Virus d'archées : interaction avec un hôte hyperthermophile, isolement d'un virus d'habitat géothermique, motifs courts exceptionnels dans les génomes

Ariane Bize

▶ **To cite this version:**

Ariane Bize. Virus d'archées : interaction avec un hôte hyperthermophile, isolement d'un virus d'habitat géothermique, motifs courts exceptionnels dans les génomes. Sciences agricoles. Université Paris-Est, 2009. Français. NNT : 2009PEST1052 . tel-00539609

HAL Id: tel-00539609

https://pastel.hal.science/tel-00539609

Submitted on 24 Nov 2010

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

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

UNIVERSITÉ PARIS-EST

Mémoire de thèse présenté pour l'obtention du titre de
DOCTEUR DE L'UNIVERSITÉ PARIS-EST

Spécialité: Biologie

par

Ariane BIZE-D'AVEZAC

Virus d'archées :
interaction avec un hôte hyperthermophile, isolement d'un virus d'habitat géothermique, motifs courts exceptionnels dans les génomes

Thèse soutenue le 3 avril 2009 devant le jury composé de :

Rapporteurs :	Mme Pascale Boulanger-Biard
	Mme Marie-Claude Serre
Président :	M. Bruno Tassin
Examinateur :	M. Vincent Maréchal
Directeurs de thèse	M. David Prangishvili
	M. Olivier Tenaillon

A mes parents,
à Tilon,
et Pol

Je souhaite sincèrement remercier les nombreuses personnes qui m'ont apporté une aide précieuse et indispensable lors de cette aventure qu'est une thèse.

Je tiens à exprimer toute ma reconnaissance et mon admiration pour Patrick Forterre, qui, lorsque j'ai frappé à sa porte, m'a accueillie avec bienveillance et m'a offert la possibilité de travailler au sein de l'unité BMGE, alors qu'il ne connaissait pas grand chose de moi. C'était une chance d'être au sein d'une unité très dynamique, nouvellement créée et riche en thématiques. Je remercie également Patrick pour ses conseils et son écoute. Sa passion, communicative, pour l'évolution, les virus d'archées et le virus du voyage, a été une grande source de motivation.

Je souhaite vivement remercier mes directeurs de thèse, David Prangishvili et Olivier Tenaillon, pour leur encadrement très complémentaire. Je remercie David de m'avoir fait participer à des projets très variés, et permis d'interagir avec des collaborateurs d'horizons divers, faisant de cette thèse une expérience très riche. Un très grand merci à Olivier, qui s'est investit dans un projet s'éloignant progressivement de ses thématiques principales de recherche. Je le remercie pour les discussions sur l'évolution des virus et son aide très efficace. Merci aussi pour son « coaching » tout au long de la thèse, sa gentillesse et sa franchise en toutes circonstances.

Je remercie vivement Sophie Schbath de m'avoir encadrée pour le projet bioinformatique. Son aide était d'une grande qualité, ses réponses très pédagogiques et les discussions toujours approfondies. C'était un réel plaisir de travailler et d'apprendre avec elle.

Je remercie Pascale Boulanger et Marie-Claude Serre d'avoir accepté le lourd travail de rapporteur. Je les remercie pour le temps qu'elles y ont consacré, pour leurs conseils et leurs encouragements. Je remercie Vincent Maréchal, d'avoir accepté de faire partie de mon jury de thèse, et pour ses commentaires instructif, et remercie Bruno Tassin, d'avoir bien voulu présider le jury et de m'avoir suivie tout au long de mon parcours à l'ENPC et en thèse.

Merci beaucoup à Simonetta Gribaldo et Tamara Basta pour la relecture critique de chapitres de ce manuscrit de thèse. Merci également à ma tutrice Monica Sala-Schaeffer.

