UNIVERSITE D'AIX-MARSEILLE

Sciences de la Vie et de la Santé

Unité de Recherche sur les Maladies Infectieuses et Tropicales Émergentes (URMITE)

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Pathologie Humaine
Spécialité : Maladies infectieuses

Haitham SOBHY

Characterization of proteins involved in the fibers of Mimivirus

Soutenue le 26/09/2014 devant le jury :

Pr. Bruno POZZETTO Rapporteur

Pr. Patrice MORAND Rapporteur

Pr. Pierre-Edouard FOURNIER Examinateur

Pr. Philippe COLSON Directeur de thès

Dedication

I would like to dedicate this work to soul of my father and Mister Ali El-Hosary, I wish them peace and merci. I also like to dedicate the work to my mother Shadia, Madam Elvie El-Hosary, Shaimaa Sobhy, Waleed Sobhy, Khalid Sobhy, Nabil Ali, Mohsen Ali, Reham Mohsen and my friends.

Avant Propos

Le format de présentation de cette thèse correspond à une recommandation de la spécialité Maladies Infectieuses et Microbiologie, à l'intérieur du Master de Sciences de la Vie et de la Santé qui dépend de l'Ecole Doctorale des Sciences de la Vie de Marseille. Le candidat est amené à respecter des règles qui lui sont imposées et qui comportent un format de thèse utilisé dans le Nord de l'Europe permettant un meilleur rangement que les thèses traditionnelles. Par ailleurs, la partie introduction et bibliographie est remplacée par une revue envoyée dans un journal afin de permettre une évaluation extérieure de la qualité de la revue et de permettre à l'étudiant de le commencer le plus tôt possible une bibliographie exhaustive sur le domaine de cette thèse. Par ailleurs, la thèse est présentée sur article publié, accepté ou soumis associé d'un bref commentaire donnant le sens général du travail. Cette forme de présentation a paru plus en adéquation avec les exigences de la compétition internationale et permet de se concentrer sur des travaux qui bénéficieront d'une diffusion internationale.

Professeur Didier RAOULT

Résumé

Les virus géants sont un groupe de virus ADN double brin caractérisés par une taille géante du virion et du génome, et un répertoire de gènes qui comprend environ 450 à 2500 gènes prédits. Une proportion importante de ces gènes (jusqu'à 93%) sont des 'ORFans', ou codent pour des protéines de fonction inconnue. Acanthamoeba polyphaga mimivirus est le premier virus géant découvert, il y a une décennie, par co-culture sur Acanthamoeba spp. Il est le membre prototype de la famille Mimiviridae. Le génome de Mimivirus code pour environ 1000 protéines, parmi lesquelles ~50% n'ont pas d'homologue connu dans les banques de séquences publiques. La capside de Mimivirus a un diamètre d'environ 500 nm et est couverte par une couche dense de fibres, à l'exception de l'un de ses sommets. Ces fibres sont d'environ 130 nm de longueur et se composent d'une tige souple et d'une tête de forme globulaire.

Dans ce travail de thèse, nous avons cherché à étudier les gènes impliqués dans la formation des fibres de Mimivirus. Dans ce but, nous avons notamment exprimé des gènes candidats dans E. coli, et nous avons mis au point une stratégie qui a utilisé l'interférence ARN afin d'étudier la fonction et la structure des protéines de Mimivirus. Nous avons annoté quatre protéines associées aux fibres. La stratégie utilisant les petits ARN interférant appliquée ici est originale et a été utilisée pour la première fois pour les virus géants qui infectent les amibes. Elle pourrait permettre de décrypter la fonction des gènes des mimivirus et d'annoter potentiellement des centaines de protéines présentes dans les bases de données publiques, et de différencier l'ADN poubelle des gènes réellement utilisés.

Mots clés : Acanthamoeba, virus géant, Mimivirus, nucleocytoplasmic large DNA virus, Megavirales, protéines associées aux fibres, interférence par ARN, virophage, entré des virus

Abstract

Giant viruses are a group of double stranded DNA viruses that are characterized by a giant virion and genome size, and gene repertoires encompassing approximately 450 to 2500 predicted genes. A substantial proportion of these genes (up to 93%) consists in ORFans, or encodes proteins with unknown functions. Acanthamoeba polyphaga mimivirus is the first giant virus that was discovered, a decade ago, after co-culturing on Acanthamoeba spp. It is the prototype member of the family Mimiviridae. Mimivirus encodes about 1000 proteins, among which ~50% have no known homolog in public sequence databases. The Mimivirus capsid is about 500 nm in diameter and is covered by a dense layer of fibers, except at one of its vertices. These fibers are about 130 nm in length and consist of a soft shaft and a globular shaped head.

In this thesis work, we aimed to study the genes involved in the formation of the Mimivirus fibers. For this purpose, we have expressed candidate genes in E. coli, and implemented a strategy that used RNA interference to study the function and structure of Mimivirus proteins. We then succeeded in annotating four proteins as fiber associated proteins. The short interfering RNA strategy that we applied here is original and has been used for the first time in giant viruses that infect amoeba. It could allow deciphering the function of the mimivirus gene repertoires and help annotating hundreds of proteins without known function found in public databases and differentiate between junk DNA and truly used genes.

Keywords: Acanthamoeba, giant virus, Mimivirus, nucleocytoplasmic large DNA virus, Megavirales, fiber associated proteins, RNA interference, virophage, virus entry

Remerciements - Acknowledgements

During the past three years, I was surrounded by kind, sympathetic and warmhearted persons. Their support started since the first day of my doctoral study. The sympathy and support that I got from my family and friends were the stimulus to advance in my work. The great support from the professors and colleagues was so generous to me.

I am grateful to “Fondation Infectiopole Sud” for their generous financial support. I like also to thank Aix-Marseille University, the doctoral school and URMITE for their kind hospitality. I like to express the gratitude to Professor Colson and Professor Raoult for their great support and advices during my research work. The discussions with them were so useful and added new experiences to me. I am grateful to Prof. La Scola and Doctor Isabelle Pagnier for their support on Mimivirus and electron microscopy work. I am so grateful for the generous and outstanding technical support from Misters and Doctors Said Azza, Claude Nappez, Guillaume Gotthard, Christophe Flaudrops and Jean-Pierre Baudoin; I like to thank them from my heart. I am thankful to Madame Micheline Pitaccolo, Madame Valérie Filosa, Madame Francine Verin and the administrative staffs; I appreciate their endless assistance. I am very grateful to Audrey Ferretti, Audrey Borg, Audrey Averna and all personnel of URMITE, including the technicians and the engineers for their help.

I would like to thank the members of the thesis committee, Professor Bruno Pozzetto, Professor Patrice Morand and Professor Pierre-Edouard Fournier, for accepting to review the thesis. Their comments were so valuable and helped to improve the thesis.

I am so grateful to my friends and colleagues Nisserine Ben Amara, Amira Ben Amara, Mondher Boughalmi, Hanène Saadi, Mahfoud Bakli, Erwin Sentausa, Fanny Balique, Tatsiana Ngounga and Niyaz Yoosuf for their generous helps and support.

Table of contents

Background and objectives

Acanthamoeba polyphaga mimivirus (APMV or Mimivirus) has been discovered in 2003 from water collected in a cooling tower in Bradford, UK, in 1992 (Raoult, Audic et al. 2004). Mimivirus was the largest known virus at the time of its discovery. Interestingly, antibodies to Mimivirus were further detected in several serological studies in pneumonia patients (La Scola, Marrie et al. 2005, Berger, Papazian et al. 2006, Dare, Chittaganpitch et al. 2008, Vincent, La Scola et al. 2010, Parola, Renvoise et al. 2012, Vanspauwen, Franssen et al. 2012, Bousbia, Papazian et al. 2013, Luyt, Brechot et al. 2014). Subsequent studies, including the recent isolation of Mimivirus close relatives from patients with unexplained pneumonia (Saadi, Pagnier et al. 2013, Saadi, Reteno et al. 2013) showed that Mimivirus was a potential agent of pneumonia (reviewed in (Colson, La Scola et al. 2013)). Noteworthy, although Mimivirus was first isolated using co-culturing on *Acanthamoeba polyphaga*, for which it carries this name, it remains unclear if this virus has other hosts, and this deserves to be elucidated by further studies. Isolation of Mimivirus has opened the door to the isolation of dozens of new giant viruses, which are closely related to Mimivirus, mostly from environmental samples, consisting in marine or freshwater or soil samples, also from an insect and, more recently, from humans by using this strategy of co-culture with *Acanthamoeba polyphaga* or *A. castellanii* (Boughalmi, Saadi et al. 2013, Pagnier, Reteno et al. 2013). Distantly related mimiviruses were also isolated from marine flagellates (Fischer, Allen et al. 2010, Santini, Jeudy et al. 2013). Mimiviruses and closely related viruses are currently classified into a new virus family, so-called *Mimiviridae*. Due to their relatively large genome and virion particle size, mimiviruses (members of family *Mimiviridae*) are known as giant viruses. Mimiviruses are linked by a small set of genes to phycodnaviruses and other large DNA viruses that compose the nucleocytoplasmic large DNA viruses (NCLDVs), which were recently proposed to be reclassified into a new viral order named the "*Megavirales*" (Iyer, Aravind et al. 2001, Colson, de Lamballerie et al. 2012, Colson, De Lamballerie et al. 2013). The Mimivirus capsid is about 500 nm in diameter and is covered by a dense layer of fibers, except at one of its vertices that harbors a structure termed the stargate, which is thought to be involved in DNA release (Raoult, Audic et al. 2004, Kuznetsov, Xiao et al. 2010). These viral fibers are about 125-140 nm in length and about 1.4 nm in diameter and consist of a soft shaft, and a globular shaped head (Xiao, Kuznetsov et al. 2009, Klose, Kuznetsov et al. 2010,

Kuznetsov, Xiao et al. 2010). The fibers are highly glycosylated, antigenic, and resistant to protease and collagenase treatment (Xiao, Kuznetsov et al. 2009). Clusters of 3-4 fibers were found to be linked via a disc-shaped base (Kuznetsov, Xiao et al. 2010). Interestingly, sub-culturing mimiviruses 150 times on germ-free amoebae led to the emergence of viruses lacking fibers, i.e., the so-called Mimivirus M4 strain (Boyer, Azza et al. 2011), which exhibited a 16% reduction in the size of its genome and a loss of about 150 genes, including fiber-associated proteins. Concurrently, the M4 strain showed reduced infectivity for *Acanthamoeba* spp. compared with wild-type mimiviruses, which suggested the role of Mimivirus fibers in entry into host cells.

The *Megavirus chilensis* has currently the largest virion particle and genome sizes amongst mimiviruses (Arslan, Legendre et al. 2011). Megavirus LBA111 virus, which is the largest known mimivirus isolated from a human, has a genome size of about 1.2 million base pair (Mbp) and is closely related to *Megavirus chilensis*; it was isolated from the bronchoalveolar lavage of a pneumonia patient (Saadi, Pagnier et al. 2013), while Shan virus was isolated from the stools of another pneumonia patient (Saadi, Reteno et al. 2013).

Marseilleviruses are other giant viruses of amoeba, with similar virion particle topology, but smaller capsid and genome compared to mimiviruses, and were recently proposed to compose a new viral family, the *Marseilleviridae* (Colson, Pagnier et al. 2013). Marseillevirus was the first *Marseilleviridae* member isolated (Boyer, Yutin et al. 2009). It exhibit about 12 nm-long fibers at its surface. Lausannevirus is another member of marseilleviruses also isolated by culturing on amoeba (Thomas, Bertelli et al. 2011). The presence of antibodies to Marseillevirus and Lausannevirus were reported in healthy humans and a patient with adenitis (Mueller, Baud et al. 2013, Popgeorgiev, Boyer et al. 2013, Popgeorgiev, Michel et al. 2013).

Recently, three giant viruses, *Pandoravirus salinus*, *Pandoravirus dulcis* and *Pithovirus sibericum*, which represent two new putative viral families, were isolated by co-culture with *Acanthamoeba* spp. (Raoult, Audic et al. 2004, Arslan, Legendre et al. 2011, Philippe, Legendre et al. 2013, Saadi, Pagnier et al. 2013, Legendre, Bartoli et al. 2014). The *Pandoravirus* spp. have capsid particle size >700 nm (Philippe, Legendre et al. 2013). Further phylogenetic analyses described them to belong to the family *Phycodnaviridae* (Yutin and Koonin 2013). The virus *Pithovirus sibericum* has the largest capsid particle, about 1.5 µM, but its genome is only 600 kilobase pairs (kbp) (Legendre, Bartoli et al. 2014). Although, the topology of this

virus resembles those of *Pandoravirus*, gene content and phylogenetic analysis reveals that the virus was closely related to iridoviruses and marseilleviruses.

Virophages, which are viruses that infect mimiviruses within their amoebal host, have a double-stranded DNA that is 18-19 kbp-long and encodes about 20 proteins (La Scola, Desnues et al. 2008, Krupovic and Cvirkaite-Krupovic 2011). Sputnik was the first virophage isolated with mimivirus and was observed attached to their fibers (La Scola, Desnues et al. 2008). Afterward, several other virophages have been isolated; including Mavirus, which was isolated with *Cafeteria roenbergensis* virus (Fischer and Suttle 2011); Organic Lake virophage, which was isolated with Organic Lake phycodnaviruses (distantly related mimiviruses, albeit first classified in family *Phycodnaviridae*) (Yau, Lauro et al. 2011); and Zamilon, which was isolated with Mont1 virus from Tunisian soil (Gaia, Benamar et al. 2014). Structurally, Sputnik has mushroom-like protrusions at its surface, which are attached to capsomers and could recognize the viral host (Sun, La Scola et al. 2010, Zhang, Sun et al. 2012). Antibodies to the Sputnik virophage were detected in human blood (Parola, Renvoise et al. 2012). Noteworthy, Mimivirus M4 strain resists virophage infection (Boyer, Azza et al. 2011).

Giant viruses of amoeba (including, mimiviruses and the other recently isolated giant viruses of amoeba) are characterized by their large genome size, if compared to those of other viruses, which is comparable to the genome size of small bacteria. Due to their huge genome size, giant viruses encode hundreds of proteins (approximately 450 to 2500). Mimivirus genomes encode about 1000 proteins, whereas, *P. salinus* and *P. dulcis* genomes contain 2.5 Mbp and 1.9 Mbp and encode about 2500 and 1500 putative proteins, respectively (Philippe, Legendre et al. 2013). Although, *P. sibericum* has a still larger particle size, its genome is about 600 kbp and encodes only 467 proteins (Legendre, Bartoli et al. 2014). Amongst the tremendous number of proteins encoded by these giant viruses, some share sequence homology with those encoded by members of the cellular domains of life, including, bacteria, archaea and eukaryotes. Giant viruses encode several proteins involved in translation. Additionally, giant virus genomes contain many ORFans and encode number of hypothetical or putative proteins, which have not been annotated yet and whose function is not known (Boyer, Gimenez et al. 2013). This can be observed in Mimivirus, in which more than half of the genes are ORFans or encode hypothetical proteins, and pandoraviruses in which 93% of the gene repertoire is composed by ORFans.

On the other hand, the surface topology of giant viruses is distinctive from those of other viruses, in particular the outer fibers present at the surface of Mimvirus. Some viruses harbor

external protrusions on the virion surface. These protrusions are either short and rigid, such as spikes, or long, forming a tail, such as those attached to the head of bacteriophages. These protrusions are usually found at the vertices of the capsid. Unlike other viruses, Mimivirus is covered by long and flexible fibers. These fibers resemble those of some organisms that belong to the cellular domains of life. In this thesis work, we aimed to study the function and structure of the proteins involved in the formation of the Mimivirus fibers. We gained a better knowledge on the function of four proteins involved in the formation of the Mimivirus fibers. Moreover, we aimed to express these proteins in E. coli and study their structure.

References

Arslan, D., M. Legendre, V. Seltzer, C. Abergel and J. M. Claverie (2011). "Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae." Proc Natl Acad Sci U S A **108**(42): 17486-17491.

Berger, P., L. Papazian, M. Drancourt, B. La Scola, J. P. Auffray and D. Raoult (2006). "Ameba-associated microorganisms and diagnosis of nosocomial pneumonia." Emerg Infect Dis **12**(2): 248-255.

Boughalmi, M., H. Saadi, I. Pagnier, P. Colson, G. Fournous, D. Raoult and B. La Scola (2013). "High-throughput isolation of giant viruses of the *Mimiviridae* and Marseilleviridae families in the Tunisian environment." Environ Microbiol **15**(7): 2000-2007.

Bousbia, S., L. Papazian, P. Saux, J. M. Forel, J. P. Auffray, C. Martin, D. Raoult and B. La Scola (2013). "Serologic prevalence of amoeba-associated microorganisms in intensive care unit pneumonia patients." PLoS One **8**(3): e58111.

Boyer, M., S. Azza, L. Barrassi, T. Klose, A. Campocasso, I. Pagnier, G. Fournous, A. Borg, C. Robert, X. Zhang, C. Desnues, B. Henrissat, M. G. Rossmann, B. La Scola and D. Raoult (2011). "Mimivirus shows dramatic genome reduction after intraamoebal culture." Proc Natl Acad Sci U S A **108**(25): 10296-10301.

Boyer, M., G. Gimenez, M. Suzan-Monti and D. Raoult (2013). "Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses." Intervirology **53**(5): 310-320.

Boyer, M., N. Yutin, I. Pagnier, L. Barrassi, G. Fournous, L. Espinosa, C. Robert, S. Azza, S. Sun, M. G. Rossmann, M. Suzan-Monti, B. La Scola, E. V. Koonin and D. Raoult (2009). "Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms." Proc Natl Acad Sci U S A **106**(51): 21848-21853.

Colson, P., X. de Lamballerie, G. Fournous and D. Raoult (2012). "Reclassification of giant viruses composing a fourth domain of life in the new order *Megavirales*." Intervirology **55**(5): 321-332.

Colson, P., X. De Lamballerie, N. Yutin, S. Asgari, Y. Bigot, D. K. Bideshi, X. W. Cheng, B. A. Federici, J. L. Van Etten, E. V. Koonin, B. La Scola and D. Raoult (2013). ""*Megavirales*", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses." Arch Virol **158**(12): 2517-2521.

Colson, P., B. La Scola and D. Raoult (2013). "Giant viruses of amoebae as potential human pathogens." Intervirology **56**(6): 376-385.

Colson, P., I. Pagnier, N. Yoosuf, G. Fournous, B. La Scola and D. Raoult (2013). ""Marseilleviridae", a new family of giant viruses infecting amoebae." Arch Virol **158**(4): 915-920.

Dare, R. K., M. Chittaganpitch and D. D. Erdman (2008). "Screening pneumonia patients for mimivirus." Emerg Infect Dis **14**(3): 465-467.

Fischer, M. G., M. J. Allen, W. H. Wilson and C. A. Suttle (2010). "Giant virus with a remarkable complement of genes infects marine zooplankton." Proc Natl Acad Sci U S A **107**(45): 19508-19513.

Fischer, M. G. and C. A. Suttle (2011). "A virophage at the origin of large DNA transposons." Science **332**(6026): 231-234.

Gaia, M., S. Benamar, M. Boughalmi, I. Pagnier, O. Croce, P. Colson, D. Raoult and B. La Scola (2014). "Zamilon, a novel virophage with *Mimiviridae* host specificity." PLoS One **9**(4): e94923.

Iyer, L. M., L. Aravind and E. V. Koonin (2001). "Common origin of four diverse families of large eukaryotic DNA viruses." J Virol **75**(23): 11720-11734.

Klose, T., Y. G. Kuznetsov, C. Xiao, S. Sun, A. McPherson and M. G. Rossmann (2010). "The three-dimensional structure of Mimivirus." Intervirology **53**(5): 268-273.

Krupovic, M. and V. Cvirkaite-Krupovic (2011). "Virophages or satellite viruses?" Nat Rev Microbiol **9**(11): 762-763.

Kuznetsov, Y. G., C. Xiao, S. Sun, D. Raoult, M. Rossmann and A. McPherson (2010). "Atomic force microscopy investigation of the giant mimivirus." Virology **404**(1): 127-137.

La Scola, B., C. Desnues, I. Pagnier, C. Robert, L. Barrassi, G. Fournous, M. Merchat, M. Suzan-Monti, P. Forterre, E. Koonin and D. Raoult (2008). "The virophage as a unique parasite of the giant mimivirus." Nature **455**(7209): 100-104.

La Scola, B., T. J. Marrie, J. P. Auffray and D. Raoult (2005). "Mimivirus in pneumonia patients." Emerg Infect Dis **11**(3): 449-452.

Legendre, M., J. Bartoli, L. Shmakova, S. Jeudy, K. Labadie, A. Adrait, M. Lescot, O. Poirot, L. Bertaux, C. Bruley, Y. Coute, E. Rivkina, C. Abergel and J. M. Claverie (2014). "Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology." Proc Natl Acad Sci U S A **2014**: 3.

Luyt, C. E., N. Brechot and J. Chastre (2014). "What role do viruses play in nosocomial pneumonia?" Curr Opin Infect Dis **27**(2): 194-199.

Mueller, L., D. Baud, C. Bertelli and G. Greub (2013). "Lausannevirus seroprevalence among asymptomatic young adults." Intervirology **56**(6): 430-433.

Pagnier, I., D. G. Reteno, H. Saadi, M. Boughalmi, M. Gaia, M. Slimani, T. Ngounga, M. Bekliz, P. Colson, D. Raoult and B. La Scola (2013). "A decade of improvements in *Mimiviridae* and Marseilleviridae isolation from amoeba." Intervirology **56**(6): 354-363.

Parola, P., A. Renvoise, E. Botelho-Nevers, B. La Scola, C. Desnues and D. Raoult (2012). "*Acanthamoeba polyphaga* mimivirus Virophage Seroconversion in Travelers Returning from Laos." Emerg Infect Dis **18**(9): 1500-1502.

Philippe, N., M. Legendre, G. Doutre, Y. Coute, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J. M. Claverie and C. Abergel (2013). "Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes." Science **341**(6143): 281-286.

Popgeorgiev, N., M. Boyer, L. Fancello, S. Monteil, C. Robert, R. Rivet, C. Nappez, S. Azza, J. Chiaroni, D. Raoult and C. Desnues (2013). "Marseillevirus-like virus recovered from blood donated by asymptomatic humans." J Infect Dis **208**(7): 1042-1050.

Popgeorgiev, N., G. Michel, H. Lepidi, D. Raoult and C. Desnues (2013). "Marseillevirus Adenitis in an Eleven-Month-Old Child." Journal of Clinical Microbiology

Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan and J. M. Claverie (2004). "The 1.2-megabase genome sequence of Mimivirus." Science **306**(5700): 1344-1350.

Saadi, H., I. Pagnier, P. Colson, J. K. Cherif, M. Beji, M. Boughalmi, S. Azza, N. Armstrong, C. Robert, G. Fournous, B. La Scola and D. Raoult (2013). "First isolation of mimivirus in a patient with pneumonia." Clin Infect Dis **57**(4): e127-134.

Saadi, H., D. G. Reteno, P. Colson, S. Aherfi, P. Minodier, I. Pagnier, D. Raoult and B. La Scola (2013). "Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia." Intervirology **56**(6): 424-429.

Santini, S., S. Jeudy, J. Bartoli, O. Poirot, M. Lescot, C. Abergel, V. Barbe, K. E. Wommack, A. A. Noordeloos, C. P. Brussaard and J. M. Claverie (2013). "Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes." Proc Natl Acad Sci U S A **110**(26): 10800-10805.

Sun, S., B. La Scola, V. D. Bowman, C. M. Ryan, J. P. Whitelegge, D. Raoult and M. G. Rossmann (2010). "Structural studies of the Sputnik virophage." J Virol **84**(2): 894-897.

Thomas, V., C. Bertelli, F. Collyn, N. Casson, A. Telenti, A. Goesmann, A. Croxatto and G. Greub (2011). "Lausannevirus, a giant amoebal virus encoding histone doublets." Environ Microbiol **13**(6): 1454-1466.

Vanspauwen, M. J., F. M. Franssen, D. Raoult, E. F. Wouters, C. A. Bruggeman and C. F. Linssen (2012). "Infections with mimivirus in patients with chronic obstructive pulmonary disease." Respir Med **106**(12): 1690-1694.

Vincent, A., B. La Scola and L. Papazian (2010). "Advances in Mimivirus pathogenicity." Intervirology **53**(5): 304-309.

Xiao, C., Y. G. Kuznetsov, S. Sun, S. L. Hafenstein, V. A. Kostyuchenko, P. R. Chipman, M. Suzan-Monti, D. Raoult, A. McPherson and M. G. Rossmann (2009). "Structural studies of the giant mimivirus." PLoS Biol **7**(4): e92.

Yau, S., F. M. Lauro, M. Z. DeMaere, M. V. Brown, T. Thomas, M. J. Raftery, C. Andrews-Pfannkoch, M. Lewis, J. M. Hoffman, J. A. Gibson and R. Cavicchioli (2011). "Virophage control of antarctic algal host-virus dynamics." Proc Natl Acad Sci U S A **108**(15): 6163-6168.

Yutin, N. and E. V. Koonin (2013). "Pandoraviruses are highly derived phycodnaviruses." Biol Direct **8**: 25.

Zhang, X., S. Sun, Y. Xiang, J. Wong, T. Klose, D. Raoult and M. G. Rossmann (2012). "Structure of Sputnik, a virophage, at 3.5-A resolution." Proc Natl Acad Sci U S A **109**(45): 18431-18436.

Chapter 1

Review of Entry pathways utilized by Nucleocytoplasmic large DNA viruses and role of fibers

Introduction

Virus adopted various mechanisms and molecules to infect their host cell. This infection includes three main steps with, notably, entry into the host cell, delivery of the genome at the site of genome replication, and a final step of virus egress from the cell. Receptors at the cell surface facilitate the binding and attachment of the virus to the cell. Virus surface proteins, including glycoproteins or fiber proteins, also facilitate the viral attachment. In adenovirus, as an example, fiber proteins have major role in virus binding and attachment to cell receptors (Cusack 2005, Zhang and Bergelson 2005, Russell 2009, Cupelli and Stehle 2011, Arnberg 2012). This chapter aims to briefly review the entry mechanisms used by members of the Nucleocytoplasmic large DNA viruses (proposed order *Megavirales*), which include mimiviruses and other giant viruses that infect *Acanthamoeba*.

References

Arnberg, N. (2012). "Adenovirus receptors: implications for targeting of viral vectors." Trends Pharmacol Sci **33**(8): 442-448.

Cupelli, K. and T. Stehle (2011). "Viral attachment strategies: the many faces of adenoviruses." Curr Opin Virol **1**(2): 84-91.

Cusack, S. (2005). "Adenovirus complex structures." Curr Opin Struct Biol **15**(2): 237-243.

Russell, W. C. (2009). "Adenoviruses: update on structure and function." J Gen Virol **90**(Pt 1): 1-20.

Zhang, Y. and J. M. Bergelson (2005). "Adenovirus receptors." J Virol **79**(19): 12125-12131.

Review of the entry pathways for giant and large *Megavirales* viruses

Abstract

Nucleocytoplasmic large DNA viruses (NCLDVs) (proposed order *Megavirales*) members are large and giant double-stranded viruses that encompass at least seven families and new putative families including families *Asfarviridae*, *Poxviridae, Ascoviridae*, *Iridoviridae*, *Phycodnaviridae*, *Mimiviridae*, and *Marseilleviridae*, as well as pandoraviruses and *Pithovirus sibericum*. Amongst these viruses, so-called giant viruses have a diameter and genome size that are larger than those of small bacteria, and they considerably expand the diversity of the viral world. Megaviruses infect a wide-range of hosts, including humans, other mammals, fishes, invertebrates, protozoa and algae. Some are a concern for human health, or livestock and aquaculture industries. Virophages are small viruses that infect mimiviruses while co-infecting their phagocytic protozoan host. We provide here a short review of entry pathways recruited by these viruses.

1. Introduction

Nucleocytoplasmic large DNA viruses (NCLDVs) or members of the proposed order *Megavirales* are double stranded (ds) giant and large DNA viruses that encompass at least seven families and new putative families including families *Asfarviridae*, *Poxviridae, Ascoviridae*, *Iridoviridae*, *Phycodnaviridae*, *Mimiviridae*, and *Marseilleviridae*, as well as pandoraviruses and *Pithovirus sibericum* (Iyer, Balaji et al. 2006, Colson, de Lamballerie et al. 2012, Philippe, Legendre et al. 2013, Yutin and Koonin 2013, Legendre, Bartoli et al. 2014). The NCLDVs were first described in 2001 (Iyer, Aravind et al. 2001). Then, during the past decade, giant viruses were linked to them including mimiviruses and marseilleviruses that were found to compose new families, and, more recently, pandoraviruses and *P. sibericum*. In addition, virophage, which are viruses of viruses, were discovered in association with mimiviruses, co-infected their protistan host (Parola, Renvoise et al. 2012).

The *Megavirales* members encompass genome size ranging from 0.1 to 2.5 megabase pairs (Mbp) and encode up to 2500 proteins (Colson, de Lamballerie et al. 2012, Philippe, Legendre et al. 2013, Legendre, Bartoli et al. 2014). Giant viruses that infect phagocytic protists are the largest viruses known so far. Overall, the megaviruses have a considerable interest in evolutionary biology and they substantially expand the definition of viruses (Raoult, Audic et al. 2004, Moreira and Brochier-Armanet 2008, Raoult and Forterre 2008, Boyer, Yutin et al. 2009, Forterre 2010). They have a gene content that include genes that in some cases are unique amongst viruses and have been considered as the trademark of cellular organisms. Moreover, their genome is composed of many ORFan genes, genes without homolog in public sequence databases, or genes without annotated functions, so-called hypothetical proteins (Raoult, Audic et al. 2004, Yutin, Wolf et al. 2009, Boyer, Gimenez et al. 2010, Fischer, Allen et al. 2010, Philippe, Legendre et al. 2013, Legendre, Bartoli et al. 2014).

These megaviruses infect a wide-range of hosts including humans, non human primates, other mammals, fish, amphibians, invertebrates, even protozoa and algae. They can be a concern for public health, livestock and aquacultures industries, which also emphasizes their biological and economic importance. These viruses replicate mostly in the nucleus of the host cells, whereas strictly cytoplasmic replication is also reported for some of them (Mutsafi, Zauberman et al. 2010). Here, we briefly review current knowledge on entry pathways recruited by the megaviruses and the mimivirus virophages to initiate the replicative cycle.

2. Main characteristics, features and taxonomy of *Megavirales*

The members of NCLDV are classified into at least seven families *Asfarviridae*, *Poxviridae, Ascoviridae*, *Iridoviridae*, *Phycodnaviridae*, *Mimiviridae*, and *Marseilleviridae* (proposed family for marseilleviruses) as well as, the recently discovered pandoraviruses and *P. sibericum* were reported to be linked to them (Iyer, Balaji et al. 2006, Boyer, Madoui et al. 2010, Williams, Embley et al. 2011, Yamada 2011, Colson, de Lamballerie et al. 2012, Yutin and Koonin 2012, Colson, De Lamballerie et al. 2013, Philippe, Legendre et al. 2013, Yutin, Colson et al. 2013, Yutin and Koonin 2013, Legendre, Bartoli et al. 2014). Recently, NCLDVs were re-classified into a proposed new order, so-called *Megavirales* (Colson, de Lamballerie et al. 2012, Colson, De Lamballerie et al. 2013, Yutin, Colson et al. 2013). These viruses have a high diversity regarding their topology, genome size and protein content. Therefore, the classification of *Megavirales* is being revisited with the discovery of new viruses. An example of these phylogenomic studies and taxonomic issues is the classification of giant viruses as a fourth domain of life (Boyer, Madoui et al. 2010). These phylogenetic studies aimed to reveal the evolutionary relationships either between *Megavirales* members, or between them and organisms from the cellular domains of life. Virophages have not been clearly classified yet (Desnues and Raoult 2010).

3. Mechanisms used by large DNA viruses of *Megavirales* to enter into their host cell

We aimed to review the different mechanisms used by large DNA members (asfarviruses, poxviruses, ascoviruses, iridoviruses and phycodnaviruses) of *Megavirales* order (Table 1) to enter their host cells.

3.1. Asfarviruses

African swine fever virus (ASFV) is a linear dsDNA virus with 170-190 kilobase pairs (kbp) long genomes that harbor about 150 genes. The replication of this virus occurs at both cytoplasmic and nuclear locations (Garcia-Beato, Salas et al. 1992, Chapman, Tcherepanov et al. 2008, Ballester, Rodriguez-Carino et al. 2011).

The early steps of binding and entry of African swine fever virus are largely unknown (Alonso, Galindo et al. 2013). The ASFV-E70 and Ba71V strains enter into Vero cells and macrophage by a dynamin- and clathrin-dependent endocytotic pathway; which requires the small GTPase Rab7 and a phosphatidylinositol 3-kinase (Hernaez and Alonso 2010, Cuesta-Geijo, Galindo et al. 2012). The Rac1, a member of Rho family of small GTPases, is critical for intracellular transport, infection and viral production (Quetglas, Hernaez et al. 2012). There were other evidence of entry through macropinocytosis, which requires actin, kinases and Na+/H+ exchange (Sanchez, Quintas et al. 2012).

3.2. Poxviruses

Poxviruses are enveloped viruses, which are widely distributed (Smith, Vanderplasschen et al. 2002, Cyrklaff, Risco et al. 2005, Condit, Moussatche et al. 2006, Moss 2006). They have about 195 kbp dsDNA genome, which encodes about 200 proteins, and replicate in the cytoplasm of the host cell (Smith, Vanderplasschen et al. 2002, Cyrklaff, Risco et al. 2005, Condit, Moussatche et al. 2006, Moss 2006). Vaccinia virus is a prototypic virus amongst poxviruses and is used as smallpox vaccine (Yang, Reynolds et al. 2011). Vaccinia virus is found in three forms (Smith, Vanderplasschen et al. 2002, Cyrklaff, Risco et al. 2005, Condit, Moussatche et al. 2006, Moss 2006): 1) Mature virion (MVs, intracellular mature virus [IMV] or INV), which is the most abundant, stable, simple form, and mediates host-host transmission; MV has brick-shaped structure, dumbbell-shaped core, DNA, lipid membrane with >20 non-glycosylated proteins, and tubular protrusions on the surface; 2) wrapped virion (WVs or intracellular enveloped virus [IEV]), which contains the MV core wrapped by two membranes. WV travels to the cell periphery via microtubules, fuses with the plasma membrane and is released by exocytosis; 3) extracellular virion (EV, cell-associated extracellular enveloped virus [CEV] or extracellular enveloped virus [EEV]), which is less abundant, fragile, and is composed of MV core with an additional membrane that contains >6 proteins and is specialized for exit and cell-to-cell transmission within host. Therefore, poxviruses use different mechanisms to enter into its host cells (Moss 2006, Moss 2012, Schmidt, Bleck et al. 2012).

The entry-fusion complex (EFC) is composed of MV membrane proteins that mediate virus-cell fusion, cell membrane disruption and cell-to-cell fusion (Senkevich, Ojeda et al. 2005, Moss 2006, Moss 2012). Four proteins (A26, A27, D8, and H3) are involved in attachment of

poxviruses, whereas 12 proteins (A16L, A21L, A28L, F9, G3L, G9R, H2, I2, J5, L1R, L5R, and O3L) are involved in entry (Moss 2012). Inhibition of any of them perturbs the virus entry.

Early studies showed neutral or low-pH fusion for MV (Vanderplasschen, Hollinshead et al. 1998, Sodeik and Krijnse-Locker 2002, Townsley, Weisberg et al. 2006, Moss 2012, Schmidt, Bleck et al. 2013). Low-pH or binding to cellular glycosaminoglycans induce the disruption of the outer membrane of EV, which exposes MV at low-pH (Vanderplasschen, Hollinshead et al. 1998, Carter, Law et al. 2005), followed by fusion. The fusion, but not the binding, is affected by inhibition of tyrosine kinases, dynamin and actin, whereas the viral core entry is affected by inhibition of endosomal acidification and membrane blabbing (Laliberte, Weisberg et al. 2011). MV and EV enter HeLa cells by dynamin, actin, cholesterol, and low-pH dependent macropinocytosis. However, they induce different signal cascades (Locker, Kuehn et al. 2000, Townsley, Weisberg et al. 2006, Huang, Lu et al. 2008, Mercer and Helenius 2008, Mercer, Knebel et al. 2010, Schmidt, Bleck et al. 2011), whereas they enter dendritic cells by dynamin- and pH-independent macropinocytosis (Sandgren, Wilkinson et al. 2010). Phosphatidylserine, actin, kinases, GTPases, and Na+/H+ exchangers are required for internalization (Mercer and Helenius 2008). The virus entry is reduced by depletion of integrins that activates phosphatidylinositol 3-kinase (Izmailyan, Hsao et al. 2012), and is inhibited by depletion of cholesterol, suggesting the role of the lipid raft (Laliberte and Moss 2010). The vaccinia virus penetration factors and the CD98 glycoprotein, associated with lipid rafts, participate in the MVs uptake (Huang, Lu et al. 2008, Schroeder, Chung et al. 2012). Knocking down CD98 in HeLa cells reduces vaccinia MV entry, but not virus attachment, which indicates that it is required in endocytosis (Schroeder, Chung et al. 2012). Gas6 protein enhances the entry of vaccinia EV (strain Western Reserve) through bridging viral phosphatidylserine to TAM (Tyro3/Axl/Mer) receptor tyrosine kinases (Morizono, Xie et al. 2011).

