout sur une vue complète de la capsidé (Figure 51), c'est pourquoi j'ai réalisé un zoom sur une face en Figure 52. La capsidé du *ts1* semble posséder les mêmes caractéristiques externes qu'une souche mature d'Adénovirus. Aucune différence majeure n'est observée au niveau des protéines majeures ni des protéines mineures. Les protéines majeures montrent un fort niveau de détails mais la structure et le positionnement global sont conservés chez le mutant *ts1*. Les structures de la Base du Penton et des Hexons sont en accord avec les précédentes reconstructions (Figure 52).

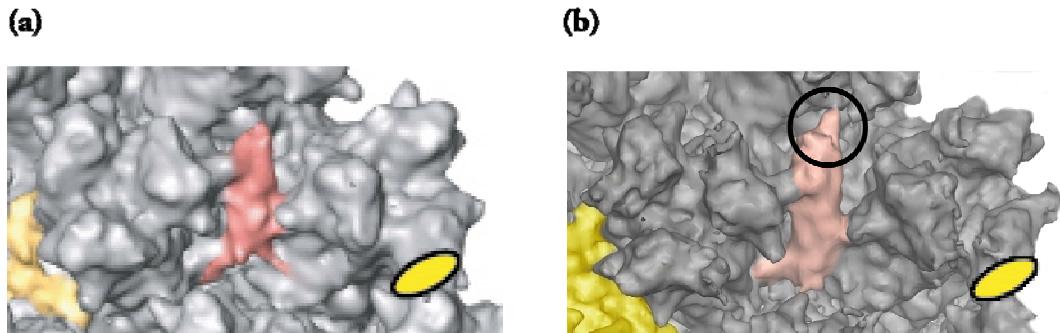


*Figure 52: Zoom sur une face de la Reconstruction 3D du mutant HAdV-2 ts1.*

Vue selon l'axe d'ordre 3 présent au centre de la face. Les Pentons sont représentés en bleu et les densités correspondantes aux extrémités N et C terminales de la protéine IX en jaune.

Les protéines mineures du coté externe de la capsidé ne montre pas non plus de différences majeures avec une souche mature d'Adénovirus (Figure 52). Nous n'observons pas d'extra densités chez le *ts1* du coté externe de la capsidé pouvant être attribuées à une zone non maturée de la capsidé. La densité attribuée à l'extrémité N terminale de la protéine IX forme, de la même manière que dans une capsidé mature, quatre trimères au niveau des GONs. La partie C terminale est aussi présente au niveau de l'extérieur des GONs (en jaune sur la Figure 52). Les contacts réalisés par la protéine IX avec les Hexons adjacents semblent aussi conservés. La seule différence représentative est localisée au niveau de la partie externe du C terminale de la protéine IX. En effet, cette zone semble plus allongée chez le mutant *ts1* (entourée sur la Figure 53). Cette différence est difficile à interpréter en raison de la résolution de la reconstruction ainsi que de l'utilisation d'un autre MET pour la prise d'images mais elle pourrait tout de même provenir d'une zone non maturée de la protéine. En effet, la taille du précurseur immature de la protéine IX est connue mais personne ne sait réellement

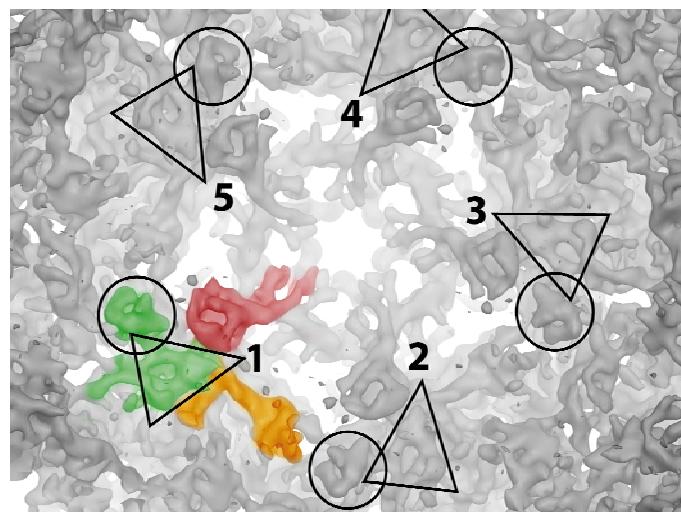
quelles parties sont incorporées dans la capsidé ni quelles zones subissent une maturation par la protéase virale.



**Figure 53: Comparaison des densités attribuées à l'extrémité C terminale de la protéine IX.**

(a) Densité attribuée à la partie C terminale de la protéine IX dans la reconstruction du HAdV-5 WT en rose (cf. Article 1). (b) Même densité dans la reconstruction du *ts1*, en rose. L'ellipse jaune marque la position d'un axe icosaédrique d'ordre 2. Le cercle montre la densité supplémentaire présente à l'extrémité la plus externe.

Dans l'ensemble, la maturation par la protéase virale ne semble pas affecter les contacts entre protéines ni avoir une grande influence sur la partie externe de la capsidé, étant donné le peu de différences observées par rapport à un virus mature. Ce n'est pas le cas de la partie interne de la capsidé (Figure 54).



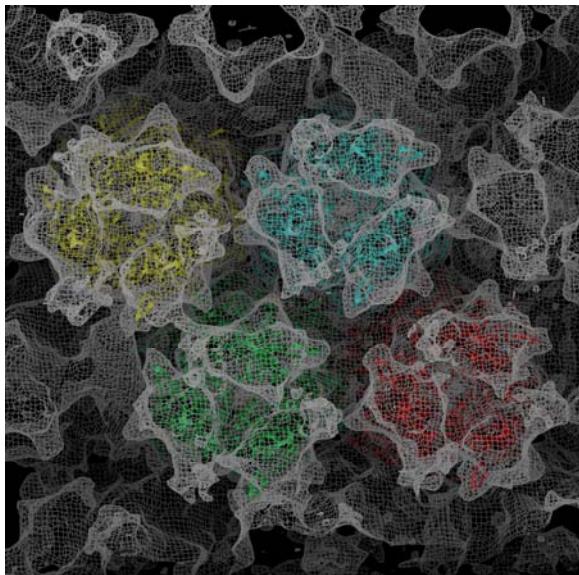
**Figure 54: Vue selon l'axe 5 de l'icosaèdre de la zone située en dessous des Pentons.**

Densités supplémentaires chez le *ts1* observées depuis l'intérieur du virus en vert. La protéine IIIa et la protéine VIII sont respectivement colorées en rouge et orange. Les numéros indiquent la présence de l'axe 5 de l'icosaèdre et marquent la répétition de cinq complexes identiques à celui présenté en couleurs.

En effet, la zone interne qui subit le plus de modifications se situe en dessous des Pentons. Cette zone semble clairement affectée par l'absence de maturation par protéolyse. Des densités supplémentaires sont visibles au contact des protéines IIIa et VIII à l'intérieur de la capsidé par rapport à la capsidé du HAdV-5 (Figure 54). Même si une reconstruction de si haute résolution reste difficile à analyser, on peut remarquer la présence d'une densité de forme globulaire en contact avec une autre densité supplémentaire de forme plutôt triangulaire (en vert sur la Figure 54). La zone triangulaire vient se placer à l'interface entre la protéine IIIa et VIII alors que la partie plus globulaire

se situe plus vers l'intérieur de la capsidé uniquement attachée à la densité triangulaire. En absence de caractérisations biologiques, la taille des protéines IIIa et VIII dans une capsidé à l'état mature n'étant pas connue, il est difficile de conclure sur l'origine de ces densités surnuméraires mais plusieurs causes peuvent être à leur origine. L'équipe du Dr. Greber a montré la présence sous une forme inactive de la protéase virale à l'intérieur de la capsidé. Les densités observées pourraient soit correspondre à la présence de la protéase virale soit à des parties non maturées des protéines IIIa et VIII. Cette souche mutante de HAdV-2 échoue à sortir de l'endosome probablement parce que la protéase n'est pas capable de dégrader la protéine VI qui est responsable de l'activité lytique des membranes lors de la phase précoce d'infection. On s'attendrait, dans ces conditions, à trouver la protéase au voisinage de la protéine VI. La localisation de la protéine VI étant très controversée depuis la publication des travaux de Saban et al., 2006, il est difficile d'aller plus loin dans l'interprétation.

Le fitting Un modèle quasi-atomique a été calculé pour le mutant ts1. « Le fitting » en corps rigide a pu être réalisé avec le programme URO (Navaza et al., 2002) pour toutes les protéines majeures : les Hexons de l'UA et la Base du Penton. Les critères de qualité obtenus en dernière phase d'affinement sont, pour les Hexons de l'UA : CC= 88.9, R=38.2, Q=10.7 et pour la Base du Penton : CC=89.6, R=38.8, Q=10.0.



*Figure 55 : Vue de dessus d'une UA du modèle quasi-atomique à travers la densité correspondante provenant de la reconstruction 3D du mutant ts1.*

La structure cristallographique des Hexons 1 à 4 de l'UA sont colorés respectivement, en jaune, cyan, vert et rouge.

Les valeurs obtenues pour les différents paramètres d'évaluation de « fitting » sont très bonnes même si elles restent légèrement inférieures à celle obtenues pour le modèle quasi-atomique du HAdV-5. Le placement de la Base du Penton et des Hexons est réalisé de façon satisfaisante. Il

n'existe pas de différences majeures que ce soit au niveau de la position des protéines majeures ou au niveau de leur orientation entre le modèle quasi-atomique du mutant *ts1* et celui du HAdV-5. On peut expliquer la légère diminution des critères de qualité par la résolution de la reconstruction 3D du *ts1*. Plus la résolution du modèle 3D obtenu en ME est élevée, plus la différence entre la structure cristallographique d'une protéine et ce modèle est marquée. Lorsqu'on observe un Hexon replacé au sein de la densité lui correspondant, on remarque que certaines zones riches en structures secondaires ne se placent pas parfaitement dans la densité. Il faudrait commencer à déformer la structure cristallographique des Hexons et de la Base du Penton par les techniques de « fitting » flexibles pour obtenir un meilleur accord avec la carte de ME. On peut aussi expliquer les résultats obtenus par l'utilisation de structures cristallographiques de protéines majeures maturées que l'on tente de replacer au sein de la carte obtenue à partir d'un virus immature. Le déplacement de certaines zones dans l'Hexon ou la Base du Penton peut être expliqué par l'absence de maturation de la capsid du *ts1*. Dans notre étude, le modèle quasi-atomique du *ts1* a surtout été utilisé pour calculer une carte de différence et éliminer la densité des protéines majeures dans le modèle 3D obtenu en MET facilitant ainsi l'observation et l'analyse de la capsid du *ts1*.

### VI.3 Conclusion

Nous avons réalisé une étude structurale en CryoME d'une souche immature dérivée du HAdV-2 : le *ts1*. Nous avons pu obtenir une reconstruction 3D à haute résolution de ce virus ainsi qu'un modèle quasi-atomique associé. Le modèle 3D, comparé avec celui de la capsid mature du HAdV-2 et 5, montre des différences structurales au niveau de la zone interne de la capsid. Nous avons pu identifier clairement la présence de densités surnuméraires chez le mutant *ts1* en dessous du Penton. Plusieurs hypothèses peuvent expliquer l'origine de ces densités mais en absence de caractérisation biochimique plus complète du *ts1*, il est difficile de conclure sur leur origine. Une caractérisation biochimique et structurale de ce mutant immature d'Adénovirus humain permettrait de comprendre le processus de maturation de la capsid ainsi que son désassemblage lors de la phase précoce d'infection. De plus, un modèle quasi-atomique plus ajusté pourrait être calculé par les techniques de « fitting » flexible basées sur l'utilisation des modes normaux mais ceci doit être réalisé en collaboration étroite avec une équipe qui met au point les programmes utilisés. Le volume important des données manipulées nécessitent un gros travail de développement informatique. Si l'on veut avancer dans la compréhension complète du cycle cellulaire des Adénovirus, il est important de concentrer les efforts de caractérisation sur des souches immatures comme cela est déjà fait pour une grande partie des virus.

## VII. Le modèle quasi atomique du dodécaèdre de l'Adénovirus Humain de type 3

Cette partie concerne l'étude en MET de la structure du dodécaèdre, une sous-particule virale dérivée de l'Adénovirus humain de type 3 (HAdV-3) composée de la Base du Penton et de la Fibre. Cette particule partage une caractéristique majeure avec le virus entier dont elle est issue : elle peut s'attacher et pénétrer à l'intérieur d'une cellule. Le dodécaèdre a, dans ce cadre, été proposé comme vecteur thérapeutique. La cavité présente au sein de la particule est accessible au solvant et peut transporter des petites molécules. Pour mettre au point un vecteur dérivé du dodécaèdre, il est nécessaire d'avoir une bonne connaissance de sa structure et de ces mécanismes d'assemblage/désassemblage afin d'y introduire de potentiels modifications. Deux types de particules sont produits de façon recombinante par le Dr. Pascal Fender. Une première particule, le dodécaèdre Base, ne contient que les Bases du Penton alors que la deuxième particule est formée à partir de la Base et de la fibre. Les images en CryoME et la reconstruction 3D des dodécaèdres Base et Fibre ont été réalisé par le Dr. Guy Schoehn. Dans cette étude, je me suis intéressée au « fitting » des structures atomiques de la base du Penton de l'Adénovirus de type 2, seule ou en complexe avec le peptide N terminal de la Fibre, à l'intérieur des reconstructions 3D obtenues en MET.

### VII.1 Article 3

#### “Structure of the dodecahedral penton particle from human adenovirus type 3”

Auteurs : P. Fuschiotti, G. Schoehn, P. Fender, C.M.S. Fabry, E.A. Hewat, J. Chroboczek, R.W.H. Ruigrok & J.F. Conway

Publié en 2006 dans *Journal of Molecular Biology*, Vol. 356, pages 510-520.

# Structure of the Dodecahedral Penton Particle from Human Adenovirus Type 3

**P. Fuschiotti<sup>1</sup>, G. Schoehn<sup>2</sup>, P. Fender<sup>3</sup>, C. M. S. Fabry<sup>2</sup>, E. A. Hewat<sup>1</sup>  
J. Chroboczek<sup>4</sup>, R. W. H. Ruigrok<sup>2</sup> and J. F. Conway<sup>1\*</sup>**

<sup>1</sup>*Laboratoire de Microscopie Electronique Structurale  
Institut de Biologie Structurale  
UMR 5075 CNRS-CEA-UJF  
38027 Grenoble cedex, France*

<sup>2</sup>*Institut de Virologie  
Moléculaire et Structurale  
FRE 2854 CNRS-UJF, BP181  
38042 Grenoble cedex 9, France*

<sup>3</sup>*Laboratoire d'Enzymologie  
Moléculaire, Institut de Biologie  
Structurale, UMR 5075  
CNRS-CEA-UJF, 38027  
Grenoble cedex, France*

<sup>4</sup>*Laboratoire de Biophysique  
Moléculaire, Institut de Biologie  
Structurale, UMR 5075  
CNRS-CEA-UJF, 38027  
Grenoble cedex, France*

The sub-viral dodecahedral particle of human adenovirus type 3, composed of the viral penton base and fiber proteins, shares an important characteristic of the entire virus: it can attach to cells and penetrate them. Structure determination of the fiberless dodecahedron by cryo-electron microscopy to 9 Å resolution reveals tightly bound pentamer subunits, with only minimal interfaces between penton bases stabilizing the fragile dodecahedron. The internal cavity of the dodecahedron is ~80 Å in diameter, and the interior surface is accessible to solvent through perforations of ~20 Å diameter between the pentamer towers. We observe weak density beneath pentamers that we attribute to a penton base peptide including residues 38–48. The intact amino-terminal domain appears to interfere with pentamer–pentamer interactions and its absence by mutation or proteolysis is essential for dodecamer assembly. Differences between the 9 Å dodecahedron structure and the adenovirus serotype 2 (Ad2) crystallographic model correlate closely with differences in sequence. The 3D structure of the dodecahedron including fibers at 16 Å resolution reveals extra density on the top of the penton base that can be attributed to the fiber N terminus. The fiber itself exhibits striations that correlate with features of the atomic structure of the partial Ad2 fiber and that represent a repeat motif present in the amino acid sequence. These new observations offer important insights into particle assembly and stability, as well as the practicality of using the dodecahedron in targeted drug delivery. The structural work provides a sound basis for manipulating the properties of this particle and thereby enhancing its value for such therapeutic use.

© 2005 Elsevier Ltd. All rights reserved.

**Keywords:** adenovirus; dodecahedron; penton; fiber; cryo-electron microscopy

\*Corresponding author

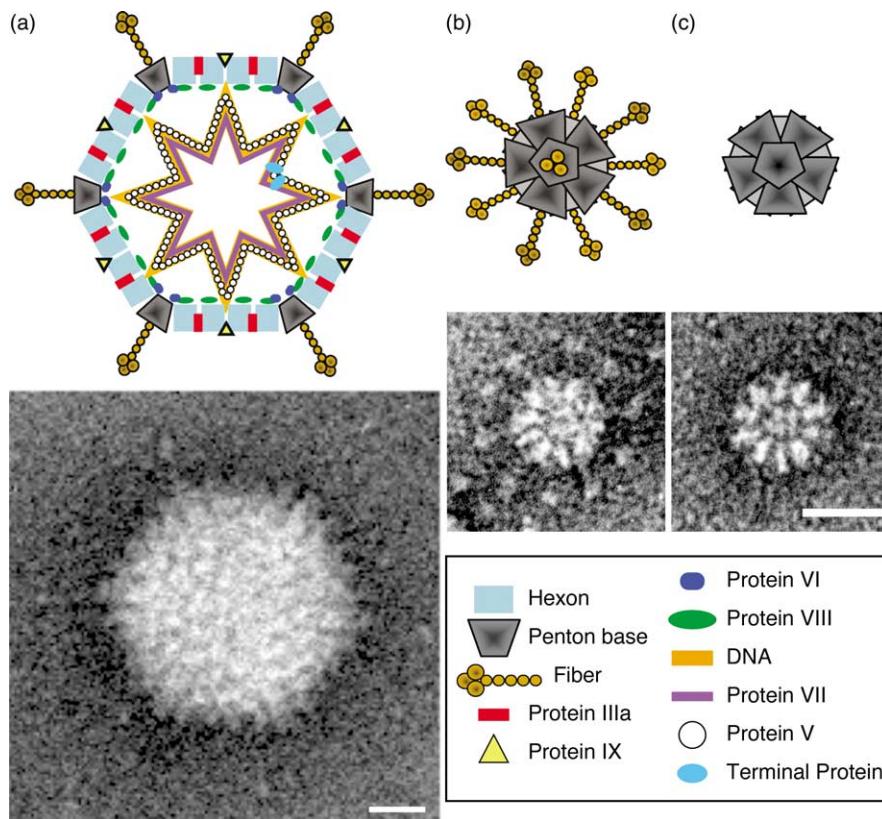
Present addresses: P. Fuschiotti, Department of Immunology, University of Pittsburgh, Room E1040, Biomedical Science Tower, 200 Lothrop Street, Pittsburgh, PA 15261, USA; J. F. Conway, Department of Structural Biology, University of Pittsburgh, Room 2047, Biomedical Science Tower 3, 3501 5th Ave, Pittsburgh, PA 15260, USA.

Abbreviations used: Ad2, Ad3, Ad5, adenovirus serotypes 2, 3 and 5, respectively; Bs-Dd, dodecahedra constructed from adenovirus penton base protein alone; Pt-Dd, dodecahedra of pentons, i.e. penton base and fiber proteins; EM, electron microscopy; cryoEM, cryo-electron microscopy; Mab, monoclonal antibody.

E-mail address of the corresponding author:  
jxc100@pitt.edu

## Introduction

Adenoviruses are a large family of non-enveloped DNA viruses that infect a wide range of human and animal cell types, most commonly causing respiratory disease but also gastroenteritis or conjunctivitis. Their characteristic morphology includes an icosahedral capsid with a diameter of 900–1000 Å, pentameric capsomers (penton bases) at the vertices and hexavalent trimers (hexons) populating the facets in a pseudo- $T=25$  arrangement.<sup>1</sup> A trimeric fiber protein extends 100–300 Å (according to serotype) from the outer capsid surface and is attached non-covalently to each penton base forming the penton. Several other minor structural proteins are involved in stabilizing



**Figure 1.** Schematics of adenovirus and its dodecahedron. (a) Diagram of adenovirus, including the major structural proteins: hexon (light blue), penton base (grey) and the trimeric fiber (gold) that extends outwards ending in a knob. Additional minor capsid proteins include the hexon “cementing” proteins IIIa (red bars), IX (yellow triangles), protein VIII associated with the hexons (green oval) and protein VI associated with the penton base (blue ovals). DNA is represented as an extended string (although in reality it is more compact than indicated here) and coated with associated proteins. Beneath is a negative-stain EM image of a viral particle. (b) Representation of the sub-viral dodecahedral particle with fiber bound (Pt-Dd) viewed down a 5-fold axis and a negative stain EM image from expressed penton base and fiber proteins where some fibers are visible, particularly the distal knobs. (b) The dodecahedral particle without fiber (Bs-Dd) and corresponding EM image. The bars represent 200 Å.

the capsid (Figure 1). The distal knob-domain of the fiber is necessary for cell recognition. Most subgroups (A–F, excluding B) bind the Coxsackie and adenovirus receptor (CAR),<sup>2,3</sup> an integral membrane protein of the tight junction in epithelial cells.<sup>4</sup> The interaction between the fiber and CAR is implicated both in viral entry and egress.<sup>5</sup> Several subgroup B adeno-viruses utilize another cell surface receptor, CD46,<sup>6,7</sup> but the primary receptor for another subgroup B member, adenovirus serotype 3 (Ad3), is still not known. Cell entry is gained by endocytosis, in most serotypes through RGD-loops extending from the penton base that bind to the  $\alpha\beta\beta 3$  or  $\beta 5$  integrins.<sup>8,9</sup> The penton base also has a role in release from endosomes and decapsidation.<sup>10,11</sup>

Models of the adenovirus capsid have been built by combining available crystallographic structures of the capsomers with lower resolution density maps of the entire capsid from cryo-electron microscopy (cryoEM). A model of the subgroup C Ad2 capsid was partially built from the trimeric hexon<sup>12</sup> and capsid reconstruction.<sup>1</sup> This model accounted for ~60% of the protein in the virion and offered insights into hexon–hexon and hexon–

penton base interfaces, as well as indicating potential locations for some of the minor proteins. More recent updates include cryo-electron microscopy density maps of the Ad2 and Ad12 capsids to 21 Å resolution where surface features such as the RGD receptor binding loops were localized through binding MAbs or small receptor fragments.<sup>13,14</sup> However, much detail was lacking, including the structure of the penton base, its interaction with the fiber and the role of the symmetry-mismatch between these two components. Our recent cryoEM study on the Ad5 capsid has extended the resolution to 10 Å, allowing more detailed modeling of the capsomer interfaces and localization of minor structural proteins by comparison of native and deletion mutant capsids.<sup>15</sup>

For certain adenovirus serotypes, penton bases can self-assemble into a sub-viral particle called dodecahedron that comprises 12 copies of the penton arranged on a  $T=1$  icosahedral lattice<sup>16,17</sup> and may include fiber molecules (penton dodecahedron, Pt-Dd) or not (base dodecahedron, Bs-Dd) (Figure 1). Analysis of the dodecahedral particle of Ad3 by cryoEM to a resolution of ~30 Å revealed

the overall shape of the particle and allowed an estimation of the dimensions of the internal cavity.<sup>18</sup> Comparison between particles including the fiber molecule (Pt-Dd) and those without (Bs-Dd) revealed apparent redistribution of density at the outermost surface of the penton at the presumed location of the mobile integrin-binding RGD loops. The fiber itself, however, was poorly resolved, possibly as a consequence of intrinsic flexibility, limited number of particles included in the reconstruction (30) or from bending imposed by a restricted depth of ice formed during preparation for microscopy. Subsequently, the dodecahedron was shown to enter cells efficiently using the same mechanism as adenovirus,<sup>19,20</sup> confirming the localization of the infection mechanisms to the penton's two constituent proteins, penton base and fiber. We have previously proposed the dodecahedron as an alternative to adenovirus for therapeutic delivery.<sup>19,20</sup> One of the potential advantages over the intact virus particle is that it cannot provoke infection directly or through recombination, as it harbors none of the viral genome. A second is the possibility of a weaker immune response compared to adenovirus because of its reduced protein content, thereby enhancing the success of multiple applications. In addition, recent work has shown that dodecahedra can penetrate a wider range of cells than the virus by entering through a heparin sulfate pathway.<sup>21</sup> However, the internal cavity of the dodecahedron is too small to accommodate more than ~100 bp of DNA and is thus too small to carry a gene, although it could accommodate drugs. Specific targeting to non-native receptors is possible through modification of the cell-recognition domains on the fiber, or combining fibers and penton bases from different adenovirus serotypes.<sup>22</sup>

Recently, a crystallographic model of the Ad2 penton base was reported. An expressed 48 N-terminal truncated form of the penton exposed to crystallization conditions lead to formation of dodecahedral particles in the crystal.<sup>23</sup> Interestingly, the full-length Ad2 penton base did not spontaneously form dodecahedra. The core of the structure is a jellyroll motif, an eight-stranded antiparallel  $\beta$ -barrel that is common to capsid proteins of certain virus families, with an extensive insertion of 308 residues (54% of the protein) between two  $\beta$ -strands of the jellyroll, and a second insertion of 54 residues between another pair of strands. The first insertion includes two external loops, one of which (the RGD loop) contains the integrin-binding RGD domain and is highly variable between serotypes in sequence and length. It is also flexible and none of the 74 residues of the loop were modeled in the crystal structure. The second loop (variable loop) is shorter, also variable between serotypes, and projects from the outer surface of the penton base into solvent. The binding pattern of the Ad2 fiber was revealed in a second dodecahedron structure including residues 11–21 from the N-terminal part of the fiber protein as five monomers bound to the penton base, albeit

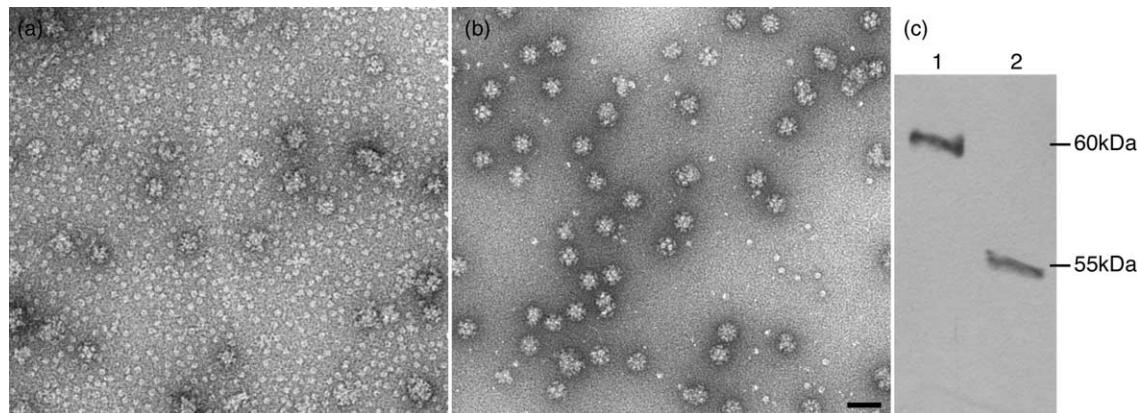
in non-stoichiometric quantities since the fiber molecule is a trimer.

