

Université du Québec
INRS-Institut Armand-Frappier

**Caractérisation de l'adaptation du coronavirus
dans le système nerveux central humain**

Par
Julien St-Jean

Thèse présentée pour l'obtention
du grade de Docteur ès sciences (Ph.D.)
en virologie et immunologie

Jury d'évaluation

Présidente du jury et
examinatrice interne : Véronika Von Messling, INRS-Institut Armand-Frappier

Examineurs externes : Éric Rassart, Université du Québec à Montréal
Michel J. Tremblay, Université Laval

Directeur de recherche : Pierre J. Talbot, INRS-Institut Armand-Frappier

RÉSUMÉ

Les souches 229E et OC43 du coronavirus humain (HCoV) sont omniprésentes dans l'environnement et sont responsables d'environ le tiers des rhumes recensés chez l'être humain. Il a été démontré expérimentalement dans le laboratoire du Dr Pierre Talbot que le HCoV-OC43 possédait des propriétés neurotropes et neuroinvasives et qu'il pouvait induire des infections persistantes dans des cultures de cellules neurales. Il pourrait donc être associé à des pathologies du système nerveux central. Bien qu'il ait également été démontré dans la présente étude que le HCoV-OC43 pouvait demeurer génétiquement stable dans l'environnement, il est connu que les coronavirus peuvent s'adapter en culture cellulaire ou en réponse à une pression de sélection donnée, telle que celle exercée par le système immunitaire.

Dans le présent projet de recherche, des infections persistantes de certaines lignées cellulaires neurales (MO3.13, H4, U-87 MG et TE-671) ont été effectuées dans le but de vérifier si la persistance du virus est à l'origine de mutations dans le gène qui code pour la protéine de surface S, laquelle pourrait être associée à l'adaptation du virus face à l'environnement neural. Dans le même ordre d'idées, le gène HE ainsi que le gène de nucléocapside (N) ont aussi été séquencés afin de déterminer si ces gènes pouvaient être reliés à une adaptation quelconque pouvant survenir en culture cellulaire. Les particules virales relarguées des cellules infectées de façon persistante ont donc été extraites puis purifiées et l'ARN génomique viral a été séquencé. Dans les différentes infections persistantes, des mutations dans le gène S ont été identifiées et conservées lors de passages subséquents, suggérant qu'elles conféreraient un avantage sélectif au virus persistant. Ces mutations ont notamment été retrouvées dans le site potentiel de liaison au récepteur, où les mutations D24Y, S83T, H183R et Y241H étaient prédominantes. Seulement quelques mutations ont été identifiées dans les gènes HE et N, suggérant que le gène S est le principal responsable de l'adaptation à l'environnement du virus, laquelle pourrait être associée au neurotropisme, à la neuroinvasion et possiblement à la neurovirulence.

Afin d'évaluer l'impact de ces mutations, des tests de modulation de tropisme et d'infectivité ont été réalisés en utilisant des lignées cellulaires dont la susceptibilité à l'endroit du HCoV-OC43 avait préalablement été déterminée. Ces analyses ont révélé que les mutations retrouvées dans le gène S pouvaient affecter certaines propriétés du virus. Ainsi, certains variants issus des infections persistantes présentaient un tropisme cellulaire élargi, un phénotype différent dans la souris et pouvaient former des plages de lyse *in vitro*, contrairement à la souche de

référence HCoV-OC43 ATCC VR759. Ces résultats suggèrent donc que le HCoV-OC43 peut évoluer en s'adaptant à son environnement cellulaire *in vitro*.

Finalement, un clone infectieux d'ADNc du HCoV-OC43 a été assemblé dans un *bacterial artificial chromosome* (BAC) sous le contrôle du promoteur du cytomégalovirus (CMV). Il a été montré que le virus recombinant possédait les mêmes propriétés de réplication *in vitro* et de neurovirulence *in vivo*. Ce clone s'avérera un outil de premier plan afin de déterminer l'impact des mutations retrouvées dans le gène S et de comprendre davantage les mécanismes de réplication et de tropisme du virus. En combinaison avec les expériences décrites précédemment, ce clone nous permettra de répondre à de nombreuses questions qui concernent les bases moléculaires de la neuropathogenèse des coronavirus humains.



Julien St-Jean



Pierre Talbot

AVANT-PROPOS

Les coronavirus humains (les souches 229E et OC43 principalement) sont omniprésents dans l'environnement et sont responsables d'environ le tiers des rhumes recensés chez l'être humain. Il y a plusieurs années, une association entre les coronavirus et certaines maladies neurodégénératives, dont la sclérose en plaques, a été établie, par l'entremise du coronavirus murin notamment. Plus récemment, plusieurs études sont venues appuyer cette hypothèse, confirmée chez l'humain par les travaux de Nathalie Arbour notamment, à l'époque étudiante au doctorat dans le laboratoire du Dr Pierre Talbot. Ces travaux ont établi que la souche humaine OC43 du coronavirus pouvait induire des infections persistantes dans différentes lignées cellulaires neurales (U-87 MG, U-373 MG, MO3.13 et H4) et que le virus possédait aussi des propriétés neuroinvasives. Lors de ces études, des mutations ont été retrouvées dans la sous-unité S1 du gène S des particules virales issues d'infections persistantes ainsi que dans le gène N de virions dont la neuroinvasion a été démontrée, suggérant une adaptation potentielle du HCoV-OC43 dans le système nerveux central et/ou le cerveau humain.

Pour faire suite à ces travaux, une étude plus approfondie portant sur l'adaptation du HCoV-OC43, induite suite à des infections persistantes dans différentes lignées cellulaires neurales, a été entreprise dans le cadre de mon projet de doctorat. Les travaux découlant de ce projet de doctorat sont donc présentés dans cette thèse. La thèse se présente principalement sous forme de cinq publications scientifiques, trois ayant déjà été publiées dans *Journal of Virology*. La publication no.1, déjà publiée, présente notamment une analyse du génome du HCoV-OC43 et de ses principales caractéristiques, une étude comparative du virus par rapport à d'autres souches coronavirales apparentées, telles que le SARS-CoV, ainsi qu'une démonstration de la stabilité et des propriétés neuroinvasives intrinsèques du virus. La publication no.2 se veut une lettre à l'éditeur envoyée en réponse au commentaire d'une équipe belge, celle du Dr Marc Van Ranst, laquelle doutait de la stabilité génétique du HCoV-OC43 rapportée dans la publication no.1. Notre réponse à l'équipe du Dr Marc Van Ranst constitue en fait la publication no.2, alors que les commentaires combinés de l'équipe belge et de notre équipe sont présentés tels que publiés, en annexe à cet ouvrage. La publication no.3 est présentement sous presse dans le journal scientifique *Advances in Experimental Medicine and Biology*. Elle traite de l'évolution du HCoV-OC43 en culture cellulaire et contient des résultats préliminaires qui concernent la caractérisation de l'adaptation du virus. La publication no.4, publiée dans *Journal of Virology*, traite de l'assemblage d'un clone infectieux d'ADNc du HCoV-OC43, alors que la publication

no.5, soumise à *Virology*, présente les résultats relatifs à l'adaptation du virus suite à des infections persistantes dans des lignées cellulaires neurales.

La contribution de chacun des auteurs en ce qui concerne la réalisation de chacune des publications de cette thèse est la suivante. Pour la publication no.1, nos collaborateurs français, les Drs Astrid Vabret et François Freymuth, ont isolé la souche clinique Paris du HCoV-OC43 et nous l'ont fait parvenir. Le Dr Marc Desforges, auteur dans les cinq publications, a participé à l'analyse des résultats et à la révision de tous les manuscrits, et a également mis au point et réalisé les expériences de transfections transitoires du clone d'ADNc du HCoV-OC43, dans la publication no.4. Le Dr Hélène Jacomy, auteure dans les publications no.1, 2, 4 et 5, a quant à elle effectué toutes les infections impliquant les cultures de cellules primaires neurales et les souris, ainsi que l'immunohistochimie rattachée à ces expériences. Finalement, nos collaborateurs espagnols, les Drs Fernando Almazán et Luis Enjuanes, ont contribué à l'assemblage du clone infectieux du HCoV-OC43, présenté dans la publication no.4. Je me suis moi-même rendu en Espagne afin de me familiariser avec le vecteur BAC et d'élaborer une stratégie permettant de débiter la construction du clone. L'équipe espagnole nous a également aidés à mettre les conditions de transfection du vecteur au point. Le reste des travaux présentés dans cette thèse a été réalisé par moi-même.

REMERCIEMENTS

Beaucoup de personnes ont joué un rôle important dans la longue entreprise que fut mon doctorat. Bien qu'on puisse quelques fois se sentir laissé à soi-même durant un tel cheminement, il n'en est rien. Plusieurs personnes ont donc contribué directement ou indirectement à cet accomplissement et je les en remercie. Tout d'abord, j'aimerais remercier mes collègues étudiants et/ou de laboratoire, qui m'ont été d'un support incroyable durant les cinq dernières années. Grâce à leur énergie et leur dynamisme, j'ai su conserver l'envie de travailler et de me dépasser, même durant les moments les plus difficiles. J'aimerais particulièrement remercier les Drs Marc Desforges et Hélène Jacomy, ainsi que Francine Lambert, avec qui j'ai étroitement collaboré durant tout mon projet de doctorat. Je remercie également mes collègues espagnols, les Drs Fernando Almazán et Luis Enjuanes, pour m'avoir aidé à démarrer et mener à terme l'assemblage du clone infectieux du HCoV-OC43, qui disons-le, était un projet qui me tenait particulièrement à cœur. Par contribution indirecte, j'entendais les gens qui m'ont supporté moralement et psychologiquement, et à cet égard, j'aimerais remercier ma tendre moitié Julie, qui sans peut-être même s'en rendre compte, s'est avérée être une complice hors pair au cours des deux dernières années. J'aimerais finalement remercier mon directeur de recherche, le Dr Pierre Talbot, pour m'avoir accueilli dans son laboratoire mais également pour avoir cru en moi durant toutes ces années. Je ne sais toujours pas ce que l'avenir me réserve, mais j'entrevois mon futur en recherche scientifique avec beaucoup d'optimisme et d'enthousiasme.

TABLE DES MATIÈRES

CHAPITRE I : REVUE BIBLIOGRAPHIQUE	1
1.0 LES CORONAVIRUS	2
1.1 Nomenclature	2
1.1.1 Le mouse hepatitis virus (MHV)	3
1.1.2 Le SARS-coronavirus	5
1.1.3 Le coronavirus humain OC43	6
1.2 Les propriétés du virion	8
1.2.1 Le génome	9
1.2.2 Les protéines structurales	11
1.2.3 Les protéines non-structurales	13
1.3 La glycoprotéine S	14
1.3.1 La structure de la glycoprotéine S	14
1.3.2 Les fonctions associées à la glycoprotéine S	16
1.4 Autres propriétés fonctionnelles	21
1.4.1 La réplicase 1ab	22
1.4.2 Le cycle répliatif	25
2.0 L'INFECTION VIRALE CHEZ L'HÔTE	34
2.1 Généralités	34
2.2 La persistance virale et l'adaptation	35
2.3 Les infections du système nerveux central	37
3.0 LES CONSÉQUENCES DE L'INFECTION	39
3.1 Les pathologies associées	39
3.2 Les réponses immunitaires associées	41
3.2.1 La réponse immunitaire générale	41
3.2.2 L'auto-immunité	43
3.3 Les traitements anti-viraux	44
4.0 LA BIOLOGIE MOLÉCULAIRE ET LES CORONAVIRUS	47
4.1 Généralités	48
4.2 Les outils disponibles	48
4.3 Les clones infectieux	49

CHAPITRE II : PUBLICATION no.1 : <i>Human respiratory coronavirus OC43 : Genetic stability and neuroinvasion</i>	51
Page-titre	52
Résumé (en français)	53
Abstract	54
Introduction	55
Materials and methods	58
Results	63
Discussion	68
Acknowledgments	73
References	74
Tables	83
Figure legends	87
Figures	90
CHAPITRE III : PUBLICATION no.2 : <i>Apparent genetic stability of human respiratory coronavirus OC43</i>	95
Page-titre	96
Contenu	97
References	101
CHAPITRE IV : PUBLICATION no.3 : <i>Genetic evolution of human coronavirus OC43 in neural cell culture</i>	103
Page-titre	104
Introduction	104
Materials and methods	105
Results	106
Discussion	107
References	109

RÉRÉRENCES BIBLIOGRAPHIQUES	193
LISTE DES CONTRIBUTIONS SCIENTIFIQUES	224
ANNEXES	226

LISTE DES TABLEAUX

Tableau 1.	Les espèces de coronavirus	4
-------------------	----------------------------	---

LISTE DES FIGURES

Figure 1.	Représentation schématique d'une particule virale du HCoV-OC43 et des protéines structurales la composant.	9
Figure 2.	Représentation schématique des ARN du HCoV-OC43.	27
Figure 3.	Représentation schématique du mécanisme de transcription des coronavirus.	29
Figure 4.	Représentation schématique des mécanismes d'assemblage du TGEV dans la cellule infectée.	33

LISTE DES ABBRÉVIATIONS

2'-O-MT : *S-adenosylmethionine-dependant 2'-O-ribose methyltransferase*

3CLpro : *3C-like protease*

ACE2 : *angiotensin-converting enzyme 2*

ADN : acide désoxyribonucléique

ADNc : ADN complémentaire

ARN : acide ribonucléique

ARNm : ARN messenger

ATP : adénosine triphosphate

BAC : *bacterial artificial chromosome*

BCoV : coronavirus bovin

CEACAM1 : *carcinoembryonic antigen-related cell adhesion molecule 1*

CI : compartiment intermédiaire

CMH : complexe majeur d'histocompatibilité

CMV : cytomégalovirus

CoV : coronavirus

CS : *core sequence*

dATP : désoxyadénosine triphosphate

DI-RNA : *defective interfering RNA*

E : enveloppe

EGFR : *epidermal growth factor receptor*

ExoN : exoribonucléase (3' vers 5')

FIPV : *feline infectious peritonitis virus*

FMDV : *foot-and-mouth disease virus*

GFP : *green fluorescent protein*

GTP : guanosine triphosphate

HCoV : coronavirus humain

HDV : *hepatitis delta virus*

HE : hémagglutinine-estérase

HR : *heptad repeat*

IBV : *infectious bronchitis virus*

IFN : interféron

IRES : *internal ribosome entry site*

kb : kilobases
kDa : kilodaltons
M : membrane
MBP : *myelin basic protein*
MHV : *mouse hepatitis virus*
Mpro : *main protease*
N : nucléocapside
NendoU : *replicative nidoviral uridylate-specific endoribonuclease*
NK : *natural killer*
nm : nanomètre
nsp : *non structural protein*
ORF : *open reading frame*
PCR : *polymerase chain reaction*
PF : peptide de fusion
PLP : *proteolipid protein*
PLP : *papain-like protease*
RdRp : *RNA-dependant RNA polymerase*
RCT : récepteur de cellules T
RE : réticulum endoplasmique
RHV : région hypervariable
RT : *reverse transcription*
S : spicule ou *spike*
siRNA : *small interfering RNA*
SLR : site de liaison au récepteur
SNC : système nerveux central
SRAS : syndrome respiratoire aigu sévère
TGEV : *transmissible gastroenteritis virus*
TRS : *transcription regulatory sequences*
TRS-G : TRS-génomme
TRS-L : TRS-*leader*
UTR : *untranslated region*
VIH : virus de l'immunodéficience humaine
VLP : *virus-like particles*

CHAPITRE I : REVUE BIBLIOGRAPHIQUE

1.0 LES CORONAVIRUS

Les premiers coronavirus humains (HCoV) ont été isolés au milieu des années 60 chez des patients atteints de rhinite (Gagneur *et al.*, 2002), et s'inscrivaient ainsi à une longue liste de virus associés à des maladies des voies respiratoires supérieures et du rhume chez l'humain (Myint, 1994). Les rhinovirus, les paramyxovirus, les orthomyxovirus, les picornavirus et les adénovirus s'ajoutaient aussi à cette liste (Fouchier *et al.*, 2004). Le coronavirus humain 229E fut découvert en 1966 (Hamre *et al.*, 1966), alors que la souche humaine OC43 fut découverte en 1967 (McIntosh *et al.*, 1967). Au début des années 80, la mise en évidence d'une association entre la neuroinvasion et la souche JHM du coronavirus murin (MHV) chez le rat, suggéra une implication potentielle du virus dans le déclenchement de certaines pathologies du système nerveux central (SNC), dont des pathologies démyélinisantes (Sorensen *et al.*, 1980). Cette hypothèse fut renforcée par l'identification d'anticorps dirigés contre les coronavirus humains OC43 et 229E chez des patients atteints de la sclérose en plaques, en 1982 (Salmi *et al.*, 1982). Depuis, de nombreuses recherches, dont plusieurs effectuées dans le laboratoire du Dr Pierre Talbot, ont validé l'hypothèse du neurotropisme (Arbour *et al.*, 1999a; Arbour *et al.*, 1999b; Bonavia *et al.*, 1997) et de la neuroinvasion (Arbour *et al.*, 2000), et ainsi démontré l'implication et le rôle prépondérant des coronavirus dans le développement des maladies neurodégénératives.

1.1 Nomenclature

Les coronavirus (CoV) sont des virus à ARN monocaténaire de polarité positive. Ces virus font partie de l'ordre des *Nidovirales* et appartiennent à la famille des *Coronaviridae* (Lai et Cavanagh, 1997). Les *Nidovirales* comprennent également deux autres familles: les *Arteriviridae* ainsi que les *Roniviridae*. Les caractéristiques principales qui distinguent les *Nidovirales* des virus appartenant à d'autres ordres sont l'arrangement niché co-terminal des ARN messagers (ARNm) sous-génomiques à leur extrémité 3' ainsi que la grande taille du gène de la réplicase 1ab, qui à lui seul occupe plus des deux tiers du génome viral (Snijder *et al.*, 2003). La famille des *Coronaviridae* comprend deux genres: les coronavirus et les torovirus. Les torovirus possèdent une structure similaire à celle des coronavirus (de Haan *et al.*, 2002b), contrairement aux *Arteriviridae* et aux *Roniviridae*, qui possèdent un génome beaucoup plus petit et restreint.

Les coronavirus sont séparés en trois groupes génétiques distincts, basés sur des critères sérologiques, antigéniques et phylogéniques (Tableau 1) (Lai et Cavanagh, 1997). Les espèces

les mieux caractérisées du groupe 1 sont le coronavirus humain 229E (HCoV-229E) ainsi que le *transmissible gastroenteritis virus* (TGEV), qui infecte principalement le porc. Pour sa part, le groupe 2 contient notamment le MHV, le HCoV-OC43 ainsi que le *severe acute respiratory syndrome coronavirus* (SARS-CoV) (Snijder *et al.*, 2003), responsable du syndrome respiratoire aigu sévère (SRAS). Le groupe 3 des coronavirus est quant à lui principalement représenté par le coronavirus aviaire, ou *infectious bronchitis virus* (IBV).

Depuis l'émergence du virus responsable du SRAS en novembre 2002 (Peiris *et al.*, 2003), l'intérêt accordé aux coronavirus s'est ravivé. Les recherches visant à découvrir de nouveaux virus, dans la nature ou chez l'humain, se sont accrues et plusieurs autres souches de coronavirus ont ainsi été découvertes. La première souche coronavirale identifiée suite à l'éruption du SARS-CoV fut le HCoV-NL63 (Van der Hoek *et al.*, 2004), aussi nommé HCoV-NL, par un autre groupe de recherche qui découvrit le virus presque au même moment (Fouchier *et al.*, 2004). Cette souche humaine du coronavirus appartient au groupe 1 et présente une identité de 65% avec le HCoV-229E, au niveau nucléotidique. Elle causerait des infections des voies respiratoires inférieures, mais moins sévères que celles provoquées par le SARS-CoV (Lassnig *et al.*, 2005). Une autre souche humaine fut ensuite découverte en 2005. Il s'agit du HCoV-HKU1, appartenant au groupe 2 selon des analyses phylogéniques (Woo *et al.*, 2005). La recherche d'un réservoir animal pour le SARS-CoV a également mené à l'identification de nouvelles souches animales. Parmi celles-ci, on compte le *bat coronavirus* (Bat-CoV), appartenant au groupe 1 et isolé en Asie chez trois espèces différentes de chauve-souris (Poon *et al.*, 2005a), ainsi que différentes espèces du virus isolées chez différentes espèces d'oiseaux (canard, pigeon et oie) et appartenant au groupe 3 (Jonassen *et al.*, 2005). Comme les autres coronavirus du groupe 3, ces virus contiennent la séquence s2m, comprise dans la région non-traduite 3' (UTR 3') du génome viral.

1.1.1 Le mouse hepatitis virus (MHV)

Le coronavirus murin (MHV) est sans contredit le coronavirus le mieux caractérisé. Cette espèce a longtemps été considérée comme le modèle de référence des coronavirus, étant donné la période de temps depuis laquelle il est étudié ainsi que la disponibilité d'un modèle animal lui étant spécifique. Puisqu'il appartient au même groupe génétique que le HCoV-OC43 et qu'il partage de nombreuses propriétés structurales et fonctionnelles avec ce dernier, le MHV

est également considéré comme étant l'homologue murin de la souche humaine OC43 (Arbour *et al.*, 1999a).

Tableau 1. Les espèces de coronavirus

Groupe génétique	Nom (en anglais) *	Abréviation
1	<i>Canine coronavirus</i>	CCoV
	<i>Feline coronavirus</i>	FCoV
	<i>Feline infectious peritonitis virus</i>	FIPV
	<i>Human coronavirus 229E</i>	HCoV-229E
	<i>Porcine epidemic diarrhea virus</i>	PEDV
	<i>Transmissible gastroenteritis virus</i>	TGEV
	<i>Porcine respiratory coronavirus</i>	PRCoV
	<i>Human coronavirus NL63</i>	HCoV-NL63
	<i>Bat coronavirus</i>	Bat-CoV
2A	<i>Bovine coronavirus</i>	BCoV
	<i>Human coronavirus OC43</i>	HCoV-OC43
	<i>Human enteric coronavirus</i>	HECoV
	<i>Murine hepatitis virus</i>	MHV
	<i>Porcine hemagglutinating encephalomyelitis virus</i>	HEV
	<i>Puffinosis coronavirus</i>	PCoV
	<i>Rat coronavirus</i>	RtCoV
	<i>Sialodacryoadenitis virus</i>	SDAV
	<i>Human coronavirus HKU1</i>	HCoV-HKU1
2B	<i>Severe acute respiratory syndrome coronavirus</i>	SARS-CoV
3	<i>Infectious bronchitis virus</i>	IBV
	<i>Pheasant coronavirus</i>	PhCoV
	<i>Turkey coronavirus</i>	TCoV

* Les noms de virus sont en anglais parce qu'il n'existe pas de traduction française pour certains d'entre eux.

L'intérêt accordé au MHV date principalement du début des années 80, alors que la souche JHM du MHV fut associée à la neuroinvasion chez le rat (Sorensen *et al.*, 1980). Cette

découverte suggérait une implication potentielle du virus dans le déclenchement de certaines pathologies du SNC, dont des pathologies démyélinisantes comme la sclérose en plaques. Elle ouvrit aussi la porte à de nombreuses études visant à établir un lien entre les coronavirus, les pathologies démyélinisantes et le système immunitaire (Watanabe *et al.*, 1983), suggérant ainsi une implication du système immunitaire dans ces pathologies. Le MHV devint alors une référence importante dans l'étude des pathologies démyélinisantes induites par le système immunitaire. Le mode de propagation du virus au SNC a également suscité beaucoup d'intérêt et demeure à ce jour controversé. Suite à des inoculations intranasales, il a été démontré que des souches neurotropes du MHV pouvaient pénétrer dans le cerveau par l'entremise du nerf olfactif et se répandre à l'intérieur du SNC de la souris (Barnett et Perlman, 1993). D'autres voies permettant au MHV d'atteindre le SNC, telles que les systèmes sanguins et lymphatiques, ont aussi été identifiées (Barthold et Smith, 1992). Étant donné son niveau de caractérisation très poussé et les similitudes qu'il partage avec le HCoV-OC43, le MHV sera souvent utilisé comme référence tout au long de cette thèse.

1.1.2 Le SARS-coronavirus

Le SARS-CoV est l'agent responsable du SRAS. Comme le MHV, il représente l'une des espèces de coronavirus les mieux caractérisées. Malgré le court intervalle de temps qui s'est écoulé depuis son isolement à l'automne 2002, des milliers d'études ont déjà porté sur le SARS-CoV étant donné son niveau de virulence très élevé ainsi que la menace toujours présente d'une nouvelle éruption du virus. Malgré la caractérisation très poussée du SARS-CoV à différents niveaux, aucune cure efficace contre le virus n'a toujours été mise sur pied. Cependant, de nombreux traitements, incluant plusieurs vaccins (Zhu, 2004), sont présentement à l'étude chez l'animal (Cinatl Jr *et al.*, 2005; Groneberg *et al.*, 2005) et à l'essai chez l'humain (Peiris *et al.*, 2004).

Le SARS-CoV a été isolé pour la première fois en novembre 2002, dans la province de Guangdong, dans le sud de la Chine (Peiris *et al.*, 2004; Ziebuhr, 2004). Suite à la pandémie provoquée par cette éruption, plusieurs mesures ont été prises, tant au niveau publique que scientifique (Marshall et Enserink, 2004), et le SARS-CoV est ainsi devenu l'un des sujets de recherche les plus exploités dans le milieu de la recherche en santé et en infectiologie. Bien que l'étendue de la maladie était surtout restreinte à l'Asie, de nombreux autres cas ont été recensés ailleurs dans le monde, dont plusieurs dans la région de Toronto, au Canada. Une deuxième

éruption du virus est aussi survenue en 2003-2004. Celle-ci était par contre très mineure par rapport à la première et aurait été provoquée suite à une contamination survenue en laboratoire (Li *et al.*, 2005a).

Le virus aurait initialement été transmis à l'homme dans un marché, à l'intérieur duquel on retrouvait des animaux vivants. Plusieurs animaux porteurs du virus auraient été identifiés, mais le principal vecteur de la maladie semblerait être la civette (Guan *et al.*, 2003). Des infections par le même virus ont aussi été détectées chez d'autres animaux et chez des humains travaillant au même marché, suggérant un mode de transmission interspèce possible. Tous les isolats animaux recensés possédaient une séquence de 29 nucléotides, qui n'était pas retrouvée chez la plupart des isolats humains. Cette distinction entre les souches humaines et animales laisse croire que la transmission des humains vers les animaux serait donc peu probable (Guan *et al.*, 2003). Bien que la civette ait transmis le virus à l'humain, elle ne représenterait pas le réservoir de la maladie, étant donné que l'animal n'est pas tout à fait asymptomatique. Le réservoir naturel du virus aurait été identifié dernièrement et serait plutôt la chauve-souris (Li *et al.*, 2005b).

Bien qu'un lien avec des troubles reliés au SNC ait été avancé (Hung *et al.*, 2003; Tsai *et al.*, 2005), le SARS-CoV est surtout reconnu pour les troubles respiratoires qu'il occasionne. Les symptômes caractéristiques du SRAS sont une forte fièvre, des malaises, des maux de tête, de la toux et les voies respiratoires inférieures sont parfois atteintes (Anand *et al.*, 2003). Le taux de mortalité de la maladie se situerait entre 10 et 15% (Babcock *et al.*, 2004) et la maladie affecterait plus particulièrement les personnes âgées et/ou les personnes ayant des prédispositions au niveau du système immunitaire. Pour les mêmes raisons que le MHV, le SARS-CoV sera également utilisé pour faire des parallèles avec le HCoV-OC43, tout au long de cet ouvrage.

1.1.3 Le coronavirus humain OC43

Le HCoV-OC43 représente le centre des travaux présentés dans cette thèse. Il fait partie de l'un des deux sérotypes humains identifiés dans le milieu des années 60. Bien que le virus soit surtout associé avec des troubles des voies respiratoires supérieures, un lien entre le déclenchement de certaines maladies neurodégénératives, comme la sclérose en plaques et le Parkinson, et le HCoV-OC43 a déjà été suggéré (Talbot, 1995; Talbot, 1997). À l'occasion, le HCoV-OC43 peut également être associé à d'autres types d'infections plus rares, tels la rhinite,

l'entérite, l'hépatite, la péritonite et l'encéphalomyélite (Compton *et al.*, 1993). Plus récemment, il a aussi été démontré, dans le laboratoire du Dr Pierre Talbot, que ce virus possédait des propriétés neurotropes et neuroinvasives puisque celui-ci était capable de causer des infections persistantes dans des lignées cellulaires neurales (Arbour *et al.*, 1999a) et que son ARN était aussi retrouvé au niveau du cerveau humain (Arbour *et al.*, 2000). Le HCoV-OC43 pourrait ainsi potentiellement être associé à des pathologies du système nerveux. Parmi les deux sérotypes humains disponibles en 2001, OC43 fut sélectionné plutôt que 229E en raison de la grande infectivité du virus en cultures primaires de cellules neurales (Bonavia *et al.*, 1997) et parce que l'association avec les désordres d'ordre neurologique était plus prépondérante (Arbour *et al.*, 2000).

En raison de la recrudescence de l'intérêt accordé aux coronavirus, survenue suite à l'émergence du SARS-CoV à l'automne 2002, une surveillance plus accrue des éruptions d'infections virales a été entreprise. Les virus plus apparentés au SARS-CoV, comme le HCoV-OC43, ont principalement fait l'objet d'une surveillance plus particulière. Des éruptions du HCoV-OC43 avaient cependant déjà été rapportées dans le passé. Par exemple, en février 2001, une épidémie de HCoV-OC43 est survenue en France, en Normandie plus précisément (Vabret *et al.*, 2003). Différentes souches virales, dont le HCoV-OC43, avaient alors été isolées chez des patients atteints du rhume. En août 2002, une éruption plus sévère de HCoV-OC43 est aussi survenue dans un centre hospitalier de Melbourne, en Australie (Birch *et al.*, 2005). Les symptômes alors rencontrés étaient très apparentés à ceux de la grippe. Des études diagnostiques par RT-PCR ont cependant révélé que 59% des patients étaient infectés par le HCoV-OC43 et que les infections n'étaient pas mixtes, démontrant que le virus peut être une cause de morbidité chez des personnes plus âgées ou prédisposées.

Tel que mentionné précédemment, le HCoV-OC43 appartient au deuxième groupe génétique des coronavirus, comme le SARS-CoV. Étant donné les similitudes qui existent au niveau des séquences nucléotidiques et en acides aminés entre d'importants domaines fonctionnels des deux virus, la souche OC43 pourrait s'avérer un excellent modèle d'étude pour le SARS-CoV. Cette idée est abordée dans la publication no.1 de cette thèse (St-Jean *et al.*, 2004) et propose l'utilisation du HCoV-OC43 comme modèle d'étude du SARS-CoV, évitant ainsi les inconvénients entraînés par un confinement de niveau 3, nécessaire à l'étude du SARS-CoV. Pour des raisons similaires, la même hypothèse est aussi supportée par d'autres spécialistes des coronavirus (Gorbalenya *et al.*, 2004), qui croient que le coronavirus bovin

(BCoV), qui présente 95% d'identité avec le HCoV-OC43 (St-Jean *et al.*, 2004), pourrait lui aussi servir de modèle au SARS-CoV. Il a également été récemment démontré que le HCoV-OC43 et le SARS-CoV partageaient de nombreuses propriétés antigéniques. Lors de cette étude, une réactivité croisée aurait été observée entre des anticorps spécifiques au SARS-CoV et des anticorps spécifiques aux HCoV-OC43 et -229E (Che *et al.*, 2005). Les antigènes du SARS-CoV et du HCoV-OC43 entraînaient cependant une plus forte réaction croisée que ceux du SARS-CoV et du HCoV-229E, bien que les déterminants antigéniques responsables de ces réactions ne soient pas connus.

Il est à noter qu'un modèle animal spécifique au HCoV-OC43 est aussi disponible. Il a été développé dans le laboratoire du Dr Pierre Talbot par le Dr Héléne Jacomy (Jacomy et Talbot, 2001). Les animaux infectés dans ce modèle sont des souris C57Bl/6 ou BALB/c, lesquelles produisent une réponse antivirale qui varie en fonction du fond génétique immunitaire qu'elles possèdent (Th1 et Th2 respectivement) (Watanabe *et al.*, 2004). Sachant que le HCoV-OC43 possède des propriétés neurotropes et neuroinvasives chez la souris, ces dernières sont infectées par voie intracrânienne de façon à contrôler plus précisément les doses administrées et ainsi favoriser une reproductibilité du modèle. L'inoculation du virus à ces souris a donc conduit à une infection généralisée du SNC, ciblant principalement les neurones, lesquels dégénéraient suite à une vacuolation (Jacomy et Talbot, 2003). Les lésions observées au SNC ne semblaient pas causées par le système immunitaire et étaient plutôt une conséquence directe de l'infection virale, qui dégénérait la majeure partie du temps en encéphalite. Jusqu'à maintenant, ce modèle a donc permis une confirmation *in vivo* des propriétés neurotropes et neuroinvasives du HCoV-OC43.

1.2 Les propriétés du virion

Les coronavirus sont des virus enveloppés (Figure 1). L'enveloppe qui les entoure dérive d'un compartiment interne de la cellule dans laquelle le virus a bourgeonné. Il a d'ailleurs déjà été démontré, pour le TGEV, que la composition lipidique du virus dépend de la cellule dans laquelle il s'est multiplié et que cette composition reflète celle de la dite cellule (Myint, 1994). Les coronavirus possèdent un génome composé d'un ARN monocaténaire de polarité positive et représentent les virus à ARN dont le génome est le plus imposant (Masters, 1999). La structure protéique du virus en fait également un virus très imposant, avec un diamètre qui varie entre 100 et 160 nm (Gagneur *et al.*, 2002). Le virus contient plusieurs protéines structurales et est

notamment entouré d'une couronne de spicules qui lui vaut son nom, étant donné l'étymologie grecque du mot *corona*, qui signifie couronne.

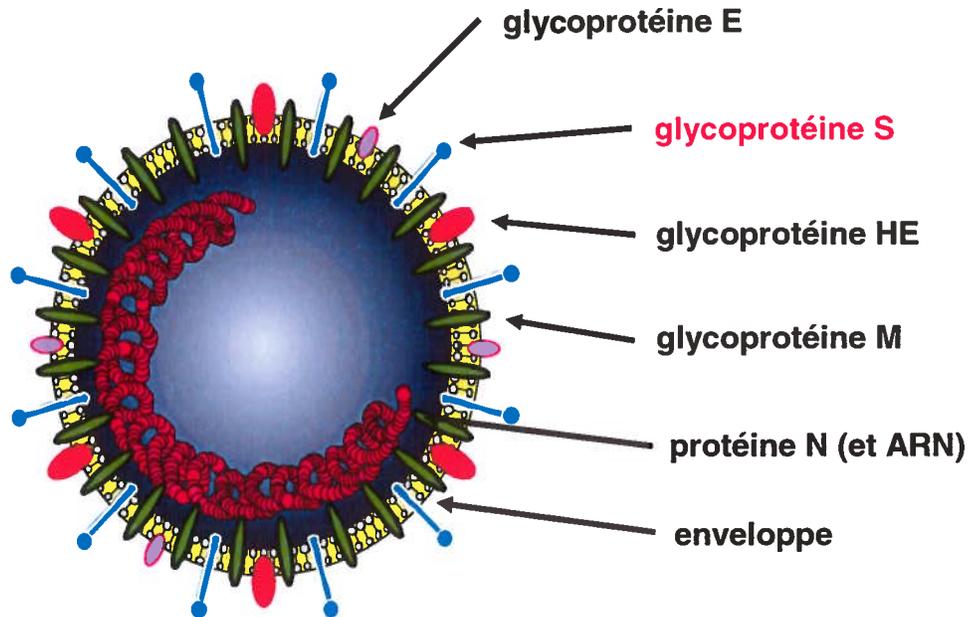


Figure 1. Représentation schématique d'une particule virale du HCoV-OC43 et des protéines structurales la composant. E, enveloppe; S, spicule; HE, hémagglutinine-estérase; M, membrane; N, nucléocapside.

1.2.1 Le génome

Les coronavirus possèdent un génome d'ARN dont la longueur varie entre 27 et 31 kilobases (kb) (Lai, 1996). L'ARN génomique viral, tout comme les ARN messagers (ARNm) sous-génomiques, est coiffé à son extrémité 5' (Ivanov *et al.*, 2004b) et possède une queue de poly(A) à son extrémité 3' (Spagnolo et Hogue, 2001). Le génome des coronavirus est aussi muni d'une séquence *leader* de 70 à 90 nucléotides à son extrémité 5' (Lai, 1996). Cette séquence *leader* est présente sur tous les ARNm viraux et est toujours identique à la séquence *leader* située en 5' du génome viral (Alonso *et al.*, 2002). La séquence *leader* joue un rôle important au niveau de la transcription et de la réplication des ARN viraux. Les fonctions attribuées à cette séquence seront cependant détaillées un peu plus loin.

Les deux extrémités du génome des coronavirus sont aussi pourvues de séquences non-traduites, communément appelées UTR. Bien qu'elles n'encodent aucun gène viral, ces séquences jouent un rôle important au niveau de la transcription des ARN viraux. La séquence nucléotidique des UTR 5' et 3' permettrait à celles-ci de se replier et de former des structures secondaires par complémentarité avec d'autres séquences qui leur sont adjacentes (Liu *et al.*, 2001). Ces séquences ont principalement été caractérisées chez le MHV, à l'extrémité 3' du génome viral. Une structure en *bulged stem-loop*, essentielle à la réplication du virus, serait donc située immédiatement en aval du gène de nucléocapside (N), suivie d'une deuxième structure, en *pseudoknot*, elle aussi essentielle à la réplication du virus. Ces deux éléments, mutuellement exclusifs l'un à l'autre, interagiraient donc ensemble et pourraient réguler une transition survenant durant la synthèse des ARN viraux (Goebel *et al.*, 2004). Les 301 nucléotides de l'UTR 3' des différentes souches de MHV seraient très conservés et présenteraient une identité de 99% avec toutes les autres souches de MHV, et 69% d'identité avec le BCoV. Les UTR 3' du BCoV et de MHV seraient d'ailleurs interchangeables (Goebel *et al.*, 2004). Certains virus, comme les picornavirus, possèdent également des *internal ribosome entry site* (IRES) dans l'UTR 5' de leur génome. Ces structures secondaires d'ARN permettent l'attachement du complexe ribosomique aux ARNm viraux et d'initier la traduction. Il a cependant été démontré que de telles structures étaient absentes du génome des coronavirus (Dove *et al.*, 2004).

Bien que l'organisation du génome de chacun des coronavirus soit très similaire, le nombre de cadres de lecture ouverts (ORF) varie selon l'espèce. Les ORF représentent des séquences du génome dont la traduction peut mener à la synthèse d'un peptide ou d'une protéine virale. Chaque ORF est habituellement représenté par un gène et on peut en compter entre 9 et 14 par génome chez les coronavirus. Du point de vue génomique, l'une des principales caractéristiques des coronavirus réside dans l'ORF 1ab, ou réplicase 1ab, qui à lui seul s'étend sur plus des deux tiers du génome viral, à l'extrémité 5' (Thiel *et al.*, 2003a). Le nombre d'ORF ainsi que la composition de la région 3' du génome varie beaucoup d'un coronavirus à l'autre, mais inclut toujours les gènes qui encodent les protéines structurales de surface (S), de la membrane (M), de l'enveloppe (E) et de la nucléocapside (N) (Snijder *et al.*, 2003).

Le génome du HCoV-OC43 a une longueur de 30 713 nucléotides, excluant la queue de poly(A), dont la taille moyenne est de 28 nucléotides (St-Jean *et al.*, 2004). Le génome du virus possède des caractéristiques qui sont surtout représentatives des coronavirus de groupe 2. Tous les détails concernant les caractéristiques du génome du HCoV-OC43 sont regroupés dans la

publication no.1 de cette thèse. Une représentation schématique du génome du HCoV-OC43 est d'ailleurs présentée dans la Figure 1 de cette publication.

1.2.2 Les protéines structurales

Le génome des coronavirus encode deux types de protéines, selon qu'elles sont encapsidées ou non dans les particules virales. Ces protéines sont respectivement définies comme étant structurales ou non-structurales (Ito *et al.*, 2005). Généralement, les protéines structurales des virus enveloppés comme le coronavirus, font partie de la capsid du virus, de l'enveloppe ou des deux. En partant de l'extrémité 5' du génome viral, les gènes de protéines structurales retrouvés chez les coronavirus sont respectivement S, E, M et N (Figure 1) (Lai et Cavanagh, 1997).

Chez certains coronavirus de groupe 2, comme le HCoV-OC43 et certains MHV, on retrouve un cinquième gène de protéine structurale, situé en aval de l'ORF 1ab (Wurzer *et al.*, 2002). Ce gène encode une hémagglutinine-estérase (HE), une enzyme présente à la surface du virion, capable d'induire l'hémagglutination des globules rouges et possédant une activité acétyl estérase (Klausegger *et al.*, 1999). Il s'agit d'une glycoprotéine de type I, qui ne traverse l'enveloppe virale qu'une seule fois (Figure 1) (Nguyen et Hogue, 1997). Chez certains coronavirus comme le MHV-A59, le gène HE serait présent dans le génome viral mais serait dépourvu du codon d'initiation de la traduction, faisant de celui-ci un pseudogène (Yokomori *et al.*, 1991). Il a été suggéré que le gène HE aurait pu être introduit dans un génome ancestral de coronavirus par recombinaison avec un ARNm codant pour la protéine HE du virus influenza C (Holmes, 2003). À cause de cette ressemblance entre les deux protéines et à cause des fonctions d'attachement au récepteur que la protéine HE du virus influenza C occupe, on a longtemps cru que la protéine HE des coronavirus partageait les mêmes fonctions que celle de l'influenza C (Myint, 1994). Cependant, il a été démontré que les fonctions d'attachement au récepteur des coronavirus résidaient dans la protéine S et que cette même protéine possédait un pouvoir d'hémagglutination plus important que celui de HE (Schultze *et al.*, 1991). Bien que la fonction précise de la protéine HE demeure encore mal comprise, celle-ci pourrait favoriser l'infectivité virale en concentrant le virus aux membranes cellulaires de l'hôte (Thackray et Holmes, 2004). Il a également été récemment démontré que la protéine HE pouvait accentuer la neurovirulence du MHV-A59, probablement en favorisant la dissémination du virus dans le SNC (Kazi *et al.*, 2005). Par contre, son absence chez plusieurs coronavirus laisse croire qu'elle occuperait un rôle plutôt

accessoire et que sa présence ne serait pas indispensable à la réplication virale. Il a d'ailleurs été démontré que la protéine HE du BCoV n'était pas essentielle à la réplication du virus et que la protéine S seule pouvait alors remplir les fonctions reliées à l'entrée virale et mener à l'infection *in vitro* (Popova et Zhang, 2002).

Toujours à partir de l'extrémité 5' du génome viral, le gène S représente le deuxième gène de protéine structurale après HE, lorsque ce dernier est présent. La protéine S des coronavirus est nommée ainsi à cause de la projection qu'elle forme à la surface des particules virales, en forme de spicule ou de *spike* (Figure 1). La description structurale et les fonctions reliées à la protéine S seront élaborées un peu plus loin, plus en détails, étant donné le rôle important que cette protéine a occupé dans les recherches effectuées au cours de ce doctorat ainsi que dans cette thèse.

Le troisième gène encodant une protéine structurale est le gène de l'enveloppe, E. Il s'agit de la plus petite protéine structurale des coronavirus avec une masse moléculaire d'environ 9,5 kilodaltons (kDa) (Figure 1) et on la retrouve aussi à des taux très faibles dans les virions (Nal *et al.*, 2005). Bien que le rôle de la protéine E n'ait pas encore été clairement défini, il semblerait qu'elle joue un rôle majeur au niveau de l'assemblage des particules virales (Lim et Liu, 2001), de leur morphogenèse (Ortego *et al.*, 2002) et de leur bourgeonnement à l'extérieur des cellules infectées (Kuo et Masters, 2003). Il a aussi été récemment démontré que la protéine E intervenait dans la formation de canaux à cations monovalents, chez le SARS-CoV notamment (Wilson *et al.*, 2004). La protéine E des coronavirus ferait partie du même groupe que les protéines M2, BM2 et NB de l'influenza, et Vpu et Vpr du virus de l'immunodéficience humaine (VIH), qui sont également des protéines qui forment des canaux à ions. On ne sait cependant toujours pas si l'activité de la protéine E des coronavirus joue un rôle dans la réplication virale. Chez l'influenza et le VIH, l'activité des canaux interviendrait au niveau de la décapsidation du virus et du transport des protéines virales durant l'assemblage (Wilson *et al.*, 2004). Bien que la protéine E ne semble pas être essentielle à la réplication du MHV *in vitro* (Kuo et Masters, 2003), il semblerait qu'elle soit cependant essentielle à la réplication du TGEV (Ortego *et al.*, 2002), ce qui rend son rôle dans la réplication des coronavirus encore plus controversé.

Le gène M encode la protéine de membrane des coronavirus. Cette glycoprotéine contient trois domaines transmembranaires (Kuo *et al.*, 2000) (Figure 1) et serait aussi la protéine virale la plus abondante (Kuo et Masters, 2002). Elle joue un rôle très important au niveau de

l'assemblage des virions (de Haan *et al.*, 1999). Elle est notamment essentielle au bourgeonnement viral, à la formation de l'enveloppe virale et interagirait avec la nucléocapside dans le but d'attacher le génome à l'enveloppe (Myint, 1994). Les protéines E et M seraient les deux seules protéines virales absolument nécessaires à l'assemblage des particules virales (Lai et Cavanagh, 1997). Lorsqu'elles sont co-exprimées, ces deux protéines peuvent d'ailleurs mener à la formation de pseudoparticules, même en absence de N et S (Nal *et al.*, 2005). Durant l'assemblage, la protéine M interagirait avec plusieurs autres protéines coronavirales, telles que la protéine E (Lim et Liu, 2001), la protéine N (Fang *et al.*, 2005; He *et al.*, 2004; Kuo et Masters, 2002) et les protéines S et HE (de Haan *et al.*, 1999). D'une certaine façon, la protéine M orchestre l'assemblage des particules virales chez les coronavirus.

Le dernier gène de protéine structurale est le gène de la nucléocapside, N. Chez la plupart des coronavirus, il est situé presque complètement en aval du génome viral, en amont de l'UTR 3'. La protéine N serait composée de trois domaines structuraux hautement conservés reliés les uns aux autres par des régions moins conservées (Parker et Masters, 1990). La localisation des fonctions d'attachement à l'ARN de N demeure cependant controversée. Ces sites pourraient varier en fonction des souches de coronavirus (Huang, Q. *et al.*, 2004). La nucléocapside est présente à l'intérieur de l'enveloppe du virion et entoure l'ARN génomique viral dans le but de le protéger (Figure 1). Lorsqu'elle est associée à l'ARN génomique, la protéine N forme un complexe nommé la ribonucléoprotéine (Lai et Cavanagh, 1997). En plus des fonctions de protection qu'elle occupe, la protéine N serait impliquée au niveau de la réplication des coronavirus, principalement dans les stades précoces, et jouerait un rôle moins important au niveau de la transcription (Schelle *et al.*, 2005). Bien que le mécanisme d'action de N ne soit pas connu, on croit que la protéine pourrait jouer un rôle direct au sein du complexe de réplication-transcription ou qu'elle jouerait un rôle au niveau de la stabilisation des réplicons génomiques du virus (Almazán *et al.*, 2004).

1.2.3 Les protéines non-structurales

Les protéines non-structurales des coronavirus sont encodées par le génome viral, mais ne sont pas incorporées à l'intérieur des virions. Il en existe deux catégories selon leur emplacement dans le génome. Les protéines de la réplicase 1ab remplissent habituellement des fonctions reliées à la réplication et à la transcription du génome viral, alors que les autres gènes de protéines non-structurales, situés dans le tiers 3' du génome, remplissent des fonctions

biologiques peu caractérisées jusqu'à présent (Ito *et al.*, 2005). Étant donné le rôle majeur que jouent les protéines de la réplicase 1ab dans la biologie des coronavirus, leur fonction sera détaillée un peu plus loin. Les gènes de protéines non-structurales du tiers 3' du génome coronaviral varient d'une espèce à l'autre, et plus spécifiquement d'un groupe génétique à l'autre. C'est pour cette raison que ces gènes sont désignés comme étant *group-specific* (de Haan *et al.*, 2002a). Le génome du HCoV-OC43 en contient seulement deux, contrairement au génome du SARS-CoV, qui en contiendrait jusqu'à huit (Thiel *et al.*, 2003a). Chez la souche OC43, ces protéines sont donc ns2 (Labonté *et al.*, 1995) et ns12,9 (St-Jean *et al.*, 2004). Les gènes qui encodent ce type de protéines non-structurales sont aussi souvent qualifiés d'accessoires, étant donné que leur délétion ne serait pas létale pour le virus *in vitro*, notamment chez le MHV (de Haan *et al.*, 2002a) et le TGEV (Ortego *et al.*, 2003). Bien qu'on ne connaisse pratiquement rien sur les fonctions des produits de ces gènes et qu'aucune activité enzymatique ne leur ait été attribuée à ce jour, les protéines accessoires contribueraient à la pathogenèse virale *in vivo* (Ito *et al.*, 2005; Ortego *et al.*, 2003) et seraient importantes pour les interactions hôte-virus, contribuant ainsi à la stabilité virale *in vivo* (Tan *et al.*, 2005).

1.3 La glycoprotéine S

La glycoprotéine S des coronavirus est de loin la protéine la plus importante structurellement et fonctionnellement. La pathogenèse des coronavirus est en grande partie déterminée par cette protéine, qui comporte une multitude de propriétés biologiques importantes, telles que la liaison au récepteur ainsi que l'induction des fusions cellule-cellule et virus-cellule, de la production d'anticorps neutralisants et de l'activation de lymphocytes T cytotoxiques (Arbour *et al.*, 1999a). La protéine S contiendrait aussi des déterminants de neurovirulence, responsables de la propagation des coronavirus, du MHV notamment, dans le SNC (Dalziel *et al.*, 1986; Fleming *et al.*, 1987). Cette protéine représente également un élément central des travaux présentés dans cette thèse. C'est pour cette raison que le rôle de la protéine S sera abordé de façon très approfondie.

1.3.1 La structure de la glycoprotéine S

Une représentation schématique de la structure primaire de la glycoprotéine S est présentée dans la publication no.5 de cette thèse, à la Figure 2. S est une protéine membranaire de type I qui est séparée en deux sous-unités. Le clivage protéolytique de S mène à la formation

d'une sous-unité amino-terminale S1, qui forme la tête globulaire du péplomère, et à une partie carboxy-terminale S2, qui forme la région de la tige (Daniel *et al.*, 1993). Une particule coronavirale possède environ 200 protéines S à sa surface (Krueger *et al.*, 2001). La protéine S des coronavirus est aussi N-glycosylée (de Haan *et al.*, 1999), d'où l'appellation de glycoprotéine S. On a longtemps cru que S se présentait sous forme de dimère (Lewicki et Gallagher, 2002), mais de plus en plus d'études démontrent que la forme trimérique serait la norme (Spiga *et al.*, 2003). En fait, une étude très approfondie à ce niveau a récemment démontré que la protéine S se présentait sous forme de trimère, mais que la zone d'attachement au récepteur se présentait sous forme de dimère et que cet état dimérique était essentiel pour la liaison au récepteur et l'entrée du virus dans la cellule (Xiao *et al.*, 2004). Cette même étude a également démontré que le domaine de trimérisation de la protéine S était situé dans le segment transmembranaire, comme c'est le cas pour le virus influenza.

Étant donné les nombreuses propriétés biologiques que la glycoprotéine S des coronavirus contrôle, cette dernière renferme de nombreux domaines fonctionnels. À son extrémité amino-terminale, la protéine possède un peptide *signal* d'environ 16 acides aminés, qui lui permet d'être exportée à la membrane cellulaire qui servira d'enveloppe virale (Spiga *et al.*, 2003). En aval du peptide *signal*, on retrouve le site de liaison au récepteur cellulaire (SLR), lui aussi présent dans la sous-unité S1, mais à des positions différentes selon les espèces de coronavirus (Tan *et al.*, 2002). Toujours dans la sous-unité S1, on retrouve la région hypervariable (RHV), reconnue pour muter facilement. Cette région jouerait potentiellement un rôle dans l'interaction avec le système immunitaire (Phillips *et al.*, 2001), notamment au niveau de l'évasion immunitaire (Rowe *et al.*, 1997). D'autres études ont aussi suggéré que des altérations de cette région pourraient influencer les interactions entre les sous-unités S1 et S2 (Phillips *et al.*, 2001; Tsai *et al.*, 1999). À l'exception des segments transmembranaire et intracytoplasmique qui la composent, la sous-unité S2 contient trois petits domaines fonctionnels, juxtaposés l'un à l'autre et impliqués au niveau des processus de fusion. Ces domaines sont respectivement le peptide de fusion (PF) (Bosch *et al.*, 2003), et les *heptad repeats* (HR) 1 et 2 (Supekar *et al.*, 2004; Xu *et al.*, 2004c).

Compte tenu de l'impact majeur de la glycoprotéine S sur la pathogenèse et les propriétés biologiques des coronavirus, de nombreux scientifiques ont récemment cherché à analyser la protéine par cristallographie (Li, F. *et al.*, 2005; Xu *et al.*, 2004a; Xu *et al.*, 2004b). Ces analyses ont permis de mieux comprendre la structure de la protéine et de caractériser davantage les

interactions potentielles de la protéine S avec d'autres molécules. Bien que le niveau de caractérisation de la protéine S des coronavirus se soit accru énormément au cours des dernières années, plusieurs des mécanismes associés à la protéine demeurent mal compris. La compréhension de ceux-ci est directement liée à la compréhension du virus et au lien qui existe entre la protéine S et la pathogenèse coronavirale.

1.3.1 Les fonctions associées à la glycoprotéine S

Tout comme la glycoprotéine S, de nombreuses autres protéines coronavirales remplissent des fonctions essentielles à la biologie du virus. Cependant, c'est la protéine S qui remplit le plus grand nombre de fonctions directement reliées à la virulence (Phillips *et al.*, 2001). Les principales fonctions de S qui seront abordées dans cette section sont donc l'attachement du virus au récepteur cellulaire et l'induction des fusions virus-cellule, lesquelles sont des processus primordiaux à l'entrée virale.

Immédiatement en aval du peptide *signal* situé à l'extrémité amino-terminale de la sous-unité S1 de la protéine S, on retrouve le SLR des coronavirus. Ce domaine est localisé à la surface des projections protéiques présentes à la surface du virion. Il est donc responsable de reconnaître un récepteur cellulaire spécifique, de le lier et d'initier la cascade de réactions menant à l'entrée virale (Cullen, 2001). En d'autres termes, la protéine S est le principal déterminant du tropisme du virus et possède par conséquent l'habileté de lier un type cellulaire précis (Gallagher et Buchmeier, 2001; Hofmann et Pöhlmann, 2004). Il est à noter que les termes tropisme et invasion ne doivent pas être confondus, l'invasion représentant la capacité du virus à cibler une région ou un organe de l'hôte et de l'infecter (van Marle et Power, 2005).

Selon l'espèce de coronavirus rencontrée, la localisation du SLR dans la protéine S ainsi que sa spécificité peuvent varier. Pour les coronavirus de groupe 1 comme le *feline infectious peritonitis virus* (FIPV), le TGEV et le HCoV-229E, le récepteur cellulaire utilisé par le virus est l'aminopeptidase N ou CD13 (Hegyí et Kolb, 1998; Lachance *et al.*, 1998; Schwegmann-Webels *et al.*, 2002). Le HCoV-NL63, un autre coronavirus de groupe 1, utilise plutôt l'*angiotensin-converting enzyme 2* (ACE2) (Hofmann *et al.*, 2005), tout comme le SARS-CoV (Li *et al.*, 2003), qui appartient cependant au groupe 2. Le seul autre récepteur utilisé par les coronavirus, qui a été identifié et caractérisé, est le récepteur du MHV, le *carcinoembryonic antigen-related cell adhesion molecule 1* (CEACAM1) (Saeki *et al.*, 1997). Parmi tous les récepteurs énumérés

précédemment, le CEACAM1 est de loin le mieux caractérisé. C'est une glycoprotéine *immunoglobulin-like* très répandue dans l'organisme et qui remplit plusieurs fonctions biologiques. Il s'agit entre autres d'un facteur de croissance des vaisseaux endothéliaux, d'un inhibiteur de croissance des cellules tumorales et d'un marqueur de l'activation des lymphocytes T murins (Tan *et al.*, 2002). Le CEACAM1 humain sert également de récepteur aux souches bactériennes *Neisseria gonorrhoeae* et *meningitidis* et *Haemophilus influenzae*. Chez la souris, il en existe quatre isoformes qui peuvent tous servir de récepteur au MHV-A59 (Tan *et al.*, 2002). La liaison de la protéine S et du CEACAM1 induirait la formation d'un nouveau pont disulfure entre les deux protéines et pourrait entraîner la séparation des sous-unités S1 et S2. Les dimères de S ne s'attacheraient qu'à une seule molécule CEACAM1, contrairement au virus influenza et au VIH, où chaque monomère des glycoprotéines trimériques HA et gp120 respectivement, peut lier un ligand (Lewicki et Gallagher, 2002).

La localisation des SLR varie d'un coronavirus à l'autre. Chez MHV, le domaine de liaison au récepteur cellulaire est situé à l'extrémité amino-terminale de la protéine S, au niveau des 330 premiers acides aminés (Matsuyama et Taguchi, 2002a). Le SLR du MHV est cependant le seul à être situé complètement à l'extrémité de la protéine. Chez le HCoV-229E, le SLR est situé entre les acides aminés 407 et 547 de la sous-unité S1 (Breslin *et al.*, 2003), alors qu'il serait situé entre les acides aminés 318 et 510 chez le SARS-CoV (Wong *et al.*, 2004). Un tel site de liaison n'a toujours pas été identifié pour plusieurs espèces de coronavirus, telles que le IBV, le BCoV et le HCoV-OC43 notamment. En ce qui concerne la souche humaine OC43, on sait que la protéine S serait capable de lier l'acide N-acétyl-9-O-neuraminique, un acide sialique très répandu sur les surfaces cellulaires (Gélinas *et al.*, 2001; Künkel et Herrler, 1996). Cependant, le récepteur cellulaire spécifique au HCoV-OC43 n'a toujours pas été identifié. En plus de lier l'aminopeptidase N, le TGEV serait aussi capable de lier l'acide sialique. Cette activité d'attachement à l'acide sialique augmenterait le taux de liaison du virus aux cellules cibles, mais n'augmenterait pas l'infectivité du TGEV (Schwegmann-Webels *et al.*, 2002).

En plus de lier leur récepteur principal, d'autres coronavirus utiliseraient également des récepteurs secondaires ou alternatifs lors de l'infection. L'attachement à ces récepteurs alternatifs ne mène cependant pas toujours à l'infection, comme c'est le cas pour le TGEV lorsque l'acide sialique est utilisé plutôt que l'aminopeptidase N porcine. Le MHV peut par exemple utiliser différents isomères du CEACAM afin d'initier l'infection, mais le niveau d'infectivité varie en fonction du récepteur utilisé, le CEACAM1a étant le récepteur le plus efficace (Matsuyama et

Taguchi, 2002a; Tan *et al.*, 2002). L'exemple le mieux caractérisé de récepteur alternatif est celui du SARS-CoV qui, en plus du ACE2, peut lier deux autres récepteurs cellulaires, soit le CD209L (L-SIGN) et le CD209 (DC-SIGN) (Jeffers *et al.*, 2004). L'attachement au CD209, situé à la surface des cellules dendritiques, n'entraînerait pas d'infection virale alors que la liaison au CD209L induirait une infection productive. Le SARS-CoV pourrait donc utiliser autant ACE2 que CD209L comme récepteur, mais CD209L serait moins efficace pour initier l'infection (Jeffers *et al.*, 2004).

Suite à la liaison de la glycoprotéine S à un récepteur cellulaire spécifique, une cascade de réaction menant à l'internalisation du virus est initiée. Selon le coronavirus impliqué et par conséquent, selon la séquence de la protéine S, il y aura clivage ou non de cette dernière. Le clivage protéolytique de la protéine S survient généralement chez les coronavirus du groupe 2, puisque les coronavirus des groupes 1 et 3 ne possèdent pas de site de clivage (Breslin *et al.*, 2003). Les exceptions rencontrées sont le HCoV-OC43 (Künkel et Herrler, 1996), le SARS-CoV (Babcock *et al.*, 2004), et quelques souches de MHV, comme le MHV-2 (Hingley *et al.*, 2002), lesquels ne possèdent pas de site de clivage protéolytique malgré leur appartenance au groupe 2. Le clivage de la protéine S serait effectué par une enzyme cellulaire nommée la furine, ou par une enzyme semblable à celle-ci (de Haan *et al.*, 2004). La furine est une protéase cellulaire du système sécrétoire de plusieurs types cellulaires, qui serait responsable de la transformation des précurseurs d'hormones en hormones matures (Hingley *et al.*, 2002). Chez MHV-A59, le clivage de la protéine S surviendrait entre les résidus 717 et 718 à la séquence RRAHR↓SVS, séquence qui ressemble au site de reconnaissance de la furine RXR/KR (de Haan *et al.*, 2004). Chez les coronavirus dont l'infection nécessite un clivage de S, la furine serait un déterminant important du tropisme cellulaire. C'est notamment le cas du virus influenza, dont la protéine de surface HA doit être clivée par la furine afin que le virus puisse initier l'infection (Garcia-Sastre *et al.*, 1998).

Le but du clivage protéolytique de la protéine S lors de l'infection est de permettre le repliement de la sous-unité S2 et l'exposition des domaines de fusion de l'extrémité carboxy-terminale (PF, HR1 et HR2) à la membrane plasmique de la cellule cible (Bosch *et al.*, 2003). Bien que certains coronavirus ne subissent pas de clivage protéolytique, il semblerait que les glycoprotéines S de ces virus soient possiblement assez flexibles pour exposer le peptide de fusion membranaire suite à l'attachement au récepteur (Breslin *et al.*, 2003). Chez les protéines S de coronavirus, et contrairement à la plupart des protéines de surface des autres virus à ARN, le clivage n'aurait pas d'incidence sur la virulence (Hingley *et al.*, 2002). Il semblerait cependant

qu'une augmentation du clivage protéolytique de S stimulerait les fusions cellule-cellule, un phénomène complètement indépendant des fusions virus-cellule (de Haan *et al.*, 2004). Bien que les fusions cellule-cellule ne soient pas nécessaires à la virulence, elles mèneraient à la formation de syncytiums et, dans certains cas, de plages de lyse *in vitro* (Haijema *et al.*, 2003). En fait, le non-clivage de la protéine S serait associé à une interaction plus stable entre S1 et S2, qui corrèlerait avec une perte de l'activité fusiogénique entre les cellules (Krueger *et al.*, 2001).

La liaison de la glycoprotéine S avec son récepteur déclencherait donc une série d'évènements entraînant la fusion des enveloppes virales et cellulaires (Gallagher et Buchmeier, 2001). Le premier évènement à survenir suite à l'attachement au récepteur serait un changement conformationnel de la protéine S (Matsuyama et Taguchi, 2002b). Lors de ce phénomène, les sous-unités S1 et S2 interagiraient ensemble dans le but de rapprocher la sous-unité S2 de la membrane plasmique (Matsuyama et Taguchi, 2002a). Suite au clivage ou au repliement de la protéine S, un nouveau changement conformationnel de la sous-unité S2 surviendrait. Celui-ci consisterait en un repliement de la protéine entre les domaines HR1 et HR2, de façon à exposer le PF à la membrane plasmique d'une manière optimale (Bosch *et al.*, 2003; Tsai *et al.*, 2003). Les domaines HR1 et HR2 seraient en fait des hélices α contenant des répétitions de sept acides aminés et qui sont caractérisés par la présence de résidus hydrophobes en positions "a" et "d" de l'hélice (Kliger et Levanon, 2003). La plus grande similitude au niveau de la séquence du gène S des coronavirus surviendrait au niveau des domaines HR, démontrant ainsi leur importance du point de vue fonctionnel (Hofmann et Pöhlmann, 2004). Lors du changement de conformation de la protéine, les domaines trimériques HR1 et HR2 s'associeraient de façon antiparallèle pour former un complexe oligomérique extrêmement stable. Ce complexe formerait une structure en bâtonnet constituée de six hélices, caractéristique des protéines de fusion de type I, et permettant au PF d'entrer en contact avec la membrane plasmique (Ingallinella *et al.*, 2004). La conformation adoptée par le complexe à cet instant serait qualifiée de *postfusion* ou de *fusion-active* et déclencherait le phénomène de fusion entre les membranes virale et cellulaire (Supekar *et al.*, 2004).

Bien que les processus de fusion virus-cellule des coronavirus ne soient pas clairement caractérisés, les similitudes qui existent entre le complexe de fusion des coronavirus et celui du virus influenza laissent croire que les deux virus partageraient des mécanismes semblables (Ingallinella *et al.*, 2004). L'hémagglutinine HA du virus influenza est la protéine de fusion de classe I la mieux caractérisée (Supekar *et al.*, 2004). Cependant, contrairement à celle-ci, la

protéine S des coronavirus serait particulière puisqu'elle ne subirait pas de clivage à l'extrémité amino-terminale du PF (Duquerroy *et al.*, 2005). Le PF des coronavirus serait localisé immédiatement en amont du domaine HR1 (Guillén *et al.*, 2005) et posséderait les caractéristiques classiques des peptides de fusion de classe I. Ces caractéristiques sont notamment l'hydrophobicité, une conformation en hélice α , la présence de séquences hydrophobes riches en alanines, glycines et phénylalanines et la présence de séquences canoniques tripeptides (YFG ou FXG) très conservées chez plusieurs virus, tels que les rétrovirus, les paramyxovirus et les filovirus (Sainz Jr *et al.*, 2005; Taguchi et Shimazaki, 2000). Au moment de la fusion, on sait qu'il y aurait une dissociation du trimère du PF afin d'induire la réaction de fusion, suivie d'une réassociation du peptide (Duquerroy *et al.*, 2005). Des régions adjacentes au PF, telles que le domaine pré-transmembranaire de la protéine S, pourraient aussi jouer un rôle très important au niveau de la fusion et pourraient être impliquées au niveau de la déstabilisation membranaire et de la formation du pore de fusion (Guillén *et al.*, 2005). Cependant, on ne connaît toujours pas précisément les relations qui existent entre les événements entourant le processus de fusion, tels que le clivage protéolytique, l'activation de la protéine S et la formation du complexe de fusion (Follis *et al.*, 2005).

Plusieurs processus plus ou moins caractérisés pourraient également intervenir durant ou après l'attachement au récepteur cellulaire chez les coronavirus. Par exemple, afin de pénétrer à l'intérieur de la cellule, certains coronavirus doivent interagir avec des radeaux lipidiques, qui consistent en des agglomérations de cholestérol présentes à la surface cellulaire, qui permettent de concentrer des protéines membranaires requérant une interaction entre elles (Thorp et Gallagher, 2004). C'est notamment le cas du MHV (Choi *et al.*, 2005) et du HCoV-229E, lequel aurait aussi besoin des radeaux lipidiques pour s'attacher au récepteur (Nomura *et al.*, 2004). Bien que le mécanisme impliquant les radeaux lipidiques durant l'infection ne soit toujours pas clair, il a été démontré que la déplétion du cholestérol n'affectait pas l'attachement du MHV, mais plutôt son entrée (Choi *et al.*, 2005). Selon une autre étude, il a également été démontré que le cholestérol pourrait jouer un rôle direct et indépendant du CEACAM1 dans l'entrée du MHV dans la cellule cible, indiquant que les radeaux lipidiques sont probablement impliqués dans l'étape d'internalisation du virus (Thorp et Gallagher, 2004). En ce qui concerne le HCoV-229E, l'internalisation se ferait par l'intermédiaire d'invaginations membranaires nommées cavéoles, suite à la liaison avec les radeaux lipidiques (Nomura *et al.*, 2004). D'autres virus comme le SV40, l'échovirus-1 et le polyomavirus pénètrent à l'intérieur des cellules par les cavéoles. Le

HCoV-OC43 utiliserait aussi un récepteur séquestré aux cavéoles afin d'amorcer son infection (Nomura *et al.*, 2004).

À ce jour, il n'y a toujours pas de consensus sur le mode d'entrée des coronavirus dans les cellules susceptibles (Peiris *et al.*, 2004). En plus de la fusion membranaire, certains coronavirus sembleraient utiliser l'endocytose afin de pénétrer à l'intérieur des cellules cibles. Bien que ce mécanisme ait été démontré chez quelques virus, dont le HCoV-229E (Blau et Holmes, 2001) et le SARS-CoV (Hofmann et Pöhlmann, 2004), l'intervention de la fusion membranaire et de l'endocytose lors d'une même entrée virale, comme c'est le cas chez le virus influenza, demeure un point controversé. Le virus de l'influenza pénètre les cellules cibles par endocytose et lorsque le pH devient très acide, la protéine HA induit une fusion avec la membrane endosomale, contrairement à ce qui est observé chez le VIH qui fusionne directement avec la membrane cellulaire en utilisant sa gp120 (Matsuyama et Taguchi, 2002b). Puisque l'entrée virale des coronavirus de groupe 1, comme le HCoV-229E, semble être un phénomène dépendant du pH et que l'entrée virale des coronavirus de groupe 2, comme le MHV, semble être indépendante du pH, il serait raisonnable de croire que les virus de groupe 1 et de groupe 2 favorisent respectivement un mécanisme d'endocytose (Blau et Holmes, 2001) et de fusion membranaire (Krueger *et al.*, 2001). Il a par contre été récemment démontré que le SARS-CoV pouvait utiliser les deux mécanismes étant donné qu'une acidification des endosomes semble être requise afin que l'entrée virale s'effectue efficacement (Simmons *et al.*, 2004). La protéine S du SARS-CoV pourrait donc induire des fusions de façon dépendante et indépendante du pH, mais les paramètres impliqués dans l'induction des signaux qui mènent au déclenchement de ces phénomènes demeurent inconnus (Hofmann et Pöhlmann, 2004). Bien que les coronavirus de groupe 1 pourraient employer un mécanisme d'entrée par endocytose, on ignore si un lien existe entre l'efficacité de clivage de la protéine S et le mode d'entrée utilisé par le virus.

1.4 Autres propriétés fonctionnelles

Les fonctions attribuées à la protéine S sont directement ou indirectement reliées à la virulence et à la pathogenèse des coronavirus (Gallagher et Buchmeier, 2001). Cependant, d'autres protéines virales jouent aussi des rôles essentiels dans la biologie des coronavirus. C'est notamment le cas de la réplicase 1ab, qui peut produire à elle seule au moins 13 protéines différentes chez tous les coronavirus (Ziebuhr *et al.*, 2000). Ces protéines remplissent toutes des

fonctions primordiales au niveau de la réplication et de la transcription des ARN génomiques et sous-génomiques viraux et jouent donc un rôle important dans le cycle réplcatif des coronavirus.

1.4.1 La réplicase 1ab

La réplicase 1ab des coronavirus est représentée par l'ORF 1ab. Il s'agit du plus long gène chez tous les coronavirus. Celui-ci s'étend sur plus des deux tiers du génome viral, sur une longueur qui dépasse 21 kb chez le HCoV-OC43 (St-Jean *et al.*, 2004). Le gène est appelé 1ab parce qu'il est le premier gène du génome viral et qu'il contient deux ORF reliés l'un à l'autre par un *ribosomal frameshift -1* (Thiel *et al.*, 2003a). Le *ribosomal frameshift* consiste en une structure secondaire d'ARN qui permettrait au complexe de traduction de l'ARN de poursuivre son chemin après l'ORF 1a. Dans la majorité des cas, la traduction s'arrête après l'ORF1a, mais occasionnellement le complexe de traduction de l'ARN synthétise une protéine énorme d'environ 800 kDa correspondant à l'ORF 1ab (Thiel *et al.*, 2003a). Les protéines résultant de cette traduction, ORF 1a et ORF 1ab, sont en fait des polyprotéines puisqu'elles sont clivées par des protéases virales qui permettent la génération de plusieurs produits de clivage selon la souche de virus rencontrée. Chez HCoV-OC43, 15 produits seraient générés suite au clivage effectué par les trois protéases du virus, soit les protéases accessoires *papain-like* (PLP) 1 et 2, et la protéase principale, la *3C-like* (3CLpro ou Mpro) (St-Jean *et al.*, 2004). Les produits de clivage ainsi produits représentent les protéines non-structurales de l'extrémité 5' du génome viral, lesquelles sont associées aux fonctions de réplication et de transcription du virus, comme le nom de réplicase l'indique (Ziebuhr *et al.*, 2000).

Une représentation schématique de la polyprotéine de la réplicase 1ab du HCoV-OC43 est présentée à la Figure 2 de la publication no.1 de cette thèse. Les domaines les plus importants dans la réplication du virus ainsi que les sites de clivage potentiels y sont présentés. La réplicase 1ab la mieux caractérisée jusqu'à présent est définitivement celle du SARS-CoV et cette dernière sera principalement utilisée comme référence lors de la description de la protéine homologue du HCoV-OC43. La principale différence entre les deux polyprotéines réside au niveau de la PLP1, qui est absente chez le SARS-CoV (Thiel *et al.*, 2003a). Comme dans le cas du IBV qui ne possède qu'une seule PLP, le clivage des trois premiers sites serait donc réalisé exclusivement par la PLP2 chez le SARS-CoV, contrairement aux autres coronavirus, chez qui les deux PLP se partageraient le clivage des trois premiers sites (Harcourt *et al.*, 2004). Chez tous les coronavirus, les positions et les séquences des sites de clivage des différentes protéases sont relativement bien

conservées, surtout pour la 3CLpro. Comme dans le cas du HCoV-OC43, la 3CLpro du SARS-CoV clive 11 sites dans la polyprotéine de la réplicase et joue un rôle très important puisqu'elle permet notamment le relargage de l'ARN polymérase ARN-dépendante (RdRp) et de l'hélicase virale (Thiel *et al.*, 2003a). Les 15 produits de clivage générés suite à la maturation de la polyprotéine de la réplicase 1ab du HCoV-OC43 seront donc énumérés et décrits à partir de l'extrémité amino-terminale de la polyprotéine.

Le premier produit de clivage de la réplicase 1ab est désigné nsp1, pour *non structural protein 1*. Bien que ce produit de clivage soit peu caractérisé et que la fonction qu'il remplit demeure inconnue, on sait qu'il serait relargué très rapidement suite à l'entrée du virus dans la cellule (Harcourt *et al.*, 2004). Il a aussi été récemment démontré que l'extrémité amino-terminale du produit nsp1 de MHV serait requise pour une réplication virale optimale *in vitro* (Brockway et Denison, 2005). Comme nsp1, nsp2 demeure également très peu caractérisée. Lors de l'infection, nsp2 pourrait notamment lier les produits de clivage nsp3, nsp8 et nsp9 (Prentice *et al.*, 2004; Tan *et al.*, 2005). Le produit de clivage nsp3 est de loin celui qui est le plus imposant chez tous les coronavirus. Il contient notamment les domaines des PLP1 et PLP2, ou seulement le domaine PLP2 chez le SARS-CoV et IBV (Thiel *et al.*, 2003a). En plus des fonctions de clivage protéolytique qui leur sont attribuées, les PLP des coronavirus pourraient potentiellement jouer un rôle au niveau de la déubiquitination de certaines protéines (Lindner *et al.*, 2005; Sulea *et al.*, 2005). Les produits nsp1, nsp2 et nsp3 sont colocalisés en périphérie du noyau et sont étroitement associés à la machinerie de réplication coronavirale (Harcourt *et al.*, 2004). Le produit de clivage nsp4 colocalise quant à lui avec les produits nsp10 et nsp15. Bien que le rôle précis de chacun de ces trois produits n'ait pas été déterminé, on sait qu'ils feraient partie du complexe de réplication-transcription du virus (Tan *et al.*, 2005). On sait aussi que nsp15 posséderait une fonction *S-adenosylmethionine-dépendant 2'-O-ribose methyltransferase* (2'-O-MT) (Ziebuhr, 2004), qui jouerait potentiellement un rôle au niveau de la formation de la coiffe, comme c'est le cas pour les virus de la rougeole et de la rage (Ferron *et al.*, 2002).