Je remercie les collaborateurs des différents projets pour tout ce qu'ils m'ont appris, notamment Soizick Lucas-Staat, sur les archées hyperthemophiles, Diego Cortez, sur la bioinformatique, Erik Karlsson et Rolf Bernander, sur la cytométrie en flux, Marie-Christine Prevost, sur la microscopie électronique, Nicole Steinmetz, sur les bionanotechnologies, Xu Peng et Roger Garrett, sur le séquençage des génomes. Merci à toutes ces personnes pour leur disponibilité malgré des emplois du temps des plus chargés, leur dynamisme et leur gentillesse. Merci tout particulièrement à Marie-Christine Prevost de m'avoir initiée à la microscopie électronique au sein de la plate-forme qu'elle dirige.

Je remercie Chantal Archambeau, Gislaine Lebon, et Evelyne Mauve, du laboratoire de préparation, avec qui il est très agréable de travailler et qui facilitent grandement la vie des chercheurs. Merci aux personnes des plates-formes et ateliers pour leur aide, notamment Gérard Pehau-Arnaudet, Pascal Lenormand, Yannick Archambeau, et Marie-Christine Wagner, ainsi qu'aux personnes du bâtiment Fernbach pour l'entre-aide et le prêt de matériel.

Je remercie chaleureusement tous les membres passés et présents de BMGE, pour leur aide informelle constante et pour leur amitié. Les discussions à bâtons rompus du traditionnel « coffe talk » permettaient aussi bien d'échanger les dernières nouvelles scientifiques que de célébrer les anniversaires, d'admirer les photos de voyage des uns et des autres et de déguster les spécialités culinaires correspondantes. Grand merci à Ana Cova pour l'attention qu'elle porte à chacun, et qui est pour moi un exemple en matière d'organisation. Grand merci à Nicole Desnoues, pour sa légendaire connaissance des techniques de biologie moléculaire et des anecdotes sur l'histoire de l'Institut Pasteur, qu'elle partage, l'une comme l'autre, volontiers. Grand merci à Soizick Lucas-Staat pour sa gentillesse et pour les séances partagées de microscopie électronique et de yoga. Merci également à Laurent Debarbieux pour la qualité de son écoute et de ses réponses. Merci à Guennadi Sezonov pour ses conseils et les discussions du soir, à l'heure à laquelle débarque le fantôme du couloir (dont le bruit rappelle étrangement un vieux frigo). Merci à Simonetta Gribaldo de m'avoir fait découvrir le monde passionnant de la phylogénie et donné l'envie de découvrir la Sicile. A Peter Redder, qui m'a fait profiter de sa grande culture en matière de biochimie, de bières, du Danemark et beaucoup d'autres choses, merci pour le « fire holding ». A Madame Tamara Basta, merci pour son aide précieuse en biologie moléculaire, et sa constante bonne humeur (mais comment fait-elle ?). A Diego Cortez, merci pour ses explications patientes sur la bioinformatique, et les discussions et sorties autour de la salsa. A Mery Pina, merci pour les soirées inoubliables, et à Nuno Peixeiro, « obrigada ». Merci également à Elie Desmond, Chloé Terras, Tomohiro Mochizuki, Alexis Criscuolo, Carole Jaubert, Damien Maura, Eric Morello, Tessa Quax, Céline Petitjean, Jean-Marie Clément, Claudine Elmerich, Mireille Leduc, Monika Häring. Merci aux personnes de l'équipe soeur BMGE Orsay et à toutes celles que je n'ai pas citées mais n'oublie pas. Merci à tous pour les événements partagés, des divers mariages et enterrements de vie de jeune fille aux soirées quizz-bières.

Un grand merci également à l'équipe d'Ecologie et évolution des microorganismes pour son accueil et son amitié, avec Erick Denamur, Alix Cavalié, Pierre-Alexis Gros. De l'Institut Pasteur, je remercie par ailleurs le bureau de l'ASIP et Nouchka Ovtchinikoff, notre très appréciée et indispensable professeur de yoga.

Je remercie « les Ponts » de m'avoir permis de bénéficier du cursus de formation par la recherche, et le Cemagref, de m'avoir accueillie et permis de finir ma thèse dans de bonnes conditions. Merci donc tout particulièrement à Théodore Bouchez.

Je remercie également l'école doctorale EDVE, notamment Catherine Alcouffe.