The available adenovirus penton base protein sequences are highly conserved outside these external loops, suggesting that the bulk of the penton base has a common fold. The Ad2 penton protein differs in sequence from that of Ad3 at the amino terminus and at the two external loops, where the Ad3 protein is some 34 residues shorter in the RGD loop and nine residues longer in the variable loop. These three variable regions are perhaps the most important outside of the conserved interface residues that direct subunit and penton–hexon organization. The amino terminus is located internally to the capsid<sup>15</sup> and holds two conserved motifs of the form PPxY that may potentially interact with host cell ubiquitine ligases containing WW domains<sup>24</sup> presumably before incorporation into the capsid or after partial decapsidation. The RGD loop is a co-receptor binding site involved in endocytosis, and the second loop has unknown function but is clearly serotype specific.

Here we report on our work extending the structural characterization of the Ad3 dodecahedron by cryo-electron microscopy, including a more detailed visualization of the fiber and its interface with, and effects on, the penton base. We model the cryoEM density with the atomic coordinates of the Ad2 dodecahedron to compare the core structures and to identify the location of residues unique to the Ad3 penton sequence as well as part of the 48 amino-terminal residues not visualized by crystallography.

## Results and Discussion

Although we published previously that dodecameric particles consisting of Ad3 penton base or Ad3 penton base plus fiber formed spontaneously upon expression,<sup>18,20</sup> new purification methods of recombinant protein produced in insect cells led to assembly of few particles, some incomplete, and with a much higher proportion of free penton bases (Figure 2(a)). We explain this discrepancy by first noting that the atomic structure of the Ad2 penton base dodecahedron was derived from protein that was N-terminally truncated, with the first 48 amino acid residues missing.<sup>23</sup> The internal space in this dodecahedron may be approximated by a sphere with diameter of 80 Å, sufficient to accommodate 225 kDa of tightly packed protein but which is too small to accommodate the 317 kDa of truncated peptides (60 copies of 48 residues). Clearly, dodecahedra cannot assemble from full-length penton base protein. Secondly, we observed that when full-length Ad3 penton base (Figure 2(c), lane 1) was kept for two weeks at room temperature, almost all the protein had assembled into dodecahedra (Figure 2(b)) and that the protein had been spontaneously degraded to a smaller size (Figure 2(c), lane 2). N-terminal sequencing showed

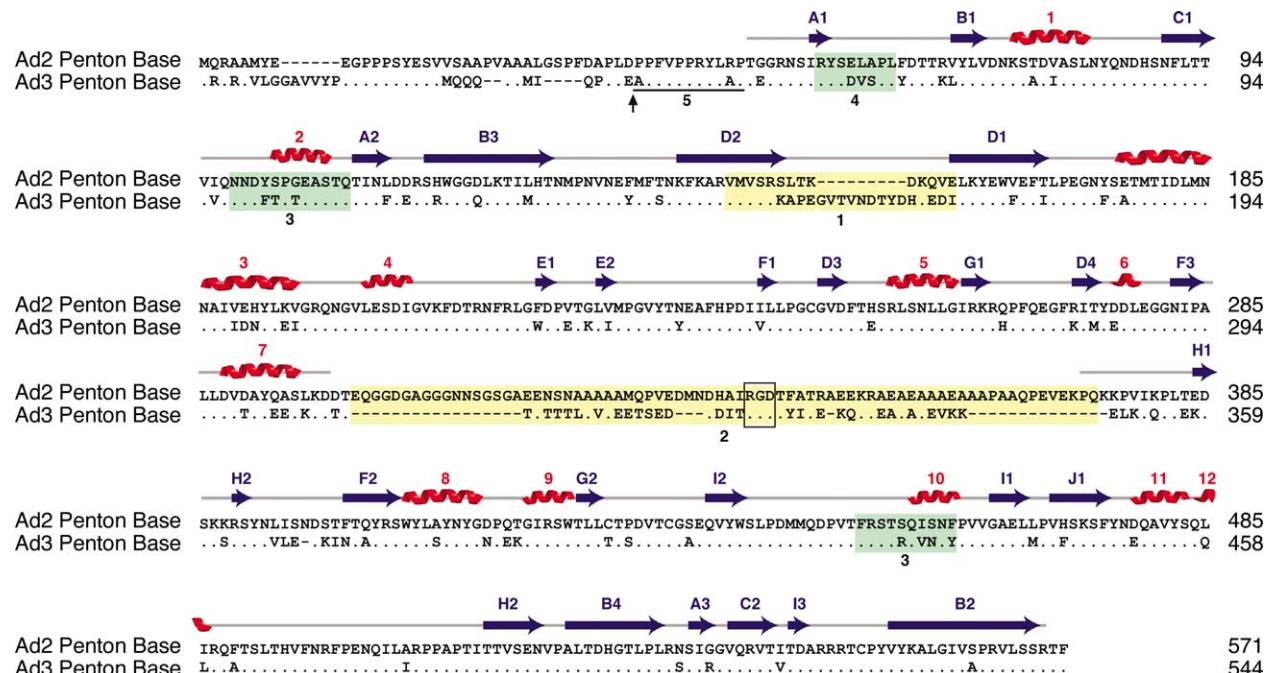


**Figure 2.** Negative stained images of expressed penton base. (a) Expressed penton base visualized immediately after expression and purification. Mainly free pentameric bases are visible with also some more or less complete dodecahedra. (b) Same sample as in (a) but visualized after two weeks at room temperature, having lost the first 37 residues. Most of the particles are well-formed dodecahedra. Only a few free penton bases are visible in the background. The scale bar represents 30 nm. (c) An SDS-PAGE gel of the two samples visualized in a, lane 1, and b, lane 2.

that the protein was N-terminally truncated between residues 37 and 38 (Figure 3). This 11 residue extension over the truncated Ad2 protein visualized by crystallography totals only 73 kDa (i.e. 60 copies of 11 residues) and is well able to pack into an 80 Å radius cavity. We conclude that our previous work on the Ad3 dodecahedron most probably involved N-terminally truncated protein resulting from inadvertent and unsuspected

proteolysis. In the experiments shown here, all penton base protein was cleaved at the same site (residues 37–38), but in principle one would expect that dodecahedra could be formed from a mixture of cleaved and full-length protein.

The observation of dodecahedra in lysates of cells infected by Ad3<sup>25</sup> but not from cells infected by Ad2 may be due to the fact that the cleaved site is not conserved between the two serotypes (Figure 3) and



**Figure 3.** Sequence alignment of penton base proteins from human adenoviruses 2 and 3. Secondary structures in the Ad2 penton base are denoted above the sequence by blue arrows ( $\beta$ -strands) and red helices ( $\alpha$ -helices). The RGD region (numbered 2, RGD is boxed into a black square) and the variable loop (numbered 1) are highlighted in orange whereas the regions involved in particle dodecamerization are in green and numbered 3 (two 2-fold related connections) and 4 (one connection on the 2-fold axis). The black arrow indicates the site of proteolysis allowing dodecahedron formation.<sup>23</sup> The number 5 and the black line above indicate residues 38–49 that are present in the Ad3 dodecahedron visualized in Figure 2(b) but not in the Ad2 dodecahedron X-ray structure. Regions 1–5 are also indicated in Figure 5.

that the Ad3 penton base is more readily cleaved in infected cells. During infection of cells by Ad2 there is production of a large excess of soluble fiber that is thought to be necessary to undo the tight junctions between cells involving the Ad2 receptor CAR.<sup>5</sup> Ad3 does not infect the same cells as Ad2 and uses a different receptor from CAR.<sup>3</sup> The fact that Ad3 dodecahedra are found in large quantities in Ad3-infected cells most likely means that these structures also play an as yet unknown role in the infection process.<sup>26</sup>

Electron cryo-microscopy and image analysis yielded models of the dodecahedron made from penton base only (Bs-Dd) and dodecahedron made of complete pentons (Pt-Dd) (Figure 4). The Bs-Dd particle was resolved at 9 Å but for the Pt-Dd particle only 16.5 Å was achieved as a consequence of both a smaller dataset of images as well as higher background noise due to the larger diameter particle being embedded in thicker ice. The Bs-Dd density map (Figure 4(d) and (e)) reveals tight packing of subunits in the penton bases but a looser association between penton bases. Although at lower resolution, the Pt-Dd map (Figure 4(f) and (g)) shows very strong density for the fiber up to the fiber head. Keeping in mind the improper imposition of 5-fold symmetry on the trimeric fiber molecule, several observations can be made. Firstly, the density level in the fiber is similar to that in the penton base, indicating that the fiber is coincident with the icosahedral 5-fold axis and so does not lose strength due to the imposed and inappropriate averaging. This strength of the fiber density is remarkable in comparison to that visualized in previous work<sup>18</sup> and likely results from embedding the particles in thicker ice, thus avoiding bending of the fibers, as well as correction of image distortions generated by the microscope (contrast transfer function).<sup>27</sup> We note that the Quantifoil grids used here appear to have rims of thick carbon around the holes that leads to thicker ice than for the "lacy" home made holey carbon films used previously. Secondly, we observe that the extent of the fiber is well delimited: small bridges of density connect it to the penton base (arrows in Figure 4(g)), and at the distal end the transition from strong to smeared density is sudden. Extending the radius of selection for the raw images does not result in any further extension of the fiber density in the reconstruction (results not shown), thus indicating the onset of either flexibility or a kink at the visible distal end. Such flexibility was observed before for negatively stained fibers of Ad2 that were seen to have a triangular shaped head with the fiber shaft lying flat on the carbon support film.<sup>28</sup> Flexibility of the fiber near the head and near the base probably assists in favorable positioning for receptor binding.<sup>29</sup> Finally, the fiber density shows an axial banding pattern with a spacing of about 15 Å. This feature will be discussed below.

The atomic structure of the Ad2 penton base was fit into the EM density maps of Ad3 Bs-Dd and Pt-Dd (Figure 5(a)–(d)). The atomic model fits very

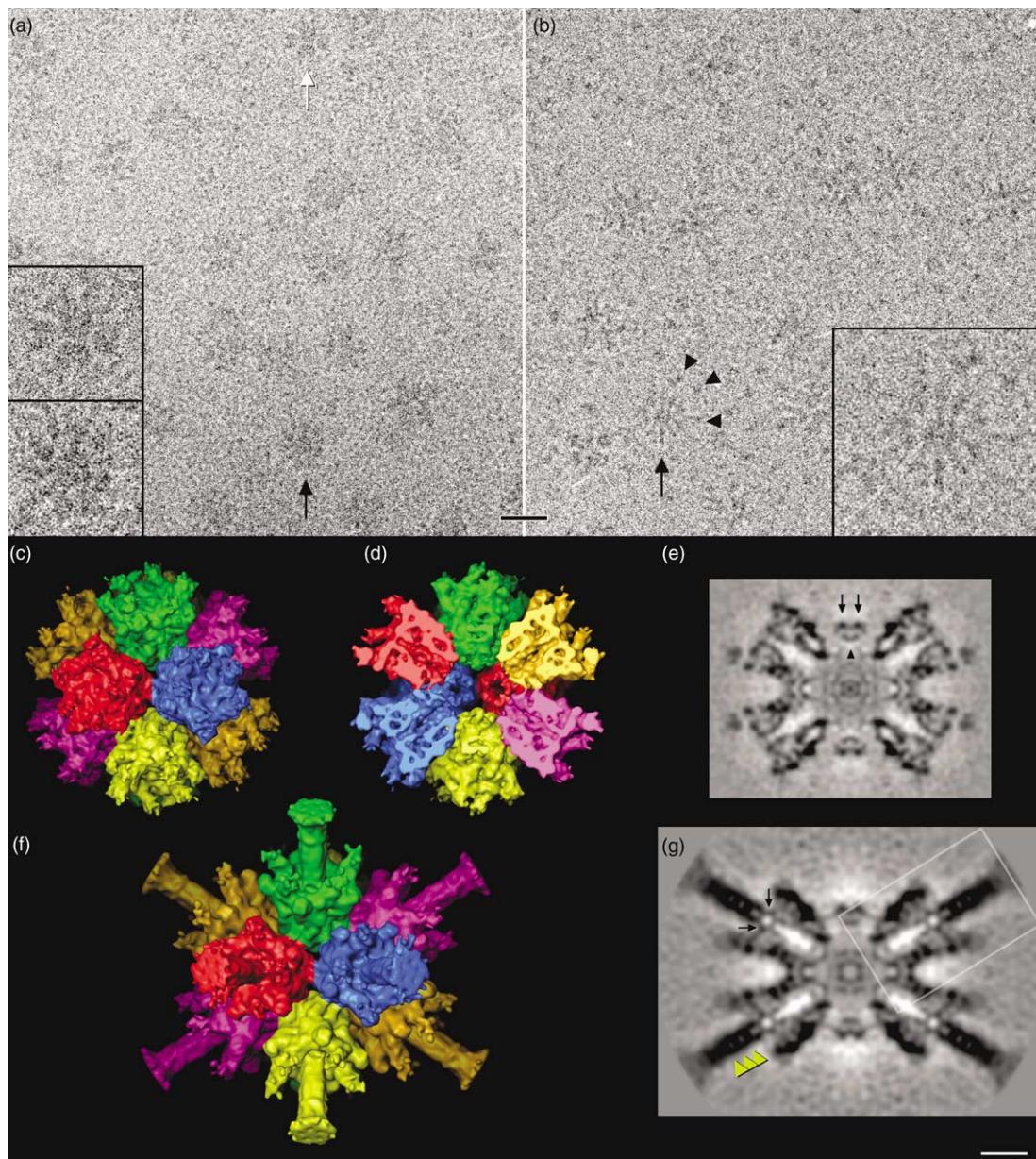
well (see Materials and Methods) apart from two loop areas where the EM density is too wide, as indicated in Figure 5(a) and (b) by 1 for the variable loop<sup>15,23</sup> and by 2 for the RGD loop (see also the sequence in Figure 3). The extra EM density in loop 1 is about 950 Å<sup>3</sup> in volume and can accommodate about 800 Da, slightly less than required for the eight amino acid insertion in the Ad3 penton base compared to that of Ad2. Less can be said about loop 2, the RGD loop, since most of it is invisible in the atomic structure of the Ad2 penton base<sup>23</sup> and in the 10 Å Ad5 structure.<sup>15</sup> Consequently, extra volume is visualized in the EM density maps corresponding to the RGD loop despite its shorter length than for Ad2 (Figure 3). This volume appears further from the penton's axis of symmetry in the Pt-Dd structure than for Bs-Dd, presumably a consequence of the fiber binding the penton base (see below).

The Ad3 dodecahedron is stabilized by three contacts between each pair of pentons. Two of these (marked 3 in Figure 5(c)–(f)) are related by the 2-fold axis and one (marked 4 in Figure 5(d) and (f)) is situated on the 2-fold axis. The corresponding positions in the Ad2 sequence are also marked by 3 and 4 in Figure 3. Because the quasi-atomic model of the Ad3 dodecahedron is made using the Ad2 X-ray structure, it is difficult to say exactly which amino acids interact across the Ad3 penton–penton interface. However, the atomic structure of the Ad2 penton base falls exactly inside the Ad3 EM density, implying that both types of penton base use the same structural elements to form the dodecahedron. Further, the high degree of homology between the sequences in these regions suggests that many of the same contacts may be made.

The EM structure shows additional density attached to the inside surface of the dodecahedron at the position of the N-terminal residue 49 of the Ad2 atomic structure (Figure 5(g) and (h)); the N-terminal residue 49 is marked by an asterisk (\*) and the extra density is marked by the number 5). We attribute this to the presence of 11 additional N-terminal residues present in the Ad3 penton base protein used for the cryoEM reconstructions, and it is a further indication that the N terminus of the penton base is sequestered inside the dodecahedron, and by inference inside the virus particle as well. Data obtained by cryoEM for the Ad5 capsid also showed extra density present under the penton and was likewise attributed to the N-terminal 48 residues of the penton base.<sup>15</sup> A corollary of these observations is that the N-terminal PPxY sequences that interact with cellular ubiquitine protein ligases<sup>24,30</sup> are not accessible in intact particles nor after assembly of new virus particles.

#### Model for the fiber density and its attachment to the base

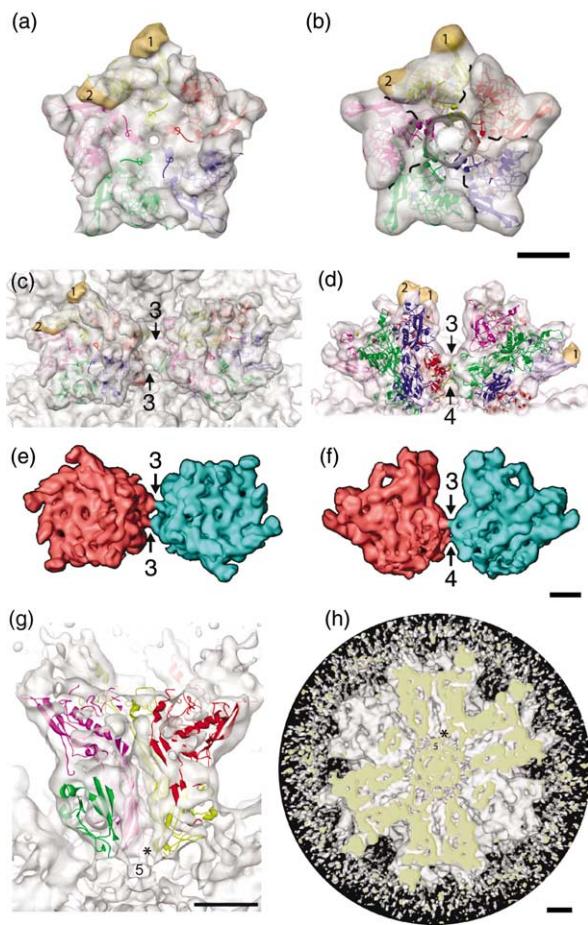
The penton base is a pentamer but the fiber molecule that binds it is a trimer. There are five binding sites on the penton base for the N-terminal



**Figure 4.** Structures of the Ad3 penton base dodecahedron with and without fiber. cryoEM images of Bs-Dd (a) and Pt-Dd (b), where arrowheads indicate fiber knobs. Several particles indicated by arrows are inset at double scale. The scale bar represents 250 Å. Surface representations of the Bs-Dd particle reconstruction viewed down a 2-fold axis and calculated from the cryoEM images are shown in (c), exterior view, and (d), interior view where the front half has been computationally removed. The penton bases are symmetrically identical, but colorized differently by a simple segmentation across the icosahedral 2-fold axes to aid in interpretation. The most peripheral density appears to be partially complete loops that may lose visibility due to flexibility. Gaps between penton bases are about 20 Å across and would allow solvent access to the interior. The interior cavity is about 80 Å in diameter and lined by the inwardly protruding loops of the penton bases. The central section corresponding to the sectioning plane in (d) is shown in grey-scale (protein is black) in (e). Densities 3 and 4 on the icosahedral 2-fold axis, and near or on the interface between penton bases, are marked with arrows and an arrowhead, respectively. These bridges are visualized in more detail in Figure 5. Equivalent views of the Pt-Dd particle are shown in (f), exterior view, and (g), central section, where density connections between the strong fiber density and the penton base are indicated with arrows. Also note the banding pattern along the axis of the fiber density, with a spacing of about 15 Å (yellow arrow heads). The boxed region is reproduced in Figure 6. An animation of the surface views is available on-line (Supplementary Data). The scale bar represents 50 Å.

tail of the fiber.<sup>23</sup> Therefore, when the fiber binds to the base, the fiber N termini will occupy at most three of the five available sites, an interesting example of symmetry mismatch but with unknown

purpose. Because icosahedral symmetry is imposed upon the dodecahedron during the image reconstruction process, the density at the vertices, including that of the trimeric fiber, is 5-fold

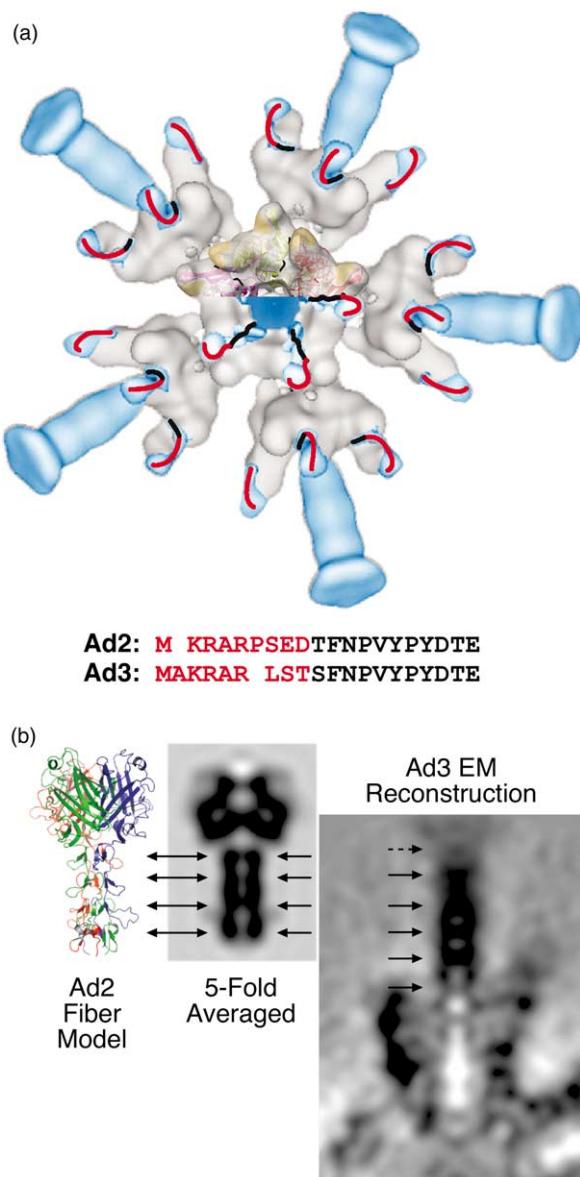


**Figure 5.** Quasi-atomic model of the Ad3 dodecahedron using the atomic structure of the Ad2 penton base. (a) Fit of the Ad2 penton base X-ray structure (PDB accession number 1X9P) into the Ad3 penton base EM density. The fit is nearly perfect at this resolution with only two unoccupied regions visible in the EM map. The first, marked 1, corresponds to the variable loop (see Figure 3) while the second, 2, is the RGD loop. (b) Fit of the fiber peptide, Ad2 penton base complex X-ray structure (PDB accession number 1X9T), into the Ad3 penton EM density. Extra densities are marked as in (a). The peptide that mimics the N terminus of the fiber (amino acid residues 10–20) is drawn in black. (c) and (d) Visualization of the contacts between two penton bases across the 2-fold axis. (c) Top-view along the icosahedral 2-fold axis, and (d) a perpendicular side view. The sequences involved in the different contacts are marked 3 and 4 as in Figure 3. Extra densities compared to the Ad2 X-ray structure are marked 1 and 2 as in (a) and (b). (e) and (f) The contact densities between penton bases are indicated in these surface views that have been rendered from a density map limited to 11 Å and contoured at a stringent level to highlight stronger densities. Arrows and numbers indicate features corresponding to those of (c) and (d) and in Figure 3. The boundaries between penton bases have been slightly adjusted from the simple 2-fold positions of Figure 4 by inspection of the density and respecting the 2-fold axis. The lower connection (4) is apparently weaker and is not quite continuous at this stringent contour level. (g) Same view of the Ad2 penton base X-ray data fitted into the Ad3 penton base as in (d) but the EM map is contoured at a lower threshold (i.e. more density is visible). It is then possible to see that the N-terminal part of the X-ray data (represented by an

asterisk (\*)) is now connected to a small cylindrical density inside the dodecahedron. We attribute this density to residues 38–49 (region 5 in Figure 3). (h) Same view as in (g) but for the entire dodecahedron. The cutting slice is colored in beige and region 5 and \* are indicated under one penton base. The bars represent 25 Å. averaged. When the EM density for Bs-Dd is subtracted from the density for Pt-Dd, several regions remain (colored light blue in Figure 6(a)), corresponding to the fiber as well as any penton base densities that are flexible in Bs-Dd but more rigid in Pt-Dd. On the top of the base, five spokes of density radiate azimuthally outwards, each consisting of three blobs. The center-most two blobs correspond to the density observed in the complete Ad5 capsid and also in the crystal structure of the Ad2 penton base plus an N-terminal fiber peptide, and are likely made up of Ad3 fiber residues 10–SFNPVYPYEDE-21.<sup>15,23</sup> The third blob is situated towards the outside of the penton base against the RGD loop and is either due to a stabilizing effect of the fiber N terminus on the flexible loop or comprises the nine N-terminal residues of the fiber tail. The volume of this blob of density, 1000 Å<sup>3</sup>, can contain 850 Da of protein, enough for most of the nine N-terminal Ad3 fiber tail residues (1-MAKRARLST-9). In a previous analysis<sup>18</sup> we interpreted the volume difference between Bs-Dd and Pt-Dd near the RGD loop as movement by this loop because at the low resolution of that analysis we could only see that the midpoint of the mass of this loop changed position. Only higher resolution data would confirm our hypothesis about the identity of the outer-most blue blobs.