Le produit de clivage nsp5 joue un rôle majeur dans la biologie des coronavirus étant donné qu'il contient le domaine de la protéase principale, la 3CLpro. Cette dernière est responsable du clivage protéolytique des 11 derniers sites de la réplicase 1ab chez le HCoV-OC43 et permet ainsi le relargage d'autres motifs très importants dans la réplication des coronavirus (Ziebuhr, 2004). Le rôle important de la 3CLpro dans la biologie des coronavirus en a d'ailleurs fait une cible de prédilection pour la thérapie anti-SRAS (Anand *et al.*, 2003). Le

nom de la protéase dérive de celui de la protéase principale des picornavirus, la 3C, qui partage de nombreuses caractéristiques avec la protéase coronavirale (Ziebuhr *et al.*, 2000). Puisque le résidu catalytique de la 3CLpro des coronavirus serait une cystéine, la protéase est désignée comme étant une cystéine protéase. Les 3CLpro coronavirales contiennent en fait des doublets catalytiques dans leur site actif, contrairement aux protéases des autres virus, qui contiennent habituellement un triplet catalytique (Anand *et al.*, 2003). Le second résidu catalytique serait donc une histidine et son emplacement dans le site actif varierait en fonction de l'espèce de coronavirus rencontrée (Hegyi *et al.*, 2002). La 3CLpro possède une taille qui varie entre 302 et 307 acides aminés, contrairement à son homologue des picornavirus dont la taille est de seulement 182 acides aminés (Ziebuhr *et al.*, 2000). La 3CLpro des coronavirus possède donc une longue portion carboxy-terminale d'environ 110 acides aminés, qui n'est pas retrouvée chez les autres protéases *3C-like* de virus à ARN (Anand *et al.*, 2002). La protéine comprend trois domaines et se présente sous forme de dimère, dans lequel un seul des deux monomères est actif et l'autre inactif (Yang *et al.*, 2003). Les propriétés d'attachement au substrat résideraient dans le deuxième domaine de la protéine, alors que le troisième domaine serait impliqué au niveau de l'activité protéolytique (Anand *et al.*, 2003). La 3CLpro des coronavirus serait d'abord clivée de façon autocatalytique, puis la protéine dirigerait ensuite le clivage des domaines situés en aval de son emplacement (Ziebuhr *et al.*, 2000). Les spécificités de substrat et de clivage seraient très bien conservées parmi les 3CLpro de tous les coronavirus. Cette conservation serait aussi vraie pour des inhibiteurs de l'enzyme, lesquels pourraient ainsi posséder un large spectre d'action et agir sur les 3CLpro de plusieurs coronavirus à la fois (Hegyi et Ziebuhr, 2002). La spécificité de substrat de la 3CLpro des coronavirus serait déterminée principalement par les positions P1 et P2, lesquelles seraient préférentiellement et respectivement occupées par les acides aminés glutamine et leucine (Anand *et al.*, 2003).

Les produits de clivage nsp6 à nsp10 sont très peu caractérisés et rempliraient pour la plupart des fonctions encore indéterminées. Tel que mentionné précédemment, nsp10 pourrait lier d'autres produits de clivage pour former des complexes de réplication-transcription, tout comme nsp8 et nsp9, lesquels se lieraient mutuellement (Tan *et al.*, 2005). Le produit de clivage nsp9 du SARS-CoV posséderait également des propriétés exclusives aux coronavirus et aurait la capacité de lier des brins d'ARN simple brin (Egloff *et al.*, 2004). Le produit de clivage nsp11 renferme quant à lui des motifs caractéristiques des RdRp virales. Ce produit ne serait pas toujours synthétisé en entier étant donné qu'il renferme le *ribosomal frameshift*. La région immédiatement en aval du *ribosomal frameshift* contient des motifs qui sont conservés pour

toutes les RdRp. Ces motifs incluent notamment un site de liaison au Mg^{2+} , un domaine d'attachement à l'ARN, un noyau catalytique et les domaines de doigts, de pouce et de paume, caractéristiques des polymérase (Brockway *et al.*, 2003). La RdRp des coronavirus possède une extrémité amino-terminale de 375 acides aminés, absente chez les autres RdRp de virus à ARN de polarité positive. Elle représente également une cible potentielle pour des thérapies antivirales étant donné le rôle central qu'elle occupe dans la biologie des coronavirus (Xu *et al.*, 2003). Le produit de clivage nsp12 est aussi très important puisqu'il renferme les motifs de l'hélicase virale, qui joue aussi un rôle majeur dans la réplication du virus. Les hélicases seraient impliquées à plusieurs niveaux lors du métabolisme de l'ARN, c'est-à-dire la transcription, l'épissage, la traduction, l'exportation, la biogenèse des ribosomes, l'expression des gènes mitochondriaux et la régulation de la stabilité des ARNm (Seybert *et al.*, 2000). L'hélicase des coronavirus fait partie de la superfamille 1 des hélicases. Elle est liée à un motif en doigt de zinc à son extrémité amino-terminale et est située en aval de la RdRp, contrairement aux autres virus à ARN de polarité positive, dans lesquels elle est située en amont de la RdRp (Seybert *et al.*, 2000). Elle pourrait dérouler autant les brins d'ARN que les brins d'ADN en direction 5' vers 3' et hydrolyserait l'ATP, le dATP et le GTP plus facilement que les autres nucléotides. Aussi, nsp12 posséderait une activité ARN-5'-triphosphatase, qui serait impliquée au niveau de la formation de la coiffe des ARN viraux (Ivanov *et al.*, 2004b). Finalement, peu de données existent sur les produits de clivage nsp13 et nsp14. Nsp13 consisterait en une exoribonucléase avec une activité 3' vers 5' (ExoN) et serait potentiellement impliquée dans la recombinaison et les mécanismes de réparation de l'ARN (Ziebuhr, 2004). Pour sa part, nsp14 aurait été identifiée comme étant la *replicative nidoviral uridylate-specific endoribonuclease* (NendoU). Bien que le rôle de cette protéine n'ait pas été déterminé précisément, on sait qu'elle serait exclusive aux nidovirus. Il a aussi été démontré que nsp14 pouvait cliver l'ARN simple brin et double brin, malgré une spécificité plus grande pour l'ARN double brin, et que sa fonction est cruciale à la réplication et à la transcription du génome des coronavirus (Ivanov *et al.*, 2004a). Tous les produits de clivage énumérés précédemment participent directement ou indirectement à la transcription des ARN viraux et jouent ainsi un rôle majeur dans le cycle réplcatif coronaviral.

1.4.2 Le cycle réplcatif

Le cycle réplcatif des coronavirus peut être abordé du point de vue génomique et structural. Cependant, les deux volets de ce cycle s'enchaînent chronologiquement et leur synchronisme permet aux coronavirus de générer une progéniture virale qui mènera à

l'établissement d'une infection productive. Suite à l'attachement du virus au récepteur cellulaire et à l'entrée du virus dans la cellule cible, la décapsidation de la particule virale survient et le cycle réplcatif est initié (Lai et Cavanagh, 1997). Bien que les mécanismes de décapsidation des coronavirus ne soient pas bien caractérisés, on sait que le protéasome pourrait intervenir dans ce phénomène et potentiellement faciliter le transfert des virions à l'intérieur du cytoplasme de la cellule infectée, du moins chez le MHV (Yu et Lai, 2005). Le cycle de réplication complet du coronavirus se déroule dans le cytoplasme et les fonctions du noyau ne seraient pas requises (Lai et Cavanagh, 1997). Suite à la décapsidation des virions, l'ORF 1ab du virus est d'abord traduit par le complexe ribosomal cellulaire. L'ARN génomique viral est ensuite répliqué et transcrit par le complexe de réplication-transcription viral nouvellement synthétisé, puis les ARNm sont traduits en utilisant le complexe de synthèse protéique de la cellule (Chen *et al.*, 2002).

Tel que mentionné précédemment, la structure des ARN coronaviraux est particulière et représente l'une des caractéristiques principales de ces virus. L'ARN génomique, pleine longueur, comprend tous les gènes du virus et représenterait environ 4% des ARN viraux lors du cycle de réplication (Thiel *et al.*, 1997). La transcription de ce dernier mène à la génération d'un ensemble d'ARNm sous-génomiques dont l'arrangement est co-terminal à l'extrémité 3'. La séquence de chaque ARNm est donc nichée ou contenue entièrement à l'intérieur du prochain ARNm, qui est plus long (Figure 2) (Méndez *et al.*, 1996). Chaque ARNm correspond généralement à un gène viral, mais il arrive que certains ARNm soient bicistroniques, comme c'est le cas chez le HCoV-OC43 et le SARS-CoV, lesquels contiennent respectivement un (St-Jean *et al.*, 2004) et cinq des ces ARNm (Thiel *et al.*, 2003a). Les coronavirus utilisent un mécanisme dans lequel seuls les ORF situés à l'extrémité 5' de chaque ARNm sont traduits (Méndez *et al.*, 1996). Lorsque les ARNm sont bicistroniques, le virus doit avoir recours à un système qui permet la traduction des ORF situés en aval du premier. Bien que ce processus n'ait pas encore été caractérisé chez le coronavirus, il est raisonnable de croire qu'un mécanisme de réinitiation de la traduction pourrait être utilisé, étant donné qu'un tel mécanisme a déjà été identifié chez le calicivirus (Meyers, 2003). La raison pour laquelle certains gènes se chevauchent et produisent des ARNm polycistroniques n'est pas connue. Il a cependant déjà été proposé qu'un tel système de chevauchement des gènes pourrait s'avérer un mécanisme à travers lequel les nidovirus préserveraient l'intégrité génétique des éléments vitaux de leurs génomes (de Vries *et al.*, 2000).

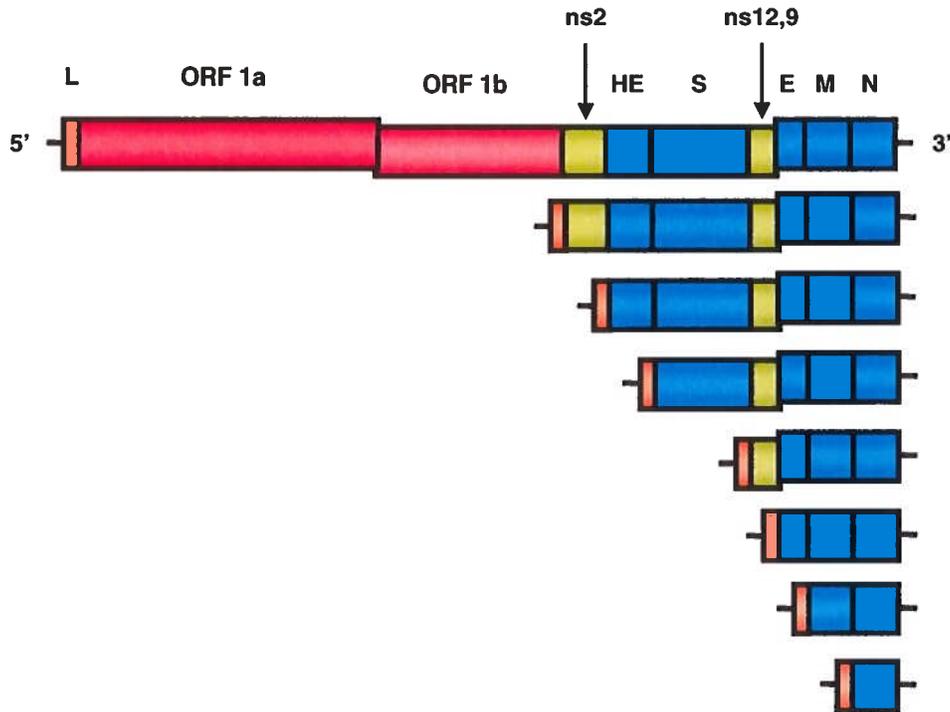


Figure 2. Représentation schématique des ARN du HCoV-OC43, dont l'arrangement est niché et co-terminal à l'extrémité 3' de chacun. En rouge, gène de la réplicase; en vert, gènes de protéines non structurales; en bleu, gènes de protéines structurales. L, leader; ns, non-structurale; HE, hémagglutinine-estérase; S, spicule; E, enveloppe; M, membrane; N, nucléocapside.

On a longtemps cru que le système de transcription utilisé par les coronavirus était le *leader-primed transcription* (Kamahora *et al.*, 1989). Ce mécanisme impliquait d'abord une transcription de l'ARN génomique viral en un ARN génomique pleine longueur de polarité négative. L'ARN ainsi généré servirait ensuite de matrice pour la synthèse des ARNm sous-génomiques de polarité positive (Kamahora *et al.*, 1989). Ce modèle a principalement été avancé parce qu'aucun ARNm de polarité négative ou intermédiaire de réplication de polarité négative n'avait alors été isolé (Sawicki et Sawicki, 1998). Le modèle de transcription actuel propose plutôt un mécanisme discontinu, dans lequel les ARNm sous-génomiques de polarité négative seraient générés directement à partir du génome de polarité positive. Ces ARNm contiendraient

les séquences *leaders* typiques des coronavirus et serviraient de matrice pour la production des ARNm sous-génomiques de polarité positive (Sawicki et Sawicki, 1998). L'élément central du mécanisme de transcription des coronavirus réside au niveau de séquences nucléotidiques appelées séquences de régulation de la transcription, ou TRS pour *transcription regulatory sequences* (Alonso *et al.*, 2002). Chez une même espèce de coronavirus, ces séquences sont généralement bien conservées et sont présentes en amont des ORF du virus qui mènent à la synthèse d'un ARNm (Curtis *et al.*, 2004). Les TRS renferment aussi des *core sequences* (CS), qui consistent en de petites séquences de 6 ou 7 bases, très conservées et essentielles à la transcription du génome d'ARN viral. Ces séquences étaient autrefois nommées séquences intergéniques, mais comme plusieurs gènes se chevauchent dans le génome et que les séquences intergéniques n'existent pas toujours, l'appellation CS a été employée (Alonso *et al.*, 2002).

Les TRS sont également présentes dans la séquence *leader* des coronavirus et cette caractéristique est primordiale pour la transcription de ces virus. Les mécanismes de transcription des coronavirus ont dernièrement été étudiés de façon plus approfondie par l'équipe du Dr Luis Enjuanes, par l'entremise d'un clone d'ADN complémentaire (ADNc) du TGEV (Zúñiga *et al.*, 2004). Suite à la formation du complexe de transcription virale, lequel permettrait de relier les extrémités 5' et 3' du génome, le TRS de la séquence *leader* (TRS-L) se retrouverait à proximité des séquences de l'extrémité 3' du génome (Figure 3A). À l'intérieur du même complexe, la synthèse d'un brin négatif d'ARN s'amorcerait pendant que le TRS-L du brin positif balayerait le brin nouvellement synthétisé, à la recherche d'un TRS lui étant complémentaire (Figure 3B) (Sola *et al.*, 2005). Chez certains coronavirus comme le BCoV, les TRS du génome (TRS-G) seraient exposées à la surface de structures secondaires d'ARN, les rendant ainsi plus accessibles (Ozdarendeli *et al.*, 2001). Selon le modèle de transcription discontinue, les TRS agiraient comme des signaux d'arrêt de la transcription (Alonso *et al.*, 2002). Suite à cet arrêt, le complexe de réplication-transcription pourrait donc poursuivre la réplication de la matrice d'ARN génomique jusqu'au prochain TRS-G ou interrompre celle-ci. Cette boucle d'évènements pourrait ainsi se perpétuer jusqu'à l'extrémité 5' du génome viral, mais les signaux qui permettraient de contrôler ce phénomène ne sont pas clairement caractérisés. Dans l'éventualité d'un arrêt de la transcription lors de la rencontre d'un TRS-G, un changement de matrice s'effectuerait et le brin négatif se répliquerait alors à partir du brin d'ARN contenant le TRS-L, donc le *leader*, ce qui mènerait à la terminaison de la transcription (Figure 3C). L'ARNm nouvellement synthétisé serait ensuite répliqué en un ARNm de polarité positive de même taille. Des ARNm sous-génomiques de différentes longueurs et des ARN génomiques pleine longueur

pourraient ainsi être synthétisés en utilisant ce mécanisme de transcription discontinue (Figure 3C) (Zúñiga *et al.*, 2004).

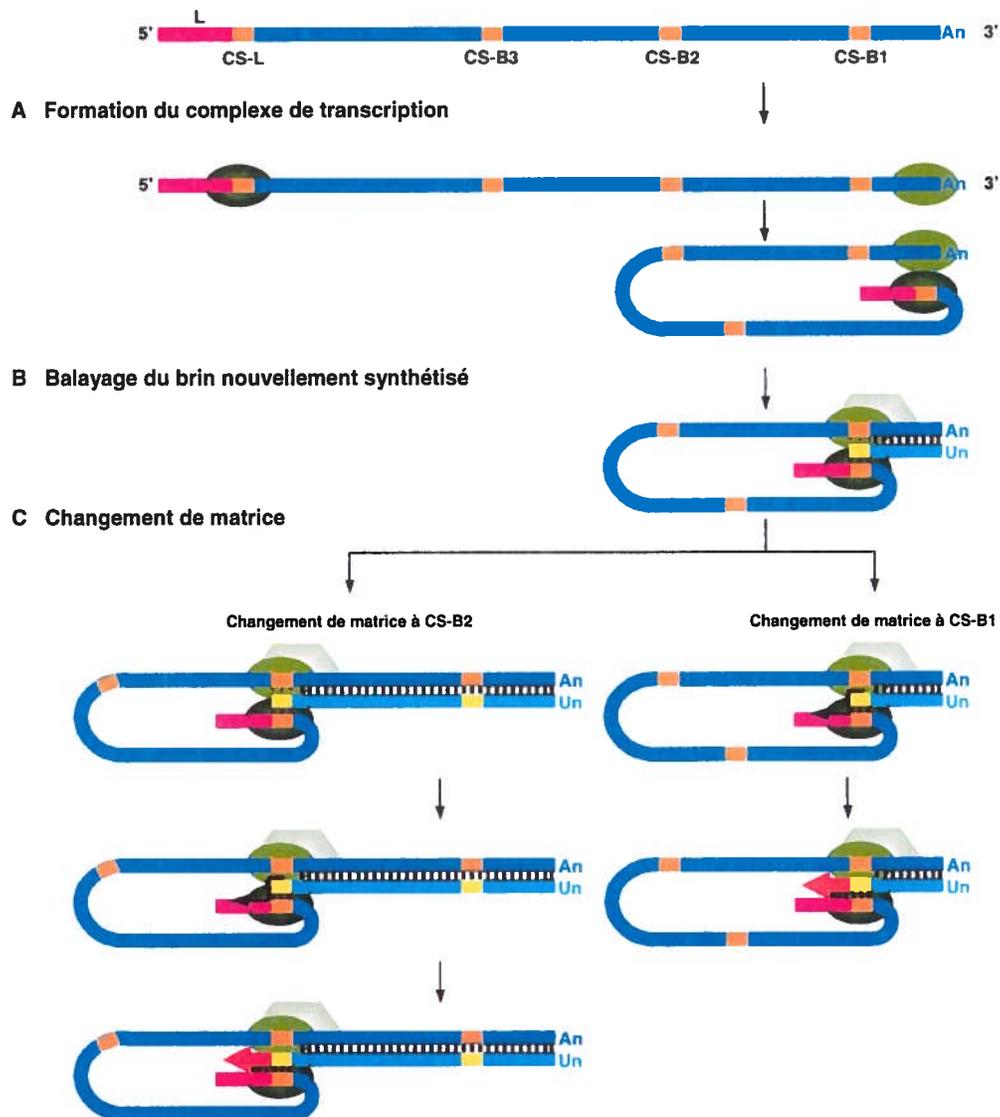


Figure 3. Représentation schématique du mécanisme de transcription des coronavirus. L, séquence *leader*; CS-L, *core sequence - leader*; CS-B1, -B2 et -B3, *core sequences* du génome viral; An, queue de poly(A); Un, queue de poly(U) (adaptée de Zúñiga *et al.*, 2004).

La complémentarité entre les TRS-L et TRS-G serait cruciale pour le contrôle des niveaux d'ARNm sous-génomiques. Le niveau de complémentarité entre les TRS-L et -G corrèlerait donc avec le taux de synthèse des ARNm viraux (Sola *et al.*, 2005). Bien que les CS contenues dans les TRS ne soient pas essentielles à la génération des ARNm sous-génomiques, il a été démontré que leur présence pouvait augmenter les niveaux de transcription de 1000 fois (Zúñiga *et al.*, 2004). Les séquences qui joueraient un rôle dans la complémentarité des TRS au moment de la transcription incluraient donc la CS ainsi que les quatre nucléotides situés de part et d'autre de cette séquence. La complémentarité de ces séquences et la composition des séquences situées en aval du CS auraient d'ailleurs une forte influence sur le changement de matrice qui survient à la fin de la transcription des ARNm de polarité négative (Sola *et al.*, 2005). L'efficacité de la transcription pourrait aussi dépendre d'un domaine de régulation situé à environ 50 nucléotides en amont de la CS, mais le mécanisme à travers lequel ce domaine interviendrait dans la transcription demeure mal compris (Curtis *et al.*, 2004).

En raison du taux d'erreur de la RdRp virale, qui peut atteindre jusqu'à 10^{-3} chez certains virus à ARN (Steinhauer *et al.*, 1992), mais qui se situe habituellement aux environs de 10^{-4} pour les coronavirus (Sanchez *et al.*, 1992), et du mécanisme de transcription discontinue utilisé par ces derniers, des mutations sont souvent introduites dans le génome viral. La recombinaison est un phénomène qui est souvent rencontré chez les coronavirus et qui résulte des caractéristiques de transcription de ces derniers (Lai, 1996). Durant l'arrêt de transcription effectué lors de la synthèse des ARN viraux, la RdRp virale, contenue dans le complexe de répllication-transcription, pourrait effectuer une erreur lors du changement de matrice et poursuivre son chemin sur une matrice inappropriée, entraînant ainsi des modifications majeures dans le génome du virus (Holmes, 2003). Lors d'un phénomène semblable, la RdRp pourrait également effectuer un "saut" lors de la transcription, sur une nouvelle matrice ou en aval de la même matrice, en transportant avec elle le brin d'ARN nouvellement synthétisé, ce qui pourrait mener à la génération de particules virales possédant des génomes défectifs. Les virus dont le génome est défectif sont en fait des mutants de délétion qui possèdent les signaux en *cis* requis pour la répllication et l'encapsidation, mais qui sont dépendants d'autres fonctions virales fournies en *trans* par un virus *helper*. Les extrémités du génome sont nécessaires puisqu'elles renferment les signaux responsables de la répllication du virus (Izeta *et al.*, 1999). La structure standard d'un ARN défectif comprend donc les extrémités 5' et 3' de la souche sauvage et des délétions internes qui justifient la taille plus petite de l'ARN. Les ARN défectifs peuvent aussi interférer avec la

réplication des virus de souche sauvage, d'où l'appellation *defective interfering RNA* (DI-RNA), qui leur est parfois attribuée (Méndez *et al.*, 1996).

La recombinaison est un mécanisme important qui contribue autant à la stabilité génétique qu'à la diversité des coronavirus dans la nature. Elle a souvent été démontrée expérimentalement, notamment avec des souches du MHV (Kamahora *et al.*, 1989) et du BCoV (Wu *et al.*, 2003). Bien que la fréquence de recombinaison des coronavirus n'ait pas été déterminée, la facilité avec laquelle les recombinants ont été isolés laisse croire que cette dernière serait relativement élevée. La fréquence est même parfois tellement élevée que les recombinants sont sélectionnés sans pression spécifique. Ce phénomène est aussi parfois observé dans la nature, mais les recombinants ainsi générés doivent cependant posséder des avantages évolutifs et/ou adaptatifs pour survivre (Lai, 1996). Le cas du gène HE des coronavirus, lequel aurait potentiellement été acquis suite à une recombinaison avec le virus influenza C, pourrait représenter un tel exemple (Holmes, 2003).

Qu'ils soient défectifs ou non, les génomes coronaviraux doivent posséder des signaux distincts afin d'être encapsidés à l'intérieur des virions (Izeta *et al.*, 1999). L'ARN génomique de ces virus contiendrait donc des signaux d'encapsidation localisés à des endroits différents selon l'espèce de coronavirus rencontrée. Ces signaux seraient habituellement situés à l'extrémité 5' du génome viral, d'où l'encapsidation possible des ARN génomiques défectifs. Puisque les ARNm sous-génomiques des coronavirus sont dépourvus de l'extrémité 5' du génome, ils ne seraient pas encapsidés non plus (Escors *et al.*, 2003). Ce signal d'encapsidation serait localisé dans les 649 premiers nucléotides du génome chez le TGEV, alors que le MHV et le BCoV pourraient potentiellement en contenir deux, l'un étant situé dans les 498 premiers nucléotides du génome, et l'autre dans l'ORF 1b, aux environs du vingtième kilobase (Escors *et al.*, 2003).

Avant que le génome viral soit encapsidé, la structure protéique du virus doit d'abord être assemblée. Les ARNm viraux qui encodent les protéines structurales HE, S, E, M et N sont donc transportés au réticulum endoplasmique (RE) et sont traduits par le complexe ribosomal cellulaire à la surface du RE (Lontok *et al.*, 2004). Les protéines structurales seraient toujours exprimées individuellement et subiraient des modifications post-traductionnelles propres à chacune dans des compartiments cellulaires donnés (Nal *et al.*, 2005). Les protéines HE, S, E et M posséderaient également des signaux de rétention au RE afin de subir leur maturation et de débiter leur assemblage (Lim et Liu, 2001). Cette propriété des protéines structurales a surtout été

caractérisée pour la protéine S, laquelle contiendrait un signal canonique de rétention au RE dans sa queue cytoplasmique, spécifique à chaque groupe génétique de coronavirus. La queue cytoplasmique de la protéine S des coronavirus de groupe 2 ne posséderait pas un tel signal cependant, suggérant une distribution différente de S et une pathogenèse différente due à une accumulation accrue de S à la surface cellulaire (Lontok *et al.*, 2004). Tel que mentionné précédemment, la protéine M est un déterminant majeur dans l'assemblage des coronavirus. Suite à la synthèse et à l'oligomérisation des protéines structurales dans le RE, une interaction de la protéine M surviendrait avec les protéines HE (Nguyen et Hogue, 1997), S (de Haan *et al.*, 1999) et E (Lim et Liu, 2001). Chez la protéine S, le site d'interaction avec la protéine M serait localisé à l'extrémité carboxy-terminale de la protéine, dans le domaine cytoplasmique plus précisément (Bosch *et al.*, 2005). Bien qu'elle soit essentielle à l'infectivité intracellulaire du virus, la protéine S ne serait pas impliquée au niveau de l'assemblage du virus (Kuo et Masters, 2002). La fonction principale de la protéine M durant l'assemblage serait donc de retenir les autres protéines structurales au site de bourgeonnement du virus et d'induire l'incorporation de ces protéines dans l'enveloppe virale (de Haan *et al.*, 1999). Une représentation schématique de l'assemblage du TGEV est d'ailleurs présentée à la Figure 4.

Contrairement aux protéines structurales mentionnées précédemment, la protéine N ne contiendrait pas de signal de rétention au RE. Celle-ci serait plutôt localisée aux sites de synthèse des ARN viraux, lesquels sont associés à des membranes cellulaires qui consisteraient en des vésicules à doubles membranes (Gosert *et al.*, 2002). La protéine N colocaliserait donc avec la RdRp virale, l'hélicase et la 3CLpro, à des sites distincts des sites d'assemblage, confirmant l'implication de N au niveau de la réplication et de la transcription (Brockway *et al.*, 2003). Bien que le cycle de réplication des coronavirus ne dépende pas des fonctions du noyau, la protéine N pourrait quand même interagir avec des composantes du nucléole, comme la nucléoline et la fibrillarine. Les coronavirus pourraient ainsi retarder le cycle cellulaire en interphase alors que la traduction des ARNm viraux est à son maximum (Chen *et al.*, 2002). La protéine N est aussi la seule protéine structurale phosphorylée des coronavirus, tel que démontré chez le TGEV et le IBV. La protéine N phosphorylée présenterait une plus grande affinité pour l'ARN viral que la protéine N non-phosphorylée (Calvo *et al.*, 2005) et comme la protéine M coronavirale n'interagirait qu'avec les protéines N liées à l'ARN génomique viral, un phénomène d'incorporation sélective des ribonucléocapsides dans les virions bourgeonnant serait observé (Kuo et Masters, 2002).

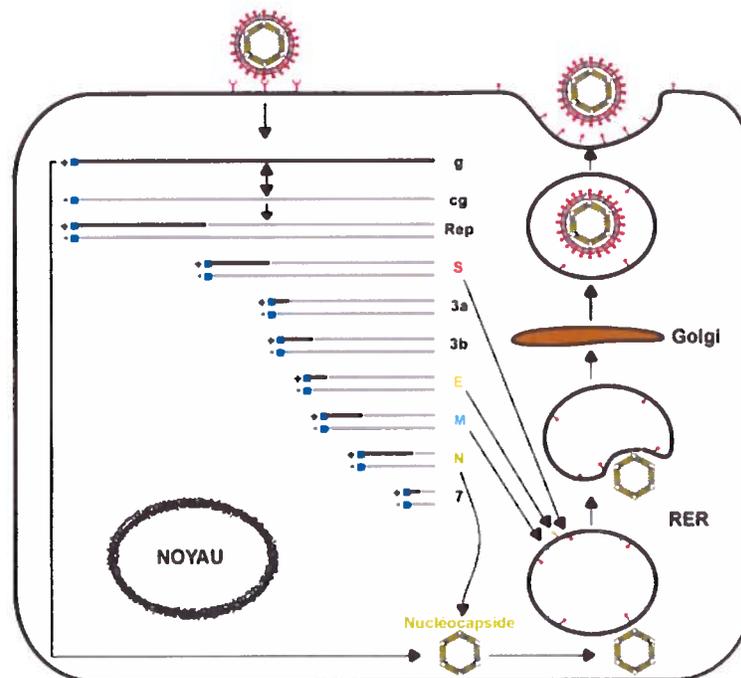


Figure 4. Représentation schématique des mécanismes d'assemblage du TGEV dans la cellule infectée. G, génomique; cg, copie génomique; Rep, réplication; S, spicule; E, enveloppe; M, membrane; N, nucléocapside; RER, réticulum endoplasmique rugueux; (source : <http://virologie.free.fr/34-Coronaviridae/coronaviridae.htm>).

Contrairement aux autres virus à ARN enveloppés, les coronavirus emploient une stratégie indépendante de la nucléocapside afin de s'assembler et de bourgeonner (Lim et Liu, 2001). Suite à l'oligomérisation des protéines structurales, celles-ci se rendraient à la membrane du compartiment intermédiaire (CI), un compartiment cellulaire situé entre le RE et l'appareil de Golgi, afin d'y être encapsidées (de Haan *et al.*, 1999). Les coronavirus termineraient donc leur assemblage dans le CI et obtiendraient leur enveloppe des membranes de ce même compartiment. La protéine M retiendrait les protéines HE, S et E dans le CI afin qu'elles ne soient pas exportées vers la surface cellulaire (Nguyen et Hogue, 1997). La protéine S du IBV contiendrait aussi un signal d'endocytose dans sa queue cytoplasmique, suggérant que les protéines S qui échappent au CI sont rapidement endocytosées lorsqu'elles atteignent la membrane plasmique. Advenant une saturation du CI, les protéines S supplémentaires pourraient cependant se rendre à la membrane

plasmique et causer des fusions cellule-cellule (Lontok *et al.*, 2004). Ce type de fusion cellulaire est indépendant des fusions membranaires, dépendantes de la présence d'un récepteur cellulaire spécifique, et fait aussi intervenir un mécanisme indépendant du pH de l'environnement cellulaire (Simmons *et al.*, 2004). Les fusions cellule-cellule sont également responsables de la formation des syncytiums, qui consistent en des cellules géantes comprenant plusieurs noyaux (Krueger *et al.*, 2001). Finalement, lorsque les génomes et les protéines structurales sont encapsidés dans les virions, la progéniture virale est libérée du CI et bourgeoise (Hofmann et Pöhlmann, 2004). Les virions sont alors transportés dans des vésicules à la membrane plasmique et sont relargués de la cellule infectée par exocytose (Blau et Holmes, 2001).

2.0 L'INFECTION VIRALE CHEZ L'HÔTE

Les propriétés structurales des coronavirus, ainsi que les mécanismes qui découlent de celles-ci et qui mènent à l'infection virale, ont été décrits dans la première partie de cet ouvrage. Cependant, suite à l'établissement de l'infection et en réponse à la présence du virus dans l'organisme, une cascade d'évènements peut survenir chez l'hôte. Les conséquences immédiates et à court terme de l'infection coronavirale seront donc abordées dans cette section du manuscrit.

2.1 Généralités

Pour la majorité des coronavirus, notamment les HCoV et le SARS-CoV, la transmission à l'hôte s'effectue par des aérosols. Les virus sont donc présents dans l'air, en suspension, et atteignent les voies respiratoires de l'hôte lorsqu'ils sont inhalés par ce dernier (Gagneur *et al.*, 2002). Bien que le site d'infection primaire de ces coronavirus soit les voies respiratoires, le site d'infection principal est déterminé par le tropisme du virus (Hofmann *et al.*, 2005). Le récepteur cellulaire utilisé par le virus représente l'un des déterminants majeurs de ce tropisme. Selon le récepteur utilisé, le virus peut donc infecter un type cellulaire spécifique et par conséquent, un tissu ou un organe précis. Cependant, le récepteur utilisé ne détermine pas entièrement le site d'infection et de réplication virale puisque deux coronavirus d'une même espèce, qui utilisent le même récepteur cellulaire, peuvent causer des infections à des sites distincts chez un même hôte. C'est notamment le cas du MHV, dont la souche A59 possède un tropisme plus établi pour le foie (Navas *et al.*, 2001), alors que la souche JHM infecte préférentiellement le SNC (Taguchi *et al.*, 1995). Dans le même ordre d'idées, différentes souches du TGEV peuvent infecter les voies

respiratoires du porc ou son système digestif (Almazán *et al.*, 2000). Chez les HCoV-OC43 et -229E, une telle dualité au niveau du tropisme aurait aussi été observée et chez une même souche par surcroît. Le HCoV-OC43 aurait donc la capacité d'infecter les cellules des voies respiratoires supérieures (Myint, 1994) ainsi que certains types cellulaires du SNC (Arbour *et al.*, 1999a). Il a également été démontré que la souche OC43 pouvait atteindre le SNC (Arbour *et al.*, 2000). La souche OC43 est donc respectivement neurotrophe et neuroinvasive (van Marle et Power, 2005), et elle représente également un candidat intéressant pour l'étude d'infections du SNC induites par des agents viraux. Contrairement à la souche JHM du MHV, laquelle peut induire le déclenchement de pathologies démyélinisantes du SNC chez la souris, un lien direct entre le HCoV-OC43 et le déclenchement de telles maladies chez l'humain n'a jamais été clairement établi. La souche OC43 serait également neurovirulente puisqu'elle possède la capacité de causer des pathologies neurologiques (van Marle et Power, 2005), dont une encéphalite aiguë chez la souris (Jacomy et Talbot, 2003).

2.2 La persistance virale et l'adaptation

Lorsque le coronavirus s'attache à un récepteur cellulaire spécifique, qu'il pénètre à l'intérieur de la cellule et que son cycle de réplication débute, l'infection est amorcée. Cependant, cette dernière ne sera productive que si la progéniture virale est encapsidée et bourgeonne à l'extérieur de la cellule. Dans le cas contraire, l'infection sera abortive (Pejawar *et al.*, 2005). Suite au relargage des nouveaux virions, une cascade d'évènements impliquant l'organisme de l'hôte ainsi que son système immunitaire survient. Cette période de l'infection, qui précède habituellement les premiers symptômes chez l'hôte, représente le début de la phase aiguë. À l'occasion, il arrive que l'infection virale ne soit pas neutralisée et complètement éliminée par le système immunitaire et c'est dans ces circonstances que l'infection aiguë devient une infection persistante. Plusieurs facteurs peuvent être à l'origine de cette persistance et plusieurs mécanismes viraux peuvent aussi induire le déclenchement d'une infection persistante. Les rétrovirus et les virus de la famille de l'herpès adoptent notamment des stratégies uniques qui leur permettent de déclencher ce type d'infections (Cullen, 2001; Williams et Crawford, 2005). Par les mécanismes qu'ils utilisent, ces virus peuvent préserver une certaine intégrité de leur génome et principalement échapper au système immunitaire (Oldstone, 1989). Bien que les coronavirus n'utilisent pas de mécanismes spécialisés dans l'établissement de l'infection persistante, il a déjà été démontré que ce type d'infection pouvait survenir chez différentes espèces, autant *in vivo* (Adami *et al.*, 1995; Addie *et al.*, 2003; Navas-Martin *et al.*, 2005) que *in*

vitro (Arbour *et al.*, 1999a; Schickli *et al.*, 2004). Les mécanismes qui mènent à de telles infections chez le coronavirus n'ont toujours pas été caractérisés, mais on sait par contre qu'aucune adaptation du virus ne serait nécessaire à l'établissement de ce phénomène, du moins pour le HCoV-OC43 (Arbour *et al.*, 1999a).

Chez les virus à ARN, les infections persistantes sont souvent associées avec l'adaptation virale (Schickli *et al.*, 1997) ainsi que la génération de quasiespèces (Adami *et al.*, 1995). Le taux d'erreur élevé des RdRp virales est principalement à l'origine de ce phénomène. Bien que la majorité des mutations introduites dans les génomes viraux soient délétères (Novella, 2003), les mutations qui ne sont pas létales pour le virus confèrent souvent un avantage évolutif ou adaptatif à ce dernier. De nouvelles populations virales, ou quasiespèces, sont ainsi générées et forment une progéniture possédant de nouvelles propriétés adaptatives, par rapport à leur contexte environnemental. La recombinaison d'ARN lors d'infections persistantes a aussi été proposée comme étant un mécanisme important dans la génération de variants chez les coronavirus (Rowe *et al.*, 1997). Les mutations acquises durant l'infection persistante permettent ainsi aux virus d'interagir différemment avec les composantes de leur environnement, ce qui leur permet notamment d'acquérir certaines propriétés reliées à la virulence (Ontiveros *et al.*, 2003) ou même d'élargir leur tropisme cellulaire (Schickli *et al.*, 1997). La publication no.5 de cette thèse traite d'ailleurs de la capacité d'adaptation du HCoV-OC43 suite à des infections persistantes en culture cellulaire. Suite à ces infections, de nombreuses mutations ont été identifiées dans le génome du virus et de nouvelles propriétés virales ont également été associées à quelques-unes de ces mutations.

Chez les coronavirus, la faculté d'adaptation serait en grande partie liée au gène de protéine structurale S (Fang *et al.*, 2005). Cette association est logique puisque la protéine S régit de nombreuses fonctions reliées à la pathogenèse, à la virulence et à l'entrée virale. Il a d'ailleurs déjà été démontré que des mutations dans le gène S, acquises suite à une infection persistante, pouvaient affecter les fonctions de fusion membranaire (Krueger *et al.*, 2001) et d'attachement au récepteur (Künkel et Herrler, 1996). De toutes les propriétés coronavirales attribuables au gène S, c'est sans aucun doute l'adaptation reliée au tropisme cellulaire qui a le plus été étudiée. Par exemple, suite à des infections persistantes chez la souris avec le MHV-A59, il a été démontré que des substitutions d'acides aminés spécifiques de la protéine S pouvaient déterminer l'habileté du virus à atteindre le foie et à causer des hépatites létales (Navas-Martin *et al.*, 2005). De plus, des études très récentes ont montré qu'une adaptation reliée au tropisme pouvait survenir durant

l'infection et que les virus impliqués pouvaient alors développer une spécificité pour un récepteur secondaire. C'est notamment le cas du MHV qui, suite à quelques passages en culture cellulaire, pouvait employer le sulfate d'héparan comme récepteur principal (de Haan *et al.*, 2005), et le cas du FIPV, lequel peut modifier son tropisme pour les macrophages durant l'infection (Rottier *et al.*, 2005). La mutagenèse dirigée a aussi été utilisée afin de démontrer que le tropisme cellulaire d'un virus pouvait être modifié en effectuant une substitution précise de quelques acides aminés (Thackray et Holmes, 2004). Par des substitutions de l'ectodomaine de la protéine S, il a même été démontré que le tropisme des coronavirus pouvait être drastiquement redirigé vers un type cellulaire précis, selon la protéine S employée (Almazán *et al.*, 2000), et que cette modification pouvait permettre au virus de franchir la barrière d'espèce, alors que les coronavirus présentent habituellement une très forte spécificité d'hôte (Haijema *et al.*, 2003; Kuo *et al.*, 2000). L'exemple le plus spectaculaire d'une telle adaptation dans la nature est sans contredit celui du SARS-CoV, qui a franchi la barrière d'espèce et infecté l'humain (Song *et al.*, 2005). Suite à l'analyse des génomes des souches de SARS-CoV qui ont infecté la civette et l'humain, il a clairement été démontré que la protéine S du virus était responsable de l'adaptation observée (Peiris *et al.*, 2004). Des analyses plus poussées des protéines S de souches isolées chez la civette et durant l'éruption de 2002-2003 et l'éruption mineure de 2003-2004 ont aussi été effectuées. Ces études ont révélé que les trois protéines S analysées liaient et utilisaient efficacement le récepteur ACE2 de civette, mais que les protéines S issues des coronavirus de civette et de l'éruption de 2003-2004 utilisaient moins efficacement l'ACE2 humain, comparativement à la souche isolée durant l'éruption de 2002-2003 (Li *et al.*, 2005b). Seulement quatre acides aminés différaient entre la souche adaptée à l'humain de 2002-2003 et la souche animale du virus. À eux seuls, ces quatre changements causeraient une augmentation de l'affinité de la protéine S pour le récepteur ACE2 humain de plus de 1000 fois (Holmes, 2005).

2.3 Les infections du système nerveux central

Chez le HCoV-OC43, deux hypothèses seraient plausibles en ce qui concerne le mode de propagation du virus vers le SNC. Dans le premier modèle, le virus atteindrait d'abord les voies respiratoires supérieures puis infecterait le SNC dans un second temps, alors que dans le deuxième modèle, le virus atteindrait directement le SNC par transport transneuronal via le nerf olfactif, qui servirait alors de site primaire d'infection. Connaissant la capacité intrinsèque du HCoV-OC43 à infecter et à se répliquer dans les cellules du système nerveux (Arbour *et al.*, 1999a; Bonavia *et al.*, 1997) et sachant que certains coronavirus peuvent utiliser le nerf olfactif

pour remonter jusqu'au SNC (Barnett et Perlman, 1993), la deuxième hypothèse de neuroinvasion semble valable. Bien qu'un modèle n'exclut pas nécessairement l'autre, certains faits font aussi pencher la balance en faveur de la première hypothèse. En effet, la capacité de certains coronavirus à causer une virémie (Ng *et al.*, 2004) combinée à la capacité du HCoV-OC43 à infecter les monocytes et les macrophages du sang périphérique (Collins, 2002) laisse croire que la première hypothèse pourrait également être plausible. Les particules virales présentes à l'intérieur des cellules du système immunitaire pourraient alors utiliser ces dernières comme véhicule afin de franchir la barrière hémato-encéphalique, laquelle sépare les éléments du sang et du SNC. D'ailleurs, une association entre le SARS-CoV et des symptômes neurologiques a été avancée chez des patients ayant préalablement été atteints du SRAS, ce qui pourrait renforcer cette hypothèse (Xu *et al.*, 2005). Dans ce modèle, on ne sait cependant pas si une adaptation du virus surviendrait lorsque celui-ci passerait des voies respiratoires au SNC.

Bien que les déterminants de neurovirulence n'aient pas été déterminés de façon précise chez les coronavirus, ces derniers seraient contenus dans les sous-unités S1 (Taguchi *et al.*, 1995) et S2 de la protéine S (Fu *et al.*, 2004; Phillips *et al.*, 2001). Le tropisme du virus, sa capacité à induire des fusions cellule-cellule ainsi que les effets cytopathiques engendrés par celui-ci pourraient donc être à l'origine des propriétés neurovirulentes des coronavirus. La RHV des coronavirus aurait également un rôle à jouer dans ces propriétés du virus étant donné son rôle potentiel dans l'évasion immunitaire (Rowe *et al.*, 1997). Les cellules neurales touchées lors des infections du SNC par le HCoV-OC43 n'ont pas été clairement identifiées. Bien que *in vitro* la souche OC43 soit capable d'infecter des cellules microgliales, astrocytaires, oligodendrocytaires et neuronales (Arbour *et al.*, 1999a; Bonavia *et al.*, 1997), ce phénomène ne serait peut-être pas respecté *in vivo*. Jusqu'à présent, l'infection *in vivo* du HCoV-OC43 n'a été caractérisée que dans un modèle animal, dans lequel le virus est inoculé directement dans le cerveau de souris C57Bl/6 ou BALB/c (Jacomy et Talbot, 2001). Contrairement à ce qui est observé lors des infections par le MHV (Lane et Buchmeier, 1997), aucune démyélinisation n'est observée dans le SNC de la souris suite à des infections par le HCoV-OC43. Alors que la pathologie induite par le MHV ressemble à la sclérose en plaques, laquelle est principalement caractérisée par des foyers de destruction de la myéline dans la matière blanche du SNC (Hunter et Rodriguez, 1995), la pathologie engendrée par le HCoV-OC43 aurait plutôt des allures d'encéphalite aiguë. L'infection cible principalement les neurones et est caractérisée par une vacuolisation et une dégénérescence de ceux-ci. L'inoculation du virus entraîne donc une pathologie plus généralisée

du SNC, lors de laquelle les dommages observés semblent être une conséquence directe de l'infection virale (Jacomy et Talbot, 2003).

3.0 LES CONSÉQUENCES DE L'INFECTION

Suite à l'établissement de l'infection coronavirale, l'organisme de l'hôte réagit et tente de s'ajuster à la présence du microorganisme et aux répercussions engendrées par ce dernier. C'est alors que le système immunitaire intervient et qu'il tente de limiter la dissémination du virus. Selon le tropisme du virus et la longévité de l'infection, des pathologies peuvent aussi survenir chez l'hôte, en réponse directe ou indirecte à l'infection. Les conséquences de l'infection coronavirale, à plus long terme, seront donc abordées dans cette section du manuscrit.

3.1 Les pathologies associées

Tel que mentionné précédemment, les pathologies liées aux infections coronavirales peuvent varier selon le tropisme du virus. Cependant, chez le HCoV-OC43, les pathologies engendrées par le virus sont d'ordre neurologique plutôt que respiratoire chez la souris. En fait, bien que la souche OC43 puisse causer des infections respiratoires à différents degrés, son niveau de persistance et sa présence dans le SNC pourraient entraîner une dégénérescence de ce dernier, laquelle résulterait en une pathologie neurologique d'une forme quelconque. Les propriétés neurotropes et neuroinvasives du HCoV-OC43 seraient donc à l'origine du déclenchement ou de l'amplification de telles pathologies, quoique l'association entre le virus et les maladies neurologiques n'ait pas été établie formellement (Talbot *et al.*, 1996). Le modèle de maladie démyélinisante fourni par le MHV laisse cependant croire qu'une telle association serait possible et que le HCoV-OC43 pourrait déclencher des pathologies neurodégénératives chez l'humain, comme la sclérose en plaques (Matthews *et al.*, 2002). Chez les patients atteints de sclérose en plaques, des infections des voies respiratoires supérieures sont souvent associées avec des rechutes de la maladie, renforçant ainsi le lien avec les coronavirus (Arbour *et al.*, 2000). Cependant, il est à noter que différents virus ont été associés avec le déclenchement de la sclérose en plaques, sans qu'une relation claire n'ait été établie avec la maladie. D'autres maladies neurodégénératives comme le Parkinson, l'Alzheimer et la sclérose amyotrophique latérale pourraient aussi avoir une origine infectieuse et possiblement coronavirale (Mattson, 2004). Les

désordres psychiatriques ont également fait l'objet d'études visant à déterminer si ces troubles étaient une conséquence d'infections virales (Jacomy et Talbot, 2003).

La sclérose en plaques représenterait en quelque sorte l'homologue humain de la maladie démyélinisante induite par le MHV et caractérisée chez la souris. Il s'agit en fait d'une pathologie inflammatoire, neurodégénérative et démyélinisante du SNC. La maladie survient principalement chez les jeunes adultes, à une moyenne d'âge de 30 ans. Les femmes représentent 70% des personnes atteintes et les hormones féminines seraient probablement à la base de cette préférence (Whitacre *et al.*, 1999). Les causes de la sclérose en plaques demeurent inexplicables, mais les données suggèrent une étiologie multifactorielle qui impliquerait à la fois une susceptibilité génétique ainsi que de facteurs environnementaux, comme des virus (Talbot, 1997). Bien que les personnes possédant un génotype HLA-DR15 semblent être plus susceptibles à la maladie (Lang *et al.*, 2002), plusieurs faits supportant l'implication de facteurs environnementaux dans le déclenchement de la maladie existeraient (Holz *et al.*, 2000). La pathologie est essentiellement caractérisée par la présence de foyers d'inflammation et par la destruction de la myéline dans le SNC. Étant donné le rôle de la myéline dans la conduction du potentiel d'action le long des axones, une atteinte de la maladie résulte principalement en des troubles nerveux reliés au système moteur (Hunter et Rodriguez, 1995). Deux modèles existent en ce qui concerne les processus de destruction de la myéline dans le SNC : le *outside-in* ainsi que le *inside-out* (Tsunoda et Fujinami, 2002). Dans le premier modèle, les dommages axonaux seraient une conséquence secondaire de la destruction des oligodendrocytes, alors que dans le deuxième, l'inverse se produirait. Bien qu'il ait longtemps été cru que la pathologie était restreinte aux oligodendrocytes et à la myéline, le deuxième modèle, dans lequel la destruction des axones déclencherait la démyélinisation, est de plus en plus supporté. Il est aujourd'hui évident que la maladie implique un processus induit par le système immunitaire (Coyle, 1996). Cette réaction auto-immunitaire n'a cependant toujours pas été caractérisée de façon précise. L'une des hypothèses les plus plausibles en ce qui concerne le mécanisme d'auto-immunité serait un mimétisme moléculaire. Dans ce modèle, des lymphocytes T activés reconnaîtraient des séquences d'acides aminés communes entre le virus et des antigènes du soi, tels que la *myelin basic protein* (MBP) ou la *proteolipid protein* (PLP), deux composantes de la myéline, provoquant ainsi une réaction immunitaire dirigée contre cette même myéline. Des peptides ne partageant aucun résidu homologue avec la MBP ou la PLP pourraient même réagir et activer les cellules T responsables d'une telle réaction immunitaire croisée (Zhou et Hemmer, 2004). La présence de myéline phagocytée chez les cellules astrogliales et microgliales exprimant un CMH

de classe II renforce également l'hypothèse de la présentation autoantigénique (Hunter et Rodriguez, 1995). Lors de ce phénomène auto-immunitaire, une infiltration de lymphocytes T CD4+ autoréactifs et de macrophages surviendrait et entraînerait une destruction de la myéline au niveau des foyers d'inflammation (Holz *et al.*, 2000).

3.2 Les réponses immunitaires associées

La réponse immunitaire générale peut varier beaucoup d'un coronavirus à l'autre, même à l'intérieur d'une même espèce. Les mécanismes antiviraux peuvent également changer selon le type de cellule infectée et le tissu infecté, comme en témoignent les infections du SNC (Ramakrishna *et al.*, 2002). Cependant, les patrons de réponse immunitaire représentatifs de l'infection coronavirale sont connus. Ces infections ont principalement été caractérisées dans le SNC suite à des inoculations de différentes souches de MHV. Le modèle de réponse immunitaire présenté ici sera donc en grande partie inspiré du modèle du MHV.

3.2.1 La réponse immunitaire générale

Les différences observées au niveau de la réponse immunitaire sont une conséquence des structures et des biologies variables des coronavirus. Les relations directes entre le coronavirus et les composantes du système immunitaire n'ont d'ailleurs jamais été caractérisées de façon précise. Les épitopes immunodominants varieraient d'un virus à l'autre, mais seraient généralement contenus à l'intérieur de la glycoprotéine S (Daniel *et al.*, 1993; Lu *et al.*, 2004).

L'immunité antivirale est habituellement contrôlée par les lymphocytes T CD8+, mais différentes composantes du système immunitaire doivent d'abord intervenir afin de limiter la dispersion du virus dans l'organisme. En réponse à l'infection virale, la cellule infectée sécrète de l'interféron (IFN) de type I, soit les IFN- α et - β . Ces cytokines sont produites par la majorité des cellules en réponse directe à une infection virale et sont responsables de l'induction de l'état antiviral de la cellule, établissant ainsi la première ligne de défense contre l'infection. Bien que les mécanismes d'action des IFN de type I ne soient que partiellement compris, on sait que l'interaction de ces molécules avec leurs récepteurs induit l'activation de centaines de gènes impliqués dans les fonctions antivirales de la cellule (Basler et Garcia-Sastre, 2002). Lorsque le virus est détecté par l'organisme, en réaction à l'état antiviral cellulaire et/ou à la cytolysse provoquée par le virus, les cellules de l'inflammation, ou de l'immunité innée, sont recrutées au

site d'infection. Les premières cellules immunitaires à atteindre le site d'infection sont habituellement les neutrophiles et les macrophages, mais dans le SNC, les macrophages résidents, les cellules microgliales, prennent l'infection en charge. Les cellules microgliales relâchent alors des cytokines et des chimiokines afin de favoriser le recrutement et l'infiltration d'autres macrophages et cellules du système immunitaire (Lane *et al.*, 2000). Parmi ces cellules, les cellules NK jouent un rôle primordial dans l'élimination du virus étant donné leur activité cytotoxique sur les cellules infectées. Par leur sécrétion accrue d'IFN- γ , les cellules NK sont également responsables de l'induction de l'expression de CMH de classe II ainsi que de l'activation des lymphocytes T et de l'immunité acquise (Marten *et al.*, 2001).

Les cellules de l'immunité acquise sont principalement représentées par les lymphocytes B et T. Suite à leur recrutement au SNC, les cellules T CD4⁺ activées continuent de sécréter de l'IFN- γ , d'autres cytokines proinflammatoires et différentes chimiokines afin de stimuler la réponse immunitaire et de poursuivre le recrutement de cellules immunitaires (Lane *et al.*, 2000; Sainz Jr *et al.*, 2004). C'est d'ailleurs l'IFN- γ qui serait la cytokine la plus importante lors de l'élimination du MHV (Houtman et Fleming, 1996). Lors de ce même processus d'élimination, les lymphocytes T CD4⁺ seraient aussi importants que les lymphocytes T CD8⁺. Contrairement aux cellules T CD8⁺, lesquelles se spécialisent dans la destruction des cellules infectées, les lymphocytes T CD4⁺ rempliraient des fonctions reliées à l'amplification de la réponse immunitaire. La stimulation de la présentation antigénique par les macrophages et/ou les cellules microgliales du SNC et de la synthèse d'anticorps par les lymphocytes B représenterait notamment des événements de l'immunité influencés par les cellules T CD4⁺ (Marten *et al.*, 2001). Les mécanismes immunitaires qui interviennent lors des infections aiguës et persistantes de coronavirus sont aussi différents. Effectivement, l'immunité cellulaire semble contrôler la réplication virale dans les stades précoces de l'infection, comme en témoigne l'apparition rapide des cellules T activées au site d'infection. De plus, chez des souris déficientes en lymphocytes B, des infections avec le MHV ont démontré que les cinétiques d'élimination virale étaient indépendantes de la présence d'anticorps, indiquant un contrôle initial de la réplication par l'immunité cellulaire (Ramakrishna *et al.*, 2002). L'immunité humorale contrôlerait plutôt la phase de persistance de l'infection coronavirale. Pour MHV, l'absence d'anticorps avant l'élimination virale indique que les anticorps jouent un rôle mineur durant l'infection aiguë. Les anticorps seraient donc synthétisés plus tard dans l'infection et pourraient réapparaître dans l'organisme de l'hôte advenant une réémergence du virus (Marten *et al.*, 2001).

3.2.2 L'auto-immunité

Bien qu'elle n'ait pas été caractérisée en profondeur, on sait maintenant que l'auto-immunité joue un rôle majeur dans la pathogenèse des infections coronavirales, chez le MHV notamment. L'auto-immunité représente également un phénomène de plus en plus observé dans différentes pathologies dont le déclenchement serait relié à une présence infectieuse (Oldstone, 1998). Tel que mentionné précédemment, la réaction croisée observée durant la pathogenèse coronavirale serait dirigée contre une protéine de la myéline, probablement la MBP ou la PLP, et entraînerait une démyélinisation des axones, comme dans le cas de la sclérose en plaques. Plusieurs observations supportent l'hypothèse de l'auto-immunité dans le modèle de pathologie démyélinisante induite par le MHV. Par exemple, il a été démontré que des souris saines qui subissent un transfert adoptif de lymphocytes T autoréactifs spécifiques à la MBP, pouvaient développer des lésions démyélinisantes. Il a aussi été démontré que suite à des infections de souris irradiées, une réplication virale était observée sans qu'il n'y ait toutefois de démyélinisation, confirmant ainsi l'implication du système immunitaire dans ce phénomène (Wege *et al.*, 1995). La raison pour laquelle la myéline pourrait être ciblée par le système immunitaire pourrait être attribuable à un défaut de maturation des lymphocytes T lors du processus de tolérance immunitaire dans le thymus ou parce que des épitopes de la myéline seraient cryptiques, comme la MBP (Hafler *et al.*, 2004). Les lymphocytes T spécifiques à la myéline seraient alors présents dans le répertoire normal de l'hôte et pourraient être activés si des épitopes de la MBP ou de la PLP étaient détectés par le système immunitaire (Barnaba et Sinigaglia, 1997). De plus, la dégénérescence du récepteur de cellules T (RCT) pourrait également avoir un rôle à jouer dans l'auto-immunité induite par les coronavirus. Dû à la taille imposante du répertoire antigénique pouvant être reconnu par le RCT, un haut degré de réaction croisée serait essentiel afin de contrôler les infections (Wucherpfennig, 2004). La flexibilité de la reconnaissance du RCT ferait donc en sorte que plus d'un peptide pourrait déclencher une réponse immunitaire, ce qui favoriserait les réactions auto-immunitaires.

Bien que le mécanisme immunitaire qui intervient dans la pathogenèse de la sclérose en plaques est peut-être dû à un mimétisme moléculaire, d'autres modèles d'auto-immunité existent aussi. Parmi ceux-ci, la propagation de déterminants s'avère un modèle qui pourrait également être impliqué dans la pathogenèse de maladies démyélinisantes. L'hypothèse derrière ce modèle propose un phénomène par lequel des cellules T autoréactives sont activées *de novo* par des antigènes du soi libérés suite à des dommages tissulaires causés par l'inflammation (Vanderlugt

et al., 2000). Autrement dit, lors de l'inflammation causée par l'infection virale, des protéases cellulaires pourraient dégrader et/ou libérer des épitopes cryptiques et favoriser leur présentation par le CMH, ce qui déclencherait une nouvelle réaction dirigée contre l'antigène nouvellement exposé (Olson *et al.*, 2001). Le *bystander immune response* représente un autre mécanisme d'auto-immunité qui pourrait intervenir lors de pathologies neurodégénératives. Ce modèle propose une réponse immunitaire indirecte et non-spécifique, déclenchée par une réaction immunitaire située à proximité (Houtman et Fleming, 1996). Parfois, la réponse survient suite à l'activation de cellules de l'immunité par des lymphocytes T, eux-mêmes activés par divers composants telles que des superantigènes (Olson *et al.*, 2001). Bien que les lymphocytes T CD8+ soient surtout reconnus pour intervenir lors d'infections virales, les cellules T CD4+ joueraient un rôle majeur dans les processus d'auto-immunité induite par des virus. Aucun des mécanismes auto-immunitaires décrits précédemment n'a été unanimement associé aux pathologies démyélinisantes du SNC, mais il est à noter que ceux-ci ne sont pas mutuellement exclusifs.

3.3 Les traitements anti-viraux

Aucun traitement anti-coronavirus efficace n'a été mis sur pied jusqu'à présent. Depuis l'éruption du SARS-CoV en novembre 2002, les efforts dédiés à cette tâche ont été décuplés, de peur qu'une nouvelle épidémie de SRAS ne refasse surface. Plusieurs approches ont été employées afin de contrôler les infections et la réplication coronavirale, autant *in vitro* que *in vivo*. Les résultats obtenus n'étaient cependant pas toujours très concluants et certaines des approches tentées requièrent encore une mise au point afin de présenter un taux d'efficacité suffisant, principalement *in vivo*. En ce moment, beaucoup d'études mettent l'accent sur la vaccination, surtout en ce qui concerne le SARS-CoV (Groneberg *et al.*, 2005). Bien que la mise au point de tels vaccins demeure un travail très laborieux, plusieurs essais tentés *in vivo* laissent croire que cette avenue pourrait s'avérer très prometteuse (Peiris *et al.*, 2004).

Au cours des trois dernières années, plusieurs molécules et traitements ont été testés chez le SARS-CoV principalement, afin d'inhiber ou de contrôler la réplication des coronavirus. Plusieurs drogues ont notamment été testées *in vitro*. Parmi celles-ci, la chloroquine et la bafilomycine A1, tous deux des inhibiteurs d'endocytose, fonctionnent relativement bien *in vitro*, mais sont moins efficaces et spécifiques *in vivo* (Blau et Holmes, 2001). Des agents anti-inflammatoires comme les corticostéroïdes ne se sont pas avérés très efficaces non plus, pas plus

la ribavirine (Xu *et al.*, 2003), un analogue de bases habituellement efficace contre plusieurs virus à ARN et ADN. L'effet de molécules du système immunitaire comme l'IFN a également été analysé chez différents coronavirus, dont le SARS-CoV. Utilisés de façon synergique, les IFN- γ et - β présentent un effet anti-viral important comparativement aux effets marginaux que ces deux molécules ont sur la réplication coronavirale lorsqu'elles sont utilisées séparément (Sainz Jr *et al.*, 2004). Dans une étude indépendante de la dernière, l'IFN- α utilisé seul possédait également un pouvoir inhibiteur important sur la réplication du SARS-CoV (Hertzog *et al.*, 2004). Jusqu'à présent, le seul inhibiteur de coronavirus efficace qui a été reporté est l'hygromycine B. Il s'agit d'un inhibiteur de la traduction qui agit au niveau des ribosomes procaryotes et eucaryotes, qui n'est cependant pas approuvé pour des usages thérapeutiques chez l'humain ou les animaux (Neuman *et al.*, 2004).

Plusieurs molécules interagissant avec les ARN viraux ont aussi été analysées pour leur capacité à inhiber la réplication virale. Parmi celles-ci, des oligomères antisens conjugués ou non à des peptides et spécifiques à des séquences du virus se sont avérés efficaces pour le MHV (Neuman *et al.*, 2004) et le SARS-CoV (Neuman *et al.*, 2005). Selon les séquences visées, le taux d'inhibition pouvait varier. Lorsque les TRS de l'UTR 5' étaient visées, le taux d'inhibition était à son plus haut et la réplication virale pouvait être réduite de 100 fois (Neuman *et al.*, 2004). Les *small interfering RNA* (siRNA) représentent une autre forme de molécules pouvant interagir avec les ARN viraux afin de réduire leur réplication et ces derniers ont principalement été utilisés pour l'étude du SARS-CoV. Les siRNA sont en fait des oligonucléotides d'ARN double brins d'environ une vingtaine de bases qui sont spécifiques à une séquence donnée. En liant leur séquence d'ARNm, les siRNA induisent un phénomène d'interférence d'ARN, lequel consiste en la dégradation de l'ARNm cible et donc en une inhibition de l'expression du gène visé (Wu *et al.*, 2004). Plusieurs gènes viraux ou régions spécifiques de l'ARN viral ont été ciblés en utilisant cette approche. Les siRNA dirigés contre la séquence *leader* des ARN viraux (Li, T. *et al.*, 2005), le gène S (Zhang *et al.*, 2004), l'UTR 3' (Wu *et al.*, 2004) et le gène N (Zhao *et al.*, 2005) se sont avérés particulièrement efficaces pour inhiber la réplication virale *in vitro*.

La caractérisation plus poussée de certains domaines protéiques des coronavirus permet aujourd'hui d'utiliser des inhibiteurs spécifiques afin de contrer ces virus. Dans la recherche d'une cure contre le SARS-CoV, plusieurs régions du virus ont été ciblées et plusieurs inhibiteurs ont été testés. Jusqu'à présent, les deux régions les plus ciblées ont été la 3CLpro et les domaines HR de la sous-unité S2. Étant donné le rôle central de la 3CLpro dans la réplication du virus,

l'inhibition de ce domaine rendrait l'infection abortive. Un composé non-peptidique qui lie le site actif de la 3CLpro, la sabadinine (Toney *et al.*, 2004), ainsi qu'un analogue de substrat, l'hexapeptidyl CMK (Yang *et al.*, 2003), se sont avérés très efficaces pour contrer le SARS-CoV *in vitro*. Des inhibiteurs spécifiques aux régions HR se sont déjà avérés très efficaces pour plusieurs virus (Ingallinella *et al.*, 2004). Au lieu de bloquer l'infection dans la cellule, ces inhibiteurs lient les domaines HR avant même l'entrée du virus dans la cellule et empêchent ainsi le phénomène de fusion membranaire de se réaliser. Plusieurs molécules efficaces ont été testées et parmi celles-ci, on retrouve notamment le TGG et la lutéoline, deux composantes particulièrement efficaces pour inhiber le SARS-CoV (Yi *et al.*, 2004). Une molécule spécifique à la protéase *papain-like* du SARS-CoV a récemment été utilisée afin d'inhiber l'activité de l'enzyme *in vitro* (Lindner *et al.*, 2005). Cette protéase pourrait donc s'avérer une cible potentielle dans la recherche d'un traitement contre le SARS-CoV.

Parmi les traitements envisagés dans la lutte aux coronavirus et au SARS-CoV, la vaccination figure définitivement en tête de liste. Les vaccins ont l'avantage d'être administrés à titre préventif et peuvent ainsi permettre de contrer des épidémies de façon drastique. Bien que leur conception demande une mise au point minutieuse, les vaccins offrent une grande gamme de possibilités. Plusieurs approches ont été tentées jusqu'à présent dans la génération de vaccins anti-SRAS notamment et plusieurs types de vaccins sont présentement à l'étude (Marshall et Enserink, 2004) et même en phase clinique (Peiris *et al.*, 2004). Parmi les approches de vaccination tentées dans le but de contrer le SARS-CoV, notons les vecteurs viraux, les vaccins d'ADN ainsi que les vaccins sous-unitaires. Les vecteurs viraux ressemblent à des vaccins sous-unitaires, sauf qu'ils sont administrés à l'hôte sous forme d'ADN, un peu comme les vaccins d'ADN d'ailleurs. Cependant, dans les vecteurs viraux, le gène de la protéine immunodominante du virus à contrer est contenu à l'intérieur du génome d'un virus atténué, alors que le gène de la protéine est introduit directement dans un plasmide en ce qui concerne les vecteurs d'ADN (Groneberg *et al.*, 2005). Récemment, une immunité protectrice a d'ailleurs été obtenue en utilisant un vecteur viral chez la souris (Bisht *et al.*, 2004). Le vecteur consistait en un virus de la vaccine atténué à l'intérieur duquel le gène S du SARS-CoV avait été introduit. Des vaccins d'ADN ont déjà démontré une certaine efficacité pour différents virus, comme le VIH, l'herpès, l'hépatite B et C, l'influenza et le virus de la rage, et cette approche a aussi été appliquée au SARS-CoV avec un succès relativement élevé chez la souris (Wang *et al.*, 2005). Lors de cette étude, les gènes M, N et deux régions du gène S ont été introduits dans le vecteur pcDNA3.1. Bien que chacun des antigènes pouvait induire une réponse immunitaire, la meilleure réponse

humorale était fournie par les fragments du gène S, alors que la meilleure réponse cellulaire était fournie par le gène M (Wang *et al.*, 2005). En ce qui concerne les vaccins sous-unitaires, plusieurs approches ont aussi été tentées. Parmi celles-ci, une approche innovatrice consistait en l'expression de la sous-unité S1 du SARS-CoV dans un plant de tomate (Pogrebnyak *et al.*, 2005). Suite à l'ingestion du fruit par des souris, une réponse immunitaire systémique et mucoale était alors induite chez ces dernières. Toutes approches confondues, c'est généralement la protéine S qui fournit le meilleur pouvoir protecteur parmi tous les antigènes viraux (Peiris *et al.*, 2004), puisqu'elle constitue une cible majeure de la réponse immunitaire humorale et cellulaire (Hofmann et Pöhlmann, 2004).

En raison du taux de recombinaison élevé des coronavirus, les vaccins viraux atténués n'ont toujours pas été mis à l'essai sérieusement. Bien que cette approche pourrait s'avérer très efficace, le danger d'une recombinaison avec un virus de souche sauvage représente un facteur limitant dans l'élaboration de tels vaccins (Holmes, 2003). Bien que la délétion de certains gènes coronaviraux soit possible par mutagenèse dirigée (de Haan *et al.*, 2002a), une approche spécifique doit être mise au point dans le but d'empêcher la recombinaison et la réversion des virus atténués. Parmi ces approches, le changement de la configuration génomique ainsi que la modification des TRS pourraient s'avérer efficaces afin de réduire drastiquement le phénomène de recombinaison. Étant donné que les protéines S, M et N sont suffisantes pour la génération de *virus-like particles* (VLP), une telle approche pourrait également s'avérer une alternative intéressante aux virus atténués (Huang, Y. *et al.*, 2004). Les VLP consistent en des particules virales défectives qui contiennent certains antigènes à leur surface et qui sont dépourvues de génome, éliminant ainsi les possibilités de recombinaison avec un autre virus. Bref, peu importe l'approche utilisée dans la génération de vaccins, la biologie des coronavirus à l'échelle moléculaire joue un rôle majeur dans l'élaboration des stratégies utilisées.

4.0 LA BIOLOGIE MOLÉCULAIRE ET LES CORONAVIRUS

Étant donné la composition du code génétique des coronavirus ainsi que la disponibilité de techniques et d'outils permettant la manipulation de l'ADN et de l'ARN, la manipulation des génomes est aujourd'hui une approche utilisée couramment et qui peut être associée à une multitude d'applications. La biologie moléculaire occupe donc une place importante dans la recherche fondamentale sur les virus et est désormais très utilisée dans différentes facettes de la

recherche en milieu médical, telles que le diagnostic, la génération de vaccins et même la thérapie anti-cancer.

4.1 Généralités

La biologie moléculaire se définit comme étant une discipline consacrée à l'étude des molécules porteuses du message héréditaire. Plusieurs techniques de biologie moléculaire sont aujourd'hui établies et utilisées de routine dans les laboratoires de recherche. Parmi celles-ci, notons la transcription inverse, le PCR, la digestion enzymatique, l'électrophorèse, la recombinaison homologe et le clonage. Bien que ces techniques répondent à des critères et des conditions précises, il n'en tient qu'à l'imagination du chercheur afin de les combiner à d'autres et de les adapter à un contexte de recherche précis. Parmi la vaste gamme de possibilités offertes par la biologie moléculaire, l'étude des coronavirus au cours des dernières années a principalement été réalisée en faisant appel à des techniques de mutagenèse dirigée et de génétique inverse.

4.2 Les outils disponibles

Bien que les techniques de mutagenèse mentionnées auparavant puissent être employées dans divers contextes, la recherche sur les coronavirus a permis de combiner ces approches de biologie moléculaires à différents outils, souvent très ingénieux. Une des limitations principales du génome des coronavirus et des autres virus à ARN réside en leur composition, laquelle est souvent très instable, ce qui permet difficilement de la modifier (Almazán *et al.*, 2000). Plusieurs approches ont néanmoins été mises sur pied dans le but de manipuler le génome de ces virus. La recombinaison d'ARN est l'une de celles-ci et a été mise au point par l'équipe du Dr Paul Masters (Kuo *et al.*, 2000). Cette approche permet la manipulation directe du génome d'ARN du virus et d'y apporter des modifications précises. Le système de recombinaison de Masters ne permet cependant pas la mutagenèse des 20 premiers kilobases du génome viral, puisque les vecteurs donneurs utilisés pour la recombinaison sont des molécules d'ARN contenant seulement les 10 derniers kilobases du génome (Sperry *et al.*, 2005). Une approche impliquant des minigénomes viraux a également été élaborée, par l'équipe du Dr Luis Enjuanes cette fois. Ces minigénomes comprennent quelques régions du génome du virus à l'étude, dont les extrémités 5' et 3', lesquelles sont contenues à l'intérieur d'un vecteur eucaryote sous le contrôle d'un promoteur quelconque (Enjuanes *et al.*, 2001; Izeta *et al.*, 1999). Ces minigénomes ont

principalement été utilisés pour étudier les mécanismes de réplication et d'encapsidation du TGEV. Enfin, l'assemblage de virus chimériques a aussi permis la caractérisation de plusieurs protéines structurales du virus, dont les protéines S et HE du BCoV (Popova et Zhang, 2002). Lors de cette étude, l'activité biologique des protéines S et HE était évaluée en transfectant les gènes de ces protéines dans des cellules DBT, lesquelles étaient ensuite infectées par le MHV-A59. Des virus MHV chimériques possédant les protéines de surface S ou HE étaient ainsi générés et évalués pour leur capacité à infecter des cellules HRT-18, susceptibles au BCoV.

4.3 Les clones infectieux

L'outil le plus spectaculaire et le plus efficace dans l'étude des coronavirus demeure le clone d'ADNc. Cet outil consiste en une copie d'ADN du génome du virus à l'étude, introduite dans un vecteur ou non, et située en aval d'un promoteur transcriptionnel puissant. Suite à la transfection du clone d'ADNc dans une cellule eucaryote, l'ADN est transcrit en un ARN qui correspond exactement au génome du virus, puis des particules virales recombinantes, identiques aux particules de souche sauvage, sont synthétisées. L'avantage de cet outil est qu'il est relativement facile à utiliser une fois assemblé, et qu'il permet la modification du génome viral de quelque façon que ce soit. La publication no.4 de cet ouvrage traite d'ailleurs de l'assemblage d'un clone infectieux d'ADNc du HCoV-OC43, lequel a été construit dans un *bacterial artificial chromosome* (BAC) sous le contrôle du promoteur du cytomégalovirus (CMV) (St-Jean *et al.*, 2006).

Les clones infectieux de virus à ARN ne datent pas d'hier. La première expérience de clone infectieux a été réalisée en 1981 chez le poliovirus (Neumann *et al.*, 2002; Racaniello et Baltimore, 1981). Depuis, plusieurs autres virus à ARN ont vu leur génome être transcrit en ADN puis cloné. Le premier clone d'ADNc coronaviral a été assemblé en 2000, dans un BAC (Almazán *et al.*, 2000), puis d'autres clones ont ensuite été assemblés pour différentes souches. L'assemblage de ces clones a été ralenti par la taille imposante du génome des coronavirus ainsi que par la présence de séquences virales instables dans les bactéries (Gonzalez *et al.*, 2002). Des stratégies d'assemblage peu communes ont donc dû être élaborées afin de contourner les problèmes engendrés par le génome viral. Jusqu'à présent, trois approches différentes ont été utilisées pour la construction des clones d'ADNc coronaviraux. L'une d'elle est la méthode du clone linéaire, pensée et mise au point par l'équipe du Dr Ralph Baric, et utilisée pour l'assemblage des clones du TGEV (Yount *et al.*, 2000), du MHV-A59 (Yount *et al.*, 2002) et du

SARS-CoV (Yount *et al.*, 2003). Dans cette approche, la copie d'ADN du génome viral n'est pas introduite dans un vecteur, d'où l'appellation de clone linéaire. L'ADNc est donc précédé du promoteur du bactériophage T7 et est suivi d'une queue de poly(A) synthétique, lesquels permettent la transcription adéquate du clone en ARN. Une autre approche employée dans la conception de clones infectieux coronaviraux est celle du vecteur de la vaccine. Cette stratégie a été élaborée par l'équipe du Dr Stuart Siddell et s'est notamment concrétisée chez le HCoV-229E (Thiel *et al.*, 2001), le IBV (Casais *et al.*, 2001) et plus récemment chez le MHV-A59 (Coley *et al.*, 2005). Dans cette approche, le génome du virus est introduit dans le génome d'un virus de la vaccine atténué, lequel est employé comme vecteur et pour assurer la propagation du génome coronaviral dans des cellules CV-1, susceptibles au virus de la vaccine (Thiel *et al.*, 2001). L'avantage majeur de cette méthode est qu'elle utilise un système non-bactérien, permettant ainsi de contourner le problème d'instabilité dans *E. coli* (Casais *et al.*, 2001).