Pour finir, je souhaite remercier ma famille et ma belle-famille pour leur affection, qui m'est indispensable et me porte chaque jour. Merci tout particulièrement à mes parents, de m'avoir encouragée dans mes études et transmis leur goût pour les sciences.

Merci infiniment à Pol, mon mari, qui m'a apporté bien plus que du soutien moral et a fait preuve d'une patience incroyable.

Résumé

Les microorganismes du domaine du vivant Archaea sont très divers sur le plan biologique et sont présents dans de nombreux types d'écosystèmes. Ils sont majoritaires dans les environnements dits extrêmes. Parmi les virus d'archées, ceux infectant les espèces d'un phylum majeur des archées, Crenarchaeota, constitué d'hyperthermophiles, forment un groupe exceptionnel. En effet, leurs morphotypes sont uniques, variés, et complexes. Le contenu de leur génome est également unique. Enfin, la plupart de ces virus se maintiennent dans la cellule hôte en état porteur, une relation chronique qui permet un équilibre entre production de virions et division cellulaire.

J'ai d'abord démontré que le virus de crenarchée *Sulfolobus islandicus* rod-shaped virus 2 est un virus virulent, et non chronique comme il avait été suggéré. Un mécanisme de lyse unique a été découvert. La paroi cellulaire est modifiée en plusieurs points, avec l'apparition de structures pyramidales saillantes. Celles-ci s'ouvrent en fin de cycle infectieux, permettant aux virions, assemblés auparavant dans le cytoplasme, de quitter la cellule. Puis j'ai travaillé sur des échantillons de sources géothermiques de la péninsule de Kamchatka (Russie) et contribué à l'isolement et la caractérisation d'un virus de morphotype filamenteux. Des protéines structurales supplémentaires ont ainsi été identifiées. Enfin, des mots courts exceptionnels ont été identifiés dans un grand nombre de génomes d'archées et de leurs éléments extra-chromosomiques. Ce sont potentiellement des motifs fonctionnels non-codants, impliqués dans des mécanismes biologiques importants. Typiquement, les motifs palindromiques sont évités dans les génomes.

Mots clés : Archaea, virus, Sulfolobus, cycle infectieux, interaction hôte-virus, motifs exceptionnels.

Thèse réalisée à l'Institut Pasteur,
dans l'unité de Biologie Moléculaire du Gène chez les Extrêmophiles,
située au 25 rue du Docteur Roux, 75724 Paris Cedex 15. France.

Title

Archaeal viruses: interaction with a hyperthermophilic host,
isolation of a virus from a geothermal environment,
short exceptional motifs in genomes.

Summary

The microorganisms from the Archaea domain are very diverse at the biological level and they are present in many types of ecosystems. They are dominant in the so-called extreme environments. Among their viruses, those infecting species of the Crenarchaeota phylum, a major archaeal phylum comprising hyperthermophiles, form an exceptional group. Indeed, their morphotypes are unique, diverse, and complex. Their genome content is also unique. Finally, most of these viruses persist in the host cell in a carrier state, a chronic relationship allowing an equilibrium between virion production and cell division.

I first proved that *Sulfolobus islandicus* rod-shaped virus 2 is a virulent virus, and not chronic, as had previously been suggested. A unique lysis mechanism was discovered. The cell wall is modified in several locations, with the appearance of pyramidal prominent structures. Those burst open at the end of the infection cycle, allowing the virions, previously assembled in the cytoplasm, to leave the cell. Then, I worked on environmental samples from geothermal springs of the Kamchatka peninsula (Russia) and contributed to the isolation and characterization of a virus of filamentous morphotype. Additional structural proteins were in particular identified. Finally, short exceptional words were identified in a great number of genomes from archaea and their extra-chromosomal elements. These are potentially functional non-coding motifs involved in important biological mechanisms. Typically, palindromic motifs are avoided in the genomes.

Key words: Archaea, virus, Sulfolobus, infection cycle, host-virus interaction, exceptional motifs.