Vaccinia virus strain Western Reserve and International Health Department-J (IHD-J) enter HeLa cells by macropinocytosis using phosphatidylserine, kinases and actin (Mercer, Knebel et al. 2010). IHD-J MV is low pH-independent and induces filopodia. Noteworthy, WR, not IHD-J, uses tyrosine kinase, phosphatidylinositol 3-kinase and Rac1 to activate blebs (Mercer, Knebel et al. 2010). The entry into HeLa, B78H1 and L cells, but not Vero and BSC-1 cells, is inhibited by soluble heparin, whereas inhibitors of endosomal acidification affect the virus entry into B78H1 and BSC-1 cells but not Vero and HeLa cells (Bengali, Townsley et al. 2009, Whitbeck, Foo et al. 2009). Bafilomycin A1 (that prevents acidification) inhibits the entry of vaccinia virus strains Western Reserve, Wyeth, cowpox virus, and monkeypoxvirus at

different levels, whereas it does not inhibit entry of IHD-J, Copenhagen, and Elstree strains, and of modified vaccinia virus Ankara strain (Bengali, Townsley et al. 2009, Bengali, Satheshkumar et al. 2012). Low-pH enhances entry of vaccinia virus strain Western Reserve, monkeypox virus, cowpox virus (Bengali, Townsley et al. 2009, Bengali, Satheshkumar et al. 2012). Vaccinia virus strains IHD-J, Copenhagen and Elstree fuse with the host cell membrane after attachment with glycosaminoglycans and do not require low pH (Bengali, Townsley et al. 2009, Bengali, Satheshkumar et al. 2012). Mutant vaccinia virus strains lacking A25/A26 fuse with HeLa cells at neutral pH, while strains Western Reserve and IHD-J with A25/A26 enter via endocytosis, suggesting the role of A25 and A26 as fusion suppressors (Chang, Chang et al. 2010, Moss 2012).

In cancer cells, entry of Western Reserve MV is accelerated by a low-pH, and requires serine/threonine-protein kinase PAK1 and a tyrosine kinase (Villa, Bartee et al. 2010). In cell lines, increased expression of vascular endothelial growth factor A activates Akt signaling pathway, which enhances entry and replication of vaccinia virus (Hiley, Chard et al. 2013). The binding of vaccinia virus to cell lines / leukocytes can be inhibited by heparin and laminin, whereas binding of myxoma virus is inhibited only by heparin, which indicates that the two viruses bind to leukocytes differently and have different tropisms (Chan, Bartee et al. 2013). This could be a consequence of the expression by primary human leukocytes of receptors specific to vaccinia virus, which triggers the virus binding (Byrd, Amet et al. 2013). Vaccinia virus entry and replication is enhanced by tumor necrosis factor receptor associated factor 2 in fibroblast or HeLa cells (Haga, Pechenick Jowers et al. 2014).

In *Drosophila* cells, vaccinia virus enters by macropinocytosis induced by a phosphatase and kinases as adenosine monophosphate-activated protein kinase (AMPK) (Moser, Jones et al. 2010). Vaccinia virus uptake by *Drosophila* S2 cells occurs via a low-pH endocytic pathway that requires EFC proteins (Bengali, Satheshkumar et al. 2011). In addition, MV of Ectromelia virus, which belong to the genus *Orthopoxvirus* and causes mousepox, can spread between hepatic cells (Ma, Xu et al. 2013).

3.3. Ascoviruses

Ascoviruses have circular dsDNA genomes, 156-186 kbp in length, that encode about 180 proteins with nuclear replication, and can be transmitted by wasps (Hussain, Garrad et al. 2009). The entry mechanism of ascoviruses into their host cells is largely unknown. The entry

of *Heliothis virescens* ascovirus-3e requires actin rearrangement (Hussain, Garrad et al. 2009). The destabilization of actin affects the viral infection but not entry.

3.4. Iridoviruses

Iridoviruses have linear dsDNA genomes, 100-200 kbp long, that are circularly-permuted and harbor redundant termini, and encode about 100-200 proteins (Eaton, Ring et al. 2010). Their hosts include fishes, crustaceans and insects (Eaton, Ring et al. 2010). Iridoviruses are enveloped or non-enveloped viruses and are divided into five genera (Chinchar, Yu et al. 2011).

The entry pathway differs among iridoviruses according to the virus topology. Generally, enveloped viruses enter by fusion, whereas non-enveloped viruses use an endocytosis pathway (Chinchar, Yu et al. 2011). The frog virus 3 that belongs to genus *Ranavirus*, enters mammalian BHK-21 cells by low pH-dependent clathrin endocytosis (Braunwald, Nonnenmacher et al. 1985). In contrast, the tiger frog virus from the same genus was shown to use a different mechanism of entry, since it entered into mammalian HepG2 cells via pH- and caveolin-dependent endocytosis that requires cholesterol, dynamin and actin (Guo, Liu et al. 2011). The infectious spleen and kidney necrosis virus (ISKNV), which belongs to genus *Megalocytivirus*, enters mandarin fish fry cells via a caveolin-endocytic pathway (Guo, Wu et al. 2012). This entry requires interaction between mandarin fish caveolin 1 and the major capsid protein (Jia, Wu et al. 2013). In large yellow croaker iridovirus, protein 037L binds to integrins at the cell surface allowing the virus uptake, while lipids from the viral outer layer facilitate the binding and fusion with the cellular membrane (Ao and Chen 2006, Wu, Chan et al. 2010).

3.5. Phycodnaviruses

Phycodnaviruses are dsDNA viruses with a 330-560 kbp genome that encodes 200-400 proteins and replicate in the nucleus (Van Etten, Graves et al. 2002). The *Phaeocystis globosa* virus PgV-16T is the largest known dsDNA virus that infect algae with a 460 kbp long genome that encodes 434 putative proteins (Santini, Jeudy et al. 2013). Although phycodnavirus hosts are algae, which have a cell topology distinct from that of other eukaryotes, their entry routes

are similar to those used by bacteriophages or some animal viruses (Van Etten, Graves et al. 2002). *Paramecium bursaria* chlorella virus (PBCV-1), which belongs to genus *Chlorovirus* (Van Etten and Dunigan 2012, Jeanniard, Dunigan et al. 2013), attaches to its host by the virus vertex and degrades the host cell wall at the site of attachment, similarly to bacteriophages (Meints, Lee et al. 1984). Then, replication initiates about 1 hour after this step. The PBCV-1 genome encodes chitinase, chitosanase, β-1,3-glucanase and alginase enzymes that lyse the cell wall (Yanai-Balser, Duncan et al. 2010). This genome also encodes potassium ion channel proteins, which have a putative role in entry and hence viral replication (Wang, Xie et al. 2011, Romani, Piotrowski et al. 2013).

The *Ectocarpus fasciculatus* virus, from genus *Phaeovirus*, infects zoospores or gametes of brown algae that lack a cell wall (Maier, Müller et al. 2002). This virus fuses with the host outer plasma membrane, leaving the capsid outside the cell surface, and injects the genome into the cytoplasm. *Emiliania huxleyi* virus 86 enters into host cells via endocytosis or fusion of the outer lipid membrane surrounding the capsid (Mackinder, Worthy et al. 2009). The intact virion is seen in the cytoplasm, then capsid breaks down releasing the viral genome.

4. Mechanisms used by giant DNA viruses of *Megavirales* to enter into their host cell

We aimed to review the different mechanisms used by giant DNA members (mimiviruses, marseilleviruses, pandoraviruses and *Pithovirus* spp.) of order *Megavirales*, and mimivirus virophages (Table 1) to enter their host cells.

4.1. Mimiviruses

Mimiviridae members are icosahedral viruses with large virion particle size that are referred as giant viruses (Raoult, Audic et al. 2004, Claverie and Abergel 2009, Fischer, Allen et al. 2010, Arslan, Legendre et al. 2011). They have circular dsDNA genomes that are predicted to encode for about 1000 proteins (Legendre, Audic et al. 2010). They replicate within the cytoplasm of their host cells, which are free-living amoebae of the genus *Acanthamoeba*, either species *castellanii* or *polyphaga* (Mutsafi, Zauberman et al. 2010). The family *Mimiviridae* encompasses several viruses all discovered during the past decade, which

were isolated from different environmental samples, consisting in fresh- or marine water or soil, and more recently from humans (La Scola, Campocasso et al. 2010).

The *Acanthamoeba polyphaga* mimivirus (APMV, mimivirus), was the first member described and prototype among mimiviruses. This virus was isolated from cooling tower in Bradford in 1992 by co-culturing on *Acanthamoeba polyphaga* phagocytic protozoan (Raoult, Audic et al. 2004). *Acanthamoeba castellanii* mamavirus, another strain of Mimivirus, was isolated by co-culturing on *A. castellanii* from cooling tower water from France (Colson, Yutin et al. 2011). *Acanthamoeba polyphaga* moumouvirus and *Megavirus chilensis* are closely related viruses that represent a second and third lineage of mimiviruses of amoeba; the latter has the largest virion and genome size amongst these viruses (Arslan, Legendre et al. 2011). Moreover, other mimiviruses were isolated including Hirudovirus isolated from *Hirudo medicinalis* Leech (Boughalmi, Pagnier et al. 2013), and recently, members of *Mimiviridae*, named LBA111 virus and Shan virus, were isolated from two patients with pneumonia (La Scola, Marrie et al. 2005, Vincent, La Scola et al. 2010, Parola, Renvoise et al. 2012, Vanspauwen, Franssen et al. 2012), while a Mimivirus protein similar to human collagen was suspected to be responsible for arthritis in humans (Shah, Hulsmeier et al. 2014). In addition, antibody to mimivirus have been detected in the serum samples from domestic mammals in Amazon region (Dornas, Rodrigues et al. 2014). *Cafeteria roenbergensis* virus, *Phaeocystis globosa* viruses and Organic Lake phycodnaviruses are distantly related mimiviruses that form a distant group amongst the family *Mimiviridae* and infect different phagocytic protists than *Acanthamoeba* spp. (Yutin, Colson et al. 2013).

Mimiviruses of amoeba are characterized by a big virion particle and a huge number of putative proteins. The function of most of these proteins is still not known, as they are products of ORFan genes or are un-annotated proteins, so-called hypothetical, for which homologs exist. These viruses have been isolated from a wide range of environmental samples and from humans. Although natural hosts are free-living amoebae, it remains unclear if other hosts exist (Dornas, Silva et al. 2014). Mimivirus was able to replicate in human macrophages (Ghigo, Kartenbeck et al. 2008) and human peripheral blood mononuclear cells (Dornas, Silva et al. 2014), and was described to induce type I interferon system (Dornas, Silva et al. 2014). Furthermore, LBA111 virus was isolated from the bronchoalveolar lavage of a two patients exhibiting pneumonia from Tunisia (Saadi, Pagnier et al. 2013). This virus is the largest virus recovered from a human and a close relative to *Megavirus chilensis*. Shan virus was subsequently isolated from a stool sample from a Tunisian patient with pneumonia (Saadi,

Pagnier et al. 2013). Noteworthy, Mimivirus has been found highly stable in environmental and clinical samples (Dornas, Silva et al. 2014), and Mimivirus can even resist extreme physical and chemical conditions, including UV light (Boratto, Dornas et al. 2013). Taken together, these findings prompt to discover new member of giant viruses in humans and study the interactions of these giant viruses with their host cells and their potential association with human diseases.

The surface of the Mimivirus is completely covered by fibers attached to the capsid through a disc-shaped structure, except at one vertex of the virus capsid, where a starfish shape gate is located that is suspected to allow the release of the genome (Xiao, Chipman et al. 2005, Xiao, Kuznetsov et al. 2009, Klose, Kuznetsov et al. 2010, Kuznetsov, Xiao et al. 2010). The fibers are strongly glycosylated and may be composed of peptidoglycan. They cannot be degraded by collagenase, thus are not composed of collagen. A head protein resists protease, which may be due to outer oligosaccharides (Xiao, Chipman et al. 2005, Xiao, Kuznetsov et al. 2009, Klose, Kuznetsov et al. 2010, Kuznetsov, Xiao et al. 2010). Surprisingly, culturing 150 times Mimivirus on amoeba in allopatric conditions, i.e. in germ free amoeba, led to the emergence of a strain, named M4, which lacks outer fibers (Boyer, Azza et al. 2011). The genome of Mimivirus M4 is shorter than that of mimivirus by about 200 kbp, and about 150 genes are missing in this strain. Interestingly, although Moumouvirus goulette and Moumouvirus Monve are closely related to Mimivirus, no fiber was observed on the outer capsid.

The entry of mimivirus into amoeba was first observed using electron microscopy and suggested that Mimivirus entry was through phagocytosis (Suzan-Monti, La Scola et al. 2007). Further experimental analyses showed that Mimivirus enters macrophages via dynamin-dependent phagocytosis, using actin and phosphatidylinositol 3-kinase, but not, rabankyrin-5 nor Na+/H+ exchange (Ghigo, Kartenbeck et al. 2008). The association of fiber with virus entry has been previously described for adenoviruses (Cusack 2005, Zhang and Bergelson 2005, Russell 2009, Cupelli and Stehle 2011, Arnberg 2012). Although the role of Mimivirus fiber in entry is not clear, they may have a role in attachment on the amoebal cell via their head (Xiao, Chipman et al. 2005, Xiao, Kuznetsov et al. 2009, Klose, Kuznetsov et al. 2010, Kuznetsov, Xiao et al. 2010). The lytic effect of bald Mimivirus M4 on amoeba is lower than that of the parental Mimivirus, which argues for the role of the fibers in attachment of mimivirus or entry into amoebal cells (Boyer, Azza et al. 2011). Protein analysis of purified mimivirus fibers (using 2D gel electrophoresis coupled with matrix-assisted laser desorption/ionization mass spectrometry) identified three proteins as associated with the

Mimivirus fibers (Boyer, Azza et al. 2011). These three proteins are a GMC-type oxidoreductase (R135) and two hypothetical proteins (L725 and L829), which are missing in the bald M4 strain. Further study identified nine proteins that are associated with glucose biosynthesis (Piacente, Marin et al. 2012). These proteins could have a role in the biosynthesis of the Mimivirus fibers, since these fibers are highly glycosylated. Knocking down three of these proteins (R135, L725, L829) and a tetratricopeptide repeat (TPR) domain-containing protein (R856) not detected by proteomics in the bald M4 strain led to the production of viruses without fiber or with shorter fibers or fibers with a modified shape [chapter 2, article no. 1]. The role of fiber in Mimivirus entry and or attachment into amoebae should be elucidated by further studies. Interestingly, Mimivirus proteins tentatively identified as involved in fiber formation have no homolog in other viruses than giant viruses of amoeba, including adenovirus fiber proteins, which highlight the uniqueness of these giant viruses.

4.2. Virophages of the giant mimiviruses

Virophages are viruses that infect giant viruses within their protistan host. They are dsDNA virus and their genome is about 18-19 kbp in length and encodes about 20 proteins (La Scola, Desnues et al. 2008, Fischer and Suttle 2011). Sputnik was the first isolated virophage, in co-infection with Mamavirus, another Mimivirus strain (La Scola, Desnues et al. 2008); then, Mavirus was isolated with *Cafeteria roenbergensis* virus (Fischer and Suttle 2011), and the Organic Lake virophage was isolated with Organic Lake phycodnaviruses (reclassified as distantly related mimiviruses) from Organic Lake in Antartica (Yau, Lauro et al. 2011). Regarding Sputnik, this virus has mushroom-like protrusions that are attached to capsomers and could function in recognizing the host (Sun, La Scola et al. 2010, Krupovic and Cvirkaite-Krupovic 2011).

The mechanism by which virophages enter their host cells and the receptors that need to be further studied. Notably, Sputnik was found associated with Mimivirus fibers (Desnues and Raoult 2010), and the bald Mimivirus M4 strain resists virophage infection, which supports the hypothesis that Sputnik can attach to Mimivirus fibers.

4.3. Marseilleviruses

Marseilleviruses are similar to mimiviruses regarding virion particle topology and are also giant viruses of amoeba, though they have capsid diameter about 2-3 times smaller than those of mimiviruses. They have circular dsDNA genomes that are also about 3 times smaller than those of mimiviruses. Marseillevirus was the first isolate from the new family *Marseilleviridae* (Colson, Pagnier et al. 2013). This virus has short fibers, about 12 nm in length, at its surface (Boyer, Yutin et al. 2009). Lausannevirus was also isolated by culturing in amoeba as well from Switzerland (Thomas, Bertelli et al. 2011). The Insectomime virus is isolated from insect *Eristalis tenax* (Diptera: *Syrphidae*) from Tunisia (Boughalmi, Pagnier et al. 2013).

Marseilleviruses infect same hosts as mimiviruses, namely *Acanthamoeba castellanii and polyphaga*. Noteworthy, Marseillevirus close relatives were recently detected from the blood and a lymph node from humans and a giant blood Marseillevirus could replicate in Jurkat cells, though was lost thereafter (Popgeorgiev, Boyer et al. 2013). The entry pathway of the marseilleviruses is largely unknown.

4.4. Pandoraviruses

The *Pandoravirus* spp. have capsid particle size >700 nm in length. The genome of *P. salinus,* isolated from coastal sediments in Chile, is about 2.5 Mbp in length and harbors about 2500 putative genes, making it the largest known virus genome. *P. dulcis* was isolated from a freshwater pond in Australia, and has a 1.9 Mbp long genome that encodes ~1500 proteins (Philippe, Legendre et al. 2013). Further phylogenetic analysis described them to be belong to the family *Phycodnaviridae* (Yutin and Koonin 2013). Entry mechanism for these viruses is unknown.

4.5. *Pithovirus sibericum*

Pithovirus sibericum virions are similar to pandoraviruses, with a larger capsid particle, about 1.5 μm in length, but a smaller genome, about 600 kbp long, which encodes ~467 proteins. This virus was isolated from Siberian permafrost by co-culturing on amoeba (*A. castellanii*) (Legendre, Bartoli et al. 2014). Although, the topology of the virion resembles those of

pandoraviruses, gene content and phylogenetic analysis revealed that the virus is most closely related to marseilleviruses and iridoviruses. Entry mechanism for this virus is unknown.

5. Conclusion

Viruses from the proposed order *Megavirales* display considerable differences in virus topology and genome size. They infect a wide spectrum of hosts and tropisms and were isolated from a wide range of environment and organisms. This viral group was enriched during the past decade by giant viruses of amoeba, which have remarkable features and considerably expand our knowledge of the virosphere. Current knowledge on entry pathways of the megaviruses is limited, and this is particularly true for the giant ones.

Several *Megavirales* members are animal pathogens, and mimiviruses, their virophages and marseilleviruses were recently detected in and isolated from humans (Yoosuf, Pagnier et al. 2013). Mimivirus was associated with pneumonia patient (Saadi, Pagnier et al. 2013, Saadi, Reteno et al. 2013), while a marseillevirus was associated with adenitis (Popgeorgiev, Boyer et al. 2013). These recent findings warrant gaining a better knowledge of the tropism and hosting range of these giant viruses. In addition, entry step of viral replication cycle is an interesting target to develop an antiviral drug that may eliminate or inhibit the infection outside the host cell, instead of inhibiting replication at a later stage inside the host cell, as a concern of cytotoxicity, which highlights the significance of studying the entry of the viruses. In addition, the increase concern of generation of mutant viruses, including for bioterrorism (Smith and Wang 2013), can also justify to study megavirus entry into their host cells. Taken together, previous data prompt to conduct future studies to elucidate the interactions with their host cells and the entry pathway of megaviruses, including giant viruses of amoebae.

References

Alonso, C., I. Galindo, M. A. Cuesta-Geijo, M. Cabezas, B. Hernaez and R. Munoz-Moreno (2013). "African swine fever virus-cell interactions: From virus entry to cell survival." Virus Res **173**(1): 42-57.

Ao, J. and X. Chen (2006). "Identification and characterization of a novel gene encoding an RGD-containing protein in large yellow croaker iridovirus." Virology **355**(2): 213-222.

Arnberg, N. (2012). "Adenovirus receptors: implications for targeting of viral vectors." Trends Pharmacol Sci **33**(8): 442-448.

Arslan, D., M. Legendre, V. Seltzer, C. Abergel and J. M. Claverie (2011). "Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae." Proc Natl Acad Sci U S A **108**(42): 17486-17491.

Ballester, M., C. Rodriguez-Carino, M. Perez, C. Gallardo, J. M. Rodriguez, M. L. Salas and F. Rodriguez (2011). "Disruption of nuclear organization during the initial phase of African swine fever virus infection." J Virol **85**(16): 8263-8269.

Bengali, Z., P. S. Satheshkumar and B. Moss (2012). "Orthopoxvirus species and strain differences in cell entry." Virology **433**(2): 506-512.

Bengali, Z., P. S. Satheshkumar, Z. Yang, A. S. Weisberg, N. Paran and B. Moss (2011). "Drosophila S2 cells are non-permissive for vaccinia virus DNA replication following entry via low pH-dependent endocytosis and early transcription." PLoS One **6**(2): e17248.

Bengali, Z., A. C. Townsley and B. Moss (2009). "Vaccinia virus strain differences in cell attachment and entry." Virology **389**(1-2): 132-140.

Boratto, P. V., F. P. Dornas, K. R. Andrade, R. Rodrigues, F. Peixoto, L. C. Silva, B. La Scola, A. O. Costa, G. M. de Almeida, E. G. Kroon and J. S. Abrahao (2013). "Amoebas as mimivirus bunkers: increased resistance to UV light, heat and chemical biocides when viruses are carried by amoeba hosts." Arch Virol **2013**: 26.

Boughalmi, M., I. Pagnier, S. Aherfi, P. Colson, D. Raoult and B. La Scola (2013). "First isolation of a giant virus from wild Hirudo medicinalis leech: Mimiviridae isolation in Hirudo medicinalis." Viruses **5**(12): 2920-2930.

Boughalmi, M., I. Pagnier, S. Aherfi, P. Colson, D. Raoult and B. La Scola (2013). "First isolation of a Marseillevirus in the Diptera Syrphidae Eristalis tenax." Intervirology **56**(6): 386-394.

Boyer, M., S. Azza, L. Barrassi, T. Klose, A. Campocasso, I. Pagnier, G. Fournous, A. Borg, C. Robert, X. Zhang, C. Desnues, B. Henrissat, M. G. Rossmann, B. La Scola and D. Raoult (2011). "Mimivirus shows dramatic genome reduction after intraamoebal culture." Proc Natl Acad Sci U S A **108**(25): 10296-10301.

Boyer, M., G. Gimenez, M. Suzan-Monti and D. Raoult (2010). "Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses." Intervirology **53**(5): 310-320.

Boyer, M., M. A. Madoui, G. Gimenez, B. La Scola and D. Raoult (2010). "Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses." PLoS One **5**(12): e15530.

Boyer, M., N. Yutin, I. Pagnier, L. Barrassi, G. Fournous, L. Espinosa, C. Robert, S. Azza, S. Sun, M. G. Rossmann, M. Suzan-Monti, B. La Scola, E. V. Koonin and D. Raoult (2009). "Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms." Proc Natl Acad Sci U S A **106**(51): 21848-21853.

Braunwald, J., H. Nonnenmacher and F. Tripier-Darcy (1985). "Ultrastructural and biochemical study of frog virus 3 uptake by BHK-21 cells." J Gen Virol **66 (Pt 2)**: 283-293.

Byrd, D., T. Amet, N. Hu, J. Lan, S. Hu and Q. Yu (2013). "Primary human leukocyte subsets differentially express vaccinia virus receptors enriched in lipid rafts." J Virol **87**(16): 9301-9312.

Carter, G. C., M. Law, M. Hollinshead and G. L. Smith (2005). "Entry of the vaccinia virus intracellular mature virion and its interactions with glycosaminoglycans." J Gen Virol **86**(Pt 5): 1279-1290.

Chan, W. M., E. C. Bartee, J. S. Moreb, K. Dower, J. H. Connor and G. McFadden (2013). "Myxoma and vaccinia viruses bind differentially to human leukocytes." J Virol **87**(8): 4445-4460.

Chang, S. J., Y. X. Chang, R. Izmailyan, Y. L. Tang and W. Chang (2010). "Vaccinia virus A25 and A26 proteins are fusion suppressors for mature virions and determine strain-specific virus entry pathways into HeLa, CHO-K1, and L cells." J Virol **84**(17): 8422-8432.

Chapman, D. A., V. Tcherepanov, C. Upton and L. K. Dixon (2008). "Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates." J Gen Virol **89**(Pt 2): 397-408.

Chinchar, V. G., K. H. Yu and J. K. Jancovich (2011). "The molecular biology of frog virus 3 and other iridoviruses infecting cold-blooded vertebrates." Viruses **3**(10): 1959-1985.

Claverie, J. M. and C. Abergel (2009). "Mimivirus and its virophage." Annu Rev Genet **43**: 49-66.

Colson, P., X. de Lamballerie, G. Fournous and D. Raoult (2012). "Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales." Intervirology **55**(5): 321-332.

Colson, P., X. De Lamballerie, N. Yutin, S. Asgari, Y. Bigot, D. K. Bideshi, X. W. Cheng, B. A. Federici, J. L. Van Etten, E. V. Koonin, B. La Scola and D. Raoult (2013). ""Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses." Arch Virol **158**(12): 2517-2521.

Colson, P., I. Pagnier, N. Yoosuf, G. Fournous, B. La Scola and D. Raoult (2013). ""Marseilleviridae", a new family of giant viruses infecting amoebae." Arch Virol **158**(4): 915-920.

Colson, P., N. Yutin, S. A. Shabalina, C. Robert, G. Fournous, B. La Scola, D. Raoult and E. V. Koonin (2011). "Viruses with more than 1,000 genes: Mamavirus, a new Acanthamoeba polyphaga mimivirus strain, and reannotation of Mimivirus genes." Genome Biol Evol **3**: 737-742.

Condit, R. C., N. Moussatche and P. Traktman (2006). "In a nutshell: structure and assembly of the vaccinia virion." Adv Virus Res **66**: 31-124.

Cuesta-Geijo, M. A., I. Galindo, B. Hernaez, J. I. Quetglas, I. Dalmau-Mena and C. Alonso (2012). "Endosomal maturation, Rab7 GTPase and phosphoinositides in African swine fever virus entry." PLoS One **7**(11): e48853.

Cupelli, K. and T. Stehle (2011). "Viral attachment strategies: the many faces of adenoviruses." Curr Opin Virol **1**(2): 84-91.

Cusack, S. (2005). "Adenovirus complex structures." Curr Opin Struct Biol **15**(2): 237-243.

Cyrklaff, M., C. Risco, J. J. Fernandez, M. V. Jimenez, M. Esteban, W. Baumeister and J. L. Carrascosa (2005). "Cryo-electron tomography of vaccinia virus." Proc Natl Acad Sci U S A **102**(8): 2772-2777.

Desnues, C. and D. Raoult (2010). "Inside the lifestyle of the virophage." Intervirology **53**(5): 293-303.

Dornas, F. P., F. P. Rodrigues, P. V. Boratto, L. C. Silva, P. C. Ferreira, C. A. Bonjardim, G. S. Trindade, E. G. Kroon, B. La Scola and J. S. Abrahao (2014). "Mimivirus Circulation among Wild and Domestic Mammals, Amazon Region, Brazil." Emerg Infect Dis **20**(3): 469-472.

Dornas, F. P., L. C. Silva, G. M. de Almeida, R. K. Campos, P. V. Boratto, A. P. Franco-Luiz, B. La Scola, P. C. Ferreira, E. G. Kroon and J. S. Abrahao (2014). "Acanthamoeba polyphaga mimivirus Stability in Environmental and Clinical Substrates: Implications for Virus Detection and Isolation." PLoS One **9**(2): e87811.

Eaton, H. E., B. A. Ring and C. R. Brunetti (2010). "The genomic diversity and phylogenetic relationship in the family iridoviridae." Viruses **2**(7): 1458-1475.

Fischer, M. G., M. J. Allen, W. H. Wilson and C. A. Suttle (2010). "Giant virus with a remarkable complement of genes infects marine zooplankton." Proc Natl Acad Sci U S A **107**(45): 19508-19513.

Fischer, M. G. and C. A. Suttle (2011). "A virophage at the origin of large DNA transposons." Science **332**(6026): 231-234.

Forterre, P. (2010). "Giant viruses: conflicts in revisiting the virus concept." Intervirology **53**(5): 362-378.

Garcia-Beato, R., M. L. Salas, E. Vinuela and J. Salas (1992). "Role of the host cell nucleus in the replication of African swine fever virus DNA." Virology **188**(2): 637-649.

Ghigo, E., J. Kartenbeck, P. Lien, L. Pelkmans, C. Capo, J. L. Mege and D. Raoult (2008). "Ameobal pathogen mimivirus infects macrophages through phagocytosis." PLoS Pathog **4**(6): e1000087.

Guo, C. J., D. Liu, Y. Y. Wu, X. B. Yang, L. S. Yang, S. Mi, Y. X. Huang, Y. W. Luo, K. T. Jia, Z. Y. Liu, W. J. Chen, S. P. Weng, X. Q. Yu and J. G. He (2011). "Entry of tiger frog virus (an Iridovirus) into HepG2 cells via a pH-dependent, atypical, caveola-mediated endocytosis pathway." J Virol **85**(13): 6416-6426.

Guo, C. J., Y. Y. Wu, L. S. Yang, X. B. Yang, J. He, S. Mi, K. T. Jia, S. P. Weng, X. Q. Yu and J. G. He (2012). "Infectious spleen and kidney necrosis virus (a fish iridovirus) enters Mandarin fish fry cells via caveola-dependent endocytosis." J Virol **86**(5): 2621-2631.

Haga, I. R., T. Pechenick Jowers, S. J. Griffiths, J. Haas and P. M. Beard (2014). "TRAF2 Facilitates Vaccinia Virus Replication by Promoting Rapid Virus Entry." J Virol **88**(7): 3664-3677.

Hernaez, B. and C. Alonso (2010). "Dynamin- and clathrin-dependent endocytosis in African swine fever virus entry." J Virol **84**(4): 2100-2109.

Hiley, C. T., L. S. Chard, R. Gangeswaran, J. R. Tysome, A. Briat, N. R. Lemoine and Y. Wang (2013). "Vascular endothelial growth factor A promotes vaccinia virus entry into host cells via activation of the Akt pathway." J Virol **87**(5): 2781-2790.

Huang, C. Y., T. Y. Lu, C. H. Bair, Y. S. Chang, J. K. Jwo and W. Chang (2008). "A novel cellular protein, VPEF, facilitates vaccinia virus penetration into HeLa cells through fluid phase endocytosis." J Virol **82**(16): 7988-7999.

Hussain, M., S. Garrad and S. Asgari (2009). "The role of actin filaments in ascovirus replication and pathology." Arch Virol **154**(11): 1737-1743.

Iyer, L. M., L. Aravind and E. V. Koonin (2001). "Common origin of four diverse families of large eukaryotic DNA viruses." J Virol **75**(23): 11720-11734.

Iyer, L. M., S. Balaji, E. V. Koonin and L. Aravind (2006). "Evolutionary genomics of nucleo-cytoplasmic large DNA viruses." Virus Res **117**(1): 156-184.

Izmailyan, R., J. C. Hsao, C. S. Chung, C. H. Chen, P. W. Hsu, C. L. Liao and W. Chang (2012). "Integrin beta1 mediates vaccinia virus entry through activation of PI3K/Akt signaling." J Virol **86**(12): 6677-6687.

Jeanniard, A., D. D. Dunigan, J. R. Gurnon, I. V. Agarkova, M. Kang, J. Vitek, G. Duncan, O. W. McClung, M. Larsen, J. M. Claverie, J. L. Van Etten and G. Blanc (2013). "Towards defining the chloroviruses: a genomic journey through a genus of large DNA viruses." BMC Genomics **14**: 158.

Jia, K. T., Y. Y. Wu, Z. Y. Liu, S. Mi, Y. W. Zheng, J. He, S. P. Weng, S. C. Li, J. G. He and C. J. Guo (2013). "Mandarin fish caveolin 1 interaction with major capsid protein of infectious spleen and kidney necrosis virus and its role in early stages of infection." J Virol **87**(6): 3027-3038.

Klose, T., Y. G. Kuznetsov, C. Xiao, S. Sun, A. McPherson and M. G. Rossmann (2010). "The three-dimensional structure of Mimivirus." Intervirology **53**(5): 268-273.

Krupovic, M. and V. Cvirkaite-Krupovic (2011). "Virophages or satellite viruses?" Nat Rev Microbiol **9**(11): 762-763.

Kuznetsov, Y. G., C. Xiao, S. Sun, D. Raoult, M. Rossmann and A. McPherson (2010). "Atomic force microscopy investigation of the giant mimivirus." Virology **404**(1): 127-137.

La Scola, B., A. Campocasso, R. N'Dong, G. Fournous, L. Barrassi, C. Flaudrops and D. Raoult (2010). "Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry." Intervirology **53**(5): 344-353.

La Scola, B., C. Desnues, I. Pagnier, C. Robert, L. Barrassi, G. Fournous, M. Merchat, M. Suzan-Monti, P. Forterre, E. Koonin and D. Raoult (2008). "The virophage as a unique parasite of the giant mimivirus." Nature **455**(7209): 100-104.

La Scola, B., T. J. Marrie, J. P. Auffray and D. Raoult (2005). "Mimivirus in pneumonia patients." Emerg Infect Dis **11**(3): 449-452.

Laliberte, J. P. and B. Moss (2010). "Lipid membranes in poxvirus replication." Viruses **2**(4): 972-986.

Laliberte, J. P., A. S. Weisberg and B. Moss (2011). "The membrane fusion step of vaccinia virus entry is cooperatively mediated by multiple viral proteins and host cell components." PLoS Pathog **7**(12): e1002446.

Legendre, M., S. Audic, O. Poirot, P. Hingamp, V. Seltzer, D. Byrne, A. Lartigue, M. Lescot, A. Bernadac, J. Poulain, C. Abergel and J. M. Claverie (2010). "mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus." Genome Res **20**(5): 664-674.

Legendre, M., J. Bartoli, L. Shmakova, S. Jeudy, K. Labadie, A. Adrait, M. Lescot, O. Poirot, L. Bertaux, C. Bruley, Y. Coute, E. Rivkina, C. Abergel and J. M. Claverie (2014). "Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology." Proc Natl Acad Sci U S A **2014**: 3.

Locker, J. K., A. Kuehn, S. Schleich, G. Rutter, H. Hohenberg, R. Wepf and G. Griffiths (2000). "Entry of the two infectious forms of vaccinia virus at the plasma membane is signaling-dependent for the IMV but not the EEV." Mol Biol Cell **11**(7): 2497-2511.

Ma, X., R. H. Xu, F. Roscoe, J. C. Whitbeck, R. J. Eisenberg, G. H. Cohen and L. J. Sigal (2013). "The mature virion of ectromelia virus, a pathogenic poxvirus, is capable of intrahepatic spread and can serve as a target for delayed therapy." J Virol **87**(12): 7046-7053.

Mackinder, L. C., C. A. Worthy, G. Biggi, M. Hall, K. P. Ryan, A. Varsani, G. M. Harper, W. H. Wilson, C. Brownlee and D. C. Schroeder (2009). "A unicellular algal virus, Emiliania huxleyi virus 86, exploits an animal-like infection strategy." J Gen Virol **90**(Pt 9): 2306-2316.

Maier, I., D. G. Müller and C. Katsaros (2002). "Entry of the DNA virus, Ectocarpus fasciculatus virus type 1 (Phycodnaviridae), into host cell cytosol and nucleus." Phycological Research **50**(3): 227-231.

Meints, R. H., K. Lee, D. E. Burbank and J. L. Van Etten (1984). "Infection of a Chlorella-like alga with the virus, PBCV-1: ultrastructural studies." Virology **138**(2): 341-346.