La troisième et dernière stratégie d'assemblage est celle du BAC et c'est d'ailleurs l'approche qui a été utilisée pour l'assemblage du clone infectieux du HCoV-OC43. Cette approche a été élaborée par l'équipe du Dr Luis Enjuanes, laquelle a collaboré avec nous pour la construction du clone d'ADNc du HCoV-OC43. Jusqu'à présent, des clones du TGEV (Almazán *et al.*, 2000), du SARS-CoV et du HCoV-OC43 (St-Jean *et al.*, 2006) ont été assemblés en utilisant cette stratégie. Les clones d'ADNc assemblés dans un BAC utilisent un système d'amplification en deux étapes dans les cellules eucaryotes. D'une part, grâce au promoteur du CMV, le clone d'ADN est transcrit en ARN par l'ARN polymérase II, puis l'ARN obtenu mène ensuite à la production de particules virales recombinantes. Bien que l'assemblage d'un clone d'ADNc dans un BAC représente une entreprise laborieuse comparativement aux autres approches disponibles, le produit fini est définitivement celui qui présente le plus d'avantages. Tout d'abord, le BAC est un vecteur dérivé du facteur F de *E. coli*, lequel peut demeurer extrêmement stable pour plus d'une centaine de générations. Le BAC est également un plasmide très facile à manipuler et nécessite des méthodes de transformation, d'extraction et de clonage relativement conventionnelles (Shizuya *et al.*, 1992). Finalement, le système de clone d'ADNc fourni par le BAC est très facile à utiliser. En ce qui concerne le HCoV-OC43, la simple transfection du clone d'ADNc dans des cellules BHK-21 mène directement à la production de particules virales recombinantes. Le clone infectieux du HCoV-OC43 ouvre la porte à de nombreuses possibilités en ce qui concerne la caractérisation du virus, mais également à l'implication de cet outil à diverses applications.

CHAPITRE II : PUBLICATION no.1 :

Human respiratory coronavirus OC43 : Genetic stability and neuroinvasion

**HUMAN RESPIRATORY CORONAVIRUS OC43:
GENETIC STABILITY AND NEUROINVASION**

Julien R. ST-JEAN¹, H  l  ne JACOMY¹, Marc DESFORGES¹, Astrid VABRET², Fran  ois
FREYMUTH², and Pierre J. TALBOT^{1*}

Running title: Genetic stability and neuroinvasion by human coronavirus OC43

¹ Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier,
531 boulevard des Prairies, Laval (Qu  bec), CANADA H7V 1B7
and

² Laboratoire de Virologie Humaine et Mol  culaire,
Centre Hospitalier R  gional et Universitaire de Caen,
Avenue Georges Cl  menceau, 14033 Caen, FRANCE

* Correspondent footnote:

Dr. Pierre J. Talbot
Laboratory of Neuroimmunovirology
INRS-Institut Armand-Frappier
531 boulevard des Prairies,
Laval (Qu  bec)
CANADA
H7V 1B7
Tel.: (450) 686-5515
Fax: (450) 686-5566
E-mail: pierre.talbot@inrs-iaf.quebec.ca

Word counts:

Abstract: 265

Text: 5659

RÉSUMÉ

Dans cette étude, les séquences complètes des génomes de la souche de laboratoire du coronavirus humain OC43 (HCoV-OC43), provenant de l'*American Type Culture Collection* (ATCC), et d'un isolat clinique du HCoV-OC43, désigné Paris, ont été obtenues. Les deux génomes ont une taille identique de 30 713 nucléotides, en excluant la queue de poly(A), et diffèrent par seulement six nucléotides. Ces six mutations sont dispersées à travers le génome du virus et entraînent seulement deux substitutions d'acides aminés, une localisée dans le gène S (I958F) et l'autre dans le gène N (V81A). Il a également été démontré que les deux variants pouvaient atteindre le système nerveux central (SNC) suite à des inoculations intranasales dans des souris BALB/c, démontrant ainsi leurs propriétés neuroinvasives. Même si la souche ATCC pouvait pénétrer le SNC plus efficacement que l'isolat Paris 2001, ces résultats suggèrent que le coronavirus respiratoire humain OC43, isolé dans les années 60, possédait des propriétés neuroinvasives intrinsèques avant même d'être propagé dans des cerveaux de souris. Ces résultats démontrent aussi que la structure moléculaire du HCoV-OC43 est très stable dans l'environnement (les deux variants ayant été isolés à 40 ans d'intervalle) malgré la transmission du virus et les chances de persistance dans l'hôte. Les génomes des deux variants OC43 présentent respectivement 71%, 53,1% et 51,2% d'identité avec les génomes du *mouse hepatitis virus* A59 (MHV-A59), de la souche Tor2 du virus du SRAS (SARS-HCoV Tor2) et du coronavirus humain 229E (HCoV-229E). HCoV-OC43 possède également des motifs très bien conservés par rapport au génome du SARS-HCoV Tor2, principalement dans l'ORF1b. Ces résultats suggèrent que le HCoV-OC43 et le SARS-HCoV partagent plusieurs propriétés fonctionnelles importantes et que le HCoV-OC43 pourrait être utilisé comme modèle pour étudier la biologie du SARS-HCoV sans la nécessité de confinement biologique de niveau 3.

ABSTRACT

The complete genome sequences of the human coronavirus OC43 (HCoV-OC43) laboratory strain from the American Type Culture Collection (ATCC), and a HCoV-OC43 clinical isolate, designated Paris, were obtained. Both genomes are 30,713 nucleotides long, excluding the poly(A) tail, and only differ by six nucleotides. These six mutations are scattered throughout the genome and give rise to only two amino acid substitutions, one in the spike protein gene (I958F) and the other in the nucleocapsid protein gene (V81A). Furthermore, the two variants were shown to reach the central nervous system (CNS) following intranasal inoculation in BALB/c mice, demonstrating neuroinvasive properties. Even though the ATCC strain could penetrate within the CNS more effectively than the Paris 2001 isolate, these results suggest that intrinsic neuroinvasive properties already existed for the HCoV-OC43 ATCC human respiratory isolate from the 1960s before it was propagated in newborn mouse brains. It also demonstrates that the molecular structure of HCoV-OC43 is very stable in the environment (the two variants were isolated about 40 years apart) despite virus shedding and chances of persistence in the host. The genomes of the two HCoV-OC43 variants display 71%, 53.1% and 51.2% identity with those of mouse hepatitis virus A59 (MHV-A59), SARS human coronavirus Tor2 strain (SARS-HCoV Tor2) and human coronavirus 229E (HCoV-229E), respectively. HCoV-OC43 also possesses well-conserved motifs with regards to the genome sequence of the SARS-HCoV Tor2, especially in ORF1b. These results suggest that HCoV-OC43 and SARS-HCoV may share several important functional properties and that HCoV-OC43 may be used as a model to study the biology of SARS-HCoV without the need for level three biological facilities.

INTRODUCTION

Human coronaviruses (HCoV), members of the *Coronaviridae* family, are ubiquitous in the environment and are responsible for up to one third of common colds (41). In the past years, we have provided experimental evidence that this virus possesses neurotropic and neuroinvasive properties: it persists in neural cell cultures (7, 8) and human brains (9). Of the two HCoV serotypes available, HCoV-OC43 was selected for further characterization of persistence in the nervous system because of a more efficient infection of primary neural cell cultures (11), as well as a trend towards association with neurological disease (9).

Coronaviruses are enveloped viruses that possess a positive-strand RNA genome of up to 31 kilobases (kb), which represents the largest known genome among all RNA viruses (35). This genome comprises several genes encoding several structural and non-structural proteins. Among these proteins, the S protein is biologically very important because it could be implicated in determination of tropism (3) and its modulation (50). Indeed, the S protein could be associated with the capacity of the virus to reach the central nervous system (CNS) and possibly trigger neurological disorders (9, 22). It could also be responsible for conferring the strong degree of host species specificity observed with coronaviruses (28).

Only the 3' one third of the HCoV-OC43 genome has been sequenced over the years. Therefore, until now, the complete sequence of the open reading frames (ORFs) 1a and 1b, known as the replicase gene, was still undetermined. This gene is essential for coronavirus survival because it contains several motifs, which could be involved in various important viral functions such as transcription, replication and pathogenesis (66). The products encoded by these two ORFs are polyprotein precursors, which are processed by two or three different proteinases encoded by ORF1a. These proteinases could include two papain-like proteases (PLP1 and PLP2) and a polio 3C-like protease (3CLpro), which presents the most important cleavage activity. The

3CLpro essential function is reflected by its capacity to cleave at many sites in the replicase polyproteins and to release the key replicative functions, such as the RNA-dependant RNA polymerase (RdRp) and the RNA helicase (67).

The HCoV-OC43 strain belongs to the second genetic group, just as SARS-HCoV apparently does (51). The latter is responsible for the severe acute respiratory syndrome (SARS), which is a life-threatening form of pneumonia (46). Since the outbreak of SARS in the fall of 2002 (60), a lot of work has been done to sequence the entire genome of the virus (34) and to understand the mechanisms underlying virus pathogenesis. As presented here, the whole genome of HCoV-OC43 has now been sequenced and since this human strain is the most related to SARS-HCoV, it could be used as a model for the study of the SARS-HCoV without the drawbacks of level three biological confinement. Comparisons with the SARS-HCoV nucleotide and amino acid sequences (34) revealed that the two viruses share extensive homology in some important motifs involved in viral replication and pathogenesis. Indeed, the most significant homology between the genomes of the HCoV-OC43 strain and the one of the SARS-HCoV Tor2 isolate is found in the ORF1b region, which comprises the RdRp and helicase motifs (16). The 3CLpro motif of HCoV-OC43 also displays an important level of identity with the one of SARS-HCoV. This finding is noteworthy since the SARS-HCoV 3CLpro so far represents the most promising target for SARS therapy (58).

In the current study, we report the complete genome sequences of the HCoV-OC43 strain from the American Type Culture Collection (ATCC), as well as a HCoV-OC43 respiratory clinical isolate, designated HCoV-OC43 Paris. Both genomes are 30,713 nucleotides (nt) long, share the same genomic organization and only differ by six nucleotides. Differences found in the genome of the HCoV-OC43 Paris isolate, compared with the genome of HCoV-OC43 ATCC, give rise to only two amino acid substitutions, which are located in the S (I958F) and the N (V81A) protein genes. Following intranasal inoculation in BALB/c mice, the HCoV-OC43

ATCC strain, as well as the Paris isolate, reached the CNS, where they replicated and disseminated, although mice were apparently more easily infected with the ATCC strain than with the Paris isolate. These results suggest that both viruses possess the ability to reach and infect neural cells *in vivo*. The fact that a natural OC43 isolate has an intrinsic capacity to invade and replicate within the mouse CNS also suggests that the HCoV-OC43 ATCC strain has not acquired its neuroinvasive properties following propagation in newborn mouse brains. Bioinformatics analysis were also performed on the HCoV-OC43 genome. These analysis revealed that this virus strain is closely related to MHV-A59 and that it displays significant identity levels with important functional domains of the SARS-HCoV. These data provide evidence that HCoV-OC43 could be used as a model for the study of other group 2 coronaviruses, including SARS-HCoV, and that it will facilitate understanding of the biology of this emerging viral strain.

MATERIALS AND METHODS

Viruses and cell lines.

The ATCC HCoV-OC43 strain (ATCC number VR-759), isolated in the 1960s, and the Paris clinical respiratory isolate, isolated in March 2001, were grown on a HRT-18 cell line (human adenocarcinoma rectal) as described previously (37). The clinical sample (HCoV-OC43 Paris) was isolated from the respiratory tract of a 68-year old immunocompromised male who was not related whatsoever to laboratory work and was not in contact with any laboratory workers who had manipulated the HCoV-OC43 ATCC virus. A RT-PCR was performed to specifically detect for the presence of the HCoV-OC43 RNA and an aliquot of the clinical sample was then used to infect the HRT-18 cell line. The HCoV-OC43 ATCC strain and the Paris isolate were never cultured at the same time and stringent laboratory precautions were used in order to eliminate possible cross-contamination.

Acute infections of cells.

Cells were infected at a multiplicity of infection (MOI) of 0.02 and 0.2 for the ATCC strain and the Paris isolate, respectively. The fifth passage of the ATCC strain and the eight passage of the Paris isolate were respectively used to perform the infections. Cell lines at 70% confluence were infected with the appropriate virus stock in the presence of *L*-1-tosylamide-2-phenylethyl-chloromethyl ketone-treated trypsin (TPCK) (10 U/mL; Sigma-Aldrich Canada Ltd.) and 1% (v/v) heat-inactivated fetal calf serum (FCS) and incubated at 33°C for four days in a 5% (v/v) CO₂ humid atmosphere.

Mice and inoculations.

In order to determine the susceptibility of mice to an infection by HCoV-OC43 ATCC and HCoV-OC43 Paris variants, MHV-seronegative 14-day post-natal BALB/c mice (Charles River Laboratories, St-Constant, Québec, Canada) were inoculated intra-nasally using 5 µL of a

virus stock solution containing 10^6 TCID₅₀/mL. Five mice, inoculated by HCoV-OC43 ATCC or HCoV-OC43 Paris variants, were sacrificed every two days post-infection (dpi) and processed for detection of infectious virus particles. Every two days, two mice infected by HCoV-OC43 ATCC were processed for immunohistochemical detection of viral antigens.

Immunohistochemistry.

Mice were perfused by intraventricular injection of 4% (v/v) paraformaldehyde, under deep ketamine-xylazine anesthesia, as previously described (22). Brains were dissected-out and sectioned at a thickness of 40 μ m with a Lancer Vibratome. Sections were collected in 0.05M TRIS-buffered saline (TBS) and then incubated for 2 h at 37°C in a 1/1000 dilution of an ascites fluid from mouse MA b 1-10C.3, directed against the spike protein of HCoV-OC43 (7). Sections were then rinsed and processed with Vectastain ABC Kit (Vector Laboratories, Burlingame, CA). Labeling was revealed with 0.03% (w/v) DAB solution (Sigma) and 0.01% (v/v) H₂O₂, which yielded a dark brown product.

Infectious virus assays.

Brain and lung were dissected, homogenized in 10% (w/v) sterile PBS, centrifuged at 4°C, 20 min at 1,000 x g, then supernatants were immediately frozen at -80°C and stored until assayed. The extracts were processed for the presence and quantification of infectious virus by an indirect immunoperoxidase assay, as previously described (22). Briefly, HCoV-OC43-susceptible HRT-18 cells were inoculated with serial logarithmic dilutions of each tissue sample. After 4 days of incubation at 33°C in a 5% (v/v) CO₂ humid atmosphere, the cells were washed in PBS and fixed with 0.3% (v/v) hydrogen peroxide (H₂O₂) in methanol. After washing with PBS, they were incubated for 2 h at 37°C in a 1/1000 dilution of an ascites fluid from mouse MA b 1-10C.3. Afterwards, cells were washed in PBS and HRP-goat anti mouse immunoglobulins

(DAKO; Diagnostics Canada Inc., Mississauga, ON) were added and incubated for 2h at 37°C. Antibody complexes were detected by incubation in DAB (Sigma), with 0.01% (v/v) H₂O₂.

RNA extraction, RT (reverse transcription) and PCR.

Following infections, cells were washed with PBS and total RNA was extracted from the cells using the GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich) as recommended by the manufacturer. The RNA was then quantified and 3 µg were directly used for RT using the Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) (Invitrogen). For each RT, 500 ng of oligo (dT) primer and 0.5 mM dNTPs (Amersham Biosciences) were used and the reactions lasted between 50 and 60 minutes at 37°C. Two µL of the RT cDNA were then used to perform the polymerase chain reaction (PCR) amplifications. The Expand High Fidelity *Taq polymerase* (Roche) was used to amplify the HCoV-OC43 genome in six segments, in combination with primers listed in Table 1. All amplifications were performed using the Cetus DNA thermal cycler (Perkin Elmer, Applied Biosystems) and appropriate annealing temperature was used for each specific reaction. Except for the PCR JUB3-12, which required a higher annealing temperature of 65°C, all other annealing temperature used corresponded to the melting temperature of the primers. For each PCR amplification, at least six reactions were performed, pooled together, migrated on a 0.8% (w/v) agarose gel (SeaKem) and gel extracted using the QIAEX[®] II Gel Extraction Kit (Qiagen) prior to sequencing.

Rapid amplification of cDNA ends (RACE) and cloning.

The RACE amplifications, cloning and sequencing were performed for both 5' and 3' ends of HCoV-OC43 ATCC strain and HCoV-OC43 Paris isolate. Primers from the kit used for the RACE are listed in Table 1. A RT reaction of the 5' end was performed using the GeneRacer™ Kit (Invitrogen) as recommended by the manufacturer whereas RT of the 3' end was performed only by using the GeneRacer™ oligo dT primer provided in the kit (Table 1). In

order to amplify both ends, primers from the kit were used in combination with primers specific for the HCoV-OC43 genome. Therefore, the GeneRacer™ 5' nested primer was used with JUB2 primer and GeneRacer™ 3' nested primer was combined to JUMO1 primer for the ATCC strain, and JUO8 primer for the Paris isolate. Amplicons of the 5' end of both viruses and of the 3' end of the Paris isolate were cloned using the Zero Blunt® TOPO PCR Cloning Kit for Sequencing (Invitrogen), whereas amplicons of the 3' end of the ATCC strain were cloned using the TOPO® XL PCR Cloning Kit. The RACE 5' clones were sequenced using M13 universal forward and reverse primers and RACEJUB1 and RACEJUB2 primers, and RACE 3' clones were sequenced using M13 universal forward and reverse primers and JUO7 primer.

Sequencing.

Sequencing reactions were performed by Bio S&T (Montreal, Canada), using the dideoxy method (Sanger) and specific primers, which are listed in Table 2. As described above, PCR products were directly sequenced for both genomes and both strands were sequenced in each case, including RACE clones. For each genome, at least two RACE 5' and 3' clones were sequenced for both isolates. Sequences obtained from chromatograms were aligned using BLAST (bl2seq) from NCBI and were analysed using the Chromas 2 software.

Bioinformatics analysis.

Bioinformatics analysis were performed by Sequence Bioinformatics (Montreal, Canada). The BLAST program was used to perform genome versus genome and gene versus genome alignments. RNA folding was analysed using MFOLD. Phylip was used for phylogenetic tree construction. The FASTA-formatted sequences of the complete genomes were aligned with clustalW (v1.82), using the default parameters for DNA alignments. The Phylip output option of clustalW was used to produce a multiple alignment file that was used as input for dnaml (v3.6), which produced an unrooted maximum-likelihood phylogenetic tree using the default parameters. ORF analysis was performed using tools from the EMBOSS suite. In the case of SARS-HCoV

and HCoV-OC43 ATCC, the extracted ORFs were submitted to HMMPFAM, of the HMMER suite, for motif detection against the PFAM database. The amino acid sequences of the known expressed proteins were also submitted to HMMPFAM and the patmatmotif tool of EMBOSS. This tool performs motif scanning against the PROSITE motif database.

Nucleotide sequence accession number.

The GenBank sequence accession numbers for the complete genome of the HCoV-OC43 ATCC strain and the Paris isolate are respectively AY585228 and AY585229.

RESULTS

Amplification and sequencing of HCoV-OC43 ATCC and Paris genomes.

The genomes of the HCoV-OC43 ATCC strain and of the Paris isolate were amplified in six fragments by RT-PCR in order to be sequenced (Fig. 1 and Table 1). The PCR products encompassed the entire genome of the viruses and overlapped each other to make sure the final sequences were complete. Primers used for the amplifications of ORF 1a and 1b were created using the sequence of the bovine coronavirus Quebec strain (BCoV Quebec) (63), which displays 97% identity in this region of the genome and was known to share 92% identity with the 3' 9 kb of the HCoV-OC43 genome (24, 26, 29, 37, 38, 39). Primers used to amplify the 3' region were designed on the basis of sequences of HCoV-OC43 available in GenBank. The gene walking approach was used to sequence the whole genome of the HCoV-OC43 ATCC strain whereas the Paris isolate was sequenced using the primers generated during the sequencing of the ATCC strain (Table 2).

Main features of the HCoV-OC43 genome.

The genomes of the two variants contain 30,713 nt, excluding the poly(A) tail, and include nine main ORFs flanked by 5'- (nt 1-209) and 3'- (nt 30426-30713) untranslated regions (UTRs) (Fig. 1 and Table 3). The genome of HCoV-OC43 contains multiple secondary ORFs, scattered throughout the genome in all frames and in both orientations (data not shown). Using the ShowORF software, it was possible to determine that the translation complex could potentially use several of these ORFs in the 5' -> 3' orientation, using for instance a translation reinitiation mechanism (20, 36). Using better characterized coronaviruses, putative transcription-regulating sequences (TRSs) of HCoV-OC43 were also identified (4, 44, 58) (Table 3). These sequences are found at the 5' end of each viral RNA, genomic or subgenomic, and represent signals for the discontinuous transcription of subgenomic mRNA (49). The identified canonical

core sequence for HCoV-OC43 was 5'-UCUAAAC-3', but it was not always perfectly conserved throughout the genome (Table 3).

Using bioinformatics tools and well-characterized coronaviruses (15, 18, 67), it was also possible to draw a precise map of the main domains contained within the polyprotein 1ab and to determine the location of the putative viral proteolytic cleavage sites (Fig. 2). Most of the main motifs found throughout the genome following bioinformatics analysis corresponded to expected motifs found in other coronaviruses. Indeed, the PLP1 and PLP2 motifs, the membrane-spanning domains (TM), the 3CLpro motif as well as the RdRp and the RNA helicase motifs were found in ORF1ab at the expected positions, compared to other coronaviruses. As HCoV-OC43 and BCoV possess a high degree of identity, the positions of the cleavage sites were determined using the BCoV model (15) and 14 sites were identified in the polyprotein 1ab. Among these cleavage sites, three are recognized by PLP1 or PLP2 and the 11 others are recognized by the 3CLpro, generating mature products containing key motifs for viral transcription and replication. A putative ribosomal frameshift -1 was also identified upstream the intersection of ORF1a and ORF1b. The slippery sequence ₁₃₃₃₄UUUAAAC₁₃₃₄₀ was found at the 3' end of ORF1a and is thought, in combination with RNA pseudoknot structures, to be implicated in the frameshift, which would occur at the C₁₃₃₄₀ nt (58).

High degree of identity between the HCoV-OC43 ATCC strain and the Paris isolate.

Differences in nucleotides and amino acids between the HCoV-OC43 ATCC strain and the Paris isolate are presented in Table 4. In total, only six nucleotides differ between the two variants. These mutations are located in the UTR5', in the ns2, S, M and N genes as well as in the UTR3'. According to the MFOLD software, mutations located in the UTRs would not affect RNA folding (data not shown) and would therefore not have any effect on viral transcription and replication. Mutations located in the ns2 and M genes would not affect virus biology since they do not give rise to any amino acid substitution. However, two mutations lead to amino acid

substitutions. The first is located at nt 26,514, in the S2 subunit of the S gene, and give rise to the I958F (ATCC → Paris) mutation whereas the second is located at nt 29,320, in the N gene, and give rise to the V81A mutation.

Neuroinvasion in BALB/c mice.

After inhalation of virus, mice were processed for histochemical labeling of HCoV-OC43 ATCC antigens. Cells positive for viral antigens were first observed around 3 dpi in the olfactory bulb (OB), as patches of labeled neurons (Fig. 3A). No cells positive for viral antigens could be seen in other part of the brain, even near perivascular blood cells. At 7 dpi, viral antigens were detected in all brain regions, indicating a rapid dissemination throughout the CNS (Fig. 3B). Five mice of each group, infected with HCoV-OC43 ATCC and Paris, were sacrificed every 48 hours and virus titers were measured in the CNS and lung. Even though mice inhaled a viral suspension, virus was rarely found in the lung (limit of detection was $10^{1.5}$ TCID₅₀/gram due to a lung extract toxicity on HRT-18 cells) and only when brain titers reached at least 10^4 TCID₅₀/gram (data not shown). HCoV-OC43 ATCC infectious virus could be detected in mouse CNS, as early as two dpi. Viral titers were maximal around 4 dpi and remained high during the whole experiments (Fig. 3C). When virus reached the brain, replication of HCoV-OC43 ATCC led to a fatal encephalitis. Infectious HCoV-OC43 Paris could only be detected in mice starting at 6 dpi (Fig. 3C) and a lower number of mice were productively infected by HCoV-OC43 Paris than by HCoV-OC43 ATCC. Nevertheless, when infectious virus reached the brain, infectious virus titers were comparable for the two HCoV-OC43 variants, suggesting that the ATCC and Paris variants both exhibit neuroinvasive and neurotropic properties.

Comparison of HCoV-OC43 with other coronaviruses.

HCoV-OC43 is part of the second genetic group of coronaviruses (30) and displays higher identity levels with virus strains that belong to this group, including SARS-HCoV. The coronavirus strains that present the highest degree of identity with HCoV-OC43 are BCoV and

MHV-A59, with 95% and 71% identities, respectively. HCoV-OC43 and BCoV are very related at the nucleotide level and most of the differences between the two genomes are found in the S1 subunit of the S gene, suggesting that the two virus strains possess similar biological properties but display a different cellular tropism. SARS-HCoV and HCoV-229E display 53.1% and 51.2% identity with regards to the HCoV-OC43 strain. Although SARS-HCoV is apparently part of group 2 (51), the overall identity level with the OC43 strain is not striking, but the two strains present a very high degree of amino acid identity in some important functional domains, such as the RdRp, the RNA helicase and the 3CLpro.

A phylogenetic unrooted tree regrouping seven coronavirus strains from the three genetic groups was obtained using the complete genome sequences of all strains (Fig. 4). This tree is the first one which includes the complete genome of the HCoV-OC43 strain. It shows that HCoV-OC43 and BCoV are evolutionary very related and that they form a clade with MHV-A59. Although SARS-HCoV is apparently part of group 2 (51), the analysis shows that it is more divergent from strains of the previous clade and that infectious bronchitis virus (IBV) and SARS-HCoV display the highest divergence among the strains analysed. Group 1 coronaviruses are also grouped in such a clade.

A BLAST analysis with coronaviruses from all three genetic groups showed that different degrees of identity exist between several regions of different virus strains, but that the most conserved region among all coronaviruses is located within ORF1b (data not shown). More stringent BLAST analysis was carried out on the genome sequences of HCoV-229E, MHV-A59 and SARS-HCoV using the HCoV-OC43 genome as a reference (data not shown). Among all genes analysed, the most significant identity levels were found in ORF1ab as well as in the S2 subunit of the S gene. Significant identity levels were observed with MHV-A59 and SARS-HCoV. With regards to the ORF1ab, the identity levels were usually lower in ORF1a than in ORF1b and were more significant in the case of MHV-A59 than for SARS-HCoV. Moreover,

identity was more significant at the amino acid level. Several domains which are essential for viral replication, such as the 3CLpro, RdRp and helicase domains, are very interesting because of their functional importance. The S2 subunit of the S gene also displayed high identity levels with its counterparts from other coronavirus. For instance, the MHV-A59 S2 subunit displayed 76% identity and 88% similarity, while the S1 subunit presented only 53% identity and 65% similarity. This result is logical since it has been shown that the membrane fusion function resides within the S2 subunit (54, 62) whereas the S1 subunit is involved in receptor binding (56) and determination of tropism (3), which is different from one virus to another.

Since SARS-HCoV is now considered as a serious pathogen that has recently emerged and that we believe HCoV-OC43 could represent an excellent model for the study of this virus, it was of interest to analyze some functionally important motifs which display significant identity levels with the HCoV-OC43 genome. The most striking identities between the two strains were found mainly in ORF1b, albeit the 3CLpro motif, in ORF1a, also presented a significant identity level. The cleavage product containing the 3CLpro motif displayed 48% identity and 64% similarity with the corresponding region of HCoV-OC43. Among the three viral proteases that play a role in the processing of the polyprotein 1ab, the 3CLpro is designated as the main protease (67). This domain of the viral genome is essential for replication since it cleaves the HCoV-OC43 polyprotein 1ab at 11 sites and allows the release of important functional domains (Fig. 2) (32). Like other coronavirus 3CLpro, the HCoV-OC43 3CLpro acts via a catalytic dyad, which is composed of a His⁴¹ and a Cys¹⁴⁵ (6). The HCoV-OC43 3CLpro is 303 amino acids long and displays an outstanding conservation among coronaviruses from the three genetic groups (Fig. 5). Seventy-nine residues are strictly conserved among sequences from six different coronaviruses, displaying 26% identity among all 3CLpro sequences analysed, whereas group 2 coronaviruses display 46.2% identity for the same motif between each other.

DISCUSSION

HCoV-OC43 belongs to the second genetic group of coronaviruses and represents the human coronavirus that is most related to SARS-HCoV. Here, we present the first report of a complete sequence of the HCoV-OC43 genome, including the complete sequence of a clinical respiratory isolate of the OC43 serotype. The two genomes are 30,713 nt long and only differ by six nucleotides including two amino acid substitutions located in the S (I958F) and the N (V81A) protein genes. The genomes of the two virus variants display 71%, 53.1% and 51.2% identity with the genomes of MHV-A59, SARS-HCoV Tor2 and HCoV-229E, respectively. Using bioinformatics tools and well-characterized coronaviruses, further characterization of the HCoV-OC43 genome was performed and these analysis revealed that HCoV-OC43 is closely related to BCoV and MHV, and that it displays significant amino acid identity levels with important functional domains of the SARS-HCoV. Like the ATCC strain that was isolated in the 1960s, HCoV-OC43 Paris, isolated in 2001, exhibited neuroinvasive properties in BALB/c mice. Although mice were more easily infected with the ATCC strain than with the Paris isolate, these results suggest that both viruses possess the intrinsic ability to infect neural cells and to reach the CNS from the periphery.

Recently, Vijgen and collaborators have submitted a complete sequence of the HCoV-OC43 genome to GenBank (NC_005147). The virus strain used for this sequencing is described as corresponding to the virus strain that was used in our laboratory (VR-759). However, comparison of our sequence with theirs shows that they differed at 33 positions, 29 mutations being located in the S gene, including two mutations in the S2 subunit. Among the four other differences, one is located at the beginning of the genome sequence, where a guanine is added with regards to our sequence, whereas the other three are scattered throughout ORF1a. Despite these differences, the availability of the complete genome sequence from a clinical isolate reinforces the validity of our sequence, since the HCoV-OC43 ATCC and Paris sequences only

differ by six nucleotides. Therefore, this observation suggests that the viral strain used by Vijgen and collaborators could have been adapted in cell culture, given the differences observed in the S gene, which is known to be associated with viral adaptation (27). No differences were noticed among ORF1b sequences between HCoV-OC43 ATCC, Paris and the one from Vijgen and collaborators. This observation suggests that this region of the genome needs a high rate of conservation in order to remain functional and that genes located downstream of the replicase gene are more permissive to sequence modifications.

Using a recent HCoV-OC43 clinical respiratory isolate, it is shown in the present study that HCoV-OC43 apparently remains genetically stable in the environment. Indeed, despite virus shedding and chances of persistence in the host, the HCoV-OC43 Paris isolate displays differences at only six positions with regard to the ATCC strain sequence, despite about 40 years have elapsed between the two isolations. Knowing that viral persistence could be associated with molecular adaptation (7, 8), the low rate of mutation observed here could be explained by the fact that the HCoV-OC43 Paris isolate has never or rarely persisted before. However, it is delicate to speculate about such an issue given that the exact origin of the virus before its isolation remains undetermined. It is also worth noting that viral persistence does not necessarily require an adaptation to the environment (2) and that despite the high rate of mutation of the coronavirus RdRp (1), 95% of the mutations engendered by RNA virus polymerases are deleterious and therefore not conserved (42).

Our observation that inhalation of HCoV-OC43 led to a generalized infection of the whole CNS in mice demonstrates neuroinvasiveness. This result confirms that human coronaviruses have neuroinvasive properties in mice, which was first shown in newborn mice (10, 22), and which is consistent with their detection in human brain (9, 12, 40, 53). After inhalation, the first infected cells were detected in the OB, illustrating that virus directly reached the brain by a transneuronal route, as already demonstrated for MHV (10, 31, 47). The HCoV-OC43 Paris

isolate, which was never propagated in mouse brain or other neurological tissue, also exhibited neuroinvasive properties in mice. Replication within the CNS was similar for the two variants, but fewer mice were infected by the HCoV-OC43 Paris isolate than by the ATCC strain. This data suggest that only one mutation in the S gene, giving rise to one amino acid modification, could partially modulate the neuroinvasiveness of one variant over the other. Indeed, a single amino acid change has already been demonstrated to influence MHV ability to spread within the CNS (43, 59).

Although the degree of sequence conservation between the genomes of the HCoV-OC43 ATCC and Paris variants is very high, their phenotype seems to differ slightly in mice, considering that the ATCC strain reached the CNS more easily. As we have demonstrated *in vitro*, using primary hippocampus and cortical cell cultures, both HCoV-OC43 ATCC and Paris variants were able to replicate in rodent neurons although the HCoV-OC43 ATCC strain yielded more infectious virus particles than the HCoV-OC43 Paris isolate. However, the two viral variants exhibited different biological properties, such as plaque formation and cytopathic effects on different cell lines (H. Jacomy and P. J. Talbot, unpublished data).

Although both mutations preserve some but not all properties of the parental residues, the I958F mutation leads to a substitute phenylalanine that does not display the same steric hindrance than the isoleucine, which could potentially affect protein folding and function. Moreover, the I958F mutation is located in the S2 subunit of the S gene and would probably be positioned in the putative fusion peptide domain (23), conferring a lot of impact to this mutation at the biological level. On its own, this mutation could therefore have the capacity to influence the phenotype of the HCoV-OC43 Paris isolate because it may interfere with the fusion process in a positive or a negative manner (43). Given the known involvement of the S protein in viral biology and pathogenesis (7, 8, 15, 48), this mutation is more likely to influence the phenotype of the Paris isolate. It has been reported that the N protein may be involved in viral RNA synthesis (30) and

that it could colocalize with nucleolar antigens and delay the cell cycle (14). However, the fact that the V81A mutation within the N gene is positioned in domain I of the protein should not influence the RNA binding properties of N, as this functional feature of the protein lies in domain II (45). Therefore, even though the role of both mutations needs to be investigated, we feel that the S mutation is more likely to influence the virus phenotype.

Comparison with better characterized coronaviruses (23, 59), suggests that the I958F mutation is located in the putative S fusion peptide and could therefore affect viral fusogenic properties and phenotype. Although no fusion peptides have formally been identified in any coronavirus S protein, predictions have located such fusion sequences near the N terminus of the heptad repeat 1 (HR1) for MHV (33). Studies with the MHV-A59 S protein also showed that mutations introduced in the HR1 region severely affected cell-cell fusion ability (33). Moreover, it has already been reported that a single mutation introduced in HR1 could influence the degree of virulence of MHV (59). Depending on the effect of the mutation on cleavage ability, the phenotype of the resulting virus could also be affected. Although the cleavage of the S protein is not absolutely required for fusion (23, 52, 55), it has been shown to enhance fusogenicity (55). Thus, inhibition of S protein cleavage would be associated with a more stable interaction between S1 and S2 and would correlate with a loss of fusogenicity (25). So, as it was observed by Tsai and collaborators (59) for the MHV-JHM strain, the I958F mutation in the S gene of the HCoV-OC43 Paris isolate could either alter the conformation of the S protein or have an incidence on its cleavage, impairing the ability of the virus to spread within the CNS.

An animal model for the HCoV-OC43 ATCC strain has recently been developed and optimized in our laboratory (22). Moreover, HCoV-OC43 may also be used as a model for the study of SARS-HCoV, not only because of the identity level the two virus strains display, but because HCoV-OC43 can also be studied without the requirement of a level three, aerosol-aware, biological confinement. Indeed, we have now demonstrated that the two virus strains present a

high level of conservation for some essential functional domains, especially within the 3CLpro, the RdRp and the RNA helicase. This result is consistent with the possible sharing of several important properties by these two viruses. All of these motifs represent potential candidates for therapy of coronavirus-mediated diseases because they are specific targets and because of the specificity they exhibit towards their substrate. Indeed, substrate specificities of all coronavirus proteases, and mainly 3CLpro, are conserved among the three established groups (19), and this is also true for SARS-HCoV. The picornavirus RdRp (21) and viral proteases (17) have notably been designated as such targets for anti-viral therapy. At present, the SARS-HCoV 3CLpro enzyme represents the most promising target for SARS therapy (58). The availability of 3CLpro crystal structures should provide a valuable tool for rapid identification of potential drugs against SARS. So far, 3CLpro crystal structures have been obtained for TGEV (5), HCoV-229E (6) and more recently for SARS-HCoV (61). A putative *in vitro* inhibitor has also been identified for TGEV (5) and SARS-HCoV (61). This inhibitor, the hexapeptidyl chloromethyl ketone (CMK), was shown to bind the 3CLpro enzyme very efficiently *in vitro* and although it provides an excellent structural basis for drug design, *in vivo* experiments need to be performed on this issue.

Now that the complete genome sequence of HCoV-OC43 has been deciphered, it will provide a very useful tool for the study of coronaviruses from all genetic groups and particularly for those of group 2, including SARS-HCoV. Indeed, the genome sequence will allow comparative studies with other coronavirus strains and RNA viruses and will also allow optimization of prediction models. This sequence will also allow the assembly of an infectious cDNA clone of HCoV-OC43, which is currently underway. So far, cDNA clones have been assembled for several coronavirus strains using different approaches. Among these clones, those of TGEV (3, 64), HCoV-229E (57), IBV (13), MHV-A59 (66) and even SARS-HCoV (65) are now available. The HCoV-OC43 clone will provide an invaluable tool to further understand the underlying mechanisms for replication and pathogenesis of human coronaviruses.

ACKNOWLEDGMENTS

We are grateful to Francine Lambert for her technical assistance with cell and virus culture. We also thank Dr D. W. Yoo, who kindly provided the ORF1ab sequence from the BCoV, which allowed us to sequence the HCoV-OC43 ORF1ab.

This work was supported by Grant No. MT-9203 from Canadian Institutes for Health Research (Institute of Infection and Immunity). Pierre J. Talbot is the holder of a Tier-I (Senior) Canada Research Chair. Julien St-Jean acknowledges a studentship from *Fonds québécois de recherche sur la nature et les technologies*.

REFERENCES

1. **Adami, C., J. Pooley, J. Glomb, E. Stecker, F. Fazal, J. O. Fleming, and S. C. Baker.** 1995. Evolution of mouse hepatitis virus (MHV) during chronic infection: quasispecies nature of the persisting MHV RNA. *Virology* **209**:337-346.
2. **Addie, D. D., I. A. T. Schaap, L. Nicolson, and O. Jarrett.** 2003. Persistence and transmission of natural type I feline coronavirus infection. *J. Gen. Virol.* **84**:2735-2744.
3. **Almazán, F., J. M. González, Z. Péntzes, A. Izeta, E. Calvo, J. Plana-Durán, and L. Enjuanes.** 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. USA* **97**:5516-5521.
4. **Alonso, S., A. Izeta, I. Sola, and L. Enjuanes.** 2002. Transcription regulatory sequences and mRNA expression levels in the coronavirus transmissible gastroenteritis virus. *J. Virol.* **76**:1293-1308.
5. **Anand, K., G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld.** 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain. *EMBO J.* **21**:3213-3224.
6. **Anand, K., J. Ziebuhr, P. Wadhvani, J. R. Mesters, and R. Hilgenfeld.** 2003. Coronavirus main proteinase (3CL^{pro}) structure: Basis for design of anti-SARS drugs. *Science* **300**:1763-1767.
7. **Arbour, N., G. Côté, C. Lachance, M. Tardieu, N. R. Cashman, and P. J. Talbot.** 1999. Acute and persistent infection of human neural cell lines by human coronavirus OC43. *J. Virol.* **73**:3338-3350.

8. **Arbour, N., S. Ekandé, G. Côté, C. Lachance, F. Chagnon, M. Tardieu, N. R. Cashman, and P. J. Talbot.** 1999. Persistent infection of human oligodendrocytic and neuroglial cell lines by human coronavirus 229E. *J. Virol.* **73**:3326-3337.
9. **Arbour, N., R. Day, J. Newcombe, and P. J. Talbot.** 2000. Neuroinvasion by human respiratory coronaviruses. *J. Virol.* **74**:8913-8921.
10. **Barthold, S. W., M. S. de Souza, and A. L. Smith.** 1990. Susceptibility of laboratory mice to intranasal and contact infection with coronaviruses of other species. *Lab. Anim. Sci.* **40**:481-485.
11. **Bonavia, A., N. Arbour, V. W. Yong, and P. J. Talbot.** 1997. Infection of primary cultures of human neural cells by human coronaviruses 229E and OC43. *J. Virol.* **71**:800-806.
12. **Burks, J. S., B. L. DeVald, L. D. Jankovsky, and J. C. Gerdes.** 1980. Two coronaviruses isolated from central nervous system tissue of multiple sclerosis patients. *Science* **209**:933-934.
13. **Casais, R., V. Thiel, S. G. Siddell, D. Cavanagh, and P. Britton.** 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J. Virol.* **75**:12359-12369.
14. **Chen, H., W. Torsten, P. Britton, G. Brooks, and J. A. Hiscox.** 2002. Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell. *J. Virol.* **76**:5233-5250.
15. **Chouljenko, V. N., X. Q. Lin, J. Storz, K. Kousoulas, and A. E. Gorbalenya.** 2001. Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. *J. Gen. Virol.* **82**:2927-2933.

16. **Denison, M. R., W. J. Spaan, Y. van der Meer, C. A. Gibson, A. C. Sims, B. Prentice, and X. T. Lu.** 1999. The putative helicase of the coronavirus mouse hepatitis virus is processed from the replicase gene polyprotein and localizes in complexes that are active in viral RNA synthesis. *J. Virol.* **73**:6862-6871.
17. **Dragovich, P. S., T. J. Prins, R. Zhou, T. O. Johnson, Y. Hua, H. T. Luu, S. K. Sakata, E. L. Brown, F. C. Maldonado, T. Tuntland, C. A. Lee, S. A. Fuhrman, L. S. Zalman, A. K. Patick, D. A. Matthews, E. Y. Wu, M. Guo, B. C. Borer, N. K. Nayyar, T. Moran, L. Chen, P. A. Rejto, P. W. Rose, M. C. Guzman, E. Z. Dovalsantos, S. Lee, K. McGee, M. Mohajeri, A. Liese, J. Tao, M. B. Kosa, B. Liu, M. R. Batugo, J. P. Gleeson, Z. P. Wu, J. Liu, J. W. Meador, 3rd, and R. A. Ferre.** 2003. Structure-based design, synthesis and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 8. Pharmacological optimization of orally bioavailable 2-pyridone-containing peptidomimetics. *J. Med. Chem.* **46**:4572-4585.
18. **Gorbalenya, A. E.** 2001. Big nidovirus genome. When count and order of domains matter. *Adv. Exp. Med. Biol.* **494**:1-17.
19. **Hegy, A., and J. Ziebuhr.** 2002. Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **83**:595-599.
20. **Horvath, C. M., M. A. Williams, and R. A. Lamb.** 1990. Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. *EMBO J.* **9**:2639-2647.
21. **Hung, M., C. S. Gibbs, and M. Tsiang.** 2002. Biochemical characterization of rhinovirus RNA-dependant RNA polymerase. *Antiviral Res.* **56**:99-114.

22. **Jacomy, H., and P. J. Talbot.** 2003. Vacuolating encephalitis in mice infected by human coronavirus OC43. *Virology* **315**:20-33.
23. **Jan Bosch, B., R. van der Zee, C. A. M. de Haan, and P. J. M. Rottier.** 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **77**:8801-8811.
24. **Kamahora, T., L. H. Soe, and M. M. C. Lai.** 1989. Sequence analysis of nucleocapsid gene and leader RNA of human coronavirus OC43. *Virus Res.* **12**:1-9.
25. **Krueger, D. K., S. M. Kelly, D. N. Lewicki, R. Ruffolo, and T. M. Gallagher.** 2001. Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *J. Virol.* **75**:2792-2802.
26. **Kündel, F., and G. Herrler.** 1993. Structural and functional analysis of the surface protein of human coronavirus OC43. *Virology* **195**:195-202.
27. **Kündel, F., and G. Herrler.** 1996 Structural and functional analysis of the S proteins of two human coronavirus OC43 strains adapted to growth in different cells. *Arch. Virol.* **141**:1123-1131.
28. **Kuo, L., G.-J. Godeke, M. J. B. Raamsman, P. S. Masters, and P. J. M. Rottier.** 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J. Virol.* **73**:1393-1406.
29. **Labonté, P., S. Mounir, and P. J. Talbot.** 1995. Sequence and expression of the ns2 protein gene of human coronavirus OC43. *J. Gen. Virol.* **76**:431-435.
30. **Lai, M. M. C., and D. Cavanagh.** 1997. The molecular biology of coronaviruses. *Adv. Virus Res.* **48**:1-97.

31. **Lavi, E., P. S. Fishman, M. K. Highkin, and S. R. Weiss.** 1988. Limbic encephalitis after inhalation of a murine coronavirus. *Lab. Invest.* **58**:31-36.
32. **Lu, Y., X. Lu, and M. R. Denison.** 1995. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *J. Virol.* **69**:3554-3559.
33. **Luo, Z., and S. R. Weiss.** 1998. Roles in cell-to-cell fusion of two conserved hydrophobic regions in the murine coronavirus spike protein. *Virology* **244**:483-494.
34. **Marra, M. A., S. J. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, Y. S. Butterfield, J. Khattra, J. K. Asano, S. A. Barber, S. Y. Chan, A. Cloutier, S. M. Coughlin, D. Freeman, N. Girn, O. L. Griffith, S. R. Leach, M. Mayo, H. McDonald, S. B. Montgomery, P. K. Pandoh, A. S. Petrescu, A. G. Robertson, J. E. Schein, A. Siddiqui, D. E. Smailus, J. M. Stott, G. S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T. F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G. A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R. C. Brunham, M. Kraiden, M. Petric, D. M. Skowronski, C. Upton, and R. L. Roper.** 2003. The genome sequence of the SARS-associated coronavirus. *Science* **300**:1399-1404.
35. **Masters, P. S.** 1999. Reverse genetics of the largest RNA viruses. *Adv. Virus Res.* **53**:245-264.
36. **Meyers, G.** 2003. Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependant reinitiation mechanism. *J. Biol. Chem.* **278**:34051-34060.
37. **Mounir, S., and P. J. Talbot.** 1992. Sequence analysis of the membrane protein gene of human coronavirus OC43 and evidence for *O*-glycosylation. *J. Gen. Virol.* **73**:2731-2736.

38. **Mounir, S., and P. J. Talbot.** 1993. Molecular characterization of the S protein gene of human coronavirus OC43. *J. Gen. Virol.* **74**:1981-1987.
39. **Mounir, S., and P. J. Talbot.** 1993. Human coronavirus OC43 RNA 4 lacks two open reading frames located downstream of the S gene of bovine coronavirus. *Virology* **192**:355-360.
40. **Murray, R. S., B. Brown, D. Brian, and G. F. Cabirac.** 1992. Detection of coronavirus RNA and antigen in multiple sclerosis brain. *Ann. Neurol.* **31**:525-533.
41. **Myint, S. H.** 1994. Human coronaviruses : a brief review. *Rev. Med. Virol.* **4**:35-46.
42. **Novella, I., S.** 2003. Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. *Curr. Opin. Microbiol.* **6**:399-405.
43. **Ontiveros, E., T. S. Kim, T. M. Gallagher, and S. Perlman.** 2003. Enhanced virulence mediated by the murine coronavirus, mouse hepatitis virus strain JHM, is associated with a glycine at residue 310 of the spike glycoprotein. *J. Virol.* **77**:10260-10269.
44. **Ozdarendeli, A., S. Ku, S. Rochat, G. D. Williams, S. D. Senanayake, and D. A. Brian.** 2001. Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. *J. Virol.* **75**:7362-7374.
45. **Parker, M. M., and P. S. Masters.** 1990. Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein. *Virology* **179**:463-468.
46. **Peiris, J. S., C. M. Chu, V. C. Cheng, K. S. Chan, I. F. Hung, L. L. Poon, K. I. Law, B. S. Tang, T. Y. Hon, C. S. Chan, K. H. Chan, J. S. Ng, B. J. Zheng, W. L. Ng, R. W. Lai, Y.**

- Guan, and K. Y. Yuen.** 2003. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**:1767-1772.
47. **Perlman, S., G. Evans, and A. Afifi.** 1990. Effect of olfactory bulb ablation on spread of a neurotropic coronavirus into the mouse brain. *J. Exp. Med.* **172**:1127-1132.
48. **Rowe, C. L., S. C. Baker, M. J. Nathan, and J. O. Fleming.** 1997. Evolution of mouse hepatitis virus : detection and characterization of spike deletion variants during persistent infection. *J. Virol.* **71**:2959-2969.
49. **Sawicki, S. G., and D. L. Sawicki.** 1998. A new model for coronavirus transcription. *Adv. Exp. Med. Biol.* **440**:215-219.
50. **Schickli, J. H., B. D. Zelus, D. E. Wentworth, S. G. Sawicki, and K. V. Holmes.** 1997. The murine coronavirus mouse hepatitis virus strain A59 from persistently infected murine cells exhibits an extended host range. *J. Virol.* **71**:9499-9507.
51. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. M. Spaan, and A. E. Gorbalenya.** 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **331**:991-1004.
52. **Stauber, R., M. Pfeleiderera, and S. Siddell.** 1993. Proteolytic cleavage of the murine coronavirus surface glycoprotein is not required for fusion activity. *J. Gen. Virol.* **74**:183-191.
53. **Stewart, J. N., S. Mounir, and P. J. Talbot.** 1992. Human coronavirus gene expression in the brains of multiple sclerosis patients. *Virology* **191**:502-505.

54. **Sturman, R. M., C. S. Ricard, and K. V. Holmes.** 1990. Conformational change of the coronavirus peplomer glycoprotein at pH 8.0 and 37°C correlates with virus aggregation and virus-induced cell fusion. *J. Virol.* **64**:3042-3050.
55. **Taguchi, F.** 1993. Fusion formation by the uncleaved spike protein of murine coronavirus JHMV variant cl-2. *J. Virol.* **67**:1195-1202.
56. **Taguchi, F.** 1995. The S2 subunit of the murine coronavirus spike protein is not involved in receptor binding. *J. Virol.* **69**:7260-7263.
57. **Thiel, V., J. Herold, B. Schelle, and S. G. Siddell.** 2001. Infectious RNA transcribed *in vitro* from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.* **82**:1273-1281.
58. **Thiel, V., K. A. Ivanov, Á. Putics, T. Hertzog, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, and J. Ziebuhr.** 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**:2305-2315.
59. **Tsai J. C., L. de Groot, J. D. Pinon, K. T. Iacono, J. J. Phillips, S. H. Seo, E. Lavi, and S. R. Weiss.** 2003. Amino acid substitutions within the heptad repeat domain 1 of murine coronavirus spike protein restrict viral antigen spread in the central nervous system. *Virology* **312**:369-380.
60. **Tsang, K. W., P. L. Ho, G. C. Ooi, W. K. Yee, T. Wang, M. Chan-Yeung, W. K. Lam, W. H. Seto, L. Y. Yam, T. M. Cheung, P. C. Wong, B. Lam, M. S. Ip, J. Chan, K. Y. Yuen, and K. N. Lai.** 2003. A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N. Eng. J. Med.* **348**:1977-1985.

61. **Yang, H., M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. Rao.** 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci.* **100**:13190-13195.
62. **Yoo, D. W., M. D. Parker, and L. A. Babiuk.** 1991. The S2 subunit of the spike glycoprotein of bovine coronavirus mediates membrane fusion in insect cells. *Virology* **180**:291-304.
63. **Yoo, D., and Y. Pei.** 2001. Full-length genomic sequence of bovine coronavirus (31 kb). Completion of the open reading frame 1a/1b sequences. *Adv. Exp. Med. Biol.* **494**:73-76.
64. **Yount, B., K. M. Curtis, and R. S. Baric.** 2000. Strategy for systematic assembly of large RNA and DNA genomes : transmissible gastroenteritis virus model. *J. Virol.* **74**:10600-10611.
65. **Yount, B., K. M. Curtis, E. A. Fritz, L. E. Hensley, P. B. Jahrling, E. Prentice, M. R. Denison, T. W. Geisbert, and R. S. Baric.** 2003. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci.* **100**:12995-13000.
66. **Yount, B., M. R. Denison, S. R. Weiss, and R. S. Baric.** 2002. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J. Virol.* **76**:11065-11078.
67. **Ziebuhr, J., E. J. Snijder, and A. E. Gorbalenya.** 2000. Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J. Gen. Virol.* **81**:853-879.

TABLE 1 : Primers used for amplifications of the HCoV-OC43 genome

Primers combination and location	Target region	Amplicon length
JUB3-JUB12 1-20 and 6071-6091	Leader, UTR5' and ORF1a	6091 bp
JUB5-JUB6 5319-5339 and 11111-11131	ORF1a	5813 bp
JUB7-JUB8 10901-10921 and 16525-16545	ORF1a and ORF1b	5645 bp
JUB9-JUB10 16309-16329 and 21544-21564	ORF1b and ns2	5256 bp
JUNSO1-JUSO2 21330-21350 and 27754-27774	ORF1b, ns2, HE and S genes	6445 bp
JUMO1-GeneRacer™ 3' nested 27649-27669 and 30742-30764	S, ns12.9, E, M, N genes and UTR3'	3116 bp *
GeneRacer™ oligo dT	5'-GCTGTCAACGATACGCTACGTAACGGCATGACAGTG(T) ₁₈ -3'	
GeneRacer™ 5' nested	5'-GGACTGACATGGACTGAAGGAGTA-3'	
GeneRacer™ 3' nested	5'-CGCTACGTAACGGCATGACAGTG-3'	

* Assuming a poly(A) tail of 28 bp.

TABLE 2 : Primers used for sequencing of the HCoV-OC43 genome

Positive strand :			Negative strand :		
Primer	Séquence (5'→ 3')	Location	Primer	Séquence (5'→ 3')	Location
JUB3	GATTGTGAGCGATTTCGCTGC	1-20	JUB12	ACATCACCTGTAGCTGTGGC	6071-6091
RACEJUB2	TGTGATGGTGGATTGTCGCGG	380-400	JUB11	CCAGTAACGCTCTGTAACCTTC	5750-5770
JUB31	GTTATATATGATTGATCCTGC	662-682	JUB111	CCATGCCTCTTGCCATTGAAC	5277-5297
JUB32	TGATTATACTGGTAGTCTTGC	983-1003	JUB112	TGAATTTAACACCATCAACAG	4918-4938
JUB33	GCACAATCTCAGGTGTTTTG	1488-1508	JUB113	ATCCTCTTGATTATTGCTAAC	4458-4478
JUB14	AGTTGCTAGGTGTCTCAGATG	1801-1821	JUB115	AGGGTTTACAACAACCTCTGC	4086-4106
JUB141	CTTATATAGTAGTGAGAGATG	2254-2274	JUB15	AACAGCCATAGAATGACTATC	3876-3896
JUB142	GTTCTGATTTTTATTAGCGG	2557-2577	JUB151	GCTTTAGGCACATACAGACCC	3395-3415
JUB143	ACAGCTCCTGAAGATTGATGAC	3093-3113	JUB152	ACCACAGCATAAAAATTCCTCC	2915-2935
JUB144	TTAACCTTATGTGATTGGCAG	3621-3641	JUB153	CAAACTCTTCTACGTCCTAATC	2399-2419
JUB161	ATGCTATGTTCTTTTATGGTG	3700-3720	JUB13	AATGCCTTGACCACTTACTGCC	1991-2011
JUB163	TATTATTGGGCATGGTATGTC	3959-3979	JUB131	CAAGCAAACCATCTATCATG	1427-1447
JUB164	AAGTTATGTATTGTTAGAGCG	4310-4330	JUB132	CTCCAAGTAGGAAATAATGCC	1049-1069
JUB165	TATTTGAGTGTACTGGAGGC	4782-4802	RACE JUB1	GGAGCAAATCATATCCACCTC	312-332
JUB166	TGCTTGCTATTACAACATGC	5128-5148			
JUB5	CCTGCTAGATTTGTAICQTIG	5319-5339	JUB6	CCTCTAAATGTCTGCTGTGAC	11111-11131
JUB51	ACTCAGCGTATTATTTAAAGCC	5913-5933	JUB61	ATAGCAGCCAACAGTGTITCC	10724-10744
JUB52	GTTGACGATGGTGGTGCAGAC	6264-6284	JUB62	AGGGTCACTGTAAGAACCAAGC	10205-10225
JUB53	AATCTACACACAGAAATTCG	6701-6721	JUB63	CTATTTAAAGCAACATCAGAC	9752-9772
JUB54	TGGCAGGATTTGATATTGTTAG	7036-7056	JUB64	AGAAAGTATGGGGTAACTTG	9453-9473
JUB17	GGTTTTA ^C CCATTGTTGCTC	7177-7197	JUB18	ATCTCTACCACAAAAGTCTCC	9213-9233
JUB171	GCTTGTCTATGATCTGTGATG	7678-7698	JUB181	TGGTAGGGACATTTAAACACTG	8761-8781
JUB172	CTGTGCTCGTAAAAGTTGTTCC	8072-8092	JUB182	GCTTTCTCAATTTATGCTGG	8345-8365
JUB173	AATAAGCAGATGGCTAATGTC	8403-8423	JUB183	CATAGACAAAAATGTTCCAC	7950-7970
JUB55	AAGGTTTTATCCGCTCTCCAG	9043-9063	JUB184	AAGTATTACCTGGTTATAAAG	7543-7563
JUB56	CACITACAATGGCTAGTTATG	9540-9560	JUB65	ACCTAACACTCCAATGTAATG	7252-7272
JUB57	CCGCTCAACTTCACTTCTGTC	9925-9945	JUB66	CTAAAAGTGTCTTAATCCAC	6926-6946
JUB58	CTTGTGGATCTGTTGGTTATG	10378-10398	JUB67	TATACTAACAGAGTCAATCC	6531-6551
			JUB68	ACATCACCTGTAGCTGTGGC	6071-6091
JUB7	GGCTTCTACATTTTTGTTTAG	10901-10921	JUB8	TAACCTGCTTAAAGGCTTCC	16525-16545
JUB71	TGTATTTCCAGATATACCTCC	11443-11463	JUB81	GATGTTTGAAGAAGCAGACAG	16117-16137
JUB72	CAGCAGTTAAACAGACTAGAG	12084-12104	JUB82	CATAATATCTGTGACAAAGG	15522-15542
JUB73	ATAATGAGGATCTGCTACTG	12547-12567	JUB83	TATAACTACAGGAACACCAGC	15050-15070
JUB74	TGTTAAACCCGATGCTACCAC	13073-13093	JUB84	AAAAGTGCCTCAGATCAACTG	14604-14624
JUB75	AGAGAGATGAAATGCTATGAG	13559-13579	JUB85	ATACTCCAGTGCTTAAAAATAC	14158-14178
JUB76	TTGGAATGGCAATTTGATGTG	14066-14086	JUB86	TCCTTGCCTACAATGTGTGGC	13648-13668
JUB77	TTGTATGATTTACGCACTTGC	14465-14485	JUB87	TCCACAAACTIGACAAAACATC	13245-13265
JUB78	CATTATCATTGAGGACGAGC	14856-14876	JUB88	TCCTGTAGTGTGTTACAACC	12843-12863
JUB79	GAGGCATGTTGTTGCGAAAGC	15230-15250	JUB89	CTTGGATAGTTTGAATCTGCC	12448-12468
JUB710	ATATAAGTGCCTTTCAACAGG	15636-15656	JUB810	ACTAGAACCAGCTCATCAAG	12048-12068
JUB711	CAAAAGTTTACTGATGATGCC	16040-16060	JUB811	CTTAAAATTAAGCATAAAGGGC	11649-11669
			JUB812	AQAACAAATCATGGTTAATGC	11245-11265
JUB9	TTATTGTGAAGAICATAAGCC	16309-16329	JUB10	AAATGGGTAAGTGGAAAATTG	21544-21564
JUB91	TCAACACATTTGGTATGAAACG	16903-16923	JUB101	AGAGCTAACTTGTCTCGAATC	21064-21084
JUB92	AGTTTCATTTGTTTAAAGGTC	17498-17518	JUB102	GCAAACACTCTTACTACCACC	20414-20434
JUB93	TATTGGTATTTCTGCTGTTAC	18034-18054	JUB103	GCGGTGGACCTCTTATCATCG	19908-19928
JUB94	TAGAAGTGGTACTATGGTTG	18562-18582	JUB104	GTATTGTAAGACTCTAAGTAC	19369-19389
JUB95	TTTGAAGGCACATAAGGACTC	18997-19017	JUB105	TACAAAAGAGTCTTAAACAGAC	18973-18993
JUB96	AGTCAAGACTGGTCAATTATAC	19498-19518	JUB106	CCCATGTAACATAGCACAAAC	18444-18464
JUB97	CGTTCTAATAATGGCGTTTAC	19859-19879	JUB107	TTATATTTGTCATCTACTGCC	17983-18003
JUB98	TAGGCTTGTACCGAAGACAGC	20313-20333	JUB108	TTATGCCACAAAGGGTTAGCC	17599-17619
JUB99	ATACTCAGTATGTCAATATC	20736-20756	JUB109	GGTAGTAAACACATACTTACG	17153-17173
JUB100	TCGAGACAAGTTAGCCTGCGG	21067-21087	JUB110	CTGCGGAACAAGCGTAGGAGC	16808-16828
JUNSO3	TTAATGATATGGTTTATTC	21396-21416	JUSO2	AAGAACITTTAACAATGCTAG	27754-27774
JUNSO6	GTGTAGAAGAATTGCATGACG	21807-21827	JUSO4	ATCAAGTGACAAAATCTGGTGC	27384-27404
JUHEO1	AACAATCTTGGTCTTCC	22226-22246	JUSO6	ATATGATTACCATTACCACAG	27020-27040
JUHEO3	TTTAGGAGTTTTCACTTTACC	22588-22608	JUSO8	TGAATAGCATAAAGGGCATTG	26678-26698
JUHEO5	ACCCTTTGTATTTTAAACGG	22959-22979	JUSO10	GCTGCCTAGACAACTAATAC	26307-26327
JUHEO7	TCATGGAGATGCTGGTTTTAC	23337-23357	JUSO12	AATGGCTCAAAAATTAGTAAAC	25937-25957
JUSO24	ATTAATA ^A CCCTGATTTACCC	23482-23502	JUSO14	TTATAATAAGTCGCATTAAAC	25553-25573
JUSO5	AATGTATAGTAGAGTTCCCTGC	23993-24013	JUSO16	ACAAGTTAAATAATTAGTACC	25188-25208
JUSO7	CTTTCACACTATTATGTCATG	24378-24398	JUSO18	ATAGTTATGCTGGAAAAACAC	24809-24829
JUSO9	TATTCAGGCAGACTCATTTAC	24752-24772	JUSO20	TAGAAGTGAGAGGTGTAACCC	24436-24456
JUSO11	AATTGAATGGTTCGTGTGATG	25123-25143	JUSO22	TATAGGATGATTTACAAAAG	24031-24051
JUSO13	GTGTTGTGTTAATATGACC	25498-25518	JUHEO2	ACATAATAAGTACCCAAAACC	23778-23798
JUSO15	TAGGTAGTGGTTACTGTGTGG	25867-25887	JUHEO4	CATTATCATACCTAAAACCGC	23408-23428
JUSO17	GAATGGTGTACTCTTAGCAC	26234-26254	JUHEO6	AGACCATAAATAACACCAGTG	23037-23057
JUSO19	TGGATGTGCTAAGTCAAAAATC	26632-26652	JUHEO8	ATGATAAGCGCTAAAATTAAC	22660-22680
JUSO21	ATTTCTGTGGTAATGGTAAATC	27016-27036	JUNSO2	GAAACAACATTTGGTAGGAGGG	22416-22436
JUSO25	TGGACCAGATTTGTCACTTG	27382-27402	JUNSO4	TTCTTAAATGGACAGTGCTGC	21917-21937
JUMO1	GTGATGATTATCTGGATACC	27649-27669	JUO2	GCAGCAAGACATCCATTCTG	30495-30514
JUMO3	TAGTTGCCATTTGGTTTATTGG	28169-28189	JUO4	TACCAAACACTGCTGAACAG	29920-29940
JUMO5	ATGTTGGATTTGTTATTTGGT	28662-28682	JUO6	ATACCATCTGGCAGCAGTTTG	29430-29450
JUO8	ATGCTTTTTACTCTGGTAAGC	29079-29100	JUMO2	CCACTTGAGGATGCCATTACC	29136-29156
JUO3	TCTACTGGTCCGCTAGTAAAC	29512-29532	JUMO4	TACACAATCCACATAATAATG	28655-28675
JUO5	CCCACAGTCCCAATCTTCTGC	29999-30019	JUMO6	TATAAAAATTTTGGCCCCAC	28150-28170
JUO7	CTCTCTATCAGAATGGATGTC	30487-30507			

Underlined nucleotides indicate mismatched bases with regard to the genome sequence of the HCoV-OC43 ATCC strain.

TABLE 3 : Organization of the HCoV-OC43 genome

Genome region	Location	TRS location	TRS sequence
Leader and UTR 5'	nt 1-209	nt 63-69	UCUAAAC ... 139 nt... <u>AUG</u>
ORF1a	nt 210-13361		
ORF1b ¹	nt 13361-21496		
Intergenic region	nt 21497-21505		
ns2 gene	nt 21506-22342	nt 21492-21498	UCUAAACUUUAAAA <u>AUG</u>
Intergenic region	nt 22343-22353		
HE gene	nt 22354-23628	nt 22339-22344	UUAACUCAGUGAAA <u>AUG</u>
Intergenic region	nt 23629-23642		
S gene	nt 23643-27704	nt 23636-23642	UCUAAACA <u>AUG</u>
Intergenic region	nt 27705-27791		
ns12.9 gene	nt 27792-28121	nt 27771-27777	UCUUAAGGCCACGCCCUAUUA <u>AUG</u>
E gene ²	nt 28108-28362		
Intergenic region	nt 28363-28376		
M gene	nt 28377-29069	nt 28367-28373	UCCAAACA <u>UAUG</u>
Intergenic region	nt 29070-29078		
N gene	nt 29079-30425	nt 29065-29070	UCUAAAUUUUAAGGA <u>AUG</u>
UTR 3'	nt 30426-30713		
Poly(A) tail of 28 nt	nt 30714-30741		

Nucleotides in bold indicate TRS sequences whereas underlined nucleotides indicate initiation codon.

¹ putative ribosomal frameshift -1 between ORF1a and ORF1b.

² ORF overlap for the ns12.9 gene and the E gene.

TABLE 4 : Sequence differences between the HCoV-OC43 ATCC strain and the Paris isolate

Mutation location	Region of mutation	Consequence of mutation (ATCC → Paris)
nt 31	UTR 5'	C → T
nt 22243	ns2 gene	TTC (Phe 246) → TTT (Phe 246) (no aa change)
nt 26514	S gene	ATC (Ile 958) → TTC (Phe 958)
nt 28808	M gene	ACT (Thr 144) → ACC (Thr 144) (no aa change)
nt 29320	N gene	GTA (Val 81) → GCA (Ala 81)
nt 30632	UTR 3'	C → A

* reference strain: HCoV-OC43 ATCC.

FIGURE LEGENDS

Figure 1. Schematic representation of the HCoV-OC43 genome and of the amplification strategy used for sequencing. The HCoV-OC43 genome is 30,713 nt long and comprises nine main open reading frames (ORFs): ORF1a, ORF1b, ns2 (gene encoding the non structural protein 2), HE (hemagglutinin-esterase gene), S (spike gene), ns12.9 (gene encoding a non structural protein of 12.9 kDa), E (small envelope gene), M (membrane gene) and N (nucleocapsid gene). The replicase gene includes both ORF1a and ORF1b. The entire genome was amplified in six fragments in order to be sequenced. Each PCR product is named according to the name of the primers used for the amplification and location in the genome is indicated above or under each PCR product. White: gene encoding the replicase polyprotein; dotted: genes encoding non structural proteins; gray: genes encoding structural proteins; black: untranslated regions (UTRs). GR, GeneRacer™.

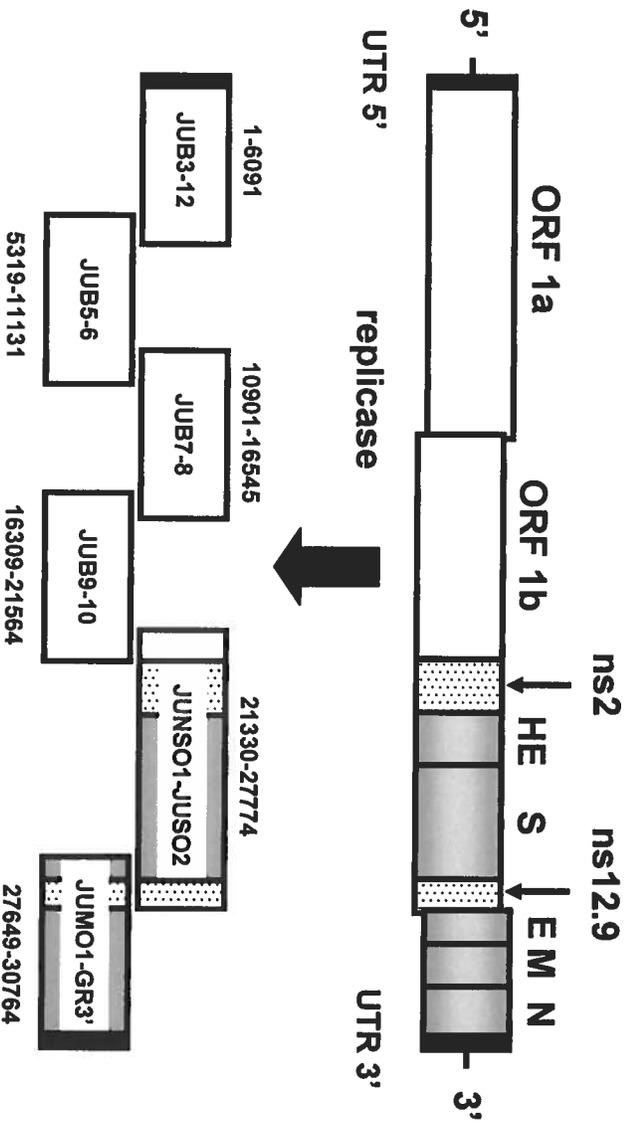
Figure 2. Schematic representation of the polyprotein 1ab putative proteolytic processing and of the main domains found in ORF1ab. The approximate positions of predicted functional domains and protease cleavage sites in ORF1ab are shown and amino acids positions are also indicated. The white arrows indicate putative cleavage sites recognized either by the PLP1 or the PLP2 whereas black arrows indicate sites recognized by the main protease, the 3CLpro. The 15 putative cleavage products generated by the proteolytic processing are named as follows: leader protein, MHV p65-like protein, nsp1 (PL1, X, PL2 and T1), T2, nsp2 (3CLpro), nsp3 (T3), nsp4, nsp5, nsp6, nsp7, nsp9 (RdRp), nsp10 (HEL), nsp11, nsp12 and nsp13 (15). A

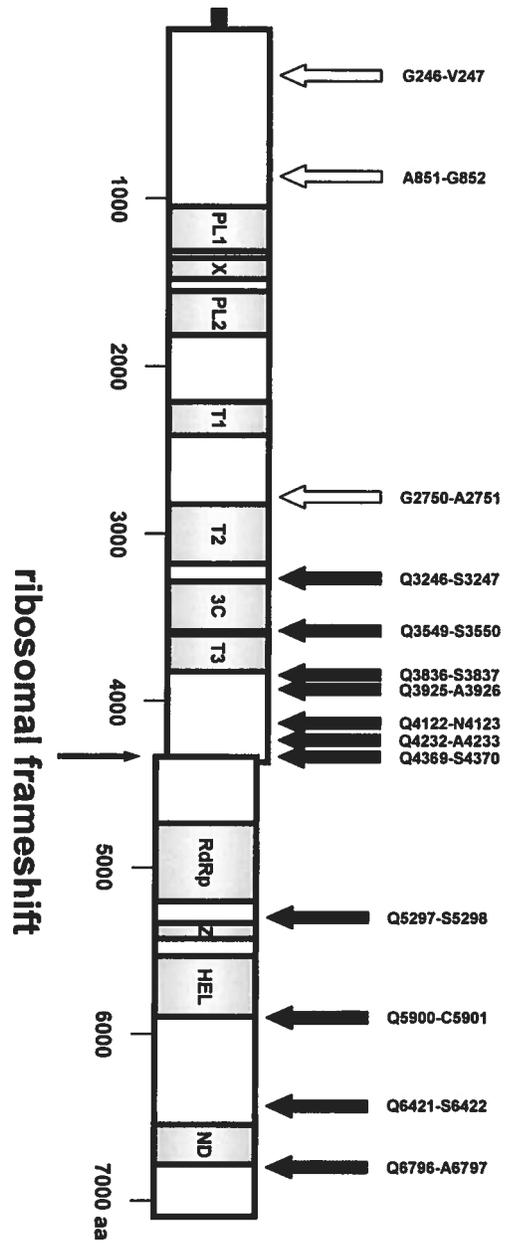
putative ribosomal frameshift -1 is presented between ORF1a and ORF1b. Upstream of the frameshift site, the slippery sequence $_{13334}\text{UUUAAAC}_{13340}$ is found. PL1 and PL2, accessory protease domains; X, conserved domain of unknown function; T1, T2 and T3, membrane-spanning (hydrophobic) domains; 3C, 3CLpro domain; RdRp, RNA-dependant RNA polymerase; Z, putative zinc finger; HEL, NTPase RNA helicase domain; ND, domain conserved exclusively in nidoviruses; nsp, non structural protein.

Figure 3. Neuroinvasive properties of HCoV-OC43 ATCC and HCoV-OC43 Paris variant in BALB/c mice after intranasal inoculation. (A) At 3 dpi, cells positive for viral antigens (arrows) were first observed in the olfactory bulb (OB). No infected cells could be detected in the cortex or other brain structures, illustrating transneuronal spreading of virus. (B) At 7 dpi, virus has disseminated to the entire CNS, illustrated by immunopositive cells throughout the brain. Original magnification of A and B: x 40. (C) Quantification of infectious virus in the brain of each individual mouse at different times post-infection. Virus titers are presented as logarithmic value of TCID₅₀ per gram of tissue (limit of detection was 100.5 TCID₅₀/g). Infection by HCoV-OC43 ATCC was detected in one mouse as early as 2 dpi, and gradually more mice became positive. HCoV-OC43 ATCC infectious particles were found between 2 to 8 dpi in mouse brain, and led to fatal encephalitis before the end of the experimentation. Antigens of the HCoV-OC43 Paris variant were first revealed in mouse brain at 6 dpi. Infectious particles were detected in some of the brains up to 10 dpi. HCoV-OC43 Paris infectious titers in susceptible animals were similar to those found after HCoV-OC43 ATCC infection, and mice positive for either variant presented all pathological and clinical signs of encephalitis.