Table des matières

Partie I – Introduction **13**

1 Introduction générale **14**

2 Archaea, troisième domaine du vivant **16**
2.1 Découverte des archées et émergence du domaine Archaea 16
2.2 Principales caractéristiques des archées 17
2.3 Diversité des archées 20
2.3.1 D'extrêmophiles et spécialisées, à mésophiles et ubiquitaires 20
2.3.2 Phylogénie et taxonomie des archées 23
2.4 Phénotype et habitat des principaux groupes d'archées 25
2.4.1 Méthanogènes 25
2.4.2 Extrêmes halophiles 26
2.4.3 Thermophiles, hyperthermophiles et acidophiles 27
2.4.4 Archées non-cultivables ou cultivables depuis peu 28
Milieu marin et plancher océanique 28
Autres milieux 29
2.5 Questions évolutives en lien avec le domaine Archaea 30
2.5.1 La classification en trois domaines, une révolution 30
2.5.2 Les relations évolutives entre les trois domaines du vivant 33
2.5.3 L'évolution des archées 34

3 Virosphère et virus d'archées **36**
3.1 La découverte des virus et quelques grandes phases de recherche 36
3.1.1 Découvert des virus infectant des hôtes eucaryotes 36
3.1.2 Découverte des virus infectant des hôtes bactériens 36
3.1.3 Et les virus d'archées ? 37
3.2 Ecologie et diversité des virus 38
3.3 Les virus d'archées 40
3.3.1 Les virus d'euryarchées 40
3.3.2 Les virus de crenarchées 42
3.4 Virus de crenarchées hyperthermophiles et évolution 45

4 Présentation du travail réalisé **48**

Partie II – Sulfolobus islandicus rod-shaped virus 2 et interaction hôte-virus chez les crenarchées **49**

1 Introduction **50**

2 Interaction hôte-virus chez les crenarchées : état de l'art 51
2.1 L'état porteur, la relation hôte-virus majoritaire chez les crenarchées 51
2.2 Des preuves expérimentales encore insuffisantes 52

3 Le modèle biologique 55
3.1 Le genre *Sulfolobus* 55
3.1.1 Principales caractéristiques 55
3.2.1 *Sulfolobus*, organisme modèle 56
3.1.3 Exemples d'avancées importantes réalisées grâce à *Sulfolobus* 57
3.2 *Sulfolobus islandicus* rod-shaped virus 2, SIRV2 59
3.2.1 Principales caractéristiques 59
3.2.2 Vers un virus modèle ? 60

4 Présentation de l'étude 62

Manuscrit soumis : « A unique virus release mechanism in Archaea » 63
References 77
Figures 81
Supporting Materials 89
Article : « Site-specific and spatially controled addressability of a new viral nanobuilding block: *Sulfolobus islandicus* rod-shaped virus 2 »
Supporting Information
Stabilité des particules virales de SIRV2 à haute température 107
Article : « *Stygiolobus* rod-shaped virus and the interplay of crenarchaeal rudiviruses with the CRISPR antiviral system »
Supplementary Table

Partie III – Isolement et caractérisation d'un virus filamenteux et de son hôte, à partir d'échantillons environnementaux de sources géothermiques de la péninsule de Kamchatka 109

1 Présentation de l'étude 110

Article : « Viruses in acidic geothermal environments of the Kamchatka Peninsula »

Partie IV – Motifs courts exceptionnels dans les génomes des archées, et de leurs éléments extra-chromosomiques 111

1 Présentation de l'étude 112
1.1 Motifs fonctionnels non-codants et cycle cellulaire 112
1.2 Motifs palindromiques 113
1.3 Approche développée 114

2 Méthodes **116**
2.1 Séquences utilisées 116
2.2 Analyse en composante principale et clustering hiérarchique 118
2.3 Détection des motifs exceptionnels 118
2.4 Exceptionnalité d'ensembles de mots dans les séquences 120
2.5 Significativité de l''évitement du mot CCTGGG dans certaines régions des génomes de *Pyrococcus* 120
2.6 Position privilégiée des mots par rapport aux gènes 121