Mercer, J. and A. Helenius (2008). "Vaccinia virus uses macropinocytosis and apoptotic mimicry to enter host cells." Science **320**(5875): 531-535.

Mercer, J., S. Knebel, F. I. Schmidt, J. Crouse, C. Burkard and A. Helenius (2010). "Vaccinia virus strains use distinct forms of macropinocytosis for host-cell entry." Proc Natl Acad Sci U S A **107**(20): 9346-9351.

Moreira, D. and C. Brochier-Armanet (2008). "Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes." BMC Evol Biol **8**: 12.

Morizono, K., Y. Xie, T. Olafsen, B. Lee, A. Dasgupta, A. M. Wu and I. S. Chen (2011). "The soluble serum protein Gas6 bridges virion envelope phosphatidylserine to the TAM receptor tyrosine kinase Axl to mediate viral entry." Cell Host Microbe **9**(4): 286-298.

Moser, T. S., R. G. Jones, C. B. Thompson, C. B. Coyne and S. Cherry (2010). "A kinome RNAi screen identified AMPK as promoting poxvirus entry through the control of actin dynamics." PLoS Pathog **6**(6): e1000954.

Moss, B. (2006). "Poxvirus entry and membrane fusion." Virology **344**(1): 48-54.

Moss, B. (2012). "Poxvirus cell entry: how many proteins does it take?" Viruses **4**(5): 688-707.

Mutsafi, Y., N. Zauberman, I. Sabanay and A. Minsky (2010). "Vaccinia-like cytoplasmic replication of the giant Mimivirus." Proc Natl Acad Sci U S A **107**(13): 5978-5982.

Parola, P., A. Renvoise, E. Botelho-Nevers, B. La Scola, C. Desnues and D. Raoult (2012). "Acanthamoeba polyphaga mimivirus Virophage Seroconversion in Travelers Returning from Laos." Emerg Infect Dis **18**(9): 1500-1502.

Philippe, N., M. Legendre, G. Doutre, Y. Coute, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J. M. Claverie and C. Abergel (2013). "Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes." Science **341**(6143): 281-286.

Piacente, F., M. Marin, A. Molinaro, C. De Castro, V. Seltzer, A. Salis, G. Damonte, C. Bernardi, J. M. Claverie, C. Abergel and M. Tonetti (2012). "Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine)." J Biol Chem **287**(5): 3009-3018.

Popgeorgiev, N., M. Boyer, L. Fancello, S. Monteil, C. Robert, R. Rivet, C. Nappez, S. Azza, J. Chiaroni, D. Raoult and C. Desnues (2013). "Marseillevirus-like virus recovered from blood donated by asymptomatic humans." J Infect Dis **208**(7): 1042-1050.

Quetglas, J. I., B. Hernaez, I. Galindo, R. Munoz-Moreno, M. A. Cuesta-Geijo and C. Alonso (2012). "Small rho GTPases and cholesterol biosynthetic pathway intermediates in African swine fever virus infection." J Virol **86**(3): 1758-1767.

Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan and J. M. Claverie (2004). "The 1.2-megabase genome sequence of Mimivirus." Science **306**(5700): 1344-1350.

Raoult, D. and P. Forterre (2008). "Redefining viruses: lessons from Mimivirus." Nat Rev Microbiol **6**(4): 315-319.

Romani, G., A. Piotrowski, S. Hillmer, J. Gurnon, J. L. Van Etten, A. Moroni, G. Thiel and B. Hertel (2013). "A virus-encoded potassium ion channel is a structural protein in the chlorovirus Paramecium bursaria chlorella virus 1 virion." J Gen Virol **94**(Pt 11): 2549-2556.

Russell, W. C. (2009). "Adenoviruses: update on structure and function." J Gen Virol **90**(Pt 1): 1-20.

Saadi, H., I. Pagnier, P. Colson, J. K. Cherif, M. Beji, M. Boughalmi, S. Azza, N. Armstrong, C. Robert, G. Fournous, B. La Scola and D. Raoult (2013). "First isolation of Mimivirus in a patient with pneumonia." Clin Infect Dis **57**(4): e127-134.

Saadi, H., D. G. Reteno, P. Colson, S. Aherfi, P. Minodier, I. Pagnier, D. Raoult and B. La Scola (2013). "Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia." Intervirology **56**(6): 424-429.

Sanchez, E. G., A. Quintas, D. Perez-Nunez, M. Nogal, S. Barroso, A. L. Carrascosa and Y. Revilla (2012). "African swine fever virus uses macropinocytosis to enter host cells." PLoS Pathog **8**(6): e1002754.

Sandgren, K. J., J. Wilkinson, M. Miranda-Saksena, G. M. McInerney, K. Byth-Wilson, P. J. Robinson and A. L. Cunningham (2010). "A differential role for macropinocytosis in mediating entry of the two forms of vaccinia virus into dendritic cells." PLoS Pathog **6**(4): e1000866.

Santini, S., S. Jeudy, J. Bartoli, O. Poirot, M. Lescot, C. Abergel, V. Barbe, K. E. Wommack, A. A. Noordeloos, C. P. Brussaard and J. M. Claverie (2013). "Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes." Proc Natl Acad Sci U S A **110**(26): 10800-10805.

Schmidt, F. I., C. K. Bleck, A. Helenius and J. Mercer (2011). "Vaccinia extracellular virions enter cells by macropinocytosis and acid-activated membrane rupture." EMBO J **30**(17): 3647-3661.

Schmidt, F. I., C. K. Bleck and J. Mercer (2012). "Poxvirus host cell entry." Curr Opin Virol **2**(1): 20-27.

Schmidt, F. I., C. K. Bleck, L. Reh, K. Novy, B. Wollscheid, A. Helenius, H. Stahlberg and J. Mercer (2013). "Vaccinia virus entry is followed by core activation and proteasome-mediated release of the immunomodulatory effector VH1 from lateral bodies." Cell Rep **4**(3): 464-476.

Schroeder, N., C. S. Chung, C. H. Chen, C. L. Liao and W. Chang (2012). "The lipid raft-associated protein CD98 is required for vaccinia virus endocytosis." J Virol **86**(9): 4868-4882.

Senkevich, T. G., S. Ojeda, A. Townsley, G. E. Nelson and B. Moss (2005). "Poxvirus multiprotein entry-fusion complex." Proc Natl Acad Sci U S A **102**(51): 18572-18577.

Shah, N., A. J. Hulsmeier, N. Hochhold, M. Neidhart, S. Gay and T. Hennet (2014). "Exposure to mimivirus collagen promotes arthritis." J Virol **88**(2): 838-845.

Smith, G. L., A. Vanderplasschen and M. Law (2002). "The formation and function of extracellular enveloped vaccinia virus." J Gen Virol **83**(Pt 12): 2915-2931.

Smith, I. and L. F. Wang (2013). "Bats and their virome: an important source of emerging viruses capable of infecting humans." Curr Opin Virol **3**(1): 84-91.

Sodeik, B. and J. Krijnse-Locker (2002). "Assembly of vaccinia virus revisited: de novo membrane synthesis or acquisition from the host?" Trends Microbiol **10**(1): 15-24.

Sun, S., B. La Scola, V. D. Bowman, C. M. Ryan, J. P. Whitelegge, D. Raoult and M. G. Rossmann (2010). "Structural studies of the Sputnik virophage." J Virol **84**(2): 894-897.

Suzan-Monti, M., B. La Scola, L. Barrassi, L. Espinosa and D. Raoult (2007). "Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga Mimivirus." PLoS One **2**(3): e328.

Thomas, V., C. Bertelli, F. Collyn, N. Casson, A. Telenti, A. Goesmann, A. Croxatto and G. Greub (2011). "Lausannevirus, a giant amoebal virus encoding histone doublets." Environ Microbiol **13**(6): 1454-1466.

Townsley, A. C., A. S. Weisberg, T. R. Wagenaar and B. Moss (2006). "Vaccinia virus entry into cells via a low-pH-dependent endosomal pathway." J Virol **80**(18): 8899-8908.

Van Etten, J. L. and D. D. Dunigan (2012). "Chloroviruses: not your everyday plant virus." Trends Plant Sci **17**(1): 1-8.

Van Etten, J. L., M. V. Graves, D. G. Muller, W. Boland and N. Delaroque (2002). "Phycodnaviridae--large DNA algal viruses." Arch Virol **147**(8): 1479-1516.

Vanderplasschen, A., M. Hollinshead and G. L. Smith (1998). "Intracellular and extracellular vaccinia virions enter cells by different mechanisms." J Gen Virol **79 (Pt 4)**: 877-887.

Vanspauwen, M. J., F. M. Franssen, D. Raoult, E. F. Wouters, C. A. Bruggeman and C. F. Linssen (2012). "Infections with mimivirus in patients with chronic obstructive pulmonary disease." Respir Med **106**(12): 1690-1694.

Villa, N. Y., E. Bartee, M. R. Mohamed, M. M. Rahman, J. W. Barrett and G. McFadden (2010). "Myxoma and vaccinia viruses exploit different mechanisms to enter and infect human cancer cells." Virology **401**(2): 266-279.

Vincent, A., B. La Scola and L. Papazian (2010). "Advances in Mimivirus pathogenicity." Intervirology **53**(5): 304-309.

Wang, K., S. Xie and B. Sun (2011). "Viral proteins function as ion channels." Biochim Biophys Acta **1808**(2): 510-515.

Whitbeck, J. C., C. H. Foo, M. Ponce de Leon, R. J. Eisenberg and G. H. Cohen (2009). "Vaccinia virus exhibits cell-type-dependent entry characteristics." Virology **385**(2): 383-391.

Williams, T. A., T. M. Embley and E. Heinz (2011). "Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses." PLoS One **6**(6): e21080.

Wu, J., R. Chan, M. R. Wenk and C. L. Hew (2010). "Lipidomic study of intracellular Singapore grouper iridovirus." Virology **399**(2): 248-256.

Xiao, C., P. R. Chipman, A. J. Battisti, V. D. Bowman, P. Renesto, D. Raoult and M. G. Rossmann (2005). "Cryo-electron microscopy of the giant Mimivirus." J Mol Biol **353**(3): 493-496.

Xiao, C., Y. G. Kuznetsov, S. Sun, S. L. Hafenstein, V. A. Kostyuchenko, P. R. Chipman, M. Suzan-Monti, D. Raoult, A. McPherson and M. G. Rossmann (2009). "Structural studies of the giant mimivirus." PLoS Biol **7**(4): e92.

Yamada, T. (2011). "Giant viruses in the environment: their origins and evolution." Curr Opin Virol **1**(1): 58-62.

Yanai-Balser, G. M., G. A. Duncan, J. D. Eudy, D. Wang, X. Li, I. V. Agarkova, D. D. Dunigan and J. L. Van Etten (2010). "Microarray analysis of Paramecium bursaria chlorella virus 1 transcription." J Virol **84**(1): 532-542.

Yang, Z., S. E. Reynolds, C. A. Martens, D. P. Bruno, S. F. Porcella and B. Moss (2011). "Expression profiling of the intermediate and late stages of poxvirus replication." J Virol **85**(19): 9899-9908.

Yau, S., F. M. Lauro, M. Z. DeMaere, M. V. Brown, T. Thomas, M. J. Raftery, C. Andrews-Pfannkoch, M. Lewis, J. M. Hoffman, J. A. Gibson and R. Cavicchioli (2011). "Virophage control of antarctic algal host-virus dynamics." Proc Natl Acad Sci U S A **108**(15): 6163-6168.

Yoosuf, N., I. Pagnier, G. Fournous, C. Robert, B. La Scola, D. Raoult and P. Colson (2013). "Complete genome sequence of Courdo11 virus, a member of the family Mimiviridae." Virus Genes **2013**: 1.

Yutin, N., P. Colson, D. Raoult and E. V. Koonin (2013). "Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family." Virol J **10**: 106.

Yutin, N. and E. V. Koonin (2012). "Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes." Virol J **9**: 161.

Yutin, N. and E. V. Koonin (2013). "Pandoraviruses are highly derived phycodnaviruses." Biol Direct **8**: 25.

Yutin, N., Y. I. Wolf, D. Raoult and E. V. Koonin (2009). "Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution." Virol J **6**: 223.

Zhang, Y. and J. M. Bergelson (2005). "Adenovirus receptors." J Virol **79**(19): 12125-12131.

Table 1. Entry mechanisms used by *Megavirales* members.

Family	Lineage, genus or species	Hosts or isolation sources	Entry mechanism
Asfarviridae	ASF virus	Swine	Clathrin endocytosis or macropinocytosis
Poxviridae	*Orthopoxvirus*	Human, primates, camels, rodent	Fusion and macropinocytosis
	Leporipoxvirus	Rabbit	
	Squirrelpox virus species	Squirrel	
	Crocodylidpoxvirus	Nile crocodile	
	Molluscipoxvirus	Immunosuppressed human	
	Parapoxvirus	Superorder Laurasiatheria	
	Yatapoxvirus	Primate	
	Suipoxvirus	Swine	
	Cervidpoxvirus	Deer	
Ascoviridae	Ascovirus	Invertebrates	Unkown
Iridoviridae	*Chloriridovirus*	Mosquitos	Fusion, dynamin- / caveolin- / raft-endocytosis or Low-pH /clathrin endocytosis
	Iridovirus	Crustaceans, insects	
	Lymphocystivirus	Fishes	
	Megalocytivirus	Fishes	
	Ranavirus	Amphibians, reptiles	
Phycodnaviridae	*Chlorovirus*	Marine protozoa and Algae	Fusion and Cell wall degradation
	Coccolithovirus		
	Phaeovirus		
	Prasinovirus		
	Prymnesiovirus		
	Raphidovirus		
Mimiviridae	Mimivirus	Human, insect or protozoa	Phagocytosis-like
	Moumouvirus		
	Megavirus chiliensis		
	Cafeteria roenbergensis virus		
	Phaeocystis globosa virus		
Marseilleviridae	Marseillevirus	Human, insect or protozoa	Unknown
	Lausannevirus		
	Tunisvirus		
Unassigned	*Pandoravirus salinus* and *P. dulcis*	*Acanthamoeba* spp.	Unknown
Unassigned	*Pithovirus sibericum*	*Acanthamoeba* spp.	Unknown

Chapter 2

Study of Mimivirus genes associated with fibers using short interfering RNA

Introduction

Based on the presence of external protrusion, viruses can be classified into four categories: (i) Viruses that show no evidence for presence of protrusion or harbor short glycoproteins at their surface, such as poxviruses; (ii) Viruses that harbor short spikes, such as astroviruses, microviruses, and reoviruses; (iii) Viruses that harbor a tail fiber, such as those in bacteriophages or fuselloviruses; (iv) Viruses that harbor fibers, including members from the *Adenoviridae*, *Guttaviridae*, *Rudiviridae* and *Mimiviridae* families. Surprisingly, most members of the family *Mimiviridae* harbor fibers at their surface, such as Mimivirus, whereas other viruses harbor short fiber, such as Marseillevirus, or no apparent fiber, such as Lausannevirus.

Mimivirus, the founder and prototype member of the family *Mimiviridae*, is covered by long fibers about 130 nm in length. These fibers are distinct from those of other viruses by their topology. Unlike adenovirus fibers, mimivirus fibers are longer, softer and more flexible, and they cover the whole mimivirus capsid except at one vertex, where a star-shaped structure is located.

In this study, the presence or absence of fibers among mimiviruses was identified using electron microscopy. Then, using RNA interference, we provided evidence that four proteins are involved in fibers either as component or by playing role in fiber biosynthesis. We identified homologs of these mimivirus proteins in other giant viruses and members from the cellular domains of life. In addition, this work is a proof of concept for the use of RNA interference to study the functionality and role of ORFan genes and genes encoding hypothetical proteins that represent more than half of the Mimivirus gene repertoire.

Research article no. 1

Short title: RNAi of Mimivirus fiber-associated genes

Full-length title: Identification of giant Mimivirus protein functions using RNA interference

Author list: Haitham SOBHY[1], Bernard LA SCOLA[1,2], Isabelle PAGNIER[1], Didier RAOULT[1,2], Philippe COLSON[1,2] *

[1] Aix-Marseille Univ., URMITE UM63 CNRS 7278 IRD 198 INSERM U1095, Marseille, France; [2]Fondation IHU Méditerranée Infection, Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Centre Hospitalo-Universitaire Timone, Assistance Publique-Hôpitaux de Marseille, Marseille, France

* To whom correspondence should be addressed.

Author contributions: H.S., P.C., B.L.S. and D.R. conceived and designed the experiments; H.S., B.L.S., I.P., D.R. and P.C. analyzed the data; H.S. performed the experiments; and H.S., P.C. and D.R. wrote the manuscript.

The authors declare no conflict of interest.

Abstract: 207 words; text: 3,330 words

Tables: 5; Figures: 5

Abstract

Genomic analysis of giant viruses, such as Mimivirus, has revealed that more than half of the putative genes have no known functions (ORFans). We knocked down Mimivirus genes using short interfering RNA (siRNA) as a proof of concept to determine the functions of giant virus ORFans. As fibers are easy to observe, we targeted a gene encoding a protein absent in a Mimivirus mutant devoid of fibers as well as 3 genes encoding products identified in a protein concentrate of fibers, including one ORFan and one gene of unknown function. We found that knocking down these four genes was associated with depletion or modification of the fibers. In contrast, an siRNA targeting the capsid protein (L425) did not modify the fibers. Our strategy of silencing ORFan genes in giant viruses opens a way to identify its complete gene repertoire and may clarify the role of these genes, differentiating between junk DNA and truly used genes. Using this strategy, we were able to annotate 4 proteins in Mimivirus and 30 homologous proteins in other giant viruses. In addition, because 3 proteins share sequence homology with homologs from organisms of cellular domains of life, we were able to annotate >500 proteins from cellular organisms and 100 from metagenomic databases.

Introduction

Acanthamoeba polyphaga mimivirus was the first member discovered of the viral family *Mimiviridae*, which encompasses viruses that infect *Acanthamoeba* spp. (La Scola, Audic et al. 2003, Raoult, Audic et al. 2004). Subsequently, dozens of Mimivirus relatives have been isolated from environmental samples and, more recently, from humans (La Scola, Desnues et al. 2008, Boyer, Yutin et al. 2009, Fischer, Allen et al. 2010, Arslan, Legendre et al. 2011, Yoosuf, Yutin et al. 2012, Saadi, Pagnier et al. 2013, Saadi, Reteno et al. 2013). Other giant viruses that infect protozoa were also subsequently discovered, including marseilleviruses (Boyer, Yutin et al. 2009, Boughalmi, Pagnier et al. 2013, Boughalmi, Saadi et al. 2013, Aherfi, Boughalmi et al. 2014), pandoraviruses (Philippe, Legendre et al. 2013) and *Pithovirus sibericum* (Legendre, Bartoli et al. 2014). Mimiviruses have been linked, along with the marseilleviruses, to the nucleocytoplasmic large DNA viruses (NCLDVs), which were recently proposed to be reclassified into a new viral order named the "Megavirales" (Colson, De Lamballerie et al. 2013). These giant viruses have raised considerable interest in the field of evolutionary biology because of their unexpectedly large size, as well as the fact that they contain genes encoding functions previously believed to be the in domain of cellular organisms, such as aminoacyl-tRNA synthetases or translations factors. They have challenged the definition of a virus (Raoult, Audic et al. 2004, Moreira and Brochier-Armanet 2008, Raoult and Forterre 2008, Forterre 2010).

The genomes of *Megavirales* members contain a large number of predicted genes annotated either as hypothetical proteins or ORFan genes, i.e., genes without homologs in sequence databases (Raoult, Audic et al. 2004, Boyer, Gimenez et al. 2010). For example, genes encoding hypothetical proteins occupy more than 50% of the Mimivirus genome. The functions and evolutionary origins of these genes are not known. To date, the functions of only a few Mimivirus proteins have been studied experimentally, including amino-acyl-tRNA synthetases (Abergel, Chenivesse et al. 2005, Abergel, Rudinger-Thirion et al. 2007) and proteins involved in sugar biosynthesis (Bandaru, Zhao et al. 2007, Benarroch, Qiu et al. 2009, Piacente, Marin et al. 2012). Thus, the large majority of Mimivirus genes have no known function and make up a 'functional dark matter'.

The Mimivirus capsid, which is approximately 500 nm in size, is covered by a dense layer of fibers. These viral fibers are approximately 125-140 nm in length and approximately 1.4 nm in diameter and consist of a soft shaft and a globular shaped head (Xiao, Kuznetsov et al. 2009, Klose, Kuznetsov et al. 2010, Kuznetsov, Xiao et al. 2010). Clusters of 3-4 fibers were found to

be linked via a disc shaped base. They are highly glycosylated, antigenic, and resistant to protease and collagenase treatment (Xiao, Kuznetsov et al. 2009, Boyer, Azza et al. 2011) . A putative GMC-type oxidoreductase (R135), and two hypothetical proteins (L725, which is the product of an ORFan, and L829) were identified in purified fibers by gel electrophoresis coupled with matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) (Boyer, Azza et al. 2011). Sub-culturing Mimivirus 150 times on germ-free amoebae led to the emergence of a mutant "M4" strain lacking fibers, and with a genome reduced by 16% and missing 150 genes (Boyer, Azza et al. 2011). Comparative proteomics of M4 and the original Mimivirus strain showed a deletion of the R135 and L829 proteins as well as a tetratricopeptide repeat (TPR) containing protein (R856). In addition, nine proteins have been proposed to be involved in sugar biosynthesis and fiber formation (Piacente, Marin et al. 2012) (Supplemental Table S1).

Here, we aimed to apply RNA interference (RNAi) to the identification of the function of Mimivirus proteins. We targeted four Mimivirus genes associated with fiber formation, as fibers can be easily observed by electron microscopy.

Experimental procedures

Cell Culture, siRNA and identification of morphological changes

Targeted genes and siRNA. We targeted the Mimivirus genes R135, L725, L829 and R856 using short interfering RNA (siRNA). These genes were either identified in purified fibers or deleted in the M4 strain (Boyer, Azza et al. 2011) (Figure 1).

Cell culture and RNAi. A culture of *A. polyphaga* in 10 ml of PYG medium was seeded for 24-48 h. Then, 100 µl of Lipofectamine RNAiMAX (Invitrogen, USA) and 0.25 µg of duplex siRNA (designed and purchased from Invitrogen; for sequences, see Supplemental Table S2) were used according to the manufacturer's instructions. To improve the siRNA specificity, we used duplex siRNA and we performed BLASTn searches against Mimivirus genes. We used the recommended concentration of siRNA. One ml of Mimivirus (≈10^6 viruses) was added to the culture, and incubated for 11 h at 32 °C. *A. polyphaga* were harvested by centrifugation (2,000 rpm/10 min) and analyzed by electron microscopy. For protein analysis, four 10-ml flasks of Mimivirus culture for 24 h at 32 °C were used. Cells were completely lysed. Then, the culture medium was centrifuged at 2000 rpm for 10 min, and the supernatant was filtered through a

1.2-µm filter to eliminate *A. polyphaga.* Cell debris was removed by centrifugation at 10,000 rpm for 15 min. The Mimivirus pellet was washed twice with PBS, and the purified viruses were used in further investigations. For the negative control gene, the same experimental procedures were applied to the L425 gene encoding capsid protein.

Electron microscopy*.* The preparation of samples for electron microscopy was previously described (Boyer, Azza et al. 2011). Briefly, the samples were fixed with glutaraldehyde and cacodylate buffer, cut into 70-nm sections using an ultramicrotome (UC7; Leica), collected on 400-mesh nickel grids with formvar carbon, and stained for electron microscopy (FCF-400-Ni, Electron Microscopy Sciences). The samples were then viewed with a Philips electron microscope (Morgagni 268D) at 80 keV. Cross sections of all pictures that were selected for the analysis were positioned at the middle of the virions and characterized by a dense cluster (black mass). The observed mimiviruses were those that were available and cultured in our laboratory.

Analysis of protein content of knocked down viruses

Antibodies preparation. Fibers were purified from the virus as previously described in (Jensen, Houthaeve et al. 1996). Previous analysis of the fiber by 2D-gel coupled with MALDI-TOF MS (Boyer, Azza et al. 2011) revealed three proteins (R135, L725 and L829) were associated with Mimivirus fibers. For anti-L725 antibodies, L725 protein fused with thioredoxin was expressed in *Escherichia coli* (*E. coli*) and purified using ÄKTA avant 25 (GE Healthcare, USA). The purified L725 protein was then injected into mice to obtain anti-L725 polyclonal antibodies, as previously described (Boyer, Azza et al. 2011).

Immunogold labeling*.* Grids were immersed in NH_4Cl (50 mM), diluted in PBS three times for 5 min (3 x 5 min), washed in PBS for 5 min, and then immersed twice in blocking buffer (1% normal goat serum (NGS), 1% bovine serum albumin (BSA), and 0.2% Tween 20 diluted in PBS) (2 x 10 min). The grids were incubated with anti-fiber polyclonal antibody that was diluted 1:100 in blocking buffer overnight at 4 °C. After four 10-min washes, the grids were incubated for 90 min in biotin (Beckman Coulter, USA) that was diluted 1:100 in blocking buffer. Then, the grids were washed with 0.1% BSA-PBS (4 x 5 min) and then in 0.01% cold water fish skin (CWFS) gelatin-PBS (3 x 10 min), and incubated with streptavidin (labeled by 10-nm gold nano-particles; Aurion, The Netherlands) that was diluted 1:100 in 0.01% CWFS gelatin-PBS for 90 min and washed with PBS. After incubating with PBS-glutaraldehyde 2.5% for 15 min, the grids were washed with PBS (2 x 10 min) and distilled water for 10 min.

Finally, the grids were contrasted by adding uranyl acetate for 20 min, immersed in water 60 times, and analyzed using electron microscopy. The number of gold particles that were bound to fibers in each image was counted. The experiments included several steps that were performed on successive days. For each step, we used Mimivirus treated without siRNA with the same experimental conditions as the siRNA-treated Mimivirus as negative control.

Proteomic analysis. All proteomic analysis (sample preparation, 1D and 2D gel electrophoresis, silver staining, and western blotting) was performed as previously described (Azza, Cambillau et al. 2009). Briefly, Mimivirus was solubilized in 40 mM Tris-HCl, pH 7.5, supplemented with 2% (wt/vol) sodium dodecyl sulfate (SDS; Sigma-Aldrich) and 60 mM dithiothreitol (DTT), followed by 5 min of heating at 95°C. The insoluble fraction was removed by centrifugation (12,000 x g, 4 °C, 10 min), and soluble proteins were precipitated using a PlusOne 2-D Clean-Up kit (GE Healthcare, USA) to remove SDS. The final pellet was re-suspended in solubilization buffer [7 M urea, 2 M thiourea, 4% (wt/vol) 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS)] and stored at -80 °C until use. The protein concentration was measured by Bradford assay (Bio-Rad, USA). Immobiline DryStrips (13 cm, pH 3–10; GE Healthcare) were rehydrated overnight using 250 µl rehydration buffer [8 M urea, 2% (w/v) CHAPS, 60 mM DTT, 0,5% (v/v) IPG buffer (GE Healthcare)] containing 20 µg of solubilized APMV proteins and isoelectric focusing (IEF) was carried out according to the manufacturer's protocol (IPGphor II, GE Healthcare). Before the second dimension electrophoresis was performed, strips were equilibrated twice in 5 ml equilibration buffer [30% (v/v) glycerol, 3% (w/v) SDS, 6 M urea, 50 mM Tris-HCl, bromophenol blue, pH 8.8] for 15 min. This buffer was supplemented with 65 mM DTT for the first equilibration and with 100 mM iodoacetamide for the second one. The strips were then embedded in 0.5% agarose and the proteins resolved by 10% SDS-PAGE (Protean II XL, Bio-Rad). Gels were stained either with silver or transferred onto nitrocellulose membranes for western blot analysis using anti-Mimivirus, anti-L725 or anti-fiber primary polyclonal antibodies. Then, the membrane was washed three times with PBS-Tween and probed for 2 h with horseradish peroxidase-conjugated goat anti-mouse secondary antibodies.

MALDI-TOF MS biotyping. One microliter of purified virus (in PBS) was deposited onto a MALDI-TOF target plate (Bruker Daltonik, Germany), and four deposits were made for each virus. Then, the preparation was overlaid with 2 µl of matrix solution [a saturated solution of a-HCCA (alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile, 2.5% tri-fluoracetic-acid], and the matrix sample was crystallized by air drying at room temperature for 5 min on a

Microflex mass spectrometer (Bruker Daltonik), as previously described (La Scola, Campocasso et al. 2010).

Comparative genomics and phylogenetic tree reconstruction

Protein sequences of mimiviruses were retrieved from the NCBI GenBank non-redundant protein sequence database (nr) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). BLASTp searches were performed with 0.01 as the e-value cutoff. The best hits were collected and aligned using ClustalW (Larkin, Blackshields et al. 2007). The multiple sequence alignments were trimmed by Gblock (Castresana 2000). Phylogenetic tree reconstructions were performed using the Maximum likelihood method of the FastTree tool with default parameters (Price, Dehal et al. 2010).

PCR testing

The presence or absence of Mimivirus genes in the purified viral solution was determined by qPCR. The most conserved sites were identified, and universal primers and probes were designed using the Gemi tool (Sobhy and Colson 2012) (Supplemental Table S3). The 25 µl-real-time PCR mixture contained 5 µl of extracted DNA, 12.5 µl qPCR Mastermix (Eurogentec, Belgium), 0.5 µl of each primer (10 nmol/µl; Eurogentec), and 0.5 µl probe (3 nmol/µl; Applied Biosystems UK). The PCR thermal cycling conditions were: a hold at 50 °C for 2 min, a hold at 95 °C for 5 min, and then 45 cycles of 30 sec at 95 °C then 1 min at 60 °C.

Results

Consequences of silencing targeted Mimivirus genes on fiber formation

We knocked down the Mimivirus genes encoding the R135, L725, L829, R856 and L425 proteins using siRNA. We then compared the fibers from the viruses produced to those from control viruses produced in the absence of siRNA, searching for any abnormal feature of the fibers, such as short, prone (procumbent), or non-stretched and curved fibers (Table 1; Figure 2). To measure length ratio for silenced versus control viruses, we selected 4-8 viruses that harbored ≥30 fibers and measured the lengths of the fibers in each condition (Table 2). To determine protein contribution in fiber formation, we counted the number of gold particle conjugated with anti-fiber antibodies, hence to Mimivirus fibers (Table 3; Figure 3).

Control Mimivirus fibers. The average length of control fibers was 131 nm (Table 2; Figure 2A, 3A-C, Supplemental Fig.S1). Silencing the gene encoding the L425 protein did not affect fiber length or topology (Figure 2B, 3D, Supplemental Figure S2).

Fibers after R856 gene silencing. Approximately 60% of si-R856 viruses (meaning viruses treated with an siRNA targeting the R856 gene) harbored abnormal or short fibers (Table 1; Figure 2C, Supplemental Fig.S3). We observed that the average length of fibers from si-R856 viruses was 48 nm, which was 64% shorter than control viruses ($p<1e-6$) (Table 2). The number of gold particles bound to fibers was decreased by 81% after silencing R856 (Table 3; Figure 3).

Fibers after L725 gene silencing. Si-L725 viruses harbored approximately 50% abnormal curved fibers, which were 15% shorter than control fibers. The number of gold particles bound to fibers was decreased by 86% after silencing the L725 gene (Figure 2D, 3D; Supplemental Fig.S4).

Fibers after L829 gene silencing. The fiber length of si-L829 viruses was 90 nm, which was 30% shorter than the control virus fibers. In addition, the fiber layer was sparse in si-L829 viruses and the gold particle count was decreased by 57% (Figure 2E, 3E; Supplemental Fig.S5).

Fibers after R135 gene silencing. The fiber length of the si-R135 viruses was 12% shorter than the control and 30% of these fibers were curved (Figure 2F, Supplemental Fig.S6).

These findings indicate that (i) the presence of short or sparse fibers after siRNA treatment was due to the siRNAs, and not to lipid carrier or suboptimal experimental conditions, and that these siRNAs were specific to their targeted genes as we observed different characteristics of the Mimivirus fibers and their reactivity with antibodies depending on which gene was targeted by siRNAs; (ii) the R135, L725, L829, and R856 proteins are either principal elements of Mimivirus fibers or play a key role during fiber biosynthesis and can be functionally annotated as fiber associated proteins (FAPs); and (iii) the L725 and R856 proteins are major contributors to fiber formation.

Consequences of silencing targeted genes on protein content

Western blot analysis was performed to validate these results and revealed that the reactivity of antibodies to fiber associated proteins was reduced against viruses whose genes

were silenced compared to control viruses (Figure 4). Thus, Figures 4A and B show that the bands corresponding to R135, L725 and L829 molecular weights were reduced in silenced viruses. To confirm these results, nitrocellulose membranes were incubated with anti-Mimivirus and anti-L725 antibodies (Figure 4 C and D, respectively), which showed reduced reactivity against silenced viruses. These results indicate that the R135, L725, L829 and R856 proteins are associated with Mimivirus fiber formation and changes observed in the Mimivirus fiber layer are due to depletion of proteins (post-transcriptional events). Additionally, a comparison of Figure 4 B and C indicates that L725 is the principal protein of Mimivirus fibers.

2D-gel electrophoresis western blots performed for both si-L829 and si-R856 viruses also revealed a reduction in anti-fiber antibodies bound to the targeted fiber associated proteins compared to control viruses (Figure 5). Thus, protein content concentration (as measured by Bradford assay) was reduced by 60% when silencing the R856 gene and 30% when silencing the L829 gene (Supplemental Figure S7). A MALDI-TOF MS biotyping was then performed for the silenced and control viruses, which showed one extra spectrum for the si-L829 virus compared to the control virus (Supplemental Figure S8).

In conclusion, the R135, L725, L829 and R856 protein levels were depleted after knocking down their encoding genes, which is evidence that these proteins play a role in fiber formation, and can be functionally annotated as fiber associated proteins.

Comparative genomics and protein re-annotation

We propose here that the R135, L725, L829 and R856 proteins can be annotated as Mimivirus fiber associated proteins (FAPs), and their names can be abbreviated as FAP1, FAP2, FAP3 and FAP4, respectively.

Searching for sequence homology of the FAPs with proteins from giant virus *Megavirales* members and other organisms from public sequence databases revealed that these proteins are conserved in most of the giant viruses, but do not share sequence homology with any fiber or spike protein encoded by any virus, including adenoviruses (Table 4, 5; Supplemental Table S1, 9). FAP2 (L725) is an ORFan only present in mimiviruses. FAP4 (R856) contains seven TPR domains and shares sequence similarity with hypothetical proteins encoded by archaea, bacteria, choanoflagellida, ciliophora, metazoa (including rotifera, cnidaria, and hydra), and from metagenomes, but not with any protein encoded by any virus (Table 5, Supplemental Table S1, Supplemental Fig.S10-12). FAP1 (R135) shares homology with

oxidoreductases and hypothetical proteins encoded by *Acanthamoeba*, metazoa, fungi, and bacteria, including proteobacteria, as well as *P. sibericum*, and metagenomes. Finally, FAP3 (L829) is encoded by mimiviruses, marseilleviruses, *Pandoravirus* spp., and shares similarity with hypothetical proteins encoded by bacteria, and eukaryotes, including amoebozoa and fungi. Phylogenetic analyses indicate that FAP4 is widely distributed among environmental and aquatic species (Supplemental Figure S12). In addition, tree topologies suggest that FAP1 and 4 may have been subject to horizontal gene exchange with cellular organisms.

Taken together, these data indicate that Mimivirus fiber associated proteins are divergent from proteins that are encoded by other viruses, including Megavirales members other than giant viruses of amoeba, and might share a common ancestor or have been exchanged through horizontal gene transfer with proteins from cellular organisms. Moreover, with our siRNA-based strategy, we are able to functionally annotate 30 proteins from mimiviruses, as well as re-annotate 108 proteins from metagenomic (dark matter) databases and approximately 1,000 hypothetical proteins archived in public sequence databases and encoded by archaea, bacteria and eukaryotes (Table 5).

Discussion

We demonstrated, using siRNA, that four proteins are involved in Mimivirus fiber formation. A disturbance in the expression of one of these proteins significantly altered the size or shape of these fibers, which indicates that these proteins are either elements of the fiber or involved in fiber formation. To our knowledge, this is the first study that described a modification of Mimivirus virions, and used siRNA to determine the function of a Mimivirus gene.