Figure 4. Phylogenetic unrooted tree regrouping seven coronavirus complete genomes from the three genetic groups. Circles regroup members of each three genetic groups. The 0.1 sliding bar represents the genetic distance between the species (i.e. nucleotide substitution units per studied site). MHV-A59, mouse hepatitis virus A59 strain (NC_001846); BCoV, bovine coronavirus Quebec strain (AF220295); SARS-HCoV, SARS-HCoV Tor2 strain (AY274119); IBV, infectious bronchitis virus Beaudette strain (NC_001451); TGEV, transmissible gastroenteritis virus (NC_002306); HCoV-229E (NC_002645).

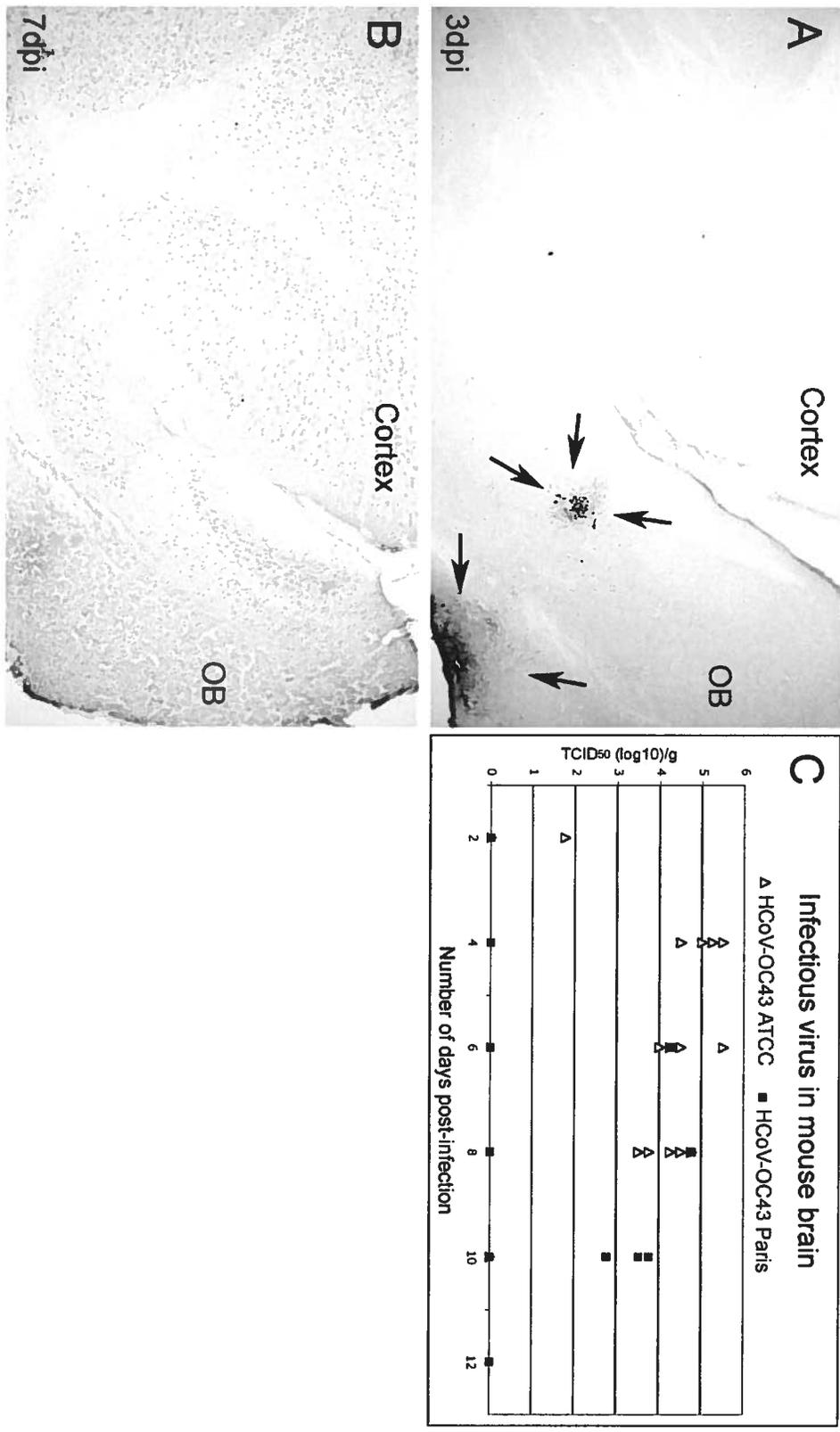
Figure 5. Multiple alignments of amino acids of the main proteases of coronaviruses from all three genetic groups. Positions with absolute conservation are shadowed whereas residues of the putative catalytic dyad, His⁴¹ and Cys¹⁴⁵, are boxed. Conservation level among group 2 coronaviruses was about 46.2% whereas all strains displayed 26% identity. OC43, HCoV-OC43 (group 2); BCoV, BCoV Quebec (group 2); MHV-A59, mouse hepatitis virus (group 2); SARS Tor2, SARS-HCoV Tor2 (group2); 229E, HCoV-229E (group 1); IBV, infectious bronchitis virus (group 3).

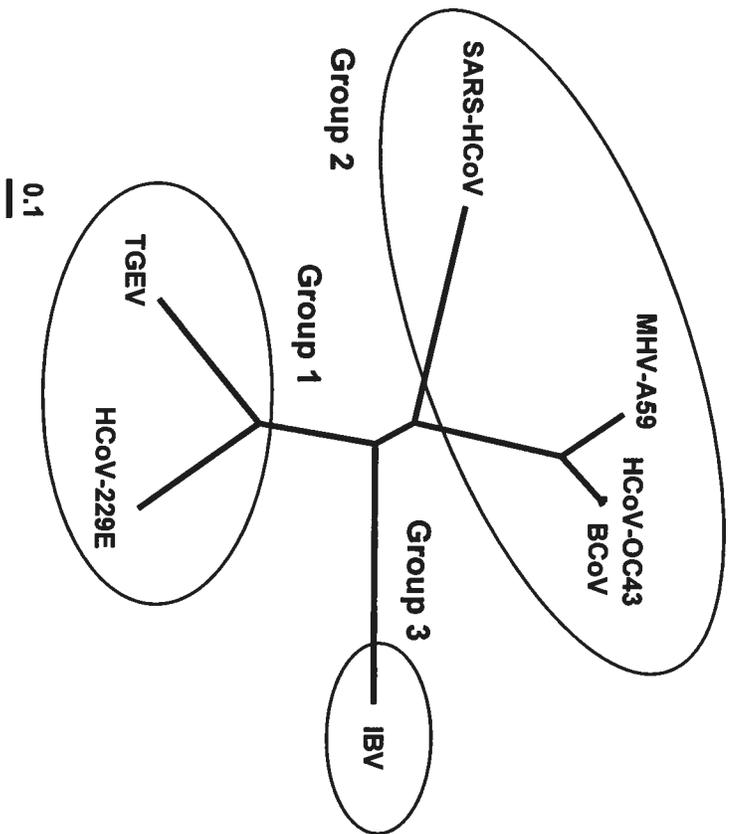




St-Jean et al. Fig. 2

St-Jean et al. Fig. 3





St-Jean et al. Fig. 4

```

OC43          SGIIVKAVNPT SKVEPCVSVV TYGNMTLNGL WLDKVKVYCPR HYICASADMT NPDYTNLLCR VTSSDFVLF DR-LSLTWMS
BCoV          SGIIVKAVNPT SKVEPCIVSV TYGNMTLNGL WLDKVKVYCPR HYICASADMT NPDYTNLLCR VTSSDFVLF DR-LSLTWMS
MHV-A59      SGIIVKAVSPT SKVEPCIVSV TYGNMTLNGL WLDKVKVYCPR HYICSSADMT DPDYTNLLCR VTSSDFCVMS GR-MSLTWMS
SARS Tor2    SGRFRKMAEPS GKVEGCWVQV TCGFTTLNGL WLDKVTVCYPR HYICTAEDML NPNYEDLLIR KSNHSFLVOA GN-VQLEVTIG
229E         AGLIRKMAOPS GFVEKCVVRV CYGNTVLNGL WLGDIVYCPR HYIAS-NTTS AIDYDHEYSI MRLHNEFITS GT-AFLGVVG
IBV          SGFKKIVSPTS SAVEKCVSVV SYRGNLNLGL WLGDITVYCPR HYIIGK---FS GDQWADVNL ANNHFEFVTT QHGVTLNVVS

OC43          YQMRGCMIVL TVTLQNSRTP KYTEGVKRPB EFTTVLAAYN GKPQGAHVHT MRSSYTIKGS FLGSGGSSVG YIIMGCVKE
BCoV          YQMRGCMIVL TVTLQNSRTP KYTEGVKRPB EFTTVLAAYN GKPQGAHVHT MRSSYTIKGS FLGSGGSSVG YIIMGCVKE
MHV-A59      YQMRGCMIVL TVTLQNSRTP KYTEGVKRPB EFTTVLAAYN GKPQGAHVHT MRSSYTIKGS FLGSGGSSVG YIIMGCVKE
SARS Tor2    HSMQNCILRL KVDTSNPKTP KYKFAVRIQBG QTESVLACYN GSPSSVYQCA MRPNHTIKGS FLNGSGSSVG ENIDYDCVSF
229E         ATMHGVTLKI KVSQTNMHTP RSHSFTLKSQ EGNIIACVYD GCAQGVFVGN MRTNWTIRGS FINGAGSSPG YNLKNGEVEF
IBV          RRLKGAVILL QPAVANAEFP KYKFIKANGG DSFTIACAYG GTVVGLYVPT MRSNGTIRAS FLAGHGGSSVG ENIEKGVVNE

OC43          VYMHQLELST GCHTGTDENG DEFYGPYKKAQ VVQLLIODYI QSVNFVAMLY AAILNNCN---WFOVSD KCSVEIDENVM
BCoV          VYMHQLELST GCHTGTDENG DEFYGPYKKAQ VVQLLPVQDYI QSVNFVAMLY AAILNNCN---WFOVSD KCSVEIDENVM
MHV-A59      VYMHQLELST GCHTGTDFEG NELYGPRDAQ VVQLLPVQDYT QTVNVVAMLY AAILNRCN---WFOVSD SCSLEEENVM
SARS Tor2    CYMHHELELPT GVHAGTDFEG KFYGPVVDKQ TAQAGTDTT ITLVNLAFLY AAVINNGDR---WFLNRE TTTLNDENLV
229E         VYMHQLELIGS GSHVGSSEFDG VMYGGFEEDQP NLQVESANQM LTVNVVAFLY AAILNGCT---WMLKGE KLFVEHYNEW
IBV          FYMHHELELPN ALHTGTDLMG EPHYGGYVDEE VAQRVPPDNL VTNNIVAWLY AAILSVKESS FSLPKWLEST TVSVDDYKMW

OC43          ALSNGFSQVK SDLV--IDAL ASMTGVSLST LLAAIKRLK-NGFOGRQIMG SCSFEDELTP SDVYQQLAGI KIQ
BCoV          ALSNGFSQVK SDLV--IDAL ASMTGVSLST LLAAIKRLK-NGFOGRQIMG SCSFEDELTP SDVYQQLAGI KIQ
MHV-A59      AMTNGFSSIK ADLV--LDAL ASMTGVTVBQ VLAAIKRLH-SGFQGRQIMG SCVLEDELTP SDVYQQLAGV KIQ
SARS Tor2    AMKRYNEPLT QDHVDILGPL SAQGTIAVD MCAALKKELQ NGMNGRITLG STILEDFTP FDVVRQCSGV TFO
229E         AQANGFTAMN GEDA--FSIL AAKGVCVVER LHAIQVYN-NGFGKQILG YSSLNDEFST NEVVKQMFVY NIQ
IBV          AGDNGFTPFS TSTA--ITKL SALTGVDVCK LKRTIMVKN-SQWGDPIIG QYNFEDELTP ESVENQIGGV RIQ
    
```

CHAPITRE III : PUBLICATION no.2 :

Apparent genetic stability of human respiratory coronavirus OC43

REPLY TO LETTER TO THE EDITOR FROM VIJGEN ET AL.**APPARENT GENETIC STABILITY
OF HUMAN RESPIRATORY CORONAVIRUS OC43**

Julien R. ST-JEAN¹, H  l  ne JACOMY¹, Marc DESFORGES¹, Astrid VABRET²,
Fran  ois FREYMUTH², and Pierre J. TALBOT^{1*}

Running title: Apparent genetic stability of human respiratory coronavirus OC43

¹ Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier,
531 boulevard des Prairies, Laval (Qu  bec), CANADA H7V 1B7
and

² Laboratoire de Virologie Humaine et Mol  culaire,
Centre Hospitalier R  gional et Universitaire de Caen,
Avenue Georges Cl  menceau, 14033 Caen, FRANCE

* Correspondent footnote:

Dr. Pierre J. Talbot
Laboratory of Neuroimmunovirology
INRS-Institut Armand-Frappier
531 boulevard des Prairies,
Laval (Qu  bec)
CANADA
H7V 1B7
Tel.: (450) 686-5515
Fax: (450) 686-5566
E-mail: pierre.talbot@inrs-iaf.quebec.ca

Word counts: Text: 1676 (including references) 1273 (without references)

In their comment regarding our recently published paper (11), Vijgen *et al.* present different arguments to underline their hypothesis that the genetic stability of the human coronavirus HCoV-OC43 that we report should be seen as an highly improbable event.

According to Vijgen *et al.*, statistical analysis based on RNA virus evolution models reveals that the mutation rate per site per year for HCoV-OC43 should be between 0.97×10^{-4} and 2.12×10^{-4} (95% confidence interval), with a maximum likelihood of 1.54×10^{-4} .

Based on their arguments, the six-nucleotide difference we found between the HCoV-OC43 ATCC strain and the HCoV-OC43 Paris isolate obtained in March 2001 (11), represents a mutation rate of 5.7×10^{-6} , which is 30-fold lower than expected from theoretical statistical calculations. Furthermore, using the Poisson process model, Vijgen *et al.* calculated that the probability of this result (only six nucleotide differences between the genome of the HCoV-ATCC strain and the Paris isolate) is only 1.9×10^{-16} .

The statistical arguments presented by Vijgen *et al.* appear very interesting and convincing in terms of statistical analysis based on some previously reported evolutionary rates of RNA viruses. Furthermore, as the authors of the paper comparing these two HCoV-OC43 variants (11), we do indeed acknowledge that our published data of apparent genetic stability is surprising. However, we are putting forward herein some hypotheses that could contribute to explain the similarity of the genome of the two HCoV-OC43 variants described in our paper (11), which may explain this apparent genetic stability of HCoV-OC43.

- 1) Before the sequencing of its genome, the HCoV-OC43 Paris isolate was passaged on the HRT-18 cell line six times. The human coronavirus contained in the original clinical isolate, which comes from the upper respiratory tract of a 68-year old immunocompromised male who was not related whatsoever to laboratory work and was not in contact with any

laboratory workers who had manipulated the HCoV-OC43 ATCC virus (citation from Materials and Methods section of our paper), consisted, as a typical RNA virus, of a complex mixture of variants, or quasispecies (2, 3, 6). As reported by Elena (3), some sort of evolutionary constraint may exist for adaptation in a particular environment. According to this assumption, by having been cultured within the HRT-18 cell line for six passages, the HCoV-OC43 Paris strain may have been subject to this type of constraint, which would have highly favored its replication in the HRT-18 cells. In other words, the six passages on the HRT-18 cells would have selected a particular variant (with only a six-nucleotide difference from HCoV-OC43 ATCC), that was present in the quasispecies mixture in the clinical isolate. Furthermore, we can now add the supplemental data that the Paris variant was the only one among several clinical isolates that was able to replicate in HRT-18 cells, meaning that this particular isolate could have already been very similar to HCoV-OC43 ATCC at the onset of the viral isolation procedure (Vabret and Freymuth, unpublished results).

- 2) As already described in our paper (11), highly stringent laboratory precautions were used in order to eliminate a possible cross-contamination between the two viruses (citation from the Materials and Methods section of our paper: “The HCoV-OC43 ATCC strain and the Paris isolate were never cultured at the same time, and stringent laboratory precautions were used in order to eliminate possible cross-contamination.”). We still are highly confident that all the manipulations that served to either isolate or amplify the HCoV-OC43 Paris variant, were performed properly, as we were always acutely aware of the potential problems of cross-contamination.
- 3) As Vijgen *et al.* are suggesting a possible contamination of the HRT-18 cell line used to propagate the HCoV-OC43 ATCC variant, we have performed RT-PCR assays on the HRT-18 cells that were used to replicate the HCoV-OC43 and were not able to detect any HCoV-OC43 RNA even after molecular hybridization using a probe specific for the M gene

of the virus (Vabret and Freymuth, unpublished results). A preexisting undetected infection of the HRT-18 cell line used to propagate HCoV-OC43 therefore appears highly unlikely.

- 4) The point mutation located in the M gene of the HCoV-OC43 Paris variant (T432C) is quite interesting. Indeed, while it is absent in the M gene of the HCoV-OC43 ATCC variant, this precise mutation is found in all 20 samples of HCoV-OC43 that were isolated during the Normandy outbreak of 2001 (13) and directly sequenced by Vabret *et al.* without any passage on any cell line (Vabret and Freymuth, unpublished results). Furthermore, the T432C mutation is found neither in the first reported sequence of the M gene, which was reported by our laboratory (8), nor in the complete genome sequences of the HCoV-OC43 so far available (11; Vijgen *et al.*, GenBank accession no. NC_005147). Therefore, this mutation appears to be a marker for the clinical isolates from the outbreak of HCoV-OC43 in Normandy during winter 2001 (13) and is not found in HCoV-OC43 ATCC.
- 5) As reported by Vijgen *et al.*, the evolutionary rates of RNA viruses in general (6), and of coronaviruses in particular (10), are usually higher than the results presented in our paper (11). However, some recent publications reported a higher-than-expected genetic stability for RNA viruses. For example, it has been reported that the usually highly variable regions of the VP proteins of Hepatitis A virus (4) and the highly variable *env* gene of feline immunodeficiency virus (FIV) (7) are shown experimentally to be more genetically stable than predicted statistically. Indeed, even though the latter is a retrovirus that could establish a latent infection, its mutation rate is usually comparable to that of HIV-1, at about 0.34% per site per year (5). However, Ikeda *et al.* (7) showed that FIV could remain very stable despite a 10-year infection in cats, as the mutation rate within the *env* gene was only 0.015% per year (7). This result represents an unexpected genetic stability, as this rate of 0.015% is about 23-fold lower than what was previously reported (5). This finding is in the same order of magnitude as the 30-fold less-than-expected variability calculated by Vijgen *et al.* on the

basis our published experimental data (11). Moreover, even though, the SARS-HCoV was pointed out by Vijgen *et al.* as an example to illustrate the extreme genetic variability of coronaviruses (9), another very interesting paper on SARS-HCoV indicates instead that this virus shows a relative genetic stability, with only 2 mutations in the S gene and 1 mutation in the N gene found in 10 different clinical isolates over a 3-month period (12). Finally, another recent study on a coronavirus, the feline coronavirus, also indicated that this virus could be more genetically stable than theoretically expected. Indeed, the sequencing of a small portion of the S gene from samples isolated from five persistently infected cats revealed low rates of mutations, ranging from none over a period of 17 months for one cat to only nine nucleotide changes over a 5-year period for another animal (1).

In conclusion, given the above comments and supportive arguments from the published literature, and even though they are, as underlined by the comment of Vijgen *et al.*, unexpected, we do stand by our results concerning the apparent genetic stability of the HCoV-OC43 virus.

REFERENCES

1. **Addie, D. D., I. A. T. Schaap, L. Nicolson, and O. Jarrett.** 2003. Persistence and transmission of natural type I feline coronavirus infection. *J. Gen. Virol.* **84**:2735-2744.
2. **Domingo, E.** Rapid evolution of viral RNA genomes. 1997. *J. Nutr.* **127 (suppl 5)**:958S-961S.
3. **Elena, S.F.** 2002. Restrictions to RNA virus adaptation: An experimental approach. *Antonie van Leeuwenhoek.* **81**:135-142.
4. **Gabrieli, R., G. Sanchez, A. Macaluso, F. Cenko, S. Bino, L. Palombi, E. Buonomo, R. M. Pinto, A. Bosch, and M. Divizia.** 2004. Hepatitis in Albanian children: molecular analysis of hepatitis A virus isolates. *J. Med. Virol.* **72**:533-537.
5. **Green, W.K., J. Meers, G. del Fierro, P.R. Carnegie, and W.F. Robinson.** 1993. Extensive sequence variation of feline immunodeficiency virus env genes in isolates from naturally infected cats. *Arch. Virol.* **133**:51-62.
6. **Holland, J., and E. Domingo.** 1998. Origin and evolution of viruses. *Virus Genes.* **16**:13-21.
7. **Ikeda, Y., T. Miyazawa, Y. Nishimura, K. Nakamura, Y. Tohya, and T. Mikami.** 2004. High genetic stability of TM1 and TM2 strains of subtype B feline immunodeficiency virus in long-term infection. *J. Vet. Med. Sci.* **66** :287-289.
8. **Mounir, S., and P. J. Talbot.** 1992. Sequence analysis of the membrane protein gene of human coronavirus OC43 and evidence for *O*-glycosylation. *J. Gen. Virol.* **73**:2731-2736.

9. **Salemi, M., W.M. Fitch, M. Cicozzi, M.J. Ruiz-Alvarez, G. Rezza, and M.J. Lewis.** 2004. Severe acute respiratory syndrome coronavirus sequence characteristics and evolutionary rate estimate from maximum likelihood analysis. *J. Virol.* **78** :1602-1603.
10. **Sanchez, C.M., F. Gebauer, C. Sune, A. Mendez, J. Dopazo, L. Enjuanes.** 1992. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology.* **109** :92-105.
11. **St-Jean, J. R., H. Jacomy, M. Desforges, A. Vabret, F. Freymuth, and P. J. Talbot.** 2004. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J. Virol.* **78**:8824-8834.
12. **Tong, S., J. R. Lingappa, Q. Chen, B. Shu, A. C. LaMonte, B. T. Cook, C. Birge, S. W. Chern, X. Liu, R. Galloway, L. Q. Mai, W. F. Ng, J.-Y. Yang, J. Butany, J. A. Comer, S. S. Monroe, S. R. Beard, T. G. Ksiazek, D. Erdman, P. A. Rota, M. A. Pallansch, and L. J. Anderson.** 2004. Direct sequencing of SARS-coronavirus S and N genes from clinical specimens shows limited variation. *J. Infect. Dis.* **190**:1127-1131.
13. **Vabret, A., T. Mourez, S. Gouarin, J. Petitjean, and F. Freymuth.** 2003. An outbreak of coronavirus OC43 respiratory infection in Normandy, France. *Clin. Infect. Dis.* **36**:985-989.

CHAPITRE IV : PUBLICATION no.3 :

Genetic evolution of human coronavirus OC43 in neural cell culture

GENETIC EVOLUTION OF HUMAN CORONAVIRUS OC43 IN NEURAL CELL CULTURE

Julien R. St-Jean, Marc Desforges, and Pierre J. Talbot ¹

¹ Pierre J. Talbot, INRS-Institut Armand-Frappier, 531 boul. des Prairies, Laval (Québec),
CANADA H7V 1B7, Tel : (450) 686-5515; Fax : (450) 686-5566; Pierre.Talbot@iaf.inrs.ca

1. INTRODUCTION

Human coronaviruses (HCoV) are ubiquitous in the environment and are responsible for up to one third of common colds. HCoV-OC43 possesses a genome that comprises genes encoding various structural and non-structural proteins. Amongst these proteins, the S protein is biologically very important because it could be involved in determination of viral tropism. Indeed, it could for instance be associated with the capacity of the virus to reach the central nervous system (CNS) and possibly trigger neurological disorders. It could also confer the host species specificity observed with coronaviruses. In the past years, we have shown that HCoV-OC43 is neurotropic and neuroinvasive, as it persistently infects neural cell cultures (Arbour *et al.*, 1999) and human brains (Arbour *et al.*, 2000). Although we have suggested that OC43 could remain genetically surprisingly stable in the environment (St-Jean *et al.*, 2004), it is known that coronaviruses can adapt in cell culture or under selection pressure, for instance related to immune system evasion.

2. MATERIALS AND METHODS

2.1 Viruses, cell lines and persistent infections

The ATCC HCoV-OC43 strain (VR-759) was grown on the HRT-18 rectal tumor cell line. Persistent infection were carried out in those HRT-18 cells, as well as in the MO3.13 oligodendrocytic (McLaurin *et al.*, 1995), H4 neuroglial, U-87 MG astrocytic and TE-671 rhabdomyosarcoma cell lines (ATCC). Other cell lines used for virus susceptibility are described in Table 2. Four infections were performed in the H4 cell line, whereas the HRT-18, MO3.13 and H4 cell lines were acutely infected as controls.

2.2 Virus purification

Virus from persistent infections was purified at different passages. Prior to purification, virus was clarified and precipitated with polyethylene glycol (PEG) 8000 (Sigma). Accudenz (Accurate Chemicals) was used to perform gradient purification.

2.3 RT-PCR and sequencing

Viral RNA was extracted using the GenElute Direct mRNA Miniprep Kit (Sigma) and reverse transcribed with MmuLV reverse transcriptase (Invitrogen). The Expand High-Fidelity *Taq* polymerase (Roche) was used to perform PCR. Primers specific to the HE, S and N genes were used to amplify target regions (St-Jean *et al.*, 2004). PCR amplicons were purified using the Qiaex II gel extraction kit (Qiagen) prior to sequencing, which was carried out by Bio S&T (Montréal, Québec, Canada).

2.4 Assays for viral susceptibility and modulation of tropism and infectivity

Prior to performing assays for modulation of tropism and infectivity, susceptibility of different cell lines to HCoV-OC43, ATCC strain, was determined (Table 2). The same cell lines

were then infected with virus isolated from different purifications (HRT-18 P33, P54, P110 and P155; H4 P47 and P90; H4 P56.1, P56.2, P56.3, P116.1, P116.2 and P116.3; TE-671 P38 and P79; U-87 MG P35, and MO3.13 P5, P6 and P22) in order to correlate the observed mutations with a modulation of tropism or infectivity. Supernatants were titrated using an indirect immunoperoxidase assay (IPA), as previously described (Jacomy and Talbot, 2003).

3. RESULTS

Persistent infections of neural cell lines were beforehand carried out to verify whether virus carrying mutations in genes encoding the surface protein S originated from viral persistence. The HE protein gene and the nucleocapsid protein gene N were also sequenced in order to determine if these genes could be related to any adaptation in cell culture. Viral particles released from persistently infected neural cell lines were isolated and purified by gradient centrifugation, and genomic RNA was sequenced. Results showed various mutations in the S gene but very few in HE and N genes, suggesting that the S gene is responsible for adaptation to the cellular environment, which could be associated with neurotropism, neuroinvasion, and presumably neuropathogenesis (Table 1). Almost every acquired mutation (Table 1) was conserved at subsequent passages, suggesting that they could confer an adaptive advantage and a stable phenotype to the virus. Five mutations were predominant and were found in almost all persistent infections (D24Y, S83T, H183R, Y241H and N489H). The first four mutations are located in the receptor binding site, whereas the fifth one is located within the hypervariable region.

To correlate the observed mutations in the S gene with viral replication and tropism, assays for modulation of tropism and infectivity (Schickli *et al.*, 1997) were performed using cell lines originating from various human tissues as well as from various animal species, for which susceptibility to HCoV-OC43 infection was previously determined (Table 2). These analyses

revealed that mutations found throughout the S gene could affect the latter viral properties in certain cell lines. Amongst the virus variants obtained following persistent infections and virus purifications, five showed extended cellular tropism and increased replication titers *in vitro*: U87-MG P35, H4 P47, H4 P56.3, H4 P116.1 and H4 P116.2 (data not shown). Furthermore, some variants isolated from persistent infections were more virulent in mice and could form plaques, in opposition to the ATCC HCoV-OC43 reference strain VR759 (data not shown).

Table 1. Location of S mutations at various passages of persistently infected cells

HRT-18	H4	H4	H4	H4	TE-671	U-87 MG	MO3.13
P155**	P90	P116.1	P116.2	P116.3	P79	P35	P22
D30H*	D24Y	D24Y	D24Y	N25Y	N27Y	D24Y	D115H
S83T	V161V	P35S	P35S	P35S	P34S	S83T	T148I
L85Q	H183R	S83T	S147Y (D)	S83T	L85R	H183R	Y241H
D115H	V240V	E170K	H183R	Y119H	S258R	Y241H	M670T
T148I	Y241H	H183R	Y241H	S147P	A373V		P973S
H183Q	N441K	Y241H	N441K	H183R	R757S		A1090V
S258R	Q541L	A469V	E460D	Y241H	G785D		V1213A
S366G	R570P	R570R	H482Y	N489H	P972L		
N413T	N639N	T855I	F683Y	K506T	P973S		
F420S	T855I	N880K (I)	L693F	T641S	A978S		
N489H	D875H	L893H	A759E	N768T	T1086N		
K506N	L893R	S959C	S898S (I)	E896K	D1170A		
T536N	A965V	W974L	V980A	S901F			
Q541L	T975A	T975P	N1203 (D)	W974L			
R757H	I1227T	V980A	I1227T	F982L			
E896D	T1245I	S1093S		V986I			
C897G		G1169G		G1169D			
E933G		M1222K		E1236A			
F982L		D1232Y					
S1192R		P1249L					
T1225I		I1304I					
P1228S							

* D, deletion; I, insertion. ** Passages (and purification numbers) are indicated below the cell line.

4. DISCUSSION

We have identified several mutations in the S gene of the HCoV-OC43 genome following persistent infections in different cell lines. These mutations will help us to further characterize

viral adaptation during persistence and to understand mechanisms that are implicated in viral tropism and infectivity. Future studies will be carried out using an infectious cDNA clone of the OC43 strain assembled in a BAC vector (Almazán *et al.*, 2000; St-Jean *et al.*, submitted). The construction of this clone was performed in collaboration with F. Almazán and L. Enjuanes and will provide an invaluable tool to further understand the underlying mechanisms for viral replication and tropism and, in combination with the experiments described above, will help us get useful data about the molecular basis for human coronavirus neuropathogenesis.

Table 2. Viral susceptibility of various cell lines to the HCoV-OC43 ATCC strain

Cell line	Origin	Tissue	Type	IPA ¹	Susceptibility ²
HeLa	Human	Uterus	Epithelial	3.75	Low
MT4	Human	Bone marrow	T lymphocyte	2.0	No
U937	Human	Bone marrow	Monocyte	≤0.5	No
Jurkat E6.1	Human	Bone marrow	T lymphocyte	≤1.5	No
Raji	Human	Bone marrow	B lymphocyte	4.5	Yes
HL-60	Human	Bone marrow	Monocyte	3.25	Low
WI-38	Human	Lung	Fibroblast	4.0	Yes
L132	Human	Lung	Epithelial	3.25	Low
Caki-2	Human	Kidney	Epithelial	≤1.5	No
SW 156	Human	Kidney	Epithelial	≤1.5	No
NCI-N87	Human	Stomach	Epithelial	≤1.75	No
Arpe-19	Human	Eye	Epithelial	≤1.5	No
FHs 74 Int	Human	Intestine	Epithelial	3.5	Low
TK6	Human	Spleen	T lymphocyte	≤1.75	No
17 Cl-1	Mouse	Embryo	Fibroblast	≤1.75	No
L929	Mouse	Subcutaneous	Fibroblast	≤1.5	No
N-11	Mouse	Brain	Microglial	≤1.5	No
DBT	Mouse	Brain	Glial	≤2.25	No
J774 A.1	Mouse	Bone marrow	Macrophage	4.5	Yes
A20	Mouse	Bone marrow	B lymphocyte	3.25	Low
S.END.1	Mouse	Skin	Endothelial	≤1.5	No
Cos-7	Monkey	Kidney	Fibroblast	≤1.5	No
Vero	Monkey	Kidney	Epithelial	≤1.75	No
Vero E.6	Monkey	Kidney	Epithelial	≤1.5	No
B104	Rat	Brain	Fibroblast	3.25	Low
BHK-21	Hamster	Kidney	Fibroblast	4.0	Yes

¹ Indirect immunoperoxidase assay (infectious titers in TCID₅₀/mL).

² Titers from 0 to 3: not susceptible; titers over 3 and under 4: low susceptibility; titers of 4 and over: susceptible.

5. REFERENCES

- Almazán, F., Gonzalez, J. M., Péntzes, Z., Izeta, A., Calvo, E., Plana-Duran, J., Enjuanes, L. 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci.* **97**:5516-5521.
- Arbour, N., Côté, G., Lachance, C., Tardieu, M., Cashman, N. R., Talbot, P. J. 1999. Acute and persistent infection of human neural cell lines by human coronavirus OC43. *J. Virol.* **73**:3338-3350.
- Arbour, N., Day, R., Newcombe, J., Talbot, P. J. 2000. Neuroinvasion by human respiratory coronaviruses. *J. Virol.* **74**:8913-8921.
- Jacomy, H., Talbot, P. J. 2003. Vacuolating encephalitis in mice infected by human coronavirus OC43. *Virology* **315**:20-33.
- McLaurin, J., Trudel, G. C., Shaw, I. T., Antel, J. P., Cashman, N. R. 1995. A human glial hybrid cell line differentially expressing genes subserving oligodendrocyte and astrocyte phenotype. *J. Neurobiol.* **26**:183-193.
- Mounir, S., Talbot, P. J. 1992. Sequence analysis of the membrane protein gene of human coronavirus OC43 and evidence for *O*-glycosylation. *J. Gen. Virol.* **73**:2731-2736.
- Schickli, J. H., Zelus, B. D., Wentworth, D. E., Sawicki, S. G., Holmes, K. V. 1997. The murine coronavirus mouse hepatitis virus strain A59 from persistently infected murine cells exhibits an extended host range. *J. Virol.* **71**:9499-9507.
- St-Jean, J. R., Jacomy, H., Desforges, M., Vabret, A., Freymuth, F., Talbot, P. J. 2004. Human respiratory coronavirus OC43 : Genetic stability and neuroinvasion. *J. Virol.* **78**:8824-8834.

CHAPITRE V : PUBLICATION no.4 :

Recovery of a neurovirulent human coronavirus OC43 from an infectious cDNA clone

RECOVERY OF A NEUROVIRULENT HUMAN CORONAVIRUS OC43
FROM AN INFECTIOUS cDNA CLONE

Julien R. ST-JEAN ¹, Marc DESFORGES ¹, Fernando ALMAZÁN ², Hélène JACOMY ¹, Luis ENJUANES ², and Pierre J. TALBOT ^{1*}

Running title: Infectious cDNA clone of HCoV-OC43

¹ Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier,
531 boulevard des Prairies, Laval (Québec), CANADA H7V 1B7
and

² Department of Molecular and Cell Biology,
Centro Nacional de Biotecnología, CSIC, Campus Universidad Autónoma,
Darwin 3, 28049 Madrid, Spain

* Correspondent footnote:

Dr. Pierre J. Talbot
Laboratory of Neuroimmunovirology
INRS-Institut Armand-Frappier
531 boulevard des Prairies,
Laval (Québec)
CANADA
H7V 1B7

Tel.: (450) 686-5515
Fax: (450) 686-5566
E-mail: pierre.talbot@iaf.inrs.ca

Word counts:

Abstract: 107

Text: 1267

RÉSUMÉ

Cette étude décrit l'assemblage d'un clone d'ADNc pleine longueur du coronavirus humain (HCoV)-OC43 dans un *bacterial artificial chromosome* (BAC). Le BAC contenant l'ADNc infectieux pleine longueur, le pBAC-OC43^{FL}, a été assemblé en utilisant une stratégie en deux parties. La première étape consistait en l'introduction de chacune des extrémités du génome viral dans le BAC, combinées à des séquences accessoires permettant une transcription adéquate du génome viral. La deuxième étape consistait en l'insertion du génome complet d'ADNc du HCoV-OC43 dans le BAC. Afin de produire des particules virales recombinantes, le pBAC-OC43^{FL} a été transfecté dans des cellules BHK-21. Les virus recombinants présentaient les mêmes propriétés phénotypiques que le virus de souche sauvage lorsque les titres viraux en culture cellulaire et la neurovirulence chez la souris étaient comparés.

ABSTRACT

This study describes the assembly of a full-length cDNA clone of human coronavirus (HCoV)-OC43 in a bacterial artificial chromosome (BAC). The BAC containing the full-length infectious cDNA (pBAC-OC43^{FL}), was assembled using a two-part strategy. The first step consisted in the introduction of each end of the viral genome into the BAC with accessory sequences allowing proper transcription. The second step consisted in the insertion of the whole HCoV-OC43 cDNA genome into the BAC. To produce recombinant viral particles, pBAC-OC43^{FL} was transfected into BHK-21 cells. Recombinant virus displayed the same phenotypical properties as wild type virus, including infectious viral titers produced in cell culture and neurovirulence in mice.

Human coronaviruses (HCoV) are positive-stranded enveloped RNA viruses that are mainly associated with respiratory infections and are responsible for up to one third of common colds (10), and possibly neurologic disease (3, 8). Like SARS-CoV, which was identified as the causative agent of severe acute respiratory syndrome (SARS) (11), HCoV-OC43 (ATCC VR-759 from the mid 1960s; GenBank accession number AY585228) (17) belongs to the second genetic group of the *Coronaviridae* (7, 14).

Several coronavirus infectious clones have been assembled, although this has proven difficult given the length of the viral genome (9), as well as the presence of unstable viral cDNA regions in bacteria (2, 6, 21). Uncommon strategies were therefore necessary to circumvent technical limitations engendered by the viral genome. Among the strategies available for the construction (2, 4, 5, 18, 19, 20, 21), the bacterial artificial chromosome (BAC) approach was selected given the availability of the technology as well as the number of possibilities it offers for downstream applications (1, 2, 15, 16, 22). This approach uses a two-step amplification system that couples viral RNA expression in the nucleus from the cytomegalovirus (CMV) immediate-early promoter, with a second amplification step in the cytoplasm driven by the viral polymerase.

Construction of the full length HCoV-OC43 cDNA clone.

Prior to introducing the full-length viral cDNA into the pBeloBAC11 (13), pBAC-OC43-5'-3' was first assembled. This plasmid includes the first 720 nt of the genome under the control of the CMV promoter, and the last 1651 nt of the viral RNA, followed by a 25 bp synthetic poly(A), the hepatitis delta virus ribozyme (HDV) and the bovine growth hormone termination and polyadenylation sequences (BGH) to lead to an accurate 3' end. In order to join the CMV promoter with the viral 5' end and the viral 3' end with the HDV-BGH sequences in a contiguous and very precise way, overlapping PCRs were performed.

Primers used to perform the overlapping PCRs are listed in Table 1. The vector pBAC-TGEV-5'-3', which contains the accessory sequences CMV, HDV and BGH, was used as a template to amplify these latter sequences. Following two rounds of amplification, an overlapping PCR was performed for each end of the viral genome (Fig. 1A). Both overlapping PCR amplicons were directly introduced into the pBeloBAC11 plasmid in order to create the pBAC-OC43-5'-3' (Fig. 1B). After each cloning step, the new BAC vector was sequenced to assess that no undesired mutations were introduced into the new construct.

Before performing the assembly of the full-length cDNA clone, RACE 5' was carried out with RNA extracted from BHK-21 cells transfected with pBAC-OC43-5'-3' in order to check that the transcription machinery was functional (data not shown). The HDV cleavage efficiency was also determined using a RT-PCR approach and as expected, it was estimated at approximately 60% (data not shown) (12).

The HCoV-OC43 infectious cDNA clone was assembled using a five-step cloning approach, starting with the pBAC-OC43-5'-3' as a backbone. Given the low availability of unique restriction sites in the HCoV-OC43 genome, three additional restriction sites (*MluI*, *AsiSI* and *SacII*) were introduced into the sequence of the cDNA clone. This modification required the introduction of six silent point mutations (T6479C, T6482G, C12143T, T12146C, A24068G, T24071G). These sequence changes were used as genetic markers to identify the recombinant virus (2, 19). For cDNA assembly, we took advantage of the presence of the unique *SfoI* restriction site in both the BAC vector and the HCoV-OC43 genome (Fig. 1B and 2A).

The genome of HCoV-OC43 was therefore amplified in five distinct overlapping regions, named A to E, which spanned the entire viral genome, except for the 5' and 3' ends (Fig. 2A). Each region was generated in such a way that it contained a given cloning site at one end and a *SfoI* site at the other end. Primers used for the generation of region A to E are described in

Table 1. Each amplicon was first inserted into the pSMART™ HCKan vector in order to generate plasmids pSMART-A to -E (Fig. 2A) and positive clones with the appropriate sequence were obtained.

Viral DNA regions contained in clones pSMART-A to -E were used to fill the pBAC-OC43-5'-3' and to generate the pBAC-OC43^{FL} (Fig 2B). Regions A to E were respectively introduced into the BAC vector using the restriction sites they were sharing with the construct (Fig 2A). Region E, which contained the disrupted unstable segment of ORF1a, was introduced at the last cloning step to minimize the toxicity problem usually observed with this region of the genome (6, 21). Junction sites created for each new construct were sequenced to make sure no unexpected mutations were introduced into the genome.

Recovery, replication and neurovirulence of recombinant HCoV-OC43.

One of the major advantage provided by the BAC system is that the recovery of recombinant viral particles is very efficient and simple. This was especially true for pBAC-OC43^{FL} given that the cell line BHK-21 provided a good transfection rate and was also susceptible to the infection by HCoV-OC43. For rescue of recombinant virus, 5 µg of pBAC-OC43^{FL} were directly transfected into BHK-21 cells using Lipofectamine 2000 (Invitrogen). Viral titers obtained for the recombinant virus reached 6.5 TCID₅₀/mL, which are comparable to those obtained following infection with the HCoV-OC43 laboratory strain. To make sure rescued viruses were indeed recombinant particles, RT-PCR were performed and the presence of the newly introduced restriction sites, which are exclusive to the recombinant virus, was assessed (data not shown).

To test whether the recombinant virus displayed the same growth kinetics as the parental strain in HRT-18 cells, replication was assessed at different times post-infection. Wild type and recombinant virus shared very similar patterns of replication, reaching viral titers of

approximately $10^{6.0}$ TCID₅₀/mL after four days of infection, although the replication of the recombinant virus was slightly delayed during the first two days (Fig. 3A). As expected and like the wild type HCoV-OC43 strain, recombinant virus did not form any plaques (data not shown). Taken together, these data confirm that the recombinant virus shares phenotypical properties with the wild type HCoV-OC43 strain in cell culture.

To determine whether the HCoV-OC43 strain as well as the recombinant virus displayed the same phenotype *in vivo*, their neurovirulence was assessed in mice. C57Bl/6 mice and BALB/c mice were infected intra-cerebrally with either virus (8). In C57Bl/6 mice, a survival rate of 30% was achieved for both viruses whereas 70% and 80% of BALB/c mice survived the infection with the recombinant virus and the parent strain respectively (Fig. 3B). Both recombinant and wild type virus displayed very similar levels of virulence following injection into the brain. Furthermore, both viruses showed very similar patterns of replication in mouse brain (data not shown).

Overall, the virus recovered from the infectious cDNA clone exhibited the same genotypical and phenotypical properties as the HCoV-OC43 from ATCC, both *in vitro* and *in vivo*.

The cDNA infectious clones represent very interesting tools for the study of potential vaccine candidates, like genetically attenuated viruses, but also allow multiple possibilities for the characterization of any viral functions and mechanisms by reverse genetics. The pBAC-OC43^{FL} will be invaluable in the characterization of the adaptation of the HCoV-OC43 virus in a neural environment and will provide an opportunity to better understand the biology of the HCoV-OC43 as well as other related coronaviruses, like SARS-CoV.

We are grateful to Francine Lambert for her technical assistance with cell and virus culture. We thank Dr. R. C. Lévesque (Université Laval, Québec, Canada) for having provided us with vector pBeloBAC11 and Dr. Luc Gagnon (Ecopia Biosciences Inc., Montréal, Québec, Canada) for his precious advice regarding cloning techniques in the BAC vector.

This work was supported by Grant No. MT-9203 from Canadian Institutes for Health Research (Institute of Infection and Immunity) to Pierre J. Talbot, who is the holder of a Tier-1 Canada Research Chair, and by a grant from the Department of Education and Science of Spain (BIO2004-00636) to Luis Enjuanes. Julien R. St-Jean acknowledges a studentship from *Fonds québécois de recherche sur la nature et les technologies* (FQRNT).

REFERENCES

1. **Almazán, F., C. Galán, and L. Enjuanes.** 2004. The nucleoprotein is required for efficient coronavirus genome replication. *J. Virol.* **78**:12683-12688.
2. **Almazán, F., J. M. González, Z. Péntes, A. Izeta, E. Calvo, J. Plana-Durán, and L. Enjuanes.** 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. USA* **97**:5516-5521.
3. **Arbour, N., R. Day, J. Newcombe, and P. J. Talbot.** 2000. Neuroinvasion by human respiratory coronaviruses. *J. Virol.* **74**:8913-8921.
4. **Casais, R., V. Thiel, S. G. Siddell, D. Cavanagh, and P. Britton.** 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J. Virol.* **75**:12359-12369.
5. **Coley, S. E., E. Lavi, S. G. Sawicki, L. Fu, B. Schelle, N. Karl, S. G. Siddell, and V. Thiel.** 2005. Recombinant mouse hepatitis virus strain A59 from cloned, full-length cDNA replicates to high titers in vitro and is fully pathogenic in vivo. *J. Virol.* **79**:3097-3106.
6. **Gonzalez, J. M., Z. Péntes, F. Almazán, E. Calvo, and L. Enjuanes.** 2002. Stabilization of a full-length infectious cDNA clone of transmissible gastroenteritis coronavirus by insertion of an intron. *J. Virol.* **76**:4655-4661.
7. **Gorbalenya, A. E., E. J. Snijder, and W. J. Spaan.** 2004. Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. *J. Virol.* **78**:7863-7866.
8. **Jacomy, H., and P. J. Talbot.** 2003. Vacuolating encephalitis in mice infected by human coronavirus OC43. *Virology* **315**:20-33.
9. **Masters, P. S.** 1999. Reverse genetics of the largest RNA viruses. *Adv. Virus Res.* **53**:245-264.
10. **Myint, S. H.** 1994. Human coronaviruses : a brief review. *Rev. Med. Virol.* **4**:35-46.
11. **Peiris, J. S., C. M. Chu, V. C. Cheng, K. S. Chan, I. F. Hung, L. L. Poon, K. I. Law, B. S. Tang, T. Y. Hon, C. S. Chan, K. H. Chan, J. S. Ng, B. J. Zheng, W. L. Ng, R. W. Lai, Y.**

- Guan, and K. Y. Yuen.** 2003. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**:1767-1772.
12. **Sharmeen, L., M. Y. P. Kuo, G. Dinter-Gottlieb, and J. Taylor.** 1988. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J. Virol.* **62**:2674-2679.
13. **Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon.** 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci.* **89**:8794-8797.
14. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. M. Spaan, and A. E. Gorbalenya.** 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **331**:991-1004.
15. **Sola, I., S. Alonso, S. Zúñiga, M. Balasch, J. Plana-Durán, and L. Enjuanes.** 2003. Engineering the transmissible gastroenteritis virus genome as an expression vector inducing lactogenic immunity. *J. Virol.* **77**:4357-4369.
16. **Sola, I., J. L. Moreno, S. Zúñiga, S. Alonso, and L. Enjuanes.** 2005. Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J. Virol.* **79**:2506-2516.
17. **St-Jean, J. R., H. Jacomy, M. Desforges, A. Vabret, F. Freymuth, and P. J. Talbot.** 2004. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J. Virol.* **78**:8824-8834.
18. **Thiel, V., J. Herold, B. Schelle, and S. G. Siddell.** 2001. Infectious RNA transcribed *in vitro* from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.* **82**:1273-1281.
19. **Yount, B., K. M. Curtis, and R. S. Baric.** 2000. Strategy for systematic assembly of large RNA and DNA genomes : transmissible gastroenteritis virus model. *J. Virol.* **74**:10600-10611.

20. **Yount, B., K. M. Curtis, E. A. Fritz, L. E. Hensley, P. B. Jahrling, E. Prentice, M. R. Denison, T. W. Geisbert, and R. S. Baric.** 2003. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci.* **100**:12995-13000.
21. **Yount, B., M. R. Denison, S. R. Weiss, and R. S. Baric.** 2002. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J. Virol.* **76**:11065-11078.
22. **Zúñiga, S., I. Sola, S. Alonso, and L. Enjuanes.** 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J. Virol.* **78**:980-994.

TABLE 1 : Primers used for the construction of the HCoV-OC43 cDNA clone

Primer pairs and location	Sequence (5'→3')
Primers for pBAC-OC43-5'-3'	
CMV5' (7716-7736 of pBAC-TGEV-5'-3') ^{1,4}	GCGGAATTCGTGCACTTGACATTGATTATTGACTAG ²
CMV3' (8297-8316 of pBAC-TGEV-5'-3')	GCACGCAAATCGCTCACAAATACGGTTCACATAAACGAGCTC
OC43L-5' (1-41 of OC43 genome)	ATTGTGAGCGATTTGCGTGCATCCCGCTTCACTGATC
OC43L-3' (699-719 of OC43 genome) ⁴	GCGGAATTCGTGCACCCCAACCCACGAATTGACCTG
OC43R-5' (29064-29083 of OC43 genome) ⁴	GCGCAAGCTTATCTAAATTTTAAGGATGTC
OC43R-3' (30683-30713 of OC43 genome)	GTGATTCTTCCAATTGGCCATAATTAACCTC
HDV5' (30697-30746 of OC43 genome and 217-227 of pBAC-TGEV-5'-3' respectively)	CCAATTGGAAGAATCACAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAGGGTCCGGCATG
BGH3' (506-525 of pBAC-TGEV-5'-3') ⁴	GCGCAAGCTTGCTCTCCCCAGCATGCCTGC
Primers for pBAC-OC43^{FL}	
JUB1 (28-47 of OC43 genome)	CCCCTTCACTGATCTCTTG
JUBMlu reverse (6469-6491 of OC43 genome)	TCGTCCGGCGCCTCTGCTCAA <u>CGC</u> TTAGCAGTTC ³
JUBMlu forward (6469-6492 of OC43 genome)	GAACTGCTAA <u>CGC</u> TTGAGCAGAG
JUBAsiSI reverse (12127-12154 of OC43 genome)	AGGTTGGGCGCCGCTACAGC <u>CGA</u> TCGCGTTCATAAGCAG
JUBAsiSI forward (12127-12154 of OC43 genome)	CTGCTTATGAACCGGA <u>TCG</u> CGCTGTAGC
JUB105 (18973-18993 of OC43 genome)	TACAAAAGAGTCTTAACAGAC
JUB94 (18562-18582 of OC43 genome)	TAGAACTGGTTACTATGGTTG
JUBSac2 reverse (24056-24082 of OC43 genome)	GAATTGATTGT <u>CCG</u> CGGTTGTACTACC
JUBSac2 forward (24058-24082 of OC43 genome)	TAAGGTGGCGCCTAGTACAAC <u>CCG</u> GACAATCAATTC
JUO2 (30495-30514 of OC43 genome)	GCAGCAAGACATCCATTCTG

¹ The sequence of pBAC-TGEV-5'-3' was provided by Luis Enjuanes;

² Nucleotides in bold contain restriction sites and are not included in the primer location;

³ Underlined nucleotides indicate mutated bases with regards to the genome sequence of the HCoV-OC43 ATCC strain (GenBank accession number AY585228).

⁴ External primers used for the generation of overlapping PCRs.

FIGURE LEGENDS

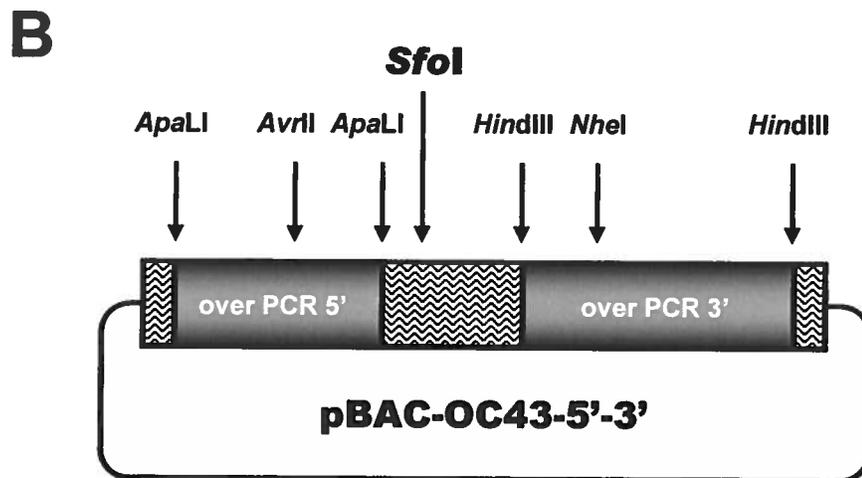
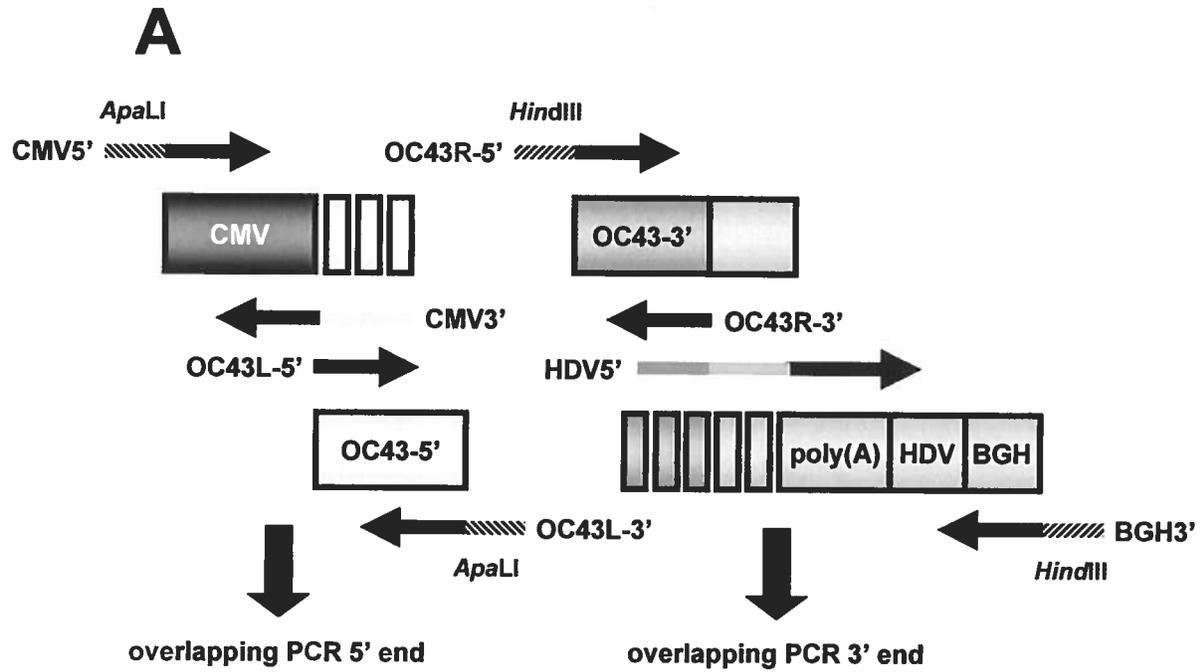
Figure 1. Assembly of the pBAC-OC43-5'-3'. (A) Schematic representation of the overlapping PCR method, used to juxtapose each viral end to accessory sequences. The viral 5' end was joined to the CMV promoter in order to generate the overlapping PCR 5' end, whereas the viral 3' end was joined to the HDV-BGH sequences to generate the overlapping PCR 3' end. In the first PCR round, two products were generated for each end using specific primers. In the second round of amplification, the overlapping extension, CMV and OC43-5' PCR products were pooled and fused as they slightly overlap each other, generating overlapping PCR 5'. The same method was used for the generation of overlapping PCR 3'. Primers used for the generation of the final products were the external ones (Table 1), which contained restriction sites for subsequent cloning into the BAC vector. (B) Schematic representation of the pBAC-OC43-5'-3'. Overlapping PCR 5' end was introduced in the *Apa*LI restriction site whereas overlapping PCR 3' end was introduced using the *Hind*III site. The vector contains the central *Sfo*I restriction site as well as *Nhe*I and *Avr*II sites, which were used to start the assembly with regions A and B respectively.

Figure 2. Assembly of the pBAC-OC43^{FL}. (A) Schematic representation of the five-step cloning strategy used for the construction of pBAC-OC43^{FL}. Primers used for the amplification of regions A to E are shown (Table 1). Regions A, B and C allowed the introduction of *Sac*II, *Mlu*I and *Asi*SI restriction sites respectively. These same regions also contained a mutagenized *Sfo*I site (diagonal stripes) whereas regions D and E contained the original *Sfo*I site, located at position 18585 of the viral genome. Other specific restriction sites, introduced by site-directed mutagenesis using primers

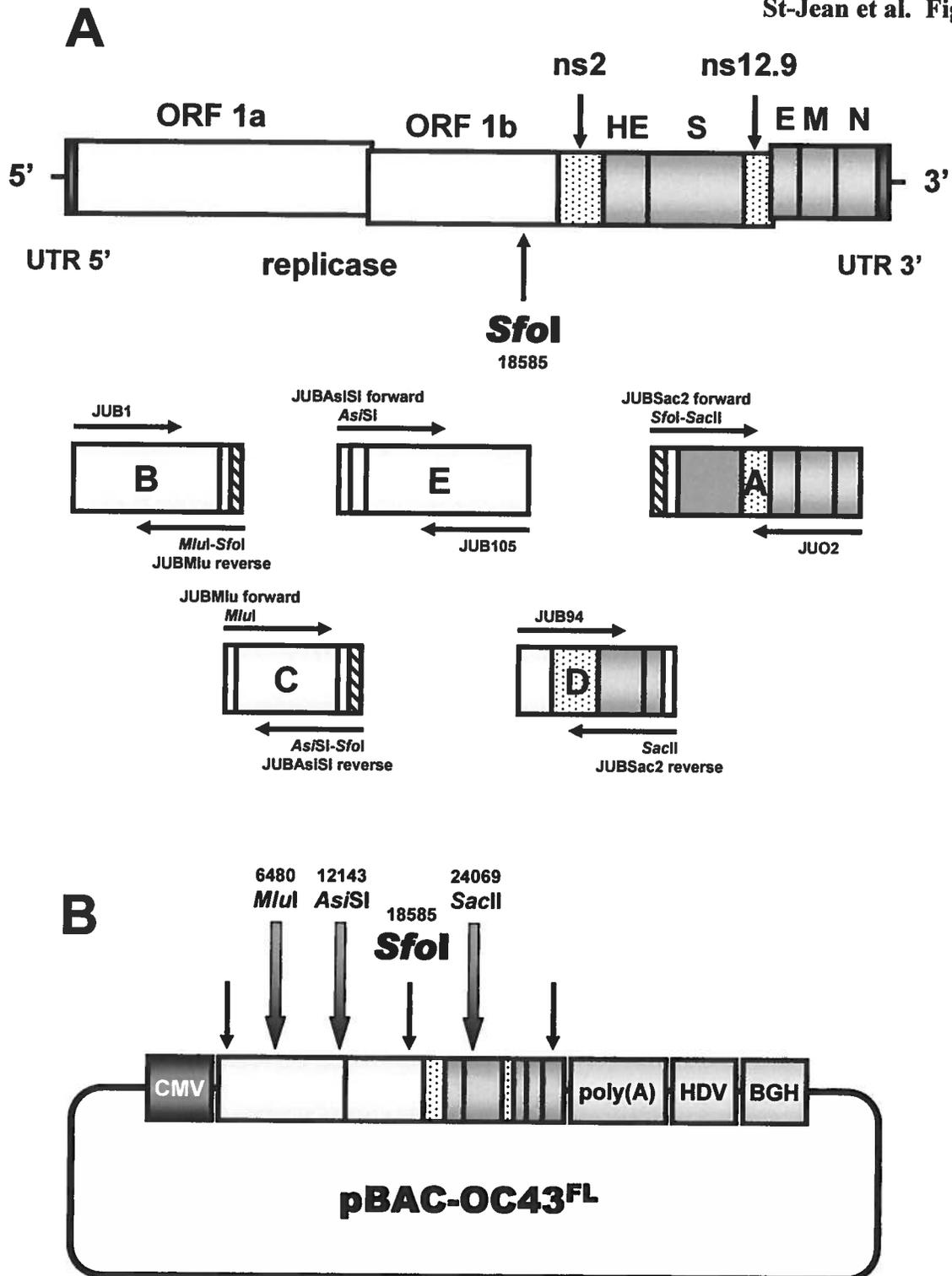
JUBSac2 forward and reverse, JUBMlu forward and reverse and JUBAsiSI forward and reverse, are shown in white. (B) Schematic representation of the full-length infectious cDNA clone. Restriction sites introduced by mutagenesis are represented by large arrows. The *SfoI* restriction site is also indicated by a small arrow. ORF, open reading frame; ns, accessory genes encoding non structural proteins; HE, hemagglutinin-esterase protein gene; S, spike protein gene; E, envelope protein gene; M, membrane protein gene; N, nucleocapsid protein gene; UTR, untranslated region; CMV, cytomegalovirus promoter; HDV, hepatitis delta virus ribozyme; BGH, bovine growth hormone termination and polyadenylation sequences.

Figure 3. *In vitro* and *in vivo* properties of the recombinant virus. (A) Viral growth kinetics of both the HCoV-OC43 ATCC strain and the recombinant HCoV-OC43 virus. Culture supernatants from HRT-18 infected cells were harvested at the indicated times post-infection. Infectious titers were determined using the immunoperoxidase assay (IPA) and are expressed in log TCID₅₀/mL. Both viruses displayed similar patterns of replication. (B) Neurovirulence of the recombinant HCoV-OC43 in mice. Two different strains of MHV-seronegative female mice, BALB/c and C57Bl/6, were inoculated with the same dose of virus.

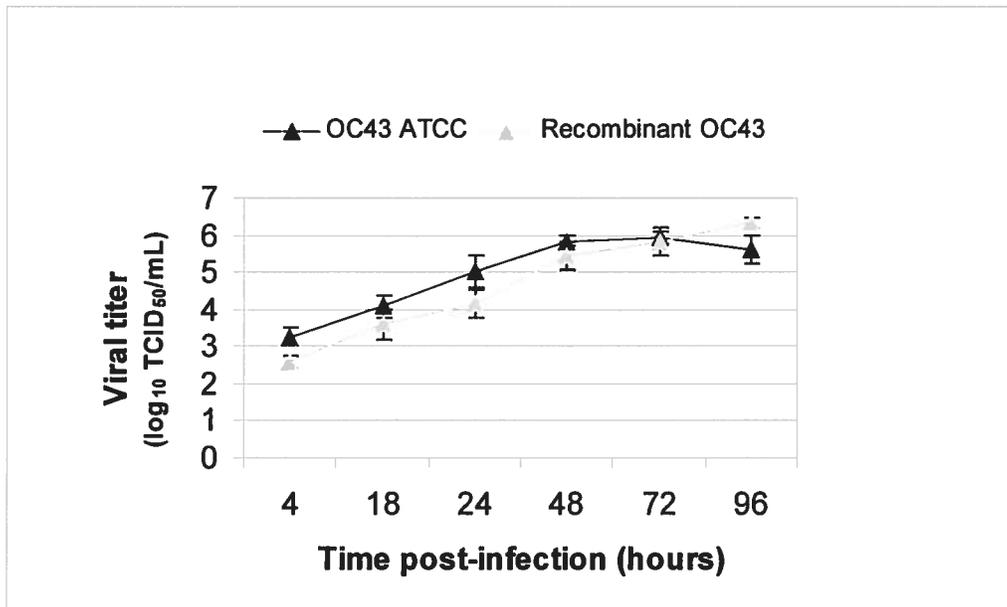
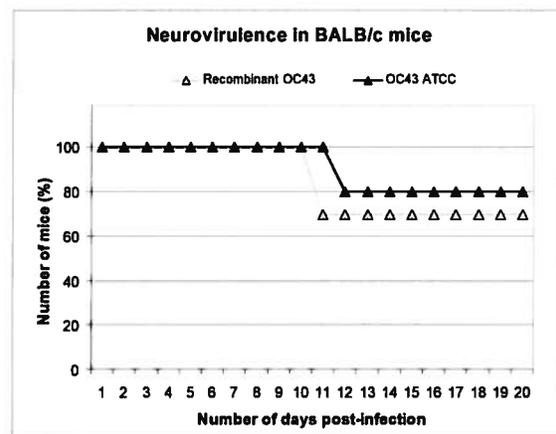
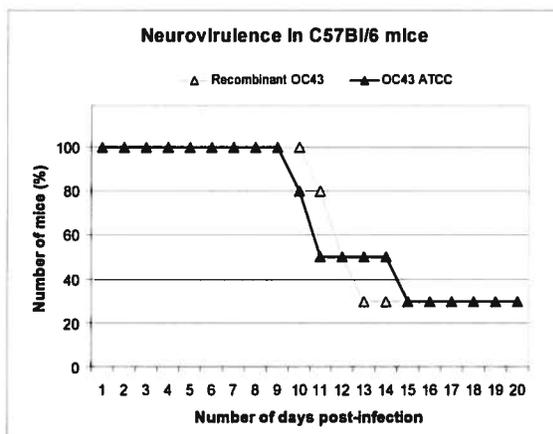
St-Jean et al. Fig. 1



St-Jean et al. Fig. 2



St-Jean et al. Fig. 3

A**B**

SUPPLEMENTAL MATERIAL

ST-JEAN ET AL.

Construction of the pBAC-OC43-5'-3'.

In order to obtain viral RNA of HCoV-OC43, a low-passage stock of the virus was first produced. HRT-18 cells were infected at a multiplicity of infection (MOI) of 0.005 with the fourth passage of HCoV-OC43 ATCC strain. Following infection, total mRNA was extracted using the Dynabeads[®] mRNA DIRECT™ Kit (DYNAL Biotech) as recommended by the supplier. The RNA was then quantified and reverse transcribed using the SuperScript™ First-Strand Synthesis System for RT-PCR (Invitrogen) and specific primers JUNSO2 and JUO2, according to the region to amplify.

To generate pBAC-OC43-5'-3', four polymerase chain reaction (PCR) amplifications were required and performed using primers listed in Table 1. Each set of primers (CMV, OC43L, OC43R and HDV-BGH) was used in combination with the Expand High Fidelity PCR System (Roche) according to the manufacturer's instructions, at an annealing temperature of 60°C for 35 cycles. The vector pBAC-TGEV-5'-3' was used to amplify the CMV, HDV and BGH sequences, whereas viral cDNA was used for amplifications of both 5' and 3' ends. In order to join the CMV promoter with the 5' viral end and the 3' viral end with the HDV-BGH sequences in a contiguous and very precise way, overlapping PCRs were performed. Amplicons CMV and OC43L were then pooled together at a 1:1 molar ratio, with primers CMV5' and OC43L-3' in order to generate overlapping PCR 5', as were amplicons OC43R and HDV-BGH with primers OC43R-5' and BGH3', to generate overlapping PCR 3'. Both overlapping PCR amplicons were then purified by gel extraction using the QIAEX[®] II Gel Extraction Kit (Qiagen), digested with the appropriate enzyme and directly introduced into the BAC. The overlapping PCR 5' amplicon was first inserted into the vector, using the *Apa*LI restriction site in order to generate the pBAC-OC43-5',

whereas the overlapping PCR 3' amplicon was introduced into the latter vector using the *Hind*III restriction site to generate the pBAC-OC43-5'-3'. After each cloning step, the new BAC vector was electroporated into DH10B cells, extracted and purified using the QIAGEN® Large-Construct Kit (Qiagen), as recommended by the manufacturer, and sequenced to make sure no mutations were introduced into the new construct.

Construction of the full-length HCoV-OC43 cDNA clone.

The second part of the experimental strategy used for the construction was a five-step cloning approach. The *Sfo*I restriction site, which was located in the BAC polylinker as well as in the nsp11 motif of ORF1b, was used for all five cloning steps and played a central role in the construction of pBAC-OC43^{FL}. Viral genomic RNA was first reverse transcribed as previously described and amplified using primers listed in Table 1. These primers allowed the introduction of three new restriction sites into the genome sequence: *Mlu*I, *Asi*SI and *Sac*II at positions 6480, 12143 and 24069 respectively. Two bases were mutated for each new restriction site. The mutated nucleotides were therefore T6479C, T6482G, C12143T, T12146C, A24068G and T24071G. The genome was amplified in five distinct regions, named A to E. Some of the primers, such as JUBSac2 forward, JUBMlu reverse and JUBAsiSI reverse, also allowed the insertion of a *Sfo*I site at one end of fragments A, B and C respectively, which could be used at each cloning step and allowed the sequential cloning of these fragments in the *Sfo*I site of pBAC-OC43-5'-3'. Each amplicon was generated using the PfuUltra™ High-Fidelity DNA Polymerase (Stratagene) as recommended by the manufacturer for larger targets and then purified using the Strataprep™ DNA Gel Extraction Kit (Stratagene). Purified amplicons A to E were then inserted into the pSMART™ HCKan vector (Lucigen® Corporation) in order to generate plasmids pSMART-A to -E and transformed as described above. Positive pSMART clones were extracted using the GenElute™ High Performance (HP) Plasmid Maxiprep Kit (Sigma-Aldrich), as recommended by the manufacturer. Each new vector was sequenced to make sure no mistakes

were introduced into the clones, except for mutations allowing the introduction of new restriction sites.

Viral DNA regions contained in clones pSMART-A to -E were used to fill the pBAC-OC43-5'-3' and to complete the sequence of the HCoV-OC43 genome. Prior to each cloning step, pSMART clones were therefore digested with a combination of two restriction enzymes, one being a specific enzyme and the other always being *SfoI*. Likewise, the BAC vector was digested with *SfoI* and the corresponding specific enzyme. Following digestion, the BAC vectors were dephosphorylated using 2 U of Shrimp Alkaline Phosphatase (Roche) according to the manufacturer's instructions, without gel purification, whereas digested fragments from the pSMART clones were gel purified using the QIAEX[®] II Gel Extraction Kit (Qiagen) and quantified before ligation with T4 DNA Ligase (Invitrogen). Again, following each cloning step, the new BAC vector was electroporated in DH10B cells, extracted and purified using the QIAGEN[®] Large-Construct Kit (Qiagen). Junction sites created in each new BAC constructs were sequenced to make sure no mutations were introduced.

Transfection of pBAC-OC43-5'-3' and pBAC-OC43^{FL} and recovery of recombinant HCoV-OC43 particles.

Transfections were performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Briefly, BHK-21 cells were plated at 8×10^5 cells in a 6-well plate and, 24 hours later, transfection of pBAC-OC43^{FL} was performed using 5 μ g of DNA mixed with 12 μ l of Lipofectamine 2000 in OPTIMEM medium (Invitrogen) supplemented with 10% (v/v) FCS for 72 hours. Afterwards, the OPTIMEM medium was replaced by standard MEM medium supplemented with 1% (v/v) FCS and transfected cells were incubated at 33°C for 3 days in a 5% (v/v) CO₂ humid atmosphere, which represents the standard protocol for optimal production of infectious HCoV-OC43 virus particles. Supernatants containing recombinant viral particles were

clarified, quantified by the immunoperoxidase assay (IPA) and aliquoted for downstream experiments.

Viral growth kinetics in cell culture.

The HRT-18 cell line was used for the viral growth kinetics. Briefly, cells were plated into 12-well plates and infected in triplicates at a MOI of 0.2 with either the HCoV-OC43 ATCC laboratory strain or the recombinant HCoV-OC43 virus, when cell confluence reached 70%. The viral growth kinetics were performed at the same time for both viruses but using separate laminar flow hoods to prevent any possible cross-contamination. Supernatants containing virus particles were harvested at different times post-infection and virus titers were determined using the IPA. The rhabdomyosarcoma cell line RD-151 was used to perform plaque assays with both the recombinant virus and the laboratory strain.

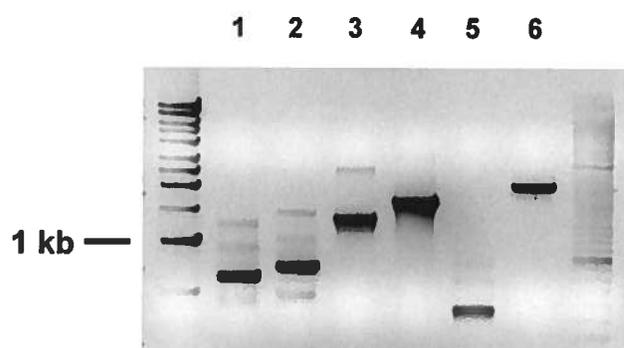
Virulence assays.

In order to compare the susceptibility of mice to HCoV-OC43 ATCC and to the recombinant HCoV-OC43 virus, two different strains of MHV-seronegative female mice, BALB/c and C57Bl/6 (Jackson Laboratories, Bar Harbor, ME, U.S.A.), were inoculated with the same amount of virus. As previously described (Reference 8: Jacomy and Talbot, 2003), 21 days post-natal (DPN) mice were infected with an intra-cerebral (IC) inoculation of 10 μ L, containing 10 TCID₅₀ for C57Bl/6 and 10,000 TCID₅₀ for BALB/c mice, of either HCoV-OC43 stock. The viral dose was administered under deep anesthesia of ketamine-xylamine (ketamine at 200mg/kg and xylamine at 10mg/kg). Twenty mice of each strain were used to establish survival curves.

FIGURE LEGEND

Figure S1. Assembly of pBAC-OC43-5'-3'. DNA agarose gel containing the four intermediate PCR products and the two final overlapping PCR products. Left lane, KiloBase DNA Marker (Amersham Biosciences); 1, CMV PCR product; 2, OC43-5' PCR product; 3, overlapping PCR 5' end; 4, OC43-3' PCR product; 5, HDV-BGH PCR product; 6, overlapping PCR 3' end; right lane, 100 Base-Pair Ladder (Amersham Biosciences).

St-Jean et al. Fig. S1



CHAPITRE VI : PUBLICATION no.5 :

Mutations in the S gene of human coronavirus OC43 isolated from persistently infected neural cells correlate with adaptation to the target host cell environment

**MUTATIONS IN THE S GENE OF HUMAN CORONAVIRUS OC43 ISOLATED FROM
PERSISTENTLY INFECTED NEURAL CELLS CORRELATE WITH ADAPTATION
TO THE TARGET HOST CELL ENVIRONMENT**

Julien R. ST-JEAN ¹, H el ene JACOMY ¹, Marc DESFORGES ¹,
and Pierre J. TALBOT ^{1*}

¹Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier,
531 boulevard des Prairies, Laval (Qu ebec), CANADA H7V 1B7

* Corresponding author:

Dr. Pierre J. Talbot
Laboratory of Neuroimmunovirology
INRS-Institut Armand-Frappier
531 boulevard des Prairies,
Laval (Qu ebec)
CANADA
H7V 1B7

Tel.: (450) 686-5515
Fax: (450) 686-5566
E-mail: pierre.talbot@iaf.inrs.ca

Word counts:

Abstract: 241

Text: 5415

RÉSUMÉ

La glycoprotéine de surface (S) des coronavirus est impliquée dans la liaison au récepteur et l'entrée virale, contribuant ainsi à l'adaptation du virus à son environnement. Dans cette étude, des infections persistantes de lignées cellulaires neurales ont été effectuées afin de vérifier si la persistance du virus est à l'origine des mutations retrouvées dans les gènes S, HE ou N et pour vérifier si ces mutations pouvaient être reliées à l'adaptation en culture cellulaire. Les virions produits lors d'infections persistantes de cellules neurales ont été isolés et purifiés par centrifugation sur gradient et l'ARN génomique viral a été séquencé. Plusieurs mutations ont été identifiées dans le gène S, contrairement aux gènes HE et N. La conservation des mutations dans S aux passages subséquents suggère que celles-ci confèrent un avantage adaptatif ainsi qu'un phénotype stable au virus, assurant ainsi une adaptation adéquate à l'environnement cellulaire de l'hôte. Quatre mutations dans S se sont avérées prédominantes et ont été retrouvées dans plusieurs infections persistantes différentes (D24Y, S83T, H183R and Y241H). Afin d'analyser le rôle des mutations de S sur le tropisme viral et la réplication, des tests de modulation de tropisme et d'infectivité ont été effectués. Parmi les variants isolés des infections persistantes, cinq ont été sélectionnés pour leur tropisme cellulaire accru et une augmentation de leur réplication. Contrairement à la souche de référence HCoV-OC43 VR-759, ces variants pouvaient tous former des plages de lyse en culture cellulaire, et deux d'entre eux présentaient une modulation de tropisme des neurones vers les astrocytes dans des cultures de cellules primaires neurales murines. Ces résultats suggèrent que les mutations retrouvées dans le gène S du HCoV-OC43 sont impliquées dans l'évolution virale et l'adaptation du virus à son environnement cellulaire, ce qui pourrait résulter en une modulation de la virulence et de la neuropathogénèse.