The cross-section of the EM density of the fiber in Figure 4(g) shows a series of striations along its length. The distances between these striations is 15(±2) Å and thus could correspond to the 15 amino acid pseudo-repeats in the fiber shaft.<sup>28,31–33</sup> Indeed, when the atomic structure of the distal shaft and knob region of the trimeric Ad2 fiber<sup>31</sup> is 5-fold averaged, a similar pattern of striations is observed from averaging of the repeating three pairs of antiparallel β-strands that are close to the fiber axis (Figure 6(b); β-strands and corresponding striations are arrowed). Analysis of the sequence of the Ad2 and Ad3 fibers shows that they have N-terminal tails of similar length (45 residues for Ad3 and 46 for Ad2) followed by three shaft repeats in common.<sup>32</sup> Although the short Ad3 fiber has only six repeats in total compared to the 22 for the longer Ad2 fiber,<sup>32,34</sup> the regular spacing of repeats can be directly compared with those of Ad2 that are present in the crystal structure of the fiber (Figure 6(b)). The cross-section of the fiber density of Ad3 Pt-Dd shows five clear bands, with the sixth striation from the last repeat being smeared out, as is the density for the knob, implying some flexibility is likely between the fifth and sixth repeat. The distance between the first striation and the innermost blob of blue density shown in Figure 6(a) is about 15 Å and is evidently made up of the proline

asterisk (\*)) is now connected to a small cylindrical density inside the dodecahedron. We attribute this density to residues 38–49 (region 5 in Figure 3). (h) Same view as in (g) but for the entire dodecahedron. The cutting slice is colored in beige and region 5 and \* are indicated under one penton base. The bars represent 25 Å.



**Figure 6.** (a) The fiber of adenovirus. Difference mapping between Pt-Dd and Bs-Dd structures filtered at 18 Å resolution. Bs-Dd is in grey and the difference density is in blue. One half of the central penton shows the same fit as in Figure 5(b). The N-terminal fiber peptide (residues 11–21 of the Ad2 fiber<sup>23</sup>) is in black and superposes exactly with extra density visible in Pt-Dd (the two radially oriented cylinders of blue density). We interpret a third blue density localized on the outside part of the RGD loop (region 2 in Figures 3 and 5) as the N-terminal part of the fiber (Ad3 fiber residues 1–9) that is modeled by a red line connected to the black line (Ad2 fiber residues 11–21 seen in the crystal structure; Figure 5(b)). The sequence alignment of the first 20 amino acid residues of the Ad2 and Ad3 fibers indicates in black the amino acids that were modeled in the Ad2 crystallographic structure of the penton base with the bound fiber peptide, and in red those that were disordered. (b) Comparison of fiber density from the crystallographic structure of the Ad2 fiber head domain (residues 312–582)<sup>31</sup> as indicated, and from the Ad3 Pt-Dd density map. The Ad2 atomic model is averaged to mimic the Ad3 icosahedral reconstruction by filtering to 16 Å followed by imposition of 5-fold symmetry about an axis

and glycine-rich sequence 21-SSSQHPFINPG FISPDGFTQSPNGV-45 that must form the well-defined link visible in the cross-section (Figure 4(g), arrows).

The loose packing between penton bases and the accessibility of the dodecahedron interior to solvent are features of the subviral particle that contrast sharply with the adenovirus capsid, where penton bases interact with hexon and other minor structural proteins to form a closed capsid. In addition, the insight into the need for truncation of the N-terminal 37 residues (48 for Ad2) in order to form dodecamers highlights another difference with the *in situ* penton base and has been crucial to understanding how to optimize particle production. These factors are important for our goal of developing the dodecahedron as a therapeutic vector (P.F. *et al*, unpublished results). Loose packing is more likely to allow our manipulation of the assembly status for the purpose of encapsidating a drug or other "cargo", although maintaining assembly buffer conditions will be crucial for particle stability. The large trefoil perforations also present an interesting challenge for retaining any packaged molecules, and clearly the size of the molecule must be sufficient to avoid diffusion through the pores. This may be achieved by polymerization of the molecules, thus increasing the diameter of the payload, or by modifying the interior surface of the dodecahedron cavity to attract the molecules but without affecting assembly. Finally, the need for truncation is essential for assembling particles in the first place, but also suggests that the amino termini might be suitable sites to introduce modifications for enhancing recruitment and retention of specific drugs and in sufficient quantity to be useful. Clearly much needs to be done to develop this system, but the structural work presented here gives some clear directions for future work.

## Materials and Methods

### Protein preparation

Adenovirus 3 dodecahedra were prepared using the baculovirus expression system as described.<sup>20</sup> Briefly, a double expression vector pAcUW31 (Clontech, Palo Alto, CA) was used for cloning the Ad3 penton base protein (60 kDa) and fiber protein (35 kDa) genes into the baculovirus genome. Infectivity titres were determined on monolayers of *Spodoptera frugiperda* (Sf21) cells by

parallel with the fiber shaft. The section view shows four parallel striations corresponding to β-sheet crossovers, as indicated by arrows, and large loops between the lowest sets of β-sheet crossovers correspond to wisps of density between the lowest two striations in the section. At right is a corresponding section through the Ad3 fiber density of the Pt-Dd structure (boxed region from Figure 4(g)) and in approximate alignment by the distal ends presents similar striations. In the Ad3 fiber five striations are visible corresponding to five N-terminal motifs<sup>32</sup> while that for the distal sixth crossover is too smeared out to be clearly defined (broken arrow).

standard procedures. *Trichoplusia ni* moth cells (High Five; Invitrogen) were infected with the recombinant baculovirus isolates at a multiplicity of infection of 5 plaque-forming units/cell. Lysates of infected High Five cells were separated on a 15%–40% (w/w) sucrose gradient containing 10 mM Tris-HCl, 150 mM NaCl, 10% glycerol, 2 mM EDTA (pH 6.6) for 18 h in a Beckman SW41 rotor at 40,000 rpm. Expressed protein was visualized by SDS-PAGE and Coomassie stain of the collected fractions at weights of 60 kDa and 35 kDa corresponding to the penton base protein and fiber protein, respectively. After the gradient, the protein-containing fractions localized in 38%–40% sucrose were dialyzed against 25 mM phosphate buffer (pH 6.6) and further purified by anionic exchange chromatography (Econo-Pac High Q Cartridge; BioRad) using a linear 0 to 1 M NaCl gradient in the phosphate buffer.

### Electron microscopy

Particles were visualized by negative-stain EM to assess their assembly status using 1% (w/v) sodium silicotungstate (pH 7.0) as stain. cryoEM for structural analysis was performed according to standard low-dose methods. Bs-Dd particles were imaged on a JEOL JEM2010F microscope equipped with an Oxford CT3500 cryoholder operated at 200 kV and 50,000 $\times$  nominal magnification. Pt-Dd particles were imaged with an FEI CM200 microscope with a Gatan 626 cryoholder operated at 200 kV and 38,000 $\times$  nominal magnification. Quantifoil R2/1 grids (Quantifoil Micro Tools GmbH, Germany) were loaded with 4  $\mu$ l of sample at 1 mg/ml, blotted and rapidly frozen in liquid ethane within a liquid nitrogen bath using a Zeiss cryo-plunger. Grids were maintained at liquid nitrogen temperatures during transfer into the microscope and imaged with low dose techniques on Kodak SO163 film. Fields were imaged twice at different defocus values with the closer-to-focus exposure being made first. Negatives were developed in full strength D19 for 12 min, screened by optical diffraction to reject those with weak or astigmatic power spectra, and digitized at a 7  $\mu$ m sampling step on a Z/I Imaging PhotoScan (previously the Zeiss SCAI).

### Image analysis

Particle images were selected manually from the digitized micrographs, corrected for contrast transfer function effects and processed as described<sup>35</sup> with several modifications as noted. The Bs-Dd micrographs (seven defocus pairs) yielded 5597 image pairs of which 1849 were selected for the final reconstruction; for Pt-Dd, 2141 image pairs were collected from six defocus pairs of micrographs, and 1071 were included in the final density map. Particle orientations were determined with the PFT protocol<sup>36</sup> with modifications (D.M. Belnap, J.B. Heymann & J.F.C., unpublished results), and three-dimensional density maps were calculated using Fourier–Bessel methods. Resolution was estimated by Fourier Shell Correlation<sup>37</sup> calculated between independent half-dataset maps, and applying a correlation limit of 0.3: for Bs-Dd this was at 9.3  $\text{\AA}$  and for Pt-Dd, 16.5  $\text{\AA}$  (see Supplementary Data, Figure 1). Density maps were manipulated with BSOFT utilities<sup>38</sup> and visualized with ROBEM (Rob Ashmore, Purdue University, Indiana), Amira (Mercury Computer Systems Inc, Merignac, France), and the UCSF Chimera software<sup>†</sup>.

Density maps have been deposited in the Macromolecular Structure Database<sup>‡</sup> with accession numbers 1178 for Bs-Dd and 1179 for Pt-Dd.

### Docking of the atomic structure of the Ad2 penton base into the Ad3 penton base dodecahedron

The atomic structures of the Ad2 penton base alone (PDB accession number 1X9P) and in complex with the N-terminal fiber peptide (PDB accession number 1X9T) were fit into the EM model for Ad3 Bs-Dd and Pt-Dd, respectively. The atomic structures were placed manually into one penton base using the program O.<sup>39</sup> The optimization procedure was carried out using URO<sup>40</sup> with the pentamer treated as a single molecule (i.e. only applying icosahedral 2-fold and 3-fold symmetry). The scale factor was also optimized to determine the real magnification of the microscopes. For Bs-Dd the quality of the final fit is given by a correlation coefficient of 80.4%, a crystallographic *R* factor of 44.5% and a quadratic misfit of 17.9. The fitted scale factor was 0.98 of the original, corresponding to a calibrated pixel size in the cryoEM images of 1.37  $\text{\AA}$ . For Pt-Dd the quality of the final fit after masking out the fiber shaft and head domains is given by a correlation coefficient of 93%, a crystallographic *R* factor of 41.9% and a quadratic misfit of 14.9. The computed scale was 0.93 of the original, or 1.72  $\text{\AA}$ /pixel.

### Difference imaging

Both the Bs-Dd and Pt-Dd maps were converted into SPIDER format,<sup>41</sup> adjusted to a common pixel size of 1.72  $\text{\AA}$  and subsequently filtered to 18  $\text{\AA}$  resolution. The two maps were then normalized (same min/max and standard deviation) and the correct threshold calculated to include the expected molecular mass of the particle (excluding the fiber for Pt-Dd). Pt-Dd was subtracted from Bs-Dd but no significant positive densities were visible (not shown). When Bs-Dd was subtracted from Pt-Dd, positive difference density resulted as seen in Figure 6(a).

### Five-fold averaging of the fiber

The PDB file corresponding to the distal region of the Ad2 fiber (residues 319–582; PDB accession number 1QIU) was loaded into SPIDER and filtered to 16  $\text{\AA}$ . 5-fold symmetry was imposed along the fiber shaft to allow comparison with the corresponding density in the cryoEM model. A central slice through the resulting density is shown in Figure 6(b).

### Acknowledgements

We thank Evelyne Gout for technical help, Dr Fred Metz and Rémi Pinck for computational support, Drs David Belnap and Bernard Heymann for help with reconstruction software, and Dr Richard Wade for support and provision of microscopy equipment. J.F.C. acknowledges

<sup>†</sup> <http://www.cgl.ucsf.edu/chimera/>

<sup>‡</sup> <http://www.ebi.ac.uk/msd/>

support from the CNRS through an ATIP grant and by a fellowship for P. F.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2005.11.048](https://doi.org/10.1016/j.jmb.2005.11.048)

## References

1. Stewart, P. L., Fuller, S. D. & Burnett, R. M. (1993). Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy. *EMBO J.* **12**, 2589–2599.
2. Bergelson, J. M., Cunningham, J. A., Drogouet, G., Kurt-Jones, E. A., Krithivas, A., Hong, J. S. *et al.* (1997). Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science*, **275**, 1320–1323.
3. Roelvink, P. W., Lizonova, A., Lee, J. G., Li, Y., Bergelson, J. M., Finberg, R. W. *et al.* (1998). The coxsackievirus-adenovirus receptor protein can function as a cellular attachment protein for adenovirus serotypes from subgroups A, C, D, E, and F. *J. Virol.* **72**, 7909–7915.
4. Cohen, C. J., Shieh, J. T., Pickles, R. J., Okegawa, T., Hsieh, J. T. & Bergelson, J. M. (2001). The coxsackievirus and adenovirus receptor is a transmembrane component of the tight junction. *Proc. Natl Acad. Sci. USA*, **98**, 15191–15196.
5. Walters, R. W., Freimuth, P., Moninger, T. O., Ganske, I., Zabner, J. & Welsh, M. J. (2002). Adenovirus fiber disrupts CAR-mediated intercellular adhesion allowing virus escape. *Cell*, **110**, 789–799.
6. Gaggar, A., Shayakhmetov, D. M. & Lieber, A. (2003). CD46 is a cellular receptor for group B adenoviruses. *Nature Med.* **9**, 1408–1412.
7. Sirena, D., Lilienfeld, B., Eisenhut, M., Kalin, S., Boucke, K., Beerli, R. R. *et al.* (2004). The human membrane cofactor CD46 is a receptor for species B adenovirus serotype 3. *J. Virol.* **78**, 4454–4462.
8. Wickham, T. J., Mathias, P., Cheresh, D. A. & Nemerow, G. R. (1993). Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell*, **73**, 309–319.
9. Albinsson, B. & Kidd, A. H. (1999). Adenovirus type 41 lacks an RGD alpha(v)-integrin binding motif on the penton base and undergoes delayed uptake in A549 cells. *Virus Res.* **64**, 125–136.
10. Greber, U. F., Willetts, M., Webster, P. & Helenius, A. (1993). Stepwise dismantling of adenovirus 2 during entry into cells. *Cell*, **75**, 477–486.
11. Seth, P. (1994). Adenovirus-dependent release of choline from plasma membrane vesicles at an acidic pH is mediated by the penton base protein. *J. Virol.* **68**, 1204–1206.
12. Roberts, M. M., White, J. L., Grutter, M. G. & Burnett, R. M. (1986). Three-dimensional structure of the adenovirus major coat protein hexon. *Science*, **232**, 1148–1151.
13. Stewart, P. L., Chiu, C. Y., Huang, S., Muir, T., Zhao, Y., Chait, B. *et al.* (1997). Cryo-EM visualization of an exposed RGD epitope on adenovirus that escapes antibody neutralization. *EMBO J.* **16**, 1189–1198.
14. Chiu, C. Y., Mathias, P., Nemerow, G. R. & Stewart, P. L. (1999). Structure of adenovirus complexed with its internalization receptor, alphavbeta5 integrin. *J. Virol.* **73**, 6759–6768.
15. Fabry, C. M., Rosa-Calatrava, M., Conway, J. F., Zubieta, C., Cusack, S., Ruigrok, R. W. & Schoehn, G. (2005). A quasi-atomic model of human adenovirus type 5 capsid. *EMBO J.* **24**, 1645–1654.
16. Pettersson, U. & Hoglund, S. (1969). Structural proteins of adenoviruses. 3. Purification and characterization of the adenovirus type 2 penton antigen. *Virology*, **39**, 90–106.
17. Norrby, E. (1968). Comparison of soluble components of adenovirus types 3 and 11. *J. Gen. Virol.* **2**, 135–142.
18. Schoehn, G., Fender, P., Chroboczek, J. & Hewat, E. A. (1996). Adenovirus 3 penton dodecahedron exhibits structural changes of the base on fibre binding. *EMBO J.* **15**, 6841–6846.
19. Fender, P., Schoehn, G., Foucaud-Gamen, J., Gout, E., Garcel, A., Drouet, E. & Chroboczek, J. (2003). Adenovirus dodecahedron allows large multimeric protein transduction in human cells. *J. Virol.* **77**, 4960–4964.
20. Fender, P., Ruigrok, R. W., Gout, E., Buffet, S. & Chroboczek, J. (1997). Adenovirus dodecahedron, a new vector for human gene transfer. *Nature Biotechnol.* **15**, 52–56.
21. Vives, R. R., Lortat-Jacob, H., Chroboczek, J. & Fender, P. (2004). Heparan sulfate proteoglycan mediates the selective attachment and internalization of serotype 3 human adenovirus dodecahedron. *Virology*, **321**, 332–340.
22. Mizuguchi, H. & Hayakawa, T. (2004). Targeted adenovirus vectors. *Hum. Gene Ther.* **15**, 1034–1044.
23. Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005). The structure of the human adenovirus 2 penton. *Mol. Cell*, **17**, 121–135.
24. Galinier, R., Gout, E., Lortat-Jacob, H., Wood, J. & Chroboczek, J. (2002). Adenovirus protein involved in virus internalization recruits ubiquitin-protein ligases. *Biochemistry*, **41**, 14299–14305.
25. Norrby, E. (1969). The structural and functional diversity of Adenovirus capsid components. *J. Gen. Virol.* **5**, 221–236.
26. Fender, P., Boussaid, A., Mezin, P. & Chroboczek, J. (2005). Synthesis, cellular localization, and quantification of penton-dodecahedron in serotype 3 adenovirus-infected cells. *Virology*, **340**, 167–1673.
27. Conway, J. F. & Steven, A. C. (1999). Methods for reconstructing density maps of “single” particles from cryoelectron micrographs to subnanometer resolution. *J. Struct. Biol.* **128**, 106–118.
28. Ruigrok, R. W., Barge, A., Albiges-Rizo, C. & Dayan, S. (1990). Structure of adenovirus fibre. II. Morphology of single fibres. *J. Mol. Biol.* **215**, 589–596.
29. Wu, E., Pache, L., Von Seggern, D. J., Mullen, T. M., Mikyas, Y., Stewart, P. L. & Nemerow, G. R. (2003). Flexibility of the adenovirus fiber is required for efficient receptor interaction. *J. Virol.* **77**, 7225–7235.
30. Chroboczek, J., Gout, E., Favier, A. L. & Galinier, R. (2003). Novel partner proteins of adenovirus penton. *Curr. Top. Microbiol. Immunol.* **272**, 37–55.
31. van Raaij, M. J., Louis, N., Chroboczek, J. & Cusack, S. (1999). Structure of the human adenovirus serotype 2 fiber head domain at 1.5 Å resolution. *Virology*, **262**, 333–343.
32. Chroboczek, J., Ruigrok, R. W. & Cusack, S. (1995). Adenovirus fiber. *Curr. Top. Microbiol. Immunol.* **199**, 163–200.
33. Green, N. M., Wrigley, N. G., Russell, W. C., Martin, S. R. & McLachlan, A. D. (1983). Evidence for a repeating

- cross-beta sheet structure in the adenovirus fibre. *EMBO J.* **2**, 1357–1365.
34. Signas, C., Akusjarvi, G. & Pettersson, U. (1985). Adenovirus 3 fiber polypeptide gene: implications for the structure of the fiber protein. *J. Virol.* **53**, 672–678.
35. Conway, J. F. & Steven, A. C. (1999). Methods for reconstructing density maps of single particles from cryo-electron micrographs to subnanometer resolution. *J. Struct. Biol.* **128**, 106–118.
36. Baker, T. S. & Cheng, R. H. (1996). A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J. Struct. Biol.* **116**, 120–130.
37. van Heel, M. (1987). Similarity measures between images. *Ultramicroscopy*, **48**, 95–100.
38. Heymann, J. B. (2001). Bsoft: image and molecular processing in electron microscopy. *J. Struct. Biol.* **133**, 156–169.
39. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. sect. A*, **47**, 110–119.
40. Navaza, J., Lepault, J., Rey, F. A., Alvarez-Rua, C. & Borge, J. (2002). On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr. sect. D*, **58**, 1820–1825.
41. Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199.

Edited by Sir A. Klug

(Received 7 September 2005; received in revised form 4 November 2005; accepted 15 November 2005)  
Available online 9 December 2005

## VII.2 Conclusion

Nous avons étudié la structure du dodécaèdre dérivé du HAdV-3. Il est formé par l'assemblage, selon un arrangement dodécaédrique de 12 Pentons, reliés de façon non covalente. Une analyse biochimique nous a tout d'abord permis de confirmer qu'un clivage protéolytique de la région N terminale de la Base du Penton est nécessaire à la formation de la particule. Cette observation est importante pour l'optimisation de la production du dodécaèdre en laboratoire.

Par la suite, nous avons réalisé une reconstruction 3D en ME à haute résolution de la particule complète et de celle ne possédant pas de fibre. Un modèle quasi-atomique a pu être reconstitué pour les deux types de particules en recalant, à l'intérieur de la densité issue de la ME, la structure atomique de la Base du Penton du HAdV-2 en complexe ou non avec le peptide N-terminal de la fibre. Notre principal objectif dans cette étude structurale était de comprendre comment une protéine possédant un axe d'ordre 3 (La Fibre) peut réaliser un complexe de forte affinité avec une protéine pentamérique (La Base du Penton). Le processus de reconstruction 3D introduit un biais en imposant un axe 5 au niveau de la fibre. Nous avons tout de même pu donner un modèle d'attachement de la fibre à la base du Penton en étudiant la carte de différence entre les modèles obtenus pour le dodécaèdre complet et le dodécaèdre Base. La partie N terminale de la fibre vient se positionner juste en dessous de la boucle flexible où se trouve le motif RGD responsable de l'internalisation du virus chez le HAdV-2. L'analyse des cartes de différence entre les différents modèles 3D en ME et quasi-atomiques obtenus a permis de révéler d'importantes différences entre la structure de la Base du Penton du HAdV-2 et du HAdV-3 notamment au niveau des boucles RGD et variables. En effet, la boucle variable est de plus grande taille chez le HAdV-3 alors que la boucle RGD est beaucoup plus courte que dans le HAdV-2. De plus, grâce aux modèles quasi-atomiques, nous avons analysé les contacts responsables de la formation de la particule et étudié la cavité interne du dodécaèdre dans laquelle résident les extrémités N terminales clivées des Pentons. Cette région de la protéine n'est pas impliquée dans la stabilisation du dodécaèdre. Elle pourrait donc faire l'objet de modifications afin de libérer de l'espace dans la cavité et/ou de retenir des molécules thérapeutiques.

# DISCUSSION GENERALE SUR LA STRUCTURE DE LA CAPSIDE DES ADENOVIRUS

---

Cette partie marque la fin de mon manuscrit. J'aimerais à ce titre réaliser une revue sur l'évolution de la structure de la capsid des Adénovirus entre le début et la fin de mes travaux de thèse. J'aimerais tout d'abord insister sur l'apport de la ME associée aux techniques de « fitting » dans l'étude de la structure des virus en particulier sur l'étude de l'Adénovirus. Les Adénovirus sont des virus de grande taille ce qui rend très difficile leur étude par cristallographique aux rayons X ou RMN. Des structures cristallographiques sont disponibles pour les protéines majeures de la capsid mais pour l'instant la structure atomique du virus entier n'est pas connue. Il a donc fallu développer des techniques afin d'observer le virus à des résolutions de plus en plus importantes. La Microscopie Electronique et le « fitting » jouent ici un grand rôle. La majorité des connaissances structurales disponibles sur les virus proviennent de la ME. Les progrès réalisés pour l'automatisation des microscopes électroniques ainsi que l'utilisation de caméra CCD en complément des logiciels d'analyse et de correction des images ont permis l'obtention de structures 3D de plus en plus résolues. Un pas a été réalisé par l'informatisation du ME. Le pilotage par ordinateur du Microscope permet de réaliser une meilleure correction des astigmatismes, ce qui a contribué de manière forte à l'amélioration des images enregistrées. De plus, couplée à l'utilisation de caméra CCD, l'informatisation a permis l'enregistrement automatique d'images numériques. La structure d'un HAdV-5 à 6 Å de résolution obtenue par Phoebe Stewart à l'aide d'un microscope équipé d'une caméra CCD et d'un gestionnaire automatique de prise d'images confirme les progrès apportés par cette méthodologie (Saban et al., 2006). Les efforts réalisés dans le traitement et la correction des images rendent maintenant possible la correction de l'astigmatisme dans les images ainsi que la détermination de l'orientation des particules dans les images même quand le contraste est faible. Ces développements contribuent aussi fortement à l'obtention de structures 3D à des résolutions de plus en plus hautes. Avec l'amélioration de la résolution, on a pu combiner les informations atomiques obtenues par cristallographique aux rayons X avec les données obtenues en ME à plus basse résolution. De là est né la technique « fitting » basée sur le traitement informatique des données. Cette technique de calcul a permis une étude précise des objets manipulés en ME. En absence de

structure cristallographique de la capsidé entière, le calcul de modèles quasi-atomiques nous a apporté beaucoup d'informations structurales notamment en facilitant l'analyse des cartes 3D haute résolution par calcul de cartes de différence plus précises. Avec le développement de cette technique, les programmes utilisés sont de plus en plus précis dans le recalage des structures cristallographiques au sein des modèles obtenus en ME et sont de plus en plus simples d'utilisation. Leur simplicité rend d'autant plus facile l'analyse d'un modèle 3D obtenu par ME même pour des scientifiques n'ayant pas de connaissances importantes en informatique. Ceci a pour conséquence la démocratisation du « fitting » au sein de la communauté scientifique y compris chez les non-virologistes.