In this article, we identified, using RNAi, the function of four proteins, including the L725 and L829 proteins with previously unknown function, a putative oxidoreductase (R135) and a TPR-containing protein (R856). Comparative genomic analyses indicated that the L725 encoding gene is an ORFan, while R135, L829, and R856 are unique amongst viruses to mimiviruses, but have homologs in amoeba, bacteria, fungi, and metazoa, and might have been exchanged by horizontal gene transfer. It is noteworthy that Mimivirus protein R856 belongs to the TPR superfamily of proteins that were reported to be involved in protein-protein interactions (Das, Cohen et al. 1998, Blatch and Lassle 1999, D'Andrea and Regan 2003, Cortajarena and Regan 2006), virus-host interactions (Callahan, Handley et al. 1998, Jeshtadi,

Burgos et al. 2010), and regulation of virus replication (Lin, Mendu et al. 2012, Tani, Shimamoto et al. 2013). We provide evidence for a new function of this protein in the formation of Mimivirus fibers.

ORFan and un-annotated genes occupy more than 50% of the gene repertoire of Mimivirus. Here, in addition to the successful identification of the function of four proteins as fiber associated proteins (FAPs) by siRNA, our new strategy allowed us to re-annotate 30 proteins in mimiviruses and closely related giant viruses that share sequences homology with FAPs. In addition, this strategy opens a way to re-annotate proteins from sequence databases and genomic dark matter, as was the case here for >500 bacterial proteins and approximately 100 proteins from metagenomes, (Table 5).

In this work, we targeted genes of Mimivirus fibers that are easy to observe by electron microscopy, immunogold and proteomics. However, our strategy of silencing ORFan genes in giant viruses opens the way to identify the function of their complete gene repertoires. In particular, the proteins of giant viruses of amoeba, like those from other intracellular species such as *Mycoplasma*, are poorly expressed and difficult to crystallize, making their functional analysis difficult. This proposed approach will lead to the annotation of hundreds of proteins without known function found in public databases and differentiate between junk DNA and truly used genes.

Funding

Haitham Sobhy received a grant from the Fondation Infectiopole Sud.

Acknowledgements

We are thankful to Said Azza, Jean-Pierre Baudoin and Claude Nappez for technical help.

References

Abergel, C., S. Chenivesse, D. Byrne, K. Suhre, V. Arondel and J. M. Claverie (2005). "Mimivirus TyrRS: preliminary structural and functional characterization of the first amino-acyl tRNA synthetase found in a virus." Acta Crystallogr Sect F Struct Biol Cryst Commun **61**(Pt 2): 212-215.

Abergel, C., J. Rudinger-Thirion, R. Giege and J. M. Claverie (2007). "Virus-encoded aminoacyl-tRNA synthetases: structural and functional characterization of mimivirus TyrRS and MetRS." J Virol **81**(22): 12406-12417.

Aherfi, S., M. Boughalmi, I. Pagnier, G. Fournous, B. La Scola, D. Raoult and P. Colson (2014). "Complete genome sequence of Tunisvirus, a new member of the proposed family Marseilleviridae." Arch Virol **2014**: 26.

Arslan, D., M. Legendre, V. Seltzer, C. Abergel and J. M. Claverie (2011). "Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae." Proc Natl Acad Sci U S A **108**(42): 17486-17491.

Azza, S., C. Cambillau, D. Raoult and M. Suzan-Monti (2009). "Revised Mimivirus major capsid protein sequence reveals intron-containing gene structure and extra domain." BMC Mol Biol **10**: 39.

Bandaru, V., X. Zhao, M. R. Newton, C. J. Burrows and S. S. Wallace (2007). "Human endonuclease VIII-like (NEIL) proteins in the giant DNA Mimivirus." DNA Repair (Amst) **6**(11): 1629-1641.

Benarroch, D., Z. R. Qiu, B. Schwer and S. Shuman (2009). "Characterization of a mimivirus RNA cap guanine-N2 methyltransferase." Rna **15**(4): 666-674.

Blatch, G. L. and M. Lassle (1999). "The tetratricopeptide repeat: a structural motif mediating protein-protein interactions." Bioessays **21**(11): 932-939.

Boughalmi, M., I. Pagnier, S. Aherfi, P. Colson, D. Raoult and B. La Scola (2013). "First isolation of a Marseillevirus in the Diptera Syrphidae Eristalis tenax." Intervirology **56**(6): 386-394.

Boughalmi, M., H. Saadi, I. Pagnier, P. Colson, G. Fournous, D. Raoult and B. La Scola (2013). "High-throughput isolation of giant viruses of the Mimiviridae and Marseilleviridae families in the Tunisian environment." Environ Microbiol **15**(7): 2000-2007.

Boyer, M., S. Azza, L. Barrassi, T. Klose, A. Campocasso, I. Pagnier, G. Fournous, A. Borg, C. Robert, X. Zhang, C. Desnues, B. Henrissat, M. G. Rossmann, B. La Scola and D. Raoult (2011). "Mimivirus shows dramatic genome reduction after intraamoebal culture." Proc Natl Acad Sci U S A **108**(25): 10296-10301.

Boyer, M., G. Gimenez, M. Suzan-Monti and D. Raoult (2010). "Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses." Intervirology **53**(5): 310-320.

Boyer, M., N. Yutin, I. Pagnier, L. Barrassi, G. Fournous, L. Espinosa, C. Robert, S. Azza, S. Sun, M. G. Rossmann, M. Suzan-Monti, B. La Scola, E. V. Koonin and D. Raoult (2009). "Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms." Proc Natl Acad Sci U S A **106**(51): 21848-21853.

Callahan, M. A., M. A. Handley, Y. H. Lee, K. J. Talbot, J. W. Harper and A. T. Panganiban (1998). "Functional interaction of human immunodeficiency virus type 1 Vpu and Gag with a novel member of the tetratricopeptide repeat protein family." J Virol **72**(6): 5189-5197.

Castresana, J. (2000). "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis." Mol Biol Evol **17**(4): 540-552.

Colson, P., X. De Lamballerie, N. Yutin, S. Asgari, Y. Bigot, D. K. Bideshi, X. W. Cheng, B. A. Federici, J. L. Van Etten, E. V. Koonin, B. La Scola and D. Raoult (2013). ""Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses." Arch Virol **158**(12): 2517-2521.

Colson, P., N. Yutin, S. A. Shabalina, C. Robert, G. Fournous, B. La Scola, D. Raoult and E. V. Koonin (2011). "Viruses with more than 1,000 genes: Mamavirus, a new Acanthamoeba polyphaga mimivirus strain, and reannotation of Mimivirus genes." Genome Biol Evol **3**: 737-742.

Cortajarena, A. L. and L. Regan (2006). "Ligand binding by TPR domains." Protein Sci **15**(5): 1193-1198.

D'Andrea, L. D. and L. Regan (2003). "TPR proteins: the versatile helix." Trends Biochem Sci **28**(12): 655-662.

Das, A. K., P. W. Cohen and D. Barford (1998). "The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions." EMBO J **17**(5): 1192-1199.

Fischer, M. G., M. J. Allen, W. H. Wilson and C. A. Suttle (2010). "Giant virus with a remarkable complement of genes infects marine zooplankton." Proc Natl Acad Sci U S A **107**(45): 19508-19513.

Forterre, P. (2010). "Giant viruses: conflicts in revisiting the virus concept." Intervirology **53**(5): 362-378.

Jensen, O. N., T. Houthaeve, A. Shevchenko, S. Cudmore, T. Ashford, M. Mann, G. Griffiths and J. Krijnse Locker (1996). "Identification of the major membrane and core proteins of vaccinia virus by two-dimensional electrophoresis." J Virol **70**(11): 7485-7497.

Jeshtadi, A., P. Burgos, C. D. Stubbs, A. W. Parker, L. A. King, M. A. Skinner and S. W. Botchway (2010). "Interaction of poxvirus intracellular mature virion proteins with the TPR domain of kinesin light chain in live infected cells revealed by two-photon-induced fluorescence resonance energy transfer fluorescence lifetime imaging microscopy." J Virol **84**(24): 12886-12894.

Klose, T., Y. G. Kuznetsov, C. Xiao, S. Sun, A. McPherson and M. G. Rossmann (2010). "The three-dimensional structure of Mimivirus." Intervirology **53**(5): 268-273.

Kuznetsov, Y. G., C. Xiao, S. Sun, D. Raoult, M. Rossmann and A. McPherson (2010). "Atomic force microscopy investigation of the giant mimivirus." Virology **404**(1): 127-137.

La Scola, B., S. Audic, C. Robert, L. Jungang, X. de Lamballerie, M. Drancourt, R. Birtles, J. M. Claverie and D. Raoult (2003). "A giant virus in amoebae." Science **299**(5615): 2033.

La Scola, B., A. Campocasso, R. N'Dong, G. Fournous, L. Barrassi, C. Flaudrops and D. Raoult (2010). "Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry." Intervirology **53**(5): 344-353.

La Scola, B., C. Desnues, I. Pagnier, C. Robert, L. Barrassi, G. Fournous, M. Merchat, M. Suzan-Monti, P. Forterre, E. Koonin and D. Raoult (2008). "The virophage as a unique parasite of the giant mimivirus." Nature **455**(7209): 100-104.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Legendre, M., J. Bartoli, L. Shmakova, S. Jeudy, K. Labadie, A. Adrait, M. Lescot, O. Poirot, L. Bertaux, C. Bruley, Y. Coute, E. Rivkina, C. Abergel and J. M. Claverie (2014). "Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology." Proc Natl Acad Sci U S A **111**(11): 4274-4279.

Lin, J. Y., V. Mendu, J. Pogany, J. Qin and P. D. Nagy (2012). "The TPR domain in the host Cyp40-like cyclophilin binds to the viral replication protein and inhibits the assembly of the tombusviral replicase." PLoS Pathog **8**(2): e1002491.

Moreira, D. and C. Brochier-Armanet (2008). "Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes." BMC Evol Biol **8**: 12.

Philippe, N., M. Legendre, G. Doutre, Y. Coute, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J. M. Claverie and C. Abergel (2013). "Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes." Science **341**(6143): 281-286.

Piacente, F., M. Marin, A. Molinaro, C. De Castro, V. Seltzer, A. Salis, G. Damonte, C. Bernardi, J. M. Claverie, C. Abergel and M. Tonetti (2012). "Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine)." J Biol Chem **287**(5): 3009-3018.

Price, M. N., P. S. Dehal and A. P. Arkin (2010). "FastTree 2--approximately maximum-likelihood trees for large alignments." PLoS One **5**(3): e9490.

Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan and J. M. Claverie (2004). "The 1.2-megabase genome sequence of Mimivirus." Science **306**(5700): 1344-1350.

Raoult, D. and P. Forterre (2008). "Redefining viruses: lessons from Mimivirus." Nat Rev Microbiol **6**(4): 315-319.

Renesto, P., C. Abergel, P. Decloquement, D. Moinier, S. Azza, H. Ogata, P. Fourquet, J. P. Gorvel and J. M. Claverie (2006). "Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products." J Virol **80**(23): 11678-11685.

Saadi, H., I. Pagnier, P. Colson, J. K. Cherif, M. Beji, M. Boughalmi, S. Azza, N. Armstrong, C. Robert, G. Fournous, B. La Scola and D. Raoult (2013). "First isolation of Mimivirus in a patient with pneumonia." Clin Infect Dis **57**(4): e127-134.

Saadi, H., D. G. Reteno, P. Colson, S. Aherfi, P. Minodier, I. Pagnier, D. Raoult and B. La Scola (2013). "Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia." Intervirology **56**(6): 424-429.

Sobhy, H. and P. Colson (2012). "Gemi: PCR primers prediction from multiple alignments." Comp Funct Genomics **2012**: 783138.

Tani, J., S. Shimamoto, K. Mori, N. Kato, K. Moriishi, Y. Matsuura, H. Tokumitsu, M. Tsuchiya, T. Fujimoto, K. Kato, H. Miyoshi, T. Masaki and R. Kobayashi (2013). "Ca(2+) /S100 proteins regulate HCV virus NS5A-FKBP8/FKBP38 interaction and HCV virus RNA replication." Liver Int **33**(7): 1008-1018.

Xiao, C., Y. G. Kuznetsov, S. Sun, S. L. Hafenstein, V. A. Kostyuchenko, P. R. Chipman, M. Suzan-Monti, D. Raoult, A. McPherson and M. G. Rossmann (2009). "Structural studies of the giant mimivirus." PLoS Biol **7**(4): e92.

Yoosuf, N., N. Yutin, P. Colson, S. A. Shabalina, I. Pagnier, C. Robert, S. Azza, T. Klose, J. Wong, M. G. Rossmann, B. La Scola, D. Raoult and E. V. Koonin (2012). "Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage." Genome Biol Evol **4**(12): 1324-1330.

Table 1. Number of viruses with fibers with normal or abnormal features

Silenced gene	Total no. of viruses observed	Abnormality %	Average abnormality %	No. of viruses with short fibers (%)	No. of viruses with sparse/curved fibers (%)
Control	91	0	0	0 (0)	0 (0)
si-L425 (control)	90	0	0	0 (0)	0 (0)
si-R135	123	16 - 54	35	20 (16)	47 (38)
si-L725	52	21 - 71	46	11 (21)	26 (50)
si-L829	90	13 - 49	31	12 (13)	32 (36)
si-R856	101	58 - 63	61	59 (58)	5 (5)

Normal fibers are those with classical shape and >100 nm-long, dense and stretched fibers. Sparse/curved fibers are fibers that are not as thick or dense as the normal or non-stretched fibers. Short fibers are those with a length <100 nm. The percentage of abnormality indicates the percentage of viruses with either short or non-stretched and sparse fibers.
si-, virus after silencing the target gene

Table 2. Length of the Mimivirus fibers according to each siRNA experimental condition

Virus	No. of viruses observed	No. of fibers	Mean length of fibers ± standard deviation	Relative length of fibers compared to control (%) (silenced / control fiber *100)	p-value [1]
Control	8	30	131.4±23.8	100.0	-
si-R135	6	34	115.0±27.5	87.5	0.0138
si-L725	4	31	110.8±26.0	84.4	0.0021
si-L829	7	33	91.8±25.0	69.9	<1e-6
si-R856	8	29	47.8±18.8	36.4	<1e-6

[1] P-values were calculated using the ANOVA test (http://www.openepi.com).
si-, virus after silencing the target gene

Table 3. Number of gold nanoparticles per virus in immunogold

Type	No. of viruses observed	Total no. of gold particles	Mean no. of gold particles per virus ± standard deviation	Decrease in particles per virus (%)
Control 1 (no siRNA)	30	880	29±7	0
si-L829	48	611	13±6	57
si-R856	44	249	6±3	81
si-L725	31	129	4±3	86
Control 2 (no anti-fiber antibodies)	39	0	0±0	100

The percentage of reduction measures the drop or reduction in gold particles after knocking down each gene.
si-, virus after silencing the target gene

Table 4. The distribution of fibers among viruses that infect *Acanthamoeba* spp. and were isolated in our laboratory or by other teams.

Virus	Fibers	R135 (FAP1)	L725 (FAP2)	L829 (FAP3)	R856 (FAP4)	L136	L137	L138	R139	L140	R141	L142	L143	R641	L872	mg878
APMV§	Yes	+ *	+ *	+ *	+ *	+ *	+	+	+ *	+	+	+ *	+	+ *	+	+
APMV-M4§	No	#	+ *	#	#	#	#	#	#	#	#	//	+	+ *	#	
ACMaV§	Yes	+ *	+ *	+ *	+ *	+ *	+	+	+ *	+	+	+ *	+	+ *	+	+
APLenV§	Short	+	+	+	+	+	+	+	+	+	+	+	+		+	+
APMoV§	Yes	+	+	+							II				+	+
Monv§	No	+	+	+							II				+	+
Goul§	No	+	+	+	+						II				+	+
Crdo11§	Yes	+	+	+			I	II		II	I		+			+
Crdo7§	Yes	+	+	+			I	I		I	I		+			+
MegCV	Yes	+	+	+			I	II		II	I		+			+
CroV											II					
MarsV§	No			I												
LauV	No			I												
PsV	Yes			I												+
PdV	Yes			I												
Pvs		I														
Sputnik§	Yes															

(+) means detected as a hit in the NCBI BLASTp search, with an e-value <0.01 and identity >40%. (I) identity is 20-30%; (II) identity is 30-40%; * Reliable qPCR threshold cycle (Ct) values; # deleted gene; // split gene; § the virus was isolated in our laboratory (supplemental Figure S9).

APMV: *Acanthamoeba polyphaga* mimivirus; APMV-M4: *Acanthamoeba polyphaga* mimivirus isolate M4; ACMaV: *Acanthamoeba castellanii* mamavirus; APMoV: *Acanthamoeba polyphaga* moumouvirus; CroV: *Cafeteria roenbergensis* virus BV-PW1; APLenV: *Acanthamoeba polyphaga* lentillevirus; LauV: Lausannevirus; MarsV: Marseillevirus; MegCV: *Megavirus chiliensis*; Crdo11: Courdo11 virus; Crdo7: Ccourdo7 virus; Goul: Moumouvirus goulette; and Monv: Moumouvirus monve; PsV: *Pandoravirus salinus*; PdV: *Pandoravirus dulcis*; Pvs: *Pithovirus sibericum* (La Scola, Desnues et al. 2008, Boyer, Yutin et al. 2009, Fischer, Allen et al. 2010, La Scola, Campocasso et al. 2010, Arslan, Legendre et al. 2011, Boyer, Azza et al. 2011, Colson, Yutin et al. 2011, Yoosuf, Yutin et al. 2012, Philippe, Legendre et al. 2013).

Table 5. Number of BLASTp hits corresponding to each Mimivirus fiber-associated protein

	Mimiviridae[1]	Viruses	*Archaea*	*Bacteria*	*Eukaryota*	Metagenome[2]
R135 (FAP1)	12	0	0	415	71	88
L725 (FAP2)	5	0	0	0	0	0
L829 (FAP3)	14	0	0	18	4	0
R856 (FAP4)	2	0	7	174	304	20

The numbers represent the number of hits per group of organisms using BLASTp against the NCBI or UniProt Archive (UniParc)[2] sequence databases. For details, see Supplemental Table S1. L725 gene is an ORFan and conserved only in mimiviruses. FAP means fiber associated protein.

[1] Includes hits in mimiviruses and closely related viruses, including pandoraviruses and *Pithovirus sibericum*.

[2] Retrieved from BLASTp search against UniProt Archive (UniParc) database, which includes UniProt Metagenomic and Environmental Sequences (UniMES) database.

Figure 1. Flow chart of the strategy implemented to characterize Mimivirus gene function using siRNA.

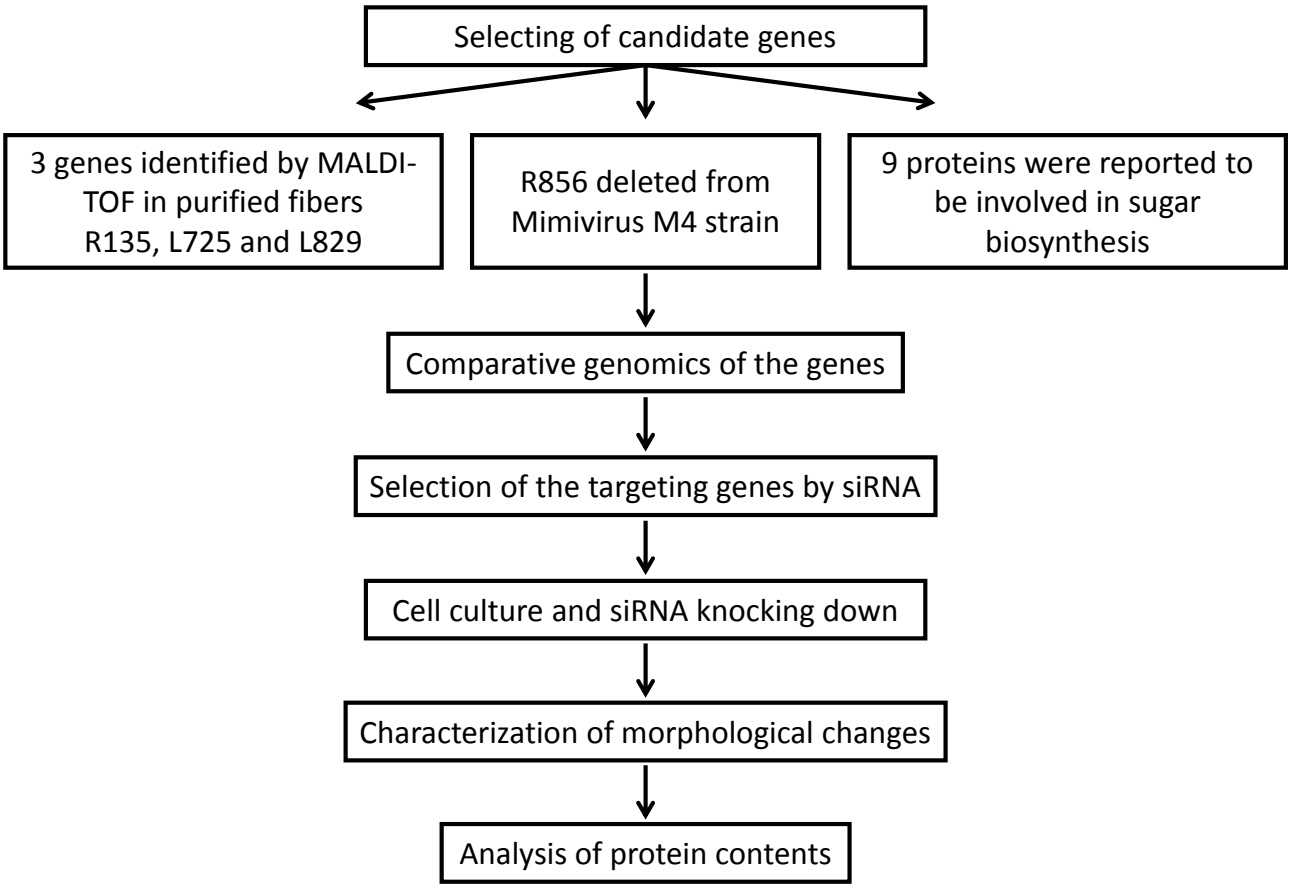
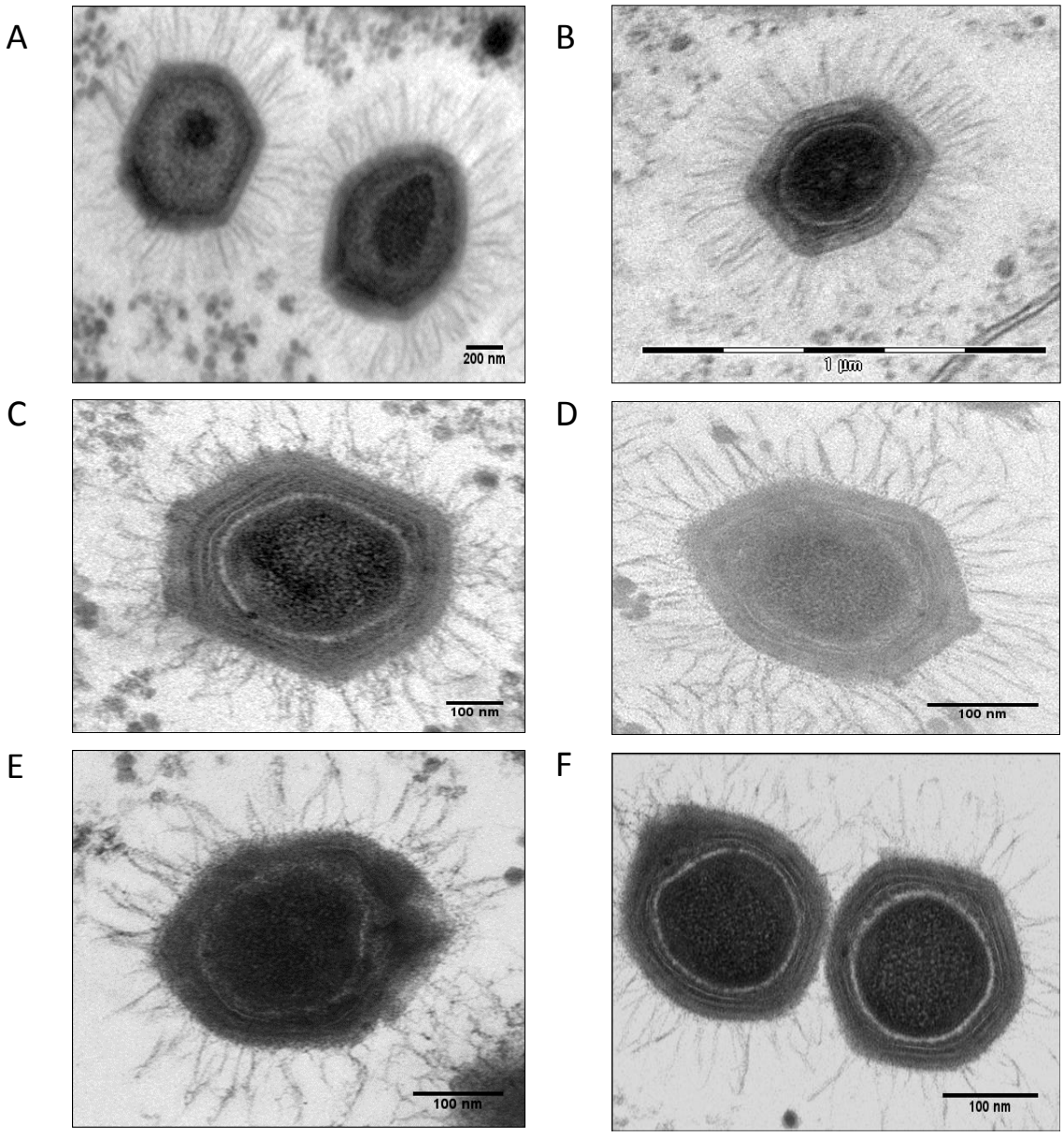
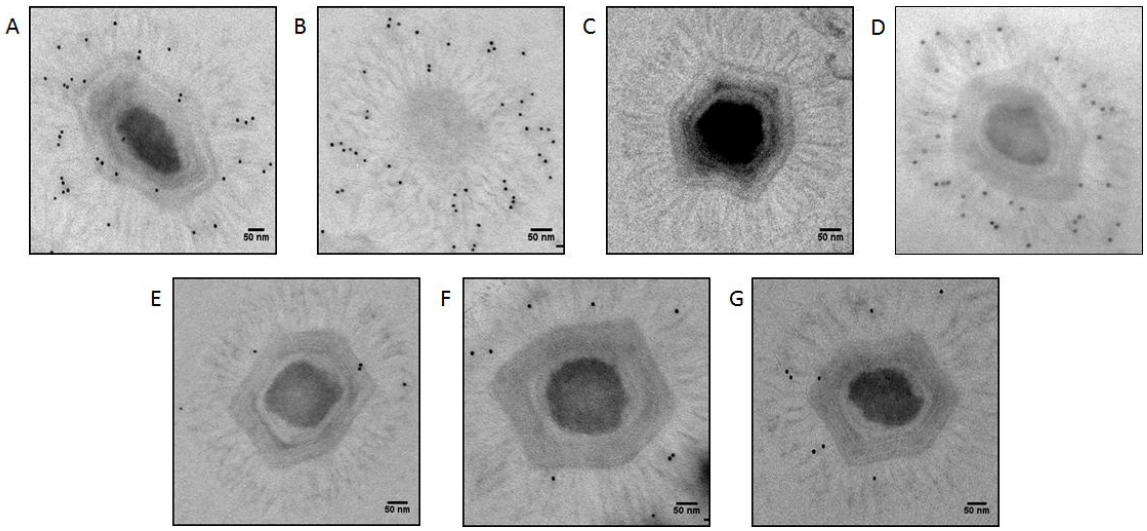


Figure 2. Electron micrographs of control Mimivirus and Mimivirus after knocking down genes encoding fiber-associated proteins.



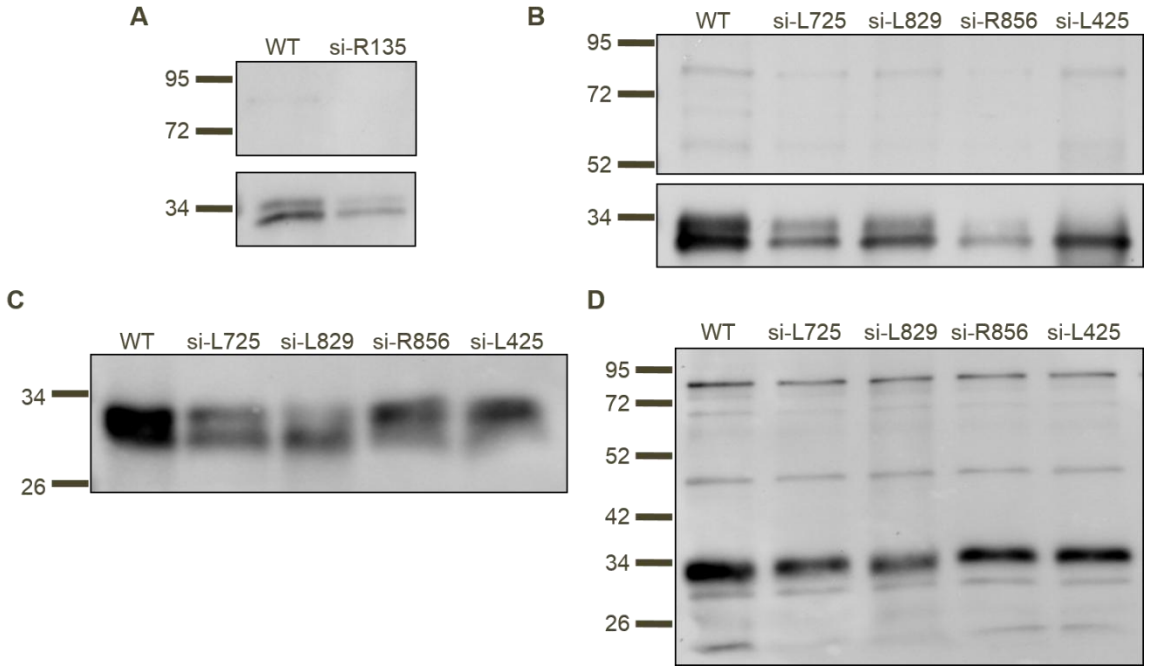
(A) Control; (B) Control si-R425 Mimivirus (meaning Mimvirus treated with siRNA targeting the R425 gene); (C) si-R856 Mimivirus; (D) si-L725 Mimivirus; (E) si-L829 Mimivirus; (F) si-R135 Mimivirus.

Figure 3. Electron microscopy with immunogold targeting of Mimivirus fiber proteins using anti-fiber antibodies (1:100).



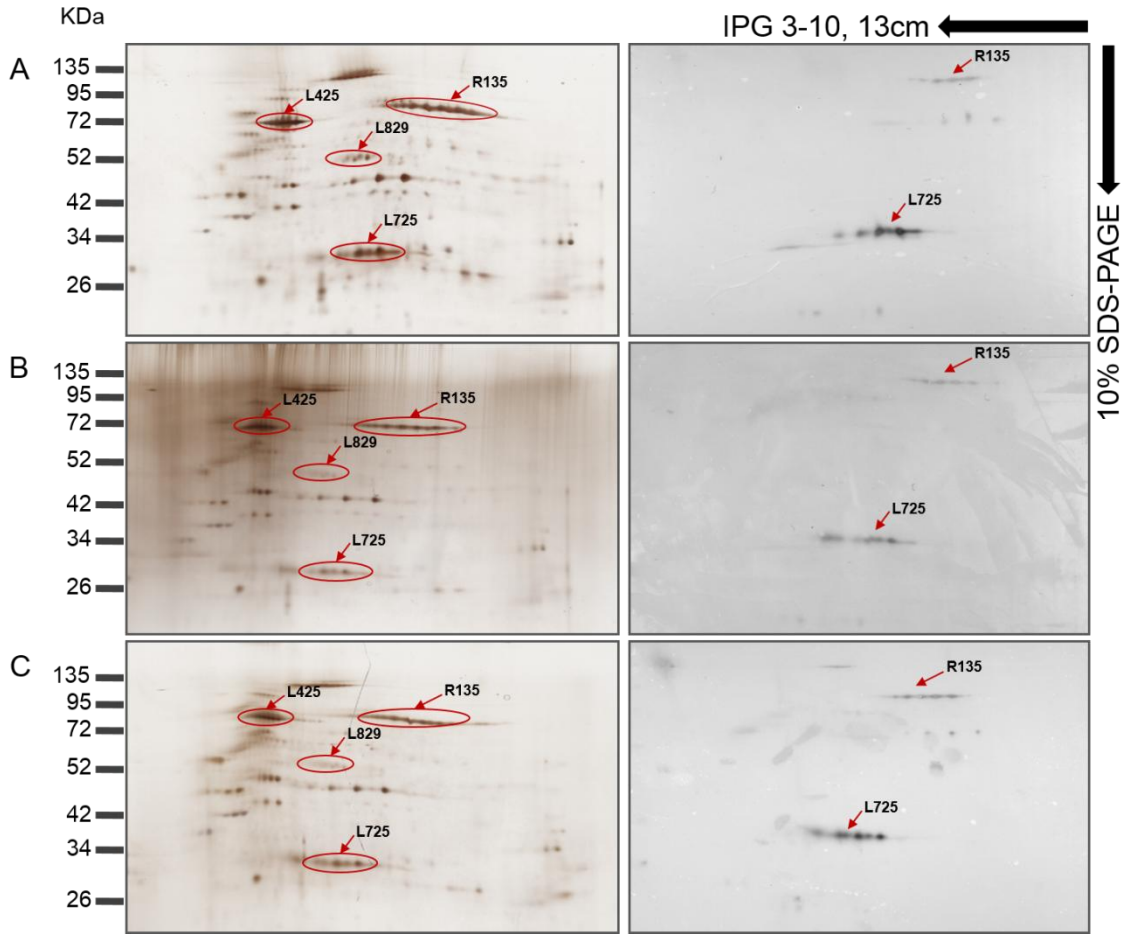
(A) Positive control; (B) Positive control (view from one of the capsid vertices); (C) Negative control with only secondary, not primary, antibodies; (D) si-L425 Mimivirus (meaning Mimvirus treated with siRNA targeting the L425 gene); (E) si-L725 Mimivirus; (F) si-L829 Mimivirus; (G) si-R856 Mimivirus.

Figure 4. Western blot analysis of fibers from silenced Mimiviruses and control Mimivirus.



(A) Reduction in anti-fiber antibodies (1:1000) binding to fiber proteins of si-R135; and (B) Reduction in anti-fiber antibodies (1:1000) binding to fiber proteins of si-L725, si-L829, si-R856 and si-L425 Mimiviruses. (C) Reduction in anti-L725 antibodies (1:5000) binding to si-L725, si-L829, si-R856 and si-L425 Mimiviruses. (D) Reduction in anti-Mimivirus antibodies (1:1000) binding to si-L725, si-L829, si-R856 and si-L425 Mimiviruses.
si- indicates the virus after silencing a target gene. Molecular weights are indicated on the left; Molecular weights for fiber-associated proteins are as follows: L725: 27 KDa; R856: 40 KDa; L829: 50 KDa; and R135: 77 KDa.

Figure 5. Mimivirus protein profiles as shown by 2D-gel electrophoresis and western blot with anti-fiber antibodies (1:5000).



(A) Silver stained and western blotted gel electrophoresis of Mimivirus in the absence of siRNA. (B) Silver stained and western blotted gel electrophoresis of si-L829 Mimivirus. (C) Silver stained and western blotted gel electrophoresis of si-R856 Mimivirus. The decrease in spot intensity indicates the depletion of fiber proteins. Immunoreactive protein spots are shown using arrows and the locus names; the spots were previously identified in (Boyer et al., 2011; Renesto et al., 2006).

Supplementary Information SI-1

Supplemental results

Table S1. The Excel "xls" spreadsheet contains the protein description, reference sequence (RefSeq accession), accession numbers, and full protein and gene names. The "BLASTp 15 aa Vs mimiviruses" spreadsheet includes the results from a BLASTp search of the 15 candidate proteins from a member of *Mimiviridae* against the entire proteome. "BLASTp of 6 prot vs nr" contains results from a BLASTp search of the four proteins involved in fiber formation and additionally two proteins that are conserved within *Mimiviridae* family against the NCBI nr database. "BLASTp of EFC vs Mimivirus aa" contains results from a BLASTp search of poxvirus entry fusion complex (EFC) proteins against Mimivirus proteins.

Table S2. The double-stranded RNA sequences for the tested Mimivirus genes.