3 Résultats et discussion **122**
3.1 Composition en mots courts des séquences de virus et de plasmides 122
3.1.1 Observations générales 122
3.1.2 Interprétation et concaténation des séquences 123
3.1.3 Remarques sur l'évolution des virus et plasmides 126
3.2 Analyse des palindromes 126
3.2.1 Exceptionnalité des palindromes correspondant à des systèmes de restriction-modification connus 126
3.2.2 Rangs de palindromes et pseudo-palindromes de tailles 4 à 6 dans les séquences 128
3.2.3 Évitement global des palindromes 132
3.3 Analyse des mots symétriques 134
3.4 Analyse d'autres motifs très exceptionnels 136
3.4.1 A la recherche de motifs impliqués dans des fonctions générales, de type maintenance du génome 136
3.4.2 Répétitions d'une même lettre 137
3.4.3 AATCC et les CRISPRs dans les génomes de *Sulfolobus* 139
3.4.4 CCTGGG/CCCAGG dans les génomes de *Pyrococcus* 139

4 Conclusions **142**

Tableaux détaillés **143**

Partie V – Discussion **151**

1 Résumé des résultats obtenus **152**
1.1 Virus SIRV2 152
1.1.1 Interaction hôte-virus 152
1.1.2 Données sur la particule virale de SIRV2 153
1.2 Enrichissements d'échantillons environnementaux de sources chaudes 154
1.3 Motifs courts exceptionnels dans les génomes d'archées et de leurs éléments extra-chromosomiques 154

2 Des pistes pour poursuivre sur le virus SIRV2 **157**
2.1 Mise au point d'outils de génétique 157
2.2 Mise au point d'un nouveau protocole de purification du virus 157

2.3 Etude du cycle infectieux 158
2.4 Aspects moléculaires 159

3 Interaction hôte-virus chez les crenarchées 162
3.1 La sortie des virus, un problème mécanique 162
3.2 Les observations effectuées sur le virus SIRV2 ne peuvent être directement généralisées 162
3.3 Virulence, environnement et évolution 163

Partie VI – Conclusion 165
Virus d'archées hyperthermophiles et nouveaux mécanismes biologiques 166

Bibliographie 169

Abbréviations 186
Noms des virus 187
Table des illustrations 188
Figures 188
Tableaux 189

Introduction

1
Introduction générale

La découverte des archées par Carl Woese, dans les années 70, a bouleversé notre conception du monde vivant, de sa diversité et de son évolution. Elle a mené à une nouvelle classification des organismes en trois domaines, Archaea, Bacteria et Eucarya. Un nouveau champ de recherche en microbiologie s'est ouvert.

Une première approche est centrée sur la biologie des archées et vise à caractériser les mécanismes moléculaires fondamentaux. Il s'agit par exemple de déterminer ce qu'elles ont en commun avec les organismes des autres domaines et ce qui leur est propre, afin de mieux comprendre les relations évolutives entre les trois domaines. Ou bien encore, de décrypter les mécanismes et les propriétés moléculaires, potentiellement uniques, qui permettent à de nombreuses archées de coloniser les environnements considérés comme les plus extrêmes.

Une seconde approche, tournée vers l'écologie, consiste à explorer la diversité des migroorganismes des trois domaines dans l'environnement, par des approches de séquençage. L'utilisation des séquences de la petite sous-unité de l'ARN ribosomique (PSU-ARNr), développée dans les années 70 et ayant mené à la découverte des archées, permet de répondre à la question : quels microorganismes trouve-t-on dans les environnement, et comment sont-ils distribués ? Les progrès réalisés dans les techniques de séquençage permettent depuis peu d'aborder la question de la fonction des microorganismes au sein des écosystèmes (métabolisme), par des approches spécifiques de métagénomique.

Les deux approches sont complémentaires. La découverte de nouveaux clades dans l'environnement stimule les travaux d'isolement puis de caractérisation de nouveaux microorganismes. Par ailleurs, le fruit de toutes ces recherches alimente la réflexion sur l'évolution du monde vivant, et la compréhension des mécanismes de cette évolution. Les travaux en cours de Rachel Whitaker, sur l'archée acidothermophile *Sulfolobus islandicus*, constituent un bon exemple de synthèse (écologie évolutive des populations microbiennes).