ABSTRACT

The spike (S) glycoprotein of coronaviruses is involved in receptor binding and viral entry, which can contribute to viral adaptation to its environment. In this study, persistent infections of neural cell lines were performed to verify whether viruses carrying mutations in the S, HE or N gene originated from viral persistence and to determine whether these mutations could be related to adaptation in cell culture. Virions released from persistently infected neural cells were isolated and purified by gradient centrifugation and genomic RNA was sequenced. Several S gene mutations were identified, unlike the HE and N genes. Conservation of S mutations in subsequent passages suggested that they confer an adaptive advantage and a stable phenotype to the virus, thereby ensuring proper adaptation to the target host cell environment. Four S mutations were predominant and found in several different persistent infections (D24Y, S83T, H183R and Y241H). To investigate the role of these S mutations on viral tropism and replication, assays for modulation of tropism and infectivity were performed. Among the virus variants isolated from persistent infections, five were selected for their extended cellular tropism and their increased replication titers. In contrast to the HCoV-OC43 ATCC reference strain VR-759, these variants could all form plaques in cell culture, and two of them showed a modulation of tropism from neurons to astrocytes in murine primary neural cell cultures. These results suggest that mutations in the S gene of HCoV-OC43 are involved in viral evolution and adaptation to its target host cell environment, which could result in modulation of viral virulence and neuropathogenesis.

Keywords: Human coronavirus; coronavirus; OC43; HCoV-OC43; persistence; viral mutants; S gene; mutations; cell culture; neural cells; tropism; adaptation

INTRODUCTION

Human coronaviruses (HCoV) are mainly associated with respiratory infections and are responsible for up to one third of common colds (Myint, 1994). In the past years, we have provided experimental evidence that the OC43 strain possesses neurotropic and neuroinvasive properties. Indeed, we have shown that it infects human primary cells from the central nervous system (CNS) (Bonavia et al., 1997) and persists in human neural cell lines (Arbour et al., 1999) and in human brains (Arbour et al., 2000). Although we have also provided recent evidence that suggests that HCoV-OC43 could remain relatively genetically stable in the environment when placed under certain selection pressure (St-Jean et al., 2004), it is also known that coronaviruses can adapt in cell culture or under a given selection pressure, for instance related to evasion from the immune system.

Like SARS-CoV, the causative agent of severe acute respiratory syndrome (SARS) (Peiris et al., 2003; Skowronski et al., 2005), HCoV-OC43 belongs to the second genetic group of the *Coronaviridae* (Gorbalenya et al., 2004; Snijder et al., 2003). The HCoV-OC43 ATCC strain possesses a positive-stranded RNA genome of 30,713 nts, excluding the poly(A) tail (St-Jean et al., 2004). This genome comprises genes encoding structural and non-structural proteins. The non-structural protein genes are mostly contained within the 5' two thirds of the genome and allow the synthesis of the replicase polyprotein. The processing of the latter polyprotein yields to the generation of key replicative functions, such as the main protease (3CLpro), the RNA-dependant RNA polymerase (RdRp) and the RNA helicase (Ziebuhr et al., 2000). The structural protein genes are clustered in the last 3' third of the genome. They include the hemagglutinin-esterase protein gene (HE), the surface or spike protein gene (S), the small envelope protein gene (E), the membrane protein gene (M) and the nucleocapsid protein gene (N). Among these gene products, the S glycoprotein is biologically particularly important. Indeed, for other coronaviruses, the S glycoprotein was shown to be involved in attachment to cellular receptors

(Babcock et al., 2004; Hofmann et al., 2005; Lachance et al., 1998; Lai and Cavanagh, 1997; Saeki et al., 1997) as well as in viral entry since it promotes membrane fusion between the virus and the target cell (Bosch et al., 2003; de Haan et al., 2004; Duquerroy et al., 2005; Guillén et al., 2005; Miura et al., 2004; Sainz Jr. et al., 2005; Taguchi and Shimazaki, 2000). For these reasons, the S protein is likely to be involved in the determination of coronavirus tropism (Almazán et al., 2000; Haijema et al., 2003) and its modulation (Schickli et al., 1997) and therefore have a direct impact on virulence and pathogenesis. Likewise, the S protein could be associated with the capacity of the virus to reach the CNS and trigger neurological disorders (Arbour et al., 2000; Jacomy and Talbot, 2003; Jacomy et al., 2006).

Viruses harbouring mutations in the S gene, either natural or experimentally induced, have been extensively used for the characterization of the importance of the S protein as a coronavirus virulence factor. For instance, a single amino acid substitution in the S glycoprotein was reported to influence the biology of the entire virus (Leparc-Goffart et al., 1997; MacNamara et al., 2005; Ontiveros et al., 2003; Shen et al., 2004; Tsai et al., 2003; Yi et al., 2005). Previous studies have also focused on the S gene by the use of deletion or substitution mutants in order to characterize various viral properties, such as induction of demyelination (Fu et al., 2004; Sarma et al., 2000), viral entry and tropism (Almazán et al., 2000; Kuo et al., 2000; Matsuyama and Taguchi, 2002; Navas-Martin et al., 2005; Schickli et al., 2004; Thackray and Holmes, 2004; Thackray et al., 2005) and virulence (Ontiveros *et al.*, 2003), and, more recently, virus-induced fusion (Follis et al., 2005). All these data point to the importance of the S glycoprotein on the virulence and biology of coronaviruses.

In the present study, persistent infections of human neural cell lines were performed to verify whether viruses carrying mutations in the S, HE or N gene originated from viral persistence and to determine whether these mutations could be related to adaptation in cell culture. Cell lines MO3.13, H4, U-87 MG and TE-671 were therefore persistently infected in order to identify

potential mutations in the S genes that could be associated to viral adaptation, which could have arisen in cell culture. The HE and N genes were targeted in parallel to provide a comparative basis. Viral particles released from infected neural cell lines were isolated and purified, and genomic RNA was sequenced. Several mutations from different persistent infections were identified in the S gene and only a few were found in HE and N genes, suggesting that the S gene is responsible for adaptation to the target host cell environment, which could be associated with neurotropism, neuroinvasion, and presumably neurovirulence and neuropathogenesis. Of these mutations, four (D24Y, S83T, H183R and Y241H) were predominant as they were found in several different persistent infections. It is noteworthy that these four mutations were all located within the putative receptor-binding site. Following assays for modulation of tropism and infectivity, which were carried out using cell lines originating from various human tissues and animal species for which susceptibility to HCoV-OC43 infection was first determined, variants U-87 MG P35, H4 P47, H4 P56.3, H4 P116.1 and H4 P116.2 were selected for further studies given their extended cellular tropism and increased replication titers. The replication of variants H4 P116.1 and U-87 MG P35 was studied in primary murine neural cell cultures and shown to display phenotypes that diverged from the HCoV-OC43 ATCC reference strain, with regards notably to cellular tropism. These results suggest that HCoV-OC43 can adapt to its target host cell environment *in vitro* through virus-host cell interactions involving the S glycoprotein.

RESULTS

Generation of HCoV-OC43 variants

In this study, neural cell lines MO3.13, H4, U-87 MG and TE-671 were persistently infected in order to identify mutations in the HE, S and N genes that may relate to virus adaptation to its cellular environment. The non-neural HRT-18 cell line was introduced in the study because it replicates the HCoV-OC43 very efficiently and could therefore be used as a control. One persistent infection per cell line was carried out for the experiment, except for the H4 cell line, for which four persistent infections were performed, three of them being performed in parallel in order to monitor the kinetics of introduction of mutations in viral genes. Several virus purifications were performed for each persistent infection, at randomly selected passages, which led to the generation of the following virus variants: HRT-18 passage 33 (P33), P54, P110 and P155; H4 P47 and P90; H4 P56.1, P56.2, P56.3, P116.1, P116.2 and P116.3; TE-671 P38 and P79; U-87 MG P35 and MO3.13 P5, P6 and P22. Given that subgenomic mRNAs could contain unwanted mutations introduced by the viral polymerase, virus purifications were performed each time to ensure that the analyzed sequences were from genomic viral RNA rather than subgenomic intracellular viral RNA which may not be packaged into mature virions (Escors et al., 2003; Arbour et al., 1999).

All 18 virus variants generated following persistent infections and purifications had their HE, S and N genes amplified and sequenced, and a representative example is shown in Fig. 1. Primers used for the amplification are described in Table 1. The S gene was amplified using two sets of primers, specific to the S1 and S2 subunits. During PCR, several controls were used to monitor purification as well as amplification efficiencies. Primers specific to the cellular gene GAPDH (lane 6) and viral mRNAs (lane 7) were used to monitor the presence of contaminants. Although small proportions of the GAPDH gene were detected in most of the viral purifications, no viral

mRNAs were detected, indicating that the achieved level of purification was appropriate. Controls of mRNA amplifications were performed on non-purified virus, illustrating that primers were functional and specific. Such an amount of the GAPDH gene could be explained by the very high sensitivity of the primers, which were used in a one-shot assay, at an annealing temperature lower than their fusion temperature. Primers specific to the 5' end of the replicase gene (lane 5) were also used to make sure that the analyzed sequences stemmed from genomic RNA, since this gene is only present on the full-length viral genome (Ziebuhr et al., 2000).

Several mutations from different persistent infections were identified in the S gene (115 mutations in eight persistent infections; Fig. 2) and only a few were found in HE and N genes (10 and 14 mutations respectively; data not shown), suggesting that the S protein gene is the main factor responsible for viral adaptation to the cellular environment. It is noteworthy that no mutations were found in the HE, S and N genes of viruses that stemmed from acutely (four days without any passage) infected cell lines (HRT-18, MO3.13 and H4) when compared to our reference strain (GenBank accession number AY585228). This result confirms that adaptation did not arise upon acute infection and that it required a long-term process associated with viral persistence. Mutations found throughout the S gene were mostly concentrated in three regions : the putative receptor binding site (RBS), taking mouse hepatitis virus (MHV) as a model (Matsuyama and Taguchi, 2002), the putative hypervariable region (HVR) (Gélinas et al., 2001) and regions involved in fusion processes, located in the S2 subunit, that are the fusion peptide (FP) as well as the heptad repeats (HR) 1 and 2 (Bosch et al., 2003) (Fig. 2). Among the observed mutations, four were predominant and were found in several persistent infections, and were located within the putative RBS (D24Y, S83T, H183R and Y241H). Interestingly, most of the mutations found throughout the S gene were located in or close to regions that are thought to be involved in specific viral functions, such as receptor binding and/or viral fusion (Fig. 2).

Assays for viral susceptibility and modulation of tropism and infectivity

Prior to performing assays for possible modulation of tropism and infectivity, susceptibility of various cell lines to infection by the HCoV-OC43 ATCC strain was determined (Table 2). Cell lines used for this experiment originated from various human tissues as well as from various animal species. Infectious viruses were detected using an immunoperoxidase assay (IPA), which is very sensitive and allows the direct detection of infectious viral particles released following a productive infection.

Criteria used to determine the susceptibility level of each cell line were established by performing viral growth kinetics using different cell lines (data not shown). Growth curves obtained from different cell lines allowed us to determine thresholds of susceptibility. Cell lines in which virus titers were clearly increasing were considered susceptible to infection. Likewise, cell lines for which viral titers reached between $10^{3.0}$ and $10^{4.0}$ TCID₅₀/mL were considered to have a low susceptibility level, whereas cell lines from which viral titers were $10^{3.0}$ TCID₅₀/mL or under were deemed not susceptible. Cell lines for which the susceptibility level to HCoV-OC43 was ambiguous (MT4, Raji, U937 and HL-60) were further characterized by performing viral growth kinetics. Overall, all cell lines were not susceptible to HCoV-OC43 infection, except for the B104 cell line, which displayed a low susceptibility level, and for the BHK-21 cell line, which was fully susceptible (Table 2).

Given this knowledge of “baseline” susceptibility of these cell lines to infection by the reference ATCC strain of HCoV-OC43, assays for modulation of tropism and infectivity were first carried out using the 18 virus variants generated during viral persistence. Preliminary results obtained from these analyses allowed us to select five variants for which replication titers were increased and cellular tropism was extended: U-87 MG P35, H4 P47, H4 P56.3, H4 P116.1 and H4 P116.2. The same analysis was then performed in triplicates using the five latter viral variants and assays

for modulation of tropism and infectivity were again performed (Table 2). A modulation of tropism was observed when a non-susceptible cell line became fully susceptible, whereas a modulation of infectivity occurred when a cell line, susceptible or not, displayed an increase of at least $10^{2.5}$ TCID₅₀/mL in the infectious viral titer. Obviously, most of the cell lines displaying a modulation of tropism also displayed a modulation of infectivity. Briefly, results from this analysis varied according to the cell lines and viral variants tested. For instance, cell lines L132 and B104 showed, respectively, either a modulation of tropism or a modulation of infectivity for all variants. Variants H4 P116.1 and H4 P116.2 displayed the highest adaptation level, showing a modulation of tropism for ten cell lines each.

In order to characterize viral properties related to adaptation, the same five variants were assessed for their capacity to produce plaques. All five variants produced plaques of distinct morphologies on RD-151 cells (Fig. 3). The effects of the different S mutations were reflected by plaque morphologies. For instance, variants U-87 MG P35 and H4 P56.3 formed very diffuse plaques, in contrast to plaques from variants H4 P116.1 and H4 P116.2, which were much brighter than the others. Plaques from H4 P47 were also diffuse and particularly pale.

Infection of primary mouse neural cell cultures

Because of the mutations they harboured in the S gene, variants U-87 MG P35 and H4 P116.1 were selected for characterization of infection in primary mouse neural cell cultures. These variants were selected for the following reason: variant U-87 MG P35 harboured only the four predominant mutations found in the putative RBS (D24Y, S83T, H183R and Y241H), whereas variant H4 P116.1 contained 21 mutations, including these four predominant mutations and a series of 13 mutations located in the S2 subunit, in proximity to regions involved in fusion processes (Fig. 2).

At 24 hours post-infection, immunofluorescent staining revealed that neurons were the main target of the HCoV-OC43 ATCC reference strain (Fig. 4, Panels A and C), whereas both neurons and astrocytes were positive for viral antigens following U-87 MG P35 (Fig. 4, Panel B) or H4 P116.1 infections (Fig. 4, Panel D), illustrating an enlargement of tropism for these two variants. At 48 hours post-infection, cells infected by variant H4 P116.1 began to aggregate and infected cells were more densely packed (Fig. 5, Panel B) compared to infection by the HCoV-OC43 ATCC reference strain, in which neurons and astrocytes remained equally distributed (Fig. 5, Panel A and B). Formation of cell aggregates (syncytium-like structures) was observed for variant H4 P116.1 (Fig. 5, Panels B, D and F) whereas no cell fusion was observed for the HCoV-OC43 ATCC strain (Fig. 5, Panels A, C and E). At four days post-infection, fusion of infected glial cells and GFAP disruption were more evident following H4 P116.1 infection (Fig. 6). Infected neurons conserved a well-defined morphology on the top of the syncytium-like structures (Fig. 6, Panels A to C) whereas GFAP was barely detectable within the main body of the syncytium-like structure (Fig. 6, Panels D to F). Residual GFAP seemed to aggregate around the nucleus (Fig. 6, Panel G) in infected cells, which were still presenting a large cytoplasm containing viral antigens (Fig. 6, Panel H). The disrupted GFAP network appeared yellow in merged pictures (Fig. 6, Panel I).

Infectious viral titers obtained from supernatants following infection with HCoV-OC43 ATCC revealed that primary murine cell cultures produced a significant progeny (Fig. 7). Likewise, variants H4 P116.1 and U-87 MG P35 were also able to replicate in these cell cultures. In all cases, replication levels increased during the first days of infection, reaching a maximum at two days post-infection. However, although both variants were also able to infect the mouse primary cells, replication levels were lower with regard to the HCoV-OC43 ATCC strain (Fig. 7). This phenomenon could be explained by the fact that infection of primary mouse astrocytes by HCoV-OC43 ATCC was less productive than the infection of neurons (Jacomy et al., 2006).

DISCUSSION

In the current study, persistent HCoV-OC43 infections were carried out in cell lines representative of the CNS (MO3.13, H4, U-87 MG and TE-671) in order to monitor the introduction of mutations in genes potentially involved in viral adaptation to its environment, which could arise *in vitro* or *in vivo*. Following persistent infections of the cell lines, variants containing mutations in the HE, S and N genes were obtained. Of the 139 identified mutations, 115 were found in the S gene (eight of them being silent mutations), suggesting that this gene is a main factor responsible for viral adaptation to the target host cell environment. This is consistent with the known functions of this protein and the fact that it is more sensitive to sequence variations than all other coronavirus proteins (Wu and Yan, 2004). Evidence linking the S glycoprotein to viral adaptation has also been reported for several other coronaviruses, such as infectious bronchitis virus (IBV) (Fang et al., 2005) and SARS-CoV (Holmes, 2005; Li et al., 2005).

In contrast to the S gene, only a few mutations appeared in the HE and N genes. Although these genes are not known for their capacity to modulate viral adaptation, we cannot rule out the possibility that they harbored mutations which could have an influence on viral biology. For instance, mutations of the N gene could affect replication and transcription levels (Almazán et al., 2004; Schelle et al., 2005). Likewise, we cannot rule out the possibility that other mutations could be found elsewhere in the viral genome, since only the HE, S and N genes were sequenced for this study. Nevertheless, our results are indeed consistent with the critical involvement of the S glycoprotein in viral adaptation.

While keeping in mind the possible importance of concurrent mutations on viral phenotype, and the need for structure-function studies making use of our recently developed infectious clone (St-Jean et al., 2006) to definitely associate mutations with phenotype, the current study does allow very interesting observations regarding appearance of mutations in the S gene during viral

persistence, especially in neural cells. The viral clonality provided by the infectious cDNA clone will also allow us to study the impact of the S mutations in a pure viral population harboring only these selected mutations.

Among the mutations identified in the S gene, four were found to be predominant: D24Y, S83T, H183R and Y241H. With the exception of mutation S83T, these predominant mutations were exclusive to neural cells. They were all located in the putative RBS and could potentially have an influence on virus interaction with the cellular receptor. Importantly, most of the mutations were found within functional domains of the S glycoprotein (RBS, FP and HR) or close to them, as well as in the HVR (Fig. 2). Although the S glycoprotein of HCoV-OC43 has not yet been characterized precisely, murine hepatitis virus (MHV) as well as bovine coronavirus (BCoV) were used as models for the mapping of the putative functional domains. Based on the fact that the BCoV HVR is located between amino acids 452 and 593 (Gélinas et al., 2001) and that HCoV-OC43 displayed 95% identity with the latter virus (St-Jean et al., 2004), it was assumed that the HCoV-OC43 HVR was similarly localized. Likewise, given the MHV model, it was assumed that the RBS was located at the amino-terminal end of the S glycoprotein (Matsuyama and Taguchi, 2002) and that the FP, HR1 and HR2 domains were respectively located at amino acids 960 to 980 (Duquerroy et al., 2005), 981 to 1084 and 1245 to 1295 (Bosch et al., 2003).

We next aimed to characterize the impact of the observed acquisition of S gene mutations during viral persistence with viral biology. Assays for modulation of viral tropism and viral infectivity were performed using five variants selected for their extended cellular tropism and their increased replication titers in different established human cell lines: U87-MG P35, H4 P47, H4 P56.3, H4 P116.1 and H4 P116.2 (Table 2). Among these variants, H4 P116.1 and H4 P116.2 displayed the most enlarged tropism, with an observed modulation for ten cell lines each, plus a modulation of infectivity for two cell lines for H4 P116.1. Both variants contained very similar patterns of mutation in the putative RBS, except for the predominant mutation S83T, which was only

harboured by H4 P116.1. Except for the V980A mutations, no other mutations found outside the RBS were shared by both variants. This could explain the slight differences observed in tropism. Mutation V980A, which was present in H4 P116.1 and H4 P116.2 and was never found in other variants, is located within the putative FP, which could account for the high adaptation level of these two variants. However, other variants containing mutations in the FP, such as H4 P90 (A965V and T975A), H4 P116.3 (W974L), TE-671 P38 and P79 (P972L, P973S and A978S) and MO3.13 P6 and P22 (P973S) did not display such a phenotype. Although variant H4 P116.1 also contained two more mutations in the FP (W974L and T975P) as well as one mutation in the HR1 (P1249L), a correlation between enlarged viral tropism and the presence of these mutations would be speculative. Importantly, although they did not all contain mutations in functional domains of the S2 subunit, all five variants analyzed during these assays were shown to form plaques, unlike the HCoV-OC43 ATCC reference strain. Mutations located around these domains could have an effect on the folding of the protein (Bosch et al., 2003; Follis et al., 2005; Guillén et al., 2005) or on the interaction with other molecules (Bosch et al., 2005; Ingallinella et al., 2004), and possibly influence the capacity of the virus to form plaques *in vitro* (Hajjema et al., 2003).

Although it is known that the S protein of HCoV-OC43 ATCC does not necessarily undergo cleavage during viral entry in the infected cell (Künkel and Herrler, 1996), the putative site which would be responsible for this process is ${}_{754}\text{RRSRG}_{758}$. Although it has been reported that proteolytic cleavage of the S protein could activate some viral properties, such as cell-cell fusion (de Haan et al., 2004), the fusion properties of our variants are not likely to be influenced by a cleavage enhancement. Indeed, experiments performed with variant HRT-18 P110, which harbors a mutation in the putative cleavage site, R757H, showed that there was no difference in the CPE generated (cell-cell fusion) by this variant with regard to the HCoV-OC43 ATCC strain (data not shown).

It is not known whether observed enlargements of tropism are to be linked to a modification of the interaction between S and the receptor on the cell surface, to a more efficient replication level of the variants, or both. However, the hypothesis regarding the interaction with the receptor is supported by the fact that all five variants displaying an enlarged tropism, when compared to the 13 others, harboured the combination of mutations H183R-Y241H. Being located in the putative RBS, these mutations could influence receptor attachment, and possibly also influence fusion properties since it has been reported that RBS residues could interact with the S2 subunit and act on fusion properties (Matsuyama and Taguchi, 2002).

Several S gene mutations reported in various studies have been shown to modulate coronaviral biology. Several of the mutations in the S1 subunit were associated with a modification of tropism (Schickli et al., 2004) or a defect in viral entry (Yi et al., 2005), whereas other mutations were associated with neurovirulence and demyelination in MHV (Fu et al., 2004; Leparc-Goffart et al., 1997). This is particularly the case for mutation Q159L of MHV-A59, for which we identified a mutation at a homologous position in the HCoV-OC43 genome. This E170K mutation could therefore have an impact on the *in vivo* virulence of variant H4 P116.1. The S gene sequences of variants were compared with sequences obtained from Künkel and colleagues (Künkel and Herrler, 1996), for their strains CU (grown in MDCK I cells) and VA (adapted to Vero cells). From the 92 amino acid substitutions observed among HCoV-OC43 variants, many were confirmed in strains CU and VA. It is noteworthy that, among them, predominant mutations S83T and H183R, as well as a mutation at position 24, were identified. Whether similar HCoV-OC43 mutations and consequences on viral adaptation would be found *in vivo* remain to be elucidated. Therefore, we used an experimental cell culture model that is further related to an *in vivo* system: infections of primary mouse neural cell cultures.

Variants selected for this experiment were H4 P116.1 and U-87 MG P35. Unlike variant U-87 MG P35, which only possessed the four predominant mutations in the putative RBS, H4 P116.1

possessed the same four mutations as well as several other mutations, most of them located within or close to the functional domains of the S2 subunit. Whereas neurons were clearly the main target of the HCoV-OC43 ATCC reference strain, as we have previously reported (Jacomy et al., 2006), variants H4 P116.1 and U-87 MG P35 were shown to infect both neurons and astrocytes (Fig. 4). This result suggested that some of the mutations in the putative RBS domain, present in these two variants, could account for the observed switch of tropism. Unlike variant U-87 MG P35 and HCoV-OC43 ATCC, the H4 P116.1 variant induced the formation of cell aggregates (syncytium-like structures), suggesting fusion of infected glial cells (Fig. 5). Although virus infectivity did not seem affected by the formation of these syncytium-like structures, H4 P116.1 infection induced a severe disruption of the GFAP cytoskeletal network in infected astrocytes (Fig. 6). Unlike typical syncytia observed after MHV-JHM infection, where GFAP staining is continuous within the aggregates (data not shown), H4 P116.1 seemed to induce a retraction of the GFAP filamentous network, whereas the cytoplasm of infected cells remained extended (Fig. 6, Panels G to I). Several viruses have been shown to cause such cytoskeleton alterations during *in vitro* infections. For instance, Theiler's murine encephalomyelitis virus (TMEV) was shown to induce a disruption of desmin and vimentin networks (Nédellec et al., 1998). Consistent with our model (H4 P116.1), the measles virus was also shown to disrupt the GFAP cytoskeleton without any alteration of actin, tubulin or vimentin following infection of an astrocytoma cell line (Duprex et al., 2000). However, this is the first such report on cytoskeleton disruption following coronavirus infections. Given the different genotypes of variants U-87 MG P35 and H4 P116.1, it is plausible to suggest that the capacity of H4 P116.1 to induce the formation of syncytium-like structures in murine primary glial cells could be attributable to mutations located in or close to the FP and the HRs.

The current study represents the first report of an analysis of the adaptation of HCoV-OC43 to its target host cell environment. It opens the door to the study of various issues regarding adaptation

capabilities of human coronaviruses and mechanisms underlying this adaptation, such as those involved in virus tropism and infectivity, as well as virus-induced fusion of infected cells. The availability of a recently generated infectious HCoV-OC43 cDNA clone (St-Jean et al., 2006) provides an invaluable tool to more precisely characterize the molecular basis for human coronavirus neuropathogenesis.

MATERIALS AND METHODS

Viruses and cell lines

The ATCC HCoV-OC43 strain (ATCC number VR-759; GenBank accession number AY585228), initially obtained from ATCC in the mid 1980s, was grown in the HRT-18 cell line (human adenocarcinoma rectal) as described previously (Mounir and Talbot, 1992). The MO3.13 cell line was obtained following a fusion with an adult human oligodendrocyte (McLaurin et al., 1995) whereas H4 (neuroglioma; HTB-148), U-87 MG (astrocytoma; HTB-14) and TE-671 (medulloblastoma; CCL-136) cell lines were obtained from ATCC. Cell lines used for the determination of viral susceptibility to HCoV-OC43 are described in Table 2.

Persistent infections

The HCoV-OC43 ATCC strain was used for all persistent infections as well as for HCoV-OC43 susceptibility assays. Persistent infections were carried out in HRT-18, MO3.13, TE-671, H4 and U-87 MG cell lines using 10^6 TCID₅₀ of HCoV-OC43, when cells reached a confluence of 70%. Infected cells were cultured for four days at 33°C, washed with phosphate-buffered saline (PBS) and trypsinized for a first passage. Persistently infected cells were then passaged every four to seven days and cultured at 37°C. Samples of supernatants were harvested at each passage for viral titration, in order to monitor the course of infection. Four persistent infections were performed in the H4 cell line whereas the HRT-18, MO3.13 and H4 cell lines were also acutely infected, as controls.

Virus production and purification

Virus harvested from persistent infections was purified at randomly selected passages. To this end, a virus production was first carried out in sixty 150 mm diameter culture dishes for each purification, using the appropriate medium. Prior to purification, supernatant of the sixty dishes

was pooled, clarified and viral particles were precipitated after the addition of 10% (w/v) polyethylene glycol (PEG) 8000 (Sigma), supplemented with 0.5 M NaCl, following an incubation of 16 hours at 4°C. Nycodenz (Accurate Chemicals & Scientific Corp.) was used to perform two rounds of gradient purification. First, a discontinuous gradient, consisting of two cushions of Nycodenz (50% and 10%), was performed for two hours and was followed by a continuous gradient, consisting of five cushions (50, 40, 30, 20 and 10% respectively), which was performed for five hours. Centrifugations were carried out at 148,000 x *g* in a Beckman SW41 rotor. Purified virus was dialyzed against TMEN buffer pH 6.2 (0.05M Tris-maleate, 1mM EDTA, 0.1M NaCl) following each gradient purification step and stored at -80°C.

RT-PCR amplifications

Following virus purification, RNA was extracted using the GenElute Direct mRNA Miniprep Kit (Sigma) as recommended by the manufacturer. RNA was then reverse transcribed with the Moloney murine leukemia virus reverse transcriptase (Invitrogen), using 500 ng of oligo(dT) primer. The Expand High-Fidelity *Taq* polymerase (Roche) was used to perform the PCR amplifications. Primers specific to the HE gene (JUHEO1 and JUHEO2), the S gene (JUSO27 to JUSO30) and the N gene (JUO8 and JUO2) were used to amplify target regions (Table 1). Primers specific to the glyceraldehyde phosphodehydrogenase (GAPDH) gene (GAPDH-H: 5'-GTGAAGGTCGGAGTC AACG-3' and GAPDH-I: 5'-CACCTGGTGCTCAGTGTAGC-3'), the 5' region of the HCoV-OC43 genome (JUB1: 5'-CCCGCTTCACTGATCTCTTG-3' and JUB2: 5'-GCACAGGCATAAAGGCTAAGG-3') and leader regions of viral mRNA (JULO1: 5'-GCTTCACTGATCTCTTGTTAG-3', which binds to the leader sequence, and a specific reverse primer of the target gene) were also used as controls. PCR amplicons specific to HE, S and N genes were purified using the Qiaex II gel extraction kit (Qiagen), prior to sequencing.

Assays for viral susceptibility and modulation of tropism and infectivity

Prior to performing assays for modulation of tropism and infectivity, susceptibility of a panel of cell lines to infection by the HCoV-OC43 ATCC strain was determined (Table 2). To this end, cells were infected at 70% confluence at a multiplicity of infection (MOI) of 0.2. One hundred mL of viral stock were produced for each variant generated following viral purifications. These viral productions were used to perform downstream experiments, such as assays for modulation of tropism and infectivity, plaque assays as well as infections of primary cell cultures. The same cell lines used for assays for susceptibility to HCoV-OC43 ATCC were then infected with viruses harvested from each purification of persistently infected cells in order to investigate whether the observed S gene mutations correlated to a modulation of tropism or infectivity. All infections of adherent cells were carried out at 33°C in 6-well or 12-well plates, with a virus adsorption step of two hours in a small volume. Unbound virus was then removed with PBS and infections were carried out for four days at 33°C. Infections of cells in suspension were performed in 4-mL plastic tubes for four hours, with occasional shaking. Infected cells were then washed, harvested and cultured in 12-well plates for four days. Supernatants from each infection were clarified and quantified using the indirect immunoperoxidase assay (IPA), as previously described (Jacomy and Talbot, 2003).

Plaque assays

Plaque assays were performed using the RD-151 cell line (human rhabdomyosarcoma). Cells were infected at a confluence of 90-95% in 12-well plates using serial dilutions in minimum essential medium (MEM) Earles (Invitrogen) supplemented with 1% (v/v) heat-inactivated fetal calf serum (FCS). Following a 2-hour incubation in a minimal volume (200 mL), with occasional shaking at 37°C, inoculums were removed and replaced with complete MEM Earles containing

1% (w/v) methylcellulose supplemented with 10% (v/v) FCS. Plates were then incubated at 37°C for five days, in a 5% (v/v) CO₂ humid atmosphere.

Primary cortical cell cultures and infections

Embryos at 16 to 18 days of gestation were removed from pregnant anesthetized BALB/c mice. Cortical cell cultures were obtained following the modified method of Brewer and collaborators (Brewer et al., 1993). Cells were dissected in HBSS without Ca²⁺ and Mg²⁺, supplemented with 1.0 mM sodium pyruvate and 10 mM HEPES. Individual cells were isolated by trituration in the same medium and diluted with 2 volumes of HBSS with Ca²⁺ and Mg²⁺. Supernatants were then transferred to a 15-mL tube and centrifuged for 1 min at 1000 x g. Pellets were resuspended in 1 mL of HBSS per brain. Cells were grown on glass coverslips, pre-treated with poly-D-Lysine and plated at 5x10⁵/cm². Cells were plated in neurobasal medium (Invitrogen) supplemented with 0.5 mM L-glutamine, 25 μM glutamate and B27 supplement (Invitrogen). After 4 days of incubation, medium was replaced with Neurobasal/B27 without glutamate, and then replaced every two days during one to two weeks, before infection. Cortical cultures were infected at a MOI of 0.1, at 37°C, using HCoV-OC43 ATCC, HCoV-OC43 H4 P116.1 or HCoV-OC43 U-87 MG P35, as described above. Supernatants were harvested at 12, 24, 48, 72 and 96 hours post-infection and analyzed for the presence and quantification of infectious virus using the IPA described above.

Immunofluorescence staining

To visualize viral antigens and cell markers, primary cultures were washed with warm sterile PBS, and then fixed with 4% (v/v) paraformaldehyde at 4°C, for 30 min. Following washing, cells were permeabilized with 100% methanol at -20 °C for 5 min, washed again in PBS, and incubated for two hours with primary antibodies, as previously described (Jacomy et al., 2006). For viral antigens, 1/1000 dilutions of ascites fluid of the 4-E11.3 hybridoma, which secretes monoclonal antibodies specific to the nucleocapsid protein of the antigenically related

hemagglutinating encephalomyelitis virus, were used (Bonavia et al., 1997) or a rabbit-anti-BCoV antiserum specific for the nucleocapsid protein of the serologically related bovine coronavirus at 1/500 dilution (Michaud and Dea, 1993). Astrocytes were identified with a rabbit anti-glial-fibrillary-acidic protein antibody (GFAP) diluted 1/500 (DAKO) and neurons by a mouse anti-tubulin beta III isoform (Chemicon), diluted 1/1000. After several washes in PBS, cell coverslips were incubated for one hour at 37°C in the dark, with a combination of immunofluorescent secondary antibodies: Alexa Fluor 488 (Molecular Probes) at a dilution of 1/1500 and Alexa Fluor 568 (Molecular Probes) at a dilution of 1/1000. After three washes in water, cells were incubated with 4,6-diamidino-2-phenylindole dihydrochloride (DAPI; Polysciences Inc.) at a 1/100 dilution for 5 min. Following final washes in water, coverslips from each well were removed and mounted on microscope slides in Immuno-mount and observed under a fluorescence microscope.

Nucleotide sequencing

Sequencing reactions were performed by Bio S&T (Montreal, Canada) using the dideoxy method (Sanger) and specific primers (St-Jean et al., 2004). Both strands of each sample were sequenced in each case. Sequences obtained from chromatograms were aligned using BLAST (bl2seq) from NCBI and were analyzed using the Chromas 2 software.

ACKNOWLEDGMENTS

We are very grateful to Francine Lambert for her technical assistance with cell and virus culture and especially for her involvement in the assays for viral susceptibility and modulation of tropism and infectivity.

This work was supported by Grant No. MT-9203 from Canadian Institutes for Health Research (Institute of Infection and Immunity). Pierre J. Talbot is the holder of the Tier-1 Canada Research Chair in Neuroimmunovirology. Julien R. St-Jean acknowledges a studentship from *Fonds québécois de recherche sur la nature et les technologies* (FQRNT).

REFERENCES

- Almazán, F., Galán, C., Enjuanes, L., 2004. The nucleoprotein is required for efficient coronavirus genome replication. *J. Virol.* 78, 12683-12688.
- Almazán, F., González, J. M., Péntzes, Z., Izeta, A., Calvo, E., Plana-Durán, J., Enjuanes, L., 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. USA* 97, 5516-5521.
- Arbour, N., Côté, G., Lachance, C., Tardieu, M., Cashman, N. R., Talbot, P. J., 1999. Acute and persistent infection of human neural cell lines by human coronavirus OC43. *J. Virol.* 73, 3338-3350.
- Arbour, N., Day, R., Newcombe, J., Talbot, P. J., 2000. Neuroinvasion by human respiratory coronaviruses. *J. Virol.* 74, 8913-8921.
- Babcock, G. J., Esshaki, D. J., Thomas Jr., W. D., Ambrosino, D. M., 2004. Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor. *J. Virol.* 78, 4552-4560.
- Bonavia, A., Arbour, N., Yong, V. W., Talbot, P. J., 1997. Infection of primary cultures of human neural cells by human coronaviruses 229E and OC43. *J. Virol.* 71, 800-806.
- Bosch, B. J., de Haan, C. A. M., Smits, S. L., Rottier, P. J. M., 2005. Spike protein assembly into the coronavirus: exploring the limits of its sequence requirements. *Virology* 334, 306-318.
- Bosch, B. J., van der Zee, R., de Haan, C. A. M., Rottier, P. J. M., 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* 77, 8801-8811.

Brewer, G. J., Torricelli, J. R., Evege, E. K., Price, P. J., 1993. Optimized survival of hippocampal neurons in B27-supplemented Neurobasal, a new serum-free medium combination. *J. Neurosci. Res.* 35, 567-576.

de Haan, C. A. M., Stadler, K., Godeke, G.-J., Bosch, B. J., Rottier, P. J. M., 2004. Cleavage inhibition of the murine coronavirus spike protein by a furin-like enzyme affects cell-cell but not virus-cell fusion. *J. Virol.* 78, 6048-6054.

Duprex, W. P., McQuaid, S., Rima, B. K., 2000. Measles virus-induced disruption of the glial-fibrillary-acidic protein cytoskeleton in an astrocytoma cell line (U-251). *J. Virol.* 74, 3874-3880.

Duquerry, S., Vigouroux, A., Rottier, P. J. M., Rey, F. A., Bosch, B. J., 2005. Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein. *Virology* 335, 276-285.

Escors, D., Izeta, A., Capiscol, C., Enjuanes, L., 2003. Transmissible gastroenteritis coronavirus packaging signal is located at the 5' end of the virus genome. *J. Virol.* 77, 7890-1902.

Fang, S. G., Shen, S., Tay, F. P. L., Liu, D. X., 2005. Selection of and recombination between minor variants lead to the adaptation of an avian coronavirus to primate cells. *Biochem. Biophys. Res. Comm.* 336, 417-423.

Follis, K. E., York, J., Nunberg, J. H., 2005. Serine-scanning mutagenesis studies of the C-terminal heptad repeats in the SARS coronavirus S glycoprotein highlight the important role of the short helical region. *Virology* 341, 122-129.

Fu, L., Gonzales, D. M., Sarma, J. D., Lavi, E., 2004. A combination of mutations in the S1 part of the spike glycoprotein gene of coronavirus MHV-A59 abolishes demyelination. *J. Neurovirol.* 10, 41-51.

Gélinas, A.-M., Boutin, M., Sasseville, M.-J., Dea, S., 2001. Bovine coronaviruses associated with enteric and respiratory diseases in canadian dairy cattle display different reactivities to *anti*-HE monoclonal antibodies and distinct amino acid changes in their HE, S and ns 4.9 protein. *Virus Res.* 76, 43-57.

Gorbalenya, A. E., Snijder, E. J., Spaan, W. J., 2004. Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. *J. Virol.* 78, 7863-7866.

Guillén, J., Pérez-Berná, A. J., Moreno, M. R., Villalaín, J., 2005. Identification of the membrane-active regions of the severe acute respiratory syndrome coronavirus spike membrane glycoprotein using a 16/18-mer peptide scan : implications for the viral fusion mechanism. *J. Virol.* 79, 1743-1752.

Haijema, B. J., Volders, H., Rottier, P. J. M., 2003. Switching species tropism: an effective way to manipulate the feline coronavirus genome. *J. Virol.* 77, 4528-4538.

Hofmann, H., Pyrc, K., van der Hoek, L., Geier, M., Berkhout, B., Pöhlmann, S., 2005. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc. Natl. Acad. Sci. USA* 102, 7988-7993.

Holmes, K. V., 2005. Adaptation of SARS coronavirus to humans. *Science* 309, 1822-1823.

Ingallinella, P., Bianchi, E., Finotto, M., Cantoni, G., Eckert, D. M., Supekar, V. M., Bruckmann, C., Carfi, A., Pessi, A., 2004. Structural characterization of the fusion-active complex of severe acute respiratory syndrome (SARS) coronavirus. *Proc. Natl. Acad. Sci. USA* 101, 8709-8714.

Jacomy, H., Talbot, P. J., 2003. Vacuolating encephalitis in mice infected by human coronavirus OC43. *Virology* 315, 20-33.

Jacomy, H., Fragoso, G., Almazan, G., Mushynski, W.E., Talbot, P.J., 2006. Human coronavirus OC43 infection induces chronic encephalitis leading to disabilities in BALB/c mice. *Virology*, Epub ahead of print.

Künkel, F., Herrler, G., 1996. Structural and functional analysis of the S proteins of two human coronavirus OC43 strains adapted to growth in different cells. *Arch. Virol.* 141, 1123-1131.

Kuo, L., Godeke, G.-J., Raamsman, M. J. B., Masters, P. S., Rottier, P. J. M., 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J. Virol.* 73, 1393-1406.

Lachance, C., Arbour, N., Cashman, N. R., Talbot, P.J., 1998. Involvement of aminopeptidase N (CD13) in infection of human neural cells by human coronavirus 229E. *J. Virol.* 72, 6511-6519.

Lai, M. M. C., Cavanagh, D., 1997. The molecular biology of coronaviruses. *Adv. Virus Res.* 48, 1-97.

Leparc-Goffart, I., Hingley, S. T., Chua, M. M., Jiang, X., Lavi, E., Weiss, S. R., 1997. Altered pathogenesis of a mutant of the murine coronavirus MHV-A59 is associated with a Q159L amino acid substitution in the spike protein. *Virology* 239, 1-10.

Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S.-K., Huang, I.-C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y., Choe, H., Farzan, M., 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 24, 1634-1643.

MacNamara, K. C., Chua, M. M., Phillips, J. J., Weiss, S. R., 2005. Contributions of the viral genetic background and a single amino acid substitution in an immunodominant CD8⁺ T-cell epitope to murine coronavirus neurovirulence. *J. Virol.* 79, 9108-9118.

Matsuyama, S., Taguchi, F., 2002. Communication between S1N330 and a region in S2 of murine coronavirus spike protein is important for virus entry into cells expressing CEACAM1b receptor. *Virology* 295, 160-171.

McLaurin, J., Trudel, G. C., Shaw, I. T., Antel, J. P., Cashman, N. R., 1995. A human glial hybrid cell line differentially expressing genes subserving oligodendrocyte and astrocyte phenotype. *J. Neurobiol.* 26, 183-193.

Michaud, L., Dea, S., 1993. Characterization of monoclonal antibodies to bovine enteric coronavirus and antigenic variability among Quebec isolates. *Arch. Virol.* 131, 455-465.

Miura, H. S., Nakagaki, K., Taguchi, F., 2004. N-terminal domain of the murine coronavirus receptor CEACAM1 is responsible for fusogenic activation and conformational changes of the spike protein. *J. Virol.* 78, 216-223.

Mounir, S., Talbot, P. J., 1992. Sequence analysis of the membrane protein gene of human coronavirus OC43 and evidence for *O*-glycosylation. *J. Gen. Virol.* 73, 2731-2736.

Myint, S. H., 1994. Human coronaviruses : a brief review. *Rev. Med. Virol.* 4, 35-46.

Navas-Martin, S., Hingley, S. T., Weiss, S. R., 2005. Murine coronavirus evolution in vivo: functional compensation of a detrimental amino acid substitution in the receptor binding domain of the spike glycoprotein. *J. Virol.* 79, 7629-7640.

Nédellec, P., Vicart, P., Laurent-Winter, C., Martinat, C., Prevost, M. C., Brahic, M., 1998. Interaction of Theiler's virus with intermediate filaments of infected cells. *J. Virol.* 72, 9553-9560.

Ontiveros, E., Kim, T. S., Gallagher, T. M., Perlman, S., 2003. Enhanced virulence mediated by the murine coronavirus, mouse hepatitis virus strain JHM, is associated with a glycine at residue 310 of the spike glycoprotein. *J. Virol.* 77, 10260-10269.

Peiris, J. S., Chu, C. M., Cheng, V. C., Chan, K. S., Hung, I. F., Poon, L. L., Law, K. I., Tang, B. S., Hon, T. Y., Chan, C. S., Chan, K. H., Ng, J. S., Zheng, B. J., Ng, W. L., Lai, R. W., Guan, Y., Yuen, K. Y., 2003. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* 361, 1767-1772.

Saeki, K., Ohtsuka, N., Taguchi, F., 1997. Identification of spike protein residues of murine coronavirus responsible for receptor-binding activity by use of soluble receptor-resistant mutants. *J. Virol.* 71, 9024-9031.

Sainz Jr., B., Rausch, J. M., Gallagher, W. R., Garry, R. F., Wimley, W. C., 2005. Identification and characterization of the putative fusion peptide of the severe acute respiratory syndrome-associated coronavirus spike protein. *J. Virol.* 79, 7195-7206.

Sarma, J. D., Fu, L., Tsai, J. C., Weiss, S. R., Lavi, E., 2000. Demyelination determinants map to the spike glycoprotein gene of coronavirus mouse hepatitis virus. *J. Virol.* 74, 9206-9213.

Schelle, B., Karl, N., Ludewig, B., Siddell, S. G., Thiel, V., 2005. Selective replication of coronavirus genomes that express nucleocapsid protein. *J. Virol.* 79, 6620-6630.

Schickli, J. H., Thackray, L. B., Sawicki, S. G., Holmes, K. V., 2004. The N-terminal region of the murine coronavirus spike glycoprotein is associated with the extended host range of viruses from persistently infected murine cells. *J. Virol.* 78, 9073-9083.

Schickli, J. H., Zelus, B. D., Wentworth, D. E., Sawicki, S. G., Holmes, K. V., 1997. The murine coronavirus mouse hepatitis virus strain A59 from persistently infected murine cells exhibits an extended host range. *J. Virol.* 71, 9499-9507.

Shen, S., Law, Y. C., Liu, D. X., 2004. A single amino acid mutation in the spike protein of coronavirus infectious bronchitis virus hampers its maturation and incorporation into virions at the nonpermissive temperature. *Virology* 326, 288-298.

Skowronski, D.M., Astell, C., Brunham, R.C., Low, D.E., Petric, M., Roper, R.L., Talbot, P.J., Tam, T., Babiuk, L., 2005. Severe acute respiratory syndrome: a year in review. *Annu. Rev. Med.* 56: 357-281.

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J. M., Gorbalenya, A. E., 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331, 991-1004.

St-Jean, J. R., Jacomy, H., Desforges, M., Vabret, A., Freymuth, F., Talbot, P. J., 2004. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J. Virol.* 78, 8824-8834.

St-Jean, J.R., Desforges, M., Almazán, F., Jacomy, H., Enjuanes, L., Talbot, P.J., 2006. Recovery of a neurovirulent human coronavirus OC43 from an infectious cDNA clone. *J. Virol.* 80, 3670-3674.

Taguchi, F., Shimazaki, Y. K., 2000. Functional analysis of an epitope in the S2 subunit of the murine coronavirus spike protein : involvement in fusion activity. *J. Gen. Virol.* 81, 2867-2871.

Thackray, L. B., Holmes, K. V., 2004. Amino acid substitutions and an insertion in the spike glycoprotein extend the host range of the murine coronavirus MHV-A59. *Virology* 324, 510-524.

Thackray, L. B., Turner, B. C., Holmes, K. V., 2005. Substitutions of conserved amino acids in the receptor-binding domain of the spike glycoprotein affect utilization of murine CEACAM1a by the murine coronavirus MHV-A59. *Virology* 334, 98-110.

Tsai, J. C., de Groot, L., Pinon, J. D., Iacono, K. T., Phillips, J. J., Seo, S.-H., Lavi, E., Weiss, S. R., 2003. Amino acid substitutions within the heptad repeat domain 1 of murine coronavirus spike protein restrict viral antigen spread in the central nervous system. *Virology* 312, 369-380.

Wu, G., Yan, S., 2005. Reasoning of spike glycoproteins being more vulnerable to mutations among 158 coronavirus proteins from different species. *J. Mol. Model.* 11, 8-16.

Yi, C. E., Ba, L., Zhang, L., Ho, D. D., Chen, Z., 2005. Single amino acid substitutions in the severe acute respiratory syndrome coronavirus spike glycoprotein determine viral entry and immunogenicity of a major neutralizing domain. *J. Virol.* 79, 11638-11646.

Ziebuhr, J., Snijder, E. J., Gorbalenya, A. E., 2000. Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J. Gen. Virol.* 81, 853-879.

TABLE 1 : Primers used for the amplification of HE, S and N genes of HCoV-OC43

Primer pairs and location	Target region	Sequence (5'→ 3')
JUHEO1 (22226-22244 of positive strand) JUHEO2 (30495-30514 of negative strand)	HE gene	AACAATTCTTGGTTCTCC ACATAATAAGTACCCAAACC
JUSO27 (23423-23443 of positive strand) JUSO28 (25976-25996 of negative strand)	S1 subunit	ATAATGTTAGCAGTGTCTGGC CCTACAGGTTTCAAATATCG
JUSO29 (25665-25685 of positive strand) JUSO30 (27794-27814 of negative strand)	S2 subunit	TATAGCGGTCGTGTTTCTGCG TTCTCAGGTTCTCCAAATGTC
JUO8 (29079-29100 of positive strand) JUO2 (30495-30514 of negative strand)	N gene	ATGTCTTTTACTCCTGGTAAGC GCAGCAAGACATCCATTCTG

TABLE 2 : Viral susceptibility of different cell lines to the HCoV-OC43 ATCC strain and variants

Cell line	Origin	Tissue	Type	IPA ^{1,2,3}	A ^{1,3,4}	B ^{1,3,4}	C ^{1,3,4}	D ^{1,3,4}	E ^{1,3,4}
HeLa	Human	Uterus	Epithelial	2.25 (no)	3.67	4.08	3.67	7.08	4.92
MT4	Human	Bone marrow	T lymphocyte	≤1.58 (no)	4.83	3.50	2.5	1.83	3.33
U937	Human	Bone marrow	Monocyte	≤0.58 (no)	4.0	<u>3.58</u>	2.67	2.75	2.17
Jurkat E6.1	Human	Bone marrow	T lymphocyte	≤1.92 (no)	3.92	4.17	2.83	3.17	2.5
Raji	Human	Bone marrow	B lymphocyte	3.0 (no)	3.17	3.42	3.25	3.42	4.33
HL-60	Human	Bone marrow	Monocyte	≤1.58 (no)	4.25	3.83	2.83	2.0	2.83
WI-38	Human	Lung	Fibroblast	≤1.83 (no)	≤2.42	≤1.5	3.08	6.17	4.75
L132	Human	Lung	Epithelial	≤2.5 (no)	5.38	4.25	4.38	6.83	4.92
Caki-2	Human	Kidney	Epithelial	≤0.5 (no)	2.75	<u>3.0</u>	<u>3.17</u>	<u>3.83</u>	2.58
NCI-N87	Human	Stomach	Epithelial	≤1.5 (no)	≤2.17	2.58	2.58	≤2.25	≤1.5
Arpe-19	Human	Eye	Epithelial	1.83 (no)	4.92	3.33	3.25	5.5	4.5
FHs 74 Int	Human	Intestine	Epithelial	1.83 (no)	3.42	4.17	2.75	4.83	4.92
TK6	Human	Spleen	T lymphocyte	≤1.58 (no)	2.5	2.92	2.83	≤1.58	≤2.0
17 Cl-1	Mouse	Embryo	Fibroblast	≤2.75 (no)	2.83	2.83	2.17	4.58	4.08
L929	Mouse	Subcutaneous	Fibroblast	3.0 (no)	2.5	2.67	2.58	4.67	2.33
N-11	Mouse	Brain	Microglial	≤1.5 (no)	≤2.42	≤1.5	≤1.75	4.0	4.25
DBT	Mouse	Brain	Glial	2.08 (no)	2.83	2.83	2.58	3.83	2.92
J774 A.1	Mouse	Bone marrow	Macrophage	2.75 (no)	3.58	4.08	3.42	3.17	2.92
A20	Mouse	Bone marrow	B lymphocyte	≤1.58 (no)	≤2.17	3.75	3.25	≤1.67	≤2.33
S.END.1	Mouse	Skin	Endothelial	≤0.5 (no)	≤0.75	≤0.58	≤0.75	≤0.5	≤0.5
Cos-7	Monkey	Kidney	Fibroblast	≤0.58 (no)	<u>3.83</u>	<u>3.33</u>	2.92	5.25	5.08
Vero	Monkey	Kidney	Epithelial	2.33 (no)	3.75	4.5	3.92	5.67	4.5
Vero E.6	Monkey	Kidney	Epithelial	≤0.58 (no)	2.92	<u>3.33</u>	2.67	<u>3.42</u>	2.67
B104	Rat	Brain	Fibroblast	3.75 (low)	<u>7.42</u>	<u>6.92</u>	<u>6.25</u>	<u>7.5</u>	<u>6.5</u>
BHK-21	Hamster	Kidney	Fibroblast	6.67 (yes)	5.75	7.17	5.67	7.58	6.33

¹ Infectious viral titers were measured using an immunoperoxidase assay and are expressed in log₁₀ TCID₅₀/mL

² Infectious viral titers from 0 to 3: not susceptible; titers over 3 and under 4: low susceptibility; titers of 4 and over: susceptible. The susceptibility to infection by the HCoV-OC43 ATCC reference strain is indicated in parenthesis.

³ Virus A : U-87 MG P35; virus B : H4 P47; virus C : H4 P56.3; virus D : H4 P116.1; virus E : H4 P116.2.

⁴ Results in bold indicate a modulation of tropism whereas underlined results indicate a modulation of infectivity.

FIGURE LEGENDS

Figure 1: Agarose gel showing RT-PCR amplifications of the N, S and HE genes, as well as controls, from purification H4 P116.2. Target genes were amplified using specific primers (Table 1) and were gel purified prior to sequencing. Left most and right most lanes, KiloBase DNA Marker (Amersham Biosciences); lane 1, N gene; lane 2, S1 subunit; lane 3, S2 subunit; lane 4, HE gene; lane 5, JUB1-2 (5' end of replicase 1a) PCR product; lane 6, GAPDH PCR product; lane 7, JULO1-JUSO28 (leader region and 5' end of S gene mRNA) PCR product.

Figure 2: Schematic representation of the main domains of the S glycoprotein as well as the approximate locations of the mutations found throughout the gene, for each purification. Color codes indicate the passage at which the mutation (or the reversion) was first observed. RBS, receptor binding site; HVR, hypervariable region; FP, fusion peptide; HR, heptad repeat; R, reversion; D, deletion; I, insertion.

Figure 3: Plaque assays performed with variants U-87 MG P35, H4 P47, H4 P56.3, H4 P116.1 and H4 P116.2. All five variants formed plaques of different morphologies on RD-151 cells, in contrast to HCoV-OC43 ATCC. Virus A : U-87 MG P35; virus B : H4 P47; virus C : H4 P56.3; virus D : H4 P116.1; virus E : H4 P116.2; virus F : HCoV-OC43 ATCC.

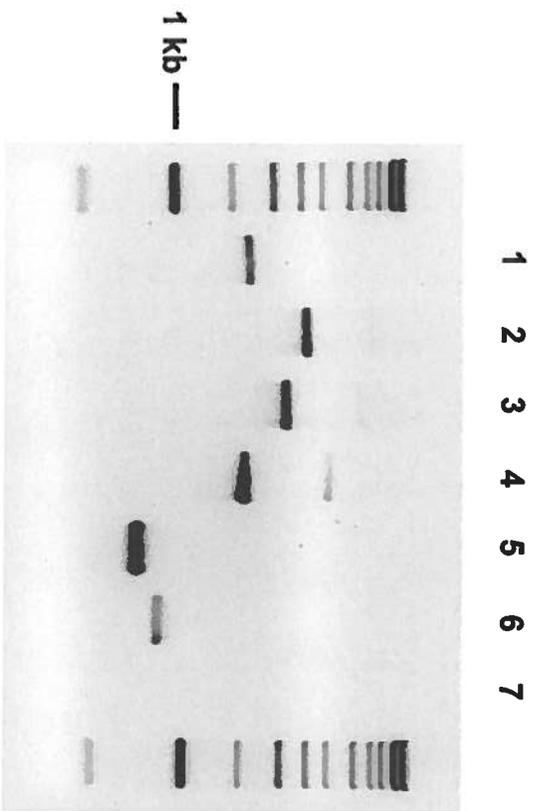
Figure 4: Susceptibility of mouse primary neural cells to HCoV-OC43 infection. Panels A and B: neural cell cultures were double stained for virus (antibody 4-E11.3, in green) and for astrocytes (GFAP; in red). Panel A: HCoV-OC43 ATCC-infected cells; Panel B: neural cells infected by U-87 MG P35; infected astrocytes appear in yellow (merge of red and green). Panels C and D: double immunofluorescent staining of neurons (Tubulin beta III isoform; in green) and

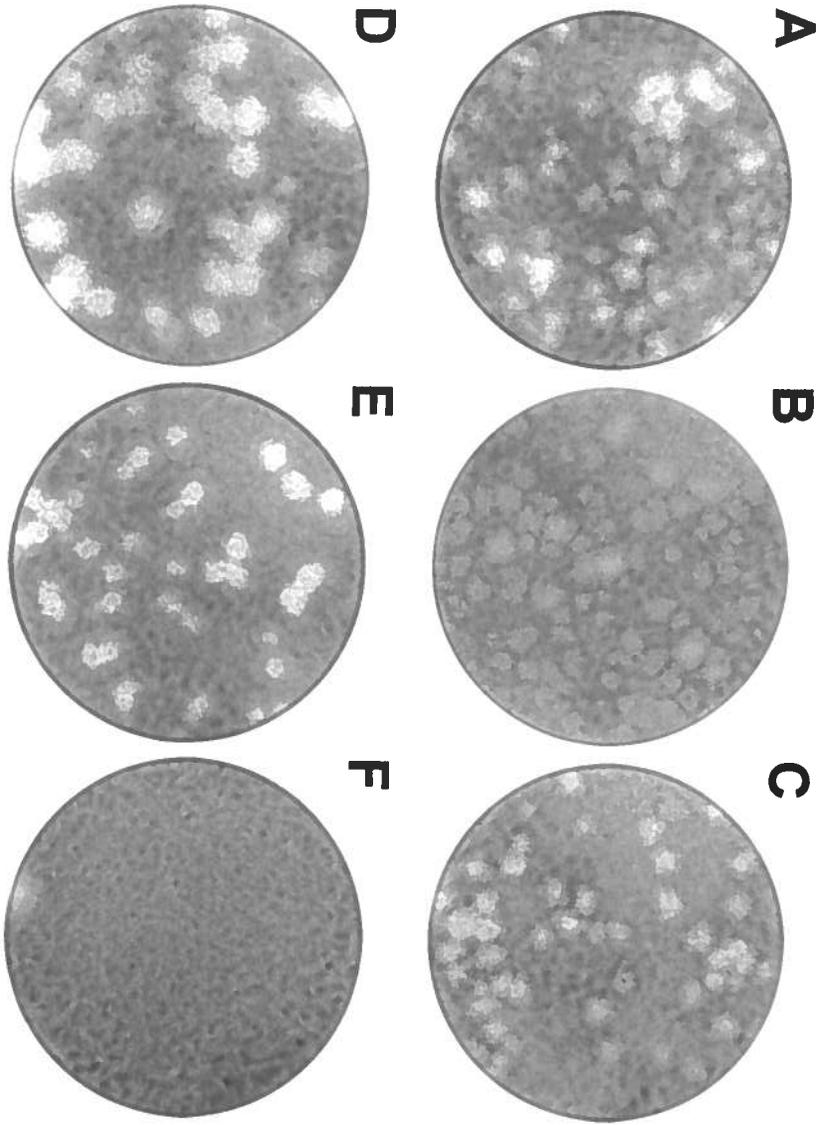
virus (antibody against related BCoV; in red) confirmed that HCoV-OC43 ATCC-infected cells (Panel C) were neurons (appearing in yellow). H4 P116.1-infected cells (Panel D) were astrocytes (appearing in red) and neurons (appearing in yellow). Original magnification x 200.

Figure 5: Infections of primary mouse neural cells with HCoV-OC43 ATCC and H4 P116.1 at 48 hours post-infection. Neural cell cultures were double stained for virus (in green) and for astrocytes (GFAP in red). The morphological appearance of infected cells differed following infection by HCoV-OC43 ATCC (Panel A) and H4 P116.1 (Panel B). In the same field, astrocytes staining in H4 P116.1-infected cultures (Panel D) contained many nuclei (arrowhead) and formed syncytium-like structures, when compared to the HCoV-OC43 ATCC-infected cultures (Panel C). Panel E: merged picture of Panels A and C. Panel F: merged picture of Panels B and D. Original magnification x 200.

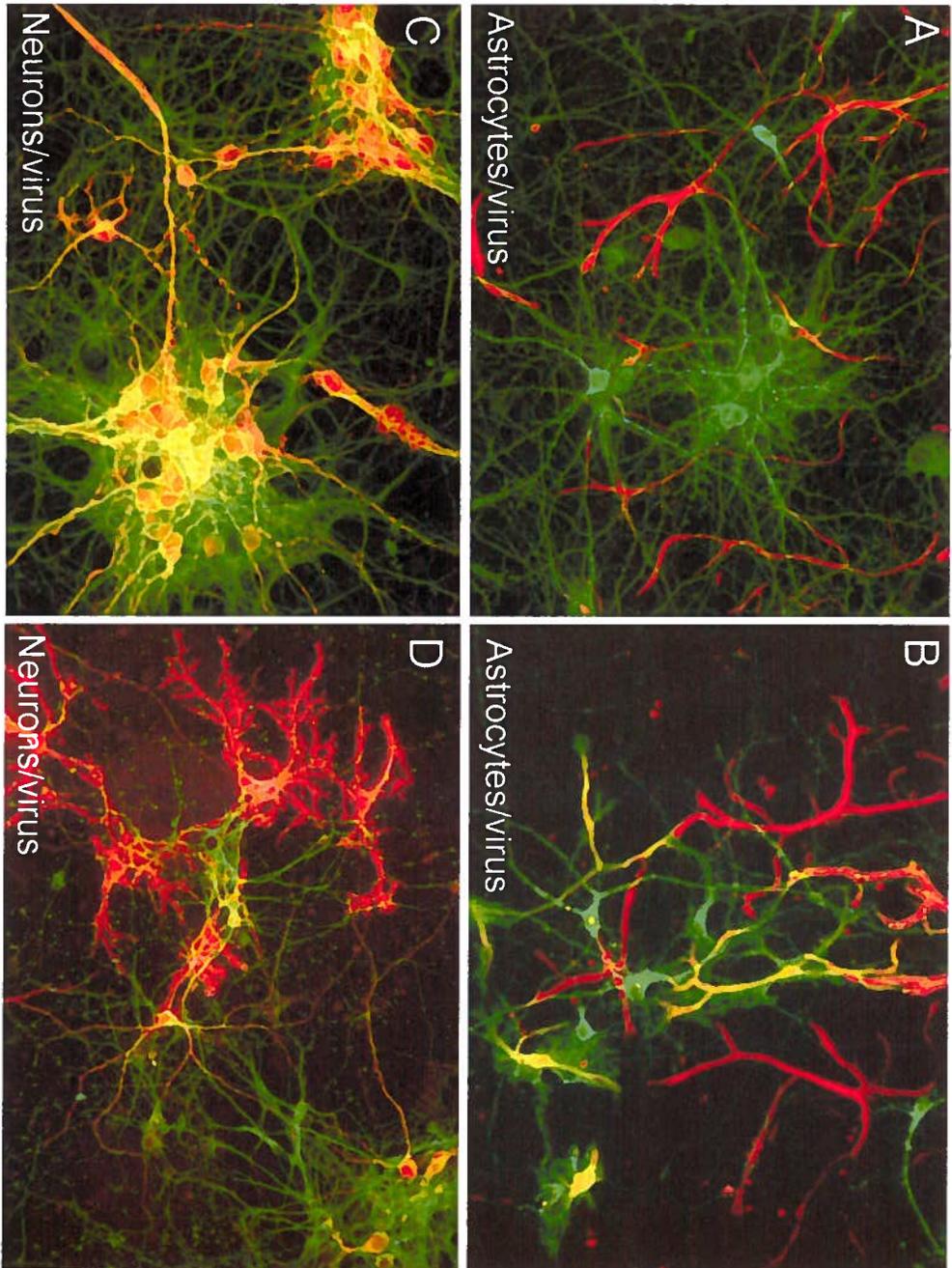
Figure 6: H4 P116.1-infected mouse primary neural cells at four days post-infection. Panels A to C: cell cultures were double stained for neurons (Panel A, in green) and for virus (Panel B, in red). Panel C: merged picture of Panels A and B. Axonal processes were well individually stained, illustrating that infected neurons (appearing yellow) were not included in the syncytium-like structures of infected cells. Panels D to F: neural cell cultures were double stained for astrocytes (Panel D, in red) and for virus (Panel E, in green), and the merged picture is displayed in Panel F. Panels G to I: magnification of Panels D to F illustrating that infected astrocytes displayed a disrupted GFAP network (Panel G, in red) following H4 P116.1 infection (Panel H, in green). Panel I: merged picture illustrating the reorganization of the GFAP network that took place in infected astrocytes upon H4 P116.1 infection. Positive GFAP staining was observed within syncytium-like formations, albeit at a very low level. Original magnification: x 200 for Panels A to F and x 400 for Panels G to I.

Figure 7: Infectious viral titers in infected mouse primary neural cell cultures following HCoV-OC43 ATCC, H4 P116.1 and U-87 MG P35 infections. Although both mutants were able to infect murine primary cells, their level of replication was lower than the HCoV-OC43 ATCC reference strain. Results are expressed as \log_{10} TCID₅₀/mL.

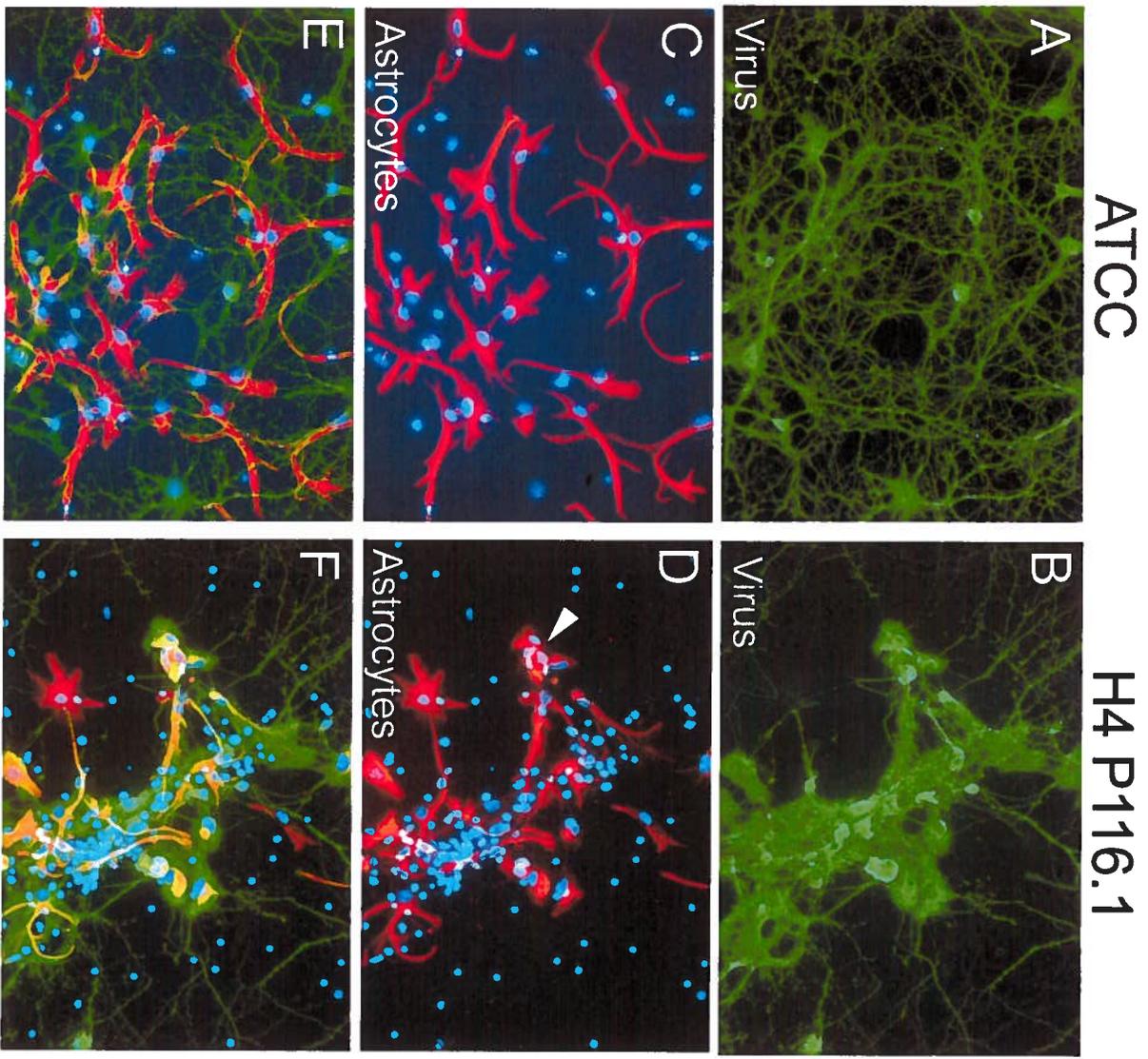




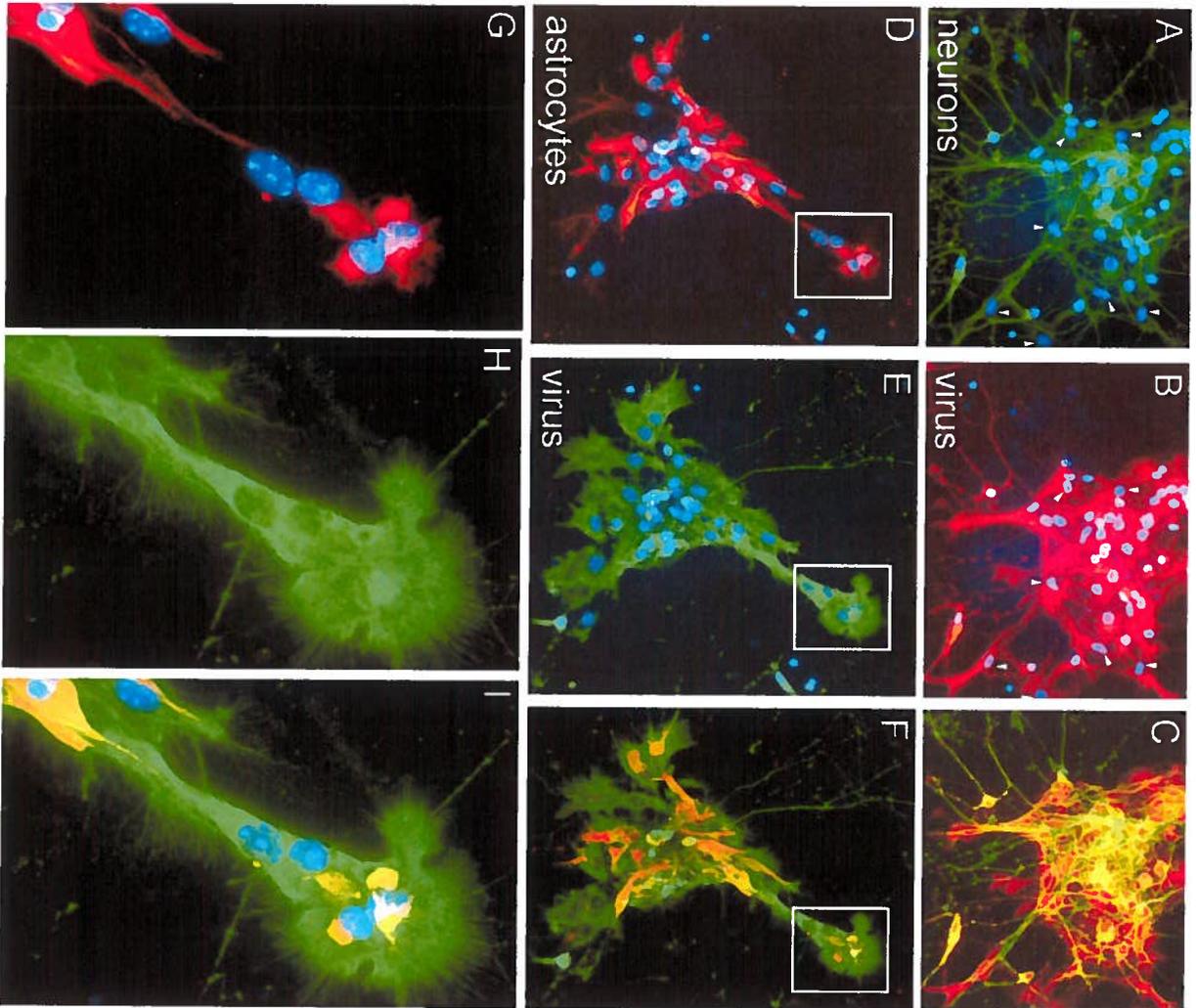
St-Jean et al. Fig. 3



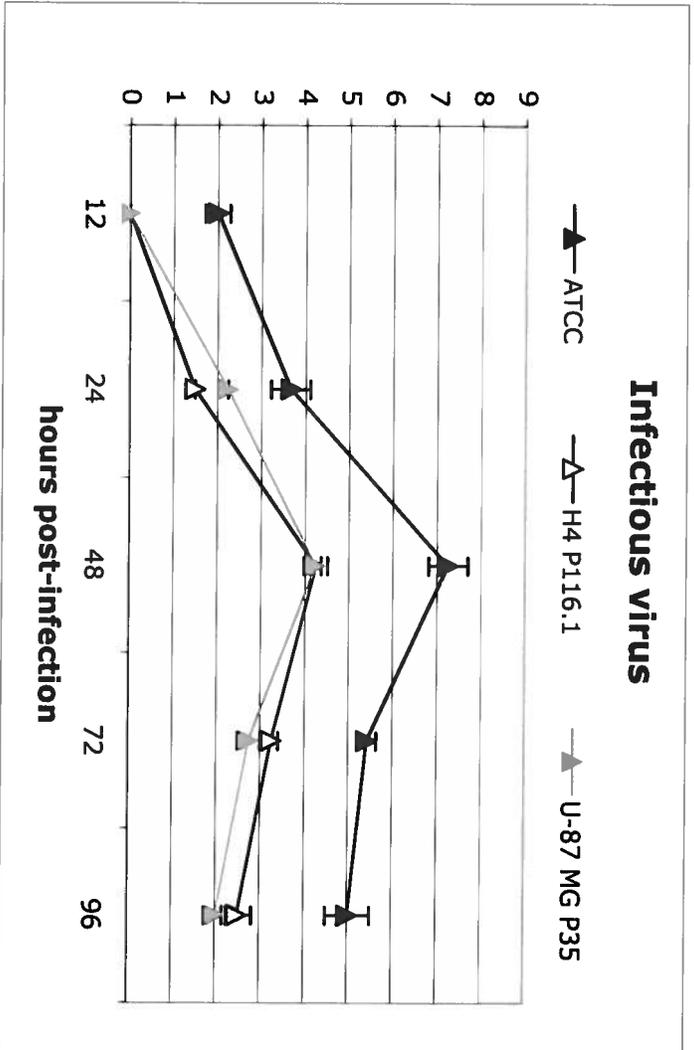
St-Jean et al. Fig. 4



St-Jean et al. Fig. 5



St-Jean et al. Fig. 6



CONCLUSION

Bien que les travaux réalisés au cours de ce doctorat aient porté sur différents aspects de la biologie des coronavirus et mené à la rédaction de manuscrits dont les sujets étaient parfois divergents, l'essence de la présente recherche découle d'une seule et même hypothèse. Cette hypothèse stipule que les mutations retrouvées dans le génome du HCoV-OC43 suite à une infection persistante représenteraient une adaptation du virus à son environnement cellulaire et qu'elles pourraient entraîner une modification des propriétés virales, dont une modulation du tropisme cellulaire. Le but de cette étude est donc de caractériser l'adaptation du HCoV-OC43 dans des cellules neurales représentatives du SNC humain. L'hypothèse du projet a été déterminée suite aux résultats obtenus lors des travaux de Nathalie Arbour, une ancienne étudiante au doctorat dans le laboratoire du Dr Pierre Talbot. Ces travaux ont notamment démontré les propriétés neurotropes et neuroinvasives du HCoV-OC43 chez l'humain (Arbour *et al.*, 1999a; Arbour *et al.*, 2000). Suite au séquençage de l'ARN viral provenant de cultures de cellules neurales et de cerveaux humains, des mutations ont été observées dans les gènes S et N respectivement, suggérant une adaptation potentielle du HCoV-OC43 dans le SNC et/ou le cerveau humain. La persistance du virus pourrait ainsi être à l'origine d'une telle adaptation, mais aucune caractérisation des mutations, ni du lien qui existe entre celles-ci et l'adaptation n'avait été effectuée à l'époque.