Gene		Sense	Anti-sense	Position
R135	1	GCUGUUGUGGAUCGAAUUATT	UAAUUCGAUCCACAACAGCTT	936
	2	GCUACUUAUGGAGCAAAUUTT	AAUUUGCUCCAUAAGUAGCTT	1752
L725	1	UCUCCUUGAUGUGAAGUAUTT	AUACUUCACAUCAAGGAGATT	138
	2	GGUGUUCACAAGGGAAAUATT	UAUUUCCCUUGUGAACACCTT	592
R829	1	CCACAUCACAAUUUGGUAATT	UUACCAAAUUGUGAUGUGGTT	575
	2	GCGGCCAAUUAUAGAGAAATT	UUUCUCUAUAAUUGGCCGCTT	1087
R856	1	GCUUUAGGCGACUAUGAUATT	UAUCAUAGUCGCCUAAAGCTT	264
	2	CCAUAUUGAAACCGCAUUUTT	AAAUGCGGUUUCAAUAUGGTT	467
L425	1	GGUCACGUUGAAUUUGCUUTT	AAGCAAAUUCAACGUGACCTT	265
	2	CCAAGAACCUGGAGGUUAUTT	AUAACCUCCAGGUUCUUGGTT	993

Table S3. Sequences of the PCR primers and probes used for the tested Mimivirus genes. 1, forward; 2 reverse; and 3, probe.

Gene	Forward	Reverse	Probe
R135	5'-TGCAAACCAATTCCGTGTAA-3'	5'-AAATGACTATGTCCAGCTTCCA-3'	5'-CCCACACCAATTCCATGTGATCCC-3'
L136	5'-TCCATCCTCATCAATATCTACA-3'	5'-GGAACAAGTGCATTACATGCA-3'	5'-TTGAGTTCTCGTCCAAGTTGGCG-3'
R139	5'-CAAACCTGATCCAGAAAGAAAA-3'	5'-CGATCTAGCCAAATTTAATCCA-3'	5'-GCTGGTGCAAGCGCCATGAA-3'
L142	5'-TTTCCTTTTCCAATAACAACACT-3'	5'-AATCTGAGCCATTTTTTTGTACA-3'	5'-TCCATTCAACATTCCCAACTGTTTG-3'
R641	5'-AAATAATATCCACAAGCTTCCAA-3'	5'-TCCAGAATTAGATTTCAAGATTG-3'	5'-AATCCACCAGTCAATCCAGATTCAA-3'
L725	5'-TGGTGCAACTTATCTTCGATCA-3'	5'-CAAGGAGACGATGGTTGACA-3'	5'-TTGGCCGCGAAAGATGAATGG-3'
L829	5'-CTATAAAGACGCTTTCCCGTT-3'	5'-GAAGCTAACCTAAGATCATCAT-3'	5'-CGAAACAGAAATGTCCGGAGTGC-3'
R856	5'-GCATTAATGGGGATTGCTTCA-3'	5'-CAAATGCGGTTTCAATATGGTT-3'	5'-AAGGAAATTACGATGAGGCACTCTC-3'
mg878	5'-AGGCTGGTTCATGGCATTTA-3'	5'-CAACACCCATACCATGAGCA-3'	5'-TGGCAAGCACCTCTCCAACCTG-3'

Supplemental figures

Figure S1. Representative figure for wild type virus (WT) with normal, stretched and long fibers that form a dense layer, which is used as experimental control. The average length is approximately 130 nm / fiber.

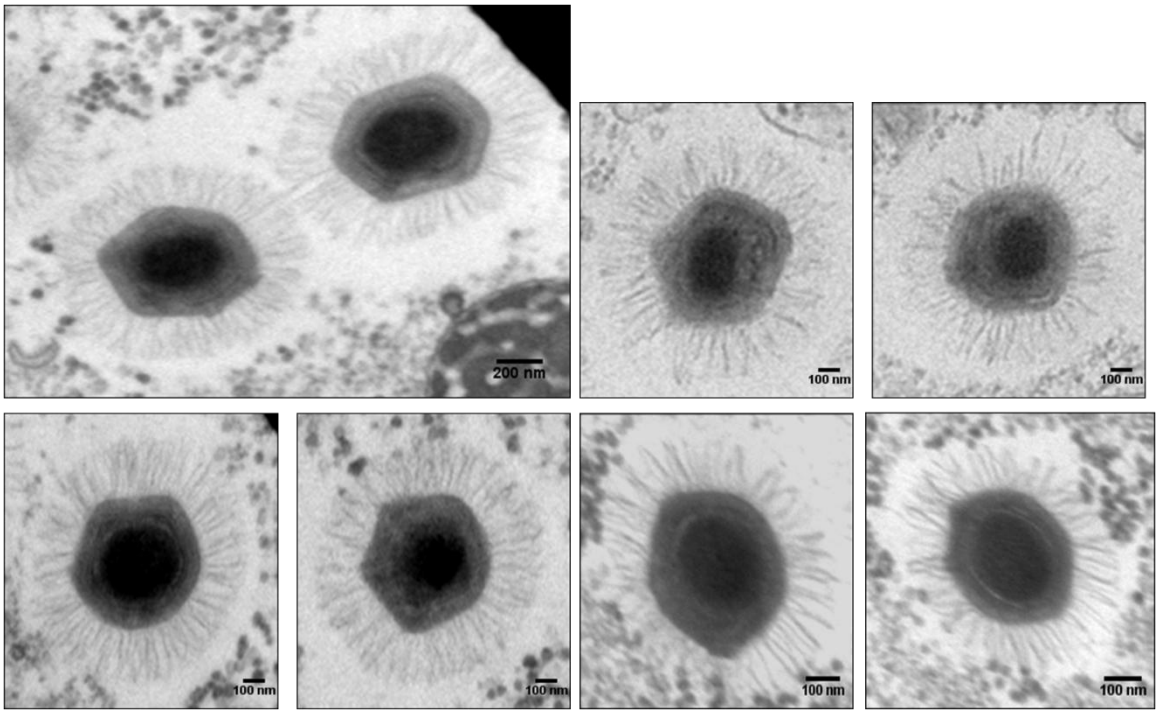


Figure S2. Representative figure for shape of the fibers after silencing the L425 capsid protein, control virus.

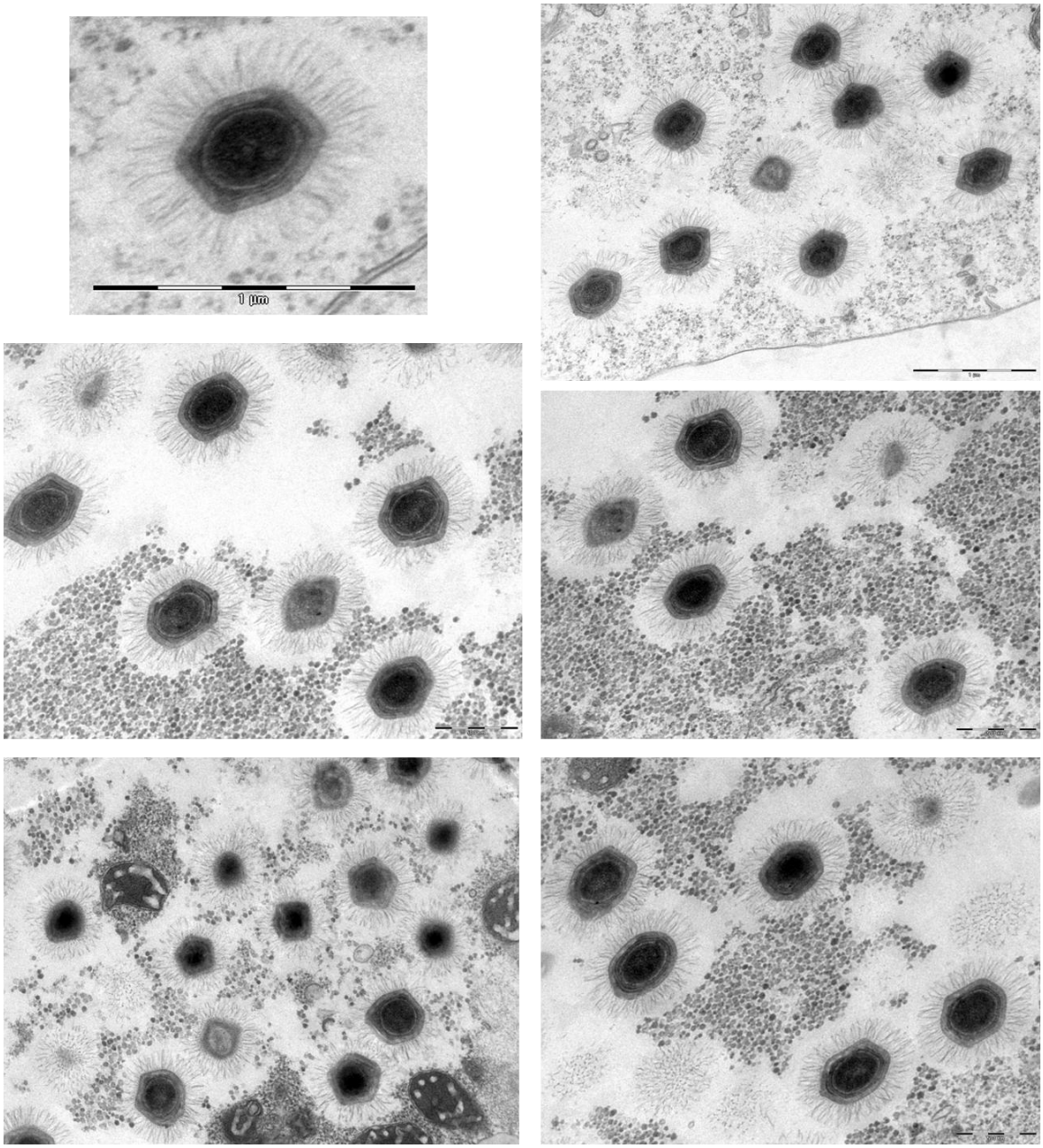


Figure S3. Representative figure for shape of the fibers after silencing the R856 protein.

The fibers are very short (<50 nm / fiber) and only 30% of the normal length of fibers.

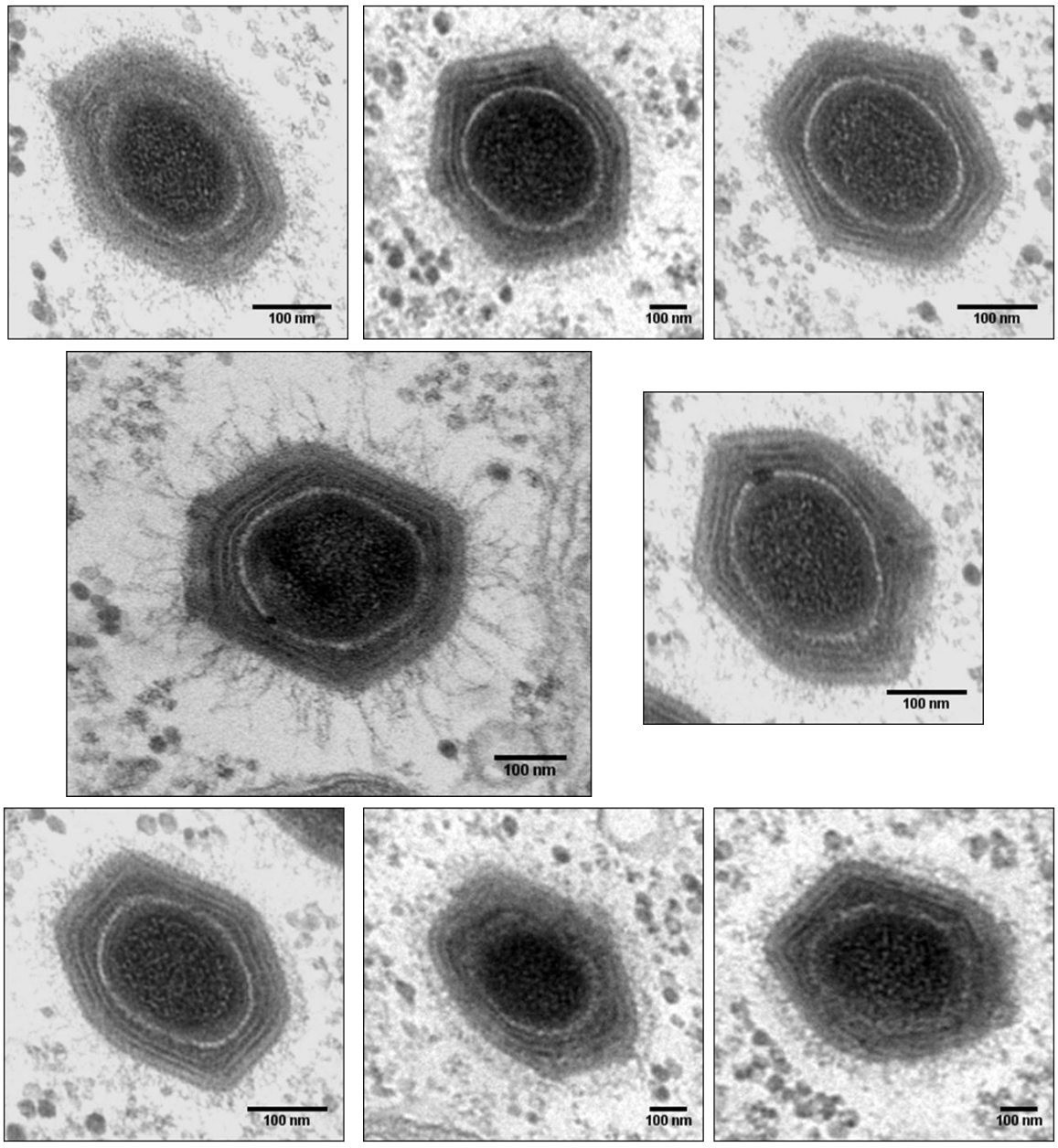


Figure S4. Representative figure for shape of the fibers after silencing the L725 protein.

The fiber layer is less dense, and the fibers are curved and shorter than those of the normal virus by approximately 15%.

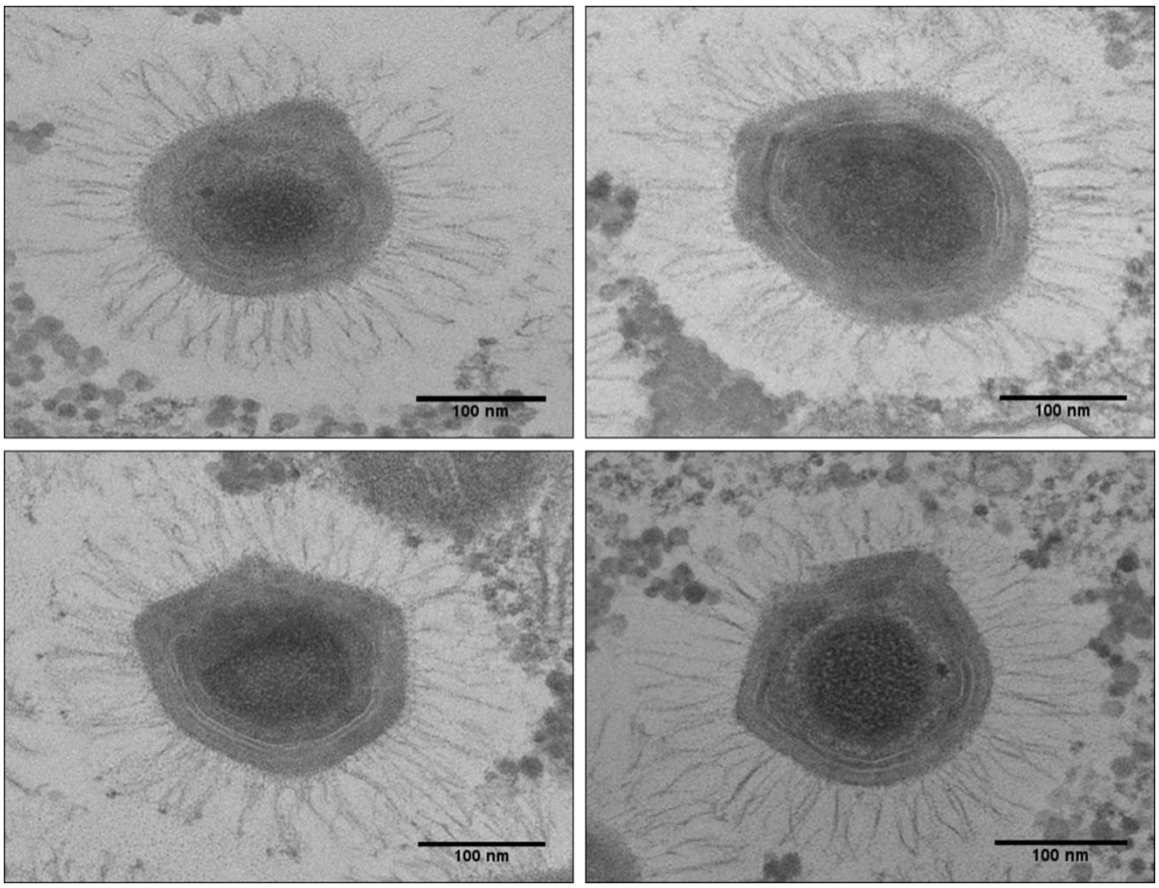


Figure S5. Representative figure for shape of the fibers after silencing the L829 protein.

The fiber layer is less dense. The fibers are curved, not stretched, and shorter than those of the normal virus by approximately 30%.

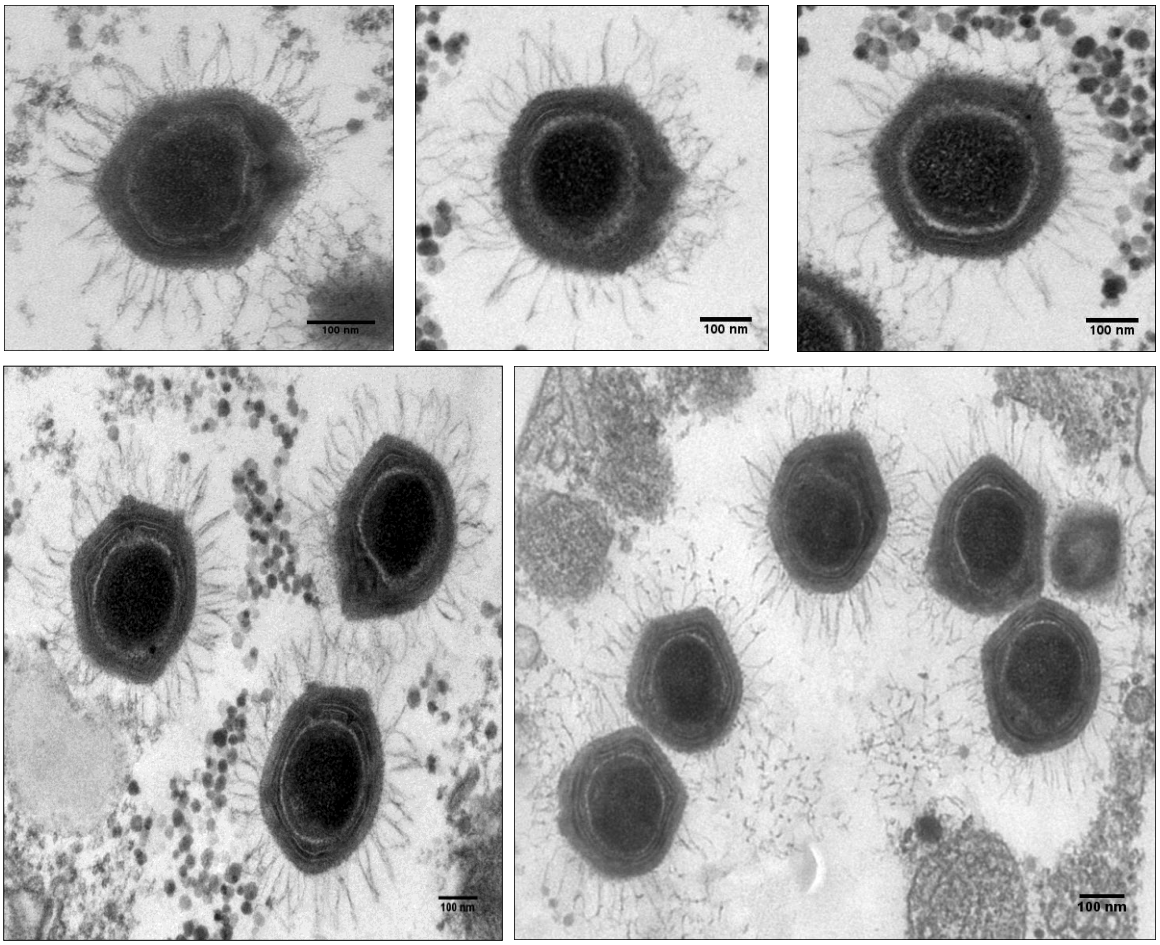


Figure S6. Representative figure for shape of the fibers after silencing the R135 protein.

The fiber layer is less dense, and the fibers are curved and shorter than those of the normal virus by approximately 12%.

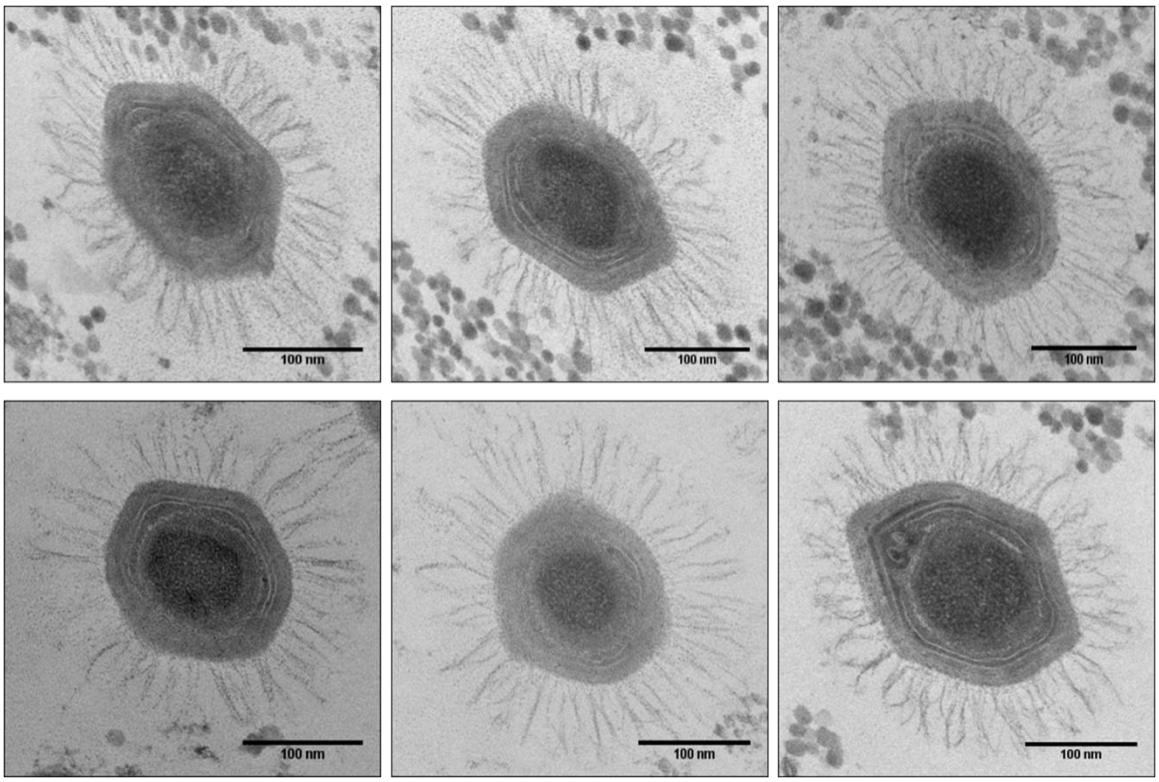


Figure S7. The protein concentration (μg/μl) after silencing two genes (L829 and R856). The first bar is control, second is si-L829 and third is si-R856.

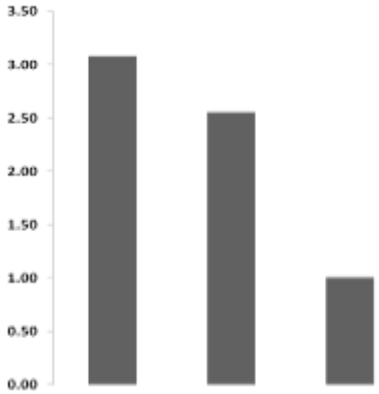


Figure S8. The MALDI-TOF MS biotyping for si-L829 (red spectrum, four replicates) that contains an extra-band compared to control (blue spectrum).

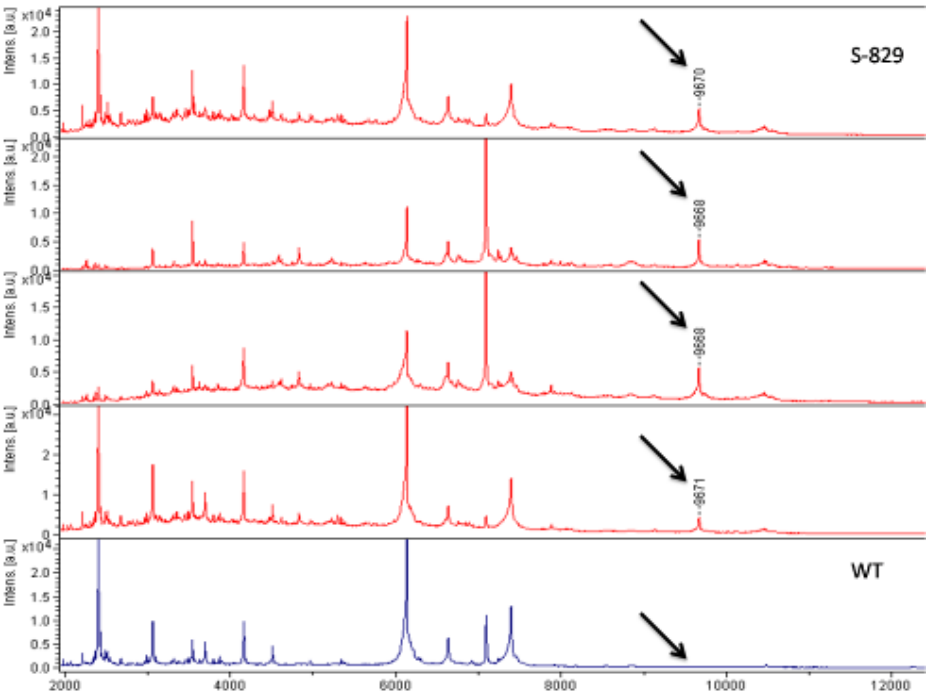
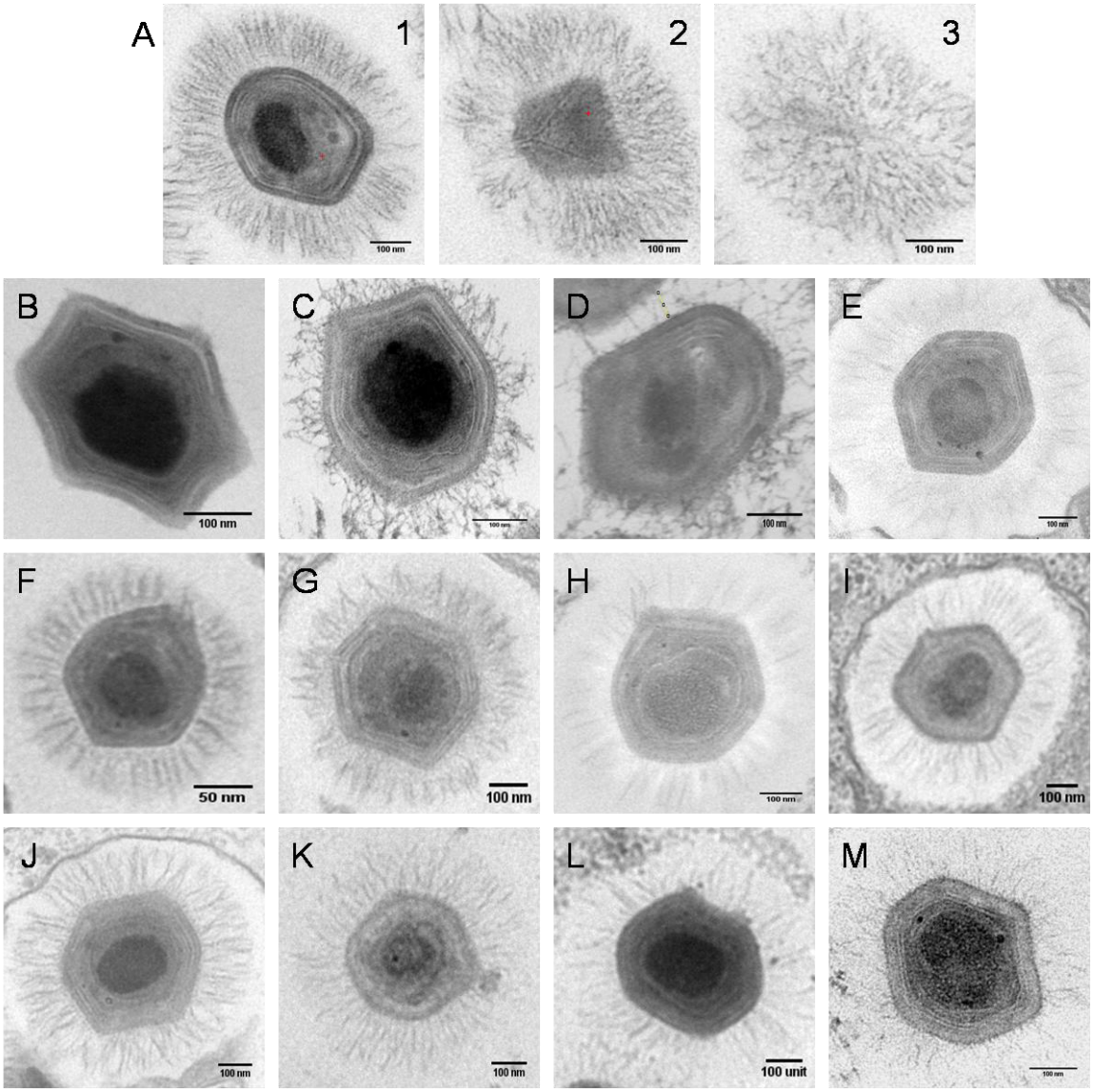
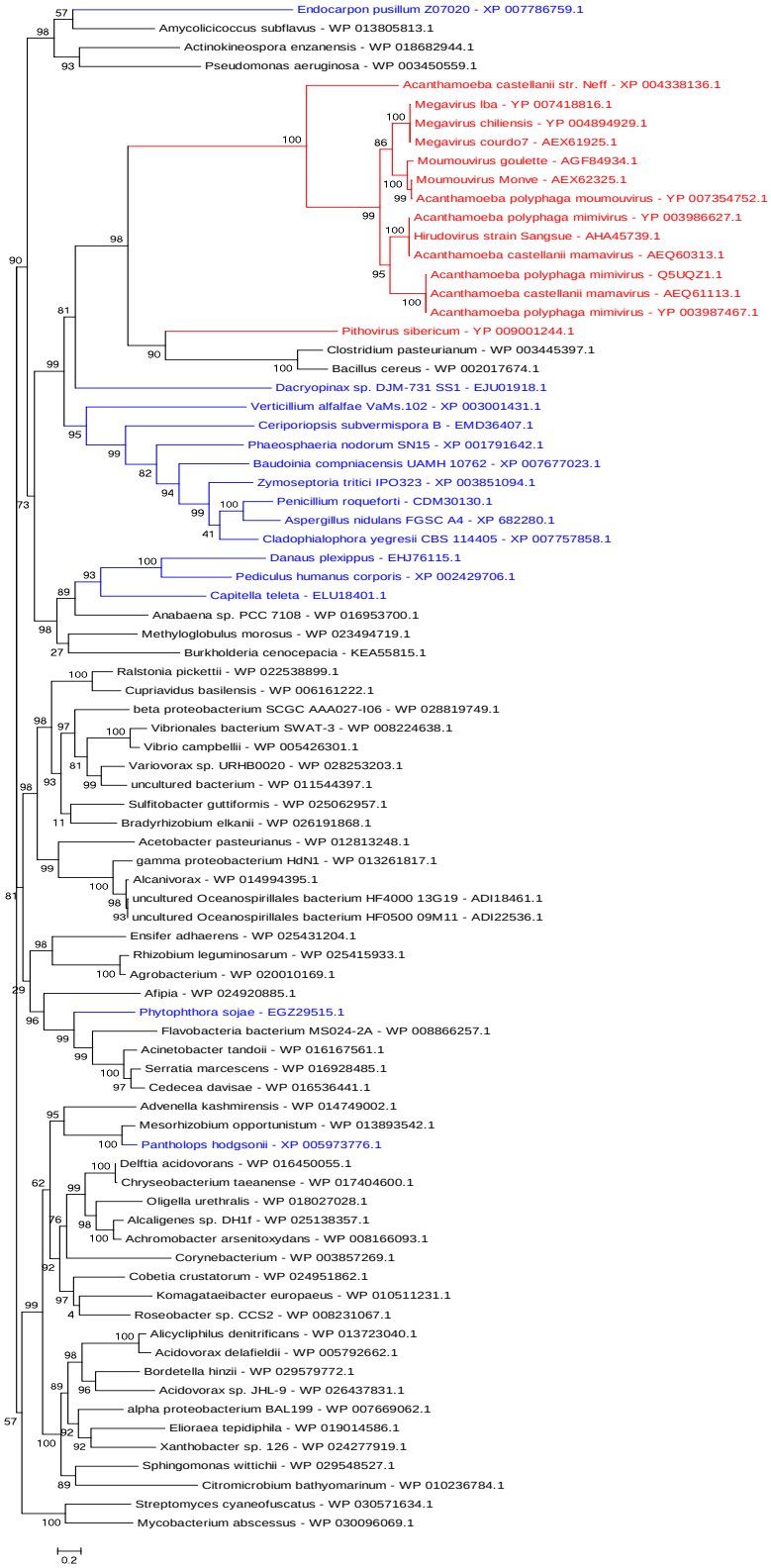


Figure S9. Electron micrograph showing fiber shape, structure, and thickness in mimiviruses.



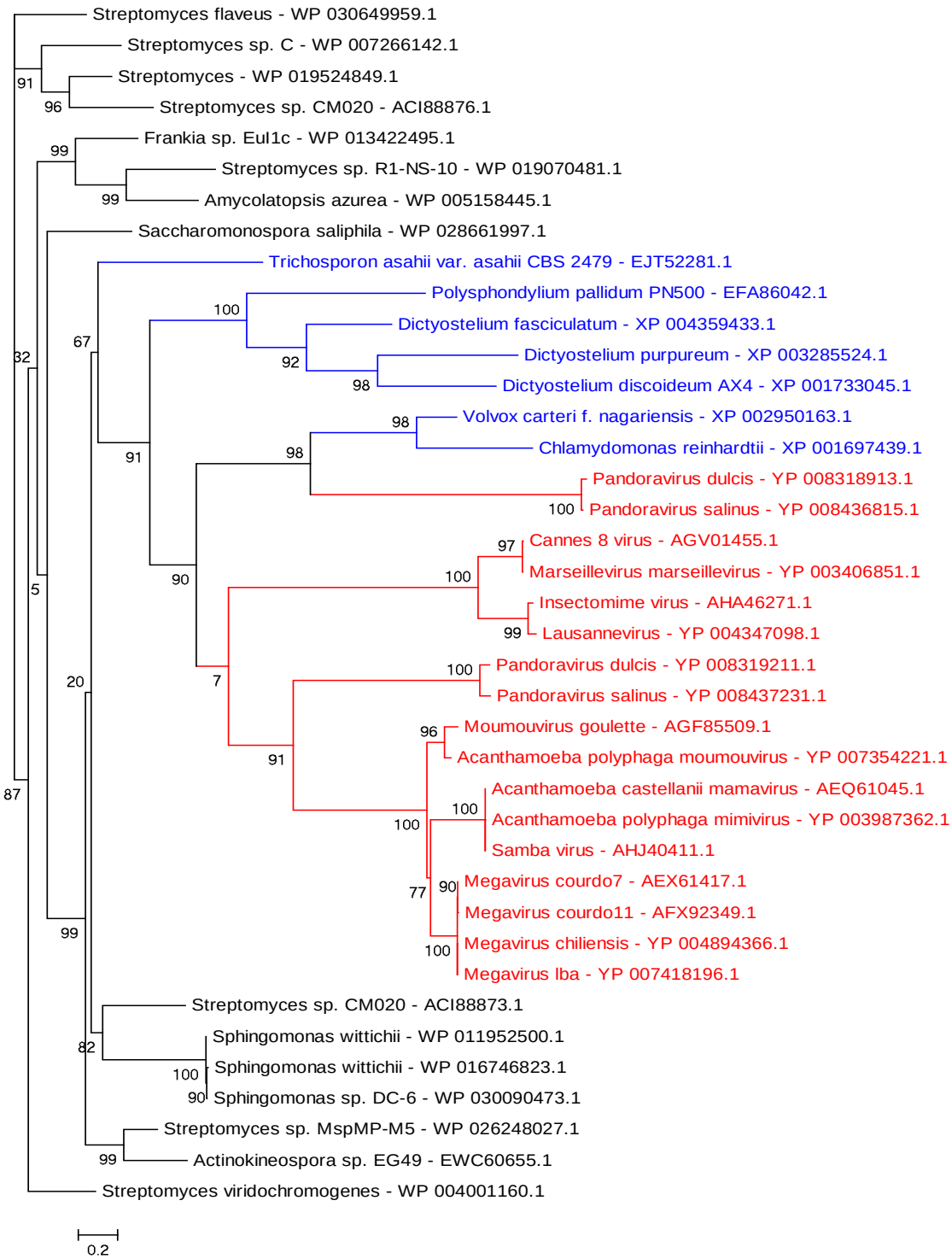
Mimivirus has a large and dense layer of fibers; the M4 strain has no fibers; Mamavirus and Lentillevirus have a short and thin layer of fibers.
(A) Mimivirus (1: transverse section of the virion; 2: star-shaped capsid vertex with no fiber; 3: top view of one of the capsid vertices showing the fibers); (B) Mimivirus M4 strain; (C) Mamavirus; (D) Lentillevirus; (E) Montpellier virus 3; (F) Courdo7 virus; (G) Courdo11 virus; (H) Terra1 virus; (I): Terra2 virus; (J) Pointe-rouge1 virus; (K) Pointe-rouge2 virus; (L) Longchamp virus; (M) Lactours2 virus.