Maintenant, j'aimerais revenir sur la structure des Adénovirus et surtout sur la composition de la capsidé en protéines mineures afin de réaliser une synthèse des différentes parties de mon manuscrit. Au début de ma thèse, la localisation des protéines mineures de la capsidé chez les Adénovirus humains était plus ou moins admise par la communauté scientifique. La reconstruction 3D obtenue par Phoebe Stewart en 1993 avait permis d'élaborer un schéma précis de la structure de la capsidé des Adénovirus. A la fin de ma thèse, ce schéma a subi des modifications majeures qui sont résumées en [Figure 56](#), avec une présentation du modèle établi depuis 1993.

*Les protéines Majeures* : La localisation et la composition de la capsidé en protéine majeure n'a pas subi de changement. Ces protéines possèdent une taille importante et ont été identifiées très rapidement. Grâce au modèle quasi-atomique que nous avons déterminé, les contacts qui règnent au sein de la capsidé entre les protéines majeures ont pu être étudiés. Ces contacts sont présents sous forme de quatre environnements distincts et illustrent parfaitement le principe de la quasi-équivalence (cf. Introduction). Nous avons aussi pu, avec l'étude structurale du dodécaèdre de l'Adénovirus de type 3, identifier la façon dont l'extrémité N terminale de la Fibre interagit avec la Base du Penton mais cela ne nous a pas permis de comprendre comment une protéine trimérique peut interagir avec une protéine pentamérique. Afin de comprendre ce phénomène, il faudrait traiter le complexe Base du Penton – Fibre comme une particule isolée et réaliser une reconstruction 3D de ce complexe sans imposer de symétries (axe de rotation d'ordre 5) lors du processus de reconstruction 3D. Ceci peut être réalisé en orientant le Penton à partir de l'orientation des particules virales complètes comme l'a proposé Briggs et collaborateurs en 2005 ([Briggs et al., 2005](#)). Cela reste tout de même un traitement informatique complexe à réaliser qui nécessiterait un grand nombre d'images du Penton avec une gamme large d'orientations. L'obtention de la structure du

Penton par ME, comme une particule isolée apporterait certainement beaucoup d'informations structurales sur l'interaction entre la Base du Penton et la Fibre chez les Adénovirus.

*Les protéines Mineures* : Contrairement aux protéines majeures, la localisation attribuée ainsi que la composition de la capsid des Adénovirus en protéines mineures a subi de nombreux changements au cours de ma thèse. Leur localisation reste encore très controversée. Ceci est dû à plusieurs raisons majeures. Tout d'abord, les protéines mineures sont des protéines de petite taille d'environ 60 kDa pour la plus grande à 1 kDa pour la plus petite. Leur petite taille rend leur étude dans le contexte capsidaire délicate et leur localisation, dans ce cas, difficile à déterminer surtout sur des reconstructions 3D de capsid virale entière à basse ou moyenne résolution. Depuis le séquençage du génome des Adénovirus, la taille des protéines non maturées, précurseur des protéines mineures est connue. La séquence consensus de maturation des protéines par la protéase virale a aussi pu être identifiée ainsi que certains sites de coupure mais très peu d'études ont été menées sur la maturation des protéines mineures. On ne connaît donc pas la taille ni l'état d'oligomérisation des protéines mineures dans le contexte de la capsid virale mature, rendant difficile leur étude structurale.

La protéine IX a pour la première fois été localisée par [Furcinitti et al., 1989](#). Cette protéine formait quatre trimères du côté externe de la capsid, stabilisant ainsi un GON. Par la suite, Stewart et collaborateurs observèrent la même densité en forme de triskelion au niveau de la partie externe des GONs et conclurent à la présence de la protéine IX dans sa totalité ([Stewart et al., 1993](#)). La reconstruction 3D à haute résolution que nous avons obtenu en 2005 (Article 1) nous a montré que la densité attribuée à la protéine IX ne pouvait contenir que l'extrémité N terminale de la protéine IX (5kDa au lieu de 10-12 kDa au total). Ce problème apparut aussi dans la reconstruction réalisée par Saban et collaborateurs en 2005 ([Saban et al., 2005](#)). Une première hypothèse fût émise pour expliquer ce phénomène. La partie C terminale manquante de la protéine IX ne réaliseraient pas de contacts du côté externe de la capsid ce qui entraînerait la formation d'une région trop flexible pour être résolue en ME (Article 1). En 2007, les travaux publiés par Saban et collaborateurs sur une reconstruction 3D du HAdV-5 à 6 de résolution (Saban et al., 2006) ainsi que les travaux de marquage de l'extrémité C terminale de la protéine IX par un Fab que j'ai réalisé au cours de ma thèse ont apporté une clarification majeure (cf. Partie V). L'extrémité C terminale de la protéine IX se situe donc à la position qui avait précédemment été attribuée à la protéine IIIa ([Figure 56](#)). Ces

résultats ont totalement remis en question la position de la protéine IIIa du côté externe de la capsidé.

Depuis 1993, la protéine IIIa formait une densité cylindrique à la liaison entre 2 GONs de deux faces adjacentes. Suite aux travaux de Saban et collaborateurs, la protéine IIIa a gagné une localisation interne au niveau du complexe protéique présent sous la Base du Penton ([Figure 56](#)). Nous n'avions pas pu identifier en 2005 la composition de ce complexe protéique. L'hypothèse des auteurs propose une explication aux densités surnuméraires que nous n'avions pas pu assigner à l'intérieur de la capsidé mais place la protéine IIIa au sein de la densité que nous avions attribuée à la protéine VI en 2005 ([Article 1](#)).

La localisation de la protéine VI reste encore très controversée et incertaine à la fin de ma thèse. En 1993, les auteurs avaient observé une densité trimérique en dessous de la Base du Penton. Ce trimère, présent en cinq exemplaires sous chaque sommet de la capsidé, avait donc été attribué à la présence de la protéine VI ([Stewart et al., 1993](#)). Sur la reconstruction 3D à plus haute résolution de la capsidé d'un Adénovirus de type 5 que nous avons obtenu en 2005, ces trimères ne sont pas présents mais une densité globulaire unique est visible au même endroit du côté interne de la capsidé. Nous avions donc confirmé la présence de protéine VI à cet endroit. Les récents travaux de Saban et al. remettent en question cette localisation interne et propose un autre emplacement situé à l'intérieur des Hexons. Cette hypothèse reste toute de même très discutable au vu de la densité supplémentaire observée au milieu des Hexons. La localisation précise de la protéine VI reste donc inconnue à ce jour ([Figure 56](#)).

Une dernière protéine mineure, la protéine VIII, dont la localisation n'avait pas pu être déterminé en 1993 de part sa petite taille a été assigné suite à l'étude structurale que nous avions mené en 2005 ([Figure 56](#)) ([Article 1](#)). L'obtention d'une reconstruction 3D à haute résolution ainsi que d'un modèle quasi-atomique de la capsidé du HAdV-5 nous a permis de proposer une localisation du côté interne de la capsidé. Cette observation a été confirmé par Saban et collaborateurs ([Saban et al., 2006](#)).

La [Figure 56](#) résume les changements subis par le modèle structurale de la capsidé des Adénovirus.

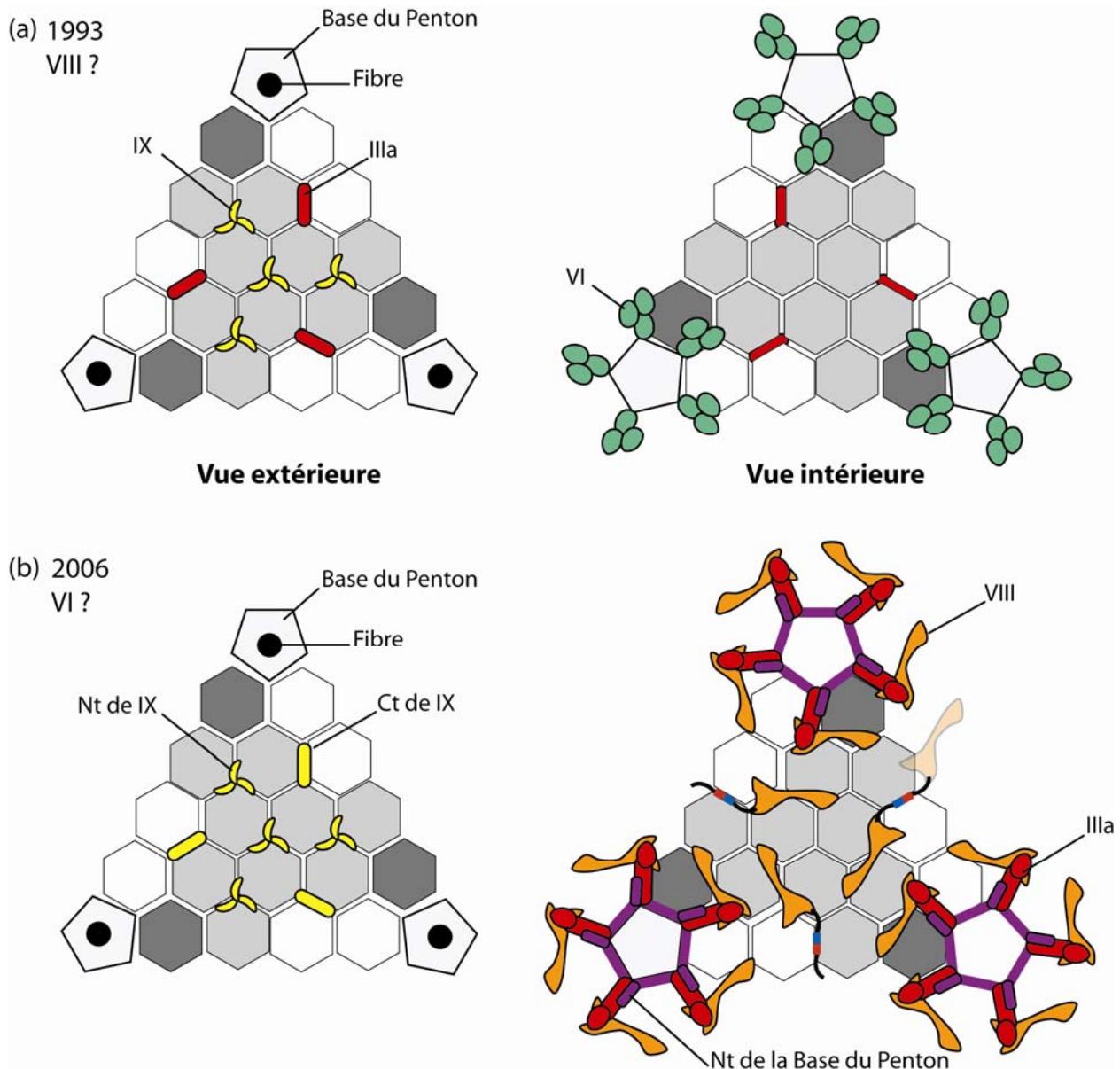


Figure 56: Schémas comparés d'une face de la capsid d'un Adénovirus humain, illustrant les changements de localisation subis par les protéines mineures.

(a) Représentation schématique en vue intérieure et extérieure d'une face de la capsid, d'après Stewart et al., 1993.

(b) Représentation schématique actuelle d'une face, en vue intérieure et extérieure, suite aux travaux réalisés sur la capsid des Adénovirus depuis 2005.

# CONCLUSION ET PERSPECTIVES

---

Dans cette étude structurale, notre but était de comprendre l'assemblage de la capsidé des Adénovirus par l'intermédiaire de deux techniques majeures de la Biologie Structurale : la microscopie électronique et la cristallographie aux rayons X. La combinaison de ces deux techniques semblait la plus appropriée pour l'étude d'un virus icosaédrique d'aussi grande taille qu'est l'Adénovirus. L'obtention d'un modèle quasi-atomique pour l'Adénovirus humain de type 5 nous a permis de décrire la capsidé de façon bien plus détaillée que les précédentes études datant de 1993. Ceci nous a conduis à repenser la localisation des protéines mineures au sein de la capsidé comme celle de la protéine VI ou VIII. Un doute persistait toujours du coté externe de la capsidé. La localisation de la protéine IIIa était encore incertaine et les travaux de Saban et collaborateurs en 2006 ont fini de semer le doute en proposant une localisation interne de la protéine IIIa. Par la suite, l'étude structurale en ME que nous avons mené sur un virus humain de type 5 en complexe avec un Fab dirigé contre l'extrémité C-terminale de la protéine IX a montré que l'extérieur de la capsidé n'est occupé que par la protéine IX ce qui suggère une localisation interne de la IIIa. Les travaux réalisés sur une souche aviaire et surtout sur la souche canine CAV-2 ont montré des différences dans l'organisation de la capsidé entre les différentes familles d'Adénovirus. En effet, chez le CAV-2, la protéine IX se situe, comme dans la capsidé d'un Adénovirus humain, en totalité du coté externe de la capsidé mais son extrémité C terminale n'occupe pas du tout la même place. Nous nous sommes aussi intéressés à un complexe protéique particulièrement important dans le virus : le Penton, formé de l'assemblage non covalent de la Fibre et de la Base du Penton. Son étude par microscopie électronique et cristallographie aux rayons X chez le HAdV-3 a permis une avancée dans la compréhension de ce complexe qui tient une part prépondérante dans le cycle d'infection de l'Adénovirus. Enfin l'étude d'un virus immature dérivé de l'Adénovirus humain de type 2 a montré une évidente différence avec un virus mature sans permettre de conclure sur la nature des changements que l'on observe dans la capsidé. Cette dernière étude ouvre la discussion concernant les processus de maturation et d'assemblage des Adénovirus. Depuis 1993, les études structurales avaient mis de coté les Adénovirus mais avec l'avènement de la thérapie génique, la communauté scientifique s'est de nouveau concentrée sur leur étude. Mon travail de thèse constitue une étude

structurale complète à haute résolution de la capsidé des Adénovirus révélant ainsi une organisation protéique nouvelle et différente de ce qui était établi.

Dans l'investigation de la capsidé des Adénovirus que ce soit de souches humaines ou animales, l'obtention d'une structure atomique de la capsidé d'un Adénovirus par cristallographie aux rayons X marquerait un pas décisif. Seule une telle structure permettrait de déterminer la composition exacte en protéines mineures ainsi que la localisation et la manière dont les protéines interagissent les unes avec les autres pour former une capsidé virale complète et stable.

# ANNEXES

---

Annexe 1: Article 4

“An archaeal peptidase assembles into two different quaternary structures: a tetrahedron and a giant octahedron”

Auteurs: G. Schoehn, F.M.D. Vellieux, M.A. Dura, V. Receveur-Bréchot, C.M.S. Fabry, R.W.H. Ruigrok, C. Ebel, A. Roussel & B. Franzetti

Publié en 2006 dans *The Journal of Biological Chemistry*, Vol. 281 (47), pages 36327–36337.

# An Archaeal Peptidase Assembles into Two Different Quaternary Structures

## A TETRAHEDRON AND A GIANT OCTAHEDRON<sup>\*§</sup>

Received for publication, May 9, 2006, and in revised form, August 3, 2006. Published, JBC Papers in Press, September 14, 2006, DOI 10.1074/jbc.M604417200

Guy Schoehn<sup>†1</sup>, Frédéric M. D. Vellieux<sup>§1</sup>, M. Asunción Durá<sup>§1,2</sup>, Véronique Receveur-Bréchot<sup>¶</sup>, Céline M. S. Fabry<sup>‡</sup>, Rob W. H. Ruigrok<sup>‡</sup>, Christine Ebel<sup>§</sup>, Alain Roussel<sup>¶13</sup>, and Bruno Franzetti<sup>§4</sup>

From the <sup>†</sup>Laboratoire de Virologie Moléculaire et Structurale c/o EMBL, FRC 2854 CNRS-UJF, 38042 Grenoble, <sup>§</sup>Laboratoire de Biophysique Moléculaire, Institut de Biologie Structurale J.-P. Ebel, UMR 5075 CNRS-CEA-UJF, 38027 Grenoble, and <sup>¶</sup>Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS, Universités Aix-Marseille I and II, 13288 Marseille, France

Cellular proteolysis involves large oligomeric peptidases that play key roles in the regulation of many cellular processes. The cobalt-activated peptidase TET1 from the hyperthermophilic Archaea *Pyrococcus horikoshii* (PhTET1) was found to assemble as a 12-subunit tetrahedron and as a 24-subunit octahedral particle. Both quaternary structures were solved by combining x-ray crystallography and cryoelectron microscopy data. The internal organization of the PhTET1 particles reveals highly self-compartmentalized systems made of networks of access channels extended by vast catalytic chambers. The two edifices display aminopeptidase activity, and their organizations indicate substrate navigation mechanisms different from those described in other large peptidase complexes. Compared with the tetrahedron, the octahedron forms a more expanded hollow structure, representing a new type of giant peptidase complex. PhTET1 assembles into two different quaternary structures because of quasi-equivalent contacts that previously have only been identified in viral capsids.

Cytosolic proteolysis is a key cellular process. In addition to its role in amino acid metabolism, it is responsible for protein quality control and the adjustment of the turnover of regulatory molecules (1, 2). Initial proteolysis is carried out by nonspecific endoproteases (1). Through oligomerization, these enzymes form barrel-like cellular subcompartments to protect cytoplasmic proteins from unwanted cleavage; the

active sites are situated in proteolytic chambers, only accessible to unfolded polypeptides. In Eukarya and Archaea, the 20 S proteasome is the core of this proteolytic system (3, 4). The proteins to be degraded are unfolded and translocated into the proteolytic complex by energy-dependent regulatory particles (5). Polypeptide degradation results in multiple endoproteolytic cleavage, yielding short peptides (6–12 residues in length) (6). The final breakdown of these fragments is probably achieved by a collection of ATP-independent exopeptidases, which have been proposed to be essential components of the protein degradation system (7, 8).

Structural studies on the Tricorn protease complex, an archaeal carboxypeptidase (peptidyl-di/tripeptidase), have shown that some exopeptidases can also form multimeric ring structures, with the active sites sequestered within an internal chamber (9). This feature of compartmentalization was also recognized in the three-dimensional structures of three ATP-independent intracellular aminopeptidases: Gal6, leucine aminopeptidase, and the D-aminopeptidase DppA (10–12). In eukaryotic cells, the tripeptidyl-peptidase II also forms a giant, well organized toroidal oligomeric structure (13). All these exopeptidases could function as scavengers of the oligopeptides generated by the ATP-dependent endoproteases (14).

TET is another type of a large hollow peptidase complex. It was characterized initially in the halophilic Archaea *Haloarcula marismortui* (15). The complex has a tetrahedral shape and is a broad substrate specificity aminopeptidase capable of degrading large peptides. In the MEROPS peptidase data base (16), TET belongs to the M42 family of metallopeptidases (clan MH).

*Pyrococcus horikoshii* is an anaerobic Archaea, which grows optimally at 98 °C. It was isolated from hydrothermal fluid samples obtained at a depth of 1395 m in the Pacific Ocean (17). We recently showed that the *P. horikoshii* TET2 protein (PhTET2),<sup>5</sup> which shares 24% identity with the HmTET, is a 12-subunit tetrahedral aminopeptidase that

\* This work was supported in part by the Groupe d'Intérêt Scientifique Génomique Marine (Roscoff-France). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2CF4) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

Density maps have been deposited in the Macromolecular Structure Database with accession numbers 1188 for PhTET1-12s and 1189 for PhTET1-24s.

§ The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. 1–3 and Movies 1 and 2.

<sup>1</sup> These authors equally contributed to this work.

<sup>2</sup> Supported by a Marie Curie Intra-European Fellowship (European Community Sixth Framework Programme).

<sup>3</sup> Present address: Centre de Biophysique Moléculaire, 45071 Orléans, France.

<sup>4</sup> To whom correspondence should be addressed: Laboratoire de Biophysique Moléculaire, Institut de Biologie Structurale J.-P. Ebel, UMR 5075 CEA-CNRS-UJF, 41 Rue Jules Horowitz, 38027 Grenoble Cedex 1, France. Tel.: 33-4-38-78-95-69; Fax: 33-4-38-78-54-94; E-mail: franzetti@ibs.fr.

<sup>5</sup> The abbreviations used are: PhTET2, *P. horikoshii* TET2; HmTET, *Haloarcula marismortui* TET; PhTET1-12s, *P. horikoshii* TET1 12-subunit complex; PhTET1-24s, *P. horikoshii* TET1 24-subunit complex; PIPES, piperazine-N,N'-bis(2-ethanesulfonic acid); pNA, 4-nitroaniline; AMC, 7-amino-4-methyl-coumarin; AAP, V. (formerly *Aeromonas*) *proteolyticus* aminopeptidase; HPLC, high pressure liquid chromatography.

displays a nonprocessive catalytic activity against N-terminal aliphatic and neutral amino acids (18). The structure of the PhTET2 edifice reveals the self-compartmentalizing nature of the complex (19, 20). In the genome of *P. horikoshii*, the open reading frame PH0519 encodes a homolog of PhTET2 that has been characterized as a cobalt-activated peptidase, suspected to be a deblocking aminopeptidase and thereby named DAP (21). The three-dimensional structures of PH0519, presented here, show that it can assemble alternatively as a tetrahedral dodecameric particle (therefore called PhTET1-12s) or as an octahedral tetracosameric edifice (PhTET1-24s). These assemblies are different from those observed in the other self-compartmentalized protease complexes of known three-dimensional structures. A mechanism of action for PhTET1 based on its peculiar compartmentalization and structural features is proposed.

## EXPERIMENTAL PROCEDURES

### Protein Expression, Purification, and Crystallization

The expression of the PH0519 protein from *P. horikoshii* in *Escherichia coli*, its purification, and the crystallization conditions are described elsewhere (22).

### Electron Microscopy

**Generation of the Model**—We started with images obtained by negatively staining a PhTET1 sample with sodium silicotungstate, pH 7.5. Two different types of particles were observed. The images of the smaller particle were similar to the images of halophilic TET aminopeptidase (15). For the larger particle, at least two different characteristic views are easy to recognize: one ring made by six spots of density and a wheel-like structure (Fig. 1A). Two-dimensional averaging of 20 particles from each type, respectively, exhibit 3- and 4-fold symmetry axes. To generate a starting three-dimensional model, the same technique as described previously (15) was used but starting with the two-dimensional averages shown in Fig. 1A.

**Cryo-electron Microscopy**—Grids were prepared as described (23) and observed with a GATAN cryotransfer stage on a Philips CM200 LaB6 microscope at 200 kV. Images were recorded on Kodak SO 163 film under low dose conditions ( $<10 \text{ e}^-/\text{\AA}^2$ ) at a nominal magnification of 50,000 times with defocus values between 1.2 and 2.5  $\mu\text{m}$ . Good micrographs were digitized on a Zeiss Photoscan TD scanner with a pixel size of 14  $\mu\text{m}$  (2.8  $\text{\AA}$  pixel $^{-1}$ ). Using x3d (24), 11,400 small particles and 8,500 big particles from 15 images were selected and interactively boxed into  $128 \times 128$  pixel boxes. These images were then corrected for the contrast transfer function as described (15). The structures of both particles were determined using a projection matching method implemented in SPIDER (15, 25). The starting models were, respectively, the three-dimensional structure of HmTET (15) and the octahedral particle determined using negative staining (see below). For the tetrahedral particle, 159 orientations equally distributed in the tetrahedral asymmetric unit (15) were used, and for the octahedral form, 58 different projections covering 1/24 of the space. The refinement (projection matching) was carried out using the phase flipped data set until no further changes were visible in the map (100

cycles for PhTET1-12s and 50 cycles for PhTET1-24s). The final reconstructions were calculated by back projection using the fully contrast transfer function corrected set of images and selecting about 75% of the images that had the highest correlation coefficient against the reprojected of the model (PhTET1-12s, 8,000 particles; PhTET1-24s, 6,000 images). The resolution of the two maps (PhTET1-12s, 14  $\text{\AA}$ ; PhTET1-24s, 15  $\text{\AA}$ ) was determined by splitting the data randomly into two halves and by comparing the resulting reconstructions by Fourier Shell Correlation (26) at 0.3 level (see Supplemental Material and supplemental Fig. 1).

### Diffraction Data Collection and Structure Determination of PhTET1-12s

X-ray diffraction data were collected from a single selenium-substituted PhTET1-12s crystal on the ID14-4 beam line at the European Synchrotron Radiation Facility. The data frames were processed with XDS (27) and further processed and scaled using SELDAT, SCALKB2, and KBAPLY from the BioMol software suite.<sup>6</sup> While structure determination was being attempted using the MAD method, a structural model became available in the Protein Data Bank with the three-dimensional structure of *Bacillus subtilis* YsdC<sup>7</sup> (Protein Data Bank code 1VHE). This protein shares 36.3% sequence identity with the PhTET1 subunit. The three-dimensional structure of PhTET1-12s was solved at 3.1  $\text{\AA}$  resolution by molecular replacement (28). The calculations were performed with PHASER (29) using all available diffraction data. The molecular replacement solution has a *Z* value of 9.62, a log-likelihood gain value of 35.93, and a proper packing inside the F4(1)32 cell, which shows the arrangement of the PhTET1-12s subunits into tetrahedral homo-dodecameric particles. Model refinement is described in the Supplemental Material. The refined model has residual indices  $R_f = 25.6\%$  and cross-validation  $R_{\text{free}} = 33.1\%$ , together with satisfactory geometry. Statistics of data collection and processing, model refinement, and model geometry are listed in the supplemental Table 1. Model quality and stereochemistry were checked using CNS (30) and PROCHECK (31). Molecular surfaces and electrostatic potentials were computed using Adaptive Poisson-Boltzmann Solver (32). The figures were prepared with PyMOL (33).