L'ordre de parution des publications contenues dans cette thèse ne reflète pas nécessairement la chronologie observée au cours de ce doctorat. Les événements et les expériences présentés dans cet ouvrage suivront par conséquent un ordre et une évolution logique par rapport à l'hypothèse de départ, présentée précédemment. Les publications no.3 et 5 de cette thèse sont directement reliées à cette hypothèse et traitent respectivement de l'évolution et de l'adaptation du HCoV-OC43, suite à des infections persistantes en cultures de cellules neurales. À l'image de ces deux notions, les publications no.3 et 5 sont plutôt complémentaires. La publication no.3 contient donc les résultats préliminaires à la poursuite de l'objectif principal, qui consiste à caractériser l'adaptation du virus. Dans cette publication, aucune analyse reliée à l'adaptation du virus n'est effectuée et les résultats présentés ne font état que de la capacité du HCoV-OC43 à évoluer en culture cellulaire. L'adaptation qui découle de cette évolution est donc abordée de façon approfondie dans la publication no.5, laquelle présente un compte rendu des mutations acquises durant les différentes infections persistantes ainsi que quelques résultats qui font état des conséquences de l'adaptation observée.

Afin de s'assurer que les mutations identifiées dans les particules virales isolées lors des infections persistantes étaient contenues dans l'ARN génomique viral plutôt que dans les ARNm sous-génomiques, le séquençage a été réalisé à partir de l'ARN extrait de virions purifiés. Comme les virions ne contiennent que l'ARN génomique pleine longueur, cette approche permettait donc d'éliminer les possibilités de contamination par les ARN indésirables. Étant donné les fonctions majeures que la protéine S des coronavirus remplit durant l'infection ainsi que son association potentielle avec le neurotropisme, la neuroinvasion et la neurovirulence, le gène S du HCoV-OC43 a donc été ciblé pour cette analyse. Le gène HE a également été analysé lors de cette étude puisqu'il pourrait lui aussi remplir des fonctions reliées à l'attachement à un récepteur cellulaire (Kazi *et al.*, 2005). Le gène N a aussi été analysé et le faible taux de mutations dans celui-ci est venu renforcer le rôle possible du gène S dans la modulation de l'adaptation du HCoV-OC43. Les résultats de cette étude ainsi que leur interprétation sont présentés en détail dans la publication no.5 de cet ouvrage. Cependant, de nombreux facteurs reliés à l'adaptation virale, tels que la capacité d'induction de la persistance et les paramètres mêmes de l'adaptation, n'ont pas été abordés et le seront ici.

Bien que le HCoV-OC43 possède la capacité intrinsèque d'atteindre le SNC (St-Jean *et al.*, 2004) et de pouvoir causer des infections persistantes, deux hypothèses sont possibles en ce qui concerne le mode de propagation du virus au SNC. Dans la première hypothèse, le virus atteindrait d'abord les voies respiratoires supérieures, qu'il utiliserait comme site primaire d'infection. Suite à une adaptation du virus dans les voies respiratoires, ou non, celui-ci gagnerait ensuite le SNC où, encore une fois, une adaptation virale pourrait survenir. Dans la seconde hypothèse, le virus emprunterait le nerf olfactif et atteindrait directement le SNC, à l'intérieur duquel une adaptation pourrait également survenir. Le site d'adaptation du virus dans l'hôte n'est donc pas connu et on ignore si ce site pourrait avoir une incidence sur le niveau ou la capacité d'adaptation du HCoV-OC43. Il a cependant déjà été démontré qu'une adaptation du virus n'était pas requise afin d'induire des infections persistantes, du moins *in vitro* (Arbour *et al.*, 1999a). Certains virus, comme le VIH et l'herpès, possèdent des mécanismes particuliers et ingénieux qui leur permettent de persister dans l'hôte et, par le fait même, d'échapper au système immunitaire. Les rétrovirus sont capables de s'intégrer dans le génome de l'hôte en utilisant une intégrase (Van Maele et Debyser, 2005), alors que le virus de l'herpès peut induire un état de latence dans lequel il peut persister durant la vie entière de l'hôte (Decman *et al.*, 2005). Bien que les coronavirus ne semblent pas posséder de tels mécanismes de persistance, ceux-ci auraient

la capacité d'induire des infections sur de très longues périodes de temps *in vitro* et *in vivo* et de se répliquer à des niveaux pratiquement indétectables chez l'hôte, du moins pour le système immunitaire (Stohlman *et al.*, 2002). *In vivo*, le niveau de persistance des coronavirus semble être en grande partie déterminé par le système immunitaire, chez le MHV notamment. Les lymphocytes T spécifiques au virus seraient cependant incapables de contrôler la persistance dans le SNC et c'est plutôt l'immunité humorale qui remplirait cette fonction. Ce phénomène est supporté par l'accumulation et la rétention de lymphocytes B spécifiques au virus suite à l'élimination virale qui survient suite à une infection aiguë (Stohlman *et al.*, 2002). Bien que les mécanismes de persistance du virus qui contrôlent directement l'infection n'aient toujours pas été élucidés *in vivo*, on sait que le HCoV-OC43 pourrait lui aussi induire des infections persistantes chez la souris. La persistance du virus a été démontrée dans le laboratoire du Dr Pierre Talbot, par le Dr Hélène Jacomy, par la détection d'ARN viral dans des cerveaux de souris ayant été inoculées un an auparavant (Jacomy *et al.*, 2006).

La survie à long terme d'un virus dans une population cellulaire viable survient habituellement lorsque la destruction cellulaire est limitée et que la réplication des génomes viraux montre un équilibre avec la multiplication des cellules hôtes (Domingo *et al.*, 1998). Certains virus pourraient donc coexister avec des types cellulaires précis, comme c'est le cas pour le réovirus et les cellules L (Ahmed *et al.*, 1981). Un tel mécanisme pourrait logiquement intervenir lors de la persistance des coronavirus dans le SNC, quoique la démonstration d'un tel phénomène n'ait pas été effectuée. Bien que le virus semble en quelque sorte devoir s'adapter à son environnement cellulaire lors de l'infection persistante, le contraire peut également survenir. Effectivement, lors d'une infection, les cellules peuvent rapidement devenir plus résistantes au virus, tel qu'il l'a été démontré pour le *foot-and-mouth disease virus* (FMDV) (Martín-Hernández *et al.*, 1994). En réponse à l'infection persistante, les cellules peuvent aussi moduler l'expression des récepteurs spécifiques au virus à la baisse. Ce mécanisme est aujourd'hui bien établi et a été rapporté chez plusieurs virus, dont le MHV (Schickli *et al.*, 1997). Il pourrait par contre entraîner une adaptation majeure du virus puisque le faible taux de récepteurs à la surface cellulaire procurerait un avantage de sélection aux virus dont l'interaction avec ces mêmes récepteurs est altérée. Une population virale qui utilise le récepteur différemment ou qui utilise un récepteur alternatif pourrait ainsi être générée (Schickli *et al.*, 1997). Un phénomène d'adaptation cellulaire a d'ailleurs été observé dans le laboratoire, lors de l'infection persistante du HCoV-OC43 dans la lignée cellulaire HRT-18. Aux environs du passage 145, les titres viraux ont commencé à chuter et sont passés sous le seuil de détection quelques passages plus tard, suite à ce qui semblait être

une adaptation de la lignée cellulaire au virus. La lignée HRT-18 adaptée était infectable par le virus de souche sauvage, indiquant qu'une modulation à la baisse du récepteur viral n'était pas survenue. Bien qu'une adaptation spécifique à la souche virale persistante aurait pu être établie par la lignée cellulaire, la possibilité d'une adaptation virale quelconque ne peut être exclue. Effectivement, une modulation à la baisse du tropisme cellulaire ou une altération du complexe de réplication auraient pu survenir lors de l'infection persistante. Un phénomène semblable a également déjà été observé dans la lignée cellulaire U-373 MG, alors que les titres viraux du HCoV-OC43 étaient devenues indétectables pour quelques passages lors d'une infection persistante (Arbour *et al.*, 1999a). Bref, peu importe la nature de cette adaptation, celle-ci est probablement survenue en réponse à une adaptation préalable du virus.

La relation qui existe entre le virus et l'organisme représente également un facteur important dans l'établissement de l'infection persistante, notamment en raison de l'implication du système immunitaire. Chez les virus à ARN, la production constante de particules virales mutantes contribue en grande partie à la survie du virus, puisque ces mutations permettent à certains virions d'échapper au système immunitaire (Domingo *et al.*, 1998). Certains virus peuvent aussi persister en étant séquestrés à des sites de l'organisme où l'immunité est moins présente, tel que le SNC, lequel est partiellement protégé du système immunitaire. L'immunosuppression, qu'elle soit induite par l'infection, une maladie ou encore un traitement, peut aussi influencer le déroulement de la persistance. Certains virus ont la capacité d'infecter des cellules du système immunitaire, telles que les macrophages, les lymphocytes et les cellules dendritiques, et utilisent cette voie d'entrée afin d'immunosupprimer l'hôte et de faciliter leur persistance (Zuniga *et al.*, 2005). Les HCoV-OC43 et 229E peuvent infecter les monocytes et les macrophages du sang périphérique (Collins, 2002), tout comme le MHV possède la capacité d'infecter les cellules dendritiques (Turner *et al.*, 2004). Il n'a cependant pas été clairement démontré si un lien existait entre l'infection de ce type de cellules par les coronavirus et l'immunosuppression.

Tel que mentionné dans la revue bibliographique de cette thèse, le gène S est le principal déterminant de l'adaptation des coronavirus. Bien que le SARS-CoV fournisse un excellent exemple de l'implication de la protéine S dans un tel phénomène, une relation entre l'adaptation et la protéine de surface de plusieurs virus à ARN a déjà été établie, principalement en ce qui concerne l'interaction avec le récepteur cellulaire (Baranowski *et al.*, 2001). Il est difficile d'établir un lien direct entre les mutations observées dans le gène S du HCoV-OC43 suite aux

différentes infections persistantes réalisées dans le cadre de ce projet et l'adaptation du virus dans un contexte *in vivo*, puisque les deux systèmes présentent des pressions de sélection complètement différentes, notamment en raison du système immunitaire. L'infection persistante du HCoV-OC43 dans un système *in vivo* n'aurait d'ailleurs probablement pas fourni exactement les mêmes résultats. Bien que la présence des mutations identifiées dans le gène S découle d'un avantage évolutif quelconque et que celles-ci confèrent un phénotype différent au virus chez la souris, l'adaptation caractérisée dans ce projet est spécifique à un système de culture cellulaire. Il est donc important de savoir comment interpréter la signification des mutations acquises en culture cellulaire lorsqu'on étudie l'évolution et l'adaptation moléculaire d'un virus (Poon *et al.*, 2005b).

L'adaptation virale représente un principe complexe, au cours duquel peut intervenir une multitude de facteurs et de variables. Plusieurs hypothèses (Ruiz-Jarabo *et al.*, 2000) et mécanismes (Sallie, 2004) sous-tendent ce phénomène et seront brièvement abordés dans cette section. L'adaptation des virus à ARN est d'abord une conséquence directe de l'absence d'activité exonucléase de la RdRp virale, laquelle est habituellement responsable de la correction de la plupart des erreurs introduites dans le génome viral. Le taux de mutations moyen des virus à ARN se situerait entre 10^{-3} et 10^{-5} substitutions par nucléotide copié (Ruiz-Jarabo *et al.*, 2000), ce qui concorde avec le taux de mutations connu des coronavirus, qui se situe à environ 10^{-4} (Sanchez *et al.*, 1992). En moyenne, une mutation serait donc introduite dans le génome des coronavirus à toutes les 10 000 bases, ce qui représente trois mutations par génome copié. En raison du taux d'erreur de la RdRp, il serait très improbable qu'une progéniture virale soit dépourvue de mutations dans une cellule infectée. Par conséquent, pour les virus à ARN et quelques virus à ADN, les variations génétiques entraînées par les mutations et la recombinaison seraient à l'origine de l'adaptation du virus à son environnement (Domingo *et al.*, 1998). En raison de leur complexité génétique limitée, due à la petite taille de leur génome, les virus à ARN sont généralement plus tolérants aux taux de mutagenèse élevés, ce qui explique leur prépondérance dans le phénomène d'adaptation virale. Lors de l'infection, les virus à ARN muteraient à un taux d'erreur maximum, mais compatible avec le maintien de l'intégrité de l'information génétique, ce qui leur permettrait d'acquérir rapidement les mutations bénéfiques requises pour l'adaptation (Elena et Sanjuán, 2005). Le taux de mutations neutres ou favorables à l'adaptation doit donc surpasser le taux de mutations ayant un effet néfaste ou délétère afin que la population virale puisse survivre et évoluer. Ce type de mécanisme est typique de ce qui est rencontré dans le système immunitaire des vertébrés, dans lequel les lymphocytes subissent une

série de mutations et de recombinaison qui mènent à la sélection et à l'expansion clonale des cellules présentant la plus haute affinité à l'antigène. L'hypermutation somatique des lymphocytes B représente un exemple de mécanisme du système immunitaire faisant intervenir une telle adaptation (Domingo *et al.*, 2002).

Les populations virales générées suite à l'incorporation de mutations et à l'adaptation qui en découle peuvent être très variables. Ces populations consistent généralement en des collections de génomes qui comprennent des mutations aléatoires et dispersées, qui font en sorte que la séquence consensus de la population issue d'une infection persistante est parfois la même que la séquence de la souche ayant causé l'infection. La séquence maîtresse est représentée par la séquence nucléotidique dominante de cette collection génomique et les distributions de mutants qui composent cette population virale sont désignées comme étant des quasiespèces virales (Domingo *et al.*, 2002). Ces quasiespèces sont constamment sujettes à une évolution et à des changements dus à la pression de sélection engendrée notamment par la compétition et l'environnement. L'influence de l'environnement est donc un facteur majeur dans la variabilité des quasiespèces, dans la sélection de nouveaux variants ainsi que dans l'émergence de nouvelles espèces virales pathogènes (Domingo *et al.*, 1998). Bien que certaines conditions de l'environnement puissent favoriser l'accumulation de mutations dans le génome viral, il a été démontré chez le VIH, l'influenza et le poliovirus qu'aucun lien n'existait entre le taux de mutations et la capacité d'adaptation d'une quasiespèce (Elena et Sanjuán, 2005).

L'environnement du virus serait également à l'origine de quelques hypothèses pouvant intervenir dans la génération de quasiespèces. L'une de ces hypothèses est la mémoire des quasiespèces virales, élaborée par le Dr Esteban Domingo (Ruiz-Jarabo *et al.*, 2000). Selon lui, lors de l'évolution d'une quasiespèce virale, les génomes enfouis ou cachés dans le spectre de mutants pourraient constituer un enregistrement moléculaire de l'histoire de la population. Autrement dit, même si un génome n'est pas dominant, il serait conservé parmi les génomes de la quasiespèce et pourrait refaire surface rapidement suite à un changement d'environnement ou face à de nouvelles contraintes de sélection, telles que celles imposées par le système immunitaire. Cette théorie a été démontrée chez le FMDV (Ruiz-Jarabo *et al.*, 2000) et fait ressortir l'importance du rôle de la clonalité et de l'individualité des mutations dans le phénomène de populations et d'adaptation. Bien que cette propriété des quasiespèces revête une certaine importance, son impact sur les populations générées suite aux infections persistantes réalisées dans le cadre de ce projet n'a pas été abordé dans la publication no.5. Ce point soulève

de nombreuses questions par rapport à la composition hétérogène des échantillons prélevés et analysés pour leur capacité d'adaptation et par rapport à l'effet de chacune des mutations retrouvées dans ces populations. Par exemple, si une quasiespèce virale devient dominante dans une lignée cellulaire précise, donc dans un environnement précis, et qu'un changement du contexte environnemental survient, comme c'est le cas lors d'un test de plages sur une nouvelle lignée cellulaire ou lors d'une infection en souris, les données peuvent changer. En effet, l'imposition de nouvelles contraintes à une population virale risque d'entraîner une modification du comportement de cette population et donc de la quasiespèce dominante. Par conséquent, les mutations analysées dans ce contexte pourraient devenir celles d'une quasiespèce minoritaire et c'est dans ce genre de situations que le phénomène de mémoire des quasiespèces virales, ou un phénomène apparenté, pourrait intervenir.

Afin d'avoir un plus grand contrôle sur les mutations analysées et de s'assurer de la clonalité de chacune, l'introduction des mutations une à une dans le génome viral serait essentielle et c'est à ce niveau que le clone infectieux d'ADNc du HCoV-OC43 intervient. Un clone d'ADNc du HCoV-OC43 a donc été assemblé dans un BAC sous le contrôle du promoteur du CMV (St-Jean *et al.*, 2006). En plus des diverses approches de mutagenèse et de clonage que ce clone permettra d'explorer, il pourra être utilisé afin de caractériser de façon plus approfondie l'adaptation du HCoV-OC43 et d'évaluer l'impact des mutations identifiées dans le génome du virus suite aux différentes infections persistantes. Afin d'entamer la construction du clone infectieux du HCoV-OC43, la séquence nucléotidique complète du génome viral a cependant dû être décryptée. Seule la séquence du tiers 3' du génome viral était connue au moment d'amorcer l'assemblage du clone infectieux, alors que la séquence de la réplicase 1ab était manquante. Le génome du HCoV-OC43 a par contre été séquencé en entier afin de s'assurer de la validité de notre séquence de référence. Les travaux relatifs à cette tâche sont présentés dans la publication no.1 de cette thèse. Cette publication présente également une analyse du génome du virus et de ses principales caractéristiques, une étude comparative du HCoV-OC43 par rapport à d'autres souches coronavirales apparentées, telles que le SARS-CoV, ainsi qu'une démonstration de la stabilité et des propriétés neuroinvasives intrinsèques du virus. Afin d'étudier la stabilité du génome viral dans l'environnement et de confirmer la neuroinvasion du HCoV-OC43, un isolat clinique du virus, désigné Paris, a été utilisé. Seulement six différences nucléotidiques, dont deux entraînant une substitution d'acides aminés, ont été identifiées entre les génomes des deux variants OC43, suggérant une stabilité génétique élevée du virus dans l'environnement. Des inoculations intranasales des deux variants ont aussi mené à des infections du SNC chez la souris,

indiquant que les deux variants possédaient la capacité d'atteindre et d'infecter des cellules neurales *in vivo*. Cette démonstration de neuroinvasion par la souche de laboratoire VR-759 et par un isolat naturel du HCoV-OC43 laisse croire que la souche de laboratoire du HCoV-OC43 posséderait des propriétés neurotropes et neuroinvasives intrinsèques. Ces propriétés n'auraient donc pas été acquises suite à la propagation du virus dans des cerveaux de souris nouveau-nés lors de sa production.

Bien que la stabilité génétique observée entre les deux variants OC43 semblait surprenante, aucun doute n'est survenu par rapport à la validité de nos résultats. Des précautions très strictes ont d'ailleurs été prises dans le laboratoire ainsi que dans celui de nos collaborateurs français afin de prévenir les possibilités de contaminations croisées entre les deux souches. Une équipe de recherche belge, celle du Dr Marc Van Ranst, a cependant trouvé que cette stabilité génétique était très improbable et a également proposé que le variant Paris soit plutôt issu d'une contamination croisée avec la souche de laboratoire VR-759. Le commentaire à l'éditeur de l'équipe du Dr Marc Van Ranst (en annexe) ainsi que notre réponse constituent la publication no.2 de cet ouvrage. Afin d'appuyer ses dires, l'équipe belge a présenté une série de calculs statistiques démontrant la faible probabilité de la stabilité génétique avancée dans la publication no.1. Bien que ces calculs étaient pertinents et réalistes et que nous partageons l'opinion de l'équipe belge à propos de la faible probabilité d'une telle stabilité génétique, l'absence d'une possibilité de contamination croisée a été maintenue dans notre réponse. Les arguments utilisés afin de défendre notre point de vue sont donc présentés dans cette publication. Parmi ceux-ci, l'argument du biais introduit par la culture du variant Paris dans la lignée cellulaire HRT-18 est très révélateur. Il fait notamment référence à l'influence des contraintes de sélection et de l'environnement sur l'adaptation d'un virus et fait du même coup indirectement référence à la présence potentielle d'une mémoire parmi les quasiespèces virales. Tel que discuté dans la lettre à l'éditeur, la présence d'un marqueur génétique dans le gène M (T432C), spécifique à tous les isolats cliniques issus de l'épidémie de Normandie (Vabret *et al.*, 2003), est également très importante et rend très peu probable la possibilité d'une contamination croisée entre les deux variants OC43.

Des hypothèses reliées aux mécanismes de persistance et de propagation du HCoV-OC43, qui n'ont pas été avancées dans la publication no.2, pourraient aussi renforcer la possibilité de la stabilité génétique du HCoV-OC43 dans l'environnement. Des observations réalisées à partir d'un picornavirus sont à la base de ces arguments (Domingo *et al.*, 1998). Un

parallèle pourrait donc être établi entre le rhinovirus humain et le coronavirus humain, puisque ces deux virus causent notamment des infections des voies respiratoires supérieures. L'analogie entre les deux virus serait basée sur le fait que le rhinovirus ne persiste pas à l'intérieur de son hôte et qu'il parvient à infecter continuellement d'autres hôtes grâce à un mode de transmission efficace par aérosols. Bien qu'il ait été démontré que l'infection persistante pouvait favoriser l'apparition de mutations et que le HCoV-OC43 pouvait persister dans des cultures de cellules neurales (Arbour *et al.*, 1999a), rien ne démontre que le même phénomène surviendrait dans les voies respiratoires, ce qui laisse croire que le virus pourrait utiliser un patron de transmission similaire aux rhinovirus. De plus, il a également été démontré chez les picornavirus que la stabilité génétique des particules virales était cruciale au succès de cette transmission (Domingo *et al.*, 1998), suggérant ainsi que le HCoV-OC43 Paris aurait pu subir une telle forme de propagation avant d'être isolé, ce qui aurait réduit considérablement le niveau de variabilité dans le génome viral. Bien que cette supposition semble contradictoire par rapport au taux de mutations élevé de la RdRp coronavirale, il est important de rappeler que la majorité des mutations introduites dans les génomes de virus à ARN sont néfastes ou délétères, et ce jusqu'à 95% du temps (Novella, 2003), et que ces mutations ne sont donc pas transmises à la progéniture virale.

Parallèlement au séquençage du génome du HCoV-OC43, une stratégie d'assemblage était mise sur pied pour la conception du clone infectieux d'ADNc du HCoV-OC43. Tel que mentionné dans la revue bibliographique, le clone a été assemblé en collaboration avec les Dr Fernando Almazán et Luis Enjuanes, du Centre National de Biotechnologie de Madrid. Les détails qui concernent l'assemblage du clone ainsi que son utilisation sont présentés dans la publication no.4 de cette thèse. Contrairement à l'équipe du Dr Luis Enjuanes, qui avait à sa disposition un génome défectif du TGEV d'environ 9,7 kb afin de démarrer la construction de leur premier clone d'ADNc, le pBAC-TGEV^{FL}, l'assemblage du pBAC-OC43^{FL} a dû être effectué en entier. Lors de l'assemblage du pBAC-TGEV^{FL}, le génome défectif a d'abord été inséré dans un BAC, puis les régions manquantes du génome, trois au total, ont ensuite été introduites dans le vecteur afin de générer un clone d'ADNc pleine longueur. Afin de permettre une transcription adéquate du génome d'ADN du TGEV en ARN, la copie d'ADN du génome du virus était flanquée du promoteur du CMV à son extrémité 5' et d'une queue de poly(A) synthétique, de la ribozyme du *hepatitis delta virus* (HDV) et des séquences de terminaison et de polyadénylation de l'hormone de croissance bovine à son extrémité 3'. Afin de démontrer l'efficacité de leur système, le gène S du virus cloné, une souche respiratoire nommée PUR46-MAD, a été remplacé

par celui d'une souche entérique et respiratoire, le PUR46-C11. Suite à cette transformation, le virus recombinant se répliquait alors abondamment dans les voies respiratoires et le tube digestif du porc et était très virulent, démontrant que le tropisme et la virulence de celui-ci pouvaient être modifiés (Almazán *et al.*, 2000).

Le pBAC-OC43^{FL} possède donc des caractéristiques très similaires au pBAC-TGEV^{FL}, mais une stratégie d'assemblage différente a dû être utilisée pour sa conception. En fait, deux stratégies ont été utilisées pour l'assemblage du pBAC-OC43^{FL}. La première consistait à introduire les extrémités 5' et 3' du génome viral dans le BAC de façon contiguë avec les séquences accessoires nécessaires à la transcription du virus et a été réalisée par *overlapping PCR*, alors que la deuxième consistait en une approche de clonage en cinq étapes, permettant d'introduire le reste du génome dans le vecteur. Une représentation schématique de l'assemblage du pBAC-OC43^{FL} est d'ailleurs fournie aux Figures 1 et 2 de la publication no.4. L'approche de clonage en cinq étapes utilisée pour l'insertion du génome du HCoV-OC43 dans le BAC est très innovatrice. Elle prend avantage de la présence de sites de restriction uniques *SfoI* dans le milieu du génome viral ainsi que dans le site de clonage multiple du BAC. Chacun des cinq fragments insérés dans la construction l'ont été par l'entremise du site *SfoI* et lorsque le site original était absent, un site temporaire était ajouté afin de permettre l'insertion du fragment, puis retiré lors de l'étape subséquente de clonage. Pour des raisons techniques, trois nouveaux sites de restriction ont également été insérés dans le génome du virus. Ces sites sont *MluI*, *AsiSI* et *SacII*, et sont respectivement situés aux positions 6480, 12 143 et 24 069 du génome viral. Il est à noter que ces sites de restriction servent également de marqueurs pour les particules virales recombinantes, lors d'expériences effectuées à partir du clone infectieux.

Tel que mentionné dans la revue bibliographique, le système du BAC comporte de nombreux avantages en soit, mais également par rapport aux autres systèmes disponibles. L'équipe du Dr Luis Enjuanes a également démontré que le pBAC-TGEV^{FL} pouvait être combiné à diverses applications de façon efficace (Almazán *et al.*, 2004; Almazán *et al.*, 2000; Galan *et al.*, 2005; Sola *et al.*, 2003; Sola *et al.*, 2005; Zúñiga *et al.*, 2004), suggérant que le pBAC-OC43^{FL} pourrait s'avérer aussi efficace et polyvalent. Du point de vue de l'assemblage, l'approche du BAC n'est pas la plus simple et est définitivement plus complexe que celle du clone d'ADNc linéaire. Lors de la construction du pBAC-OC43^{FL}, la difficulté principale fut de localiser la région instable du génome viral, laquelle est toxique chez *E. coli*, et de l'introduire au moment opportun dans le vecteur. Cette région a dû être insérée lors de la dernière étape de

clonage afin de limiter la toxicité et en ce qui nous concerne, la région toxique a été interrompue, ce qui a permis son sous-clonage dans un vecteur intermédiaire. Le tout a été rendu possible en générant des fragments de différentes longueurs de l'extrémité 3' de l'ORF1a, laquelle contiendrait la séquence virale instable. Bien que le BAC soit répliqué dans une souche *E. coli*, le faible nombre de copie du vecteur à l'intérieur de chacune des cellules bactériennes, soit une ou deux par cellules, permet de limiter la toxicité due aux séquences virales instables (Shizuya *et al.*, 1992). Cette propriété du BAC s'est d'ailleurs avérée la clé de la stratégie d'assemblage étant donné qu'un vecteur se répliquant à un taux élevé n'aurait pas pu supporter une telle toxicité. La nature exacte de cette toxicité n'a jamais été élucidée. Bien que la présence d'un promoteur cryptique aurait pu expliquer ce phénomène, il semblerait que la toxicité engendrée par la séquence coronavirale soit due à son expression (Yount *et al.*, 2002). Un promoteur cryptique présent dans le génome viral aurait pu entrer en compétition avec l'origine de répllication reconnue par *E. coli* et ainsi interférer avec la multiplication du vecteur. En créant une diminution du nombre de copies du gène de résistance à l'antibiotique présent sur le vecteur, une forme de toxicité aurait été générée pour la bactérie. Cependant, comme la région toxique du génome coronaviral est stable dans des vecteurs de levures, lesquels maintiennent un niveau d'expression constant du gène étranger, il aurait été plutôt suggéré que l'expression serait à l'origine de la toxicité observée. L'ambiguïté présente dans cette démonstration est que le vecteur utilisé possède une origine de répllication spécifique à la levure (Fabb et Ragoussis, 1995) et que son niveau de répllication est très faible, favorisant ainsi la stabilité de la région toxique.

Contrairement à l'assemblage du pBAC-OC43^{FL}, l'utilisation et la manipulation du produit fini sont relativement simples, ce qui représente un autre avantage majeur du système. Le BAC est un vecteur dont l'efficacité de clonage est extrêmement élevée. La manipulation de la séquence du génome viral, par mutagenèse dirigée ou génétique inverse, peut être effectuée de façon relativement simple en utilisant des méthodes de clonage conventionnelles. De plus, le génome du HCoV-OC43 est entièrement cloné dans des vecteurs intermédiaires en cinq régions distinctes, ce qui permet de travailler avec un segment précis du génome sans risquer d'introduire de mutations indésirables dans d'autres régions. Les segments modifiés sont ensuite excisés puis introduits dans le pBAC-OC43^{FL} en utilisant les sites de restriction uniques présents dans la construction. Cette approche facilite également les analyses du génome par séquençage et partage des similitudes avec l'approche du Dr Ralph Baric, dans laquelle chacun des segments du génome viral est compris dans un vecteur intermédiaire (Yount *et al.*, 2000).

Dans le système impliquant le BAC, la production de particules virales recombinantes est aussi relativement simple, surtout lorsque le système est comparé à celui des autres approches utilisées chez les coronavirus, ce qui constitue un avantage supplémentaire. Dans l'approche du clone linéaire par exemple, chaque production de virus recombinants nécessite le réassemblage du clone en entier suivi de sa transcription *in vitro*. L'ARN correspondant au génome du virus est ensuite transfecté dans des cellules eucaryotes et les particules virales produites (Yount *et al.*, 2000). Dans l'approche impliquant un vecteur viral, le clone d'ADNc doit d'abord subir une ronde d'amplification dans des cellules CV-1, susceptibles au virus de la vaccine. Suite à l'extraction du vecteur viral, les copies d'ADNc du génome du virus à l'étude doivent ensuite être excisées du génome du virus de la vaccine et transcrites *in vitro*, avant d'être transfectées dans une nouvelle lignée cellulaire (Thiel *et al.*, 2001). En ce qui concerne l'approche du BAC, l'ADN correspondant au pBAC-OC43^{FL} est directement transfecté dans des cellules BHK-21, lesquelles permettent à la fois la transfection du clone et l'infection par le HCoV-OC43 recombinant. Les particules virales recombinantes contenues dans les surnageants de culture peuvent alors être directement utilisées pour des expériences subséquentes. Le système du pBAC-OC43^{FL} comporte même un avantage par rapport à celui du pBAC-TGEV^{FL} puisqu'aucune étape d'amplification n'est nécessaire à la production du HCoV-OC43 recombinant. Avec le pBAC-TGEV^{FL}, les virus recombinants sont produits par la lignée cellulaire ST, susceptible au TGEV, mais le clone d'ADNc doit préalablement être transfecté dans des cellules BHK-21, les cellules ST n'étant pas transfectables (Gonzalez *et al.*, 2002). Cette étape ne représente pas un inconvénient majeur même si elle rallonge le processus de production, puisqu'elle permet de concentrer le virus et d'obtenir des titres viraux très élevés. Une telle étape d'amplification pourrait d'ailleurs être employée avec le pBAC-OC43^{FL} afin de concentrer le virus. Il a également été démontré par l'équipe du Dr Ralph Baric que la co-transfection du clone d'ADNc et du transcrit du gène N coronaviral pouvait augmenter la production et la récupération de particules virales recombinantes (Yount *et al.*, 2000). L'équipe du Dr Volker Thiel a aussi élaboré une variante de ce système en produisant son virus recombinant dans une lignée cellulaire qui exprime le gène N coronaviral de façon stable (Coley *et al.*, 2005). Une telle approche pourrait être envisageable pour la production et la récupération plus efficaces de particules recombinantes du HCoV-OC43.

Bien que le clone infectieux du HCoV-OC43 n'ait pas encore été assigné à une application particulière, le pBAC-OC43^{FL} représente un outil important, autant pour la caractérisation moléculaire du virus que pour les possibilités qu'il offre dans le domaine

thérapeutique. Connaissant également les similitudes que le HCoV-OC43 partage avec le SARS-CoV, le clone d'ADNc pourrait s'avérer un outil précieux pour la caractérisation de ce dernier, en évitant les inconvénients entraînés par un confinement de niveau 3. Le pBAC-OC43^{FL} permettra également de poursuivre la caractérisation de l'adaptation du HCoV-OC43 dans des cultures de cellules neurales et éventuellement dans le SNC et le cerveau humain. En utilisant la mutagenèse dirigée, les mutations identifiées dans le gène S des particules virales présentant le plus haut taux d'adaptation seront introduites dans le clone d'ADNc et leur impact sur la biologie du virus sera ensuite évalué. Par exemple, les mutations prédominantes D24Y, S83T, H183R et Y241H, toutes situées dans le SLR, pourraient être introduites dans le génome du virus, individuellement ou en combinaison, étant donné leur influence potentielle sur l'interaction avec le récepteur cellulaire. La combinaison des mutations H183R et Y241H pourrait également être analysée puisque ces mutations sont présentes chez les cinq souches qui démontrent la plus importante modulation de tropisme. La mutation V980A, laquelle pourrait être associée aux propriétés fusiogéniques du virus, représenterait un autre candidat intéressant pour ce genre d'analyses. L'impact de mutations situées dans des domaines fonctionnels importants de la protéine S ou qui présentent une homologie avec certaines mutations caractérisées chez d'autres coronavirus pourrait aussi être étudié. Ces analyses pourront par conséquent permettre de répondre à plusieurs questions en ce qui concerne les propriétés virales reliées à la protéine S. Le tropisme cellulaire du virus, les propriétés de fusion virus-cellule et cellule-cellule ainsi que les propriétés reliées à la virulence pourraient donc être caractérisés par l'entremise des mutations énumérées précédemment. Le clone d'ADNc pourrait également permettre la caractérisation moléculaire de régions du virus dont l'importance biologique a récemment été démontrée. Parmi les régions ou les gènes analysés, notons le gène de nucléocapside (Almazán *et al.*, 2004; Chen *et al.*, 2002; Schelle *et al.*, 2005), le gène E (Madan *et al.*, 2005; Wilson *et al.*, 2004) ainsi que certaines régions du gène de la réplicase, telles que les PLP, lesquelles posséderaient potentiellement une activité reliée à la déubiquitination (Lindner *et al.*, 2005; Sulea *et al.*, 2005). L'insertion d'un gène rapporteur tel que la *green fluorescent protein* (GFP) dans le génome coronaviral a déjà été rapporté (Bosch *et al.*, 2004; Sola *et al.*, 2003) et pourrait s'avérer une perspective intéressante pour le pBAC-OC43^{FL}. Lorsqu'elle est exprimée par le virus, la GFP permet de retracer rapidement le virus dans différents systèmes, *in vitro* et *in vivo*.

Les clones infectieux peuvent également être employés dans diverses applications thérapeutiques, telles que la vaccination, la thérapie génique ou la thérapie anti-cancer. Les vecteurs viraux sont de plus en plus utilisés dans ce genre d'approches puisque ceux-ci permettent

le transport et l'expression de gènes thérapeutiques à des sites spécifiques (Azzam et Domb, 2004). Les techniques de mutagenèse et de génétique inverse associées aux clones d'ADNc ont notamment pavé la voie au développement de vaccins vivants atténués et de vecteurs vaccinaux. L'utilisation des clones d'ADNc en tant qu'outil thérapeutique est aussi due à leur capacité d'inclure et d'exprimer des protéines étrangères de façon très stable (Neumann *et al.*, 2002). Bien qu'aucun clone infectieux coronaviral n'ait encore été développé pour la thérapie génique, il a été démontré par l'équipe du Dr Volker Thiel que le clone d'ADNc du HCoV-229E pouvait être utilisé afin d'effectuer la transduction de cellules dendritiques humaines matures et immatures, ce qui implique un rôle potentiel du vecteur dans l'immunothérapie (Thiel *et al.*, 2003b). Mais peu importe l'application qu'on veut leur assigner, les coronavirus peuvent s'avérer très attrayants comme vecteurs thérapeutiques parce que 1) la délétion de leurs gènes non essentiels est très atténuante *in vivo*, 2) l'espace ainsi créé permet l'insertion d'information étrangère et 3) leur tropisme peut être modifié en manipulant le gène S (de Haan *et al.*, 2003).

Récemment, il a également été démontré que les coronavirus pouvaient être employés comme vecteurs dans la thérapie anti-cancer et les deux approches utilisées se sont avérées plutôt concluantes *in vitro*. Cependant, il reste à démontrer que le concept élaboré puisse être appliqué *in vivo*. Dans la première approche, un anticorps simple chaîne bispécifique a été utilisé afin de cibler des cellules cancéreuses. Le FIPV et le fMHV, lequel est muni de la protéine S du FIPV, pouvaient donc être ciblés à des cellules cancéreuses humaines par l'anticorps simple chaîne, et les détruire. La liaison du virus à la cellule cancéreuse était assurée par l'anticorps, à la fois spécifique à la protéine S du coronavirus félin et au récepteur du facteur de croissance épidermal humain (EGFR; *epidermal growth factor receptor*), une molécule communément surexprimée dans plusieurs types de cancers (Würdinger *et al.*, 2005b). Dans la deuxième approche, le concept était très similaire. Une construction génétique permettant de coupler un anticorps simple chaîne spécifique au EGFR et le domaine amino-terminal du CEACAM1a, lequel est suffisant pour l'attachement de la protéine S du coronavirus murin, a été effectué. Dans un second temps, le MHV était ciblé aux cellules cancéreuses, par l'entremise de la molécule nouvellement synthétisée, et induisait la destruction de celles-ci (Würdinger *et al.*, 2005a). Ceci démontre encore une fois les vastes possibilités d'applications offertes par l'utilisation des coronavirus ou des clones d'ADNc dérivés de ceux-ci. Les clones infectieux coronaviraux représentent donc un outil puissant pour disséquer le cycle de vie du virus, l'assemblage du virus, le rôle des protéines virales dans la pathogenèse et l'interaction des protéines virales avec les

composantes de la réponse immunitaire de la cellule hôte et de l'organisme, et pourront certainement être associés à de nouvelles applications dans un avenir très rapproché.

RÉFÉRENCES BIBLIOGRAPHIQUES

1. Adami, C., Pooley, J., Glomb, J., Stecker, E., Fazal, F., Fleming, J. O., Baker, S. C. (1995) Evolution of mouse hepatitis virus (MHV) during chronic infection: quasispecies nature of the persisting MHV RNA. *Virology*, 209, 337-346.
2. Addie, D. D., Schaap, I. A. T., Nicolson, L., Jarrett, O. (2003) Persistence and transmission of natural type I feline coronavirus infection. *J. Gen. Virol.*, 84, 2735-2744.
3. Ahmed, R., Canning, W. M., Kaffman, R. S., Sharpe, A. H., Hallun, J. V., Fields, B. N. (1981) Role of the host cell in persistent viral infection: coevolution of L cells and reovirus during persistent infection. *Cell*, 25, 325-332.
4. Almazán, F., Galán, C., Enjuanes, L. (2004) The nucleoprotein is required for efficient coronavirus genome replication. *J. Virol.*, 78, 12683-12688.
5. Almazán, F., Gonzalez, J. M., Péntes, Z., Izeta, A., Calvo, E., Plana-Duran, J., Enjuanes, L. (2000) Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. USA*, 97, 5516-5521.
6. Alonso, S., Izeta, A., Sola, I., Enjuanes, L. (2002) Transcription regulatory sequences and mRNA expression levels in the coronavirus transmissible gastroenteritis virus. *J. Virol.*, 76, 1293-1308.
7. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., Hilgenfeld, R. (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain. *EMBO J.*, 21, 3213-3224.
8. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., Hilgenfeld, R. (2003) Coronavirus main proteinase (3CL^{Pro}) structure: Basis for design of anti-SARS drugs. *Science*, 300, 1763-1767.

9. Arbour, N., Côté, G., Lachance, C., Tardieu, M., Cashman, N. R., Talbot, P. J. (1999a) Acute and persistent infection of human neural cell lines by human coronavirus OC43. *J. Virol.*, 73, 3338-3350.
10. Arbour, N., Day, R., Newcombe, J., Talbot, P. (2000) Neuroinvasion by human respiratory coronaviruses. *J. Virol.*, 74, 8913-8921.
11. Arbour, N., Ékandé, S., Côté, G., Lachance, C., Chagnon, F., Tardieu, M., Cashman, N. R., Talbot, P. J. (1999b) Persistent infection of human oligodendrocytic and neuroglial cell lines by human coronavirus 229E. *J. Virol.*, 73, 3326-3337.
12. Azzam, T., Domb, A. J. (2004) Current developments in gene transfection agents. *Curr. Drug Deliv.*, 1, 165-193.
13. Babcock, G. J., Eshaki, D. J., Thomas, Jr., W. D., Ambrosino, D. M. (2004) Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor. *J. Virol.*, 78, 4552-4560.
14. Baranowski, E., Ruiz-Jarabo, C. M., Domingo, E. (2001) Evolution of cell recognition by viruses. *Science*, 292, 1102-1105.
15. Barnaba, V., Sinigaglia, F. (1997) Molecular mimicry and T cell-mediated autoimmune disease. *J. Exp. Med.*, 185, 1529-1531.
16. Barnett, E. M., Perlman, S. (1993) The olfactory nerve and not the trigeminal nerve is the major site of CNS entry for mouse hepatitis virus, strain JHM. *Virology*, 194, 185-191.
17. Barthold, S. W., Smith, A. L. (1992) Viremic dissemination of mouse hepatitis virus-JHM following intranasal inoculation of mice. *Arch. Virol.*, 122, 35-44.
18. Basler, C. F., Garcia-Sastre, A. (2002) Viruses and type I interferon antiviral system : induction and evasion. *Int. Rev. Immunol.*, 21, 305-337.

19. Birch, C. J., Clothier, H. J., Seccull, A., Tran, T., Catton, M. C., Lambert, S. B., Druce, J. D. (2005) Human coronavirus OC43 causes influenza-like illness in residents and staff of aged-care facilities in Melbourne, Australia. *Epidemiol. Infect.*, 133, 273-277.
20. Bisht, H., Roberts, A., Vogel, L., Bukreyev, A., Collins, P. L., Murphy, B. R., Subbarao, K., Moss, B. (2004) Severe acute respiratory syndrome coronavirus spike protein expressed by attenuated vaccinia virus protectively immunizes mice. *Proc. Natl. Acad. Sci. USA*, 101, 6641-6646.
21. Blau, D. M., Holmes, K. V. (2001) Human coronavirus HCoV-229E enters susceptible cells via the endocytic pathway. *Adv. Exp. Med. Biol.*, 494, 193-198.
22. Bonavia, A., Arbour, N., Yong, V. W., Talbot, P. J. (1997) Infection of primary cultures of human neural cells by human coronaviruses 229E and OC43. *J. Virol.*, 71, 800-806.
23. Bosch, B. J., de Haan, C. A. M., Rottier, P. J. M. (2004) Coronavirus spike glycoprotein, extended at the carboxy terminus with green fluorescent protein, is assembly competent. *J. Virol.*, 78, 7369-7378.
24. Bosch, B. J., de Haan, C. A. M., Smits, S. L., Rottier, P. J. M. (2005) Spike protein assembly into the coronavirus: exploring the limits of its sequence requirements. *Virology*, 334, 306-318.
25. Bosch, B. J., van der Zee, R., de Haan, C. A. M., Rottier, P. J. M. (2003) The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.*, 77, 8801-8811.
26. Breslin J. J., Mork, I., Smith, M. K., Vogel, L. K., Hemmila, E. M., Bonavia, A., Talbot, P. J., Sjöström, H., Norén, O., Holmes, K. V. (2003) Human coronavirus 229E: receptor binding domain and neutralization by soluble receptor at 37°C. *J. Virol.*, 77, 4435-4438.
27. Brockway, S. M., Clay, C. T., Lu, X. T., Denison, M. R. (2003) Characterization of the expression, intracellular localization, and replication complex association of the putative mouse hepatitis virus RNA-dependant RNA polymerase. *J. Virol.*, 77, 10515-10527.

28. Brockway, S. M., Denison, M. R. (2005) Mutagenesis of the murine hepatitis virus nsp1-coding region identifies residues important for protein processing, viral RNA synthesis, and viral replication. *Virology*, 340, 209-223.
29. Casais, R., Thiel, V., Siddell, S. G., Cavanagh, D., Britton, P. (2001) Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J. Virol.*, 75, 12359-12369.
30. Calvo, E., Escors, D., López, J. A., González, J. M., Álvarez, A., Arza, E., Enjuanes, L. (2005) Phosphorylation and subcellular localization of transmissible gastroenteritis virus nucleocapsid protein in infected cells. *J. Gen. Virol.*, 86, 2255-2267.
31. Che, X.-y., Qiu, L.-w., Liao, Z.-y., Wang, Y.-d., Wen, K., Pan, Y.-x., Hao, W., Mei, Y.-b., Cheng, V. C. C., Yuen, K.-y. (2005) Antigenic cross-reactivity between severe acute respiratory syndrome-associated coronavirus and human coronaviruses 229E and OC43. *J. Infect. Dis.*, 191, 2033-2037.
32. Chen, H., Torsten, W., Britton, P., Brooks, G., Hiscox, J. A. (2002) Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell. *J. Virol.*, 76, 5233-5250.
33. Choi, K. S., Aizaki, H., Lai, M. M. C. (2005) Murine coronavirus requires lipid rafts for virus entry and cell-cell fusion but not for virus release. *J. Virol.*, 79, 9862-9871.
34. Cinatl, Jr., J., Michaelis, M., Hoever, G., Preiser, W., Doerr, H. W. (2005) Development of antiviral therapy for severe acute respiratory syndrome. *Antiviral Res.*, 66, 81-97.
35. Coley, S. E., Lavi, E., Sawicki, S. G., Fu, L., Schelle, B., Karl, N., Siddell, S. G., Thiel, V. (2005) Recombinant mouse hepatitis virus strain A59 from cloned, full-length cDNA replicates to high titers in vitro and is fully pathogenic in vivo. *J. Virol.*, 79, 3097-3106.
36. Collins, A. R. (2002) In vitro detection of apoptosis in monocytes/macrophages infected with human coronavirus. *Clin. Diagn. Lab. Immunol.*, 9, 1392-1395.

37. Compton, S. R., Barthold, S. W., Smith, A. L. (1993) The cellular and molecular pathogenesis of coronaviruses. *Lab. Anim. Sci.*, 43, 15-28.
38. Coyle, P. K. (1996) The neuroimmunology of multiple sclerosis. *Adv. Neuroimmunol.*, 6, 143-154.
39. Cullen, B. R. (2001) Journey to the center of the cell. *Cell*, 105, 697-700.
40. Curtis, K. M., Yount, B., Sims, A. C., Baric, R. S. (2004) Reverse genetic analysis of the transcription regulatory sequence of the coronavirus transmissible gastroenteritis virus. *J. Virol.*, 78, 6061-6066.
41. Dalziel, R. G., Lampert, P. W., Talbot, P. J., Buchmeier, M. J. (1986) Site-specific alteration of murine hepatitis virus type 4 peplomer glycoprotein E2 results in reduced neurovirulence. *J. Virol.*, 59, 463-471.
42. Daniel, C., Anderson, R., Buchmeier, M. J., Fleming, J. O., Spaan, W. J. M., Wege, H., Talbot, P. J. (1993) Identification of an immunodominant linear neutralization domain on the S2 portion of the murine coronavirus spike glycoprotein and evidence that it forms part of a complex tridimensional structure. *J. Virol.*, 67, 1185-1194.
43. Decman, V., Freeman, M. L., Kinchington, P. R., Hendricks, R. L. (2005) Immune control of HSV-1 latency. *Viral Immunol.*, 18, 466-473.
44. de Haan, C. A. M., Masters, P. S., Shen, X., Weiss, S., Rottier, P. J. M. (2002a) The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host. *Virology*, 296, 177-189.
45. de Haan, C. A. M., Smeets, M., Vernooij, F., Vennema, H., Rottier, P. J. M. (1999) Mapping of the coronavirus membrane protein domains involved in interaction with the spike protein. *J. Virol.*, 73, 7441-7452.

46. de Haan, C. A. M., Stadler, K., Godeke, G.-J., Bosch, B. J., Rottier, P. J. M. (2004) Cleavage inhibition of the murine coronavirus spike protein by a furin-like enzyme affects cell-cell but not virus-cell fusion. *J. Virol.*, 78, 6048-6054.
47. de Haan, C. A. M., van Genne, L., Stoop, J. N., Volders, H., Rottier, P. J. M. (2003) Coronaviruses as vectors: position dependence of foreign gene expression. *J. Virol.*, 77, 11312-11323.
48. de Haan, C. A. M., Volders, H., Koetzner, C. A., Masters, P. S., Rottier, P. J. M. (2002b) Coronaviruses maintain viability despite dramatic rearrangements of the strictly conserved genome organization. *J. Virol.*, 76, 12491-12502.
49. de Haan, C. A. M., Zhen, L., te Lintelo, E., Bosch, B. J., Haijema, B. J., Rottier, P. J. M. (2005) Murine coronavirus with an extended host range uses heparan sulfate as an entry receptor. *J. Virol.*, 79, 14451-14456.
50. de Vries, A. A., Glaser, A. L., Raamsman, M. J., de Haan, C. A., Sarnataro, S., Godeke, G. J., Rottier, P. J. (2000) Genetic manipulation of equine arteritis virus using full-length cDNA clones: separation of overlapping genes and expression of a foreign epitope. *Virology*, 270, 84-97.
51. Domingo, E., Baranowski, E., Ruiz-Jarabo, C. M., Martín-Hernández, A. M., Sáiz, J. C., Escarmís, C. (1998) Quasispecies structure and persistence of RNA viruses. *Emerg. Infect. Dis.*, 4, 521-527.
52. Domingo, E., Ruiz-Jarabo, C. M., Sierra, S., Arias, A., Pariente, N., Baranowski, E., Escarmís, C. (2002) Emergence and selection of RNA virus variants: memory and extinction. *Virus Res.*, 82, 39-44.
53. Dove, B., Cavanagh, D., Britton, P. (2004) Presence of an encephalomyocarditis virus internal ribosome entry site sequence in avian infectious bronchitis virus defective RNAs abolishes rescue by helper virus. *J. Virol.*, 78, 2711-2721.

54. Duquerroy, S., Vigouroux, A., Rottier, P. J. M., Rey, F. A., Bosch, B. J. (2005) Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein. *Virology*, 335, 276-285.
55. Egloff, M.-P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E., Gorbalenya, A. E., Cambillau, C., Canard, B. (2004) The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc. Natl. Acad. Sci. USA*, 101, 3792-3796.
56. Enjuanes, L., Sola, I., Almazan, F., Izeta, A., Gonzalez, J. M., Alonso, S. (2001) Coronavirus derived expression systems. *Adv. Exp. Med. Biol.*, 494, 309-321.
57. Elena, S. F., Sanjuán, R. (2005) Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J. Virol.*, 79, 11555-11558.
58. Escors, D., Izeta, A., Capiscol, C., Enjuanes, L. (2003) Transmissible gastroenteritis coronavirus packaging signal is located at the 5' end of the virus genome. *J. Virol.*, 77, 7890-7902.
59. Fabb, S. A., Ragoussis, J. (1995) Yeast artificial chromosome vectors. *Mol. Cell Biol. Hum. Dis. Ser.*, 5, 104-124.
60. Fang, S. G., Shen, S., Tay, F. P. L., Liu, D. X. (2005) Selection of and recombination between minor variants lead to the adaptation of an avian coronavirus to primate cells. *Biochem. Biophys. Res. Comm.*, 336, 417-423.
61. Fang, X., Ye, L., Timani, K. A., Li, S., Zen, Y., Zhao, M., Zheng, H., Wu, Z. (2005) Peptide domain involved in the interaction between membrane protein and nucleocapsid protein of SARS-associated coronavirus. *Biochem. Mol. Biol.*, 38, 381-385.
62. Ferron, F., Longhi, S., Henrissat, B., Canard, B. (2002) Viral RNA-polymerases -- a predicted 2'-O-ribose methyltransferase domain shared by all Mononegavirales. *Trends Biochem. Sci.*, 27, 222-224.

63. Fleming, J. O., Trousdale, M. D., Bradbury, J., Stohlman, S. A., Weiner, L. P. (1987) Experimental demyelination induced by coronavirus JHM (MHV-4): molecular identification of a viral determinant of paralytic disease. *Microb. Pathog.*, 3, 9-20.
64. Follis, K. E., York, J., Nunberg, J. H. (2005) Serine-scanning mutagenesis studies of the C-terminal heptad repeats in the SARS coronavirus S glycoprotein highlight the important role of the short helical region. *Virology*, 341, 122-129.
65. Fouchier, R. A. M., Hartwig, N. G., Bestebroer, T. M., Niemeyer, B., de Jong, J. C., Simon, J. H., Osterhaus, A. D. M. E. (2004) A previously undescribed coronavirus associated with respiratory disease in humans. *Proc. Natl. Acad. Sci. USA*, 101, 6212-6216.
66. Fu, L., Gonzales, D. M., Sarma, J. D., Lavi, E. (2004) A combination of mutations in the S1 part of the spike glycoprotein gene of coronavirus MHV-A59 abolishes demyelination. *J. Neurovirol.*, 10, 41-51.
67. Gagneur, A., Legrand, M. C., Picard, B., Baron, R., Talbot, P. J., de Parscau, L., Sizun, J. (2002) Infections nosocomiales à coronavirus humains chez le nouveau-né. *Arch. Pédiatr.*, 9, 61-69.
68. Galan, C., Enjuanes, L., Almazán, F. (2005) A point mutation within the replicase gene differentially affects coronavirus genome versus minigenome replication. *J. Virol.*, 79, 15016-15026.
69. Gallagher, T. M., Buchmeier, M. J. (2001) Coronavirus spike proteins in viral entry and pathogenesis. *Virology*, 279, 371-374.
70. Garcia-Sastre, A., Durbin, R. K., Zheng, H., Palese, P., Gertner, R., Levy, D. E., Durbin, J. E. (1998) The role of interferon in influenza virus tissue tropism. *J. Virol.*, 72, 8550-8558.
71. Gélinas, A.-M., Boutin, M., Sasseville, M.-J., Dea, S. (2001) Bovine coronaviruses associated with enteric and respiratory diseases in canadian dairy cattle display different

- reactivities to *anti*-HE monoclonal antibodies and distinct amino acid changes in their HE, S and ns4.9 protein. *Virus Res.*, 76, 43-57.
72. Goebel, S. J., Hsue, B., Dombrowski, T. F., Masters, P. S. (2004) Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. *J. Virol.*, 78, 669-682.
73. Gonzalez, J. M., Pénczes, Z., Almazan, F., Calvo, E., Enjuanes, L. (2002) Stabilization of a full-length infectious cDNA clone of transmissible gastroenteritis coronavirus by insertion of an intron. *J. Virol.*, 76, 4655-4661.
74. Gorbalenya, A. E., Snijder, E. J., Spaan, W. J. M. (2004) Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. *J. Virol.*, 78, 7863-7866.
75. Gosert, R., Kanjanahaluethai, A., Egger, D., Bienz, K., Baker, S. C. (2002) RNA replication of mouse hepatitis virus takes place at double-membrane vesicles. *J. Virol.*, 76, 3697-3708.
76. Groneberg, D. A., Poutanen, S. M., Low, D. E., Lode, H., Welte, T., Zabel, P. (2005) Treatment and vaccines for severe acute respiratory syndrome. *The Lancet*, 5, 147-155.
77. Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Guan, Y. J., Butt, K. M., Wong, K. L., Chan, K. W., Lim, W., Shortridge, K. F., Yuen, K. Y., Peiris, J. S. M., Poon, L. L. M. (2003) Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. *Science*, 302, 276-278.
78. Guillén, J., Pérez-Berná, A. J., Moreno, M. R., Villalain, J. (2005) Identification of the membrane-active regions of the severe acute respiratory syndrome coronavirus spike membrane glycoprotein using a 16/18-mer peptide scan : implications for the viral fusion mechanism. *J. Virol.*, 79, 1743-1752.
79. Hafler, D. A., Slavik, J. M., Anderson, D. E., O'Connor, K. C., De Jager, P., Baecher-Allan, C. (2005) Multiple sclerosis. *Immunol. Rev.*, 204, 208-231.

80. Haijema, B. J., Volders, H., Rottier, P. J. M. (2003) Switching species tropism: an effective way to manipulate the feline coronavirus genome. *J. Virol.*, 77, 4528-4538.
81. Hamre, D., Procknow, J. J. (1966) A new virus isolated from the human respiratory tract. *Proc. Soc. Exp. Biol. Med.*, 121, 190-193.
82. Harcourt, B. H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., Severson, K. M., Smith, C. M., Rota, P. A., Baker, S. C. (2004) Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J. Virol.*, 78, 13600-13612.
83. He, R., Leeson, A., Ballantine, M., Andonov, A., Baker, L., Dobie, F., Li, Y., Bastien, N., Feldmann, H., Strocher, U., Theriault, S., Cutts, T., Cao, J., Booth, T. F., Plummer, F. A., Tyler, S., Li, X. (2004) Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. *Virus Res.*, 105, 121-125.
84. Hegyi, A., Friebe, A., Gorbalenya, A. E., Ziebuhr, J. (2002) Mutational analysis of the active center of coronavirus 3C-like proteases. *J. Gen. Virol.*, 83, 581-593.
85. Hegyi, A., Kolb, A. F. (1998) Characterization of determinants involved in the feline infectious peritonitis virus receptor function of feline aminopeptidase N. *J. Gen. Virol.*, 79, 1387-1391.
86. Hegyi, A., Ziebuhr, J. (2002) Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.*, 83, 595-599.
87. Hertzog, T., Scandella, E., Schelle, B., Ziebuhr, J., Siddell, S. G., Ludewig, B., Thiel, V. (2004) Rapid identification of coronavirus replicase inhibitors using a selectable replicon RNA. *J. Gen. Virol.*, 85, 1717-1725.
88. Hingley, S. T., Leparc-Goffart, I., Seo, S.-h., Tsai, J. C., Weiss, S. R. (2002) The virulence of mouse hepatitis virus strain A59 is not dependent on efficient spike protein cleavage and cell-to-cell fusion. *J. Neurovirol.*, 8, 400-410.

89. Hofmann, H., Pöhlmann, S. (2004) Cellular entry of the SARS coronavirus. *Trends Microbiol.*, 12, 466-472.
90. Hofmann, H., Pyrc, K., van der Hoek, L., Geier, M., Berkhout, B., Pöhlmann, S. (2005) Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc. Natl. Acad. Sci. USA*, 102, 7988-7993.
91. Holmes, K. V. (2003) SARS-associated coronavirus. *N. Engl. J. Med.*, 348, 1948-1951.
92. Holz, A., Bielekova, B., Martin, R., Oldstone, M. B. A. (2000) Myelin-associated oligodendrocytic basic protein: identification of an encephalitogenic epitope and association with multiple sclerosis. *J. Immunol.*, 164, 1103-1109.
93. Houtman, J. J., Fleming, J. O. (1996) Pathogenesis of mouse hepatitis virus-induced demyelination. *J. Neurovirol.*, 2, 361-376.
94. Huang, Q., Yu, L., Petros, A. M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S. W., Olejniczak, E. T. (2004) Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry*, 43, 6059-6063.
95. Huang, Y., Yang, Z.-y., Kong, W.-p., Nabel, G. J. (2004) Generation of synthetic severe acute respiratory syndrome coronavirus pseudoparticles: implications for assembly and vaccine production. *J. Virol*, 78, 12557-12565.
96. Hung, E. C. W., Chim, S. S. C., Chan, P. K. S., Tong, Y. K., Ng, E. K. O., Chiu, R. W. K., Leung, C.-B., Sung, J. J. Y., Tam, J. S., Lo, Y. M. D. (2003) Detection of SARS coronavirus RNA in the cerebrospinal fluid of a patient with severe acute respiratory syndrome. *Clin. Chem.*, 49, 2108-2109.
97. Hunter, S. M., Rodriguez, M. (1995) Multiple sclerosis: a unique immunopathological syndrome of the central nervous system. *Springer Semin. Immunopathol.*, 17, 89-105.

98. Ingallinella, P., Bianchi, E., Finotto, M., Cantoni, G., Eckert, D. M., Supekar, V. M., Bruckmann, C., Carfi, A., Pessi, A. (2004) Structural characterization of the fusion-active complex of severe acute respiratory syndrome (SARS) coronavirus. *Proc. Natl. Acad. Sci. USA*, 101, 8709-8714.
99. Ito, N., Mossel, E. C., Narayanan, K., Popov, V. L., Huang, C., Inoue, T., Peters, C. J., Makino, S. (2005) Severe acute respiratory syndrome coronavirus 3a protein is a viral structural protein. *J. Virol.*, 79, 3182-3186.
100. Ivanov, K. A., Hertzog, T., Rozanov, M., Bayer, S., Thiel, V., Gorbalenya, A. E., Ziebuhr, J. (2004a) Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Natl. Acad. Sci. USA*, 101, 12694-12699.
101. Ivanov, K. A., Thiel, V., Dobbe, J. C., van der Meer, Y., Snijder, E. J., Ziebuhr, J. (2004b) Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J. Virol.*, 78, 5619-5632.
102. Izeta, A., Smerdou, C., Alonso, S., Penzes, Z., Mendez, A., Plana-Duran, J., Enjuanes, L. (1999) Replication and packaging of transmissible gastroenteritis coronavirus-derived synthetic minigenomes. *J. Virol.*, 73, 1535-1545.
103. Jacomy, H., Fragoso, G., Almazan, G., Mushynski, W. E., Talbot, P. J. (2006) Human coronavirus OC43 infection induces chronic encephalitis leading to disabilities in BALB/c mice. *Virology*, Epub ahead of print.
104. Jacomy, H., Talbot, P. J. (2001) Susceptibility of murine CNS to OC43 infection. *Adv. Exp. Med. Biol.*, 494, 101-107.
105. Jacomy, H., Talbot, P. J. (2003) Vacuolating encephalitis in mice infected by human coronavirus OC43. *Virology*, 315, 20-33.
106. Jeffers, S. A., Tusell, S. M., Gillim-Ross, L., Hemmila, E. M., Achenbach, J. E., Babcock, G. J., Thomas, Jr., W. D., Thackray, L. B., Young, M. D., Mason, R. J., Ambrosino, D. M., Wentworth, D. E., DeMartini, J. C., Holmes, K. V. (2004) CD209L (L-SIGN) is a

- receptor for severe acute respiratory syndrome coronavirus. Proc. Natl. Acad. Sci. USA, 101, 15748-15753.
107. Jonassen, C. M., Kofstad, T., Larsen, I.-L., Lovland, A., Handeland, K., Follestad, A., Lillehaug, A. (2005) Molecular identification and characterization of novel coronaviruses infecting graylag geese (*Anser anser*), feral pigeons (*Columbia livia*) and mallards (*Anas platyrhynchos*). J. Gen. Virol., 86, 1597-1607.
 108. Kamahora, T., Soe, L. H., Lai, M. M. C. (1989) Sequence analysis of nucleocapsid gene and leader RNA of human coronavirus OC43. Virus Res., 12, 1-9.
 109. Kazi, L., Lissenberg, A., Watson, R., de Groot, R. J., Weiss, S. R. (2005) Expression of hemagglutinin esterase protein from recombinant mouse hepatitis virus enhances neurovirulence. J. Virol., 79, 15064-15073.
 110. Klausegger, A., Strobl, B., Regl, G., Kaser, A., Luytjes, W., Vlasak, R. (1999) Identification of a coronavirus hemagglutinin-esterase with a substrate specificity different from those of influenza C virus and bovine coronavirus. J. Virol., 73, 3737-3743.
 111. Kligler, Y., Levanon, E. Y. (2003) Cloaked similarity between HIV-1 and SARS-CoV suggests an anti-SARS strategy. BMC Microbiol., 3, 1-7.
 112. Krueger, D. K., Kelly, S. M., Lewicki, D. N., Ruffolo, R., Gallagher, T. M. (2001) Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. J. Virol., 75, 2792-2802.
 113. Künkel, F., Herrler, G. (1996) Structural and functional analysis of the S proteins of two human coronavirus OC43 strains adapted to growth in different cells. Arch. Virol., 141, 1123-1131.
 114. Kuo, L., Godeke, G.-J., Raamsman, M. J. B., Masters, P. S., Rottier, P. J. M. (2000) Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. J. Virol., 73, 1393-1406.

115. Kuo, L., Masters, P. S. (2002) Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J. Virol.*, 76, 4987-4999.
116. Kuo, L., Masters, P. S. (2003) The small envelope protein E is not essential for murine coronavirus replication. *J. Virol.*, 77, 4597-4608.
117. Labonté, P., Mounir, S., Talbot, P. J. (1995) Sequence and expression of the ns2 protein gene of human coronavirus OC43. *J. Gen. Virol.*, 76, 431-435.
118. Lachance, C., Arbour, N., Cashman, N. R., Talbot, P. J. (1998) Involvement of aminopeptidase N (CD13) in infection of human neural cells by human coronavirus 229E. *J. Virol.*, 72, 6511-6519.
119. Lai, M. M. C. (1996) Recombination in large RNA viruses : Coronaviruses. *Semin. Virol.*, 7, 381-388.
120. Lai, M. M. C., Cavanagh, D. (1997) The molecular biology of coronaviruses. *Adv. Virus Res.*, 48, 1-97.
121. Lane, T. E., Buchmeier, M. J. (1997) Murine coronavirus infection : a paradigm for virus-induced demyelinating disease. *Trends Microbiol.*, 5, 9-14.
122. Lane, T. E., Liu, M. T., Chen, B. P., Asensio, V. C., Samawi, R. M., Paoletti, A. D., Campbell, I. L., Kunkel, S. L., Fox, H. S., Buchmeier, M. J. (2000) A central role for CD4⁺ T cells and RANTES in virus-induced central nervous system inflammation and demyelination. *J. Virol.*, 74, 1415-1424.
123. Lang, H. L. E., Jacobsen, H., Ikemizu, S., Andersson, C., Harlos, K., Madsen, L., Hjorth, P., Sondergaard, L., Svejgaard, A., Wucherpfennig, K., Stuart, D. I., Bell, J. I., Jones, E. Y., Fugger, L. (2002) A functional and structural basis for TCR cross-reactivity in multiple sclerosis. *Nature Immunol.*, 3, 940-943.

124. Lassnig, C., Sanchez, C. M., Egerbacher, M., Walter, I., Majer, S., Kolbe, T., Pallares, P., Enjuanes, L., Müller, M. (2005) Development of a transgenic mouse model susceptible to human coronavirus 229E. *Proc. Natl. Acad. Sci. USA*, 102, 8275-8280.
125. Lewicki, D. N., Gallagher, T. M. (2002) Quaternary structure of coronavirus spikes in complex with carcinoembryonic antigen-related cell adhesion molecule cellular receptors. *J. Biol. Chem.*, 277, 19727-19734.
126. Li, F., Li, W., Farzan, M., Harrison, S. C. (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*, 309, 1864-1868.
127. Li, T., Zhang, Y., Fu, L., Yu, C., Li, X., Li, Y., Zhang, X., Rong, Z., Wang, Y., Ning, H., Liang, R., Chen, W., Babiuk, L. A., Chang, Z. (2005) siRNA targeting the leader sequence of SARS-CoV inhibits virus replication. *Gene Ther.*, 12, 751-761.
128. Li, W., Moore, M. J., Vasilieva, N., Sui, J., Kee Wong, S., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., Farzan, M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426, 450-454.
129. Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Cramer, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B. T., Zhang, S., Wang, L. F. (2005a) Bats are natural reservoirs of SARS-like coronaviruses. *Science*, Epub ahead of print.
130. Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S.-K., Huang, I.-C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y., Choe, H., Farzan, M. (2005b) Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.*, 24, 1634-1643.
131. Lim, K. P., Liu, D. X. (2001) The missing link in coronavirus assembly. *J. Biol. Chem.*, 276, 17515-17523.

132. Lindner, H. A., Fotouhi-Ardakani, N., Lytvyn, V., Lachance, P., Sulea, T., Ménard, R. (2005) The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. *J. Virol.*, 79, 15199-15208.
133. Liu, Q., Johnson, R. F., Leibowitz, J. L. (2001) Secondary structural elements within the 3' untranslated region of mouse hepatitis virus strain JHM genomic RNA. *J. Virol.*, 75, 12105-12113.
134. Lontok, E., Corse, E., Machamer, C. E. (2004) Intracellular targeting signals contribute to localization of coronavirus spike proteins near the virus assembly site. *J. Virol.*, 78, 5913-5922.
135. Lu, L., Manopo, I., Leung, B. P., Chng, H. H., Ling, A. E., Chee, L. L., Ooi, E. E., Chan, S.-W., Kwang, J. (2004) Immunological characterization of the spike protein of the severe acute respiratory syndrome coronavirus. *J. Clin. Microbiol.*, 42, 1570-1576.
136. Madan, V., Garcia, M. de J., Sanz, M. A., Carrasco, L. (2005) Viroporin activity of murine hepatitis virus E protein. *FEBS Lett.*, 579, 3607-3612.
137. Marshall, E., Enserink, M. (2004) Caution urged on SARS vaccines. *Science*, 303, 944-946.
138. Marten, N. W., Stohlman, S. A., Bergmann, C. C. (2001) MHV infection of the CNS: mechanisms of immune-mediated control. *Viral Immunol.*, 14, 1-18.
139. Martín-Hernández, A. M., Carillo, E. C., Sevilla, N., Domingo, E. (1994) Rapid cell variation can determine the establishment of a persistent viral infection. *Proc. Natl. Acad. Sci. USA*, 91, 3705-3709.
140. Masters, P. S. (1999) Reverse genetics of the largest RNA viruses. *Adv. Virus Res.*, 53, 245-264.

141. Matsuyama, S., Taguchi, F. (2002a) Communication between S1N330 and a region in S2 of murine coronavirus spike protein is important for virus entry into cells expressing CEACAM1b receptor. *Virology*, 295, 160-171.
142. Matsuyama, S., Taguchi, F. (2002b) Receptor-induced conformational changes of murine coronavirus spike protein. *Journal of Virology*, 76 (23), 11819-11826.
143. Matthews, A. E., Weiss, S. R., Paterson, Y. (2002) Murine hepatitis virus -- a model for virus-induced CNS demyelination. *J. Neurovirol.*, 8, 76-85.
144. Mattson, M. P. (2004) Infectious agents and age-related neurodegenerative disorders. *Ageing Res. Rev.*, 3, 105-120.
145. McIntosh, K., Dees, J. H., Becker, W. B., Kapikian, A. Z., Chanock, R. M. (1967) Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci. USA*, 57, 933-940.
146. Méndez, A., Smerdou, C., Izeta, A., Gebauer, F., Enjuanes, L. (1996) Molecular characterization of transmissible gastroenteritis coronavirus defective interfering genomes : packaging and heterogeneity. *Virology*, 217, 495-507.
147. Meyers, G. (2003) Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependant reinitiation mechanism. *J. Biol. Chem.*, 278, 34051-34060.
148. Myint, S. H. (1994) Human coronaviruses : a brief review. *Rev. Med. Virol.*, 4, 35-46.
149. Nal, B., Chan, C., Kien, F., Siu, L., Tse, J., Chu, K., Kam, J., Staropoli, I., Crescenzo-Chaigne, B., Escriou, N., van der Werf, S., Yuen, K.-Y., Altmeyer, R. (2005) Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. *J. Gen. Virol.*, 86, 1423-1434.
150. Navas, S., Seo, S.-H., Chua, M. M., Das Sarma, J., Lavi, E., Hingley, S. T., Weiss, S. R. (2001) Murine coronavirus spike protein determines the ability of the virus to replicate in the liver and cause hepatitis. *J. Virol.*, 75, 2452-2457.

151. Navas-Martin, S., Hingley, S. T., Weiss, S. R. (2005) Murine coronavirus evolution in vivo: functional compensation of a detrimental amino acid substitution in the receptor binding domain of the spike glycoprotein. *J. Virol.*, 79, 7629-7640.
152. Neuman, B. W., Stein, D. A., Kroeker, A. D., Churchill, M. J., Kim, A. M., Kuhn, P., Dawson, P., Moulton, H. M., Bestwick, R. K., Iversen, P. L., Buchmeier, M. J. (2005) Inhibition, escape, and attenuated growth of severe acute respiratory syndrome coronavirus treated with antisense morpholino oligomers. *J. Virol.*, 79, 9665-9676.
153. Neuman, B. W., Stein, D. A., Kroeker, A. D., Paulino, A. D., Moulton, H. M., Iversen, P. L., Buchmeier, M. J. (2004) Antisense morpholino-oligomers directed against the 5' end of the genome inhibit coronavirus proliferation and growth. *J. Virol.*, 78, 5891-5899.
154. Neumann, G., Whitt, M. A., Kawaoka, Y. (2002) A decade after the generation of a negative-sense RNA virus from cloned cDNA - what have we learned? *J. Gen. Virol.*, 83, 2635-2662.
155. Ng, L. F. P., Wong, M., Koh, S., Ooi, E.-E., Tang, K.-F., Leong, H.-N., Ling, A.-E., Agathe, L. V., Tan, J., Liu, E. T., Ren, E.-C., Ng, L.-C., Hibberd, M. L. (2004) Detection of severe acute respiratory syndrome coronavirus in blood of infected patients. *J. Clin. Microb.*, 42, 347-350.
156. Nguyen, V.-P., Hogue, B. G. (1997) Protein interactions during coronavirus assembly. *J. Virol.*, 71, 9278-9284.
157. Nomura, R., Kiyota, A., Suzaki, E., Kataoka, K., Ohe, Y., Miyamoto, K., Senda, T., Fujimoto, T. (2004) Human coronavirus 229E binds to CD13 in rafts and enters the cell through caveolae. *J. Virol.*, 78, 8701-8708.
158. Novella, I. S. (2003) Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. *Curr. Opin. Microbiol.*, 6, 399-405.
159. Oldstone, M. B. A. (1989) Viral persistence. *Cell*, 56, 517-520.