Comme tous les organismes cellulaires, les archées sont infectées par des virus. La recherche sur les virus d'archées s'est développée dans le même élan : les deux types d'approches (décortiquer les mécanismes biologique et évaluer la diversité) sont aujourd'hui développés en parallèle par quelques laboratoires dans le monde. Les premiers virus d'archées ont été isolés et étudiés en tant que tels dès le début des années 80, sous l'impulsion de Wolfram Zillig. Les virus infectant l'un des trois phylums majeurs des archées, les crenarchées hyperthermophiles, sont apparus comme un groupe très étonnant, par la diversité, l'unicité et la complexité de leurs morphotypes. Cette découverte enthousiasmante a dynamisé les recherches : plusieurs systèmes hôtes-virus ont été isolés et sont aujourd'hui déjà bien maîtrisés. Ils sont caractérisés en détail, autant que faire se peut. Ainsi, les aspects génomiques, structuraux et mécanistiques (transcription, réplication) sont abordés. D'autre part, l'effort d'observation, d'isolement et de caractérisation de nouveaux virus continue, et des études de métagénomique environnementale des virus sont entreprises.

Au cours de ma thèse, réalisée au sein de l'unité de Biologie Moléculaire du Gène chez les Extrêmophiles (BMGE), j'ai eu recours aux deux types d'approches. Contribuer à la caractérisation des relations hôte-virus chez les crenarchées hyperthermophiles, et tenter d'identifier des motifs fonctionnels non-codants dans les génomes des archées, de leurs virus et leurs plasmides, vise d'abord à mieux comprendre le fonctionnement de ces organismes. Le travail d'enrichissement d'échantillons de sources géothermiques, dans le but d'isoler des virus, s'inscrit, lui, dans une optique de caractérisation de la diversité.

En introduction de ce travail transversal, j'expose des données générales sur le domaine des archées et leurs virus, en insistant sur des éléments historiques, sur les aspects de diversité et d'évolution. Les éléments de contexte plus spécifiques sont intégrés aux divers chapitres, avant la présentation des résultats correspondants.

2
Archaea, troisième domaine du vivant

La découverte des archées est récente à l'échelle de l'histoire de la recherche en biologie. C'est au niveau moléculaire qu'elles se distinguent des autres organismes, et non en termes morphologiques, aussi leur découverte a-t-elle coïncidé avec le développement des techniques de biologie moléculaire (2.1). Les archées sont, encore aujourd'hui, moins bien caractérisées que les deux autres domaines du vivant et parfois moins connues au sein de la communauté des biologistes. Je présenterai donc en détail les caractéristiques propres à ce domaine du vivant (2.2) et mentionnerai le fait que les archées sont ubiquitaires et impliquées dans les cycles biogéochimiques de la planète, alors qu'elles étaient au départ considérées comme des organismes à niches très particulières (2.3). J'insisterai également sur leur diversité (2.3, 2.4). Pour finir, je présenterai certaines questions évolutives en rapport avec la découverte du domaine Archaea (2.5), car celle-ci pose ou repose de nombreuses questions.

2.1 Découverte des archées et émergence du domaine Archaea

Les archées ont été découvertes en 1977 par Carl Woese et ses collaborateurs (Woese and Fox, 1977). Considérant que la dichotomie en vigueur, entre eucaryotes et procaryotes, reflétait l'évolution de l'organisation des cellules (présence ou absence de noyau), et non l'évolution des composants cellulaires de base, Carl Woese développa des méthodes permettant de comparer des organismes, mêmes distants, et d'établir des phylogénies du vivant ; le but était d'obtenir une classification reflétant fidèlement l'évolution moléculaire.