### Quasi-atomic Models

Docking the dodecamer obtained by x-ray into the tetrahedral particle was done with SITUS (34) (unique result with a cross-correlation coefficient of 0.35). The fitting into the octahedral particle was done using URO (35) as described (23). The x-ray structure of one monomer was manually placed into the density corresponding to one monomer of the octahedral EM map that makes up the asymmetric unit. URO (35) was used to generate a preliminary quasi-atomic model of the octahedral particle by applying all the internal symmetries. The optimization procedure was also performed by using URO (35). The quality of the final fit is given by a correlation coefficient of 0.98,

<sup>6</sup> Groningen Protein Crystallography Group, unpublished data.

<sup>7</sup> Structural GenomiX Inc., San Diego, unpublished results.

a crystallographic *R* factor of 16.2% and a quadratic misfit of 1.3. The fitted scale factor was 0.91.

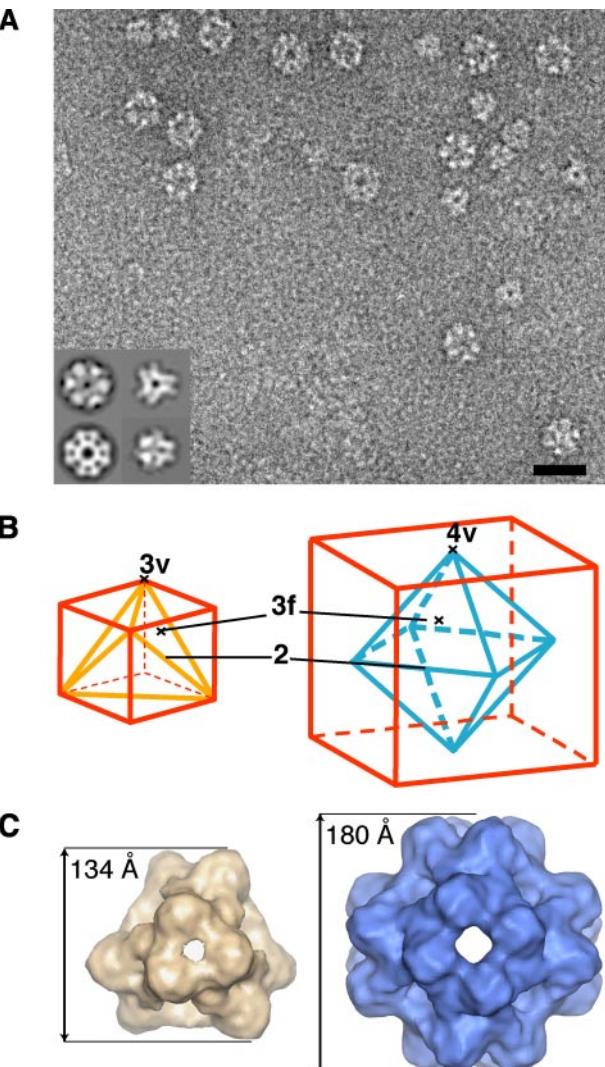
### PhTET1 Activity Assays

PhTET1 hydrolytic activity was determined as follows. Reactions were initiated by addition of the enzyme (5  $\mu$ g/ml, in the case of PhTET1-12s, or 10  $\mu$ g/ml, for PhTET1-24s) to a pre-warmed mixture containing the potential substrate in 50 mM PIPES, 150 mM KCl, 1 mM  $\text{CoSO}_4$ , pH 7.5, and covered by a layer of mineral oil to avoid water evaporation. Incubations were performed at 90 °C. To measure the evolution of the quantity of amino acids and di- and tripeptides in the reaction mixture, 80- $\mu$ l aliquots were removed from the reaction mixture at 5-min intervals and added to 200  $\mu$ l of cold acetonitrile to stop the reaction and precipitate the protein, which was separated by centrifugation. Two hundred fifty microliters of supernatant were dried under vacuum. Solids were dissolved in the same volume of Beckman analysis buffer, and the amino acids and small peptides present in each sample were quantified on a model 7300 Beckman amino acid analyzer, using the standard sodium citrate elution buffer system. To follow the progress of peptidyl- and aminoacyl-*p*NA compounds in the reaction mixture, the solids remaining after drying were dissolved in 125  $\mu$ l of chromatography initial solvent (eluent A). Fifty microliters were injected on a  $\mu$ RPC C2/C18 ST column (4.6  $\times$  100 mm) (Amersham Biosciences) in an ÄKTA purifier system (Amersham Biosciences). Chromatographic conditions were dependent of the peptide mixture, but in all cases eluent A contained 0.065% (v/v) trifluoroacetic acid and 2% (v/v) acetonitrile in water, and eluent B contained 0.05% (v/v) trifluoroacetic acid in 80% (v/v) acetonitrile. Flow was adjusted at 0.5 ml/min, and runs were carried out at room temperature. Calibrations were made by using authentic standards. Each sample was accompanied by an appropriate enzyme blank. Direct assay of generated *p*NA and AMC is described in the Supplemental Material.

## RESULTS

**Cryo-electron Microscopy Revealed Tetrahedral and Octahedral Hollow Particles**—In gel filtration, the PhTET1 protein yields two peaks with apparent molecular masses of 400 and 800 kDa, respectively (22). Analytical ultracentrifugation studies, performed on the partially purified complexes, showed that in addition to a particle with a sedimentation coefficient  $s_{20,w}$  of 15 S, *i.e.* the same value as dodecameric PhTET2 (18), a large PhTET1 particle was present. This particle displayed an  $s_{20,w}$  of 23 S, consistent with an assembly of 24 subunits (results not shown).

The PhTET1 complexes were analyzed by negative staining electron microscopy showing two populations of homogeneous particles (Fig. 1A). Cryo-electron microscopy images were obtained, and their respective quaternary structures were determined by image analysis and three-dimensional reconstruction. The three-dimensional structures of the tetrahedral and octahedral particles, at resolutions of  $\sim$ 15 Å, are shown in Fig. 1C, *left* and *right* panel, respectively. The 12-subunit assembly is a tetrahedron of largest dimension of  $\sim$ 134 Å, which is similar to the HmTET protease (15) and to the TET2 homolog from *P. horikoshii* (18–20). Four 20 Å openings (called 3f) are

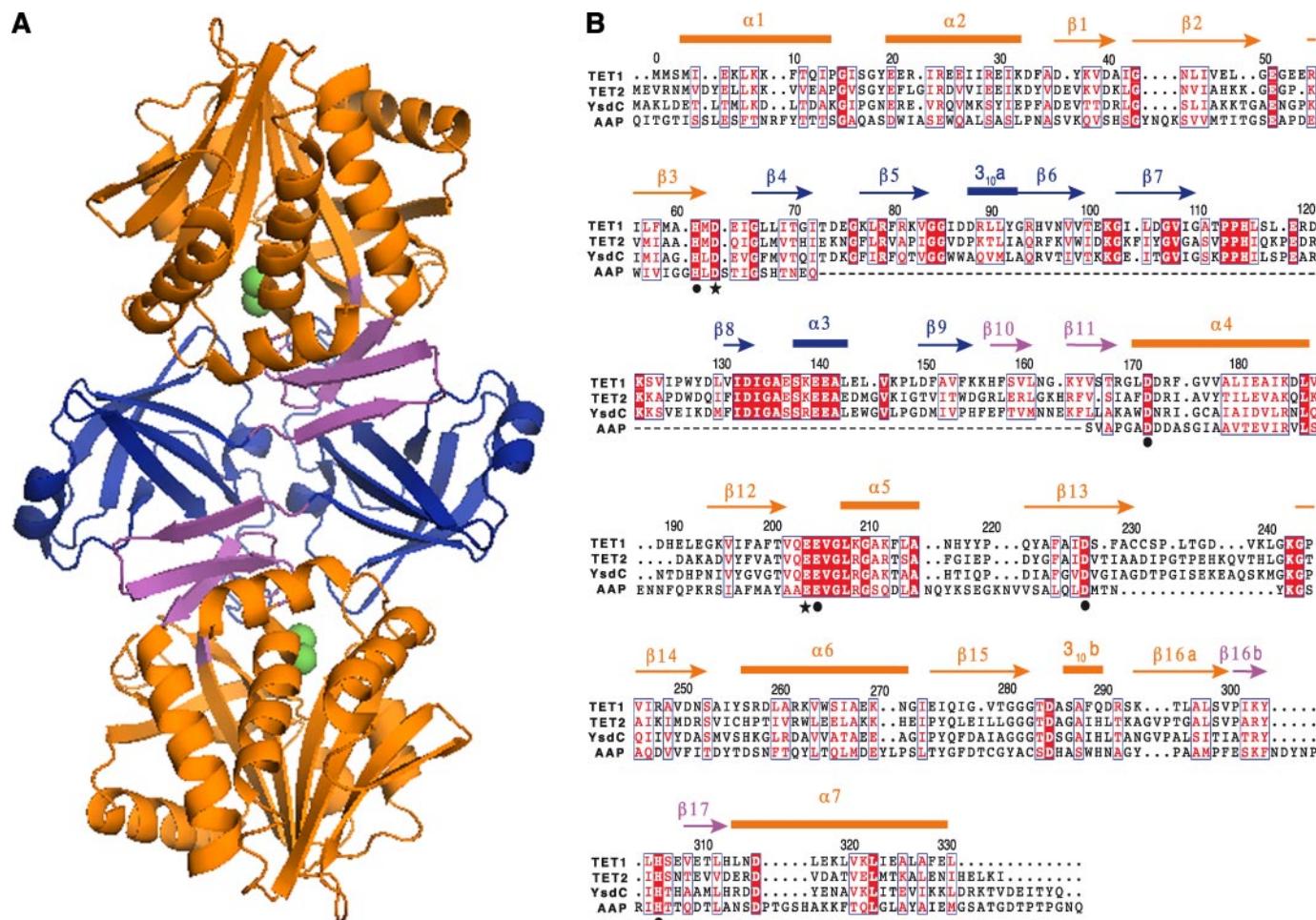


**FIGURE 1. Electron microscopy of PhTET1.** *A*, image in negative stain (sodium silicotungstate) of PhTET1 showing two types of particles. *Bottom left*, two-dimensional averages of 20 characteristic views of the large and the small particles. The large particle exhibits 3-fold (top) and 4-fold (bottom) symmetry axes. The small particle is similar to HmTET (15). The scale bar represents 20 nm. *B*, schematic views of a tetrahedron and an octahedron inscribed in a cube. The symmetry axes are indicated and called 3v or 3f for 3-fold on the vertex and on the facet, respectively. The abbreviation 4v is for 4-fold symmetry axis on the vertex. *C*, *left panel*, three-dimensional reconstruction of PhTET1-12s at 14 Å resolution viewed down the 3-fold symmetry axis on the vertex (3v); the diameter of the particle is 134 Å. *Right panel*, three-dimensional reconstruction of PhTET1-24s at 15 Å resolution viewed down the 4-fold symmetry axis on the vertex (4v); the diameter of the particle is 180 Å.

situated in each face of the tetrahedron, and four 15 Å pores (3v) are located at each vertex (Fig. 1, *B* and *C*). The tetracosameric assembly is a hollow octahedron that occupies a spherical volume of 180 Å diameter. Two types of holes penetrate the particle: six openings of 25 Å diameter at the vertices or 4-fold symmetry axes of the octahedron (4v) and eight orifices of 20 Å diameter in the center of the facets or 3-fold symmetry axes of the octahedron (3f) (Fig. 1, *B* and *C*).

**Topology of the PhTET1 Subunit**—We crystallized the 12-subunit tetrahedral complex. The three-dimensional structure of PhTET1-12s was solved at 3.1 Å resolution by molecular replacement, using the 1.9 Å resolution model of

## Peptidase Tetrahedral and Octahedral Assemblies



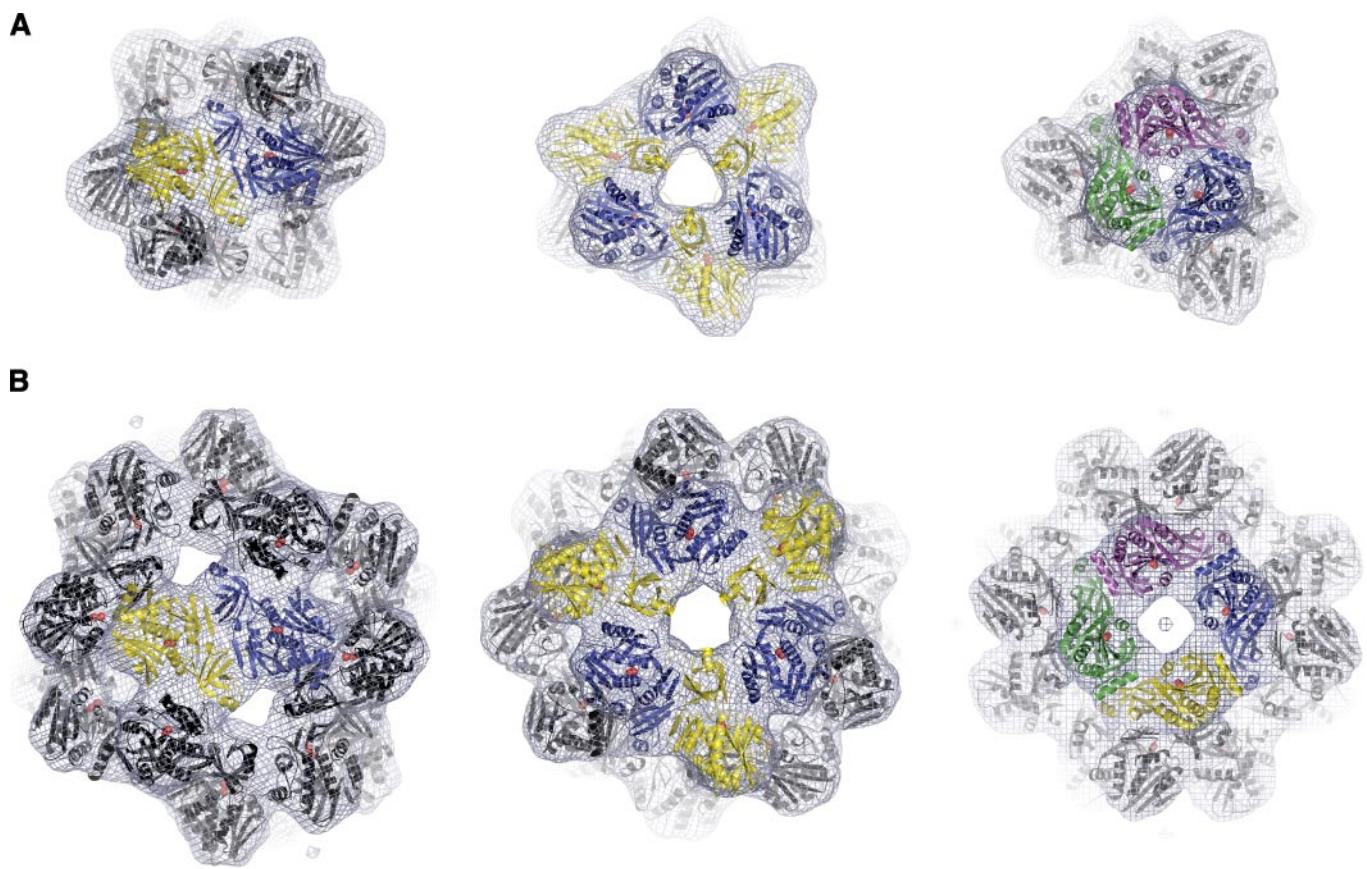
**FIGURE 2. Secondary structure elements of PhTET1.** *A*, schematic representation of the PhTET1-12s dimer along the 2-fold symmetry axis, as it appears viewed from the exterior of the dodecameric complex (see Fig. 3*A*, *left*). In each monomer, catalytic domains are colored in *orange* and *magenta*, dimerization domains in *blue*, and cobalt ions in *green*. Dimerization is achieved through formation of contacts between the dimerization domain of one monomer and a mixed four-stranded  $\beta$ -sheet (*magenta*) present in the catalytic domain of the other monomer. *B*, multiple sequence alignment indicating the secondary structure elements in PhTET1, assigned with DSSP (38), colored as in *A*. The abbreviations and GenBank™ accession numbers of the sequences are as follows: TET1, *P. horikoshii* TET1 (AP000002); TET2, *P. horikoshii* TET2 (AP000006); YsdC, *B. subtilis* YsdC protein (Z75208, Z99118); and AAP, aminopeptidase Ap1 from *Vibrio proteolyticus* (M85159, Z11993). The conserved metal-binding residues are indicated by a *ball* and the catalytic residues by a *star* (according to the MEROPS data base assignment).

the monomer from the homologous *B. subtilis* protein YsdC<sup>7</sup> (Protein Data Bank code 1VHE) as the search probe. The crystal asymmetric unit contains one monomer of PhTET1 of  $\sim 37$  kDa. The electron density maps showed no density for the first two N-terminal SeMet residues; the model thus includes only the 330 C-terminal residues of the polypeptide. The subunit is a wedge-shaped parallelepiped of dimensions  $\sim 71 \times 47$  Å, arranged in two well separated domains (Fig. 2*A*) as follows: (i) a larger catalytic domain, comprising residues 2–65 and 156–331, and (ii) a smaller dimerization domain (residues 66–153).

The catalytic domain is a flattened globule with the topology of a three-layer  $\alpha\beta\alpha$  sandwich arranged around a central, mixed eight-stranded  $\beta$ -sheet (Fig. 2*A*, *orange*). In addition, a supplementary mixed four-stranded  $\beta$ -sheet formed by strands  $\beta$ 10,  $\beta$ 11,  $\beta$ 17, and the bent C-terminal end of strand  $\beta$ 16 ( $\beta$ 16b) of the central  $\beta$ -sheet (Fig. 2*A*, *magenta*) is present at the periphery of the catalytic domain, adjacent to the smaller dimerization domain. This additional motif interacts with the dimerization domain of a second subunit to form the dimer

(Fig. 2*A*; supplemental movie 1). Using the coordinates for the catalytic domain only, a similarity search for other three-dimensional structures using the protein structure comparison server SSM (36) returned a positive match to the three-dimensional structure of AAP (37). The amino acid sequence alignments of PhTET1, PhTET2, *Bacillus subtilis* YsdC, and AAP are shown in Fig. 2*B*. AAP lacks almost all the elements of the dimerization domain, in agreement with its monomeric organization. A three-dimensional structure superposition between the 0.95 Å resolution model of AAP (Protein Data Bank code 1RTQ) and the PhTET1 subunit gives a root mean square difference of 1.08 Å for 138 equivalent pairs of  $\alpha$ -carbon. This confirms the postulated structural similarity between PhTET1 and clan MH aminopeptidases (39).

The crescent-shaped dimerization domain (Fig. 2*A*, *blue*) has a mostly  $\beta$ -barrel topology. It is made up of a six-stranded antiparallel  $\beta$ -barrel, completed by two helical segments on the outside of the barrel. The antiparallel  $\beta$ -barrel and one of the helices form the body and one of the tips of the crescent. The other tip is formed by the second helix and by a



**FIGURE 3. Atomic and quasi-atomic model of the two PhTET1 particles.** *A*, atomic model of the PhTET1-12s complex, as determined by x-ray crystallography, fitted as a dodecamer into the EM envelope. From *left* to *right*, the edifice is viewed down 2-fold and two opposite sides of 3-fold symmetry axes. The EM envelope is depicted as a *mesh*, and the cobalt ions are represented by *red spheres*. The interactions between two monomers on the 2-fold symmetry axis (*left*), between six monomers on a facet (*middle*), and between three monomers on a vertex of the particle (*right*) are shown. *B*, quasi-atomic model of the PhTET1-24s particle depicting the 24 monomers fitted into the EM envelope. From *left* to *right*, the assembly is viewed down 2-, 3-(facet), and 4-fold (vertex) symmetry axes. The EM envelope and cobalt ions are represented as in *A*. The monomers interacting on the 2-, 3- and 4-fold symmetry axes are highlighted.

disordered loop (residues 111–130), where discontinuous electron density is observed for residues 115–122.

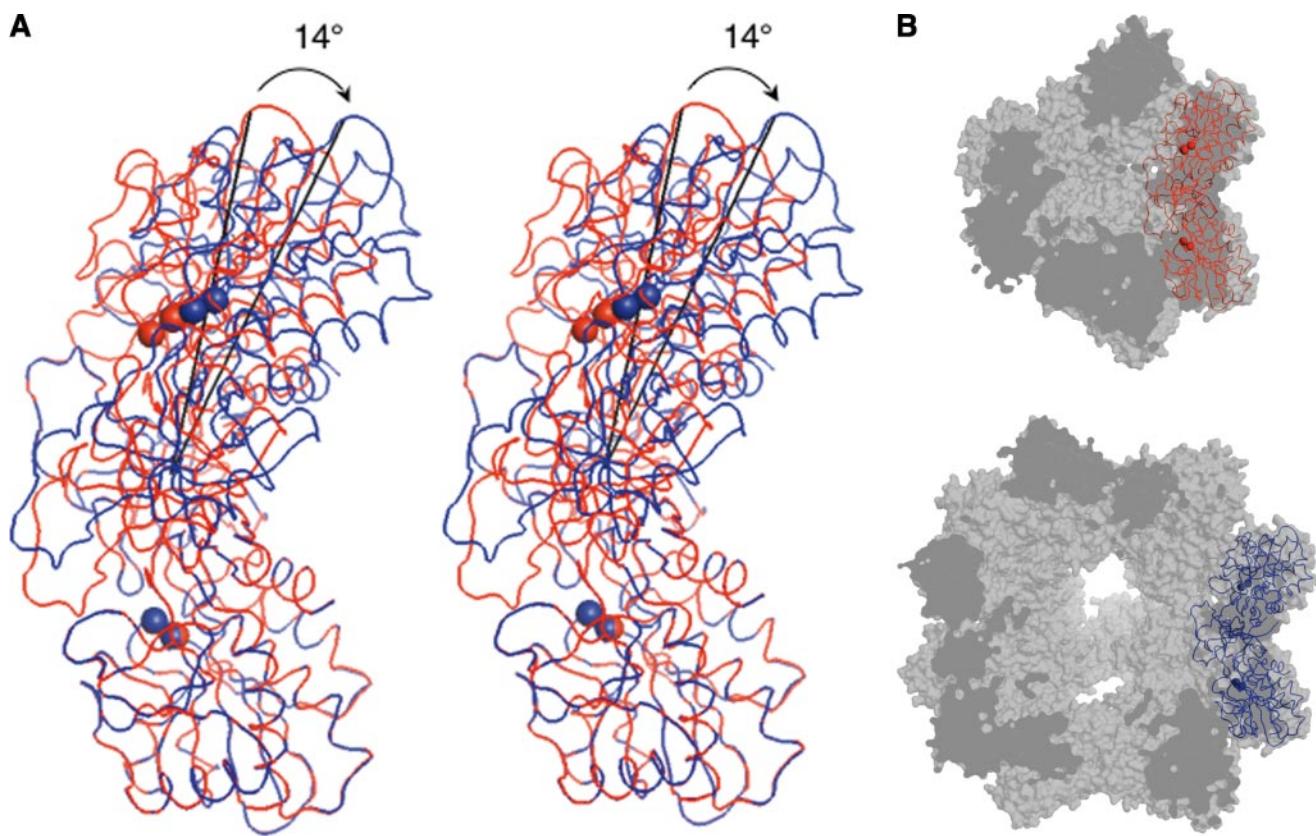
The initial difference electron density map showed, in the active site region, a very large positive density feature. This was interpreted as caused by the presence of metal ions in the active site. With divalent cobalt present in the crystallization solutions, the positive density disappeared when two divalent cobalt ions were modeled. The metal ligands are identical to those present in AAP (37) and PhTET2 (19, 20), with an identical type of coordination. Therefore, the catalytic mechanism of hydrolysis by AAP, PhTET2, and PhTET1 should be extremely similar, if not identical (19, 20, 40).

The tertiary structures of PhTET1 and PhTET2, although very similar, are not identical. PhTET2 has 22 amino acids more than PhTET1, and consequently, there are several secondary structure elements in PhTET2 that are not found in PhTET1. In particular, and following the assignment of Russo and Bauermann (19), PhTET2 contains five  $3_{10}$  helices ( $\eta_1$ ,  $\eta_2$ ,  $\eta_4$ ,  $\eta_5$ , and  $\eta_6$ ), one loop (between  $\eta_3$  and  $\beta_6$ ), and one  $\beta$ -strand ( $\beta_{15}$ ) more than PhTET1. Moreover, the  $3_{10}b$  helix in PhTET1 is an  $\alpha$ -helix ( $\alpha_7$ ) in PhTET2, and some common elements are formed by a different number of amino acids.