160. Oldstone, M. B. A. (1998) Molecular mimicry and immune-mediated diseases. *FASEB J.*, 12, 1255-1265.
161. Olson, J. K., Croxford, J. L., Miller, S. D. (2001) Virus-induced autoimmunity: potential role of viruses in initiation, perpetuation, and progression of T-cell-mediated autoimmune disease. *Viral Immunol.*, 14, 227-250.
162. Ontiveros, E., Kim, T. S., Gallagher, T. M., Perlman, S. (2003) Enhanced virulence mediated by the murine coronavirus, mouse hepatitis virus strain JHM, is associated with a glycine at residue 310 of the spike glycoprotein. *J. Virol.*, 77, 10260-10269.
163. Ortego, J., Escors, D., Laude, H., Enjuanes, L. (2002) Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome. *J. Virol.*, 76, 11518-11529.
164. Ortego, J., Sola, I., Almazán, F., Ceriani, J. E., Riquelme, C., Balasch, M., Plana, J., Enjuanes, L. (2003) Transmissible gastroenteritis coronavirus gene 7 is not essential but influences *in vivo* virus replication and virulence. *Virology*, 308, 13-22.
165. Ozdarendeli, A., Ku, S., Rochat, S., Williams, G. D., Senanayake, S. D., Brian, D. A. (2001) Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. *J. Virol.*, 75, 7362-7374.
166. Parker, M. M., Masters, P. S. (1990) Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein. *Virology*, 179, 463-468.
167. Peiris, J. S., Chu C. M., Cheng, V. C., Chan, K. S., Hung, I. F., Poon, L. L., Law, K. I., Tang, B. S., Hon, T. Y., Chan, C. S., Chan, K. H., Ng, J. S., Zheng, B. J., Ng, W. L., Lai, R. W., Guan, Y., Yuen, K. Y. (2003) Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet*, 361, 1767-1772.

168. Peiris, J. S. M., Guan, Y., Yuen, K. Y. (2004) Severe acute respiratory syndrome. *Nature Medicine*, 10, S88-S97.
169. Pejawar, S. S., Parks, G. D., Alexander-Miller, M. A. (2005) Abortive versus productive viral infection of dendritic cells with a paramyxovirus results in differential upregulation of select costimulatory molecules. *J. Virol.*, 79, 7544-7557.
170. Phillips, J. J., Chua, M., Seo, S.-h., Weiss, S. R. (2001) Multiple regions of the murine coronavirus spike glycoprotein influence neurovirulence. *J. Neurovirol.*, 7, 421-431.
171. Pogrebnyak, N., Golovkin, M., Andrianov, V., Spitsin, S., Smirnov, Y., Egolf, R., Koprowski, H. (2005) Severe acute respiratory syndrome (SARS) S protein production in plants : Development of recombinant vaccines. *Proc. Natl. Acad. Sci. USA*, 102, 9062-9067.
172. Poon, L. L., Chu, D. K. W., Chan, K. H., Wong, O. K., Ellis, T. M., Leung, Y. H. C., Lau, S. K. P., Woo, P. C. Y., Suen, K. Y., Yuen, K. Y., Guan, Y., Peiris, J. S. M. (2005a) Identification of a novel coronavirus in bats. *J. Virol.*, 79, 2001-2009.
173. Poon, L. L. M., Leung, C. S. W., Chan, K. H., Yuen, K. Y., Guan, Y., Peiris, J. S. M. (2005b) Recurrent mutations associated with isolation and passage of SARS coronavirus in cells from non-human primates. *J. Med. Virol.*, 76, 435-440.
174. Popova, R., Zhang, X. (2002) The spike but not the hemagglutinin/esterase protein of bovine coronavirus is necessary and sufficient for viral infection. *Virology*, 294, 222-236.
175. Prentice, E., McAuliffe, J., Lu, X., Subbarao, K., Denison, M. R. (2004) Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. *J. Virol.*, 78, 9977-9986.
176. Racaniello, V. R., Baltimore, D. (1981) Cloned poliovirus complementary DNA is infectious in mammalian cells. *Science*, 214, 916-919.

177. Ramakrishna, C., Stohlman, S. A., Atkinson, R. D., Shlomchik, M. J., Bergmann, C. C. (2002) Mechanisms of central nervous system viral persistence: the critical role of antibody and B cells. *J. Immunol.*, 168, 1204-1211.
178. Rottier, P. J. M., Nakamura, K., Schellen, P., Volders, H., Haijema, B. J. (2005) Acquisition of macrophage tropism during the pathogenesis of feline infectious peritonitis is determined by mutations in the feline coronavirus spike protein. *J. Virol.*, 79, 14122-14130.
179. Rowe, C. L., Baker, S. C., Nathan, M. J., Fleming, J. O. (1997) Evolution of mouse hepatitis virus : detection and characterization of spike deletion variants during persistent infection. *J. Virol.*, 71, 2959-2969.
180. Ruiz-Jarabo, C. M., Arias, A., Baranowski, E., Escarmís, C., Domingo, E. (2000) Memory in viral quasispecies. *J. Virol.*, 74, 3543-3547.
181. Saeki, K., Ohtsuka, N., Taguchi, F. (1997) Identification of spike protein residues of murine coronavirus responsible for receptor-binding activity by use of soluble receptor-resistant mutants. *J. Virol.*, 71, 9024-9031.
182. Sainz, Jr., B., Mossel, E. C., Peters, C. J., Garry, R. F. (2004) Interferon-beta and interferon-gamma synergistically inhibit the replication of severe acute respiratory syndrome-associated coronavirus (SARS-CoV). *Virology*, 329, 11-17.
183. Sainz, Jr., B., Rausch, J. M., Gallagher, W. R., Garry, R. F., Wimley, W. C. (2005) Identification and characterization of the putative fusion peptide of the severe acute respiratory syndrome-associated coronavirus spike protein. *J. Virol.*, 79, 7195-7206.
184. Sallie, R. (2004) Replicative homeostasis: a mechanism of viral persistence. *Med. Hypotheses*, 63, 515-523.
185. Salmi, A., Ziola, B., Hovi, T., Reunanen, M. (1982) Antibodies to coronaviruses OC43 and 229E in multiple sclerosis patients. *Neurology*, 32, 292-295.

186. Sanchez, C. M., Gebauer, F., Sune, C., Mendez, A., Dopazo, J., Enjuanes, L. (1992) Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology*, 190, 92-105.
187. Sawicki, S. G., Sawicki, D. L. (1998) A new model for coronavirus transcription. *Adv. Exp. Med. Biol.*, 440, 215-219.
188. Schelle, B., Karl, N., Ludewig, B., Siddell, S. G., Thiel, V. (2005) Selective replication of coronavirus genomes that express nucleocapsid protein. *J. Virol.*, 79, 6620-6630.
189. Schickli, J. H., Thackray, L. B., Sawicki, S. G., Holmes, K. V. (2004) The N-terminal region of the murine coronavirus spike glycoprotein is associated with the extended host range of viruses from persistently infected murine cells. *J. Virol.*, 78, 9073-9083.
190. Schickli, J. H., Zelus, B. D., Wentworth, D. E., Sawicki, S. G., Holmes, K. V. (1997) The murine coronavirus mouse hepatitis virus strain A59 from persistently infected murine cells exhibits an extended host range. *J. Virol.*, 71, 9499-9507.
191. Schultze, B., Gross, H. J., Brossmer, R., Herrler, G. (1991) The S protein of bovine coronavirus is a hemagglutinin recognizing 9-O-acetylated sialic acid as a receptor determinant. *J. Virol.*, 65, 6232-6237.
192. Schwegmann-Webels, C., Zimmer, G., Laude, H., Enjuanes, L., Herrler, G. (2002) Binding of transmissible gastroenteritis coronavirus to cell surface sialoglycoproteins. *J. Virol.*, 76, 6037-6043.
193. Seybert, A., Hegyi, A., Siddell, S. G., Ziebuhr, J. (2000) The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity. *RNA*, 6, 1056-1068.
194. Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y., Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA*, 89, 8794-8797.

195. Simmons, G., Reeves, J. D., Rennekamp, A. J., Amberg, S. M., Piefer, A. J., Bates, P. (2004) Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proc. Natl. Acad. Sci. USA*, 101, 4240-4245.
196. Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. M., Guan, Y., Rozanov, M., Spaan, W. J. M., Gorbalenya, A. E. (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.*, 331, 991-1004.
197. Sola, I., Alonso, S., Zúñiga S., Balasch, M., Plana-Durán, J., Enjuanes, L. (2003) Engineering the transmissible gastroenteritis virus genome as an expression vector inducing lactogenic immunity. *J. Virology*, 77, 4357-4369.
198. Sola, I., Moreno, J. L., Zúñiga, S., Alonso, S., Enjuanes, L. (2005) Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J. Virol.*, 79, 2506-2516.
199. Song, H.-D., Tu, C.-C., Zhang, G.-W., Wang, S.-Y., Zheng, K., Lei, L.-C., Chen, Q.-X., Gao, Y.-W., Zhou, H.-Q., Xiang, H., Zheng, H.-J., Chern, S.-W. W., Cheng, F., Pan, C.-M., Xuan, H., Chen, S.-J., Luo, H.-M., Zhou, D.-H., Liu, Y.-F., He, J.-F., Qin, P.-Z., Li, L.-H., Ren, Y.-Q., Liang, W.-J., Yu, Y.-D., Anderson, L., Wang, M., Xu, R.-H., Wu, X.-W., Zheng, H.-Y., Chen, J.-D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L.-Y., Li, H., Chen, F., Di, B., He, L.-J., Lin, J.-Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X.-J., Spiga, O., Guo, Z.-M., Pan, H.-Y., He, W.-Z., Manuguerra, J.-C., Fontanet, A., Danchin, A., Niccolai, N., Li, Y.-X., Wu, C.-I., Zhao, G.-P. (2005) Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci. USA*, 102, 2430-2435.
200. Sorensen, O., Perry, D., Dales, S. (1980) In vivo and in vitro models of demyelinating diseases. III. JHM virus infection of rats. *Arch. Neurol.*, 37, 478-484.

201. Spagnolo, J. F., Hogue, B. G. (2001) Requirement of the poly(A) tail in coronavirus genome replication. *Adv. Exp. Med. Biol.*, 494, 467-474.
202. Sperry, S. M., Kazi, L., Graham, R. L., Baric, R. S., Weiss, S. R., Denison, M. R. (2005) Single-amino-acid substitutions in open reading frame (ORF) 1b-nsp14 and ORF 2a proteins of the coronavirus mouse hepatitis virus are attenuating in mice. *J. Virol.*, 79, 3391-3400.
203. Spiga, O., Bernini, A., Ciutti, A., Chiellini, S., Menciassi, N., Finetti, F., Causarone, V., Anselmi, F., Prischi, F., Niccolai, N. (2003) Molecular modelling of S1 and S2 subunits of SARS coronavirus spike glycoprotein. *Biochem. Biophys. Res. Commun.*, 310, 78-83.
204. Steinhauer, D. A., Domingo, E., Holland, J. J. (1992) Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene*, 122, 281-288.
205. St-Jean, J. R., Desforges, M., Almazán, F., Jacomy, H., Enjuanes, L., Talbot, P. J. (2006) Recovery of a neurovirulent human coronavirus OC43 from an infectious cDNA clone. *J. Virol.*, 80, 3670-3674.
206. St-Jean, J. R., Jacomy, H., Desforges, M., Vabret, A., Freymuth, F., Talbot, P. J. (2004) Human respiratory coronavirus OC43 : Genetic stability and neuroinvasion. *J. Virol.*, 78, 8824-8834.
207. Stohlman, S. A., Ramakrishna, C., Tschen, S.-I., Hinton, D. R., Bergmann, C. C. (2002) The art of survival during viral persistence. *J. Neurovirol.*, 8, 53-58.
208. Sulea, T., Lindner, H. A., Purisima, E. O., Ménard, R. (2005) Letter to the Editor : Deubiquitination, a new function of the severe acute respiratory syndrome coronavirus papain-like protease? *J. Virol.*, 79, 4550-4551.
209. Supekar, V. M., Bruckmann, C., Ingallinella, P., Bianchi, E., Pessi, A., Carfi, A. (2004) Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein. *Proc. Natl. Acad. Sci. USA*, 101, 17958-17963.

210. Taguchi, F., Shimazaki, Y. K. (2000) Functional analysis of an epitope in the S2 subunit of the murine coronavirus spike protein : involvement in fusion activity. *J. Gen. Virol.*, 81, 2867-2871.
211. Taguchi, F., Kubo, H., Takahashi, H., Suzuki, H. (1995) Localization of neurovirulence determinant for rats on the S1 subunit of murine coronavirus JHMV. *Virology*, 208, 67-74.
212. Talbot, P. (1995) Implication des virus dans la sclérose en plaques. *Médecine/Sciences*, 11, 837-843.
213. Talbot, P. J. (1997) Virus-induced autoimmunity in multiple sclerosis : the coronavirus paradigm. *Adv. Clin. Neurosci.*, 7, 215-233.
214. Talbot, P. J., Paquette, J.-S., Ciurli, C., Antel, J. P., Ouellet, F. (1996) Myelin basic protein and human coronavirus 229E cross-reactive T cells in multiple sclerosis. *Ann. Neurol.*, 39, 233-240.
215. Tan, K., Zelus, B. D., Meijers, R., Liu, J.-h., Bergelson, J. M., Duke, N., Zhang, R., Joachimiak, A., Holmes, K. V., Wang, J.-h. (2002) Crystal structure of murine sCEACAM1a[1,4]: a coronavirus receptor in the CEA family. *EMBO J.*, 21, 2076-2086.
216. Tan, Y.-J., Lim, S. G., Hong, W. (2005) Characterization of viral proteins encoded by the SARS-coronavirus genome. *Antiviral Res.*, 65, 69-78.
217. Thackray, L. B., Holmes, K. V. (2004) Amino acid substitutions and an insertion in the spike glycoprotein extend the host range of the murine coronavirus MHV-A59. *Virology*, 324, 510-524.
218. Thiel, V., Herold, J., Schelle, B., Siddell, S. G. (2001) Infectious RNA transcribed *in vitro* from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.*, 82, 1273-1281.
219. Thiel, V., Ivanov, K. A., Putics, Á., Hertzog, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E., Ziebuhr, J. (2003a)

- Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.*, 84, 2305-2315.
220. Thiel, V., Karl, N., Schelle, B., Disterer, P., Klagge, I., Siddell, S. G. (2003b) Multigene RNA vector based on coronavirus transcription. *J. Virology*, 77, 9790-9798.
221. Thiel, V., Rashtchian, A., Herold, J., Schuster, D. M., Guan, N., Siddell, S. G. (1997) Effective amplification of 20-kb DNA by reverse transcription PCR. *Anal. Biochem.*, 252, 62-70.
222. Thorp, E. B., Gallagher, T. M. (2004) Requirements for CEACAMs and cholesterol during murine coronavirus cell entry. *J. Virol.*, 78, 2682-2692.
223. Toney, J. H., Navas-Martin, S., Weiss, S. R., Koeller, A. (2004) Sabadinine : a potential non-peptide anti-severe acute-respiratory-syndrome agent identified using structure-aided design. *J. Med. Chem.*, 47, 1079-1080.
224. Tsai, C.-W., Chang, S. C., Chang, M.-F. (1999) A 12-amino acid stretch in the hypervariable region of the spike protein S1 subunit is critical for cell fusion activity of mouse hepatitis virus. *J. Biol. Chem.*, 274, 26085-26090.
225. Tsai, J. C., de Groot, L., Pinon, J. D., Iacono, K. T., Phillips, J. J., Seo, S.-h., Lavi, E., Weiss, S. R. (2003) Amino acid substitutions within the heptad repeat domain 1 of murine coronavirus spike protein restrict viral antigen spread in the central nervous system. *Virology*, 312, 369-380.
226. Tsai, L. K., Hsieh, S. T., Chang, Y. C. (2005) Neurological manifestations in severe acute respiratory syndrome. *Acta. Neurol. Taiwan*, 14, 113-119.
227. Tsunoda, I., Fujinami, R. S. (2002) Inside-out versus outside-in models for virus induced demyelination: axonal damage triggering demyelination. *Springer Semin. Immunopathol.*, 24, 105-125.

228. Turner, B. C., Hemmila, E. M., Beauchemin, N., Holmes, K. V. (2004) Receptor-dependant coronavirus infection of dendritic cells. *J. Virol.*, 78, 5486-5490.
229. Vabret, A., Mourez, T., Gouarin, S., Petitjean, J., Freymuth, F. (2003) An outbreak of coronavirus OC43 respiratory infection in Normandy, France. *Clin. Infect. Dis.*, 36, 985-989.
230. Van der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J. M., Wolthers, K. C., Wertheim-van Dillen, P. M. E., Kaandorp, J., Spaargaren, J., Berkhout, B. (2004) Identification of a new human coronavirus. *Nature Medecine*, 10, 368-373.
231. Vanderlugt, C. L., Neville, K. L., Nikcevich, K. M., Eagar, T. N., Bluestone, J. A., Miller, S. D. (2000) Pathologic role and temporal appearance of newly emerging autoepitopes in relapsing experimental autoimmune encephalomyelitis. *J. Immunol.*, 164, 670-678.
232. Van Maele, B., Debyser, Z. (2005) HIV-1 integration: an interplay between HIV-1 integrase, cellular and viral proteins. *AIDS Rev.*, 7, 26-43.
233. van Marle, G., Power, C. (2005) Human immunodeficiency virus type 1 genetic diversity in the nervous system: Evolutionary epiphenomena or disease determinant? *J. Neurovirol.*, 11, 107-128.
234. Wang, Z., Yuan, Z., Matsumoto, M., Hengge, U. R., Chang, Y.-F. (2005) Immune responses with DNA vaccines encoded different gene fragments of severe acute respiratory syndrome coronavirus in BALB/c mice. *Biochem. Biophys. Res. Commun.*, 327, 130-135.
235. Watanabe, H., Numata, K., Ito, T., Takagi, K., Matsukawa, A. (2004) Innate immune response in Th1- and Th2-dominant mouse strains. *Shock*, 22, 460-466.
236. Watanabe, R., Wege, H., Ter Meulen, V. (1983) Adoptive transfer of EAE-like lesions from rats with coronavirus-induced demyelinating encephalomyelitis. *Nature*, 305, 150-153.

237. Wege, H. (1995) Immunopathological aspects of coronavirus infections. *Springer Semin. Immunopathol.*, 17, 133-148.
238. Whitacre, C. C., Reingold, S. C., O'Looney, P. A. (1999) A gender gap in autoimmunity. *Science*, 283, 1277-1278.
239. Williams, H., Crawford, D. H. (2005) Epstein-barr virus: impact of scientific advance on clinical practice. *Blood*, Epub ahead of print.
240. Wilson, L., Mckinlay, C., Gage, P., Ewart, G. (2004) SARS coronavirus E protein forms cation-selective ion channels. *Virology*, 330, 322-331.
241. Wong, S. K., Li, W., Moore, M. J., Choe, H., Farzan, M. (2004) A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J. Biol. Chem.*, 279, 3197-3201.
242. Woo, P. C. Y., Lau, S. K. P., Chu, C.-m., Chan, K.-h., Tsoi, H.-w., Huang, Y., Wong, B. H. L., Poon, R. W. S., Cai, J. J., Luk, W.-k., Poon, L. L. M., Wong, S. S. Y., Guan, Y., Malik Peiris, Yuen, K.-y. (2005) Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.*, 79, 884-895.
243. Wu, C.-J., Huang, H.-W., Liu, C.-Y., Hong, C.-F., Chan, Y.-L. (2004) Inhibition of SARS-CoV replication by siRNA. *Antiviral Res.*, 65, 45-48.
244. Wu, H.-Y., Guy, J. S., Yoo, D., Vlasak, R., Urbach, E., Brian, D. A. (2003) Common RNA replication signals exist among group 2 coronaviruses: evidence for in vivo recombination between animal and human coronavirus molecules. *Virology*, 315, 174-183.
245. Wucherpfennig, K. W. (2004) T cell receptor crossreactivity as a general property of T cell recognition. *Mol. Immunol.*, 40, 1009-1017.
246. Würdinger, T., Verheije, M. H., Broen, K., Bosch, B. J., Haijema, B. J., de Haan, C. A. M., van Beusechem, V. W., Gerritsen, W. R., Rottier, P. J. M. (2005a) Soluble receptor-

- mediated targeting of mouse hepatitis coronavirus to the human epidermal growth factor receptor. *J. Virol.*, 79, 15314-15322.
247. Würdinger, T., Verheije, M. H., Raaben, M., Bosch, B. J., de Haan, C. A. M., van Beusechem, V. W., Rottier, P. J. M., Gerritsen, W. R. (2005b) Targeting non-human coronaviruses to human cancer cells using a bispecific single-chain antibody. *Gene Ther.*, 12, 1394-1404.
248. Wurzer, W. J., Obojes, K., Vlasak, R. (2002) The sialate-4-O-acetyltransferases of coronaviruses related to mouse hepatitis virus: a proposal to reorganize group 2 Coronaviridae. *J. Gen. Virol.*, 83, 395-402.
249. Xiao, X., Feng, Y., Chakraborti, S., Dimitrov, D. S. (2004) Oligomerization of the SARS-CoV S glycoprotein: dimerization of the N-terminus and trimerization of the ectodomain. *Biochem. Biophys. Res. Commun.*, 322, 93-99.
250. Xu, J., Zhong, S., Liu, J., Li, L., Li, Y., Wu, X., Li, Z., Deng, P., Zhang, J., Zhong, N., Ding, Y., Jiang, Y. (2005) Detection of severe acute respiratory syndrome coronavirus in the brain: potential role of the chemokine mig in pathogenesis. *Clin. Infect. Dis.*, 41, 1089-1096.
251. Xu, X., Liu, Y., Weiss, S., Arnold, E., Sarafianos, S. G., Ding, J. (2003) Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res.*, 31, 7117-7130.
252. Xu, Y., Bai, Z., Qin, L., Li, X., Gao, G., Rao, Z. (2004a) Crystallisation and preliminary crystallographic analysis of the fusion core of the spike protein of the murine coronavirus mouse hepatitis virus (MHV). *Acta Cryst.*, D60, 2013-2015.
253. Xu, Y., Su, N., Qin, L., Bai, Z., Gao, G. F., Rao, Z. (2004b) Crystallization and preliminary crystallographic analysis of the heptad-repeat complex of SARS coronavirus spike protein. *Acta Crystallogr. D Biol. Crystallogr.*, 60, 2377-2379.

254. Xu, Y., Zhu, J., Liu, Y., Lou, Z., Yuan, F., Liu, Y., Cole, D. K., Ni, L., Su, N., Qin, L., Li, X., Bai, Z., Bell, J. I., Pang, H., Tien, P., Gao, G. F., Rao, Z. (2004c) Characterization of the heptad repeat regions, HR1 and HR2, and design of a fusion core structure model of the spike protein from severe acute respiratory syndrome (SARS) coronavirus. *Biochemistry*, 43, 14064-14071.
255. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R., Rao, Z. (2003) The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. USA*, 100, 13190-13195.
256. Yi, L., Li, Z., Yuan, K., Qu, X., Chen, J., Wang, G., Zhang, H., Luo, H., Zhu, L., Jiang, P., Chen, L., Shen, Y., Luo, M., Zuo, G., Hu, J., Duan, D., Nie, Y., Shi, X., Wang, W., Han, Y., Li, T., Liu, Y., Ding, M., Deng, H., Xu, X. (2004) Small molecules blocking the entry of severe acute respiratory syndrome coronavirus into host cells. *J. Virol.*, 78, 11334-11339.
257. Yokomori, K., Banner, L. R., Lai, M. M. (1991) Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. *Virology*, 183, 647-657.
258. Yount, B., Curtis, K. M., Baric, R. S. (2000) Strategy for systematic assembly of large RNA and DNA genomes : transmissible gastroenteritis virus model. *J. Virol.*, 74, 10600-10611.
259. Yount, B., Curtis, K. M., Fritz, E. A., Hensley, L. E., Jahrling, P. B., Prentice, E., Denison, M. R., Geisbert, T. W., Baric, R. S. (2003) Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. USA*, 100, 12995-13000.
260. Yount, B., Denison, M. R., Weiss, S. R., Baric, R. S. (2002) Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J. Virol.*, 76, 11065-11078.

261. Yu, G.-Y., Lai, M. M. C. (2005) The ubiquitin-proteasome system facilitates the transfer of murine coronavirus from endosome to cytoplasm during virus entry. *J. Virol.*, 79, 644-648.
262. Zhang, Y., Li, T., Fu, L., Yu, C., Li, Y., Xu, X., Wang, Y., Ning, H., Zhang, S., Chen, W., Babiuk, L. A., Chang, Z. (2004) Silencing SARS-CoV spike protein expression in cultured cells by RNA interference. *FEBS Lett.*, 560, 141-146.
263. Zhao, P., Qin, Z.-L., Ke, J.-S., Lu, Y., Liu, M., Pan, W., Zhao, L.-J., Cao, J., Qi, Z.-T. (2005) Small interfering RNA inhibits SARS-CoV nucleocapsid gene expression in cultured cells and mouse muscles. *FEBS Lett.*, 579, 2404-2410.
264. Zhou, D., Hemmer, B. (2004) Specificity and degeneracy: T cell recognition in CNS autoimmunity. *Mol. Immunol.*, 40, 1057-1061.
265. Zhu, M. S. (2004) SARS immunity and vaccination. *Cell. Mol. Immunol.*, 1, 193-198.
266. Ziebuhr, J. (2004) Molecular biology of severe acute respiratory syndrome coronavirus. *Curr. Opin. Microbiol.*, 7, 412-419.
267. Ziebuhr, J., Snijder, E. J., Gorbalenya, A. E. (2000) Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J. Gen. Virol.*, 81, 853-879.
268. Zuniga, E. I., Hahm, B., Edelmann, K. H., Oldstone, M. B. A. (2005) Immunosuppressive viruses and dendritic cells: a multifront war. *ASM News*, 71, 285-290.
269. Zúñiga, S., Sola, I., Alonso, S., Enjuanes, L. (2004) Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J. Virol.*, 78, 980-994

LISTE DES CONTRIBUTIONS SCIENTIFIQUES

Publications scientifiques

JULIEN R. ST-JEAN, HÉLÈNE JACOMY, and PIERRE J. TALBOT (2006) «Human coronavirus OC43 isolated from persistently infected neural cells displays adaptation to its environment in cell culture.» *Virology* (soumis).

JULIEN R. ST-JEAN, MARC DESFORGES, and PIERRE J. TALBOT (2006) «Genetic evolution of human coronavirus OC43 in neural cell culture.» *Adv. Exp. Med. Biol.* (sous presse).

JULIEN R. ST-JEAN, MARC DESFORGES, FERNANDO ALMAZÁN, HÉLÈNE JACOMY, LUIS ENJUANES, and PIERRE J. TALBOT (2006) «Recovery of a neurovirulent human OC43 coronavirus from an infectious cDNA clone.» *J. Virol.*, 80 : 3670-3674.

JULIEN R. ST-JEAN, HÉLÈNE JACOMY, MARC DESFORGES, ASTRID VABRET, FRANÇOIS FREYMUTH and PIERRE J. TALBOT (2005) «Genetic variability of human respiratory coronavirus OC43 : Author's reply to Letter to the Editor.» *J. Virol.*, 79 : 3224-3225.

JULIEN R. ST-JEAN, HÉLÈNE JACOMY, MARC DESFORGES, ASTRID VABRET, FRANÇOIS FREYMUTH and PIERRE J. TALBOT (2004) «Human respiratory coronavirus OC43: genetic stability and neuroinvasion.» *J. Virol.*, 78 : 8824-8834.

Présentations scientifiques

JULIEN R. ST-JEAN, HÉLÈNE JACOMY, MARC DESFORGES and PIERRE J. TALBOT. «Adaptation of human coronavirus OC43 to growth in neural cells.» 25^{ème} réunion annuelle de l'American Society for Virology, Madison University, Madison, Wisconsin, États-Unis, 15-19 juillet 2006 (présentation orale)

JULIEN R. ST-JEAN, MARC DESFORGES, HÉLÈNE JACOMY and PIERRE J. TALBOT. «Genetic evolution of human coronavirus OC43 in neural cell culture.» 10^{ème} Symposium International des Nidovirus, Colorado Springs, États-Unis, juin 2005 (présentation par affiche)

JULIEN R. ST-JEAN, HÉLÈNE JACOMY, MARC DESFORGES, ASTRID VABRET, FRANÇOIS FREYMUTH and PIERRE J. TALBOT. «Human coronavirus OC43 genome sequencing and analysis reveals genetic stability.» 23^{ème} réunion annuelle de l'American Society for Virology, McGill University, Montréal, Québec, Canada, 10-14 juillet 2004 (présentation orale)

JULIEN R. ST-JEAN et PIERRE J. TALBOT. Caractérisation moléculaire du coronavirus humain OC43: séquençage du génome et assemblage d'un clone infectieux. 3^{ème} Congrès INRS-Institut Armand-Frappier, Manoir du Lac Delage, Stoneham, Québec, Canada, novembre 2003 (présentation par affiche)

JULIEN R. ST-JEAN and PIERRE J. TALBOT. «Molecular adaptation of human coronavirus during persistence in the brain.» 9^{ème} Symposium International des Nidovirus, Egmond aan Zee, Hollande, mai 2003 (présentation par affiche)

JULIEN R. ST-JEAN et PIERRE J. TALBOT. Caractérisation de la persistance du coronavirus dans le cerveau humain. 2^{ème} Congrès INRS-Institut Armand-Frappier, Hôtel du Mont Gabriel, Ste-Adèle, Québec, Canada, novembre 2001 (présentation par affiche)

ANNEXES

Human Respiratory Coronavirus OC43: Genetic Stability and Neuroinvasion

Julien R. St-Jean,¹ Hélène Jacomy,¹ Marc Desforges,¹ Astrid Vabret,²
François Freymuth,² and Pierre J. Talbot^{1*}

Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier, Laval, H7V 1B7 Quebec, Canada,¹
and Laboratoire de Virologie Humaine et Moléculaire, Centre Hospitalier Régional
et Universitaire de Caen, 14033 Caen, France²

Received 28 January 2004/Accepted 7 April 2004

The complete genome sequences of the human coronavirus OC43 (HCoV-OC43) laboratory strain from the American Type Culture Collection (ATCC), and a HCoV-OC43 clinical isolate, designated Paris, were obtained. Both genomes are 30,713 nucleotides long, excluding the poly(A) tail, and only differ by 6 nucleotides. These six mutations are scattered throughout the genome and give rise to only two amino acid substitutions: one in the spike protein gene (I958F) and the other in the nucleocapsid protein gene (V81A). Furthermore, the two variants were shown to reach the central nervous system (CNS) after intranasal inoculation in BALB/c mice, demonstrating neuroinvasive properties. Even though the ATCC strain could penetrate the CNS more effectively than the Paris 2001 isolate, these results suggest that intrinsic neuroinvasive properties already existed for the HCoV-OC43 ATCC human respiratory isolate from the 1960s before it was propagated in newborn mouse brains. It also demonstrates that the molecular structure of HCoV-OC43 is very stable in the environment (the two variants were isolated ca. 40 years apart) despite virus shedding and chances of persistence in the host. The genomes of the two HCoV-OC43 variants display 71, 53.1, and 51.2% identity with those of mouse hepatitis virus A59, severe acute respiratory syndrome human coronavirus Tor2 strain (SARS-HCoV Tor2), and human coronavirus 229E (HCoV-229E), respectively. HCoV-OC43 also possesses well-conserved motifs with regard to the genome sequence of the SARS-HCoV Tor2, especially in open reading frame 1b. These results suggest that HCoV-OC43 and SARS-HCoV may share several important functional properties and that HCoV-OC43 may be used as a model to study the biology of SARS-HCoV without the need for level three biological facilities.

Human coronaviruses (HCoVs), members of the *Coronaviridae* family, are ubiquitous in the environment and are responsible for up to one-third of common colds (41). In the past few years, we have provided experimental evidence that this virus possesses neurotropic and neuroinvasive properties: it persists in neural cell cultures (7, 8) and human brains (9). Of the two HCoV serotypes available, HCoV-OC43 was selected for further characterization of persistence in the nervous system because of a more efficient infection of primary neural cell cultures (11), as well as a trend toward association with neurological disease (9).

Coronaviruses are enveloped viruses that possess a positive-strand RNA genome of up to 31 kb, which represents the largest known genome among all RNA viruses (35). This genome comprises several genes encoding several structural and nonstructural proteins. Among these proteins, the S protein is biologically very important because it could be implicated in determination of tropism (3) and its modulation (50). Indeed, the S protein could be associated with the capacity of the virus to reach the central nervous system (CNS) and possibly trigger neurological disorders (9, 22). It could also be responsible for conferring the strong degree of host species specificity observed with coronaviruses (28).

Only the 3' one-third of the HCoV-OC43 genome has been sequenced over the years. Therefore, until now, the complete sequence of the open reading frame 1a (ORF1a) and ORF1b, known as the replicase gene, was still undetermined. This gene is essential for coronavirus survival because it contains several motifs, which could be involved in various important viral functions such as transcription, replication, and pathogenesis (66). The products encoded by these two ORFs are polyprotein precursors, which are processed by two or three different proteinases encoded by ORF1a. These proteinases could include two papain-like proteases (PLP1 and PLP2) and a poliovirus 3C-like protease (3CLpro), which presents the most important cleavage activity. The 3CLpro essential function is reflected by its capacity to cleave at many sites in the replicase polyproteins and to release the key replicative functions, such as the RNA-dependent RNA polymerase (RdRp) and the RNA helicase (67).

The HCoV-OC43 strain belongs to the second genetic group, just as SARS-HCoV apparently does (51). The latter is responsible for the severe acute respiratory syndrome (SARS), which is a life-threatening form of pneumonia (46). Since the outbreak of SARS in the fall of 2002 (60), a lot of work has been done to sequence the entire genome of the virus (34) and to understand the mechanisms underlying virus pathogenesis. As presented here, the whole genome of HCoV-OC43 has now been sequenced and, since this human strain is the most related to SARS-HCoV, it could be used as a model for the study of the SARS-HCoV without the drawbacks of level three bio-

* Corresponding author. Mailing address: Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier, 531 Boulevard des Prairies, Laval, H7V 1B7 Quebec, Canada. Phone: (450) 686-5515. Fax: (450) 686-5566. E-mail: pierre.talbot@inrs-iaf.quebec.ca.

logical confinement. Comparisons with the SARS-HCoV nucleotide and amino acid sequences (34) revealed that the two viruses share extensive homology in some important motifs involved in viral replication and pathogenesis. Indeed, the most significant homology between the genomes of the HCoV-OC43 strain and the one of the SARS-HCoV Tor2 isolate is found in the ORF1b region, which comprises the RdRp and helicase motifs (16). The 3CLpro motif of HCoV-OC43 also displays an important level of identity with the one of SARS-HCoV. This finding is noteworthy since SARS-HCoV 3CLpro thus far represents the most promising target for SARS therapy (58).

We report here the complete genome sequences of the HCoV-OC43 strain from the American Type Culture Collection (ATCC), as well as an HCoV-OC43 respiratory clinical isolate, designated HCoV-OC43 Paris. Both genomes are 30,713 nucleotides (nt) long, share the same genomic organization, and only differ by 6 nt. Differences found in the genome of the HCoV-OC43 Paris isolate, compared to the genome of HCoV-OC43 ATCC, give rise to only two amino acid substitutions, which are located in the S (I958F) and the N (V81A) protein genes. After intranasal inoculation in BALB/c mice, the HCoV-OC43 ATCC strain, as well as the Paris isolate, reached the CNS, where they replicated and disseminated, although mice were apparently more easily infected with the ATCC strain than with the Paris isolate. These results suggest that both viruses possess the ability to reach and infect neural cells *in vivo*. The fact that a natural OC43 isolate has an intrinsic capacity to invade and replicate within the mouse CNS also suggests that the HCoV-OC43 ATCC strain has not acquired its neuroinvasive properties after propagation in newborn mouse brains. Bioinformatics analyses were also performed on the HCoV-OC43 genome. These analysis showed that this virus strain is closely related to mouse hepatitis virus A59 (MHV-A59) and that it displays significant identity levels with important functional domains of the SARS-HCoV. These data provide evidence that HCoV-OC43 could be used as a model for the study of other group 2 coronaviruses, including SARS-HCoV, and that it will facilitate understanding of the biology of this emerging viral strain.

MATERIALS AND METHODS

Viruses and cell lines. The ATCC HCoV-OC43 strain (ATCC number VR-759), isolated in the 1960s, and the Paris clinical respiratory isolate, isolated in March 2001, were grown on a HRT-18 cell line (human adenocarcinoma rectal) as described previously (37). The clinical sample (HCoV-OC43 Paris) was isolated from the respiratory tract of a 68-year-old immunocompromised male who was not related whatsoever to laboratory work and was not in contact with any laboratory workers who had manipulated the HCoV-OC43 ATCC virus. A reverse transcription-PCR (RT-PCR) was performed to specifically detect the presence of the HCoV-OC43 RNA, and an aliquot of the clinical sample was then used to infect the HRT-18 cell line. The HCoV-OC43 ATCC strain and the Paris isolate were never cultured at the same time, and stringent laboratory precautions were used in order to eliminate possible cross-contamination.

Acute infections of cells. Cells were infected at a multiplicity of infection of 0.02 and 0.2 for the ATCC strain and Paris isolate, respectively. The fifth passage of the ATCC strain and the eighth passage of the Paris isolate were used to perform the infections. Cell lines at 70% confluence were infected with the appropriate virus stock in the presence of TPCK (tolylsulfonyl phenylalanyl chloromethyl ketone)-treated trypsin (10 U/ml; Sigma-Aldrich Canada, Ltd.) and 1% (vol/vol) heat-inactivated fetal calf serum and then incubated at 33°C for 4 days in a 5% (vol/vol) CO₂ humid atmosphere.

Mice and inoculations. In order to determine the susceptibility of mice to an infection by HCoV-OC43 ATCC and HCoV-OC43 Paris variants, MHV-seronegative 14-day-postnatal BALB/c mice (Charles River Laboratories, St-Constant, Quebec, Canada) were inoculated intranasally with 5 μ l of a virus stock solution containing 10⁶ 50% tissue culture infective dose(s) (TCID₅₀)/ml. Five mice, inoculated with HCoV-OC43 ATCC or HCoV-OC43 Paris variants, were sacrificed every 2 days postinfection (dpi) and processed for detection of infectious virus particles. Every 2 days, two mice infected by HCoV-OC43 ATCC were processed for immunohistochemical detection of viral antigens.

Immunohistochemistry. Mice were perfused by intraventricular injection of 4% (vol/vol) paraformaldehyde, under deep ketamine-xylazine anesthesia, as previously described (22). Brains were dissected and sectioned at a thickness of 40 μ m with a Lancer Vibratome. Sections were collected in 0.05 M Tris-buffered saline and then incubated for 2 h at 37°C in a 1/1,000 dilution of an ascites fluid from mouse MAb 1-10C.3, directed against the spike protein of HCoV-OC43 (7). Sections were then rinsed and processed with a Vectastain ABC kit (Vector Laboratories, Burlingame, Calif.). Labeling was revealed with 0.03% (wt/vol) DAB solution (Sigma) and 0.01% (vol/vol) H₂O₂, which yielded a dark brown product.

Infectious virus assays. Brain and lung were dissected, homogenized in 10% (wt/vol) sterile phosphate-buffered saline (PBS), and centrifuged at 4°C for 20 min at 1,000 \times g, and then supernatants were immediately frozen at -80°C and stored until assayed. The extracts were processed for the presence and quantification of infectious virus by an indirect immunoperoxidase assay, as previously described (22). Briefly, HCoV-OC43-susceptible HRT-18 cells were inoculated with serial logarithmic dilutions of each tissue sample. After 4 days of incubation at 33°C in a 5% (vol/vol) CO₂ humid atmosphere, the cells were washed in PBS and fixed with 0.3% (vol/vol) hydrogen peroxide (H₂O₂) in methanol. After being washed with PBS, they were incubated for 2 h at 37°C in a 1/1,000 dilution of an ascites fluid from mouse MAb 1-10C.3. Afterward, cells were washed in PBS, and horseradish peroxidase-goat anti-mouse immunoglobulins (Dako; Diagnostics Canada, Inc., Mississauga, Ontario, Canada) were added, followed by incubation for 2 h at 37°C. Antibody complexes were detected by incubation in DAB (Sigma) with 0.01% (vol/vol) H₂O₂.

RNA extraction, RT, and PCR. After infection, the cells were washed with PBS, and the total RNA was extracted from the cells by using the GenElute Mammalian Total RNA miniprep kit (Sigma-Aldrich) as recommended by the manufacturer. The RNA was then quantified, and 3 μ g was directly used for RT with Moloney murine leukemia virus reverse transcriptase (Invitrogen). For each RT, 500 ng of oligo(dT) primer and 0.5 mM deoxynucleoside triphosphates (Amersham Biosciences) were used, and the reactions lasted between 50 and 60 min at 37°C. Then, 2 μ l of the RT cDNA was then used to perform the PCR amplifications. The Expand High-Fidelity *Taq* polymerase (Roche) was used to amplify the HCoV-OC43 genome in six segments, in combination with primers listed in Table 1. All amplifications were performed by using the Cetus DNA thermal cycler (Perkin-Elmer/Applied Biosystems), and an appropriate annealing temperature was used for each specific reaction. Except for the PCR JUB3-12, which required a higher annealing temperature of 65°C, all other annealing temperature used corresponded to the melting temperature of the primers. For each PCR amplification, at least six reactions were performed, pooled together, migrated on a 0.8% (wt/vol) agarose gel (SeaKem), and gel extracted by using the Qiagex 11 gel extraction kit (Qiagen) prior to sequencing.

RACE and cloning. Rapid amplification of cDNA ends (RACE), cloning, and sequencing were performed for both 5' and 3' ends of HCoV-OC43 ATCC strain and the HCoV-OC43 Paris isolate. Primers from the kit used for the RACE are listed in Table 1. An RT reaction of the 5' end was performed by using the GeneRacer kit (Invitrogen) as recommended by the manufacturer, whereas RT of the 3' end was performed only by using the GeneRacer oligo(dT) primer provided in the kit (Table 1). In order to amplify both ends, primers from the kit were used in combination with primers specific for the HCoV-OC43 genome. Therefore, the GeneRacer 5' nested primer was used with JUB2 primer, and GeneRacer 3' nested primer was combined with JUMO1 primer for the ATCC strain and JUO8 primer for the Paris isolate. Amplicons of the 5' ends of both viruses and of the 3' end of the Paris isolate were cloned by using the Zero Blunt TOPO PCR cloning kit for sequencing (Invitrogen), whereas amplicons of the 3' end of the ATCC strain were cloned by using the TOPO XL PCR cloning kit. The RACE 5' clones were sequenced by using M13 universal forward and reverse primers and RACEJUB1 and RACEJUB2 primers, and RACE 3' clones were sequenced by using M13 universal forward and reverse primers and JUO7 primer.

Sequencing. Sequencing reactions were performed by Bio S&T (Montreal, Quebec, Canada) by using the dideoxy method (Sanger) and specific primers, which are listed in Table 2. As described above, PCR products were directly

TABLE 1. Primers used for amplification of the HCoV-OC43 genome

Primer combination and (nt location)	Target region or sequence	Amplicon length (bp)
JUB3–JUB12 (1–20 and 6071–6091)	Leader, 5'UTR, and ORF1a	6,091
JUB5–JUB6 (5319–5339 and 11111–11131)	ORF1a	5,813
JUB7–JUB8 (10901–10921 and 16525–16545)	ORF1a and ORF1b	5,645
JUB9–JUB10 (16309–16329 and 21544–21564)	ORF1b and ns2	5,256
JUNSO1–JUSO2 (21330–21350 and 27754–27774)	ORF1b, ns2, HE, and S genes	6,445
JUMO1–GeneRacer, 3' nested (27649–27669 and 30742–30764)	S; ns12.9; E, M, and N genes; and 3'UTR	3,116 ^a
GeneRacer, oligo(dT)	5'-GCTGTCAACGATACGCTACGTAACGGCATGACAGTG(T) ₁₈ -3'	
GeneRacer, 5' nested	5'-GGACTGACATGGACTGAAGGAGTA-3'	
GeneRacer, 3' nested	5'-CGCTACGTAACGGCATGACAGTG-3'	

^a This value assumes a poly(A) tail of 28 bp.

sequenced for both genomes, and both strands were sequenced in each case, including RACE clones. For each genome, at least two RACE 5' and 3' clones were sequenced for both isolates. Sequences obtained from chromatograms were aligned by using the basic local alignment search tool (BLAST; bl2seq) from the National Center for Biotechnology Information and were analyzed by using the Chromas 2 software.

Bioinformatics analyses. Bioinformatics analyses were performed by Sequence Bioinformatics (Montreal, Quebec, Canada). The BLAST program was used to perform genome versus genome and gene versus genome alignments. RNA folding was analyzed by using MFOLD. PHYLIP was used for phylogenetic tree construction. The FASTA-formatted sequences of the complete genomes were aligned with CLUSTAL W (v1.82) by using the default parameters for DNA alignments. The PHYLIP output option of CLUSTAL W was used to produce a multiple alignment file that was used as input for dnaml (v3.6), which produced an unrooted maximum-likelihood phylogenetic tree with the default parameters. ORF analysis was performed by using tools from the EMBOSS suite. In the case of SARS-HCoV and HCoV-OC43 ATCC, the extracted ORFs were submitted to HMMPFAM, of the HMMER suite, for motif detection against the PFAM database. The amino acid sequences of the known expressed proteins were also submitted to HMMPFAM and the patmatmotif tool of EMBOSS. This tool performs motif scanning against the PROSITE motif database.

Nucleotide sequence accession number. The GenBank sequence accession numbers for the complete genome of the HCoV-OC43 ATCC strain and the Paris isolate are, respectively AY585228 and AY585229.

RESULTS

Amplification and sequencing of HCoV-OC43 ATCC and Paris genomes. The genomes of the HCoV-OC43 ATCC strain and of the Paris isolate were amplified in six fragments by RT-PCR in order to be sequenced (Fig. 1 and Table 1). The PCR products encompassed the entire genome of the viruses and overlapped each other to make sure the final sequences were complete. Primers used for the amplifications of ORF 1a and ORF1b were created by using the sequence of the bovine coronavirus Quebec strain (BCoV Quebec) (63), which displays 97% identity in this region of the genome and was known to share 92% identity with the 3' 9 kb of the HCoV-OC43 genome (24, 26, 29, 37, 38, 39). Primers used to amplify the 3' region were designed on the basis of sequences of HCoV-OC43 available in GenBank. The gene-walking approach was used to sequence the whole genome of the HCoV-OC43 ATCC strain, whereas the Paris isolate was sequenced by using

the primers generated during the sequencing of the ATCC strain (Table 2).

Main features of the HCoV-OC43 genome. The genomes of the two variants contain 30,713 nt, excluding the poly(A) tail, and include nine main ORFs flanked by 5' (nt 1 to 209)- and 3' (nt 30426 to 30713)-untranslated regions (UTRs) (Fig. 1 and Table 3). The genome of HCoV-OC43 contains multiple secondary ORFs, scattered throughout the genome in all frames and in both orientations (data not shown). By using ShowORF software, it was possible to determine that the translation complex could potentially use several of these ORFs in the 5'→3' orientation with, for instance, a translation reinitiation mechanism (20, 36). By using better-characterized coronaviruses, putative transcription-regulating sequences (TRSs) of HCoV-OC43 were also identified (4, 44, 58) (Table 3). These sequences are found at the 5' end of each viral RNA, genomic or subgenomic, and represent signals for the discontinuous transcription of subgenomic mRNA (49). The identified canonical core sequence for HCoV-OC43 was 5'-UCUAAAC-3', but it was not always perfectly conserved throughout the genome (Table 3).

Using bioinformatics tools and well-characterized coronaviruses (15, 18, 67), it was also possible to draw a precise map of the main domains contained within the polyprotein 1ab and to determine the location of the putative viral proteolytic cleavage sites (Fig. 2). Most of the main motifs found throughout the genome after bioinformatics analysis corresponded to expected motifs found in other coronaviruses. Indeed, the PLP1 and PLP2 motifs, the membrane-spanning domains (TM), and the 3CLpro motif as well as the RdRp and the RNA helicase motifs, were found in ORF1ab at the expected positions compared to other coronaviruses. Since HCoV-OC43 and BCoV possess a high degree of identity, the positions of the cleavage sites were determined with the BCoV model (15), and 14 sites were identified in the polyprotein 1ab. Among these cleavage sites, three are recognized by PLP1 or PLP2, and the 11 others are recognized by 3CLpro, generating mature products containing key motifs for viral transcription and replication. A putative ribosomal -1 frameshift was also identified upstream

TABLE 2. Primers used for sequencing of the HCoV-OC43 genome^a

Positive-strand primer	Sequence (5'→3')	Location (nt)	Negative-strand primer	Sequence (5'→3')	Location (nt)
JUB3	GATTGTGAGCGATTGCGTGC	1-20	JUB12	ACATCACCTGTAGCTGTTGGC	6071-6091
RACEJUB2	TGIGATGGTGGATTGTCCGCCG	380-400	JUB11	CCAGTAACGCTGTGTAACCTTC	5750-5770
JUB31	GTTATATAATGATTGATCCTGC	662-682	JUB111	CCATGCCTCTTGCCATTGAAC	5277-5297
JUB32	TGATTATACTGGTAGTCTTGC	983-1003	JUB112	TGAATTTAACACCACTCAACAG	4918-4938
JUB33	GCACAATCTCAGGTGTTTGG	1488-1508	JUB113	ATCCTCTTGATTATGCTAAC	4458-4478
JUB14	AGTTGCTAGGTGTGTCAAGATG	1801-1821	JUB115	AGGGTTTACAACAACCTTCTGC	4086-4106
JUB141	CTTATAATAGTAGTGGAGAGTG	2254-2274	JUB15	AACAGCCATAGAATGACTATC	3876-3896
JUB142	GTTCIGATTTTTCATTAGCCGG	2557-2577	JUB151	GCTTIAGGCACATACAGACCC	3395-3415
JUB143	ACAGCTCCTGAAGATGATGAC	3093-3113	JUB152	ACCACAGCATAAAAATTCCTCC	2915-2935
JUB144	TTAACCTTATGTAGATTGGCAG	3621-3641	JUB153	CAAATTTCTTCTACGTTCCAATC	2399-2419
JUB161	ATGCTATGTTCTTTTATGGTG	3700-3720	JUB13	AATGCTTGACCACTTACTGCC	1991-2011
JUB163	TATTATGGGCATGGTATGTG	3959-3979	JUB131	CAAGCAAAACCATCTATCATG	1427-1447
JUB164	AAGTTATGTAATTTAGAGCG	4310-4330	JUB132	CTCCAAGTAGGAAATAATGCC	1049-1069
JUB165	TATTTTGGAGTGTACTGGAGGC	4782-4802	RACE JUB1	GGAGCAAATCATATCCACCTC	312-332
JUB166	TGCTTGCCATTACAACATGC	5128-5148	JUB6	CCCTCTAAATGTCGTGTGTAC	11111-11131
JUB5	CCTGCTAGATTTGTAICGTTG	5319-5339	JUB61	ATAGCAGCCAACAGTGTTTCC	10724-10744
JUB51	ACTCAGCGTATTAATAAAGCC	5913-5933	JUB62	AGGGTCACTGTAAGAACAAGC	10205-10225
JUB52	GTTGACGATGGTGGTGCAGC	6264-6284	JUB63	CTATTAAGAAGCAACTCAGAC	9752-9772
JUB53	AATCTACACCACAGAAATTGC	6701-6721	JUB64	AGAAAGTATGGGGTAAACCTTG	9453-9473
JUB54	TGGCAGGATTTGATATGTTAG	7036-7056	JUB18	ATCTCTACCACAAAAGTCC	9213-9233
JUB17	GGTTTACCCATGTTTGTCTC	7177-7197	JUB181	TGGTAGGGACATTAACACTG	8761-8781
JUB171	GCTTTCTTCAATGATCGTATG	7678-7698	JUB182	GCTTTCTTCAATTTATGCTGG	8345-8365
JUB172	CTGTGCTCGTAAAGTGTGTT	8072-8092	JUB183	CATAGCAAAAATGTTATCCAC	7950-7970
JUB173	AATAAGCAGATGGCTAATGTC	8403-8423	JUB184	AAGTATTACCTGGTTTAAAG	7543-7563
JUB55	AAGGTTTTATCCGCTTCCAG	9043-9063	JUB65	ACCTAACACTCCAATGTAATG	7252-7272
JUB56	CACCTACAATGGCTAGTTATG	9540-9560	JUB66	CTAAAAGTGTCTTAAATCCAC	6926-6946
JUB57	CCGTCTCAACTTCATTTCTGC	9925-9945	JUB67	TATACTAACAAAGATCATACC	6531-6551
JUB58	CTTGTGGATCTGTGGTTAATG	10378-10398	JUB68	ACATCACCTGTAGCTGTTGGC	6071-6091
JUB7	GGCTTCTACATTTTGTATTAG	10901-10921	JUB8	TAACCTGTCTTAAAGGCTTCC	16525-16545
JUB71	TGTATTTCACAGATATACTCC	11443-11463	JUB81	GATGTTTTGAGAAGAGCAGACC	16117-16137
JUB72	CAGCAGTTAAAACAGCTAGAG	12084-12104	JUB82	CATAATATCCTGTGACAAAGC	15522-15542
JUB73	ATAATGAGGTATCTGCTACTG	12547-12567	JUB83	TATAACTACAGGAACACAGC	15050-15070
JUB74	TGTTAAACCCGATGCTACCAC	13073-13093	JUB84	AAAAGTGCTTCAGATCAACTG	14604-14624
JUB75	AGAGAGATGAAATGCTATGAG	13559-13579	JUB85	ATACTCCAGTGCCTAAAATAC	14158-14178
JUB76	TTGGATTGCGAATTTGATGTG	14066-14086	JUB86	TCTTTCGCTACAATGTTGGC	13648-13668
JUB77	TTGTATGATTTACCGACTTGC	14465-14485	JUB87	TCCACAAACTTGACAAAATC	13245-13265
JUB78	CATTATCTTTGAGGACGAGG	14856-14876	JUB88	TCTTGCTAGTGTGTACAACC	12843-12863
JUB79	GAGGCATGTTGTCGCAAAGC	15230-15250	JUB89	CTTGGATAGTTTGAATCTGCC	12448-12468
JUB710	ATAFAAGTGCCTTTCAACAGG	15636-15656	JUB810	ACTAGAACACGCTCATCAAG	12048-12068
JUB711	CAAAAAGTTTACTGATGAGTCC	16040-16060	JUB811	CTFAAAATTAAGCATAAAGGGC	11649-11669
JUB9	TTATTGTGAAGATCATAAGCC	16309-16329	JUB812	AGAACAAATCATGGTTAATGC	11245-11265
JUB91	TCAACACATTTGGTATGAAACG	16903-16923	JUB10	AAATGGGTAAGTGGAAAATTG	21544-21564
JUB92	AGTTACTTTGTTTATAGGGTC	17498-17518	JUB101	AGAGCTAACTTGCTCGAATC	21064-21084
JUB93	TATTGGTGAATCTGCTGTTAC	18034-18054	JUB102	GCAAACACTCTTACTACCAC	20414-20434
JUB94	TAGAACTGGTTACTATGTTG	18562-18582	JUB103	GCGGTGGACCTCTTATCATCG	19908-19928
JUB95	TTTTGAGGCACATAAGGACTC	18997-19017	JUB104	GTATTGTAAAGACTCTAAGTAC	19369-19389
JUB96	AGTCAAGACTGGTCAATTAAC	19498-19518	JUB105	TACAAAAGAGTCTTAAACAGAC	18973-18993
JUB97	CGTTCTAATAATGGCGTTTAC	19859-19879	JUB106	CCCATGTAAGTCAACCAACAC	18444-18464
JUB98	TAGGCTTGTACCGAAGACAGC	20313-20333	JUB107	TTATATTGTCTACTACTGCC	17983-18003
JUB99	ATACTGAGTTATGTCAAATATC	20736-20756	JUB108	TTATGCCACAAAGGGTTAGCC	17599-17619
JUB100	TCGAGACAAGTTAGCTCTGGG	21067-21087	JUB109	GGTATGTAACACACTACTTACG	17153-17173
JUNSO3	TTAATGATATGGTTTATTCCC	21396-21416	JUB110	CTGCGGAACAAGCGTAGGAGC	16808-16828
JUNSO6	GTGTAGAAGAAATGCATGACC	21807-21827	JUSO2	AAGAACTTTAACAATGCTAG	27754-27774
JUHEO1	AACAATTTCTGGTTCTTCC	22226-22246	JUSO4	ATCAAGTGACAAAATCTGGTGC	27384-27404
JUHEO3	TTTAGGAGTTTTCACTTTACC	22588-22608	JUSO6	ATATGATTACCATTACCACAG	27020-27040
JUHEO5	ACCCTTTGTATTTTTAACGG	22959-22979	JUSO8	TGAATAGCATAAAGGGCATTG	26678-26698
JUHEO7	TCATGGAGATGCTGGTTTAC	23337-23357	JUSO10	GCTGCCTAGACAACCTAATAC	26307-26327
JUSO24	ATTAATAACCCGATTTACCC	23482-23502	JUSO12	AATGGCTCAAAAATTAGTAAAC	25937-25957
JUSO5	AATGATATAGTGAGTCCCTGC	23993-24013	JUSO14	TTATAATAAGTCCGATTACCC	25553-25573
JUSO7	CTTTCACACTATTATGTCTATG	24378-24398	JUSO16	ACAAGTTAAATAATTAGTACC	25188-25208
JUSO9	TATTCAGGCAGACTCAATTTAC	24752-24772	JUSO18	ATAGTTATGCTGGAAACACAC	24809-24829
JUSO11	AATTGAATGGTTCGTGTGTAG	25123-25143	JUSO20	TAGAAGTGAGAGGTGTAACCC	24436-24456
JUSO13	GTGTTTGTGTTAAATATGACC	25498-25518	JUSO22	TATAGGATGTATTTACAAAAG	24031-24051
JUSO15	TAGGTAGTGGTTACTGTGTGG	25867-25887	JUHEO2	ACATAATAAGTACCCAAACC	23778-23797
JUSO17	GAATGGTGTACTCTTAGCAC	26234-26254	JUHEO4	CATTATCATACCTAAAAACGC	23408-23428
JUSO19	TGGATGTGCTAAGTCAAAATC	26632-26652	JUHEO6	AGACCAATAATAACACCAAGTG	23037-23057
JUSO21	ATTTCTGTGGTAAATGGTAATC	27016-27036	JUHEO8	ATGATAAGGCGTAAAATTAAC	22660-22680
JUSO25	TGGCACCAGATTTGTCACTTGG	27382-27402	JUNSO2	GAAACAACATTTGGTAGGAGGG	22416-22436
JUMO1	GTGATGATATACTGGATACC	27649-27669	JUNSO4	TTCTTAAATGGACAGTCTGTC	21917-21937
JUMO3	TAGTTGCCATTTGTTTATTGG	28169-28189	JUO2	GCAGCAAGACATCCATTCTG	30495-30514
JUMO5	ATGTGGATTTGTATTTTGTG	28662-28682	JUO4	TACCAAAACACTGCTGAACAG	29920-29940
JUO8	ATGTCTTTTACTCCTGGTAAGC	29079-29100	JUO6	ATACCATCGTGGCAGCAGTTG	29430-29450
JUO3	TCTACTGGTCTGCTAGTAACC	29512-29532	JUMO2	CCACTTGAGGATGCCATTACC	29136-29156
JUO5	CCCACAGTCCCAATTTCTTG	29999-30019	JUMO4	TACACAAATCCACATAATAATG	28655-28675
JUO7	CTCTCTATCAGAAATGGATGTC	30487-30507	JUMO6	TATAAAAATTTATGGCCAC	28150-28170

^a Underlined nucleotides indicate mismatched bases with regard to the genome sequence of the HCoV-OC43 ATCC strain.

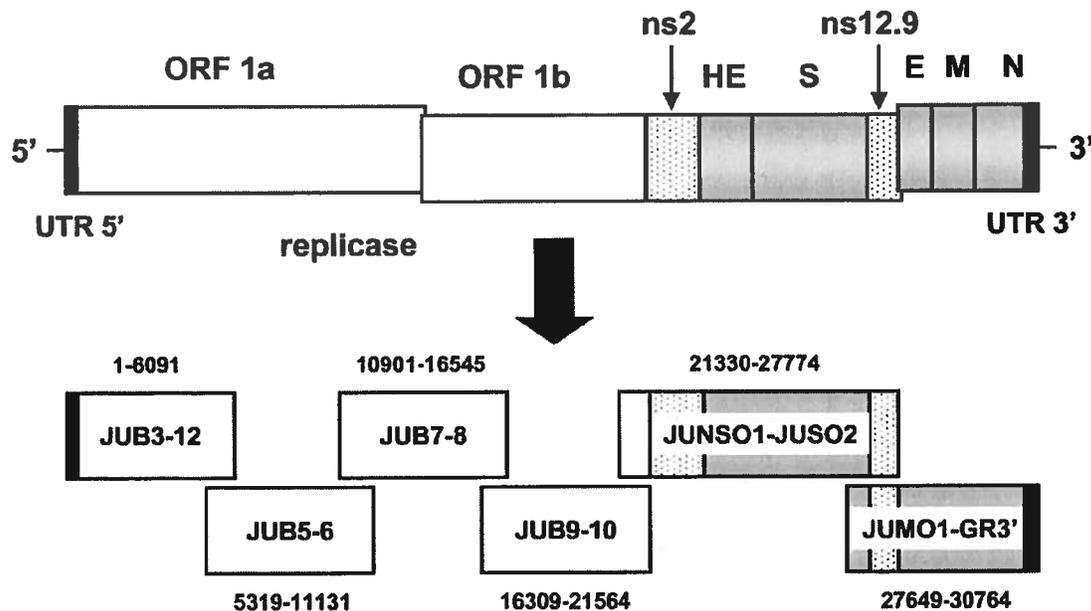


FIG. 1. Schematic representation of the HCoV-OC43 genome and of the amplification strategy used for sequencing. The HCoV-OC43 genome is 30,713 nt long and comprises nine main ORFs: ORF1a, ORF1b, ns2 (the gene encoding the nonstructural protein 2), HE (hemagglutinin-esterase gene), S (spike gene), ns12.9 (the gene encoding a nonstructural protein of 12.9 kDa), E (small envelope gene), M (membrane gene), and N (nucleocapsid gene). The replicase gene includes both ORF1a and ORF1b. The entire genome was amplified in six fragments in order to be sequenced. Each PCR product was named according to the name of the primers used for the amplification, and the location in the genome is indicated above or below each PCR product. Boxes: open, gene encoding the replicase polyprotein; dotted, genes encoding nonstructural proteins; shaded, genes encoding structural proteins; black, UTRs. GR, GeneRacer.

of the intersection of ORF1a and ORF1b. The slippery sequence $_{13334}\text{UUUAAAAC}_{13340}$ was found at the 3' end of ORF1a and is thought to be involved, in combination with RNA pseudoknot structures, in the frameshift, which would occur at the C_{13340} nt (58).

High degree of identity between the HCoV-OC43 ATCC strain and the Paris isolate. Differences in nucleotides and amino acids between the HCoV-OC43 ATCC strain and the

Paris isolate are presented in Table 4. In all, only 6 nt differ between the two variants. These mutations are located in the 5'UTR, in the ns2, S, M, and N genes, and in the 3'UTR. According to the MFOLD software, mutations located in the UTRs would not affect RNA folding (data not shown) and would therefore not have any effect on viral transcription and replication. Mutations located in the ns2 and M genes would not affect virus biology since they do not give rise to any amino

TABLE 3. Organization of the HCoV-OC43 genome

Genome region	Location (nt)	TRS location (nt)	TRS sequence ^a
Leader and 5'UTR	1–209	63–69	UCUAAAC ...139 nt... <u>AUG</u>
ORF1a	210–13361		
ORF1b ^b	13361–21496		
Intergenic region	21497–21505		
ns2 gene	21506–22342	21492–21498	UCUAAACUUUAAAAAUG
Intergenic region	22343–22353		
HE gene	22354–23628	22339–22344	UUAAACUCAGUGAAAAUG
Intergenic region	23629–23642		
S gene	23643–27704	23636–23642	UCUAAACAUG
Intergenic region	27705–27791		
ns12.9 gene	27792–28121	27771–27777	UCUUAAAGGCCACGCCCUAUUAAUG
E gene ^c	28108–28362		
Intergenic region	28363–28376		
M gene	28377–29069	28367–28373	UCCAAACAUAUG
Intergenic region	29070–29078		
N gene	29079–30425	29065–29070	UCUAAUUUUAAAGGAUG
3'UTR	30426–30713		
Poly(A) tail of 28 nt	30714–30741		

^a Nucleotides in boldface indicate TRS sequences, whereas underlined nucleotides indicate the initiation codon.

^b Putative ribosomal –1 frameshift between ORF1a and ORF1b.

^c ORF overlap for the ns12.9 gene and the E gene.

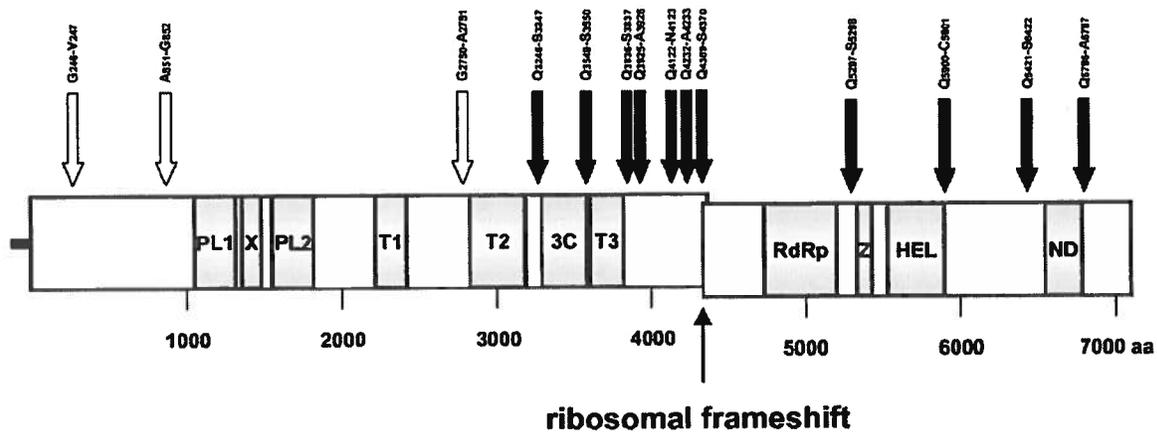


FIG. 2. Schematic representation of the polyprotein lab putative proteolytic processing and of the main domains found in ORF1ab. The approximate positions of predicted functional domains and protease cleavage sites in ORF1ab are shown, and amino acids positions are also indicated. The white arrows indicate putative cleavage sites recognized either by the PLP1 or the PLP2, whereas black arrows indicate sites recognized by the main protease, 3CLpro. The 15 putative cleavage products generated by the proteolytic processing are named as follows: leader protein, MHV p65-like protein, nsp1 (PL1, X, PL2, and T1), T2, nsp2 (3CLpro), nsp3 (T3), nsp4, nsp5, nsp6, nsp7, nsp9 (RdRp), nsp10 (HEL), nsp11, nsp12, and nsp13 (15). A putative ribosomal -1 frameshift is indicated between ORF1a and ORF1b. Upstream of the frameshift site, the slippery sequence ₁₃₃₃₄UUUAAAC₁₃₃₄₀ is found. PL1 and PL2, accessory protease domains; X, conserved domain of unknown function; T1, T2, and T3, membrane-spanning (hydrophobic) domains; 3C, 3CLpro domain; Z, putative zinc finger; HEL, NTPase RNA helicase domain; ND, domain conserved exclusively in nidoviruses. nsp, Nonstructural protein.

acid substitution. However, two mutations lead to amino acid substitutions. The first is located at nt 26514, in the S2 subunit of the S gene, and gives rise to the I958F (ATCC→Paris) mutation, whereas the second is located at nt 29320, in the N gene, and gives rise to the V81A mutation.

Neuroinvasion in BALB/c mice. After inhalation of virus, mice were processed for histochemical labeling of HCoV-OC43 ATCC antigens. Cells positive for viral antigens were first observed ca. 3 dpi in the olfactory bulb as patches of labeled neurons (Fig. 3A). No cells positive for viral antigens could be seen in other part of the brain, even near perivascular blood cells. At 7 dpi, viral antigens were detected in all brain regions, indicating a rapid dissemination throughout the CNS (Fig. 3B). Five mice of each group, infected with HCoV-OC43 ATCC and Paris, were sacrificed every 48 h, and virus titers were measured in the CNS and lung. Even though mice inhaled a viral suspension, virus was rarely found in the lung (limit of detection was $10^{1.5}$ TCID₅₀/g due to a lung extract toxicity on HRT-18 cells) and only when brain titers reached at least 10^4 TCID₅₀/g (data not shown). HCoV-OC43 ATCC infectious virus could be detected in mouse CNS, as early as 2

dpi. Virus titers were maximal ca. 4 dpi and remained high throughout the experiments (Fig. 3C). When virus reached the brain, replication of HCoV-OC43 ATCC led to a fatal encephalitis. Infectious HCoV-OC43 Paris could only be detected in mice starting at 6 dpi (Fig. 3C), and a lower number of mice were productively infected by HCoV-OC43 Paris than by HCoV-OC43 ATCC. Nevertheless, when infectious virus reached the brain, infectious virus titers were comparable for the two HCoV-OC43 variants, suggesting that the ATCC and Paris variants both exhibit neuroinvasive and neurotropic properties.

Comparison of HCoV-OC43 with other coronaviruses. HCoV-OC43 is part of the second genetic group of coronaviruses (30) and displays higher identity levels with virus strains that belong to this group, including SARS-HCoV. The coronavirus strains that present the highest degree of identity with HCoV-OC43 are BCoV and MHV-A59, with 95 and 71% identities, respectively. HCoV-OC43 and BCoV are very related at the nucleotide level, and most of the differences between the two genomes are found in the S1 subunit of the S gene, suggesting that the two virus strains possess similar biological properties but display a different cellular tropism. SARS-HCoV and HCoV-229E display 53.1 and 51.2% identity with regard to the HCoV-OC43 strain. Although SARS-HCoV is apparently part of group 2 (51), the overall identity level with the OC43 strain is not striking, but the two strains present a very high degree of amino acid identity in some important functional domains, such as the RdRp, the RNA helicase, and 3CLpro.

A phylogenetic unrooted tree regrouping seven coronavirus strains from the three genetic groups was obtained by using the complete genome sequences of all strains (Fig. 4). This tree is the first one that includes the complete genome of the HCoV-OC43 strain. It shows that HCoV-OC43 and BCoV are evolutionary very related and that they form a clade with MHV-

TABLE 4. Sequence differences between the reference strain HCoV-OC43 ATCC and the Paris isolate

Mutation location (nt)	Region of mutation	Consequence of mutation (ATCC→Paris)
31	5' UTR	C→T
22243	ns2 gene	TTC (Phe246)→TTT (Phe246) (no amino acid change)
26514	S gene	ATC (Ile958)→TTC (Phe958)
28808	M gene	ACT (Thr144)→ACC (Thr144) (no amino acid change)
29320	N gene	GTA (Val81)→GCA (Ala81)
30632	3' UTR	C→A

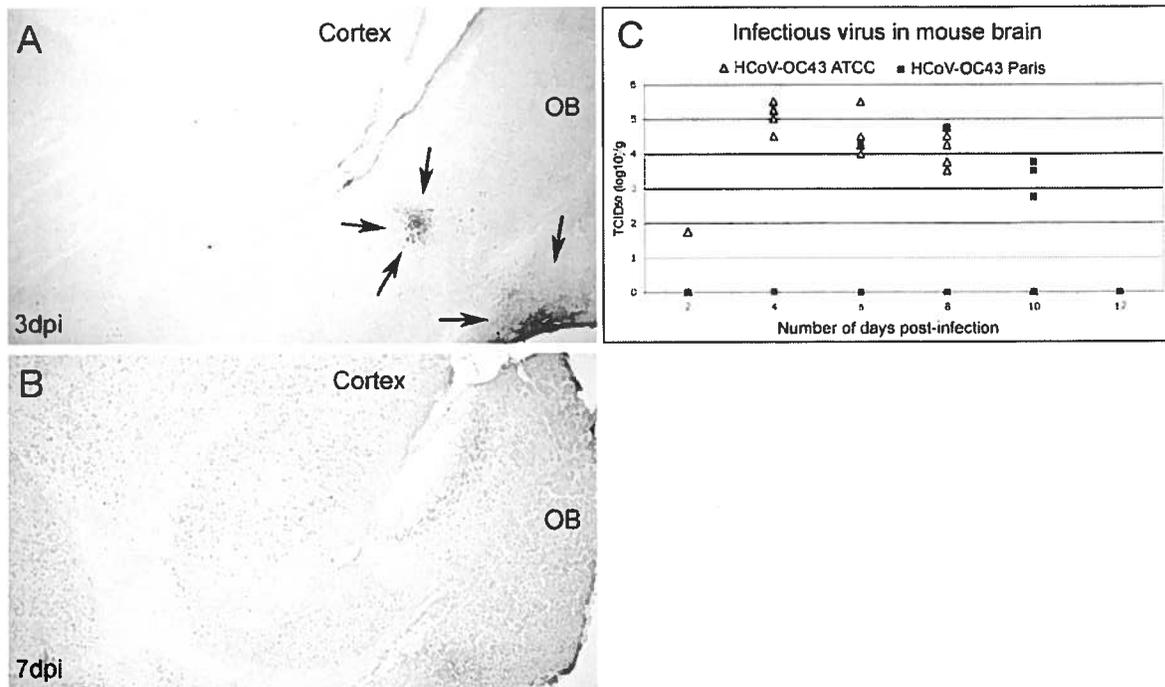


FIG. 3. Neuroinvasive properties of HCoV-OC43 ATCC and HCoV-OC43 Paris variant in BALB/c mice after intranasal inoculation. (A) At 3 dpi, cells positive for viral antigens (arrows) were first observed in the olfactory bulb (OB). No infected cells could be detected in the cortex or other brain structures, illustrating transneuronal spreading of the virus. (B) At 7 dpi, the virus has disseminated to the entire CNS, as illustrated by the presence of immunopositive cells throughout the brain. Magnification (A and B), $\times 32$. (C) Quantification of infectious virus in the brain of each mouse at different times postinfection. Virus titers are presented as logarithmic value of TCID₅₀ per gram of tissue (the limit of detection was $10^{0.5}$ TCID₅₀/g). Infection by HCoV-OC43 ATCC was detected in one mouse as early as 2 dpi, and gradually more mice became positive. HCoV-OC43 ATCC infectious particles were found between 2 to 8 dpi in mouse brain and led to fatal encephalitis before the end of the experimentation. Antigens of the HCoV-OC43 Paris variant were first revealed in mouse brain at 6 dpi. Infectious particles were detected in some of the brains up to 10 dpi. HCoV-OC43 Paris infectious titers in susceptible animals were similar to those found after HCoV-OC43 ATCC infection, and mice positive for either variant presented all pathological and clinical signs of encephalitis.

A59. Although SARS-HCoV is apparently part of group 2 (51), the analysis shows that it is more divergent from strains of the previous clade and that infectious bronchitis virus (IBV) and SARS-HCoV display the highest divergence among the strains analyzed. Group 1 coronaviruses are also grouped in such a clade.