Figure S10. Phylogenetic tree reconstruction using Maximum likelihood method for R135 protein (FAP1).



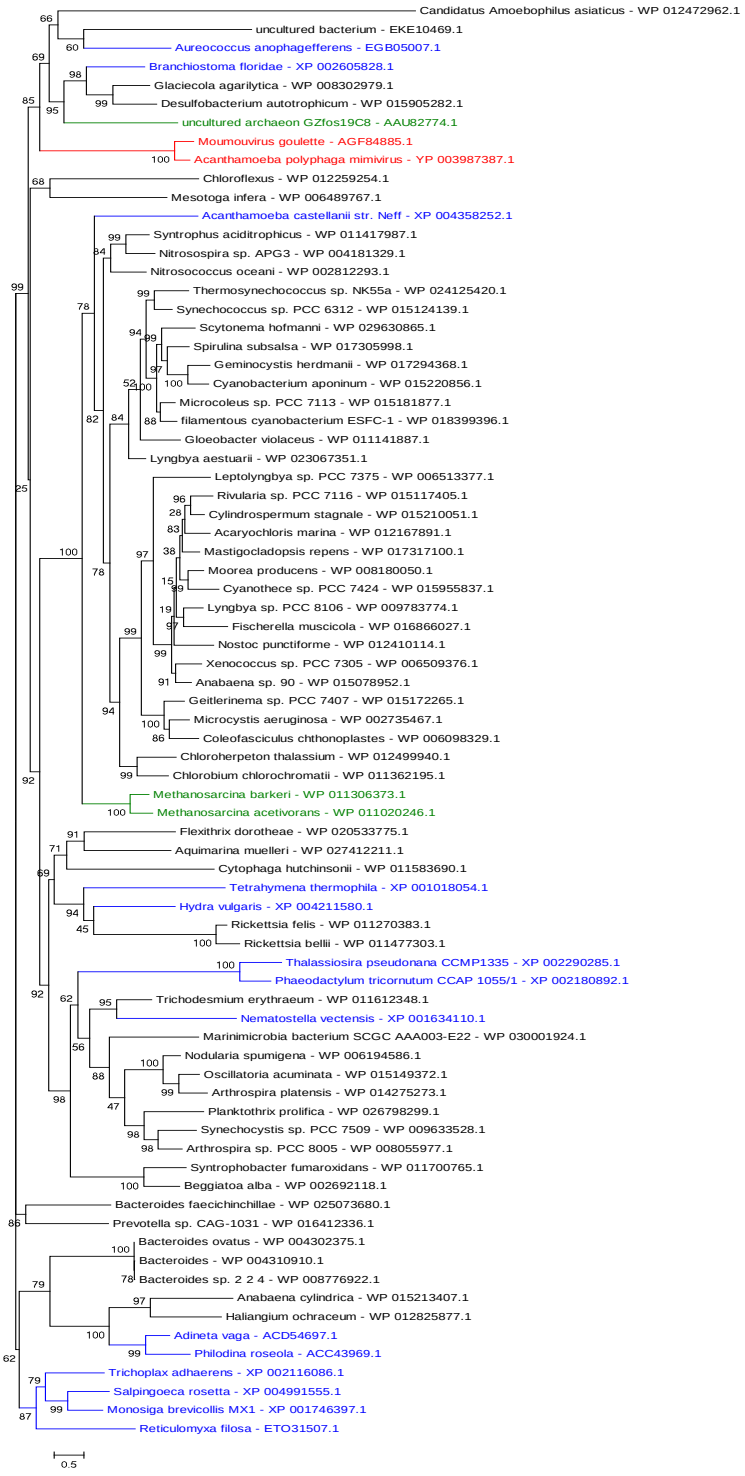
Virus, Bacteria and Eukarya are shown in red, black and blue respectively. The number represents the bootstrapping values and the scale bar represents the number of estimated changes per position. The branch label shows the taxon name and protein GenBank accession number.

Figure S11. Phylogenetic tree reconstruction using Maximum likelihood method for L829 protein (FAP3).



Virus, Bacteria and Eukarya are shown in red, black and blue respectively. The number represents the bootstrapping values and the scale bar represents the number of estimated changes per position. The branch label shows the taxon name and protein GenBank accession number.

Figure S12. Phylogenetic tree reconstruction using Maximum likelihood method for R856 protein (FAP4).

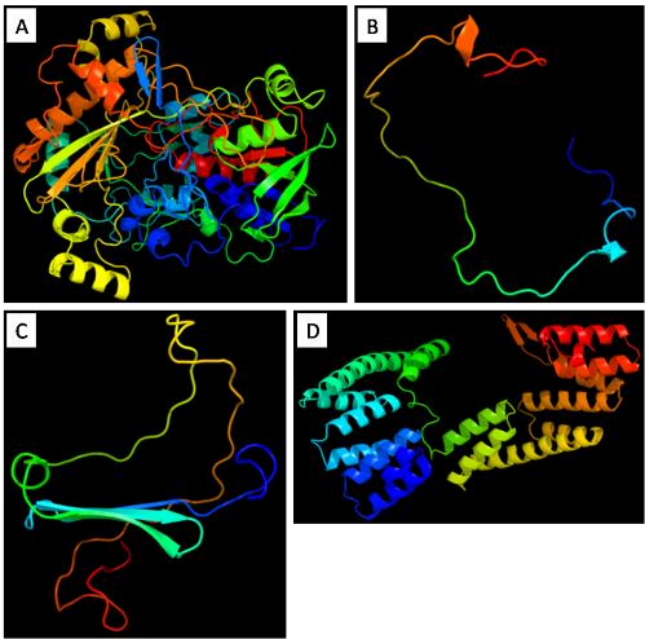


Virus, Bacteria, Eukarya and Archaea are shown in red, black, blue and green respectively. The number represents the bootstrapping values and the scale bar represents the number of estimated changes per position. The branch label shows the taxon name and protein GenBank accession number.

Figure S13. The R856 protein contains seven TPR motifs; each consisting of 34 residues (yellow shadows). Source, UniProtKB; entry number, Q5UQQ7.

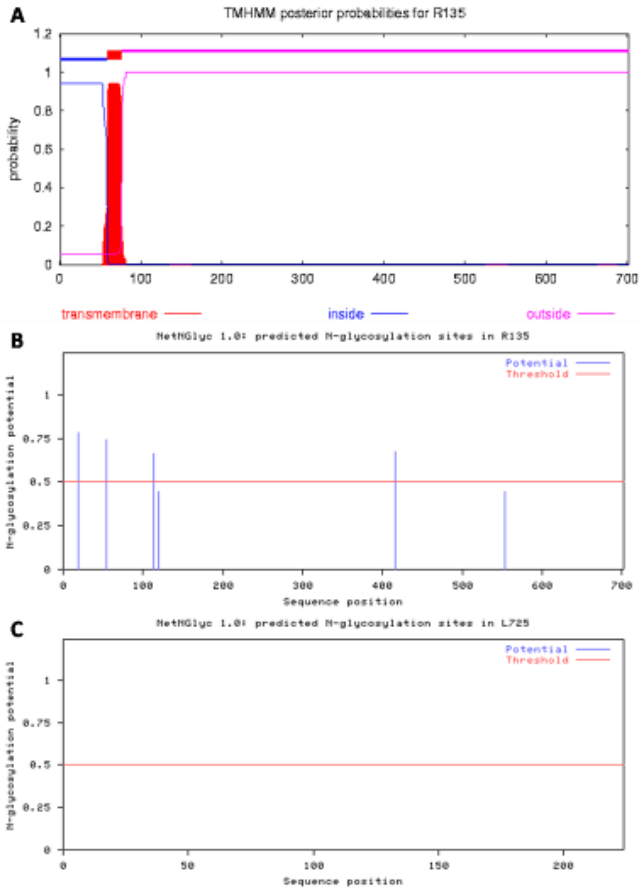
```
        10         20         30         40         50         60
MEIINACLKN VGVEENMIEK VIETFQKEND ECYNLVHIFN KAAIVFHRNG QHKKSLEMYE

        70         80         90        100        110        120
KAFGNIFNGE FALSDLFYSV NGMASMYQAL GDYDIAIKKY NSVIKIIKDM CLDNNSDLVY

       130        140        150        160        170        180
ALMGIASISQ IKGNYDEALS KYNEALEINE KLYGRNHIET AFVLNRLGML YHELDDNDKS

       190        200        210        220        230        240
IDHFNESLKI YREKYPNKLF NIAFTISRLA QSLLKMGNDS EALEKYQESI DIFNKIFTIS

       250        260        270        280        290        300
HQAVAFSLYG IGTVYEFRSE YSKALEKYQE SLQTYKNVYE RSEKYQHYDI ASCLYKIGLV

       310        320        330        340
YKLSGNDNES TTYLNQANQM FESTSTNIND KNYQACKKFL QD
```

Figure S14. The protein structure using the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2).



L725 structure is difficult to be predicted.
(A) The predicted structure of the R135 protein; confidence = 100.0%, coverage = 80%; (B) The predicted structure of the L725 protein; confidence = 28.0%, coverage = 22%; (C) The predicted structure of the L829 protein; confidence = 90.0%, coverage = 19%; (D) The predicted structure of the R856 protein; confidence = 99.9%, coverage = 90%.

Figure S15. Trans-membrane and glycosylation sites of fiber associated proteins.



(A) R135 harbors a transmembrane domain; predicted using the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/). (B) R135 is glycosylated, as well as L829 and R856 (not shown); whereas, (C) L725 is not glycosylated; predicted using the NGlyc server (http://www.cbs.dtu.dk/services/NetNGlyc/).

Research article no. 2

Study of the effect silencing mimivirus fiber protein on sputnik virophage infection

Introduction

Virophages are viruses that infect giant viruses within their host. They have double-stranded DNA genome that are about 18-19 kilobase pairs in length and encode about 20 proteins (La Scola, Desnues et al. 2008, Fischer and Suttle 2011, Krupovic and Cvirkaite-Krupovic 2011, Desnues, Boyer et al. 2012). The first discovered virophage was Sputnik, which was isolated with Mamavirus (La Scola, Desnues et al. 2008). Then, Mavirus was isolated with Cafeteria roenbergensis virus (Fischer and Suttle 2011), and the Organic Lake virophage was isolated from Organic Lake in Australia with Organic Lake phycodnaviruses, first described as phycodnaviruses but later reclassified as mimiviruses (Yau, Lauro et al. 2011). Afterward, Sputnik 2 and Sputnik 3 virophages (Gaia, Pagnier et al. 2013), and Zamilon (Gaia, Benamar et al. 2014), were also isolated. Structurally, Sputnik virions have mushroom-like protrusions at their surface, which are attached to capsomers and could function in recognizing the host (Sun, La Scola et al. 2010, Zhang, Sun et al. 2012).

Interestingly, Sputnik was detected in association with purified Mimivirus fibers and was seen in electron microscopy linked to these fibers (Desnues and Raoult 2010). Therefore, to date, virophages are only known in association with mimiviruses. Moreover, the Mimivirus M4 strain that emerged after 150 passages on germ-free amoebae and is devoid of fibers resists to infection by the Sputnik virophage infection, which suggests the role of the Mimivirus fibers in virophage replication (Boyer, Azza et al. 2011).

The way by which virophages enter their amoebal host and can enter into mimiviruses is worthy to be elucidated. Moreover, antibodies to the Sputnik virophage were detected in two patients returning from Laos, and one of them exhibited to seroconversion (Parola, Renvoise et al. 2012). Here, we aimed to study the role of fiber proteins in Sputnik virophage replication by silencing the genes previously shown to be associated with fibers.

Materials and methods

Cell culture and RNA interference. The method was previously described in [chapter 2, article no. 1] (Figure 1). Briefly, a culture of A. polyphaga in 10 ml of PYG medium was seeded for 24-48 h. Then, 100 µl of Lipofectamine RNAiMAX (Invitrogen, USA) and 0.25 µg of duplex siRNA (designed and purchased from Invitrogen) were used according to the manufacturers' instructions. Then, Mimivirus was harvested and purified, referred here as first generation Mimivirus (Mimi_1).

For Sputnik virophage co-infection, these purified viruses as well as Sputnik 3 virophage were introduced in a culture of A. polyphaga in 10 ml of PYG medium. After 24 h and complete lysis of the amoebae, cell debris was eliminated by centrifugation at 2,000 rpm for 10 min and filtration using filters with pores 1.2 µm in diameter. The Mimivirus (second generation Mimivirus; Mimi_2) was harvested by centrifugation at 10,000 rpm for 10 min, followed by filtration using filter with 0.22 µm in diameter and ultracentrifugation at 20,000 rpm for 60 min, to harvest the Sputnik virophages. The purified viruses were conserved in PBS for further analysis.

PCR testing. The by quantitative PCR (qPCR) was applied to determine the DNA in the tested virophages compared to control one. The primers and probe were designed using the Gemi tool (Sobhy and Colson 2012) (Table 1). The 20 µl real-time PCR mixture contained 2.5 µl of extracted DNA, 10 µl qPCR Takyon master mix (Eurogentec, Belgium), 2 µl of each primer (10 nmol/µl; Eurogentec), 2 µl probe (3 nmol/µl; Applied Biosystems UK). The PCR thermal cycles included a hold at 50°C for 2 min, a hold at 95°C for 3 min, then 40 cycles comprising 10 sec at 95°C then 1 min at 55°C, and were performed on the CFX96 Touch instrument (Bio-Rad, USA).

Enzyme-linked immunosorbent assay (ELISA). 100 µl of viruses were coated in a flat-bottomed 96-well ELISA microplate and incubated overnight at 4ºC, then washed four times with a washing buffer (PBS-Tween 0.1%) and incubated with blocking buffer (PBS with 5% bovine serum album (BSA; Sigma-Aldrich, USA)) for 60 min and washed three times. The viruses were then incubated with primary anti-Sputnik virophage antibodies diluted 1:1000 in PBS-Tween 0.1%-BSA 5% for 60 min, and washed, then incubated for 1 hour with secondary horse-radish peroxidase-conjugated anti-mouse IgG at a 1:5000 dilution (Jackson ImmunoResearch, UK). Detection was performed at 490 nm using o-phenylenediamine dihydrochloride substrate (Sigma-Aldrich, USA). For negative controls, the first and second

controls were manipulated with the same procedure except that virus sample and primary antibodies, respectively, were not added.

Proteomic analysis. All steps from the proteomic analyses (including sample preparation, 1D gel electrophoresis and western blotting) were performed as previously described [chapter 2, article no. 1]. The area of bands were measured, allowing semi-quantification of protein concentrations, using ImageJ and Image Studio Lite (IS Lite, USA) tools (http://imagej.nih.gov/ij/ and http://www.licor.com/).

Results

Mimiviruses with silenced fiber proteins were first obtained (first replicate). The silenced Mimiviruses were then co-cultured on amoebae in co-infection with the Sputnik virophage. Sputnik replication was studied using qPCR, ELISA, and immunoblot, to assess differences in the quantity of virophage with or without prior silencing of Mimivirus genes encoding fiber associated proteins.

Silencing of the Mimivirus genes R135, L725, L829, and R856 encoding proteins involved in fiber formation was confirmed by western blotting using anti-Mimivirus fiber antibodies (Figure 2). The western blot showed that these antibodies bound to purified Mimivirus and that the intensity of the bands was lower when any of these genes was silenced using short interfering RNA, which indicates that these genes were successfully silenced (Figure 2).

Quantification of the Sputnik virophage DNA by real-time PCR showed that an increase in concentrations of Sputnik in presence of silenced than non-silenced (control) Mimivirus. In addition, the titer of Sputnik in co-infection with Mimivirus after silencing of its R856 gene was higher than the titer of Sputnik in co-infection with control Mimivirus or Mimivirus after silencing of the other fiber-associated genes (Figure 3). To validate the qPCR results, we performed ELISA and western blot, which showed relative increases of the concentration of the Sputnik virophage in co-infected with silenced compared to non-silenced Mimivirus (Figure 4 and 5). Indeed, antibody binding to the Sputnik virophage was increased by about 1.5 time after silencing R135 and L725 genes, and 1.8 time after silencing the R856 gene, which indicates that the titer of the Sputnik virophage was increased after silencing the fiber-associated genes from Mimivirus.

Conclusion

This preliminary study conducted here is the first to combine RNA interference with Mimivirus-virophage co-infection of *Acanthamoeba* spp. The results indicate that silencing of the Mimivirus genes encoding fiber associated proteins impacted the Sputnik replication. The results observed here provide further evidence of the role of proteins previously identified as involved in Mimivirus fiber formation in the life cycle of the Mimivirus virophage (Boyer, Azza et al. 2011). Nevertheless, the increase in concentrations of the Sputnik virophage measured in co-infection with Mimivirus after silencing of four of its genes suspected to be involved in fiber formation was unexpected. Indeed, in a previous study, the bald Mimivirus M4 strain showed resistance to infection by the Sputnik virophage (Boyer, Azza et al. 2011). The increase in Sputnik replication observed here could owe to at least two main reasons including the increase in virophage uptake by the host amoebal cells, or other steps of the virophage life cycle within the host. Further studies, including transcriptomic as well as proteomic analyses concurrently with RNA interference, are needed to elucidate host-pathogen interactions between Sputnik, Mimivirus and Acanthamoeba host.

References

Boyer, M., S. Azza, L. Barrassi, T. Klose, A. Campocasso, I. Pagnier, G. Fournous, A. Borg, C. Robert, X. Zhang, C. Desnues, B. Henrissat, M. G. Rossmann, B. La Scola and D. Raoult (2011). "Mimivirus shows dramatic genome reduction after intraamoebal culture." Proc Natl Acad Sci U S A **108**(25): 10296-10301.

Desnues, C., M. Boyer and D. Raoult (2012). "Sputnik, a virophage infecting the viral domain of life." Adv Virus Res **82**: 63-89.

Desnues, C. and D. Raoult (2010). "Inside the lifestyle of the virophage." Intervirology **53**(5): 293-303.

Fischer, M. G. and C. A. Suttle (2011). "A virophage at the origin of large DNA transposons." Science **332**(6026): 231-234.

Gaia, M., S. Benamar, M. Boughalmi, I. Pagnier, O. Croce, P. Colson, D. Raoult and B. La Scola (2014). "Zamilon, a novel virophage with mimiviridae host specificity." PLoS One **9**(4): e94923.

Gaia, M., I. Pagnier, A. Campocasso, G. Fournous, D. Raoult and B. La Scola (2013). "Broad spectrum of mimiviridae virophage allows its isolation using a mimivirus reporter." PLoS One **8**(4): e61912.

Krupovic, M. and V. Cvirkaite-Krupovic (2011). "Virophages or satellite viruses?" Nat Rev Microbiol **9**(11): 762-763.

La Scola, B., C. Desnues, I. Pagnier, C. Robert, L. Barrassi, G. Fournous, M. Merchat, M. Suzan-Monti, P. Forterre, E. Koonin and D. Raoult (2008). "The virophage as a unique parasite of the giant mimivirus." Nature **455**(7209): 100-104.

Parola, P., A. Renvoise, E. Botelho-Nevers, B. La Scola, C. Desnues and D. Raoult (2012). "Acanthamoeba polyphaga mimivirus virophage seroconversion in travelers returning from Laos." Emerg Infect Dis **18**(9): 1500-1502.

Sobhy, H. and P. Colson (2012). "Gemi: PCR primers prediction from multiple alignments." Comp Funct Genomics **2012**: 783138.

Sun, S., B. La Scola, V. D. Bowman, C. M. Ryan, J. P. Whitelegge, D. Raoult and M. G. Rossmann (2010). "Structural studies of the Sputnik virophage." J Virol **84**(2): 894-897.

Yau, S., F. M. Lauro, M. Z. DeMaere, M. V. Brown, T. Thomas, M. J. Raftery, C. Andrews-Pfannkoch, M. Lewis, J. M. Hoffman, J. A. Gibson and R. Cavicchioli (2011). "Virophage control of antarctic algal host-virus dynamics." Proc Natl Acad Sci U S A **108**(15): 6163-6168.

Zhang, X., S. Sun, Y. Xiang, J. Wong, T. Klose, D. Raoult and M. G. Rossmann (2012). "Structure of Sputnik, a virophage, at 3.5-A resolution." Proc Natl Acad Sci U S A **109**(45): 18431-18436.

Table 1. qPCR oligos for the capsid encoding gene of the Sputnik virophage

Oligo sequence	**Nucleotide position in the capsid gene**	**Type**	**Oligo length (nucleotides)**
TATCACTTTTACTGCCAATCCATCTCATG	221-250	Forward primer	29
CCTGGTCGTGATGCATTCCGTGCATTC	274-304	Probe	27
GCAAGTTCAATATTCACAGGGAAACCATTT	341-371	Reverse primer	30

Figure 1. Representative figure for method used to assess the role of fibers in Sputnik virophage replication.

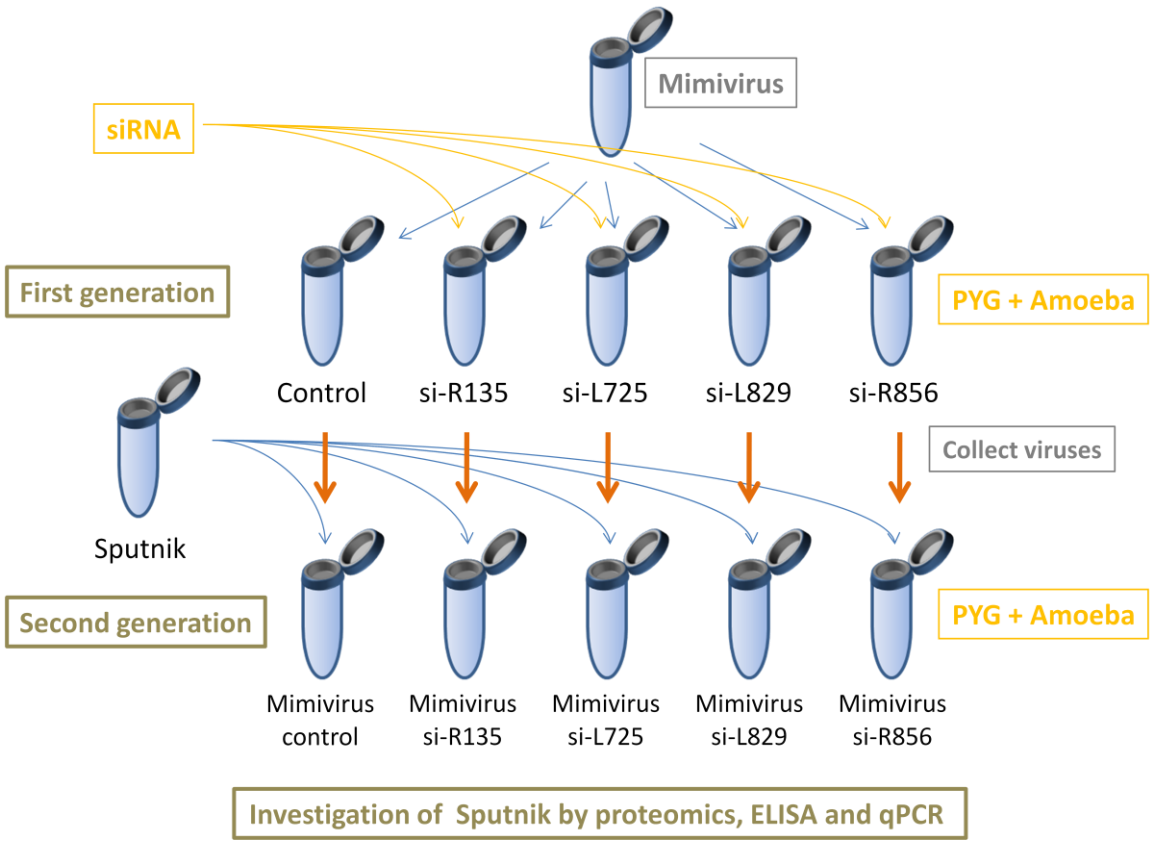
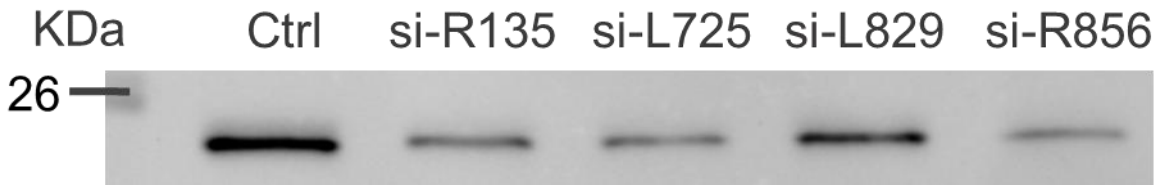
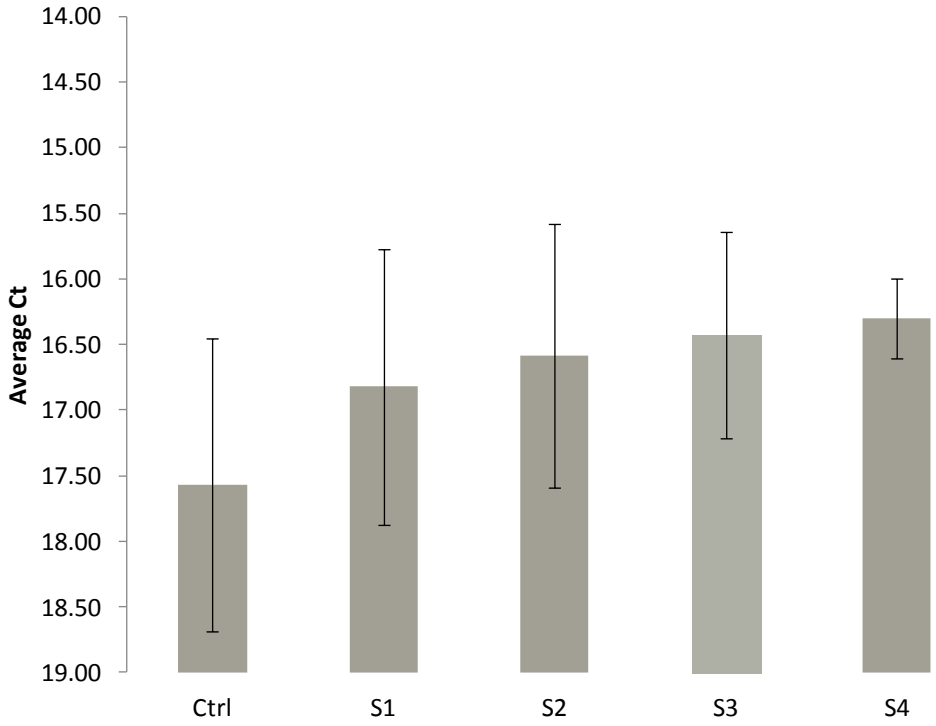


Figure 2. Western blot analysis of Mimivirus proteins targeted by siRNA.



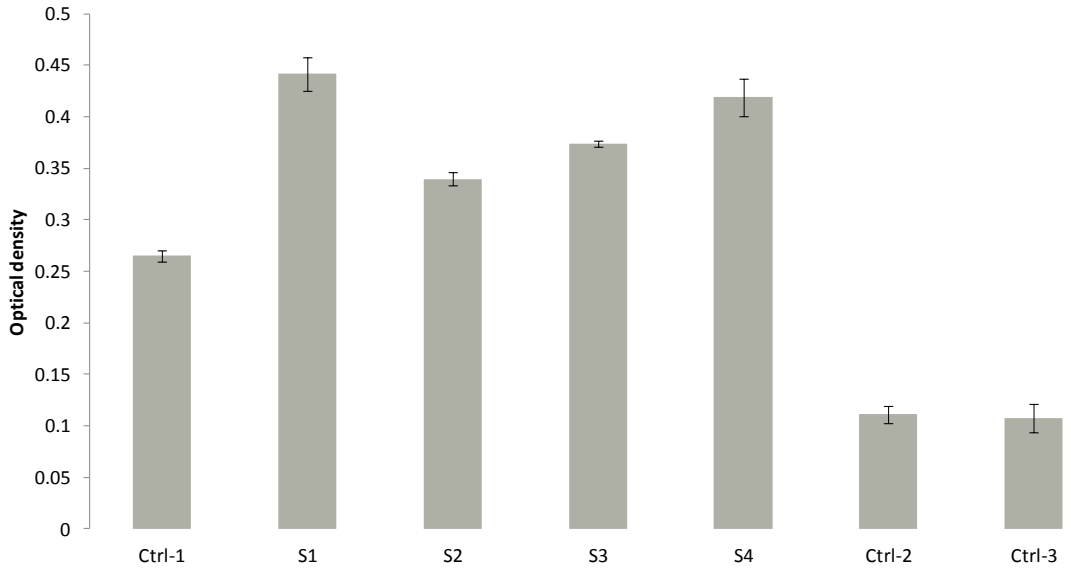
The figure shows depletion of quantity of fiber proteins after silencing with siRNA. Ctrl is control Mimivirus without siRNA; si means Mimivirus treated with siRNA targeting genes R135, L725, L829, and R856. Anti-fiber antibodies at dilution 1:5000 were used.

Figure 3. qPCR analysis of Sputnik virophage DNA co-infected with Mimivirus with and without silencing of the genes encoding proteins associated with Mimivirus fibers.



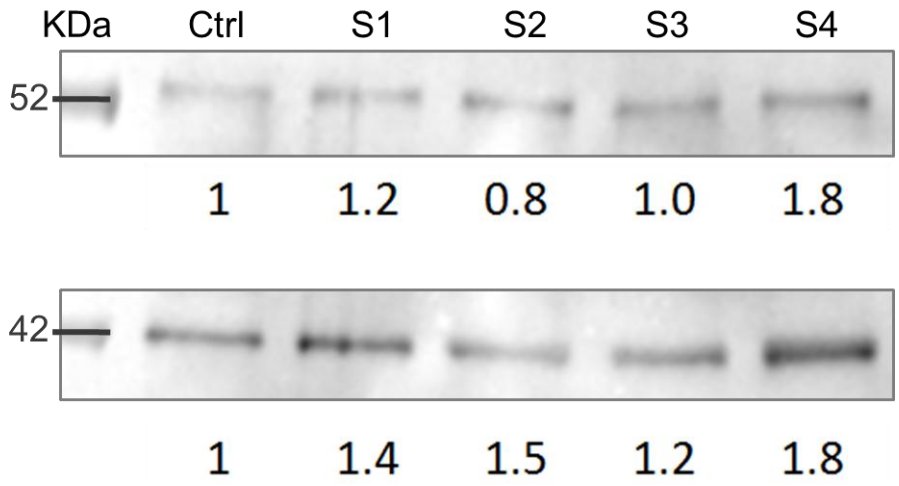
Ctrl is control Sputnik virophage in co-infection with control Mimivirus (in absence of gene silencing); S1: Sputnik co-infected with si-R135 (si means Mimivirus treated with siRNA targeting R135); S2, S3, and S4 are Sputnik co-infected with si-L725, si-L829, and si-R856 respectively.

Figure 4. ELISA analysis of Sputnik virophage in co-infection with Mimivirus with and without prior silencing of the Mimivirus genes encoding proteins associated with Mimivirus fibers.



Ctrl-1 is control Sputnik virophage in co-infection with control Mimivirus (in absence of gene silencing); S1: Sputnik co-infected with si-R135 (si means Mimivirus treated with siRNA targeting R135); S2, S3, and S4 are Sputnik co-infected with si-L725, si-L829, and si-R856 respectively. Ctrl-2 and Ctrl-3 are negative controls, which were manipulated with the same procedure except addition of virus sample or primary antibodies, respectively. Antibodies to Sputnik virophage were used at dilution 1:1000.

Figure 5. Western blot analyses with and without silencing of the genes encoding proteins associated with Mimivirus fibers.



Ctrl is control Sputnik in co-infection with control Mimivirus; S1: Sputnik co-infected with si-R135 (si means Mimivirus treated with siRNA targeting R135); S2, S2, and S3 are Sputnik co-infected with si-L725, si-L829, and si-R856 respectively. The number below each band represents the relative area (measured by ImageJ and confirmed by IS Lite) to the control virus. Antibodies to Sputnik virophage were used at dilution 1:5000; R21 hypothetical protein Sputnik3 (~52 kDa) and minor virion protein (~43 kDa).

Chapter 3

Mimivirus protein expression in *Escherichia coli*

Introduction

In chapter 2 of this study, we provided evidence using functional genomics coupled with proteomic analysis and immunassays that four Mimivirus proteins (R135, L725, L829, and R856) are component of the Mimivirus fibers or play a role in their biosynthesis.

In this part of the study, we aimed to determine the structure of the Mimivirus fiber proteins, particularly, proteins R135, L725, and L829. We expressed L725 and L829 genes in Escherichia coli, after optimizing expression by using Origami, Rosetta, C41, BL21(DE3)pLysS cells and three different temperatures (17, 25 and 37 °C) in two different media (Lysogeny broth (LB) and ZYP-5052 autoinduction media). The L725 gene was expressed, and then its product was purified on a Nickel column and through separation by size-exclusion chromatography (ÄKTA, GE Healthcare, USA). Nevertheless, the protein concentration was low and this protein could not be crystallized. Regarding the L829 gene, it could not be expressed in E. coli either, or encoded for a non soluble protein. Lastly, attempt to express the R135 gene in E. coli failed, which is likely related to the presence in the protein of a transmembrane domain.

In this chapter, we report that purified recombinant L725 self-assembled in vitro after its cleavage from fused proteins and formed a fiber-like structure. This result provides further evidence that the L725 protein is a component of the Mimivirus fibers.

Research article no. 3

Short title: Mimivirus L725 ORFan gene product

Full-length title: The Mimivirus L725 ORFan gene product can self-assemble *in vitro* and form a fiber

Author list: Haitham Sobhy[1], Guillaume Gotthard[1], Eric Chabrière[1], Didier Raoult[1,2] and Philippe Colson[1,2*]

1. Aix-Marseille University, URMITE UM63 CNRS 7278 IRD 198 INSERM U1095, Facultés de Médecine et de Pharmacie, Marseille, France

2. Fondation IHU Méditerranée Infection, Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Centre Hospitalo-Universitaire Timone, Assistance Publique-Hôpitaux de Marseille, Marseille, France

* To whom correspondence should be addressed.

Author contributions: HS, GG and EC conceived and designed the experiments; HS, GG, EC, DR and PC analyzed the data; HS performed the experiments; and HS and PC wrote the manuscript.