The crystal structure of PhTET2 contains an S1 hydrophobic specificity pocket adjacent to the metal-binding site (20), which explains the preferential activity of this aminopeptidase on pep-

tides having an N-terminal hydrophobic or noncharged residue (18). The PhTET2 specificity pocket is lined with side chains of  $Ile^{238}$ ,  $Leu^{293}$ ,  $Thr^{237}$ ,  $Glu^{291}$ ,  $Lys^{261}$ ,  $Ile^{322}$ , and  $Gly^{296}$ . Part of this pocket ( $Leu^{293}$ ,  $Ile^{238}$ , and  $Gly^{296}$ ) is involved in the binding of the inhibitor amastatin (20). The corresponding pocket in PhTET1 is constituted by the stretch  $Gly^{280}$ – $Gly^{281}$ , the guanidinium moiety of  $Arg^{248}$ , and the main chains of  $Phe^{229}$ ,  $Ala^{230}$ , and  $Thr^{280}$ . It is therefore less hydrophobic than the PhTET2 specificity pocket. In the dodecameric particle, this region of the active site also includes residues 114–118, belonging to the disordered loop of another subunit, which form a positively charged patch. These differences in the active site pockets and subpockets can account for the differences found in the substrate specificity for the two enzymes (see Refs. 21 and 18; see below).

*Quaternary Structures of PhTET1; the Subunits Can Assemble into Two Alternative Arrangements*—The three-dimensional crystallographic model of the PhTET1-12s subunit was fitted into the cryo-EM density maps of the 12- and 24-subunit assemblies to obtain the quasi-atomic structures of the complexes (Fig. 3). The PhTET1-12s EM model at 14 Å resolution is identical to the crystal structure. Octahedra and tetrahedra are Platon solids that exhibit triangular facets related by a 2-fold symmetry axis. The difference between these two volumes is located on their vertices (4- and 3-fold symmetry axes, respec-



**FIGURE 4. Dimer structure in the two PhTET1 complexes.** *A*, ribbon representation of a dimer extracted from the tetracosamer quasi-atomic model (blue) superimposed to a dimer from the dodecamer x-ray structure (red). Cobalt ions are shown as spheres. The bottom monomers of each dimer were forced to match, hence the stereo picture illustrates the rotation of  $\sim 14^\circ$  of the upper PhTET1-24s monomer with respect to the center of mass of the PhTET1-12s dimer. *A* and Fig. 2*A* are related by a  $90^\circ$  rotation around the vertical axis. *B*, cut open surface representations of the PhTET1 edifices showing the position of one dimer in the tetrahedral complex (*top*) and in the octahedral assembly (*bottom*). The dimers are depicted and oriented as in *A*.

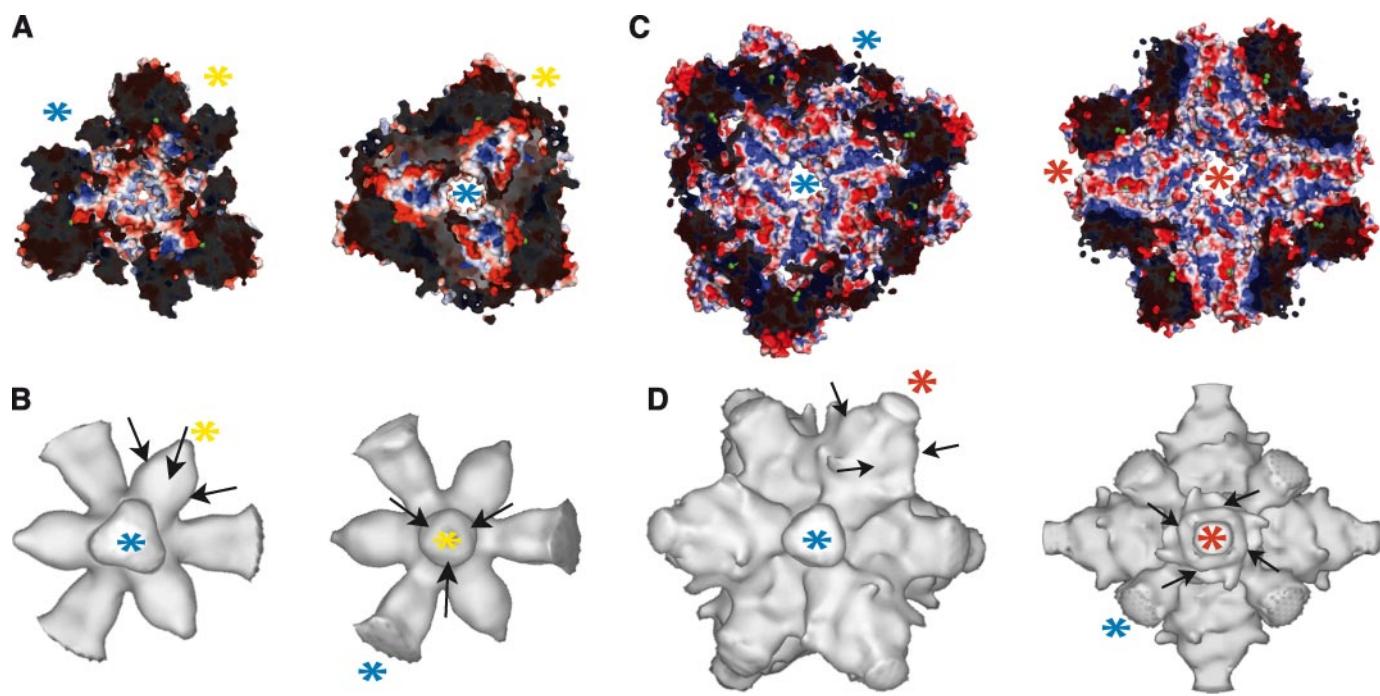
tively; Fig. 1*B*). The building blocks used to generate both types of particles can therefore either be dimers (lying on each edge) or trimers (forming each faces of the tetrahedron or octahedron). The atomic structure of the tetrahedron and the calculated quasi-atomic model of the octahedron show that each facet is generated using a combination of six monomers (Fig. 3; supplemental movies 1 and 2). The building block for both structures is thus more likely to be the dimer.

On each tetrahedron vertex (3v), three subunits are arranged as trimers in a circular fashion. The junction found in each trimer delimits the smallest of the two types of holes present in the particle (Fig. 3*A*, *right*). The distance between the OH group of the successive Tyr<sup>219</sup> residues situated at the border of the hole is  $\sim 7$  Å. On the tetrahedron faces (3f), wider openings ( $\sim 13$  Å between the O-82 atoms of successive Asp<sup>74</sup> residues) are located at the junction between six subunits (Fig. 3*A*, *middle*). These faces appear to be similar to those of the octahedron (Fig. 3*B*, *middle*). However, in the octahedral particle, the association that defines the apex small holes involves four PhTET1 monomers instead of three (Fig. 3*B*, *right*). The consequence is the generation of six wider (20 Å) novel types of openings (4v).

**PhTET1 Assembles into Tetrahedral and Octahedral Particles Using Quasi-equivalent Contacts That Are Reminiscent of Those Found for Virus Capsids**—In the two different quaternary arrangements of PhTET1, the dimeric building blocks are situated across the 2-fold symmetry axes, forming the assembly

edges. Nevertheless, these building blocks are differently structured in the octahedral and the tetrahedral particles (Fig. 4). A rotation of  $\sim 14^\circ$  of one of the monomers forming the dimer with reference to the center of mass of the PhTET1-12s dimer is evident in the building block of the tetracosameric assembly (Fig. 4*A*). In the two types of PhTET1 particles, the interactions between monomers to form the dimer involve the dimerization domain plus the four-stranded  $\beta$ -sheet present at the surface of the catalytic domain, which face each other in a head to tail fashion (Fig. 2*A*). However, the rigid body fitting of the PhTET1-12s monomer into the EM PhTET1-24s structure shows that the dimer making up the octahedron contains fewer contacts between the two monomers. In particular, strands  $\beta$ 5,  $\beta$ 14, and  $\beta$ 15 participate in dimerization in the PhTET1-12s building block, but they do not seem to contribute to any intermonomer contact in the dimer from the 24-subunit complex. This allows the  $14^\circ$  rotation, which tilts one monomer with respect to the other.

This rotation has two important consequences. First, since the angle between the dimer monomers decreases in the side oriented toward the exterior of the particle (Fig. 4*B*), the resulting dimers when positioned in the final complex give rise to wider apex holes, allowing the insertion of a new dimer. This modification changes the geometry of the particle from tetrahedral to octahedral, and doubles the number of subunits of the assembly. The insertion of the new dimer is



**FIGURE 5. Inside the PhTET1 complexes.** *A*, two halves of the tetrahedral edifice in cut-open surface representation. The particle surface is color-coded according to its electrostatic potential contoured from  $-5 \text{ kT/e}$  (intense red) to  $+5 \text{ kT/e}$  (intense blue) computed at 363 K, pH 7, 150 mM ionic strength, with a solute dielectric of 2 and a solvent dielectric of 80. The assembly is viewed down an access channel leading into the positively charged catalytic chamber (left) and in the opposite sense (right). Cobalt ions are colored in green. Blue and yellow stars are used to indicate where large (3f) and small (3v) openings, respectively, are situated. *B*, representation of the interior void volume of the dodecameric complex as obtained from the electron microscopy three-dimensional reconstruction. Arrows indicate the position of three active sites in a catalytic chamber. The orientations of the figures and the star codes are the same as in *A*. *C*, two halves of the quasi-atomic model of the octahedral edifice in cut-open surface representation viewed down the 3-fold (left) and the 4-fold (right) symmetry axes. The particle surface and cobalt ions are color-coded as in *A*. Blue and red stars indicate, respectively, where facet (3f) and vertex (4v) openings are located. *D*, representation of the interior void volume of the tetracosameric complex as obtained from the electron microscopy three-dimensional reconstruction. Arrows locate the active sites in a catalytic chamber. The orientations of the figures and the star codes are the same as in *C*.

possible because of a second change in the inter-subunit contacts. In the dodecameric particle, the formation of the trimer situated at the apex of the complex is achieved by interactions involving helices  $\alpha 2$ ,  $\alpha 3$ , and  $\alpha 5$  and strands  $\beta 1$ ,  $\beta 4$ , and  $\beta 5$  from one PhTET1 subunit, and helices  $\alpha 6$  and  $3_{10}b$  and strands  $\beta 14$ ,  $\beta 15$ , and  $\beta 16a$  from the adjacent subunit. In the rigid body fitted tetracosameric particle, fewer contacts are again observed, because helices  $\alpha 2$ ,  $\alpha 5$ , and  $3_{10}b$  and strands  $\beta 1$ ,  $\beta 5$ ,  $\beta 14$ , and  $\beta 16$  do not interact to form the apical tetramer. Instead, a new contact appears to be formed in which the loop between strands  $\beta 4$  and  $\beta 5$  is implicated (not shown). To date, quasi-equivalent contacts have been mainly found in icosahedral viral particles.

The second consequence of the  $14^\circ$  rotation is caused by the relative movement away from the catalytic domain of one monomer and the oligomerization domain of the other that results in a wider access to the active site in the octahedral particle (Fig. 4*B*). This improved accessibility probably has a direct effect on the catalytic properties of the complex. This issue is currently being studied in our laboratory.

Comparison of the position and type of the inter-subunit contacts that lead to dimerization in PhTET1-12s and in PhTET2 reveals that dimer formation is not achieved in the same way in the two molecules. In particular, a cluster of interactions is present in PhTET2 that is very different in PhTET1-12s (supplemental Fig. 2). There, only one interaction (hydrogen bond or salt bridge) is found, which connects Lys<sup>77</sup> in

strand  $\beta 5$  from one subunit to Lys<sup>303</sup> in strand  $\beta 16b$  from the other subunit. This interaction is not present in the PhTET1-24s dimer. The PhTET2 cluster is situated at the periphery and is repeated twice per dimer. It includes interactions that link the loop between strands  $\beta 4$  and  $\beta 5$  (following the assignment of Russo and Baumann, see Ref. 19) and strand  $\beta 8$  belonging to one subunit to the loop between strand  $\beta 13$  and helix  $\eta 5$ , and the loop between strands  $\beta 17$  and  $\beta 18$  from the other subunit. This cluster contains numerous hydrophobic interactions, hydrogen bonds, and also a cation-electron  $\pi$  interaction. Therefore, its existence may lead to a higher rigidity of the PhTET2 dimer that could prevent the relative displacement of one monomer with respect to the other.

**The PhTET1 Tetrahedral and Octahedral Particles Define a Network of Access Channels and Catalytic Chambers**—The examination of the PhTET1-12s interior revealed a complex compartmentalization. The four large openings (3f) situated in each face of the tetrahedron give access to four wide channels that cross each other in the middle of the particle (Fig. 5, *A* and *B*; supplemental movie 1). Four funnel-shaped sub-compartments are enclosed in the apices of the complex and extend those channels (Fig. 5, *A* and *B*). These compartments are catalytic chambers each containing three active sites. Note that, with the exception of the active site pocket, the  $\sim 23 \text{ \AA}$ -long catalytic chamber is highly basic compared with the channels that are more contrasted regarding the charge of the surfaces. The four funnel-shaped catalytic chambers open up to the exte-

# Peptidase Tetrahedral and Octahedral Assemblies

TABLE 1

## Activity of PhTET1 on different peptides

Substrate peptides were used at a 5 mM initial concentration. Generation of products was quantified by reversed phase HPLC (see "Experimental Procedures").

Substrate	Product	PhTET1-12s		PhTET1-24s	
		At 5 min	At 25 min	At 5 min	At 25 min
<i>mmol of product/g of enzyme</i>					
Asp-Ala	Asp	0.60	3.0	0.24	1.18
	Ala	0.60	2.8	0.26	1.24
Ala-Ala	Ala	3.5	17.3	8.5	43
Ala-Ala-Ala	Ala-Ala	24.1	138	20.0	100
	Ala	24.3	149	20.4	141
Ala-Ala- <i>p</i> NA	Ala- <i>p</i> NA	38	81	22.5	46
	<i>p</i> NA	0	0	0.13	0
Ala-Ala-Ala- <i>p</i> NA	Ala-Ala- <i>p</i> NA	38	170	15.9	60
	Ala- <i>p</i> NA	0.54	9.9	0.15	3.7
	<i>p</i> NA	0	0	0	0

rior by the narrow holes situated at each apex. These pores are not wide enough to allow the easy passage of an extended peptide. Accordingly, these orifices appear well adapted for the expulsion of the free amino acids generated by the aminopeptidase reaction.

Each 3f channel consists of three electrostatic layers (Fig. 5A, *left*) as follows: 1) a mixed entrance layer comprising the side chains of Asp<sup>74</sup>, Glu<sup>75</sup>, and Lys<sup>240</sup>; 2) a positively charged layer formed by the side chain of Arg<sup>79</sup>; and 3) a negatively charged layer composed of the side chains of Glu<sup>119</sup> and Asp<sup>129</sup>. These channels are ~15 Å long (channel length is defined as the distance between the O-δ2 atom of Asp<sup>74</sup> and the Asp<sup>129</sup> O-δ1). The channel intersection gives rise to a basic "spherical" central layer, which consists of the side chain of Lys<sup>122</sup>. From there, the positively charged catalytic chambers can be accessed. These compartments are lined by the side chains of Arg<sup>120</sup>, His<sup>115</sup>, Asn<sup>252</sup>, and Lys<sup>212</sup>. On the way to the small apical hole (Fig. 5A, *right*), a negatively charged layer formed by the side chain of Asp<sup>291</sup> appears before the final exit mixed layer composed of the side chains of Asn<sup>216</sup> and Tyr<sup>219</sup>. All the residues are repeated three times per circular layer, because the channels and catalytic chambers are at the interface between three PhTET1 subunits, except for the central layer in which the 12 Lys<sup>122</sup> of the structure are present.

In the octahedral particle, the number of holes situated in the facets (3f) is doubled compared with the TET1-12s particle. Although these orifices have a triangular shape similar to that found in the dodecameric complex, the quasi-atomic model of the octahedral-type organization of the subunits moves some acidic areas, placed on the surface of the tetrahedral complex, toward the inside of the particle. Hence the entrances become negatively charged (Fig. 5C; supplemental movie 2). In the channels to which these holes give access, the first two layers appear to be very similar to layers 1 and 2 of the dodecameric particle. However, unlike the 12-subunit complex, beyond these two layers the channel opens to a vast central cavity (~900 nm<sup>3</sup>), which would contain alternating positive- and negative-charged areas formed by combination of layer 3, the central layer, and part of the catalytic compartment described above.

The PhTET1-24s complex holds six catalytic chambers (each containing four active sites) that form wide dome roof-like structures where the active sites are organized as a ring (*arrows* in Fig. 5D). These catalytic compartments are mostly positively

charged like those observed in the dodecameric edifice (Fig. 5, A and C). However, in the tetracosameric complex they are broader and open to the exterior through wider orifices (20 Å).

**The Octahedral PhTET1 Particle Is Catalytically Active—**Unlike the PhTET2 protein (18), PhTET1 activity cannot be detected by using chromogenic aminoacyl compounds (AMC or *p*NA), probably because the X-AMC and X-*p*NA molecules cannot fit into the active site pocket or because these compounds fit but are not hydrolyzed. In fact, HPLC analysis of Ala-Ala-*p*NA degradation by the two PhTET1 oligomers showed that the generation of Ala-*p*NA was not linear with time (Table 1), suggesting product inhibition. Moreover, PhTET1 can cleave the N-terminal amino acid of larger peptides, even when the *p*NA moiety is in the C-terminal position, and can also hydrolyze dipeptides like Ala-Ala or Asp-Ala (Table 1). Therefore, PhTET1 seems to be unable to cleave a peptidic bond when *p*NA or AMC is in the P1' position, making activity assays difficult.

PhTET1-12s enzymatic behavior has been partially characterized by Ando *et al.* (21). As they described, PhTET1-12s activity increases with substrate chain length (Table 1). However, unlike their results, we found no hydrolytic activity toward *N*-acetyl-Leu-*p*NA (*i.e.* deblocking activity). No *p*NA was detected by spectrophotometric analysis, and no *N*-acetyl-Leu, Leu-*p*NA, nor *p*NA could be measured by reversed phase HPLC. The same applies to the compound *N*-succinyl-Ala-Ala-*p*NA, because no Ala-Ala-*p*NA, Ala-*p*NA, nor *p*NA could be detected in the reaction mixture after a long incubation with PhTET1-12s. Comparable results were obtained when the deblocking activity of TET1-24s was assayed. Therefore, the deblocking activity of the enzyme, if present, is very low, and consequently the name DAP (for deblocking aminopeptidase) is ill suited to PhTET1.

The two PhTET1 oligomeric forms are similar with respect to their activity, but some differences exist. The dodecameric complex was in general more active than the 24s-subunit edifice (Table 1). Among the assayed compounds, Ala-Ala-Ala-*p*NA was found to be the best substrate for PhTET1-12s, whereas the tetracosameric complex showed maximal activity against Ala-Ala-Ala and Ala-Ala-*p*NA.

## DISCUSSION

The quaternary structures of PhTET1 described in this paper reveal a novel type of elaborated self-compartmentalized orga-

nization for a peptidase complex. Both versions of the edifice (PhTET1-12s and PhTET1-24s) are structured in a related fashion with two types of subcompartments: the entry channels (4 and 8, respectively) and the proteolytic chambers (4 and 6, respectively). However, the dodecamer is more compact than the tetracosamer.

The tetrahedral organization of PhTET1-12s indicates a function as a peptide scavenger, the geometry and electrostatic charge repartition in the channels being involved in the substrate peptide binding and orientation toward the catalytic chambers. The channel dimensions are in the same range as those described for barrel-shaped peptidases that process unfolded polypeptides such as the 20 S proteasome (41) and ClpP (42), and are large enough to let even partially folded polypeptides (such as single helices or helical coils) enter the particle. The succession of positively and negatively charged patches present on the surface of PhTET1-12s channels would trigger the translocation of the polypeptides toward the catalytic chambers by maintaining the substrate (also containing different types of charged areas) in continuous motion. By contrast, the catalytic chambers are highly positively charged with the exception of the active sites. This feature would facilitate a greater mobility of the substrate N-terminal extremity, when it enters the catalytic compartment, because of the electrostatic repulsion, and thus would permit a flip to one of the three highly acidic bi-metallic active sites. The catalytic chambers provide a high concentration of active sites located on the same plane. Note that these compartments have a funnel-like shape that could prompt the ejection of the released free amino acids by the nearby narrow pore located at the apex of the complex (Fig. 5, *A* and *B*). The strong negative charge created by the aspartate residues (Asp<sup>291</sup>) preceding the apex openings could serve to attract the positively charged N termini of the generated amino acids, which could be then hooked up by the tyrosine residues (Tyr<sup>219</sup>) surrounding the apex holes. In fact, if an alternate conformer of the side chains of these residues is taken (together with an alternate conformer for the side chain of Asp<sup>291</sup> that does not provoke steric clashes with the remainder of the polypeptide), the exit hole becomes wider with ~16 Å between the C-β positions of two successive Tyr<sup>219</sup> side chains (supplemental Fig. 3). This tyrosine side chain swing would allow the transfer of the free amino acid to the exterior, providing a product ejection mechanism. The organization of the quaternary structure of the PhTET1-12s particle thus appears to be optimal for the elimination of the N-terminal residue from a polypeptidic chain.

The mechanism of substrate navigation based on the highly compartmentalized nature of PhTET1-12s and on its surface electrostatic potential features is clearly different from that proposed for PhTET2 (20). Such a discrepancy in the putative mode of action of the two peptidases could be explained by the differences in their respective quaternary structures. The PhTET2 particle encloses an internal cavity that appears to be an open space harboring all the catalytic sites. PhTET2 does not contain catalytic chambers as those described for PhTET1. The apices of the PhTET2 complex are almost closed, and three small holes, which were proposed to be the exit channels of the generated free amino

acids, are disposed around the pore situated in the facets of the complex (20). Those three small holes, each of which is connected directly to one active site, do not exist in the PhTET1 edifices. Consequently, despite their apparent similarity, the two peptidases have specific features regarding their quaternary structures and mode of action. Note that the enzymatic properties of the different TET proteins characterized to date are not identical; HmTET is a broad aminopeptidase (15); PhTET2 acts preferentially on neutral N-terminal residues (18), and PhTET1 is unable to cleave a peptide bond when *p*NA or AMC is in the P1' position and seems to exhibit a narrow specificity.<sup>8</sup> Nevertheless, the apical holes of the PhTET2 dodecameric complex are lined by the side chains of three phenylalanine residues (20), which could undergo a swing similar to that described for the tyrosine residue side chains in PhTET1-12s. Because PhTET2 acts preferentially on aliphatic N-terminal amino acids, a releasing strategy involving hydrophobic interactions between the side chain of the cleaved amino acid and the phenylalanine aromatic ring seems possible. In this case, the mechanisms of substrate channeling and processing, and also of product release by the two TET peptidases, would be comparable, and differences regarding the substrate specificity would be easily explainable considering the divergences existing in their specificity pockets. Although the TET peptidases can break down large peptides into free amino acids (15, 18, 21), the process is not sequential, as we demonstrated for PhTET2 (18). The lack of processivity is evident also from the analysis of Table 1, because degradation of the longest peptides involves the accumulation of shorter intermediary peptides, although they are substrates (*e.g.* Ala-Ala, Ala-Ala-*p*NA). Therefore, the peptide moiety is detached from the active site once the N-terminal residue has been cleaved, probably to allow the discharge of the free amino acid.

The Tricorn proteases (TRI) are another type of large proteolytic complex generally present in archaeal lineages where TET-like peptidases are absent. It is thus quite likely that the two systems represent functional homologs. Tricorn activity generates di- and tripeptides, and hence, the synergistic action of three aminopeptidase factors is needed to yield free amino acids (43). Biochemical studies have suggested that this modular proteolytic system would work in synergy with the proteasome (44). TRI assembles as a 720-kDa complex that is composed of six identical copies of a 120-kDa polypeptide. Despite an apparent similarity between the external shapes of the TET and TRI complexes (20), the PhTET1 structure examination reveals a completely different internal organization. The Tricorn particle is a flattened complex that encloses six active sites in a central cavity, which is accessible by two main holes (44). In the case of PhTET1 there is no cavity *per se* but a network of channels and catalytic chambers.

A striking feature of PhTET1 is its ability to assemble into two different quaternary structures, involving quasi-equivalent contacts between monomers. In viruses, this characteristic can

<sup>8</sup> M. A. Durá and B. Franzetti, unpublished results.

## Peptidase Tetrahedral and Octahedral Assemblies

result not only in differences in the contacts among capsomers in one viral particle but also in the formation of several sized particles. For example, the hepatitis B core particle can assemble into  $T = 4$  or  $T = 3$  particles comprising 120 and 90 dimers, respectively (45). The PhTET1 edifices represent the first non-filamentous and nonviral objects where the same kind of phenomena are found. Although Tricorn can form a giant icosahedral capsid, this assembly consists of 20 copies of the 720-kDa TRI complex (46), and therefore it does not imply any change in the contacts between monomers composing the TRI hexamer.

What is the role of a giant hollow octahedral peptidase, larger than the 20 S proteasome, within the cells? We have shown that the PhTET1-24s particle is nearly as active as the dodecameric form. The biological relevance of such an alternative way of assembly remains unknown. At this stage, one can only propose that a more expanded quaternary structure would facilitate the trapping of small peptides (2–6 residues) and may facilitate their free diffusion toward the active sites. The more compact, channeled tetrahedral complex would be a complementary system, streamlined for the processing of longer polypeptides because of its better capacity to guide them toward the catalytic chambers. PhTET1-12s together with PhTET2 could therefore be responsible for the complete breakdown of 6–20-residue proteasomal products, whereas PhTET1-24s could be involved preferentially in the nutritional pathway by hydrolyzing small peptides imported from the external environment. This hypothesis is supported by the fact that Pyrococcales are proteolytic Archaea and that complete DNA microarray data from *Pyrococcus furiosus* have shown that the gene encoding for a TET1 homologous protein is dramatically up-regulated in peptide-grown cells (47). Furthermore, the fold of the catalytic domain of PhTET1 resembles that of AAP and *Streptomyces griseus* aminopeptidase (48), two peptidases from family M28 in clan MH. This family contains several bi-metallic secreted aminopeptidases. All of them are monomeric. In the case of the TET peptidases, the dimerization domain was probably added subsequently during evolution to prevent unwanted proteolytic damage, allowing their cytosolic localization.