A BLAST analysis with coronaviruses from all three genetic groups showed that different degrees of identity exist between several regions of different virus strains but that the most conserved region among all coronaviruses is located within ORF1b (data not shown). More stringent BLAST analysis was carried out on the genome sequences of HCoV-229E, MHV-A59, and SARS-HCoV with the HCoV-OC43 genome as a reference (data not shown). Among all genes analyzed, the most significant identity levels were found in ORF1ab, as well as in the S2 subunit of the S gene. Significant identity levels were observed with MHV-A59 and SARS-HCoV. With regards to the ORF1ab, the identity levels were usually lower in ORF1a than in ORF1b and were more significant in the case of MHV-A59 than for SARS-HCoV. Moreover, identity was more significant at the amino acid level. Several domains which are essential for viral replication, such as the 3CLpro, RdRp, and helicase domains, are very interesting because of their functional importance. The S2 subunit of the S gene also dis-

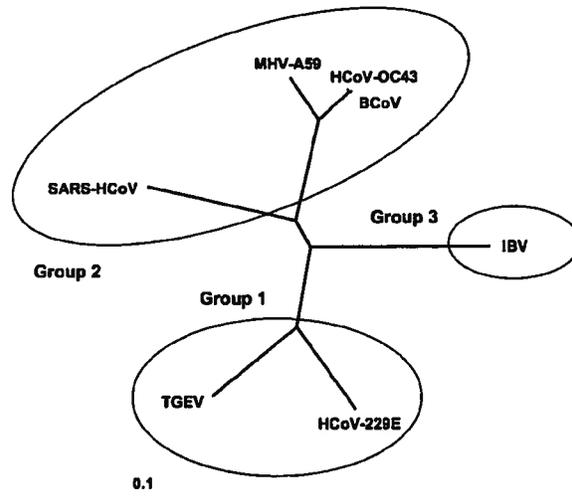


FIG. 4. Phylogenetic unrooted tree regrouping seven coronavirus complete genomes from the three genetic groups. Circles regroup members of each three genetic groups. The 0.1 sliding bar represents the genetic distance between the species (i.e., nucleotide substitution units per studied site). Strains: MHV-A59 (NC_001846); BCoV, bovine coronavirus Quebec strain (AF220295); SARS-HCoV, SARS-HCoV Tor2 strain (AY274119); IBV, IBV Beaudette strain (NC_001451); TGEV (NC_002306); HCoV-229E (NC_002645).

OC43	SGIVKMNPT	SKVEPCVSV	TYGNMTLNL	WLDDKVYCP	HVICASDMT	NPDYTNLLCR	VTSSDFTVLF	DR-LSLTVMS
BCoV	SGIVKMNPT	SKVEPCIVSV	TYGNMTLNL	WLDDKVYCP	HVICASDMT	NPDYTNLLCR	VTSSDFTVLF	DR-LSLTVMS
MHV-A59	SGIVKMVSPT	SKVEPCIVSV	TYGNMTLNL	WLDDKVYCP	HVICASDMT	DPDYPNLLCR	VTSSDFCVMS	GR-MSLTVMS
SARS Tor2	SGFRKMAFPS	GKVEGCMVQV	TCGTTTLNL	WLDDTVYCP	HVICTAEDML	NPNYEDLLIR	KSNHSLVQVA	GN-VQLRVIG
229E	AGLRKMAQPS	GFVEKCVVRV	CYGNVTLNL	WLGDIVYCP	HVIAS-NTTS	AIDYDHEYSI	MRLHNFSIIS	GT-AFLGVVG
IBV	SGFKKLVSPS	SAVEKCIIVSV	SYRGNLNL	WLGDITVYCP	HVLGK--FS	GDQWVDVNL	ANNHEFEVTT	QHGVTLNVS
OC43	YQMRGCMVLV	TVTLQNSRTP	KYTFGVVVKPG	ETFTVLAAYN	GKPGQAFHVT	MRSSYTIKGS	FLCGSGSSVG	YVIMGDCVKF
BCoV	YQMQGCMVLV	TVTLQNSRTP	KYTFGVVVKPG	ETFTVLAAYN	GKPGQAFHVT	MRSSYTIKGS	FLCGSGSSVG	YVIMGDCVKF
MHV-A59	YQMQGQQLVL	TVTLQNPNT	KYSFGVVKPG	ETFTVLAAYN	GRPGQAFHVT	LRSSHTIKGS	FLCGSGSSVG	YVLTGDSVRF
SARS Tor2	HSMQNCLLRL	KVDTSNPKTP	KYKRVRIQPG	QTFSVLACYN	GSPSGVYQCA	MRENHTIKGS	FLNGSGSSVG	FNIDYDCVSF
229E	ATMHGVTLKI	KVSQTNMHTP	RHSFRTLKSG	EGFNILACYD	GCAQGVGVN	MRTNWTIRGS	PINGACGSPG	YNLKNGEVEF
IBV	RRLKGAVALIL	QTAVANAETP	KYKFIKANGC	DSFTIACAYG	GTVVGLYPVT	MRSNGTIRAS	FLAGGSGSSVG	FNIEKGVVNF
OC43	VYMHQLELST	GCHTGTDFNG	DFYGPYKDAQ	VVQLLIQDYI	QSVNFWAWLY	AAILNNCN--	----WVQSD	KCSVEDFNW
BCoV	VYMHQLELST	GCHTGTDFNG	DFYGPYKDAQ	VVQLPVQDYI	QSVNFWAWLY	AAILNNCN--	----WVQSD	KCSVEDFNW
MHV-A59	VYMHQLELST	GCHTGTDFSG	NFYGPRDAQ	VVQLPVQDYT	QTVNVVAVWLY	AAIFNRCN--	----WVQSD	SCSLEEFNVW
SARS Tor2	CYMHMELPT	GVHAGTDLEG	KFYGPFVDRQ	TAQAAGTDTT	ITLNVLAWLY	AAVINGDR--	----WFLNRF	TTTTLNDFNLV
229E	VYMHQIELGS	GSHVGSFSDG	VYGGFEDQP	NLQVESANQM	LTVNVVAVFLY	AAILNGCT--	----WWLKE	KLFVEHYNEW
IBV	FYMHLELFPN	ALHTGTDLMG	EFYGGYVDEE	VAQRVPDNL	VTNNIVAWLY	AAIISVKESS	FSLPKWLEST	TVSVDYKWK
OC43	ALSNGFSQVK	SDLV--IDAL	ASMTGVSLET	LLAAIKRLK-	NGFQGRQIMG	SCSFEDELTP	SDVYQQLAGI	KLQ
BCoV	ALSNGFSQVK	SDLV--IDAL	ASMTGVSLET	LLAAIKRLK-	NGFQGRQIMG	SCSFEDELTP	SDVYQQLAGI	KLQ
MHV-A59	AMTNGFSSIK	ADLV--LDAL	ASMTGVTVEQ	VLAAIKRLH-	SGFQKQILG	SCVLEDELTP	SDVYQQLAGV	KLQ
SARS Tor2	AMKYNIEPLT	QDHVDILGPL	SAQTGIAVLD	MCAALKELLQ	NGMNGRTILG	STILEDEFTP	FDVVRQCSGV	TFQ
229E	AQANGFTAMN	GEDA--FSIL	AAKTGVVER	LLHAIQVLN-	NGFGKQILG	YSSLNDEFSI	NEVVKQMGV	NLQ
IBV	AGDNGFTPFS	TSTA--ITKL	SAITGVDVCK	LLRTIMVKN-	SQWGGDPILG	QYNFEDELTP	ESVFNQIGV	RLQ

FIG. 5. Multiple alignments of amino acids of the main proteases of coronaviruses from all three genetic groups. Positions with absolute conservation are shadowed, whereas residues of the putative catalytic dyad, His⁴¹ and Cys¹⁴⁵, are boxed. Conservation level among group 2 coronaviruses was ca. 46.2%, whereas all strains displayed 26% identity. Strains: OC43, HCoV-OC43 (group 2); BCoV, BCoV Quebec group 2; MHV-A59, MHV group 2; SARS Tor2, SARS-HCoV Tor2 group 2; 229E, HCoV-229E group 1; IBV, IBV group 3.

played high identity levels with its counterparts from other coronaviruses. For instance, the MHV-A59 S2 subunit displayed 76% identity and 88% similarity, whereas the S1 subunit presented only 53% identity and 65% similarity. This result is logical since it has been shown that the membrane fusion function resides within the S2 subunit (54, 62), whereas the S1 subunit is involved in receptor binding (56) and determination of tropism (3), which is different from one virus to another.

Since SARS-HCoV is now considered as a serious pathogen that has recently emerged and that we believe HCoV-OC43 could represent an excellent model for the study of this virus, it was of interest to analyze some functionally important motifs that display significant identity levels with the HCoV-OC43 genome. The most striking identities between the two strains were found mainly in ORF1b, albeit the 3CLpro motif, in ORF1a, also presented a significant identity level. The cleavage product containing the 3CLpro motif displayed 48% identity and 64% similarity with the corresponding region of HCoV-OC43. Of the three viral proteases that play a role in the processing of the polyprotein 1ab, 3CLpro is the main protease (67). This domain of the viral genome is essential for replication since it cleaves the HCoV-OC43 polyprotein 1ab at 11 sites and allows the release of important functional domains (Fig. 2) (32). Like other coronavirus 3CLpros, HCoV-OC43 3CLpro acts via a catalytic dyad, which is composed of a His⁴¹ and a Cys¹⁴⁵ (6). The HCoV-OC43 3CLpro is 303 amino acids

long and displays an outstanding conservation among coronaviruses from the three genetic groups (Fig. 5). Seventy-nine residues are strictly conserved among sequences from six different coronaviruses, displaying 26% identity among all 3CLpro sequences analyzed, whereas group 2 coronaviruses display 46.2% identity for the same motif between each other.

DISCUSSION

HCoV-OC43 belongs to the second genetic group of coronaviruses and represents the HCoV that is most related to SARS-HCoV. Here, we present the first report of a complete sequence of the HCoV-OC43 genome, including the complete sequence of a clinical respiratory isolate of the OC43 serotype. The two genomes are 30,713 nt long and only differ by 6 nt, including two amino acid substitutions located in the S (I958F)- and N (V81A)-protein genes. The genomes of the two virus variants display 71, 53.1, and 51.2% identity with the genomes of MHV-A59, SARS-HCoV Tor2, and HCoV-229E, respectively. Using bioinformatics tools and well-characterized coronaviruses, further characterization of the HCoV-OC43 genome was performed, and these analyses revealed that HCoV-OC43 is closely related to BCoV and MHV and that it displays significant amino acid identity levels with important functional domains of the SARS-HCoV. Like the ATCC strain that was isolated in the 1960s, HCoV-OC43 Paris, isolated in 2001, exhibited neuroinvasive properties in BALB/c mice. Although

mice were more easily infected with the ATCC strain than with the Paris isolate, these results suggest that both viruses possess the intrinsic ability to infect neural cells and to reach the CNS from the periphery.

Recently, L. Vijgen and coworkers have submitted a complete sequence of the HCoV-OC43 genome to GenBank (NC_005147). The virus strain used for this sequencing is described as corresponding to the virus strain that was used in our laboratory (VR-759). However, comparison of our sequence with theirs show that they differed at 33 positions, 29 mutations being located in the S gene, including two mutations in the S2 subunit. Of the four other differences, one is located at the beginning of the genome sequence, where a guanine is added with respect to our sequence, whereas the other three are scattered throughout ORF1a. Despite these differences, the availability of the complete genome sequence from a clinical isolate reinforces the validity of our sequence, since the HCoV-OC43 ATCC and Paris sequences only differ by 6 nt. Therefore, this observation suggests that the viral strain used by Vijgen and collaborators could have been adapted in cell culture, given the differences observed in the S gene, which is known to be associated with viral adaptation (27). No differences were noticed among ORF1b sequences between HCoV-OC43 ATCC, Paris, and the one from Vijgen and coworkers. This observation suggests that this region of the genome needs a high rate of conservation in order to remain functional and that genes located downstream of the replicase gene are more permissive to sequence modifications.

Using a recent HCoV-OC43 clinical respiratory isolate, we showed here that HCoV-OC43 apparently remains genetically stable in the environment. Indeed, despite virus shedding and chances of persistence in the host, the HCoV-OC43 Paris isolate displays differences at only six positions with regard to the ATCC strain sequence, despite about 40 years have elapsed between the two isolations. Since the viral persistence could be associated with molecular adaptation (7, 8), the low rate of mutation observed here could be explained by the fact that the HCoV-OC43 Paris isolate has never or rarely persisted before. However, it is too soon to speculate about such an issue given that the exact origin of the virus before its isolation remains undetermined. It is also worth noting that viral persistence does not necessarily require an adaptation to the environment (2) and that, despite the high rate of mutation of the coronavirus RdRp (1), 95% of the mutations engendered by RNA virus polymerases are deleterious and therefore not conserved (42).

Our observation that inhalation of HCoV-OC43 led to a generalized infection of the whole CNS in mice demonstrates neuroinvasiveness. This result confirms that HCoVs have neuroinvasive properties in mice, which was first shown in newborn mice (10, 22) and which is consistent with their detection in human brain (9, 12, 40, 53). After inhalation, the first infected cells were detected in the olfactory bulb, illustrating that virus directly reached the brain by a transneuronal route, as already demonstrated for MHV (10, 31, 47). The HCoV-OC43 Paris isolate, which was never propagated in mouse brain or other neurological tissue, also exhibited neuroinvasive properties in mice. Replication within the CNS was similar for the two variants, but fewer mice were infected by the HCoV-OC43 Paris isolate than by the ATCC strain. These data suggest that

only one mutation in the S gene, giving rise to one amino acid modification, could partially modulate the neuroinvasiveness of one variant over the other. Indeed, a single amino acid change has already been demonstrated to influence MHV ability to spread within the CNS (43, 59).

Although the degree of sequence conservation between the genomes of the HCoV-OC43 ATCC and Paris variants is very high, their phenotypes seem to differ slightly in mice, since the ATCC strain reached the CNS more easily. As we have demonstrated *in vitro* with primary hippocampus and cortical cell cultures, both HCoV-OC43 ATCC and Paris variants were able to replicate in rodent neurons, although the HCoV-OC43 ATCC strain yielded more infectious virus particles than the HCoV-OC43 Paris isolate. However, the two viral variants exhibited different biological properties, such as plaque formation and cytopathic effects on different cell lines (II. Jacomy and P. J. Talbot, unpublished data).

Although both mutations preserve some but not all properties of the parental residues, the I958F mutation leads to a substitute phenylalanine that does not display the same steric hindrance than the isoleucine, which could potentially affect protein folding and function. Moreover, the I958F mutation is located in the S2 subunit of the S gene and would probably be positioned in the putative fusion peptide domain (23), conferring a lot of impact to this mutation at the biological level. On its own, this mutation could therefore have the capacity to influence the phenotype of the HCoV-OC43 Paris isolate because it may interfere with the fusion process in a positive or a negative manner (43). Given the known involvement of the S protein in viral biology and pathogenesis (7, 8, 15, 48), this mutation is more likely to influence the phenotype of the Paris isolate. It has been reported that the N protein may be involved in viral RNA synthesis (30) and that it could colocalize with nucleolar antigens and delay the cell cycle (14). However, the fact that the V81A mutation within the N gene is positioned in domain I of the protein should not influence the RNA binding properties of N, since this functional feature of the protein lies in domain II (45). Therefore, even though the role of both mutations needs to be investigated, we feel that the S mutation is more likely to influence the virus phenotype.

Comparison with better-characterized coronaviruses (23, 59) suggests that the I958F mutation is located in the putative S fusion peptide and could therefore affect viral fusogenic properties and phenotype. Although no fusion peptides have formally been identified in any coronavirus S protein, predictions have located such fusion sequences near the N terminus of the heptad repeat 1 (HR1) for MHV (33). Studies with the MHV-A59 S protein also showed that mutations introduced in the HR1 region severely affected cell-cell fusion ability (33). Moreover, it has already been reported that a single mutation introduced in HR1 could influence the degree of MHV virulence (59). Depending on the effect of the mutation on cleavage ability, the phenotype of the resulting virus could also be affected. Although the cleavage of the S protein is not absolutely required for fusion (23, 52, 55), it has been shown to enhance fusogenicity (55). Thus, inhibition of S-protein cleavage would be associated with a more stable interaction between S1 and S2 and would correlate with a loss of fusogenicity (25). So, as observed by Tsai et al. (59) for the MHV-JHM strain, the I958F mutation in the S gene of the HCoV-OC43

Paris isolate could either alter the conformation of the S protein or have an incidence on its cleavage, impairing the ability of the virus to spread within the CNS.

An animal model for the HCoV-OC43 ATCC strain has recently been developed and optimized in our laboratory (22). Moreover, HCoV-OC43 may also be used as a model for the study of SARS-HCoV, not only because of the identity level the two virus strains display but because HCoV-OC43 can also be studied without the requirement of a level three, aerosol-aware, biological confinement. Indeed, we have now demonstrated that the two virus strains present a high level of conservation for some essential functional domains, especially within 3CLpro, the RdRp, and the RNA helicase. This result is consistent with the possible sharing of several important properties by these two viruses. All of these motifs represent potential candidates for therapy of coronavirus-mediated diseases because they are specific targets and because of the specificity they exhibit toward their substrate. Indeed, substrate specificities of all coronavirus proteases, and mainly 3CLpro, are conserved among the three established groups (19), and this is also true for SARS-HCoV. The picornavirus RdRp (21) and viral proteases (17) have notably been designated as such targets for antiviral therapy. At present, the SARS-HCoV 3CLpro enzyme represents the most promising target for SARS therapy (58). The availability of 3CLpro crystal structures should provide a valuable tool for rapid identification of potential drugs against SARS. Thus far, 3CLpro crystal structures have been obtained for transmissible gastroenteritis virus (TGEV) (5), HCoV-229E (6) and, more recently, for SARS-HCoV (61). A putative *in vitro* inhibitor has also been identified for TGEV (5) and SARS-HCoV (61). This inhibitor, hexapeptidyl chloromethyl ketone, was shown to bind the 3CLpro enzyme very efficiently *in vitro* and, although it provides an excellent structural basis for drug design, *in vivo* experiments need to be performed on this issue.

Now that the complete genome sequence of HCoV-OC43 has been deciphered, it will provide a very useful tool for the study of coronaviruses from all genetic groups and particularly for those of group 2, including SARS-HCoV. Indeed, the genome sequence will allow comparative studies with other coronavirus strains and RNA viruses and will also allow optimization of prediction models. This sequence will also allow the assembly of an infectious cDNA clone of HCoV-OC43, which is currently under way. Thus far, cDNA clones have been assembled for several coronavirus strains by using different approaches. Among these clones, those of TGEV (3, 64), HCoV-229E (57), IBV (13), MHV-A59 (66), and even SARS-HCoV (65) are now available. The HCoV-OC43 clone will provide an invaluable tool to further understand the underlying mechanisms for replication and pathogenesis of HCoVs.

ACKNOWLEDGMENTS

We are grateful to Francine Lambert for technical assistance with cell and virus culture. We also thank D. W. Yoo, who kindly provided the ORF1ab sequence from the BCoV, which allowed us to sequence the HCoV-OC43 ORF1ab.

This study was supported by grant MT-9203 from the Canadian Institutes for Health Research (Institute of Infection and Immunity). P.J.T. is the holder of a Tier-I (Senior) Canada Research Chair. J.S.J. acknowledges a studentship from the Fonds Québécois de Recherche sur la Nature et les Technologies.

REFERENCES

- Adami, C., J. Pooley, J. Glomb, E. Stecker, F. Fazal, J. O. Fleming, and S. C. Baker. 1995. Evolution of mouse hepatitis virus (MHV) during chronic infection: quasispecies nature of the persisting MHV RNA. *Virology* **209**: 337-346.
- Addie, D. D., I. A. T. Schaap, L. Nicolson, and O. Jarrett. 2003. Persistence and transmission of natural type I feline coronavirus infection. *J. Gen. Virol.* **84**:2735-2744.
- Almazán, F., J. M. González, Z. Pénzes, A. Izeta, E. Calvo, J. Plana-Durán, and L. Enjuanes. 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. USA* **97**: 5516-5521.
- Alonso, S., A. Izeta, I. Sola, and L. Enjuanes. 2002. Transcription regulatory sequences and mRNA expression levels in the coronavirus transmissible gastroenteritis virus. *J. Virol.* **76**:1293-1308.
- Anand, K., G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld. 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain. *EMBO J.* **21**:3213-3224.
- Anand, K., J. Ziebuhr, P. Wadhvani, J. R. Mesters, and R. Hilgenfeld. 2003. Coronavirus main proteinase (3CL^{pro}) structure: basis for design of anti-SARS drugs. *Science* **300**:1763-1767.
- Arbour, N., G. Côté, C. Lachance, M. Tardieu, N. R. Cashman, and P. J. Talbot. 1999. Acute and persistent infection of human neural cell lines by human coronavirus OC43. *J. Virol.* **73**:3338-3350.
- Arbour, N., S. Ekandé, G. Côté, C. Lachance, F. Chagnon, M. Tardieu, N. R. Cashman, and P. J. Talbot. 1999. Persistent infection of human oligodendrocytic and neuroglial cell lines by human coronavirus 229E. *J. Virol.* **73**:3326-3337.
- Arbour, N., R. Day, J. Newcombe, and P. J. Talbot. 2000. Neuroinvasion by human respiratory coronaviruses. *J. Virol.* **74**:8913-8921.
- Barthold, S. W., M. S. de Souza, and A. L. Smith. 1990. Susceptibility of laboratory mice to intranasal and contact infection with coronaviruses of other species. *Lab. Anim. Sci.* **40**:481-485.
- Bonavia, A., N. Arbour, V. W. Yong, and P. J. Talbot. 1997. Infection of primary cultures of human neural cells by human coronaviruses 229E and OC43. *J. Virol.* **71**:800-806.
- Burks, J. S., B. L. DeVald, L. D. Jankovsky, and J. C. Gerdes. 1980. Two coronaviruses isolated from central nervous system tissue of multiple sclerosis patients. *Science* **209**:933-934.
- Casais, R., V. Thiel, S. G. Siddell, D. Cavanagh, and P. Britton. 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J. Virol.* **75**:12359-12369.
- Chen, H., W. Torsten, P. Britton, G. Brooks, and J. A. Hiscox. 2002. Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell. *J. Virol.* **76**:5233-5250.
- Chouljenko, V. N., X. Q. Lin, J. Storz, K. Kousoulas, and A. E. Gorbalenya. 2001. Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. *J. Gen. Virol.* **82**:2927-2933.
- Denison, M. R., W. J. Spaan, Y. van der Meer, C. A. Gibson, A. C. Sims, B. Prentice, and X. T. Lu. 1999. The putative helicase of the coronavirus mouse hepatitis virus is processed from the replicase gene polyprotein and localizes in complexes that are active in viral RNA synthesis. *J. Virol.* **73**:6862-6871.
- Dragovich, P. S., T. J. Prins, R. Zhou, T. O. Johnson, Y. Hua, H. T. Luu, S. K. Sakata, E. L. Brown, F. C. Maldonado, T. Tunland, C. A. Lee, S. A. Fuhrman, L. S. Zelman, A. K. Patick, D. A. Matthews, E. Y. Wu, M. Guo, B. C. Borer, N. K. Nayyar, T. Moran, L. Chen, P. A. Rejto, P. W. Rose, M. C. Guzman, E. Z. Doval Santos, S. Lee, K. McGee, M. Mohajeri, A. Liese, J. Tao, M. B. Kosa, B. Liu, M. R. Batugo, J. P. Gleeson, Z. P. Wu, J. Liu, J. W. Meador III, and R. A. Ferre. 2003. Structure-based design, synthesis and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 8. Pharmacological optimization of orally bioavailable 2-pyridone-containing peptidomimetics. *J. Med. Chem.* **46**:4572-4585.
- Gorbalenya, A. E. 2001. Big nidovirus genome. When count and order of domains matter. *Adv. Exp. Med. Biol.* **494**:1-17.
- Hegyí, A., and J. Ziebuhr. 2002. Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **83**:595-599.
- Horvath, C. M., M. A. Williams, and R. A. Lamb. 1990. Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. *EMBO J.* **9**:2639-2647.
- Hung, M., C. S. Gibbs, and M. Tsiang. 2002. Biochemical characterization of rhinovirus RNA-dependent RNA polymerase. *Antivir. Res.* **56**:99-114.
- Jacomy, H., and P. J. Talbot. 2003. Vacuolating encephalitis in mice infected by human coronavirus OC43. *Virology* **315**:20-33.
- Jan Bosch, B., R. van der Zee, C. A. M. de Haan, and P. J. M. Rottier. 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **77**:8801-8811.
- Kamahora, T., L. H. Soe, and M. M. C. Lai. 1989. Sequence analysis of

- nucleocapsid gene and leader RNA of human coronavirus OC43. *Virus Res.* 12:1–9.
25. Krueger, D. K., S. M. Kelly, D. N. Lewicki, R. Ruffolo, and T. M. Gallagher. 2001. Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *J. Virol.* 75:2792–2802.
 26. Kündel, F., and G. Herrler. 1993. Structural and functional analysis of the surface protein of human coronavirus OC43. *Virology* 195:195–202.
 27. Kündel, F., and G. Herrler. 1996. Structural and functional analysis of the S proteins of two human coronavirus OC43 strains adapted to growth in different cells. *Arch. Virol.* 141:1123–1131.
 28. Kuo, L., G.-J. Godeke, M. J. B. Raamsman, P. S. Masters, and P. J. M. Rottier. 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J. Virol.* 73:1393–1406.
 29. Labonté, P., S. Mounir, and P. J. Talbot. 1995. Sequence and expression of the ns2 protein gene of human coronavirus OC43. *J. Gen. Virol.* 76:431–435.
 30. Lai, M. M. C., and D. Cavanagh. 1997. The molecular biology of coronaviruses. *Adv. Virus Res.* 48:1–97.
 31. Lavi, E., P. S. Fishman, M. K. Highkin, and S. R. Weiss. 1988. Limbic encephalitis after inhalation of a murine coronavirus. *Lab. Invest.* 58:31–36.
 32. Lu, Y., X. Lu, and M. R. Denison. 1995. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *J. Virol.* 69:3554–3559.
 33. Luo, Z., and S. R. Weiss. 1998. Roles in cell-to-cell fusion of two conserved hydrophobic regions in the murine coronavirus spike protein. *Virology* 244:483–494.
 34. Marra, M. A., S. J. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, Y. S. Butterfield, J. Khatter, J. K. Asano, S. A. Barber, S. Y. Chan, A. Cloutier, S. M. Coughlin, D. Freeman, N. Girn, O. L. Griffith, S. R. Leach, M. Mayo, H. McDonald, S. B. Montgomery, P. K. Pandoh, A. S. Petrescu, A. G. Robertson, J. E. Schein, A. Siddiqui, D. E. Smailus, J. M. Stott, G. S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T. F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G. A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R. C. Brunham, M. Krajden, M. Petric, D. M. Skowronski, C. Upton, and R. L. Roper. 2003. The genome sequence of the SARS-associated coronavirus. *Science* 300:1399–1404.
 35. Masters, P. S. 1999. Reverse genetics of the largest RNA viruses. *Adv. Virus Res.* 53:245–264.
 36. Meyers, G. 2003. Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependent reinitiation mechanism. *J. Biol. Chem.* 278:34051–34060.
 37. Mounir, S., and P. J. Talbot. 1992. Sequence analysis of the membrane protein gene of human coronavirus OC43 and evidence for O glycosylation. *J. Gen. Virol.* 73:2731–2736.
 38. Mounir, S., and P. J. Talbot. 1993. Molecular characterization of the S protein gene of human coronavirus OC43. *J. Gen. Virol.* 74:1981–1987.
 39. Mounir, S., and P. J. Talbot. 1993. Human coronavirus OC43 RNA 4 lacks two open reading frames located downstream of the S gene of bovine coronavirus. *Virology* 192:355–360.
 40. Murray, R. S., B. Brown, D. Brian, and G. F. Cabirac. 1992. Detection of coronavirus RNA and antigen in multiple sclerosis brain. *Ann. Neurol.* 31:525–533.
 41. Myint, S. H. 1994. Human coronaviruses: a brief review. *Rev. Med. Virol.* 4:35–46.
 42. Novella, I., S. 2003. Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. *Curr. Opin. Microbiol.* 6:399–405.
 43. Ontiveros, E., T. S. Kim, T. M. Gallagher, and S. Perlman. 2003. Enhanced virulence mediated by the murine coronavirus, mouse hepatitis virus strain JHM, is associated with a glycine at residue 310 of the spike glycoprotein. *J. Virol.* 77:10260–10269.
 44. Ozdarendeli, A., S. Ku, S. Rochat, G. D. Williams, S. D. Senanayake, and D. A. Brian. 2001. Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. *J. Virol.* 75:7362–7374.
 45. Parker, M. M., and P. S. Masters. 1990. Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein. *Virology* 179:463–468.
 46. Peiris, J. S., C. M. Chu, V. C. Cheng, K. S. Chan, I. F. Hung, L. L. Poon, K. I. Law, B. S. Tang, T. Y. Hon, C. S. Chan, K. H. Chan, J. S. Ng, B. J. Zheng, W. L. Ng, R. W. Lai, Y. Guan, and K. Y. Yuen. 2003. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* 361:1767–1772.
 47. Perlman, S., G. Evans, and A. Afifi. 1990. Effect of olfactory bulb ablation on spread of a neurotropic coronavirus into the mouse brain. *J. Exp. Med.* 172:1127–1132.
 48. Rowe, C. L., S. C. Baker, M. J. Nathan, and J. O. Fleming. 1997. Evolution of mouse hepatitis virus: detection and characterization of spike deletion variants during persistent infection. *J. Virol.* 71:2959–2969.
 49. Sawicki, S. G., and D. L. Sawicki. 1998. A new model for coronavirus transcription. *Adv. Exp. Med. Biol.* 440:215–219.
 50. Schickli, J. H., B. D. Zelus, D. E. Wentworth, S. G. Sawicki, and K. V. Holmes. 1997. The murine coronavirus mouse hepatitis virus strain A59 from persistently infected murine cells exhibits an extended host range. *J. Virol.* 71:9499–9507.
 51. Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. M. Spaan, and A. E. Gorbalenya. 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331:991–1004.
 52. Stauber, R., M. Pfeleiderera, and S. Siddell. 1993. Proteolytic cleavage of the murine coronavirus surface glycoprotein is not required for fusion activity. *J. Gen. Virol.* 74:183–191.
 53. Stewart, J. N., S. Mounir, and P. J. Talbot. 1992. Human coronavirus gene expression in the brains of multiple sclerosis patients. *Virology* 191:502–505.
 54. Sturman, R. M., C. S. Ricard, and K. V. Holmes. 1990. Conformational change of the coronavirus peplomer glycoprotein at pH 8.0 and 37°C correlates with virus aggregation and virus-induced cell fusion. *J. Virol.* 64:3042–3050.
 55. Taguchi, F. 1993. Fusion formation by the uncleaved spike protein of murine coronavirus JHMV variant cl-2. *J. Virol.* 67:1195–1202.
 56. Taguchi, F. 1995. The S2 subunit of the murine coronavirus spike protein is not involved in receptor binding. *J. Virol.* 69:7260–7263.
 57. Thiel, V., J. Herold, B. Schelle, and S. G. Siddell. 2001. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.* 82:1273–1281.
 58. Thiel, V., K. A. Ivanov, Á. Patic, T. Hertzog, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, and J. Ziebuhr. 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* 84:2305–2315.
 59. Tsai, J. C., L. de Groot, J. D. Pinon, K. T. Iacono, J. J. Phillips, S. H. Seo, E. Lavi, and S. R. Weiss. 2003. Amino acid substitutions within the heptad repeat domain 1 of murine coronavirus spike protein restrict viral antigen spread in the central nervous system. *Virology* 312:369–380.
 60. Tsang, K. W., P. L. Ho, G. C. Ooi, W. K. Yee, T. Wang, M. Chan-Yeung, W. K. Lam, W. H. Seto, L. Y. Yam, T. M. Cheung, P. C. Wong, B. Lam, M. S. Ip, J. Chan, K. Y. Yuen, and K. N. Lai. 2003. A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* 348:1977–1985.
 61. Yang, H., M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. Rao. 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. USA* 100:13190–13195.
 62. Yoo, D. W., M. D. Parker, and L. A. Babiuk. 1991. The S2 subunit of the spike glycoprotein of bovine coronavirus mediates membrane fusion in insect cells. *Virology* 180:291–304.
 63. Yoo, D., and Y. Pei. 2001. Full-length genomic sequence of bovine coronavirus (31 kb): completion of the open reading frame 1a/1b sequences. *Adv. Exp. Med. Biol.* 494:73–76.
 64. Yount, B., K. M. Curtis, and R. S. Baric. 2000. Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model. *J. Virol.* 74:10600–10611.
 65. Yount, B., K. M. Curtis, E. A. Fritz, L. E. Hensley, P. B. Jahrling, E. Prentice, M. R. Denison, T. W. Geisbert, and R. S. Baric. 2003. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. USA* 100:12995–13000.
 66. Yount, B., M. R. Denison, S. R. Weiss, and R. S. Baric. 2002. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J. Virol.* 76:11065–11078.
 67. Ziebuhr, J., E. J. Snijder, and A. E. Gorbalenya. 2000. Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J. Gen. Virol.* 81:853–879.

Letter to the Editor

Genetic Variability of Human Respiratory Coronavirus OC43

A recent paper by J. R. St-Jean and colleagues reported the complete genome sequences of the human coronavirus OC43 (HCoV-OC43) laboratory strain from the American Type Culture Collection (ATCC) and an HCoV-OC43 clinical isolate, designated Paris (7). The ATCC HCoV-OC43 (VR759) strain originated in 1967 and was passaged several times in suckling mouse brain and in cell culture. The contemporary HCoV-OC43 Paris strain was isolated in 2001 and was subsequently cultured in an HRT-18 cell line by the authors. A high degree of genetic stability was stated for HCoV-OC43 since only six nucleotide variations in the whole genome could be observed between both HCoV-OC43 strains, with isolation dates 34 years apart. There are, however, some arguments to suggest that the HCoV-OC43 genome is not as stable as suggested by the authors and that the HCoV-OC43 Paris strain might not be a contemporary strain but might be a result of cross-contamination with the ATCC HCoV-OC43 strain.

A first argument is based on the reported evolutionary rates of RNA viruses in general and coronaviruses in particular. An evolutionary rate of 4.0×10^{-4} nucleotide substitutions per site per year was estimated for severe acute respiratory syndrome (SARS) coronavirus by using ORF1ab sequence data (5), and an evolutionary rate of 7.5×10^{-4} nucleotide substitutions per site per year was estimated for porcine transmissible gastroenteritis virus by using S gene sequence data (6). Most RNA viruses have been reported to have evolutionary rates in the range of 10^{-4} to 10^{-3} substitutions per site per year, although slightly slower mutation rates have also been described (1, 2). Here, we estimated the evolutionary rate for the complete genome sequence data of HCoV-OC43 (one strain; GenBank accession number NC_005147 [isolated in 1967]) and BCoV (four strains; GenBank accession numbers U00735 and AF220295 [isolated in 1972] and NC_003045 and AF391542 [isolated in 1998]) using a maximum likelihood (ML) approach. A ML phylogenetic tree for the complete genomes was reconstructed with PAUP version 4.10 (8), and the rate of nucleotide substitution was estimated by using Rhino software version 1.2 (<http://evolve.zoo.ox.ac.uk/>), which implements a molecular clock model accommodating serially sampled sequences (4). This approach resulted in an estimate of 1.54×10^{-4} nucleotide substitutions per site per year (95% confidence interval, 0.97×10^{-4} to 2.12×10^{-4}) (4), which is 30-fold higher than the mutation rate (5.7×10^{-6} nucleotide substitutions per site per year) corresponding to the data presented by St-Jean et al. The maximum likelihood evolutionary rate was used to assess the probability that the contemporary HCoV-OC43 strain described by St-Jean and colleagues was effectively sampled in 2001. Assuming that nucleotide substitutions follow a Poisson process (3), the probability of observing only six mutations ($n = 6$) between two isolates sampled 34 years apart ($t = 34$) can be calculated by using the following model:

$$P(n, t|\lambda) = \frac{e^{-\lambda t}}{n!} (\lambda t)^n$$

where λ denotes the evolutionary rate in nucleotide substitutions per year.

For the maximum likelihood estimate of the evolutionary rate, the expected number of mutations after 34 years of evolution is 163. The probability of observing only six mutations after 34 years of evolution is 4.98×10^{-61} (the probability of observing six or fewer mutations is 5.17×10^{-61}). Even for the lower 95% confidence interval limit of the evolutionary rate, the probability of observing only six mutations or fewer during this time is extremely low ($P = 4.92 \times 10^{-36}$) and therefore we believe that it seems unlikely that the Paris HCoV-OC43 strain is truly a circulating strain from 2001. It should be noted that this approach is conservative since it ignores any shared ancestry of the two viral strains.

Furthermore, in a paper by Vabret and colleagues from the same laboratories as St-Jean et al. describing an HCoV-OC43 outbreak in Normandy, France, in 2001, M gene sequence data analysis revealed an estimated genetic distance of up to approximately 32 nucleotide changes per 1,000 nucleotides between some of the clinical isolates (9). For the whole M gene, this observation would apply to up to approximately 22 nucleotide variations between some of the 2001 isolates, whereas St-Jean and coworkers found only one nucleotide change in the M gene between two HCoV-OC43 strains isolated 34 years apart. Also, St-Jean and colleagues found only one variation in the S gene, the most variable coronavirus gene. Comparison of this observation to the data of the Normandy outbreak would imply that the S gene of two HCoV-OC43 strains isolated 34 years apart is more conserved than the M gene of several clinical HCoV-OC43 isolates, all from the same year (2001), which seems improbable.

In conclusion, we believe that the findings of St-Jean et al. regarding the genetic stability of human coronavirus OC43 are unlikely. We doubt that the HCoV-OC43 contemporary strain presented by the authors is a circulating strain from 2001, which is supported by an extremely low probability value. We propose that this might be due to contamination with the HCoV-OC43 ATCC strain of the cell line used to propagate the clinical isolate.

This work was supported by a fellowship from the Flemish Fonds voor Wetenschappelijk Onderzoek (FWO) to Leen Vijgen and by FWO grant G.0288.01.

REFERENCES

1. Domingo, E., and J. J. Holland. 1988. High error rates, population equilibrium and evolution of RNA replication systems, p. 3–36. In E. Domingo, J. J. Holland, and P. Ahlquist (ed.), RNA genetics, vol. 3. CRC Press, Boca Raton, Fla.
2. Jenkins, G. M., A. Rambaut, O. G. Pybus, and E. C. Holmes. 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J. Mol. Evol.* 54:156–165.
3. Kimura, M. 1983. The neutral allele theory of molecular evolution. Cambridge University Press, Cambridge, United Kingdom.
4. Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* 16:395–399.
5. Salemi, M., W. M. Fitch, M. Ciccozzi, M. J. Ruiz-Alvarez, G. Rezza, and M. J. Lewis. 2004. Severe acute respiratory syndrome coronavirus sequence characteristics and evolutionary rate estimate from maximum likelihood analysis. *J. Virol.* 78:1602–1603.
6. Sanchez, C. M., F. Gebauer, C. Sune, A. Mendez, J. Dopazo, and L. Enjuanes.

1992. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology* 190:92-105.
7. St-Jean, J. R., H. Jacomy, M. Desforges, A. Vabret, F. Freymuth, and P. J. Talbot. 2004. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J. Virol.* 78:8824-8834.
8. Swofford, D. L. 1998. PAUP* 4.0: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Massachusetts.
9. Vabret, A., T. Mourez, S. Gouarin, J. Petitjean, and F. Freymuth. 2003. An outbreak of coronavirus OC43 respiratory infection in Normandy, France. *Clin. Infect. Dis.* 36:985-989.

Leen Vijgen

Philippe Lemey

Els Keyaerts

Marc Van Ranst*

Laboratory of Clinical and Epidemiological Virology

Department of Microbiology and Immunology

Rega Institute for Medical Research

University of Leuven, Leuven, Belgium

Minderbroedersstraat 10

B-3000 Leuven, Belgium

*Phone: (32 16) 347908

Fax: (32 16) 347900

E-mail: marc.vanranst@uz.kuleuven.ac.be

Authors' Reply

In their comment regarding our recently published paper (11), Vijgen et al. present different arguments to underline their hypothesis that the genetic stability of the human coronavirus HCoV-OC43 that we report should be seen as an highly improbable event.

According to Vijgen et al., statistical analysis based on RNA virus evolution models reveals that the mutation rate per site per year for HCoV-OC43 should be between 0.97×10^{-4} and 2.12×10^{-4} (95% confidence interval), with a maximum likelihood of 1.54×10^{-4} .

Based on their arguments, the six-nucleotide difference we found between the HCoV-OC43 ATCC strain and the HCoV-OC43 Paris isolate obtained in March 2001 (11) represents a mutation rate of 5.7×10^{-6} , which is 30-fold lower than expected from theoretical statistical calculations. Furthermore, using the Poisson process model, Vijgen et al. calculated that the probability of this result (only six nucleotide differences between the genome of the HCoV-ATCC strain and the Paris isolate) is only 1.9×10^{-16} (4.92×10^{-36} for six nucleotides or fewer).

The statistical arguments presented by Vijgen et al. appear very interesting and convincing in terms of statistical analysis based on some previously reported evolutionary rates of RNA viruses. Furthermore, as the authors of the paper comparing these two HCoV-OC43 variants (11), we do indeed acknowledge that our published data of apparent genetic stability is surprising. However, we are putting forward herein some hypotheses that may contribute to explain the similarity of the genome of the two HCoV-OC43 variants described in our paper (11), which may explain the apparent genetic stability of HCoV-OC43.

Before the sequencing of its genome, the HCoV-OC43 Paris isolate was passaged on the HRT-18 cell line six times. The human coronavirus contained in the original clinical isolate, which comes from the upper respiratory tract of a 68-year-old immunocompromised male who was not related whatsoever to laboratory work and was not in contact with any laboratory workers who had manipulated the HCoV-OC43 ATCC virus (citation from the Materials and Methods section of our paper), consisted, as a typical RNA virus, of a complex mixture of variants, or quasispecies (2, 3, 6). As reported by Elena (3),

some sort of evolutionary constraint may exist for adaptation in a particular environment. According to this assumption, by having been cultured within the HRT-18 cell line for six passages, the HCoV-OC43 Paris strain may have been subject to this type of constraint, which would have highly favored its replication in the HRT-18 cells. In other words, the six passages on the HRT-18 cells would have selected a particular variant (with only a six-nucleotide difference from HCoV-OC43 ATCC) that was present in the quasispecies mixture in the clinical isolate. Furthermore, we can now add the supplemental data that the Paris variant was the only one among several clinical isolates that was able to replicate in HRT-18 cells, meaning that this particular isolate could have already been very similar to HCoV-OC43 ATCC at the onset of the viral isolation procedure (A. Vabret and F. Freymuth, unpublished results).

As already described in our paper (11), highly stringent laboratory precautions were used in order to eliminate a possible cross-contamination between the two viruses (citation from the Materials and Methods section of our paper: "The HCoV-OC43 ATCC strain and the Paris isolate were never cultured at the same time, and stringent laboratory precautions were used in order to eliminate possible cross-contamination."). We are still highly confident that all of the manipulations that served to either isolate or amplify the HCoV-OC43-Paris variant were performed properly, as we were always acutely aware of the potential problems of cross-contamination.

As Vijgen et al. are suggesting a possible contamination of the HRT-18 cell line used to propagate the HCoV-OC43-ATCC variant, we have performed RT-PCR assays on the HRT-18 cells that were used to replicate the HCoV-OC43 and were not able to detect any HCoV-OC43 RNA even after molecular hybridization using a probe specific for the M gene of the virus (Vabret and Freymuth, unpublished). A preexisting undetected infection of the HRT-18 cell line used to propagate HCoV-OC43 therefore appears highly unlikely.

The point mutation located in the M gene of the HCoV-OC43 Paris variant (T432C) is quite interesting. Indeed, while it is absent in the M gene of the HCoV-OC43 ATCC variant, this precise mutation is found in all 20 samples of HCoV-OC43 that were isolated during the Normandy outbreak of 2001 (13) and directly sequenced by Vabret et al. without any passage on any cell line (Vabret and Freymuth, unpublished). Furthermore, the T432C mutation is found neither in the first reported sequence of the M gene, which was reported by our laboratory (8), nor in the complete genome sequences of the HCoV-OC43 so far available (11; Vijgen et al., GenBank accession no. NC_005147). Therefore, this mutation appears to be a marker for the clinical isolates from the outbreak of HCoV-OC43 in Normandy during winter 2001 (13) and is not found in HCoV-OC43 ATCC.

As reported by Vijgen et al., the evolutionary rates of RNA viruses in general (6), and of coronaviruses in particular (10), are usually higher than the results presented in our paper (11). However, some recent publications reported a higher-than-expected genetic stability for RNA viruses. For example, it has been reported that the usually highly variable regions of the VP proteins of hepatitis A virus (4) and the highly variable *env* gene of feline immunodeficiency virus (FIV) (7) are shown experimentally to be more genetically stable than predicted statistically. Indeed, even though the latter is a retrovirus that could establish a latent infection, its mutation rate is usually comparable to that of human immunodeficiency virus type 1, at about 0.34% per site per year (5). However, Ikeda et al. (7)

showed that FIV could remain very stable despite a 10-year infection in cats, as the mutation rate within the *env* gene was only 0.015% per year (7). This result represents an unexpected genetic stability, as the rate of 0.015% is about 23-fold lower than what was previously reported (5). This finding is in the same order of magnitude as the 30-fold less-than-expected variability calculated by Vijgen et al. on the basis of our published experimental data (11). Moreover, even though the SARS-HCoV was pointed out by Vijgen et al. as an example to illustrate the extreme genetic variability of coronaviruses (9), another very interesting paper on SARS-HCoV indicates instead that this virus shows a relative genetic stability, with only 2 mutations in the S gene and 1 mutation in the N gene found in 10 different clinical isolates over a 3-month period (12). Finally, another recent study on a coronavirus, the feline coronavirus, also indicated that this virus could be more genetically stable than theoretically expected. Indeed, the sequencing of a small portion of the S gene from samples isolated from five persistently infected cats revealed low rates of mutations, ranging from none over a period of 17 months for one cat to only nine nucleotide changes over a 5-year period for another animal (1).

In conclusion, given the above comments and supportive arguments from the published literature, and even though they are, as underlined by the comment of Vijgen et al., unexpected, we do stand by our results concerning the apparent genetic stability of the HCoV-OC43 virus.

REFERENCES

1. Addie, D. D., I. A. T. Schaap, L. Nicolson, and O. Jarrett. 2003. Persistence and transmission of natural type I feline coronavirus infection. *J. Gen. Virol.* **84**:2735–2744.
2. Domingo, E. 1997. Rapid evolution of viral RNA genomes. *J. Nutr.* **127**(suppl. 5):958S–961S.
3. Elena, S. F. 2002. Restrictions to RNA virus adaptation: an experimental approach. *Antonie Leeuwenhoek* **81**:135–142.
4. Gabrieli, R., G. Sanchez, A. Macaluso, F. Cenko, S. Bino, L. Palombi, E. Buonomo, R. M. Pinto, A. Bosch, and M. Divizia. 2004. Hepatitis in Albanian children: molecular analysis of hepatitis A virus isolates. *J. Med. Virol.* **72**:533–537.
5. Green, W. K., J. Meers, G. del Fierro, P. R. Carnegie, and W. F. Robinson. 1993. Extensive sequence variation of feline immunodeficiency virus *env* genes in isolates from naturally infected cats. *Arch. Virol.* **133**:51–62.
6. Holland, J., and E. Domingo. 1998. Origin and evolution of viruses. *Virus Genes* **16**:13–21.
7. Ikeda, Y., T. Miyazawa, Y. Nishimura, K. Nakamura, Y. Tohya, and T. Mikami. 2004. High genetic stability of TM1 and TM2 strains of subtype B feline immunodeficiency virus in long-term infection. *J. Vet. Med. Sci.* **66**:287–289.
8. Mounir, S., and P. J. Talbot. 1992. Sequence analysis of the membrane protein gene of human coronavirus OC43 and evidence for O-glycosylation. *J. Gen. Virol.* **73**:2731–2736.
9. Salemi, M., W. M. Fitch, M. Cicozzi, M. J. Ruiz-Alvarez, G. Rezza, and M. J. Lewis. 2004. Severe acute respiratory syndrome coronavirus sequence characteristics and evolutionary rate estimate from maximum likelihood analysis. *J. Virol.* **78**:1602–1603.
10. Sanchez, C. M., F. Gebauer, C. Sune, A. Mendez, J. Dopazo, L. Enjuanes. 1992. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology* **109**:92–105.
11. St-Jean, J. R., H. Jacomy, M. Desforges, A. Vabret, F. Freymuth, and P. J. Talbot. 2004. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J. Virol.* **78**:8824–8834.
12. Tong, S., J. R. Lingappa, Q. Chen, B. Shu, A. C. LaMonte, B. T. Cook, C. Birge, S. W. Chern, X. Liu, R. Galloway, L. Q. Mai, W. F. Ng, J.-Y. Yang, J. Butany, J. A. Comer, S. S. Monroe, S. R. Beard, T. G. Ksiazek, D. Erdman, P. A. Rota, M. A. Pallansch, and L. J. Anderson. 2004. Direct sequencing of SARS-coronavirus S and N genes from clinical specimens shows limited variation. *J. Infect. Dis.* **190**:1127–1131.
13. Vabret, A., T. Mourez, S. Gouarin, J. Petitjean, and F. Freymuth. 2003. An outbreak of coronavirus OC43 respiratory infection in Normandy, France. *Clin. Infect. Dis.* **36**:985–989.

Julien R. St-Jean
Hélène Jacomy
Marc Desforges
Pierre J. Talbot*

*Laboratory of Neuroimmunovirology
 INRS-Institut Armand-Frappier
 531 boulevard des Prairies
 Laval, Québec H7V 1B7, Canada*

Astrid Vabret
François Freymuth
*Laboratoire de Virologie Humaine et Moléculaire
 Centre Hospitalier Régional et Universitaire de Caen
 Avenue Georges Clémenceau
 14033 Caen, France*

*Phone: (450) 686-5515
 Fax: (450) 686-5566
 E-mail: pierre.talbot@inrs-iaf.quebec.ca

Recovery of a Neurovirulent Human Coronavirus OC43 from an Infectious cDNA Clone†

Julien R. St-Jean,¹ Marc Desforages,¹ Fernando Almazán,² Hélène Jacomy,¹
Luis Enjuanes,² and Pierre J. Talbot^{1*}

Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier, 531 Boulevard des Prairies, Laval, Québec H7V 1B7, Canada,¹ and Department of Molecular and Cell Biology, Centro Nacional de Biotecnología, CSIC, Campus Universidad Autónoma, Darwin 3, 28049 Madrid, Spain²

Received 23 October 2005/Accepted 9 January 2006

This study describes the assembly of a full-length cDNA clone of human coronavirus (HCoV)-OC43 in a bacterial artificial chromosome (BAC). The BAC containing the full-length infectious cDNA (pBAC-OC43^{FL}) was assembled using a two-part strategy. The first step consisted in the introduction of each end of the viral genome into the BAC with accessory sequences allowing proper transcription. The second step consisted in the insertion of the whole HCoV-OC43 cDNA genome into the BAC. To produce recombinant viral particles, pBAC-OC43^{FL} was transfected into BHK-21 cells. Recombinant virus displayed the same phenotypic properties as the wild-type virus, including infectious virus titers produced in cell culture and neurovirulence in mice.

Human coronaviruses (HCoVs) are positive-stranded enveloped RNA viruses that are mainly associated with respiratory infections and are responsible for up to one-third of common colds (10) and possibly neurologic disease (3, 8). Like SARS-CoV, which was identified as the causative agent of severe acute respiratory syndrome (SARS) (11), HCoV-OC43 (ATCC VR-759 from the mid 1960s; GenBank accession number AY585228) (17) belongs to the second genetic group of the *Coronaviridae* (7, 14).

Several coronavirus infectious clones have been assembled, although this has proven difficult given the length of the viral genome (9), as well as the presence of unstable viral cDNA regions in bacteria (2, 6, 21). Uncommon strategies were therefore necessary to circumvent technical limitations engendered by the viral genome. Among the strategies available for the construction (2, 4, 5, 18, 19, 20, 21), the bacterial artificial chromosome (BAC) approach was selected given the availability of the technology, as well as the number of possibilities it offers for downstream applications (1, 2, 15, 16, 22). This approach uses a two-step amplification system that couples viral RNA expression in the nucleus from the cytomegalovirus (CMV) immediate-early promoter, with a second amplification step in the cytoplasm driven by the viral polymerase.

Construction of the full-length HCoV-OC43 cDNA clone. Prior to introducing the full-length viral cDNA into the pBeloBAC11 (13), pBAC-OC43-5'-3' was first assembled. This plasmid includes the first 720 nucleotides (nt) of the genome under the control of the CMV promoter, and the last 1,651 nt of the viral RNA, followed by a 25-bp synthetic poly(A) tail, the hepatitis delta virus ribozyme (HDV) and the bovine growth hormone termination and polyadenylation sequences (BGH) to lead to

an accurate 3' end. In order to join the CMV promoter with the viral 5' end and the viral 3' end with the HDV-BGH sequences in a contiguous and very precise way, overlapping PCRs were performed.

Primers used to perform the overlapping PCRs are listed in Table 1. The vector pBAC-TGEV-5'-3', which contains the accessory sequences CMV, HDV, and BGH, was used as a template to amplify these latter sequences. After two rounds of amplification, an overlapping PCR was performed for each end of the viral genome (Fig. 1A). Both overlapping PCR amplicons were directly introduced into the pBeloBAC11 plasmid in order to create the pBAC-OC43-5'-3' (Fig. 1B). After each cloning step, the new BAC vector was sequenced to assess that no undesired mutations were introduced into the new construct.

Before performing the assembly of the full-length cDNA clone, 5' rapid amplification of cDNA ends was carried out with RNA extracted from BHK-21 cells transfected with pBAC-OC43-5'-3' in order to check that the transcription machinery was functional (data not shown). The HDV cleavage efficiency was also determined by using a reverse transcription-PCR approach and, as expected, it was estimated at approximately 60% (data not shown) (12).

The HCoV-OC43 infectious cDNA clone was assembled using a five-step cloning approach, starting with the pBAC-OC43-5'-3' as a backbone. Given the low availability of unique restriction sites in the HCoV-OC43 genome, three additional restriction sites (MluI, AsiSI, and SacII) were introduced into the sequence of the cDNA clone. This modification required the introduction of six silent point mutations (T6479C, T6482G, C12143T, T12146C, A24068G, and T24071G). These sequence changes were used as genetic markers to identify the recombinant virus (2, 19). For cDNA assembly, we took advantage of the presence of the unique SfoI restriction site in both the BAC vector and the HCoV-OC43 genome (Fig. 1B and 2A).

The genome of HCoV-OC43 was therefore amplified in five distinct overlapping regions, named A to E, which spanned the entire viral genome, except for the 5' and 3' ends (Fig. 2A). Each

* Corresponding author. Mailing address: Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier, 531 Boulevard des Prairies, Laval H7V 1B7, Québec, Canada. Phone: (450) 686-5515. Fax: (450) 686-5566. E-mail: pierre.talbot@iaf.inrs.ca.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

TABLE 1. Primers used for the construction of the HCoV-OC43 cDNA clone

Primer pairs and location	Sequence (5'→3')
Primers for pBAC-OC43-5'-3'	
CMV5' (7716-7736 of pBAC-TGEV-5'-3') ^{a,d}	GCGGAATTCGTGCACTTGACATTGATTATTGACTAG ^b
CMV3' (8297-8316 of pBAC-TGEV-5'-3')	GCACGAAATCGCTCACAATACGGTTCACTAAACGAGCTC
OC43L-5' (1-41 of the OC43 genome)	ATIGTGAGCGATTTCGCGTGCATCCCGCTTCACTGATC
OC43L-3' (699-719 of the OC43 genome) ^d	GCGGAATTCGTGACCCCAACCCACGAATTGACCTG
OC43R-5' (29064-29083 of the OC43 genome) ^d	GCGCAAGCTTATCTAAATTTTAAGGATGTC
OC43R-3' (30683-30713 of the OC43 genome)	GTGATTCTTCCAATTGGCCATAATTAACTTC
HDV5' (30697-30746 of the OC43 genome and 217-227 of pBAC-TGEV-5'-3', respectively)	CCAATTGGAAGAATCACAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAGGGTCGGCATG
BGH3' (506-525 of pBAC-TGEV-5'-3') ^d	GCGCAAGCTTGCTCTCCCCAGCATGCCTGC
Primers for pBAC-OC43^{FL}	
JUB1 (28-47 of the OC43 genome)	CCCCTTCACTGATCTCTTG
JUBMlu reverse (6469-6491 of the OC43 genome)	TCGTCCGGCGCTCTGCTCAACGCGTTAGCAGTTC
JUBMlu forward (6469-6492 of the OC43 genome)	GAACTGCTAACGCGTTGAGCAGAG
JUBAsISI reverse (12127-12154 of the OC43 genome)	AGGTTGGGCGCCGTACAGCGCGATCGCGTTCATAAGCAG
JUBAsISI forward (12127-12154 of the OC43 genome)	CTGCTTATGAACGCGATCGCGCTGTAGC
JUB105 (18973-18993 of the OC43 genome)	TACAAAAGAGTCTTAACAGAC
JUB94 (18562-18582 of the OC43 genome)	TAGAACTGGTTACTATGGTTG
JUBSac2 reverse (24056-24082 of the OC43 genome)	GAATTGATTGTCCGCGTTGTACTACC
JUBSac2 forward (24058-24082 of the OC43 genome)	TAAGGTGGCGCTAGTACAACCGCGGACAATCAATTC
JUO2 (30495-30514 of the OC43 genome)	GCAGCAAGACATCCATTCTG

^a The sequence of pBAC-TGEV-5'-3' was provided by Luis Enjuanes.

^b Nucleotides in boldface contain restriction sites and are not included in the primer location.

^c Underlined nucleotides indicate mutated bases with regard to the genome sequence of the HCoV-OC43 ATCC strain (GenBank accession number AY585228).

^d External primer used for the generation of overlapping PCRs.

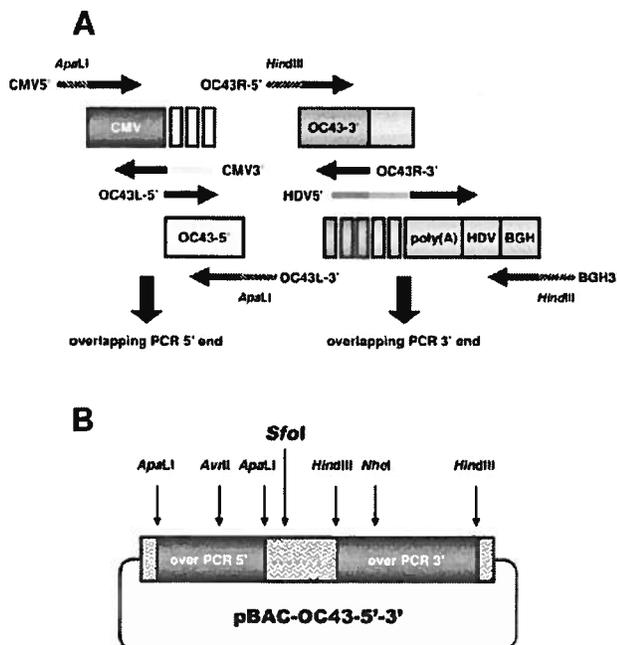


FIG. 1. Assembly of the pBAC-OC43-5'-3'. (A) Schematic representation of the overlapping PCR method, used to juxtapose each viral end to accessory sequences. The viral 5' end was joined to the CMV promoter in order to generate the overlapping PCR 5' end, whereas the viral 3' end was joined to the HDV-BGH sequences to generate the overlapping PCR 3' end. In the first PCR round, two products were generated for each end using specific primers. In the second round of amplification, the overlapping extension, CMV and OC43-5' PCR products were pooled and fused since they slightly overlap each other, generating overlapping PCR 5' products. The same method was used for the generation of overlapping PCR 3'. The primers used for the generation of the final products were the external ones (Table 1), which contained restriction sites for subsequent cloning into the BAC vector. (B) Schematic representation of pBAC-OC43-5'-3'. The overlapping PCR 5' end was introduced in the ApaI restriction site, whereas the overlapping PCR 3' end was introduced using the HindIII site. The vector contains the central SfoI restriction site, as well as the NheI and AvrII sites, which were used to start the assembly with regions A and B, respectively. Supplemental material is available at <http://jvi.asm.org>.

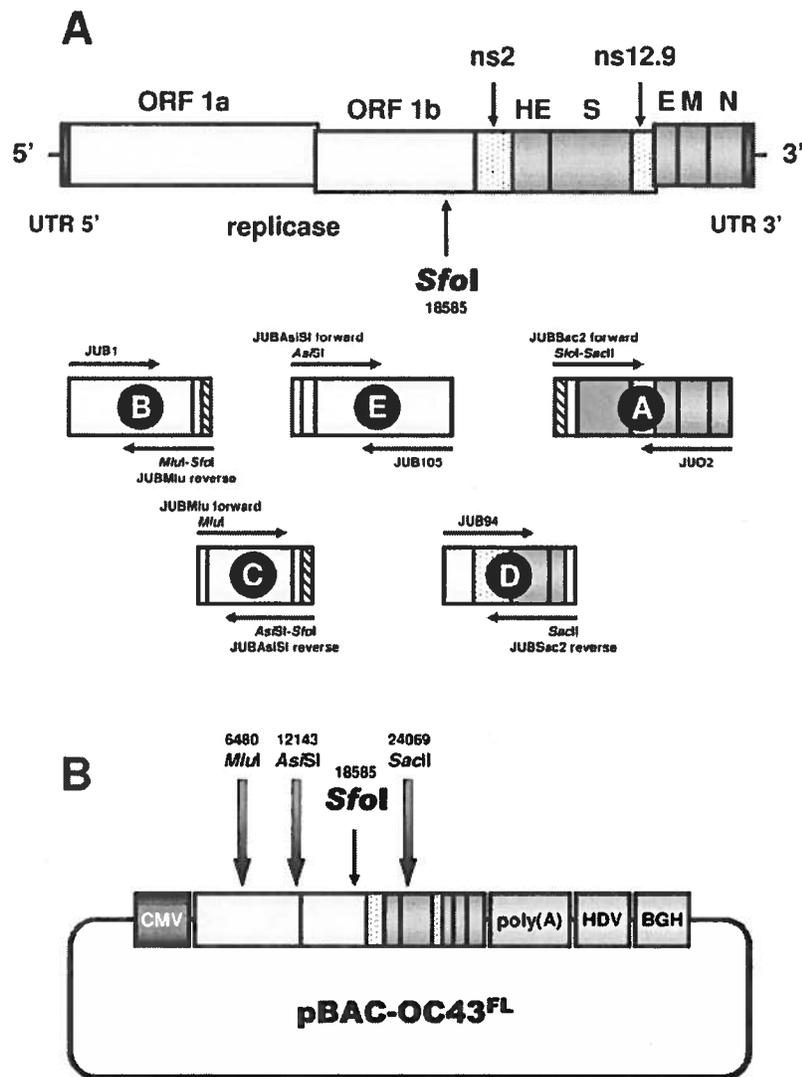


FIG. 2. Assembly of the pBAC-OC43^{FL}. (A) Schematic representation of the five-step cloning strategy used for the construction of pBAC-OC43^{FL}. Primers used for the amplification of regions A to E are shown (Table 1). Regions A, B, and C allowed the introduction of the SacII, MluI, and AsiSI restriction sites, respectively. These same regions also contained a mutagenized SfoI site (⊞), whereas regions D and E contained the original SfoI site, located at position 18585 of the viral genome. Other specific restriction sites, introduced by site-directed mutagenesis using primers JUBSac2 forward and reverse, JUBMlu forward and reverse and JUBAsiSI forward and reverse, are shown in white. (B) Schematic representation of the full-length infectious cDNA clone. Restrictions sites introduced by mutagenesis are represented by large arrows. The SfoI restriction site is also indicated by a small arrow. ORF, open reading frame; ns, accessory genes encoding nonstructural proteins; HE, hemagglutinin-esterase protein gene; S, spike protein gene; E, envelope protein gene; M, membrane protein gene; N, nucleocapsid protein gene; UTR, untranslated region; CMV, cytomegalovirus promoter; HDV, hepatitis delta virus ribozyme; BGH, bovine growth hormone termination and polyadenylation sequences. Supplemental material is available at <http://jvi.asm.org>.

region was generated in such a way that it contained a given cloning site at one end and a SfoI site at the other end. Primers used for the generation of region A to E are described in Table 1. Each amplicon was first inserted into the pSMART HCKan vector in order to generate plasmids pSMART-A to -E (Fig. 2A) and positive clones with the appropriate sequence were obtained.

Viral DNA regions contained in clones pSMART-A to -E were used to fill the pBAC-OC43-5'-3' and to generate the pBAC-OC43^{FL} (Fig. 2B). Regions A to E were respectively introduced into the BAC vector using the restriction sites they were sharing with the construct (Fig. 2A). Region E, which contained the

disrupted unstable segment of ORF1a, was introduced at the last cloning step to minimize the toxicity problem usually observed with this region of the genome (6, 21). Junction sites created for each new construct were sequenced to make sure no unexpected mutations were introduced into the genome.

Recovery, replication, and neurovirulence of recombinant HCoV-OC43. One of the major advantage provided by the BAC system is that the recovery of recombinant viral particles is very efficient and simple. This was especially true for pBAC-OC43^{FL} given that the cell line BHK-21 provided a good transfection rate and was also susceptible to infection by HCoV-

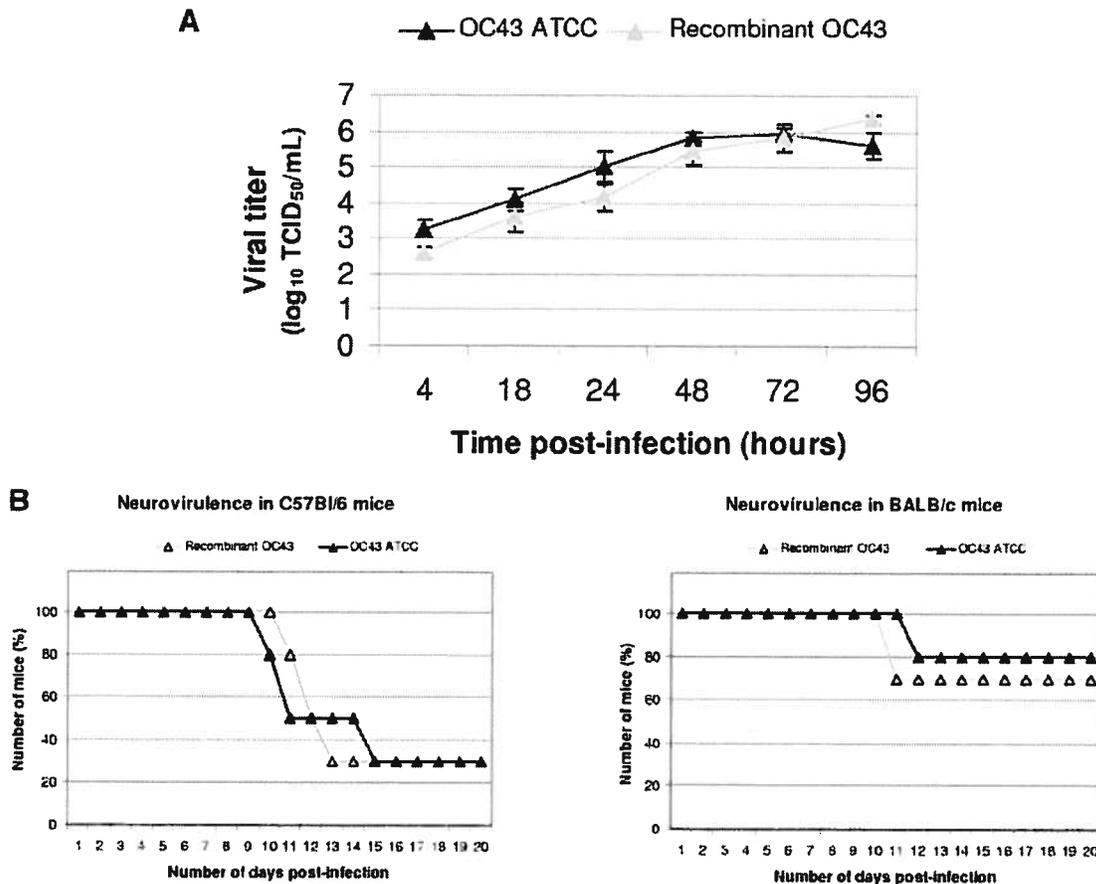


FIG. 3. In vitro and in vivo properties of the recombinant virus. (A) Viral growth kinetics of both the HCoV-OC43 ATCC strain and the recombinant HCoV-OC43 virus. Culture supernatants from HRT-18-infected cells were harvested at the indicated times postinfection. Infectious titers were determined using the immunoperoxidase assay and are expressed in log TCID₅₀/ml. Both viruses displayed similar patterns of replication. (B) Neurovirulence of the recombinant HCoV-OC43 in mice. Two different strains of MHV-seronegative female mice, BALB/c and C57BL/6, were inoculated with the same dose of virus.

OC43. For rescue of recombinant virus, 5 μ g of pBAC-OC43^{FL} were directly transfected into BHK-21 cells using Lipofectamine 2000 (Invitrogen). Virus titers obtained for the recombinant virus reached $10^{6.5}$ 50% tissue culture infective doses (TCID₅₀)/ml, which are comparable to those obtained after infection with the HCoV-OC43 laboratory strain. To make sure rescued viruses were indeed recombinant particles, reverse transcription-PCR were performed and the presence of the newly introduced restriction sites, which are exclusive to the recombinant virus, was assessed (data not shown).

To test whether the recombinant virus displayed the same growth kinetics as the parental strain in HRT-18 cells, replication was assessed at different times postinfection in a multi-cycle growth curve. Wild-type and recombinant virus shared very similar overall patterns of replication, reaching virus titers of approximately $10^{6.0}$ TCID₅₀/ml after 4 days of infection, although the replication of the recombinant virus was slightly delayed during the first 2 days (Fig. 3A). As expected and like the wild-type HCoV-OC43 strain, recombinant virus did not form any plaques (data not shown). Taken together, these data

confirm that the recombinant virus shares phenotypic properties with the wild type HCoV-OC43 strain in cell culture.

To determine whether the HCoV-OC43 strain, as well as the recombinant virus, displayed the same phenotype in vivo, their neurovirulence was assessed in mice. C57BL/6 and BALB/c mice were infected intracerebrally with either virus (8). In C57BL/6 mice, a survival rate of 30% was achieved for both viruses, whereas 70 and 80% of BALB/c mice survived the infection with the recombinant virus and the parent strain, respectively (Fig. 3B). Both recombinant and wild-type virus displayed very similar levels of virulence following injection into the brain. Furthermore, both viruses showed very similar patterns of replication in mouse brain (data not shown).

Overall, the virus recovered from the infectious cDNA clone exhibited the same genotypic and phenotypic properties as the HCoV-OC43 from ATCC, both in vitro and in vivo.

The cDNA infectious clones represent very interesting tools for the study of potential vaccine candidates, like genetically attenuated viruses, but also allow multiple possibilities for the characterization of any viral functions and mechanisms by re-

verse genetics. The pBAC-OC43^{FL} will be invaluable in the characterization of the adaptation of the HCoV-OC43 virus in a neural environment and will provide an opportunity to better understand the biology of the HCoV-OC43, as well as other related coronaviruses, such as SARS-CoV.

We are grateful to Francine Lambert for technical assistance with cell and virus culture. We thank R. C. Lévesque (Université Laval, Québec, Canada) for providing vector pBeloBAC11 and Luc Gagnon (Ecopia Biosciences, Inc., Montréal, Québec, Canada) for his valuable advice regarding cloning techniques in the BAC vector.

This work was supported by grant MT-9203 from Canadian Institutes for Health Research (Institute of Infection and Immunity) to P.J.T., who is the holder of a Tier-1 Canada Research Chair, and by a grant from the Department of Education and Science of Spain (BIO2004-00636) to L.E. J.R.S.-J. acknowledges a studentship from the Fonds Québécois de Recherche sur la Nature et les Technologies.

REFERENCES

- Almazán, F., C. Galán, and L. Enjuanes. 2004. The nucleoprotein is required for efficient coronavirus genome replication. *J. Virol.* **78**:12683–12688.
- Almazán, F., J. M. González, Z. Péntzes, A. Izeta, E. Calvo, J. Plana-Durán, and L. Enjuanes. 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. USA* **97**:5516–5521.
- Arbour, N., R. Day, J. Newcombe, and P. J. Talbot. 2000. Neuroinvasion by human respiratory coronaviruses. *J. Virol.* **74**:8913–8921.
- Casais, R., V. Thiel, S. G. Siddell, D. Cavanagh, and P. Britton. 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J. Virol.* **75**:12359–12369.
- Coley, S. E., E. Lavi, S. G. Sawicki, L. Fu, B. Schelle, N. Karl, S. G. Siddell, and V. Thiel. 2005. Recombinant mouse hepatitis virus strain A59 from cloned, full-length cDNA replicates to high titers in vitro and is fully pathogenic in vivo. *J. Virol.* **79**:3097–3106.
- Gonzalez, J. M., Z. Péntzes, F. Almazán, E. Calvo, and L. Enjuanes. 2002. Stabilization of a full-length infectious cDNA clone of transmissible gastroenteritis coronavirus by insertion of an intron. *J. Virol.* **76**:4655–4661.
- Gorbalenya, A. E., E. J. Snijder, and W. J. Spaan. 2004. Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. *J. Virol.* **78**:7863–7866.
- Jacomy, H., and P. J. Talbot. 2003. Vacuolating encephalitis in mice infected by human coronavirus OC43. *Virology* **315**:20–33.
- Masters, P. S. 1999. Reverse genetics of the largest RNA viruses. *Adv. Virus Res.* **53**:245–264.
- Myint, S. H. 1994. Human coronaviruses: a brief review. *Rev. Med. Virol.* **4**:35–46.
- Peiris, J. S., C. M. Chu, V. C. Cheng, K. S. Chan, I. F. Hung, L. L. Poon, K. I. Law, B. S. Tang, T. Y. Hon, C. S. Chan, K. H. Chan, J. S. Ng, B. J. Zheng, W. L. Ng, R. W. Lai, Y. Guan, and K. Y. Yuen. 2003. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**:1767–1772.
- Sharmeen, L., M. Y. P. Kuo, G. Dinter-Gottlieb, and J. Taylor. 1988. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J. Virol.* **62**:2674–2679.
- Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**:8794–8797.
- Snijder, E. J., P. J. Bredenbeck, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. M. Spaan, and A. E. Gorbalenya. 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **331**:991–1004.
- Sola, I., S. Alonso, S. Zúñiga, M. Balasch, J. Plana-Durán, and L. Enjuanes. 2003. Engineering the transmissible gastroenteritis virus genome as an expression vector inducing lactogenic immunity. *J. Virol.* **77**:4357–4369.
- Sola, I., J. L. Moreno, S. Zúñiga, S. Alonso, and L. Enjuanes. 2005. Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J. Virol.* **79**:2506–2516.
- St-Jean, J. R., H. Jacomy, M. Desforges, A. Vabret, F. Freymuth, and P. J. Talbot. 2004. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J. Virol.* **78**:8824–8834.
- Thiel, V., J. Herold, B. Schelle, and S. G. Siddell. 2001. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.* **82**:1273–1281.
- Yount, B., K. M. Curtis, and R. S. Baric. 2000. Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model. *J. Virol.* **74**:10600–10611.
- Yount, B., K. M. Curtis, E. A. Fritz, L. E. Hensley, P. B. Jahrling, E. Prentice, M. R. Denison, T. W. Geisbert, and R. S. Baric. 2003. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. USA* **100**:12995–13000.
- Yount, B., M. R. Denison, S. R. Weiss, and R. S. Baric. 2002. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J. Virol.* **76**:11065–11078.
- Zúñiga, S., I. Sola, S. Alonso, and L. Enjuanes. 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J. Virol.* **78**:980–994.