Ainsi, Carl Woese développa la première méthode basée sur les PSU-ARNr (ex : Fox et al., 1977). Elle consistait à extraire les PSU-ARNr, les digérer avec l'enzyme RNase T1, et à séparer les fragments obtenus par électrophorèse bi-dimensionnelle (Figure 1). Une distance était définie et calculée entre les organismes deux à deux, directement à partir des motifs de restriction observés sur les gels bi-dimensionnels, et, peu après, à partir des séquences des fragments de digestion séparés par électrophorèse. Cette deuxième méthode permettait d'associer un catalogue d'oligonucléotides à chaque organisme. L'analyse des distances obtenues permettait de distinguer les différents groupes formés par les organismes.

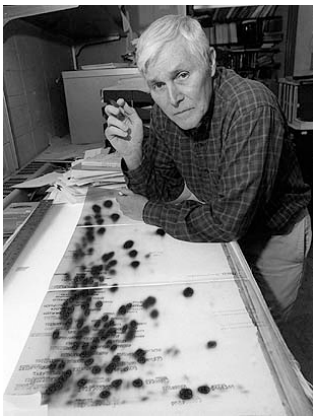


Figure 1 : Carl Woese photographié a côté d'un gel d'électrophorèse bi-dimensionnel

Il s'avéra que les microorganismes alors connus et étudiés sous le nom de « bactéries méthanogènes » constituaient un groupe (« urkingdom ») aussi distant des autres procaryotes que des « cytoplasmes d'eucaryotes » (Woese and Fox, 1977). Ce groupe fut nommé « archaeabacteria », évoquant le caractère potentiellement très ancien de ses membres. Cette idée était basé sur le fait que l'atmosphère ancienne aurait contenu beaucoup de méthane. Or le groupe comprenait alors des représentants des Methanobacteriales, notamment l'organisme thermophile *Methanobacterium thermoautotrophicum*, et des Methanosarcinales.

Par la suite, l'exploration du monde procaryote se poursuivit et l'arbre du vivant s'enrichit (Fox et al, 1980) de nouvelles « archaeabacteria », comme des halophiles extrêmes (Magrum et al., 1978) et des hyperthermophiles (Stetter et al., 1987). Enfin, en 1990, la classification des organismes vivants en 3 domaines, Archaea, Bacteria et Eucarya, fut proposée (Woese et al., 1990). Cette classification est celle utilisée de nos jours. La proposition était étayée par des comparaisons de séquences et des phylogénies (PSU-ARNr), ainsi que sur la comparaison des structures tri-dimensionnelles des PSU-ARNr et d'autres protéines. Chaque domaine était subdivisé en groupes. Pour le domaine Archaea, deux règnes (« kingdom ») étaient définis. Le règne Crenarchaeota comprenait les genres *Pyrodictium* et *Thermoproteus* ; le règne Euryarchaeota comprenait les Thermococcales, Methanococcales, Methanobacteriales, Methanomicrobiales et les halophiles extrêmes.

C'est donc grâce à des critères moléculaires (notamment, propriétés des PSU-ARNr) que les archées ont été découvertes, et que le domaine correspondant a été défini.

2.2 Principales caractéristiques des archées

La spécificité des archées n'est pas limitée aux propriétés des PSU-ARNr. Bien qu'elles présentent des phénotypes divers et colonisent des milieux très variés, les archées ont des caractéristiques communes, liées à une même origine évolutive (groupe monophylétique). Ces propriétés sont bien résumées dans Gribaldo and Brochier-Armanet, 2006. Par ailleurs, le Tableau 1 présente une comparaison synthétique.