Abstract

Mimivirus was the first discovered giant virus. Its tremendous gene repertoire was revealed to encompass several genes unique amongst viruses, and a majority of ORFans or genes encoding hypothetical proteins. The Mimivirus virions are covered by a dense layer of about 130 nm-long fibers, the length and shape of which diverge from those of other viruses. The protein encoded by the Mimivirus L725 ORFan was previously identified by proteomics with purified viral fibers. Here, we report that recombinant L725 protein expressed in *E. coli* aggregated after cleavage of fusion proteins, and formed a fiber-like structure as observed by electron microscopy and tomography. These results provide further evidence that protein L725 might be a component of the Mimivirus fibers and is one of the few studies that expressed a Mimiviral gene and the only one, to our knowledge, which expressed a Mimiviral ORFan.

Introduction

Acanthamoeba polyphaga mimivirus was the first amoebal giant virus discovered, a decade ago, and founded the family *Mimiviridae* that now encompasses dozens of members (Raoult, Audic et al. 2004, Pagnier, Reteno et al. 2013). These giant viruses are characterized by their huge virion particle and genome sizes, which are ~500 nm in diameter and 1.2 megabase pairs, respectively. In addition, their genome harbor a considerable proportion of ORFans or genes encoding putative proteins of unknown function, which account for 54% of the Mimivirus gene repertoire (Raoult, Audic et al. 2004, Kuznetsov, Xiao et al. 2010).

One specific feature of Mimivirus as compared to other viruses is the presence of a dense layer of fibers, the morphology and size of which differs from those in other viruses. These fibers are ~125-140 nm in length and ~1.4 nm in diameter, and consist of a soft shaft and a head globular shape, with a basal disc shape that links a group of 3-4 fibers (Xiao, Kuznetsov et al. 2009, Klose, Kuznetsov et al. 2010, Kuznetsov, Xiao et al. 2010). They cover the entire capsid surface, except at one of its vertices, where is located a star shape structure identified as a portal for DNA release (Xiao, Kuznetsov et al. 2009, Klose, Kuznetsov et al. 2010, Kuznetsov, Xiao et al. 2010), and they were described as glycosylated and resisting protease and collagenase treatment (Xiao, Kuznetsov et al. 2009, Boyer, Azza et al. 2011). Culturing Mimivirus 150 times in germ-free amoeba led to the emergence of a viral strain devoid of fibers, so called M4 strain, concurrently with a 16% reduction of the viral genome (Boyer, Azza et al. 2011). This Mimivirus M4 strain exhibited reduced pathogenicity for its amoebal host, and resistance to infection by virophages, which suggested the role of fibers in both host and virophage infections (Boyer, Azza et al. 2011).

Several Mimivirus proteins have been involved in the formation of the fibers, though their precise structure and composition has not been fully resolved. The 224- amino acid-long product of ORFan gene L725 (YP_003987254) (supplementary Figure 1 and 2) was among the three Mimivirus proteins, with R135 (a putative GMC-type oxidoreductase) and L829 (an hypothetical protein) identified from purified fibers by 2D-gel electrophoresis coupled with by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (Boyer, Azza et al. 2011). Here, we report that recombinant purified L725 protein (pr-L725) can self-assemble *in vitro* to form fibers.

Materials and methods

Protein expression and purification

The Mimivirus L725 gene was cloned into a Gateway pETG20A expression vector (Invitrogen, USA) used for the transformation of various *Escherichia coli* strains. This vector included thioredoxin (Trx), an hexa histidine-tag (6xHis) and *Tobacco etch* virus (Tev) protease, connected upper-stream to L725 protein. Protein expression was tested in *E. coli* Origami, Rosetta, C41 and BL21(DE3)pLysS strains (cloned and prepared in-house), and the Rosetta strain was selected (Supplementary Figure 3). Then, *E. coli* Rosetta strain containing the expression vector were grown in Luria-Bertani (LB) medium at a 1:20 ratio at 37°C overnight. Afterwards, these cultures were moved to ZYP media (with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol) and kept at 37°C until an optical density corresponding to 600 nmol/1 was reached, and then at 17°C overnight. Subsequently, *E. coli* strains were harvested by centrifugation (3,000 g at 4°C for 10 min), re-suspended in lysis buffer (150 mM NaCl, 0.25 mg/ml lysozyme, 0.1 mM PMSF, 10 µg/ml DNAse I and 20 mM MgSO4) and stored at -80°C for at least 1 hour. Frozen cultures were then incubated at 37°C to lyse bacteria by heat shock, and bacterial debris were removed by centrifugation (12,000 g, 30 min, 4°C). Lastly, protein purification was performed on a Nickel-affinity column (HisTrap HP, GE Healthcare) and size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) (both on ÄKTA avant 25 chromatography system (GE Healthcare, USA)). Protein concentration was measured by absorption spectroscopy with a Nanodrop instrument (Thermo Scientific). The fused proteins were cleaved by *Tobacco etch* virus (Tev) protease at a 1:20(w:w) ratio (cloned and prepared in-house) (van den Berg, Lofdahl et al. 2006).

Proteomic analysis

A mix consisting of 20 µL of sample and 40 µL of loading buffer (Laemmli solution (Bio-Rad, USA) and dithiothreitol (DTT) 350 mM) was heated at 95°C during 5 min then was loaded on a 12% sodium dodecyl sulfate (SDS) polyacrylamide gel in protein sample buffer (10X Tris-Glycine-SDS (Euromedex, France)). Then, gels were stained by Coomassie blue. For western blot, proteins were transferred to nitrocellulose membranes and incubated with primary antibodies (anti-Mimivirus fiber or anti-Mimivirus L725 polyclonal antibodies) diluted in blocking buffer (PBS, 0.3% Tween 20 and 5% nonfat dry milk; 1:5000 dilution). The

membranes were probed with horseradish peroxidase-conjugated goat anti-mouse secondary antibodies (Amersham, UK).

MALDI-TOF MS peptide mass fingerprinting

The MALDI-TOF MS analysis was previously described (Boyer, Azza et al. 2011). Briefly, particular bands were cut and conserved in ethanol 20%, washed by H_2O for 5 min and dissolved in acetonitrile. Then, samples were incubated with 10 mM DTT for 45 min at 56 °C and replaced by 55 mM iodoacetamide and incubated for 45 min into dark. The samples were washed by H_2O and acetonitrile, and centrifuged for 30 min to discard excess H_2O. The proteins were digested by trypsin (12 ng/pl; Promega, USA) and incubated for 45 min at 4 °C, then for 2-3 hours at 37 °C. Peptides were re-suspended in α-cyano-4-hydroxycinnamic acid (ACCA) matrix solution (prepared by diluting 0.3 mg of ACCA in 1 ml of 33% acetone and 67% ethanol, and 0.3% trifluoroacetic acid) and spotted on the MALDI-TOF target. Analyses were performed on an Ultraflex spectrometer (Bruker Daltonique, Wissembourg, France), and proteins were identified using the Mascot software (http://www.matrixscience.com).

Antibodies preparation

The mice were boosted by purified fibers or purified L725 protein (fused with thioredoxin) to obtain anti-fiber polyclonal antibodies and anti-L725 antibodies as previously described (Boyer, Azza et al. 2011). Briefly, three six week-old female BALB/c mice were inoculated three times intraperitoneally at 14-days interval with 2 µg proteins (fiber or L725) mixed with 400 µg aluminium hydroxide and 10 µg CpG.

Circular dichroism spectroscopy and thermal denaturation

Circular dichroism (CD) spectra of L725 (in Tris 50mM NaCl 300mM) in the far UV region (190-260 nm) were obtained with a J-815 spectropolarimeter (Jasco, Japan) using a 0.1 cm path length cuvette. Spectra were obtained using an average of 6 scans with a data pitch of 0.2 nm, a scan speed of 200 nm/min, and a bandwidth of 1 nm. Thermal denaturation was monitored from 17 to 70°C, one spectrum being obtained every 5°C in the far UV region (190-260 nm).

Electron microscopy

For transmission electron microscopy observations, samples were fixed by glutaraldehyde and cacodylate buffer and deposited on 400-mesh nickel grids coated with formvar/carbon (Electron Microscopy Sciences, Hatfield, PA, US), then viewed with the FEI Tecnai F20 microscope operating at 200 keV. For tomography, samples were labeled with nanogold particles then investigated using the FEI Tecnai F20 microscope, with dual axis tilt series. The tilt range was between -60° and +50° with an incrementation of 2°, and the defocus ranged from 1.5 to 3 mm. Images were recorded on a 4k CCD camera (Gatan) at a 50,000 X magnification, providing a pixel size of 2.17 Å, and a linear acquisition scheme. Tomograms were reconstructed from the tilt series with the IMOD software (Kremer, Mastronarde et al. 1996) using a 10-nm fiducial marker for alignment.

Results

Mimivirus L725 gene expression and protein purification

The expression of Mimivirus L725 protein, the estimated molecular weight (MW) of which is 27 kilodaltons (kDa), was achieved in *Escherichia coli* Rosetta strain, using ZYP media at 17°C and was confirmed by MALDI-TOF MS (Figure 1A and Supplementary Figure 1-3). The L725 gene product fused with Trx and 6XHis (L725-Trx) was found to be at a concentration of 0.5 mg/mL. In addition, circular dichroism analyses showed that L725-Trx exhibited thermal denaturation at 53°C. The results show that the protein was pure, yielding only one band at MW 52 kDa before treatment with Tev protease (Fig 1A, lane 1) and two bands at MW 34 kDa and 17 kDa after treatment with Tev protease (Fig 1A, lane 2). Unfortunately, concentration obtained was not sufficient for crystal growth.

L725 protein is a component of Mimivirus fibers

After incubation of purified recombinant-L725 (pr-L725) with mouse anti-Mimivirus fiber polyclonal antibodies, bands were observed at the expected MW of L725-Trx (45 kDa; Fig 1B, lane 1) and L725 (27 kDa; Fig 1B, lane 2). In addition, anti-L725 antibodies recognized a

protein of same MW (Figure 1C, D). Bands at MW approximately two-fold greater, and at a lower MW compared to that of the L725 protein were observed (Fig 1B, lanes 1 and 2, respectively). These results may be explained by protein dimerization and degradation, respectively.

L725 precipitate and forms fibers.

Adding Tev protease to the purified-L725 protein solution (final pH, 8.5) led to the formation of smoky fogs. Electron microscopy observations showed a fiber-like structure (Figure 2,3). This filament was degraded after exposure to electron beam for 10 min, indicating it had an organic nature and was not artefactual (Figure 3). The electron microscopy and 3D structure tomogram revealed a fiberdiameter of ~50 nm and an appearance of stacking with little regularity (Supplementary Figure S4, movie S1). Taken together, these data suggest that the L725 protein may self-aggregate and auto-assemble *in vitro*, and form a fiber-like structure.

Discussion and conclusion

We provide here evidence that the L725 protein encoded by a Mimivirus ORFan is a component of the Mimivirus fiber. This protein was expressed in *E. coli* and recognized by both anti-L725 and anti-Mimivirus fiber antibodies. In addition, we observed that this protein can polymerize *in vitro* to form a fiber-like structure, which is compatible with a fiber architecture. These results are congruent with previous reports indicating that the Mimivirus L725 gene product, together with at least two other proteins (R135, L829), are components of the Mimivirus fibers (Boyer, Azza et al. 2011), and the present study is one of the few studies that expressed a Mimiviral gene (Abergel, Chenivesse et al. 2005, Abergel, Rudinger-Thirion et al. 2007) and the only one, to our knowledge, that expressed a Mimiviral ORFan.

The diameter of the L725 self-assembled fiber-like structure was ≈35 times larger than that of fibers of Mimivirus virions. This could be explained by differences in *in vitro* physico-chemical and biological conditions, including pH, temperature, protein concentration, putative interactions with protein partners and post-translational modifications. Another issue is that the genome of the bald Mimivirus M4 strain harbors the L725 gene, but, notwithstanding, this

Mimiviral strain does not possess fibers (Boyer, Azza et al. 2011). This suggests that the Mimivirus fiber might be a polymer of several proteins. The oligomerization of viral proteins has been previously largely described, including for instance for the capsid protein of picornaviruses (Li, Wang et al. 2012) or adeno-associated viruses (Steinbach, Wistuba et al. 1997), the adenovirus fiber knob (Schulz, Zhang et al. 2007) or the øX174 bacteriophage scaffolding protein (Cherwa, Organtini et al. 2011).

To date, *Acanthamoeba* spp. are the only known hosts for Mimivirus, but macrophages and peripheral blood mononuclear cells were shown to allow a replicative cycle of Mimivirus (Raoult, La Scola et al. 2007, Ghigo, Kartenbeck et al. 2008, Pagnier, Reteno et al. 2013, Silva, Almeida et al. 2013). The determinants of the Mimiviral tropism remain unknown, but the role of the fibers in host infection and virion stability was suspected (Boyer, Azza et al. 2011, Boratto, Dornas et al. 2013, Dornas, Silva et al. 2014). The recent isolation of mimiviruses from humans presenting pneumonia (Saadi, Pagnier et al. 2013, Saadi, Reteno et al. 2013) is incentive to decipher the mechanisms of entry of these giant viruses into eukaryotic cells. In this view, further analyses should precise the structure of the Mimivirus fiber, and proteins interactions that allow fiber biosynthesis in mimiviruses.

Funding

Haitham Sobhy received a grant from the Infectiopole Sud foundation.

Acknowledgements

We are thankful to Said Azza, Christophe Flaudrops, Claude Nappez and Jean-Pierre Baudoin for technical help.

References

Abergel, C., S. Chenivesse, D. Byrne, K. Suhre, V. Arondel and J. M. Claverie (2005). "Mimivirus TyrRS: preliminary structural and functional characterization of the first amino-acyl tRNA synthetase found in a virus." Acta Crystallogr Sect F Struct Biol Cryst Commun **61**(Pt 2): 212-215.

Abergel, C., J. Rudinger-Thirion, R. Giege and J. M. Claverie (2007). "Virus-encoded aminoacyl-tRNA synthetases: structural and functional characterization of mimivirus TyrRS and MetRS." J Virol **81**(22): 12406-12417.

Boratto, P. V., F. P. Dornas, K. R. Andrade, R. Rodrigues, F. Peixoto, L. C. Silva, B. La Scola, A. O. Costa, G. M. de Almeida, E. G. Kroon and J. S. Abrahao (2013). "Amoebas as mimivirus bunkers: increased resistance to UV light, heat and chemical biocides when viruses are carried by amoeba hosts." Arch Virol **2013**: 26.

Boyer, M., S. Azza, L. Barrassi, T. Klose, A. Campocasso, I. Pagnier, G. Fournous, A. Borg, C. Robert, X. Zhang, C. Desnues, B. Henrissat, M. G. Rossmann, B. La Scola and D. Raoult (2011). "Mimivirus shows dramatic genome reduction after intraamoebal culture." Proc Natl Acad Sci U S A **108**(25): 10296-10301.

Cherwa, J. E., Jr., L. J. Organtini, R. E. Ashley, S. L. Hafenstein and B. A. Fane (2011). "In VITRO ASSEMBLY of the oX174 procapsid from external scaffolding protein oligomers and early pentameric assembly intermediates." J Mol Biol **412**(3): 387-396.

Dornas, F. P., L. C. Silva, G. M. de Almeida, R. K. Campos, P. V. Boratto, A. P. Franco-Luiz, B. La Scola, P. C. Ferreira, E. G. Kroon and J. S. Abrahao (2014). "Acanthamoeba polyphaga mimivirus Stability in Environmental and Clinical Substrates: Implications for Virus Detection and Isolation." PLoS One **9**(2): e87811.

Ghigo, E., J. Kartenbeck, P. Lien, L. Pelkmans, C. Capo, J. L. Mege and D. Raoult (2008). "Ameobal pathogen mimivirus infects macrophages through phagocytosis." PLoS Pathog **4**(6): e1000087.

Klose, T., Y. G. Kuznetsov, C. Xiao, S. Sun, A. McPherson and M. G. Rossmann (2010). "The three-dimensional structure of Mimivirus." Intervirology **53**(5): 268-273.

Kremer, J. R., D. N. Mastronarde and J. R. McIntosh (1996). "Computer visualization of three-dimensional image data using IMOD." J Struct Biol **116**(1): 71-76.

Kuznetsov, Y. G., C. Xiao, S. Sun, D. Raoult, M. Rossmann and A. McPherson (2010). "Atomic force microscopy investigation of the giant mimivirus." Virology **404**(1): 127-137.

Li, C., J. C. Wang, M. W. Taylor and A. Zlotnick (2012). "In vitro assembly of an empty picornavirus capsid follows a dodecahedral path." J Virol **86**(23): 13062-13069.

Pagnier, I., D. G. Reteno, H. Saadi, M. Boughalmi, M. Gaia, M. Slimani, T. Ngounga, M. Bekliz, P. Colson, D. Raoult and B. La Scola (2013). "A decade of improvements in Mimiviridae and Marseilleviridae isolation from amoeba." Intervirology **56**(6): 354-363.

Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan and J. M. Claverie (2004). "The 1.2-megabase genome sequence of Mimivirus." Science **306**(5700): 1344-1350.

Raoult, D., B. La Scola and R. Birtles (2007). "The discovery and characterization of Mimivirus, the largest known virus and putative pneumonia agent." Clin Infect Dis **45**(1): 95-102.

Saadi, H., I. Pagnier, P. Colson, J. K. Cherif, M. Beji, M. Boughalmi, S. Azza, N. Armstrong, C. Robert, G. Fournous, B. La Scola and D. Raoult (2013). "First isolation of mimivirus in a patient with pneumonia." Clin Infect Dis **57**(4): e127-134.

Saadi, H., D. G. Reteno, P. Colson, S. Aherfi, P. Minodier, I. Pagnier, D. Raoult and B. La Scola (2013). "Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia." Intervirology **56**(6): 424-429.

Schulz, R., Y. B. Zhang, C. J. Liu and P. Freimuth (2007). "Thiamine diphosphate binds to intermediates in the assembly of adenovirus fiber knob trimers in Escherichia coli." Protein Sci **16**(12): 2684-2693.

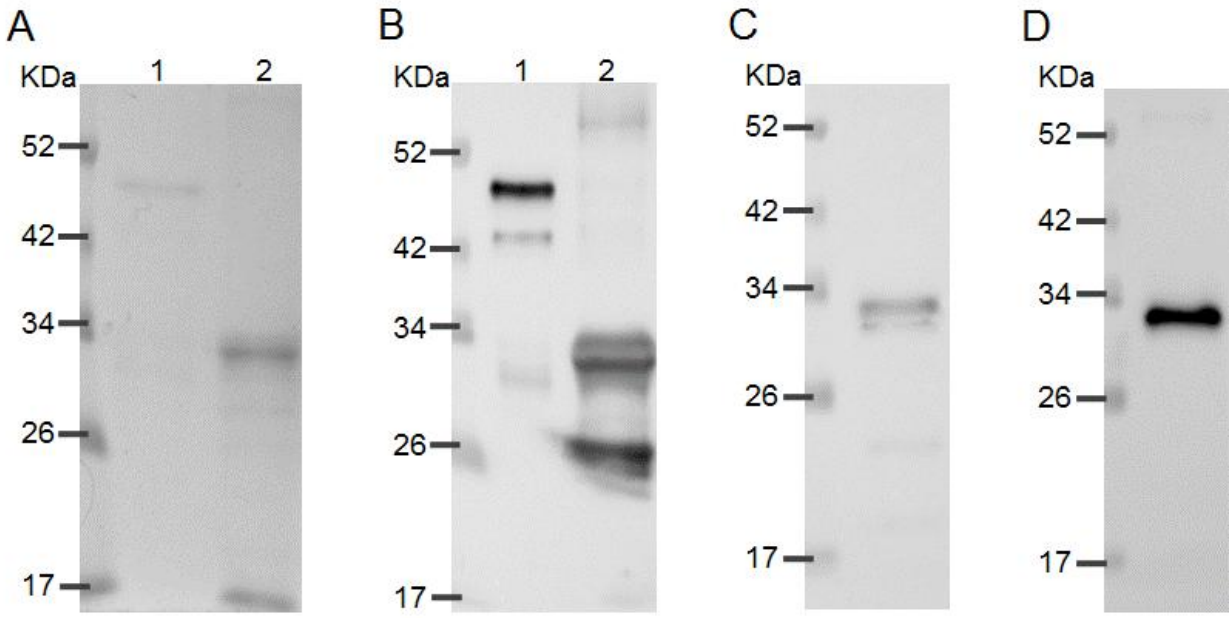
Silva, L. C., G. M. Almeida, D. B. Oliveira, F. P. Dornas, R. K. Campos, B. La Scola, P. C. Ferreira, E. G. Kroon and J. S. Abrahao (2013). "A resourceful giant: APMV is able to interfere with the human type I interferon system." Microbes Infect **4579**(13): 00253-00250.

Steinbach, S., A. Wistuba, T. Bock and J. A. Kleinschmidt (1997). "Assembly of adeno-associated virus type 2 capsids in vitro." J Gen Virol **78**(Pt 6): 1453-1462.

van den Berg, S., P. A. Lofdahl, T. Hard and H. Berglund (2006). "Improved solubility of TEV protease by directed evolution." J Biotechnol **121**(3): 291-298.

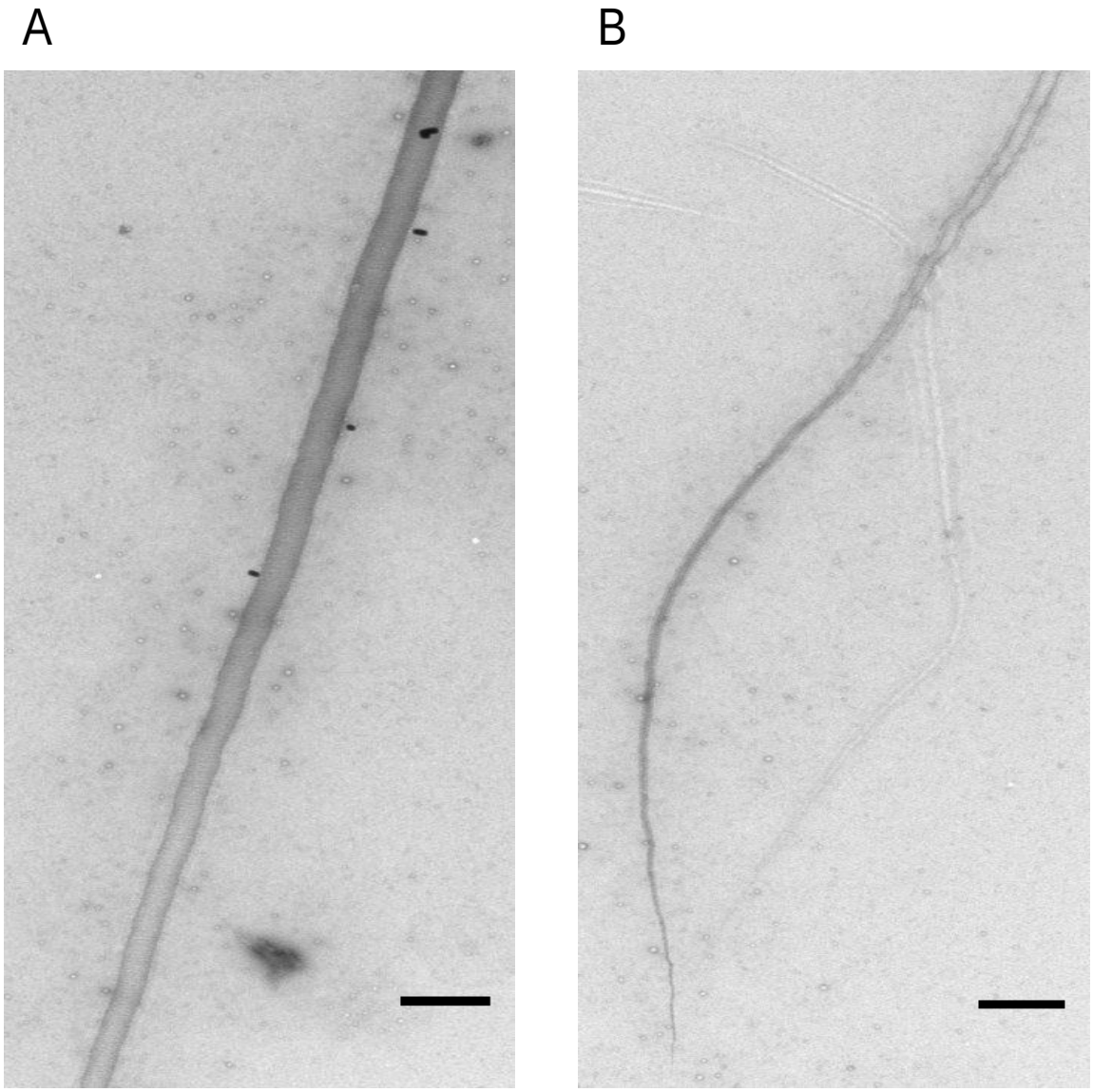
Xiao, C., Y. G. Kuznetsov, S. Sun, S. L. Hafenstein, V. A. Kostyuchenko, P. R. Chipman, M. Suzan-Monti, D. Raoult, A. McPherson and M. G. Rossmann (2009). "Structural studies of the giant mimivirus." PLoS Biol **7**(4): e92.

Figure 1. Gel analysis of purified L725 protein.



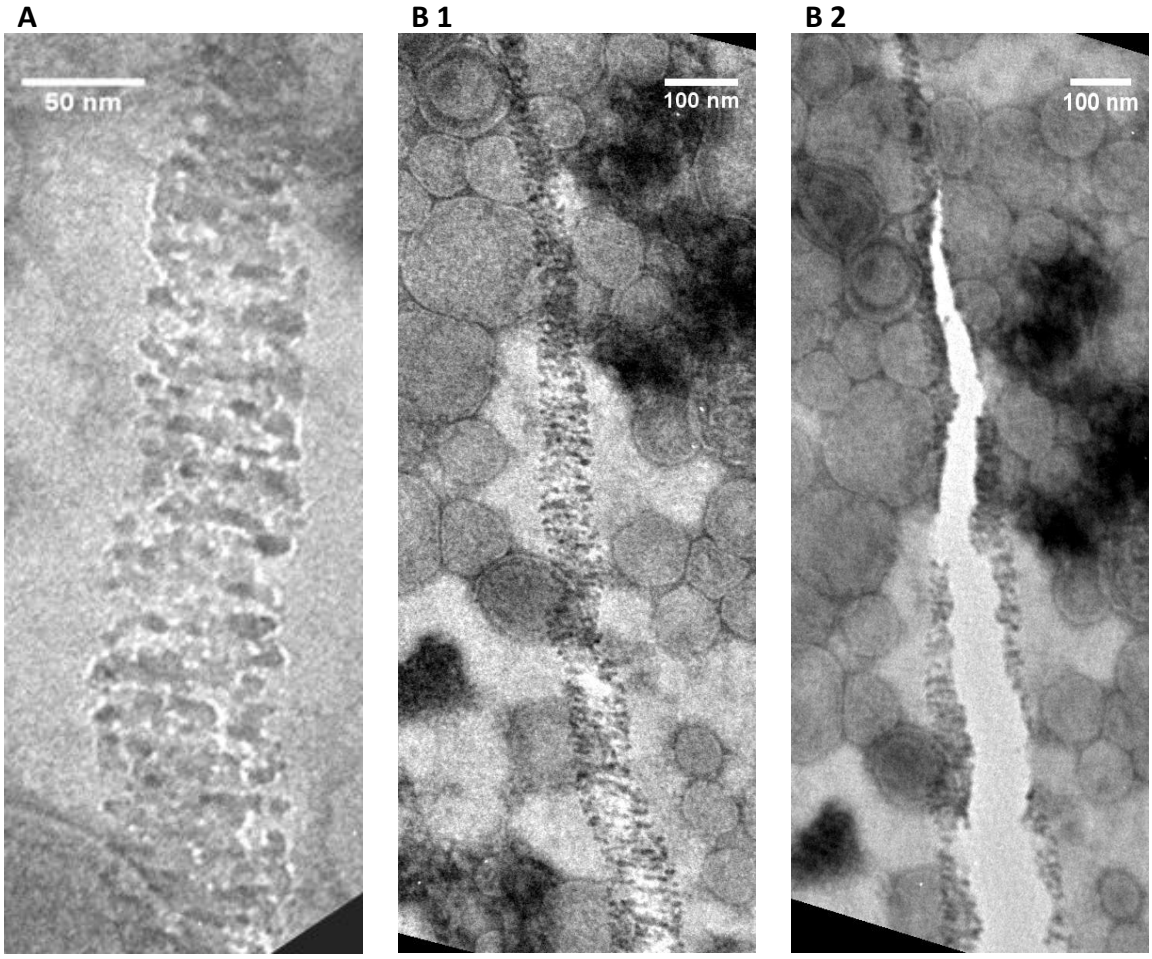
(A) SDS of purified pr-L725 protein; the bands on lane 1 and 2 were analyzed by MALDI-TOF MS. (B) Western blot of pr-L725 protein incubated with anti-Mimivirus fiber antibodies; (C) Western blot of Mimivirus proteins incubated with anti-fiber antibodies. (D) Western blot of Mimivirus proteins incubated with anti-L725 antibodies. (C) is a positive control of (D).
(Lanes 1 from Figure 1A-B) L725 fused with Trx (42 kDa); (Lanes 2 from Figure 2A-B) L725 cleaved by Tev protease; L725 (27 kDa) and Trx (~15 kDa).
Mouse polyclonal antibodies (1/5,000) were prepared against Mimivirus purified fibers, whereas anti-L725 antibodies (1/10,000) were prepared against L725 fused with Trx.

Figure 2. Electron microscopy image of the hair/fiber-like structure formed from pr-L725 proteins.



The fiber has cylindrical architectures, with two ends, (A) is the thick end and (B) is thinner; the scale bar is approximated as 200 nm.

Figure 3. Hair/fiber-like structure before and after exposure to electron beam.

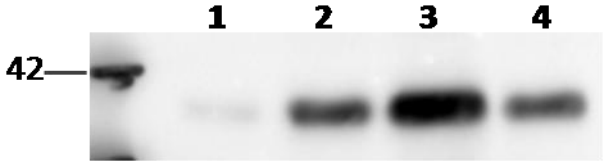


(A) The fiber appears as stacking with little regularity. (B 1 and B 2) The fiber broke down under extensive electron beam for about 10 min. Fig 3 is obtained from experiment different than Figure 2.

Supplementary Information SI-1

Supplemental results

Figure S1. Optimization of protein expression western blot using anti-Trx antibodies.



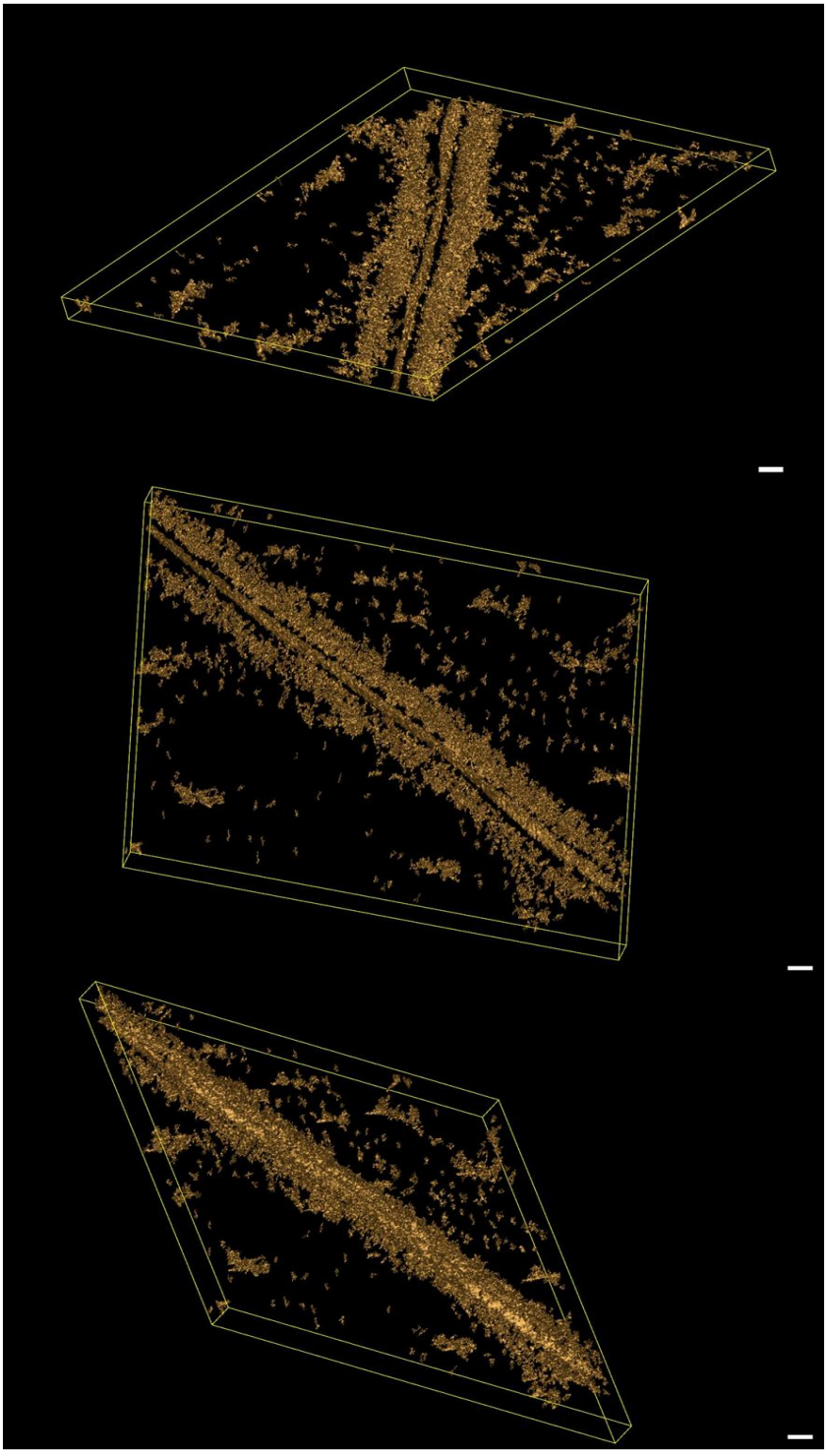
L725 protein only is expressed in E. coli, at 17°C. (1) BL21 induced by arabinose; (2) BL21; (3) Rosetta; (4) C41 strains. On the other hand, the L829 and R135 proteins did not expressed or not soluble proteins.

Figure S2. L725 protein sequence, 224 residues with 18 cysteine residues and the 154-CLPSC-158 motif conforms to the CxxxC consensus motif.

```
>lcl|NC_014649.1_cdsid_YP_003987254.1 [gene=L725] [protein=hypothetical protein]
  1 MANNLVQLIFDQFIEILEDLAAKDEWCFDFNKCDFDFRVRELVNHRLLDVKYTIKDECGRPRDVIQEIDI 70
 71 TGICYEDLTTCKWVDYLTKLAVEYINNICPPRYIIIKEEPKKCRPQLPEWNPFPCKRTTTIYRRQKPVEK 140
141 KPECEVIFEKGCECLPSCEREVPVPKEQIFIKYEPVPAKCCERTVLVRSPEQNRHSFGVHKGNIDYNNHV 210
211 WPKCCQSKKCNCAH 224
```

Figure S3. The predicted secondary structure of L725 protein using PSIPred (http://bioinf.cs.ucl.ac.uk/psipred/)

Figure S4. The 3D structure of the fiber constructed from tomogram (scale bar 50 nm).



Chapter 4

PCR primers and probes designing tool from degenerate and multiple sequences

Introduction

For most of the biologists with limited skills in bioinformatics, designing primers and probes for polymerase chain reaction (PCR) assays can be a challenging process due to the difficulty of using the design software. In addition, designing universal primers that can amplify sequences from microorganisms of different genotypes, subtypes, species or genus, or orthologous genes, is a major concern in microbiology. These challenges arise from the difficulty of finding a conserved region in aligned sequences for primers to hybridize. The genes that encode Mimivirus proteins involved in the formation of its fibers and their orthologs in other mimiviral genomes show significant variation among their sequences. This leads to difficulty in finding conserved regions for primers and probes hybridization (Figure 1).