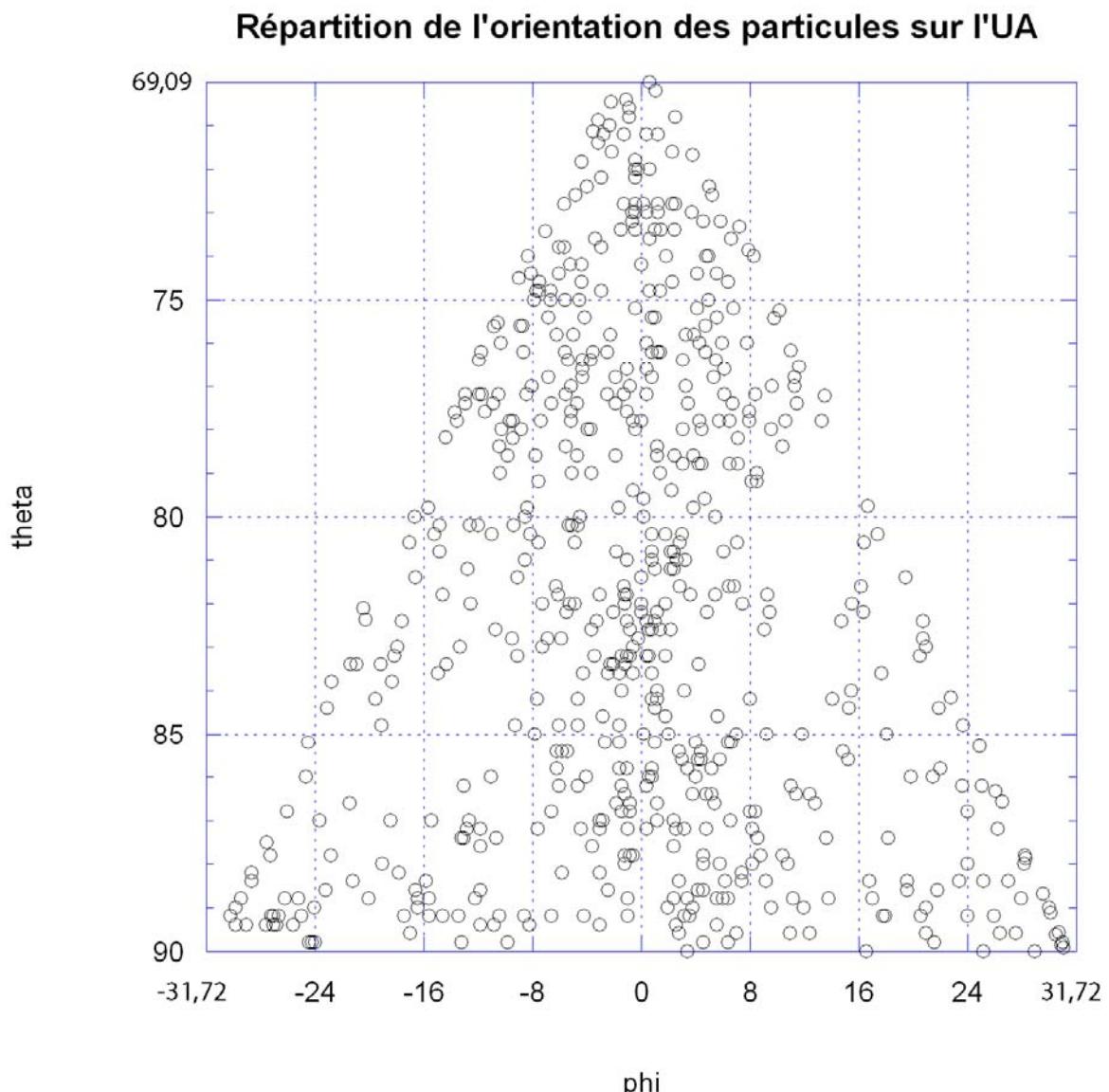
**Acknowledgments**—We thank ID14-4 beam-line staff at the European Synchrotron Radiation Facility (Grenoble), Rémi Pinck (EMBL-Grenoble), and Jacques-Philippe Colletier (Laboratoire d'Enzymologie Moléculaire-Institut de Biologie Structurale) for computational support; Dr. Elizabeth Hewat (Laboratoire de Microscopie Electronique Structurale-Institut de Biologie Structurale) and the Institut de Biologie Structurale for the use of the Philips CM200 cryo-microscope and image scanner; Jean-Pierre Andrieu (Laboratoire d'Enzymologie Moléculaire-Institut de Biologie Structurale) for amino acid analysis and computational support; and Dr. Richard H. Wade (Laboratoire de Microscopie Electronique Structurale-Institut de Biologie Structurale) for critical reading of the manuscript.

## REFERENCES

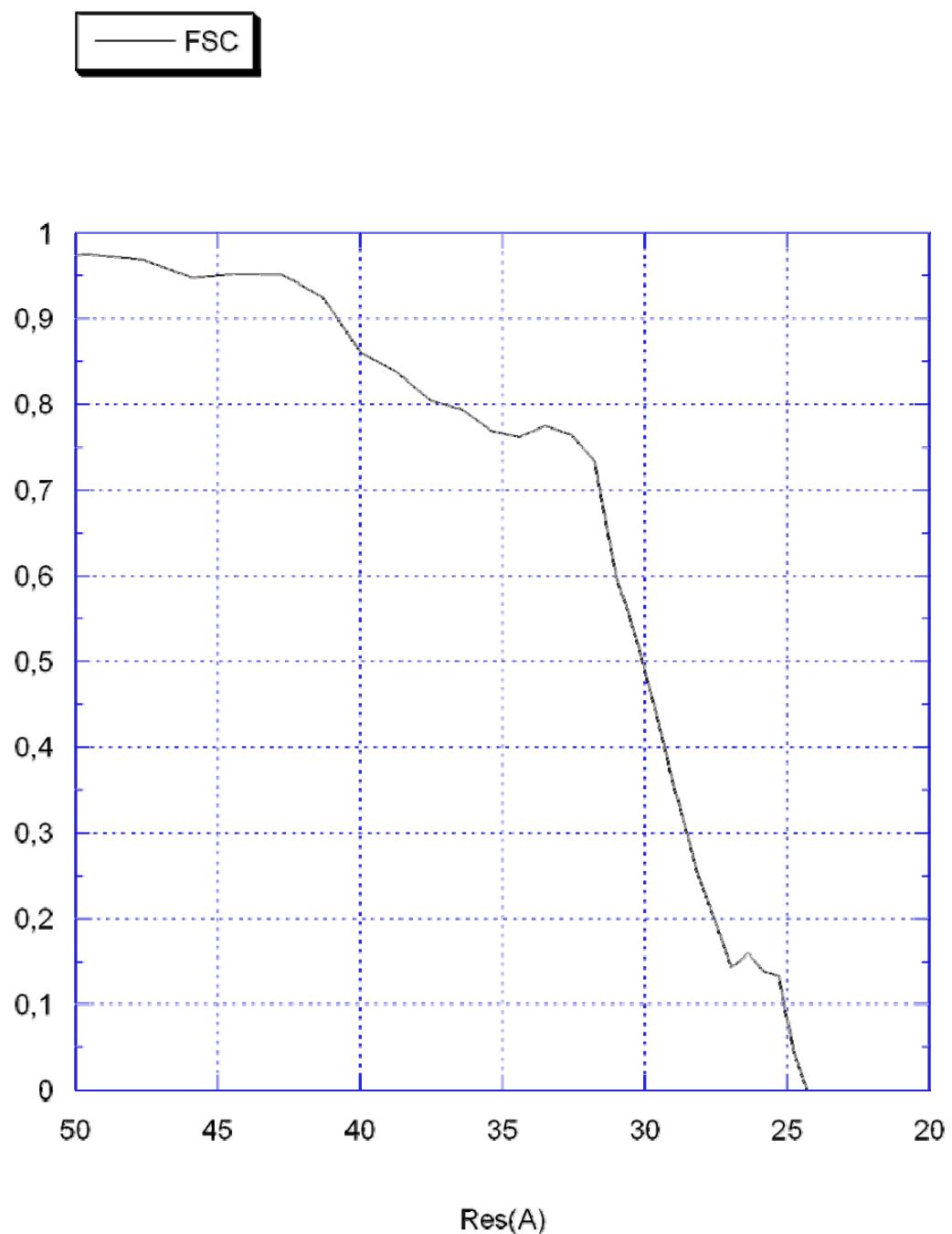
1. Kirschner, M. (1999) *Trends Cell Biol.* **9**, M42–M45
2. Gottesman, S. (2003) *Annu. Rev. Cell Dev. Biol.* **19**, 565–587
3. Baumeister, W., Walz, J., Zühl, F., and Seemüller, E. (1998) *Cell* **92**, 367–380
4. Hershko, A., and Ciechanover, A. (1998) *Annu. Rev. Biochem.* **67**, 425–479
5. Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A., Hersch, G. L., Joshi, S. A., Kenniston, J. A., Levchenko, I., Neher, S. B., Oakes, E. S., Siddiqui, S. M., Wah, D. A., and Baker, T. A. (2004) *Cell* **119**, 9–18
6. Kisselov, A. F., Akopian, T. N., Woo, K. M., and Goldberg, A. L. (1999) *J. Biol. Chem.* **274**, 3363–3371
7. Botbol, V., and Scornik, O. A. (1991) *J. Biol. Chem.* **266**, 2151–2157
8. Gonzales, T., and Robert-Baudouy, J. (1996) *FEMS Microbiol. Rev.* **18**, 319–344
9. Tamura, T., Tamura, N., Cejka, Z., Hegerl, R., Lottspeich, F., and Baumeister, W. (1996) *Science* **274**, 1385–1389
10. Burley, S. K., David, P. R., Sweet, R. M., Taylor, A., and Lipscomb, W. N. (1992) *J. Mol. Biol.* **224**, 113–140
11. Joshua-Tor, L., Xu, H. E., Johnston, S. A., and Rees, D. C. (1995) *Science* **269**, 945–950
12. Remaut, H., Bompard-Gilles, C., Goffin, C., Frère, J.-M., and van Beeumen, J. (2001) *Nat. Struct. Biol.* **8**, 674–678
13. Rockel, B., Peters, J., Muller, S. A., Seyit, G., Ringler, P., Hegerl, R., Glaeser, R. M., and Baumeister, W. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10135–10140
14. Yao, T., and Cohen, R. E. (1999) *Curr. Biol.* **9**, R551–R553
15. Franzetti, B., Schoehn, G., Hernandez, J. F., Jaquinod, M., Ruigrok, R. W., and Zaccai, G. (2002) *EMBO J.* **21**, 2132–2138
16. Rawlings, N. D., Tolle, D. P., and Barrett, A. J. (2004) *Nucleic Acids Res.* **32**, D160–D164
17. Gonzalez, J. M., Masuchi, Y., Robb, F. T., Ammerman, J. W., Maeder, D. L., Yanagibayashi, M., Tamaoka, J., and Kato, C. (1998) *Extremophiles* **2**, 123–130
18. Durá, M. A., Receveur-Brechot, V., Andrieu, J. P., Ebel, C., Schoehn, G., Roussel, A., and Franzetti, B. (2005) *Biochemistry* **44**, 3477–3486
19. Russo, S., and Baumann, U. (2004) *J. Biol. Chem.* **279**, 51275–51281
20. Borissenko, L., and Groll, M. (2005) *J. Mol. Biol.* **346**, 1207–1219
21. Ando, S., Ishikawa, K., Ishida, H., Kawarabayashi, Y., Kikuchi, H., and Kosugi, Y. (1999) *FEBS Lett.* **447**, 25–28
22. Porciero, S., Receveur-Brechot, V., Mori, K., Franzetti, B., and Roussel, A. (2005) *Acta Crystallogr. Sect. F Struct. Biol. Crystalliz. Comm.* **61**, 239–242
23. Fabry, C. M. S., Rosa-Calatrava, M., Conway, J. F., Zubieta, C., Cusack, S., Ruigrok, R. W., and Schoehn, G. (2005) *EMBO J.* **24**, 1645–1654
24. Conway, J. F., and Steven, A. C. (1999) *J. Struct. Biol.* **128**, 106–118
25. Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladadji, M., and Leith, A. (1996) *J. Struct. Biol.* **116**, 190–199
26. van Heel, M. (1987) *Ultramicroscopy* **48**, 95–100
27. Kabsch, W. (1988) *J. Appl. Crystallogr.* **21**, 916–924
28. Rossmann, M. G. (1972) in *The Molecular Replacement Method* (Rossmann, M. G., ed) Gordon and Breach Science Publishers, Inc., New York
29. Read, R. J. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 1373–1382
30. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grossé-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 905–921
31. Laskowski, L. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
32. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–10041
33. DeLano, W. L. (2002) *The PyMOL User's Manual*, DeLano Scientific, San Carlos, CA
34. Wriggers, W., and Birmanns, S. (2001) *J. Struct. Biol.* **133**, 193–202
35. Navazza, J., Lepault, J., Rey, F. A., Álvarez-Rúa, C., and Borge, J. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 1820–1825
36. Krissinel, E., and Henrick, K. (2004) in *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268
37. Chevrier, B., Schalk, C., D'Orchymont, H., Rondeau, J. M., Moras, D., and Tarnus, C. (1994) *Structure (Lond.)* **2**, 283–291
38. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577–2637
39. Onoe, S., Ando, S., Ataka, M., and Ishikawa, K. (2002) *Biochem. Biophys. Res. Commun.* **290**, 994–997
40. Stamper, C., Bennett, B., Edwards, T., Holz, R. C., Ringe, D., and Petsko, G.

- (2001) *Biochemistry* **40**, 7035–7046
41. Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995) *Science* **268**, 533–539
42. Wang, J., Hartling, J. A., and Flanagan, J. M. (1997) *Cell* **91**, 447–456
43. Tamura, N., Lottspeich, F., Baumeister, W., and Tamura, T. (1998) *Cell* **95**, 637–648
44. Kim, J.-S., Groll, M., Musiol, H. J., Behrendt, R., Kaiser, M., Moroder, L., Huber, R., and Brandstetter, H. (2002) *J. Mol. Biol.* **324**, 1041–1050
45. Wingfield, P. T., Stahl, S. J., Williams, R. W., and Steven, A. C. (1995) *Biochemistry* **34**, 4919–4932
46. Walz, J., Koster, A. J., Tamura, T., and Baumeister, W. (1999) *J. Struct. Biol.* **128**, 65–68
47. Schut, G. J., Brehm, S. D., Datta, S., and Adams, M. W. W. (2003) *J. Bacteriol.* **185**, 3935–3947
48. Greenblatt, H. M., Almog, O., Maras, B., Spungin-Bialik, A., Barra, D., Blumberg, S., and Shoham, G. (1997) *J. Mol. Biol.* **265**, 620–636

Annexe 2: Répartition de l'orientation des particules utilisées dans la reconstruction 3D du CELO

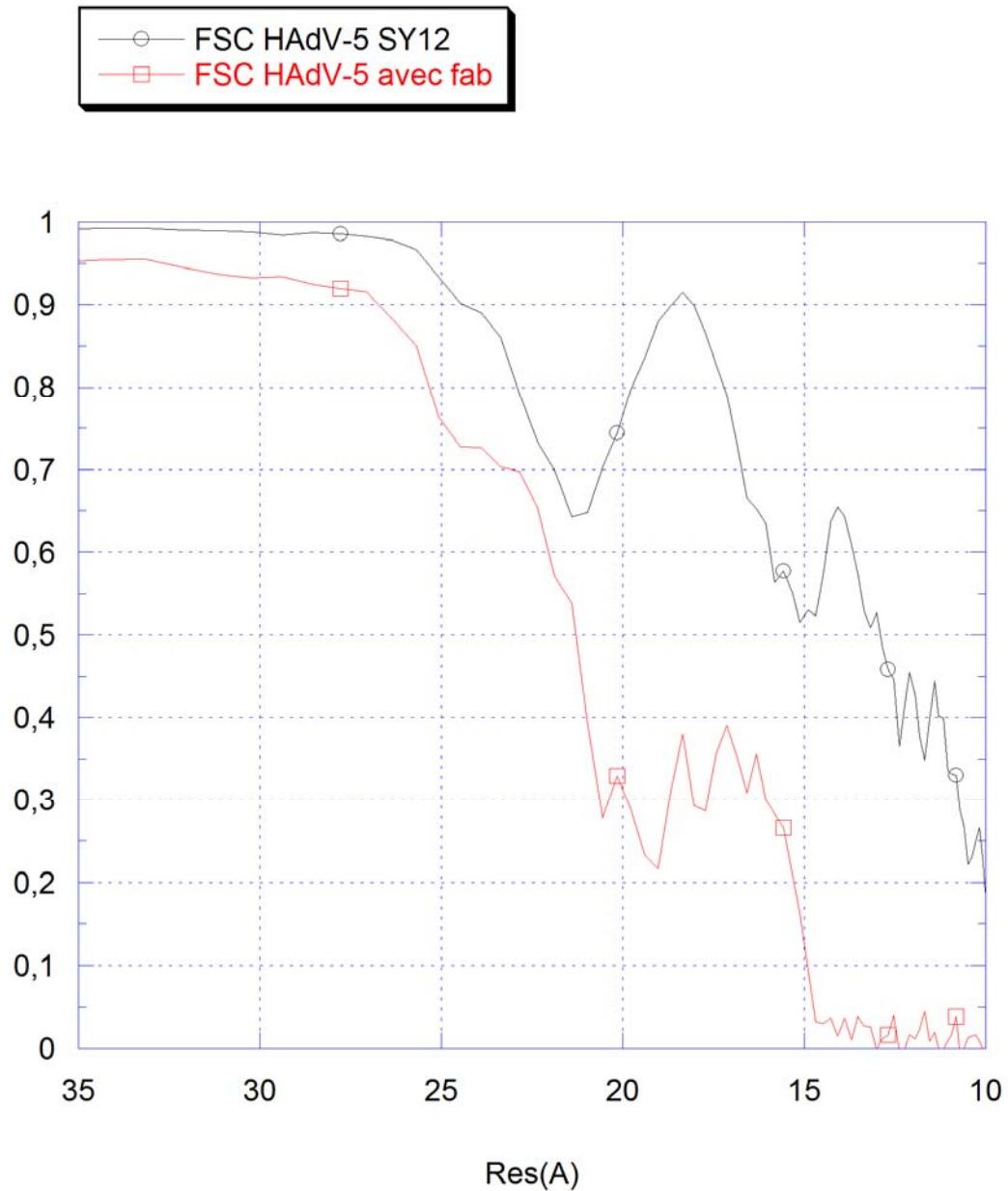


Annexe 3: Graphique des FSC en fonction de la résolution spatiale pour la reconstruction 3D obtenue pour le CELO.



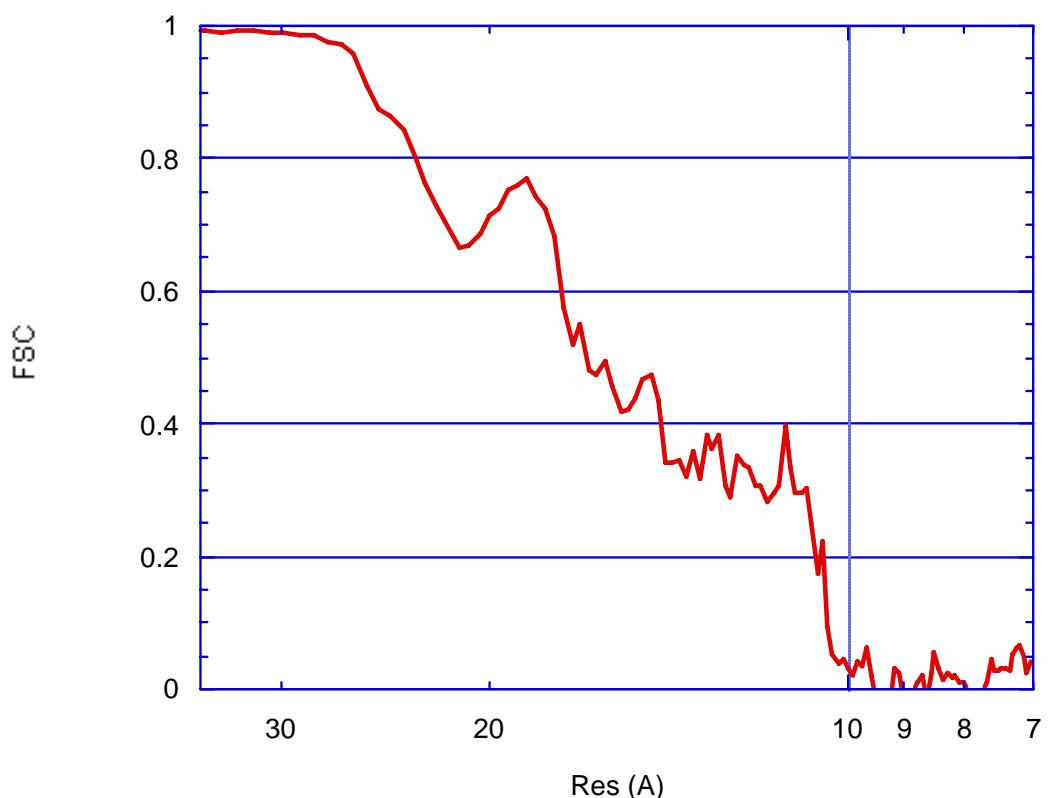
donne une résolution de 30 Å à  $FSC_{0,5}$ , et de 28 Å à  $FSC_{0,3}$

Annexe 4: Graphique des FSC en fonction de la résolution spatiale pour la reconstruction 3D obtenue pour l'échantillon de HAdV-5\_SY12 en absence et en présence de Fab.



Donne une résolution de 13 Å à  $FSC_{0,5}$  et de 11 Å à  $FSC_{0,3}$  pour le virus HAdV-5 SY12  
Pour le virus HAdV-5 SY12 en complexe avec un fab anti-SY12 : 22 Å à  $FSC_{0,5}$ .

Annexe 5: Graphique des FSC en fonction de la résolution spatiale pour la reconstruction 3D obtenue pour la souche ts1.



donne une résolution de 16 Å à  $FSC_{0,5}$ , et de 12 Å à  $FSC_{0,3}$

Annexe 6: Liste des programmes utilisés pour le traitement des données en ME et le recalage des données de cristallographie par « fitting ».

PFT, EM3DR	<a href="http://cryoem.ucsd.edu/programs/">http://cryoem.ucsd.edu/programs/</a>
PFT2, EM3DR2	<a href="http://people.chem.byu.edu/belnap/pft3dr">http://people.chem.byu.edu/belnap/pft3dr</a>
ROBEM	<a href="http://cryoem.ucsd.edu/programs/">http://cryoem.ucsd.edu/programs/</a>
SPIDER & WEB	<a href="http://www.wadsworth.org/spider_doc/spider/docs/master.html">http://www.wadsworth.org/spider_doc/spider/docs/master.html</a>
BSOFT	<a href="http://lsbr.niams.nih.gov/bsoft/bsoft.html">http://lsbr.niams.nih.gov/bsoft/bsoft.html</a>
EM2EM	<a href="http://www.imagescience.de/">http://www.imagescience.de/</a>
CCP4i	<a href="http://www ccp4.ac.uk/index.php">http://www ccp4.ac.uk/index.php</a>
URO	<a href="http://mem.ibs.fr/UROX/">http://mem.ibs.fr/UROX/</a>
SITUS	<a href="http://situs.biomachina.org/">http://situs.biomachina.org/</a>
CHIMERA	<a href="http://www.cgl.ucsf.edu/chimera/">http://www.cgl.ucsf.edu/chimera/</a>
PyMOL	<a href="http://pymol.sourceforge.net/">http://pymol.sourceforge.net/</a>

Annexe 7: La virosphère



# REFERENCES BIBLIOGRAPHIQUES

---

## A

Adrian, M., Dubochet, J., Lepault, J. and McDowall, A.W. (1984) Cryo-electron microscopy of viruses. *Nature*, **308**, 32-36.

Albertini, A.A., Wernimont, A.K., Muziol, T., Ravelli, R.B., Clapier, C.R., Schoehn, G., Weissenhorn, W. and Ruigrok, R.W. (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science*, **313**, 360-363.

Anderson, C.W., Young, M.E. and Flint, S.J. (1989) Characterization of the adenovirus 2 virion protein, mu. *Virology*, **172**, 506-512.

Arnberg, N., Edlund, K., Kidd, A.H. and Wadell, G. (2000) Adenovirus type 37 uses sialic acid as a cellular receptor. *J Virol*, **74**, 42-48.

## B

Baker, T.S. and Cheng, R.H. (1996) A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J Struct Biol*, **116**, 120-130.

Baker, T.S., Olson, N.H. and Fuller, S.D. (1999) Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol Mol Biol Rev*, **63**, 862-922, table of contents.

Bella, J., Kolatkar, P.R., Marlor, C.W., Greve, J.M. and Rossmann, M.G. (1998) The structure of the two amino-terminal domains of human ICAM-1 suggests how it functions as a rhinovirus receptor and as an LFA-1 integrin ligand. *Proc Natl Acad Sci U S A*, **95**, 4140-4145.

Benko, M. and Harrach, B. (1998) A proposal for a new (third) genus within the family Adenoviridae. *Arch Virol*, **143**, 829-837.

Benko, M. and Harrach, B. (2003) Molecular evolution of adenoviruses. *Curr Top Microbiol Immunol*, **272**, 3-35.

Bergelson, J.M., Cunningham, J.A., Drogue, G., Kurt-Jones, E.A., Krithivas, A., Hong, J.S., Horwitz, M.S., Crowell, R.L. and Finberg, R.W. (1997) Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science*, **275**, 1320-1323.

Bragg, W.L. (1929) The determination of parameters in crystal structures by means of Fourier series *Proc Roy Soc Lond A*, **123**, 537-559.

Brenner, S. and Horne, R.W. (1959) A negative staining method for high resolution electron microscopy of viruses. *Biochim Biophys Acta*, **34**, 103-110.

Briggs, J.A., Huisken, J.T., Fernando, K.V., Gilbert, R.J., Scotti, P., Butcher, S.J. and Fuller, S.D. (2005) Classification and three-dimensional reconstruction of unevenly distributed or symmetry mismatched features of icosahedral particles. *J Struct Biol*, **150**, 332-339.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**, 905-921.

Burnett, R.M. (1985) The structure of the adenovirus capsid. II. The packing symmetry of hexon and its implications for viral architecture. *J Mol Biol*, **185**, 125-143.

Burnett, R.M., Grutter, M.G. and White, J.L. (1985) The structure of the adenovirus capsid. I. An envelope model of hexon at 6 Å resolution. *J Mol Biol*, **185**, 105-123.

## C

Caspar, D.L. and Klug, A. (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol*, **27**, 1-24.

Ceulemans, H. and Russell, R.B. (2004) Fast fitting of atomic structures to low-resolution electron density maps by surface overlap maximization. *J Mol Biol*, **338**, 783-793.

Chiocca, S., Kurzbauer, R., Schaffner, G., Baker, A., Mautner, V. and Cotten, M. (1996) The complete DNA sequence and genomic organization of the avian adenovirus CELO. *J Virol*, **70**, 2939-2949.

Chiu, W., Schmid, M.F. and Prasad, B.V. (1993) Teaching electron diffraction and imaging of macromolecules. *Biophys J*, **64**, 1610-1625.

Chroboczek, J., Bieber, F. and Jacrot, B. (1992) The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology*, **186**, 280-285.

Conway, J.F. and Steven, A.C. (1999) Methods for reconstructing density maps of "single" particles from cryoelectron micrographs to subnanometer resolution. *J Struct Biol*, **128**, 106-118.

Conway, J.F., Trus, B.L., Booy, F.P., Newcomb, W.W., Brown, J.C. and Steven, A.C. (1993) The effects of radiation damage on the structure of frozen hydrated HSV-1 capsids. *J Struct Biol*, **111**, 222-233.

Coulibaly, F., Chevalier, C., Gutsche, I., Pous, J., Navaza, J., Bressanelli, S., Delmas, B. and Rey, F.A. (2005) The birnavirus crystal structure reveals structural relationships among icosahedral viruses. *Cell*, **120**, 761-772.

Crowther, R.A. (1971) Three-dimensional reconstruction and the architecture of spherical viruses. *Endeavour*, **30**, 124-129.

Crowther, R.A., Amos, L.A., Finch, J.T., De Rosier, D.J. and Klug, A. (1970) Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. *Nature*, **226**, 421-425.

## D

Davison, A.J., Benko, M. and Harrach, B. (2003) Genetic content and evolution of adenoviruses. *J Gen Virol*, **84**, 2895-2908.

Davison, A.J. and Harrach, B. (2002) *Siadenovirus*. C. A. Tidona & G. Darai. New York: Springer-Verlag.

Defer, C., Belin, M.T., Caillet-Boudin, M.L. and Boulanger, P. (1990) Human adenovirus-host cell interactions: comparative study with members of subgroups B and C. *J Virol*, **64**, 3661-3673.

Dekker, J., Kanellopoulos, P.N., Loonstra, A.K., van Oosterhout, J.A., Leonard, K., Tucker, P.A. and van der Vliet, P.C. (1997) Multimerization of the adenovirus DNA-binding protein is the driving force for ATP-independent DNA unwinding during strand displacement synthesis. *Embo J*, **16**, 1455-1463.

DeRosier, D.J. and Klug, A. (1968) Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature*, **217**, 130-134.

DeRosier, D.J. and Moore, P.B. (1970) Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. *J Mol Biol*, **52**, 355-369.

Ding, J., McGrath, W.J., Sweet, R.M. and Mangel, W.F. (1996) Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. *Embo J*, **15**, 1778-1783.

Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowall, A.W. and Schultz, P. (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys*, **21**, 129-228.

## E

El Bakkouri, M., Seiradake, E., Cusack, S., Ruigrok, R.W. and Schoehn, G. (2008) Structure of the C-terminal head domain of the fowl adenovirus type 1 short fibre. *Virology*.

Everitt, E., Lutter, L. and Philipson, L. (1975) Structural proteins of adenoviruses. XII. Location and neighbor relationship among proteins of adenovirion type 2 as revealed by enzymatic iodination, immunoprecipitation and chemical cross-linking. *Virology*, **67**, 197-208.

## F

Fender, P., Ruigrok, R.W., Gout, E., Buffet, S. and Chroboczek, J. (1997) Adenovirus dodecahedron, a new vector for human gene transfer. *Nat Biotechnol*, **15**, 52-56.

Frank, J. (1981) Three-dimensional reconstruction of single molecules. *Methods Cell Biol*, **22**, 325-344.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. and Leith, A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol*, **116**, 190-199.

Frank, J. and van Heel, M. (1982) Correspondence analysis of aligned images of biological particles. *J Mol Biol*, **161**, 134-137.

Frank, J., Verschoor, A. and Boublík, M. (1981) Computer averaging of electron micrographs of 40S ribosomal subunits. *Science*, **214**, 1353-1355.

Fuller, S.D. (1987) The T=4 envelope of Sindbis virus is organized by interactions with a complementary T=3 capsid. *Cell*, **48**, 923-934.

Fuller, S.D., Butcher, S.J., Cheng, R.H. and Baker, T.S. (1996) Three-dimensional reconstruction of icosahedral particles--the uncommon line. *J Struct Biol*, **116**, 48-55.

Furciniti, P.S., van Oostrum, J. and Burnett, R.M. (1989) Adenovirus polypeptide IX revealed as capsid cement by difference images from electron microscopy and crystallography. *Embo J*, **8**, 3563-3570.

## G

Gaggar, A., Shayakhmetov, D.M. and Lieber, A. (2003) CD46 is a cellular receptor for group B adenoviruses. *Nat Med*, **9**, 1408-1412.

Greber, U.F., Suomalainen, M., Stidwill, R.P., Boucke, K., Ebersold, M.W. and Helenius, A. (1997) The role of the nuclear pore complex in adenovirus DNA entry. *Embo J*, **16**, 5998-6007.

Greber, U.F., Webster, P., Weber, J. and Helenius, A. (1996) The role of the adenovirus protease on virus entry into cells. *Embo J*, **15**, 1766-1777.

Greber, U.F., Willetts, M., Webster, P. and Helenius, A. (1993) Stepwise dismantling of adenovirus 2 during entry into cells. *Cell*, **75**, 477-486.

Guardado-Calvo, P., Llamas-Saiz, A.L., Fox, G.C., Langlois, P. and van Raaij, M.J. (2007) Structure of the C-terminal head domain of the fowl adenovirus type 1 long fiber. *J Gen Virol*, **88**, 2407-2416.

## H

Harauz, G. and van Heel, M. (1986) Exact filters for general geometry three dimensional reconstruction. *Optik*, **73**, 146-156.

Harrison, S.C. (1996) *Principles of Virus Structure*. Lippincott Williams and Wilkins Editions, Philadelphia.

Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K. and Bricogne, G. (1978) Tomato bushy stunt virus at 2.9 [angstrom] resolution. *Nature*, **276**, 368-373.

Hewat, E.A. and Blaas, D. (1996) Structure of a neutralizing antibody bound bivalently to human rhinovirus 2. *Embo J*, **15**, 1515-1523.

Hierholzer, J.C. (1992) Adenoviruses in the immunocompromised host. *Clin Microbiol Rev*, **5**, 262-274.

Hilleman, M.R. and Werner, J.H. (1954) Recovery of new agent from patients with acute respiratory illness. *Proc Soc Exp Biol Med*, **85**, 183-188.

## J

Jiang, W. and Ludtke, S.J. (2005) Electron cryomicroscopy of single particles at subnanometer resolution. *Curr Opin Struct Biol*, **15**, 571-577.

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica Section A*, **47**, 110-119.

## K

Kanegae, Y., Makimura, M. and Saito, I. (1994) A simple and efficient method for purification of infectious recombinant adenovirus. *Jpn J Med Sci Biol*, **47**, 157-166.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. and Phillips, D.C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, **181**, 662-666.

Klug, A. and Finch, J.T. (1968) Structure of viruses of the papilloma-polyoma type. IV. Analysis of tilting experiments in the electron microscope. *J Mol Biol*, **31**, 1-12.

## L

Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V. and Rossmann, M.G. (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell*, **118**, 419-429.

Leopold, P.L., Kreitzer, G., Miyazawa, N., Rempel, S., Pfister, K.K., Rodriguez-Boulan, E. and Crystal, R.G. (2000) Dynein- and microtubule-mediated translocation of adenovirus serotype 5 occurs after endosomal lysis. *Hum Gene Ther*, **11**, 151-165.

Litvin, D.B. (1990) The Icosahedral Point Groups. *Acta Cryst. A*, **47**, 70-73.

Liu, G.Q., Babiss, L.E., Volkert, F.C., Young, C.S. and Ginsberg, H.S. (1985) A thermolabile mutant of adenovirus 5 resulting from a substitution mutation in the protein VIII gene. *J Virol*, **53**, 920-925.

## M

Mabit, H., Nakano, M.Y., Prank, U., Saam, B., Dohner, K., Sodeik, B. and Greber, U.F. (2002) Intact microtubules support adenovirus and herpes simplex virus infections. *J Virol*, **76**, 9962-9971.

Mathews, M.B. and Shenk, T. (1991) Adenovirus virus-associated RNA and translation control. *J Virol*, **65**, 5657-5662.

Meier, O. and Greber, U.F. (2004) Adenovirus endocytosis. *J Gene Med*, **6 Suppl 1**, S152-163.

Monaghan, A., Webster, A. and Hay, R.T. (1994) Adenovirus DNA binding protein: helix destabilising properties. *Nucleic Acids Res*, **22**, 742-748.

## N

Navaza, J., Lepault, J., Rey, F.A., Alvarez-Rua, C. and Borge, J. (2002) On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr D Biol Crystallogr*, **58**, 1820-1825.

Neurath, A.R. and Rubin, B.A. (1968) Interaction of p-chloromercuribenzoate with adenoviruses. Inactivation of haemagglutinins and degradation of virions of types 3, 4 and 7. *J Gen Virol*, **2**, 215-229.

Norrby, E. (1966) The relationship between the soluble antigens and the virion of adenovirus type 3. I. Morphological characteristics. *Virology*, **28**, 236-248.

Norrby, E. (1968) Identification of soluble components of adenovirus type 11. *J Gen Virol*, **2**, 123-133.

Norrby, E., Bartha, A., Boulanger, P., Dreizin, R.S., Ginsberg, H.S., Kalter, S.S., Kawamura, H., Rowe, W.P., Russell, W.C., Schlesinger, W. and Wigand, R. (1976) Adenoviridae. *Intervirology*, **7**, 117-125.

Norrby, E., Nyberg, B., Skaaret, P. and Lengyel, A. (1967) Separation and characterization of soluble adenovirus type 9 components. *J Virol*, **1**, 1101-1108.

Norrby, E. and Wadell, G. (1967) Soluble components of adenovirus type 4. *Virology*, **31**, 592-600.

## O

Olson, N.H., Kolatkar, P.R., Oliveira, M.A., Cheng, R.H., Greve, J.M., McClelland, A., Baker, T.S. and Rossmann, M.G. (1993) Structure of a human rhinovirus complexed with its receptor molecule. *Proc Natl Acad Sci U S A*, **90**, 507-511.

Orlova, E.V., Dube, P., Harris, J.R., Beckman, E., Zemlin, F., Markl, J. and van Heel, M. (1997) Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *J Mol Biol*, **271**, 417-437.

## P

Pereira, H.G. and Wrigley, N.G. (1974) In vitro reconstitution, hexon bonding and handedness of incomplete adenovirus capsid. *J Mol Biol*, **85**, 617-630.

Perez-Romero, P., Gustin, K.E. and Imperiale, M.J. (2006) Dependence of the encapsidation function of the adenovirus L1 52/55-kilodalton protein on its ability to bind the packaging sequence. *J Virol*, **80**, 1965-1971.

Perutz, M.F. (1964) The Structure of Myoglobin and Haemoglobin. *Dan Tidsskr Farm*, **38**, 113-122.

Perutz, M.F., Muirhead, H., Cox, J.M. and Goaman, L.C. (1968) Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. *Nature*, **219**, 131-139.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605-1612.

Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G.N., Cohen, S., Perrakis, A. and Noble, M. (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr*, **60**, 2288-2294.

## R

Ranson, N.A., Clare, D.K., Farr, G.W., Houldershaw, D., Horwich, A.L. and Saibil, H.R. (2006) Allosteric signaling of ATP hydrolysis in GroEL-GroES complexes. *Nat Struct Mol Biol*, **13**, 147-152.

Rekosh, D.M., Russell, W.C., Bellet, A.J. and Robinson, A.J. (1977) Identification of a protein linked to the ends of adenovirus DNA. *Cell*, **11**, 283-295.

Roberts, M.M., White, J.L., Grutter, M.G. and Burnett, R.M. (1986) Three-dimensional structure of the adenovirus major coat protein hexon. *Science*, **232**, 1148-1151.

Roberts, R.J., O'Neill, K.E. and Yen, C.T. (1984) DNA sequences from the adenovirus 2 genome. *J Biol Chem*, **259**, 13968-13975.

Rosa-Calatrava, M., Grave, L., Puvion-Dutilleul, F., Chatton, B. and Kedinger, C. (2001) Functional analysis of adenovirus protein IX identifies domains involved in capsid stability, transcriptional activity, and nuclear reorganization. *J Virol*, **75**, 7131-7141.

Roseman, A.M. (2000) Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr D Biol Crystallogr*, **56** (Pt 10), 1332-1340.

Rossmann, M.G. (2000) Fitting atomic models into electron-microscopy maps. *Acta Crystallogr D Biol Crystallogr*, **56** (Pt 10), 1341-1349.

Rowe, W.P., Huebner, R.J., Gilmore, L.K., Parrott, R.H. and Ward, T.G. (1953) Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proc Soc Exp Biol Med*, **84**, 570-573.

Ruigrok, R.W., Barge, A., Albiges-Rizo, C. and Dayan, S. (1990) Structure of adenovirus fibre. II. Morphology of single fibres. *J Mol Biol*, **215**, 589-596.

Russell, W.C. (2000) Update on adenovirus and its vectors. *J Gen Virol*, **81**, 2573-2604.

Rux, J.J. and Burnett, R.M. (2000) Type-specific epitope locations revealed by X-ray crystallographic study of adenovirus type 5 hexon. *Mol Ther*, **1**, 18-30.

## S

Saban, S.D., Nepomuceno, R.R., Gritton, L.D., Nemerow, G.R. and Stewart, P.L. (2005) CryoEM structure at 9Å resolution of an adenovirus vector targeted to hematopoietic cells. *J Mol Biol*, **349**, 526-537.

Saban, S.D., Silvestry, M., Nemerow, G.R. and Stewart, P.L. (2006) Visualization of alpha-helices in a 6-angstrom resolution cryoelectron microscopy structure of adenovirus allows refinement of capsid protein assignments. *J Virol*, **80**, 12049-12059.

San Martin, C.S., Burnett, R.M., de Haas, F., Heinkel, R., Rutten, T., Fuller, S.D., Butcher, S.J. and Bamford, D.H. (2001) Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions. *Structure (Camb)*, **9**, 917-930.

Sarma, P.S., Huebner, R.J. and Lane, W.T. (1965) Induction of tumors in hamsters with an avian adenovirus (CELO). *Science*, **149**, 1108.

Saxton, W.O. and Baumeister, W. (1982) The correlation averaging of a regularly arranged bacterial cell envelope protein. *J Microsc*, **127**, 127-138.

Schagen, F.H., Rademaker, H.J., Rabelink, M.J., van Ormondt, H., Fallaux, F.J., van der Eb, A.J. and Hoeben, R.C. (2000) Ammonium sulphate precipitation of recombinant adenovirus from culture medium: an easy method to increase the total virus yield. *Gene Ther*, **7**, 1570-1574.

Schoehn, G., Fender, P., Chroboczek, J. and Hewat, E.A. (1996) Adenovirus 3 penton dodecahedron exhibits structural changes of the base on fibre binding. *Embo J*, **15**, 6841-6846.

Schoehn, G., Mavrakis, M., Albertini, A., Wade, R., Hoenger, A. and Ruigrok, R.W. (2004) The 12 A structure of trypsin-treated measles virus N-RNA. *J Mol Biol*, **339**, 301-312.

Seiradake, E., Lortat-Jacob, H., Billet, O., Kremer, E.J. and Cusack, S. (2006) Structural and mutational analysis of human Ad37 and canine adenovirus 2 fiber heads in complex with the D1 domain of coxsackie and adenovirus receptor. *J Biol Chem*, **281**, 33704-33716.

Shenk, T.E. (1990) *Adenoviridae: The Viruses and Their replication*. Lippincott Williams & Wilkins.

Short, J.J., Pereboev, A.V., Kawakami, Y., Vasu, C., Holterman, M.J. and Curiel, D.T. (2004) Adenovirus serotype 3 utilizes CD80 (B7.1) and CD86 (B7.2) as cellular attachment receptors. *Virology*, **322**, 349-359.

Stewart, P.L., Chiu, C.Y., Huang, S., Muir, T., Zhao, Y., Chait, B., Mathias, P. and Nemerow, G.R. (1997) Cryo-EM visualization of an exposed RGD epitope on adenovirus that escapes antibody neutralization. *Embo J*, **16**, 1189-1198.

Stewart, P.L., Fuller, S.D. and Burnett, R.M. (1993) Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy. *Embo J*, **12**, 2589-2599.

Suhre, K., Navaza, J. and Sanejouand, Y.-H. (2006) NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Crystallographica Section D*, **62**, 1098-1100.

Suhre, K. and Sanejouand, Y.H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res*, **32**, W610-614.

## T

Tama, F., Miyashita, O. and Brooks, C.L., 3rd. (2004) Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol*, **147**, 315-326.

Tan, P.K., Michou, A.I., Bergelson, J.M. and Cotten, M. (2001) Defining CAR as a cellular receptor for the avian adenovirus CELO using a genetic analysis of the two viral fibre proteins. *J Gen Virol*, **82**, 1465-1472.

Trentin, J.J., Yabe, Y. and Taylor, G. (1962) The quest for human cancer viruses. *Science*, **137**, 835-841.

Tribouley, C., Lutz, P., Staub, A. and Kedinger, C. (1994) The product of the adenovirus intermediate gene IVa2 is a transcriptional activator of the major late promoter. *J Virol*, **68**, 4450-4457.

## V

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. and Schatz, M. (1996) A new generation of the IMAGIC image processing system. *J Struct Biol*, **116**, 17-24.

van Oostrum, J. and Burnett, R.M. (1985) Molecular composition of the adenovirus type 2 virion. *J Virol*, **56**, 439-448.

van Raaij, M.J., Mitraki, A., Lavigne, G. and Cusack, S. (1999) A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature*, **401**, 935-938.

Vives, R.R., Lortat-Jacob, H., Chroboczek, J. and Fender, P. (2004) Heparan sulfate proteoglycan mediates the selective attachment and internalization of serotype 3 human adenovirus dodecahedron. *Virology*, **321**, 332-340.

Volkmann, N. and Hanein, D. (1999) Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J Struct Biol*, **125**, 176-184.

## W

Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.

Wickham, T.J., Mathias, P., Cheresh, D.A. and Nemerow, G.R. (1993) Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell*, **73**, 309-319.

Wiethoff, C.M., Wodrich, H., Gerace, L. and Nemerow, G.R. (2005) Adenovirus protein VI mediates membrane disruption following capsid disassembly. *J Virol*, **79**, 1992-2000.

Wodrich, H., Guan, T., Cingolani, G., Von Seggern, D., Nemerow, G. and Gerace, L. (2003) Switch from capsid protein import to adenovirus assembly by cleavage of nuclear transport signals. *Embo J*, **22**, 6245-6255.

Wriggers, W., Milligan, R.A. and McCammon, J.A. (1999) Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol*, **125**, 185-195.

Wu, E., Pache, L., Von Seggern, D.J., Mullen, T.M., Mikyas, Y., Stewart, P.L. and Nemerow, G.R. (2003) Flexibility of the adenovirus fiber is required for efficient receptor interaction. *J Virol*, **77**, 7225-7235.

## X

Xia, D., Henry, L.J., Gerard, R.D. and Deisenhofer, J. (1994) Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 Å resolution. *Structure*, **2**, 1259-1270.

Xu, L., Benson, S.D. and Burnett, R.M. (2007) Nanoporous crystals of chicken embryo lethal orphan (CELO) adenovirus major coat protein, hexon. *J Struct Biol*, **157**, 424-431.

## Y

Yang, S., Yu, X., Galkin, V.E. and Egelman, E.H. (2003) Issues of resolution and polymorphism in single-particle reconstruction. *J Struct Biol*, **144**, 162-171.

Yates, V.J. and Fry, D.E. (1957) Observations on a chicken embryo lethal orphan (CELO) virus. *Am J Vet Res*, **18**, 657-660.

Yu, X., Jin, L. and Zhou, Z.H. (2008) 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature*, **453**, 415-419.

## Z

Zhang, W., Low, J.A., Christensen, J.B. and Imperiale, M.J. (2001) Role for the adenovirus IVa2 protein in packaging of viral DNA. *J Virol*, **75**, 10446-10454.

Zubieta, C., Schoehn, G., Chroboczek, J. and Cusack, S. (2005) The structure of the human adenovirus 2 penton. *Mol Cell*, **17**, 121-135.





## Résumé

Les Adénovirus sont des virus non enveloppés à symétrie icosaédrique. Découverts chez l'homme dans les années 50, ils constituent la famille des *Adenoviridae*. Ils existent une grande variété de souches pouvant infecter une grande variété d'hôtes allant de l'homme jusqu'au poisson. L'étude de l'Adénovirus a été stimulée par la découverte en 1962 par Trentin et collaborateurs de leur capacité à induire des tumeurs chez les bébés hamsters. Sur le plan structural, de nombreuses études en Microscopie Electronique ont été menées pour comprendre l'assemblage et la formation de la capsid. Depuis les années 90, avec la progression de l'informatique et des techniques liées à l'analyse d'images, la structure de la capsid des Adénovirus humains est de mieux en mieux comprise mais la localisation de certaines protéines mineures restaient encore incertaine ou inconnue. Nous avons, à travers l'utilisation de la Microscopie Electronique associée à l'analyse d'images, obtenu des reconstructions 3D pour plusieurs souches d'Adénovirus humains, canins et aviaires. L'obtention d'un modèle 3D à haute résolution pour l'Adénovirus humain de type 5 nous a permis de reconstituer un modèle quasi-atomique par la méthode de « fitting » et de déterminer avec plus de précisions la position de certaines protéines mineures comme la protéine IX, IIIa et VIII. L'étude d'une souche mutante d'Adénovirus de type 5 en complexe avec un Fab ainsi que d'une souche canine nous a permis de localiser précisément l'extrémité C terminale de la protéine IX. Enfin l'obtention d'une structure à haute résolution d'un virus immature dérivé de l'Adénovirus humain de type 2 a ouvert la voie sur l'étude et la compréhension de la maturation et de la phase précoce de son cycle cellulaire.

**Mots-clés:** Adénovirus, *Adenoviridae*, Microscopie Electronique, Reconstruction 3D, Fitting, Cristallographie, Virus à ADN double brin, symétries icosaédriques.

## Abstract

Adenoviruses are non-enveloped icosahedral viruses. They were found during the 50's and since constitute the huge *Adenoviridae* family. You can find a large variety of strains able to infect a large variety of hosts from humans to fishes. Fundamental studies on adenoviruses were largely stimulated since the discovery in 1962 by Trentin et al. of their ability to induce tumours in baby hamsters. From the structural point of view, many studies based on electron microscopy were done to understand how the adenovirus capsid is assembled. Since the last fifteen years, progress made in computer science and Electron Microscopy based techniques for 3D image analysis helped scientific community to get a better understanding of the adenovirus capsid structure but many minor proteins were uncertainly localised or totally unknown. Our work is based on Electron Microscopy and 3D image analysis. We obtained 3D models of different adenovirus strains like human, fowl and canine ones where two are below 10 Å resolution. A high resolution 3D reconstruction of Human Adenovirus type 5 helped us to reconstitute a quasi atomic model of the entire capsid and to determine the localisation of minor proteins like IX, IIIa and VIII. We also localised the C terminal part of protein IX from a 3D model of a mutant Human Adenovirus type 2 complexed with a Fab antibody and a canine Adenovirus. Finally, we were interested in the 3D structure of the non mature strain *ts1* derived from Human Adenovirus type 2. The “below10Å” 3D reconstruction obtained constitutes infamy into structural understanding of adenovirus maturation and early phase of infection.

**Keywords:** Adénovirus, *Adenoviridae*, Electron Microscopy, 3D reconstruction, Fitting, Crystallography, Double-stranded DNA viruses, icosahedral symmetries.