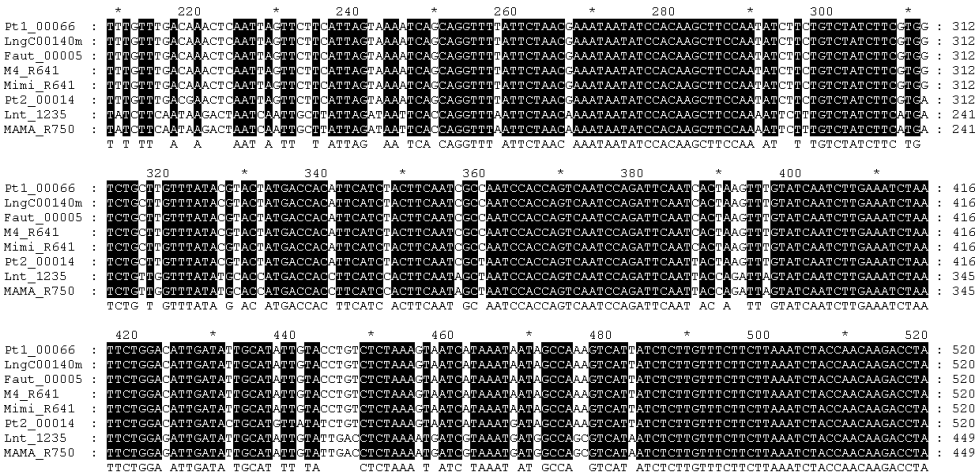


Figure 1. Representation of the sequence variation of R641 gene among mimiviruses.

In this part of the study, we developed a versatile bioinformatic tool to design primers and probes. This tool is easy to use and fast, and it is freely available online. Primers and probes designed to amplify fiber-associated genes from various mimiviruses have been tested in PCR assays and proved efficient.

Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2012, Article ID 783138, 5 pages
doi:10.1155/2012/783138

Research Article

Gemi: PCR Primers Prediction from Multiple Alignments

Haitham Sobhy[1] and Philippe Colson[1,2]

[1] *Facultés de Médecine et de Pharmacie, Aix Marseille Université, URMITE, UM 63, CNRS 7278, INSERM, U1095, 13385 Marseille Cedex 05, France*
[2] *Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, IHU Méditerranée Infection, Centre Hospitalier-Universitaire Timone, Assistance Publique-Hôpitaux de Marseille, 13385 Marseille Cedex 05, France*

Correspondence should be addressed to Haitham Sobhy, haithamsobhy@gmail.com and Philippe Colson, philippe.colson@univ-amu.fr

Received 3 August 2012; Accepted 19 October 2012

Academic Editor: Brian Wigdahl

Designing primers and probes for polymerase chain reaction (PCR) is a preliminary and critical step that requires the identification of highly conserved regions in a given set of sequences. This task can be challenging if the targeted sequences display a high level of diversity, as frequently encountered in microbiologic studies. We developed Gemi, an automated, fast, and easy-to-use bioinformatics tool with a user-friendly interface to design primers and probes based on multiple aligned sequences. This tool can be used for the purpose of real-time and conventional PCR and can deal efficiently with large sets of sequences of a large size.

1. Introduction

Polymerase chain reaction (PCR) has been increasingly used over the last two decades to detect, quantify, and/or sequence nucleic acids from various sources [1, 2]. The number of publications in PubMed referencing "PCR" has increased from 2846 in 1990 to 20,426 in 2000 and 44,231 in 2010. This method has a wide range of applications, particularly in the field of microbiology [3–5] where primers and probes are often designed with the aim to hybridize to the greatest number of genome sequences for given groups of viruses, bacteria, or parasites [5–7]. However, designs can be challenging if primers and probes are meant to hybridize to sequences with considerable nucleotide diversity; the task becomes more complicated as the nucleotide diversity increases. Thus, identifying conserved regions in the targeted nucleotide sequences is a critical step in PCR primer design [5–9].

Several tools are available to design primers [6–14]. However, these tools often present limitations in their capabilities to parse numerous and/or large sequences, which are frequently encountered situations, or to deal with degenerate positions, and some of them are not easily usable without skill in bioinformatics.

Here, we present Gemi, which means "to find" in ancient Egyptian, a simple, automated, fast, and versatile tool to find universal primers and probes within a set of multiple, variable, and long sequences. The main criterion used to identify primers and probes in Gemi is nucleotide conservation, but our tool provides the dissociation temperature (T_d), length, and GC percentage in the final output file for each of the chosen primers or probes. The application executes directly on a PC computer and provides a simple and user-friendly interface that allows designing primers easily and quickly. In addition, Gemi can parse several hundred long (>1 kilobase) sequences within seconds. We believe that our tool can be particularly useful in the field of microbiology.

2. Algorithm and Method

The input file for Gemi is a multiple aligned FASTA file. Once it is uploaded to the program, a consensus sequence will be constructed. Gemi also accepts a single sequence (that can be manually curated) and uses it as a consensus. Unlike other programs, degenerate nucleotides are included in the consensus sequence and follow the IUPAC-IUB nomenclature system [15] (see Section 1 in the supplementary

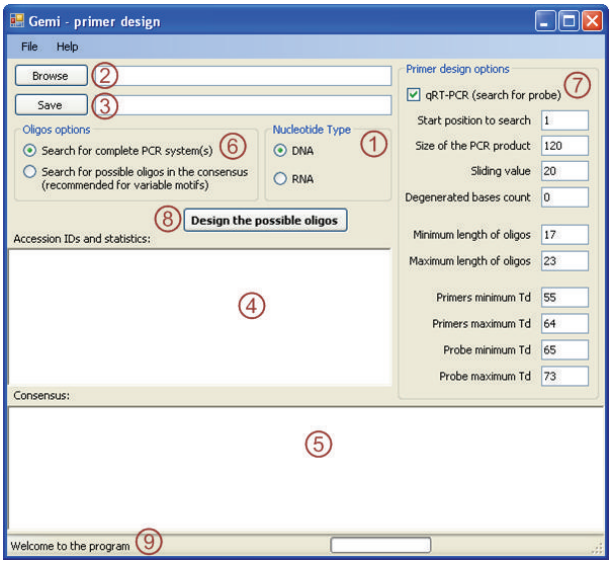


Figure 1: Screenshot of the main window of Gemi. (1) Switch from DNA to RNA, (2) browse for the input FASTA file, (3) save the output tabulated text file, (4) the accession numbers and the percentage of the conservation in the consensus will appear in this area, (5) the consensus sequence is written in this area, (6) switch between the options, (7) choose to design probes by ticking real-time PCR, the parameters can be edited in this menu, (8) finally, click to design the primers, and (9) the program progress will be seen in the status bar.

file, SI-1 of the Supplementary Material available online at doi:10.1155/2012/783138). Gemi then searches for primers and probes by sliding a window of a chosen size, which corresponds to the required size of the PCR product, along the full-length consensus sequence. The step value by which the window slides is the sliding value (Sections 2 and 3 in the supplementary file, SI-1).

Several parameters can be modified by the user from the main window, although default values are provided for each parameter (Figure 1). These parameters include the size of the sliding window, the sliding value, the number of degenerate positions, and the size and T_d of the oligos (Figure 2). The default sliding value is 20. Another default value proposed for the identification of appropriate oligos is that the number of variable sites is zero at the 3′ end positions of the primers and probes. Another criterion is that the appropriate oligos must not contain more than three variable/degenerate nucleotide positions. However, more relaxed parameters can be chosen.

Two options have been implemented in Gemi to identify potential oligos, which can be chosen by ticking boxes (Figure 1). The first option consists of delineating a size for the nucleotide fragment that will contain hybridization sites for the primers and probes. One possibility is to choose a short window size to design real-time PCR primer/probe sets, classically, <150 nucleotides. This possibility requires ticking the "search for probe" box. Another possibility is to choose a larger window size to design primers for Sanger sequencing, classically, >200 nucleotides. A second option consists of using Gemi without delineating a size for the nucleotide fragment that will contain hybridization sites for the primers and probes. This option will result in the generation of a list of all possible oligos along the consensus sequence, whatever their respective location, with the start and end positions of the oligos reported. This latter option is particularly convenient for identifying primers in highly variable sequences, when first options failed, and it allows the user to manually select the best combination of proposed oligos (Figure 2 and Section 3 in the supplementary information, SI-1).

Regardless of the chosen option, the final report presents the sequence, length, GC content, T_d, and position on the consensus sequence for each oligo. The T_d of small oligos is estimated using the Wallace rule for the dissociation temperature [16, 17]. For longer oligos, the nearest-neighbor method is used to calculate the melting temperature (T_m) [18]. Here, the T_d is calculated using the equation in [17]: $T_d = 2°\text{C} * (\#\text{A} + \#\text{T}) + 4°\text{C} * (\#\text{C} + \#\text{G})$, where "#" refers to the number of As, Cs, Gs, or Ts in the oligo.

3. Results and Discussion

We developed Gemi to supply the critical needs for the design of PCR primers and probes with an easy-to-use, fast and efficient. Several other tools for the design of PCR systems have been previously described [10]. Nonetheless, some limitations can be pointed out for these tools [19–33]. The first limitation is that some tools, such as Primer3 that is an online and powerful tool to design primers based on a single, short and conserved sequence, cannot parse sequences with degenerate bases, what can be accomplished by Gemi [19]. Other software as BatchPrimer3 or Primaclade accepts only one sequence [21, 22]. Other tools exist that can overcome this shortcoming [29–33], such as PrimerIdent, which accepts only eight sequences, one of them being used as template [29]. GeneFisher can parse multiple sequences but fails to deal with sequences with degenerate bases [30].

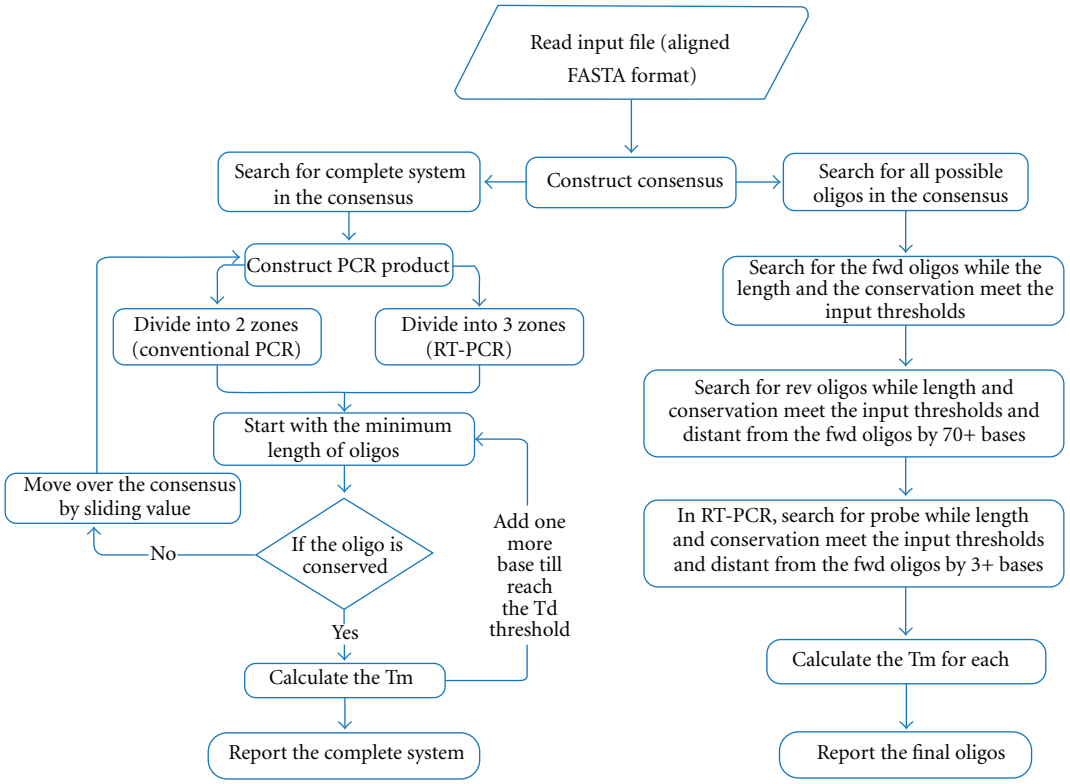


Figure 2: Flow chart explaining the procedure used by Gemi to find oligos. The first step is to load the sequences to Gemi. Then, the tool parses them and searches for the conserved regions and reports them in the final file. fwd, forward; rev, reverse; and RT-PCR, real-time PCR.

The web-based tool Greene SCPrimer designs degenerate primers from multiple sequence alignments by constructing phylogenetic tree, which is a slow process [32]. The easyPAC tool can design degenerate primers and also performs mapping to reference files for real-time PCR, but it performs slower than Gemi for the primer and probe design [33]. We previously described SVARAP for the analysis of sequence variability and primer design [6], which can analyze a maximum of 100 sequences with a maximal length of 4,000 nucleotides. Other tools have specific applications, such as PhiSiGns that identifies gene signatures in phage genomes [34]. Besides, some tools for the design of PCR systems require bioinformatics skills, such as "Prosig" [35] or the PriMux package that is based on python scripts to search for primers and probes on nonaligned multiple sequences [7].

Existing tools often search for oligos by taking into account parameters such as the GC-content, the T_m, or the formation of secondary structures. However, the most critical issue for several PCR-based assays is the identification of conserved regions where primers and probes can hybridize, in addition to the length and number of query sequences. These issues are particularly important in the field of microbiology. Moreover, the lack of user-friendly interface and cross-platform tool are challenging issues for biologist without prior knowledge of the programming tools.

Gemi has several advantages compared to other tools (Table 1). It is able to automatically and rapidly predict PCR primers for numerous long and variable sequences. Additionally, Gemi can be used to design PCR systems for both real-time PCR and sequencing. Moreover, no training in bioinformatics is required to use Gemi, which has a user-friendly interface.

Using Gemi on a PC with 512 MB RAM, it succeeded to construct consensus and to identify primers and probes among 61 aligned full-length hepatitis C virus genomes with a length of about 10000 nucleotides within few seconds (Table 1 and Section 4 in supplementary file, SI-1), while easyPAC failed to identify any primer and Greene SCPrimer hardly runs to design primers even with shorter sequences.

Some advanced options such as identification of secondary structures, and prediction of hairpins and primer-dimer formation are not presented in this version; these issues may be addressed in future versions of Gemi. Alternatively, prediction of the hairpin or dimer formation could be determined by other tools as OligoCalc tool [36].

4. Availability and Implementation

Gemi is a cross-platform application which is distributed under GNU-GPL license and is free to use for academic and research purposes. The portable desktop version of this tool facilitates its free distribution and usage. The software and documentation are freely available for research use at https://sourceforge.net/projects/gemi/. A script version of Gemi will be available upon request.

Table 1: Comparison between Gemi and other existing public tools.

Criteria	Gemi	Primer3	easyPAC
Simplicity	Yes	Yes	Yes
Fast[1]	Yes	Yes	∗
User friendly	Yes	Yes	Yes
Multiple and divergent sequences[2]	Yes	NA	#
Long sequences[3]	Yes	NA	∗
Cross-platform	Yes	Online	Yes
Probes' design	Yes	Yes	NA
GC content	Yes	Yes	Yes
Temperature	Td	Tm	Tm
Temperature range[4]	Yes	NA	#
Hairpin structure	NA	Yes	#
Parameters[5]	Basic	Advanced	Advanced

The table represents a comparison between Gemi, Primer3, and easyPAC tools. (Yes) denotes it is covered by the tool, (NA) means not offered by the tool, (∗) means offered but Gemi performs better in this function, while (#) means this option is offered referring to the paper.
[1]Gemi can retrieve primers and probes within seconds (Section 4 in supplementary document, SI-1). Primer3 searches for primers within short fragment of the sequence; its performance is relatively fast. EasyPAC performed slower than Gemi.
[2]Gemi succeeded to find primers and probes for multiple and divergent (aligned) sequences with about 30% identity, while Primer3 could not parse divergent sequences and easyPAC failed to retrieve any primer.
[3]Gemi successfully presented primers and probes for input sequences of about 10 kbp; the same process cannot be accomplished by Primer3 and easyPAC.
[4]In case of degenerate nucleotides in a position within primers, Gemi reports the temperature range of these nucleotides. Based on the paper, easyPAC reports it as well.
[5]Although Primer3 and easyPAC tools offer advanced functions, Gemi is designed to cover the basic needs of biomedical field to find reliable primers within minutes with user-friendly interface.

The tool runs on Windows 7 without any preliminary installations. For older versions, the software requires Microsoft.NET (Dot Net) Framework version 2.0, which is freely available from Microsoft website. For the Linux, Ubuntu, and Mac OS X users, please download Mono tool to run the software from http://www.mono-project.com/ or http://monodevelop.com/ (see the program's user guide).

The input file is a standard FASTA format file that contains a single sequence or a multiple sequence alignment, which can be created using any available alignment tool. The output file is generated as a tabulated text file that is easily read using any text processing program and contains the sequence of PCR product (if chosen), sequence of primers and probes (if chosen), positions of the oligos on the consensus, T_d, number of degenerate nucleotide, and GC content (see supplementary file and program's user guide).

5. Conclusions

We presented a simple, robust and fast tool, GEMI, which fulfils the regular requirements for biologists to design primers and probes. We believe that this tool can be helpful for research or diagnosis for a wide range of applications that includes detection, quantification, and genotyping in microbiology.

Acknowledgments

The authors are grateful to Prof. Didier Raoult for his advice and to Prof. Heebal Kim, Dr. Fadi Bittar, and Audrey Ferretti for their helpful comments. H. Sobhy is receiving grant from Infectiopole Sud foundation (2011–2014).

References

[1] I. M. Mackay, K. E. Arden, and A. Nitsche, "Real-time PCR in virology," *Nucleic Acids Research*, vol. 30, no. 6, pp. 1292–1305, 2002.
[2] J. M. Robertson and J. Walsh-Weller, "An introduction to PCR primer design and optimization of amplification reactions," *Methods in Molecular Biology*, vol. 98, pp. 121–154, 1998.
[3] P. E. Fournier and D. Raoult, "Prospects for the future using genomics and proteomics in clinical microbiology," *Annual Review of Microbiology*, vol. 65, pp. 169–188, 2011.
[4] D. Raoult, P. E. Fournier, and M. Drancourt, "What does the future hold for clinical microbiology?" *Nature Reviews Microbiology*, vol. 2, no. 2, pp. 151–159, 2004.
[5] M. J. Espy, J. R. Uhl, L. M. Sloan et al., "Real-time PCR in clinical microbiology: applications for routine laboratory testing," *Clinical Microbiology Reviews*, vol. 19, pp. 165–256, 2006.
[6] P. Colson, C. Tamalet, and D. Raoult, "SVARAP and aSVARAP: simple tools for quantitative analysis of nucleotide and amino acid variability and primer selection for clinical microbiology," *BMC Microbiology*, vol. 6, article 21, 2006.
[7] D. A. Hysom, P. Naraghi-Arani, M. Elsheikh, A. C. Carrillo, P. L. Williams, and S. N. Gardner, "Skip the alignment: degenerate, multiplex primer and probe design using K-mer matching instead of alignments," *PLoS One*, vol. 7, Article ID e34560, 2012.
[8] C. Frech, K. Breuer, B. Ronacher, T. Kern, C. Sohn, and G. Gebauer, "Hybseek: pathogen primer design tool for diagnostic multi-analyte assays," *Computer Methods and Programs in Biomedicine*, vol. 94, no. 2, pp. 152–160, 2009.

[9] S. N. Gardner, A. L. Hiddessen, P. L. Williams, C. Hara, M. C. Wagner, and B. W. Colston Jr., "Multiplex primer prediction software for divergent targets," *Nucleic Acids Research*, vol. 37, no. 19, pp. 6291–6304, 2009.

[10] K. A. Abd-Elsalam, "Bioinformatic tools and guideline for PCR primer design," *African Journal of Biotechnology*, vol. 2, no. 5, pp. 91–95, 2003.

[11] H. Najafabadi, N. Torabi, and M. Chamankhah, "Designing multiple degenerate primers via consecutive pairwise alignments," *BMC Bioinformatics*, vol. 9, article 55, 2008.

[12] H. S. Najafabadi, A. Saberi, N. Torabi, and M. Chamankhah, "MAD-DPD: designing highly degenerate primers with maximum amplification specificity," *BioTechniques*, vol. 44, no. 4, pp. 519–526, 2008.

[13] R. Souvenir, J. Buhler, G. Stormo, and W. Zhang, "An iterative method for selecting degenerate multiplex PCR primers," *Methods in Molecular Biology*, vol. 402, pp. 245–268, 2007.

[14] E. Gorrón, F. Rodríguez, D. Bernal et al., "A new method for designing degenerate primers and its use in the identification of sequences in Brachiaria showing similarity to apomixis-associated genes," *Bioinformatics*, vol. 26, no. 16, pp. 2053–2054, 2010.

[15] A. Cornish-Bowden, "Nomenclature for incompletely specified bases in nucleic acid sequences: rcommendations 1984," *Nucleic Acids Research*, vol. 13, no. 9, pp. 3021–3030, 1985.

[16] J. Marmur and P. Doty, "Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature," *Journal of Molecular Biology*, vol. 5, pp. 109–118, 1962.

[17] R. B. Wallace, J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, and K. Itakura, "Hybridization of synthetic oligodeoxyribonucleotides to φX 174 DNA: the effect of single base pair mismatch," *Nucleic Acids Research*, vol. 6, no. 11, pp. 3543–3558, 1979.

[18] K. J. Breslauer, R. Frank, H. Blocker, and L. A. Marky, "Predicting DNA duplex stability from the base sequence," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 11, pp. 3746–3750, 1986.

[19] S. Rozen and H. Skaletsky, "Primer3 on the WWW for general users and for biologist programmers," *Methods in Molecular Biology*, vol. 132, pp. 365–386, 2000.

[20] A. Untergasser, H. Nijveen, X. Rao, T. Bisseling, R. Geurts, and J. A. Leunissen, "Primer3Plus, an enhanced web interface to Primer3," *Nucleic Acids Research*, vol. 35, pp. W71–W74, 2007.

[21] F. M. You, N. Huo, Y. Q. Gu et al., "BatchPrimer3: a high throughput web application for PCR and sequencing primer design," *BMC Bioinformatics*, vol. 9, article 253, 2008.

[22] M. D. Gadberry, S. T. Malcomber, A. N. Doust, and E. A. Kellogg, "Primaclade—a flexible tool to find conserved PCR primers across multiple species," *Bioinformatics*, vol. 21, no. 7, pp. 1263–1264, 2005.

[23] N. Kim and C. Lee, "QPRIMER: a quick web-based application for designing conserved PCR primers from multigenome alignments," *Bioinformatics*, vol. 23, no. 17, pp. 2331–2333, 2007.

[24] N. Batnyam, J. Lee, J. Lee, S. Bok Hong, S. Oh, and K. Han, "UniPrimer: a web-based primer design tool for comparative analyses of primate genomes," *Comparative and Functional Genomics*, vol. 2012, Article ID 520732, 8 pages, 2012.

[25] A. L. Gervais, M. Marques, and L. Gaudreau, "PCRTiler: automated design of tiled and specific PCR primer pairs," *Nucleic Acids Research*, vol. 38, no. 2, pp. W308–W312, 2010.

[26] S. Arvidsson, M. Kwasniewski, D. M. Riaño-Pachón, and B. Mueller-Roeber, "QuantPrime—a flexible tool for reliable high-throughput primer design for quantitative PCR," *BMC Bioinformatics*, vol. 9, article 465, 2008.

[27] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T. Madden, "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction," *BMC Bioinformatics*, vol. 13, article 134, 2012.

[28] C. Schretter and M. C. Milinkovitch, "OligoFaktory: a visual tool for interactive oligonucleotide design," *Bioinformatics*, vol. 22, no. 1, pp. 115–116, 2006.

[29] A. M. Pessoa, S. Pereira, and J. Teixeira, "PrimerIdent: a web based tool for conserved primer design," *Bioinformation*, vol. 5, pp. 52–54, 2010.

[30] R. Giegerich, F. Meyer, and C. Schleiermacher, "GeneFisher—software support for the detection of postulated genes," *Proceedings of International Conference on Intelligent Systems for Molecular Biology*, vol. 4, pp. 68–77, 1996.

[31] R. Kalendar, D. Lee, and A. H. Schulman, "FastPCR software for PCR primer and probe design and repeat search," *Genes, Genomes and Genomics*, vol. 3, pp. 1–14, 2009.

[32] O. J. Jabado, G. Palacios, V. Kapoor et al., "Greene SCPrimer: a rapid comprehensive tool for designing degenerate primers from multiple sequence alignments," *Nucleic Acids Research*, vol. 34, no. 22, pp. 6605–6611, 2006.

[33] D. Rosenkranz, "EasyPAC: a tool for fast prediction, testing and reference mapping of degenerate PCR primers from Alignments or consensus sequences," *Evolutionary Bioinformatics Online*, vol. 8, article 151, 2012.

[34] B. Dwivedi, R. Schmieder, D. B. Goldsmith, R. A. Edwards, and M. Breitbart, "PhiSiGns: an online tool to identify signature genes in phages and design PCR primers for examining phage diversity," *BMC Bioinformatics*, vol. 13, article 37, 2012.

[35] J. D. Gans, J. Dunbar, S. A. Eichorst, L. V. Gallegos-Graves, M. Wolinsky, and C. R. Kuske, "A robust PCR primer design platform applied to the detection of Acidobacteria Group 1 in soil," *Nucleic Acids Research*, vol. 40, no. 12, article e96, 2012.

[36] W. A. Kibbe, "OligoCalc: an online oligonucleotide properties calculator," *Nucleic Acids Research*, vol. 35, pp. W43–W46, 2007.

Discussion

In this study, we first reviewed the role of external protrusion, fibers and membrane proteins in entry into host cells and virus-host interactions in large and giant DNA viruses, which include members from the proposed order *Megavirales*. Notably, we highlighted the role of membrane proteins in recognizing their host cells and attaching to their surface.

Then, we provided evidence based on RNA interference that proteins encoded by Mimivirus R135, L725, L829, and R856 genes are involved in Mimivirus fibers formation. Such evidence relied on the observation of the fiber topology by electron microscopy, on proteomic analyses coupled with immunological reactivities including immunogold analysis and 1D and 2D gel electrophoresis followed by western blotting. Our results showed that the four fiber associated candidate proteins could be components of the Mimivirus fibers or involved in their biosynthesis during assembly of the giant virions. In addition, silencing the L425 capsid protein encoding gene thaws not associate with modifications of the fiber layer, which indicates the specificity of the short interfering RNAs and that the loss or shape deformation of the Mimivirus fibers resulted from RNA interference, and was not artefactual and due to the lipid carrier or off-targeting effects.

Moreover, we provided further evidence based on comparative genomics that L725 protein is encoded only by members of the family *Mimiviridae*, whereas, the other fiber associated proteins (R135, L829, and R856) share sequence similarity with proteins of members from other domains of life, including protozoa, metazoa, bacteria, archaea, or fungi. Interestingly, these proteins do not share sequence homology with any known protein encoded by viruses except the members of the family *Mimiviridae*, and pandoraviruses and *P. sibericum*. It is not known whether Mimivirus acquired these proteins from common ancestors or lateral gene transfer. Strikingly, the Mimivirus fibers topology and length differ substantially from those of other viral fibers or protrusions. The length of these fibers also differs between members of the family *Mimiviridae*. Concurrently, the Mimivirus fibers could be compared to the external protrusions of organisms from other domains of life, including eukaryotes. To date, it remains unclear if phagocytic protists are the only host of mimiviruses. Taken together, these findings highlight the uniqueness of mimiviruses among viruses, and the importance of studying their structure, replicative cycle and evolution.

We also provided during this thesis preliminary evidence that the recombinant and purified L725 protein is capable in vitro, after cleavage from its fusion proteins, to form a fiber-like structure, which could be helicoidal. This purified protein was investigated by western blot, proteomics and electron microscopy and tomography. The possible self-assembly of the L725 gene product is another argument that it is one of the components of the Mimivirus fibers. Additionally, the role of Mimivirus fibers on co-infection of amoebae with the Sputnik virophage was studied using RNA interference of the Mimivirus genes encoding fiber candidate proteins. Taken together, results obtained by qPCR, ELISA and proteomic analyses showed that the quantity of Sputnik was greater in co-infection with silenced Mimivirus compared to non-silenced Mimivirus. This was a rather counter-intuitive preliminary result, as the virophage was suspected to interact with the Mimivirus fibers for its replication. Thus, the roles of the Mimivirus fiber-associated proteins for the replication of this giant virus and its virophage remain to be fully deciphered.

Overall, we used here a novel approach that uses RNA interference to study the expression and functions of the giant virus genes. To our knowledge, the study of the functions of giant virus proteins using siRNA was never achieved in a previous work. Similarly, studying the effect of RNA interference of Mimivirus genes on the Sputnik virophage replication was not reported previously. Our observation of the self-assembly of a Mimivirus protein was also original. Although, our studies allowed to gain a better knowledge of the implication of four Mimivirus gene products in the formation of the Mimivirus fibers, the precise structure of these fibers remain to be characterized. In addition, the precise contribution of each of these proteins in the formation of the Mimivirus fibers deserves further studies.

Perspectives and future directions

With respect to the uniqueness and amazing features of the mimiviruses of amoeba, studying their natural hosts in future researches is, beyond any doubt, of substantial interest. To date, Acanthamoeba spp. are the only known hosts for mimiviruses, though most of the details about the giant virus tropism for these amoebae are unresolved (Raoult, La Scola et al. 2007, Pagnier, Reteno et al. 2013, Silva, Almeida et al. 2013). An increasing body of evidence shows that mimiviruses are common in our environment and, most importantly, are present and potentially pathogenic in humans (Saadi, Pagnier et al. 2013, Saadi, Reteno et al. 2013), and mimivirus replication was described in macrophages and peripheral blood mononuclear cells (Ghigo, Kartenbeck et al. 2008, Silva, Almeida et al. 2013). Taken together, these findings prompt further investigations to decipher the mechanisms of entry of these viruses into eukaryotic cells and determine if they can replicate efficiently in the human body.

The method implemented in the present thesis work to knock down Mimivirus genes is promising. Thus, the study of Mimivirus fiber candidate proteins by RNA interference was a proof of concept that such approach could provide a wealth of data on the expression and function of the Mimivirus genes. It is worthy here to point out that more than half of the Mimivirus genes are ORFans, i.e. genes with no homolog in sequence databases, or unannotated genes (Raoult and Boyer 2010). This is also the case for other mimiviruses, and substantial amounts of hypothetical proteins are also found in the other *Megavirales* members, which are greater than proportions described for bacteria or archaea with a similar genome size including Mycoplasma pneumoniae (GenBank Accesssion no. NC_016807; 29%), Tropheryma whipplei (AE014184; 31%) or Ignicoccus hospitalis (CP000816; 39%). For instance, about one third of the marseillevirus genes and up to 93% of the pandoravirus genes are ORFans (Philippe, Legendre et al. 2013, Aherfi, Boughalmi et al. 2014). The function of only a few Mimivirus proteins were confirmed experimentally, including those of amino-acyl-tRNA synthetases (Abergel, Chenivesse et al. 2005, Abergel, Rudinger-Thirion et al. 2007) or proteins involved in sugar biosynthesis (Bandaru, Zhao et al. 2007, Benarroch, Qiu et al. 2009, Piacente, Marin et al. 2011). As giant viruses are characterized by their huge genome and large numbers of genes encoding unknown functions, we believe that the method presented in this study can be applied to unveil the function of other proteins from the so-called “dark

matter". Certainly, investigations should be conducted to determine the structure of the Mimivirus fibers, the reasons for the variations in the fiber topology among mimiviruses, the role of the fiber proteins in virus-host interactions and protein-protein interactions during virus entry. Notwithstanding, RNA interference could further provide experimental avenues to decipher the functions of the numerous ORFans and hypothetical proteins of giant viruses of amoeba, and provide the annotation of dozens of hypothetical proteins from sequences databases.

Besides, at least two sets of proteins strongly deserve to be studied. The first set encompasses proteins with putative functions in translation. This was very surprising when analyzing the Mimivirus genome to find genes encoding proteins from the translation apparatus (Raoult, Audic et al. 2004). The Mimivirus set of such proteins was unique amongst viruses. It comprised 4 aminoacyl-tRNA synthetases and initiation, elongation, and release factors, and 6 transfer RNAs. The putative functionality of amino-acyl-tRNA synthetases was checked experimentally (Abergel, Chenivesse et al. 2005, Abergel, Rudinger-Thirion et al. 2007). Nevertheless, it remains to be determined if, beyond being encoded, proteins putatively involved in Mimivirus protein translation are needed to complete the giant viral replicative cycle. Another set of megaviral genes that may deserve functional studies is the one composed of about 50 genes inferred to be present in the common ancestor of the *Megavirales* (Yutin, Wolf et al. 2009).

Our strategy of detecting giant virus genes to identify their functions will benefit in deciphering function of other ORFans or putative proteins. During the thesis work, we used RNA interference to knock down one of the translation proteins, a putative translation initiation factor 4a (R458; GenBank Accession ID: YP_003986965). We aim to use this strategy to target large dataset of genes encoding proteins that have putative roles in translation or other cellular process. This strategy will benefit in differentiating between junk DNA and truly encoded genes.

Finally, the homology between mimiviruses proteins and other proteins from members belonging to the cellular domains of life also highlights the importance of their investigation. This should be helpful to gain a better understanding of several aspects in virology and evolutionary biology.

References

Abergel, C., S. Chenivesse, D. Byrne, K. Suhre, V. Arondel and J. M. Claverie (2005). "Mimivirus TyrRS: preliminary structural and functional characterization of the first amino-acyl tRNA synthetase found in a virus." Acta Crystallogr Sect F Struct Biol Cryst Commun **61**(Pt 2): 212-215.

Abergel, C., J. Rudinger-Thirion, R. Giege and J. M. Claverie (2007). "Virus-encoded aminoacyl-tRNA synthetases: structural and functional characterization of mimivirus TyrRS and MetRS." J Virol **81**(22): 12406-12417.

Aherfi, S., M. Boughalmi, I. Pagnier, G. Fournous, B. La Scola, D. Raoult and P. Colson (2014). "Complete genome sequence of Tunisvirus, a new member of the proposed family Marseilleviridae." Arch Virol **2014**: 26.

Bandaru, V., X. Zhao, M. R. Newton, C. J. Burrows and S. S. Wallace (2007). "Human endonuclease VIII-like (NEIL) proteins in the giant DNA Mimivirus." DNA Repair (Amst) **6**(11): 1629-1641.

Benarroch, D., Z. R. Qiu, B. Schwer and S. Shuman (2009). "Characterization of a mimivirus RNA cap guanine-N2 methyltransferase." Rna **15**(4): 666-674.

Ghigo, E., J. Kartenbeck, P. Lien, L. Pelkmans, C. Capo, J. L. Mege and D. Raoult (2008). "Ameobal pathogen mimivirus infects macrophages through phagocytosis." PLoS Pathog **4**(6): e1000087.

Pagnier, I., D. G. Reteno, H. Saadi, M. Boughalmi, M. Gaia, M. Slimani, T. Ngounga, M. Bekliz, P. Colson, D. Raoult and B. La Scola (2013). "A decade of improvements in *Mimiviridae* and Marseilleviridae isolation from amoeba." Intervirology **56**(6): 354-363.

Philippe, N., M. Legendre, G. Doutre, Y. Coute, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J. M. Claverie and C. Abergel (2013). "Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes." Science **341**(6143): 281-286.

Piacente, F., M. Marin, A. Molinaro, C. De Castro, V. Seltzer, A. Salis, G. Damonte, C. Bernardi, J. M. Claverie, C. Abergel and M. Tonetti (2011). "Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine)." J Biol Chem **287**(5): 3009-3018.

Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan and J. M. Claverie (2004). "The 1.2-megabase genome sequence of Mimivirus." Science **306**(5700): 1344-1350.

Raoult, D. and M. Boyer (2010). "Amoebae as genitors and reservoirs of giant viruses." Intervirology **53**(5): 321-329.

Raoult, D., B. La Scola and R. Birtles (2007). "The discovery and characterization of Mimivirus, the largest known virus and putative pneumonia agent." Clin Infect Dis **45**(1): 95-102.

Saadi, H., I. Pagnier, P. Colson, J. K. Cherif, M. Beji, M. Boughalmi, S. Azza, N. Armstrong, C. Robert, G. Fournous, B. La Scola and D. Raoult (2013). "First isolation of mimivirus in a patient with pneumonia." Clin Infect Dis **57**(4): e127-134.

Saadi, H., D. G. Reteno, P. Colson, S. Aherfi, P. Minodier, I. Pagnier, D. Raoult and B. La Scola (2013). "Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia." Intervirology **56**(6): 424-429.

Silva, L. C., G. M. Almeida, D. B. Oliveira, F. P. Dornas, R. K. Campos, B. La Scola, P. C. Ferreira, E. G. Kroon and J. S. Abrahao (2013). "A resourceful giant: APMV is able to interfere with the human type I interferon system." Microbes Infect **4579**(13): 00253-00250.

Yutin, N., Y. I. Wolf, D. Raoult and E. V. Koonin (2009). "Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution." Virol J **6**: 223.