	Archaea	*Bacteria*	**Eucaryotes**
Taille typique	1 à 4 µm	1 à 4 µm	Supérieure à 5 µm
Caractéristiques physiologiques	Aérobie et anaérobie	Aérobie et anaérobie, chlorophylienne	Très largement aérobie, chlorophyllienne
Matériel génétique	Petit chromosome circulaire, plasmides et virus ; génome associé avec des histones	Petit chromosome circulaire, plasmides et virus ; pas d'histones	Noyau complexe avec plus d'un chromosome linéaire, virus ; génome associé avec des histones
Différenciation	Généralement unicellulaire, différenciation cellulaire rare	Généralement unicellulaire, différenciation cellulaire rare	Unicellulaire et pluricellulaire, différenciation cellulaire fréquente
Paroi cellulaire	Protéine, glycoprotéine, pseudomuréine, pas de paroi	Peptidoglycane (muréine) et LPS, rarement formée de protéines, formes sans parois rares	Grande variété, absence de peptidoglycane
Membrane cytoplasmique	Ether de glycérol et d'isoprénoïdes, site de la biosynthèse d'énergie	Ester du glycérol et des acides gras, site de la biosynthèse de l'énergie	Esters du glycérol et des acides gras, cholestérol fréquent
Membranes intracytoplasmiques	Généralement absentes	Généralement absentes ; quand présentes, contiennent de grandes quantités de protéines	Fréquentes dans les organites comme les mitochondries et les chloroplastes, le noyau, l'appareil de Golgi, le réticulum endoplasmique et les vacuoles ; site de la biosynthèse de l'énergie
Locomotion	Flagelle simple	Flagelle simple, glissement	Flagelles complexes, cils, pattes, nageoires, ailes
Traduction			
Synthèse protéique	Ribosome 70S, insensible au chloramphénicol et au cycloheximide, diphtamide présente dans le facteur d'élongation	Ribosome 70S, sensible au chloramphénicol, insensible au cycloheximide, diphtamide absente dans le facteur d'élongation	Ribosome 80S et 70S (organites), insensible au chloramphénicol (80S), sensible au cycloheximide (80S), diphtamide présente dans le facteur d'élongation
ARNr	5S, 16S, 23S	5S, 16S, 23S	5 S, 8S, 18S, 28S
Séquence Shine Dalgarno	Présence	Présence	Absence
Transcription			
ARN polymérase	Complexe (1 type, 12 sous-unités)	Simple (1 type, 4 sous-unités)	Complexe (3 types, nombreuses sous-unités)
Mode de fixation de l'ARN Pol sur le promoteur	Protéine TBP (TATA binding protein) Promoteur : séquence TATA située à –25 pb du site d'initiation	Facteur σ Promoteur : séquences situées à –35 pb et –10 pb du site d'initiation	Protéine TBP Promoteur : séquence TATA située à –25 pb du site d'initiation
Réplication de l'ADN			
Origine de réplication	Une ou plusieurs origines de réplication selon les espèces	Une seule origine de réplication *Ori C*	Plusieurs origines de réplication impliquant le complexe ORC
Hélicase	MCM (Minichromosome maintenance complexe)	DnaB	MCM (Minichromosome maintenance complexe)
ADN Polymérases	Famille B et D pour les *Euryarchaeota*, Famille B pour les *Crenarchaeota*	Pol I (famille A), Pol II (famille B), Pol III (famille C)	Principales ADN Pol : Polα, Polδ, Polε (famille B)
Facteur de processivité	PCNA	β clamp	PCNA

Tableau 1: Comparaison des principales caractéristiques des trois domaines du vivants. D'après Briffotaux, 2008.

Une importante spécificité des archées est la composition de leurs membranes cellulaires : les lipides consistent en de longues chaînes d'alcool isoprénoïques liées au glycérol-3-phosphate par des liaisons éther ; dans les deux autres domaines du vivant, il s'agit de deux chaînes d'acides gras assemblées au glycérol-1-phopshate par l'intermédiaire d'une liaison ester (Figure 2). De plus, les membranes des archées sont constituées d'une monocouche de lipides, alors que dans les autres domaines, il s'agit d'une bi-couche. Enfin, la paroi cellulaire des archées ne contient pas de muréine, un peptidoglycane présent chez la plupart des bactéries ; à la place, la paroi cellulaire des archées consiste souvent en la membrane lipidique associée à un « S-layer », situé du côté extracellulaire (Kandler and König, 1998). Le « S-layer » est une enveloppe de glycoprotéines d'aspect cristallin ancrée dans la membrane lipidique. Certaines bactéries ont également un S-layer, qui n'est donc pas une spécificité des archées. Par rapport aux membranes des bactéries et des eucaryotes, celles des archées conservent les propriétés essentielles à la vie cellulaire (imperméabilité aux protons) dans une plus grande gamme de paramètres physico-chimiques.

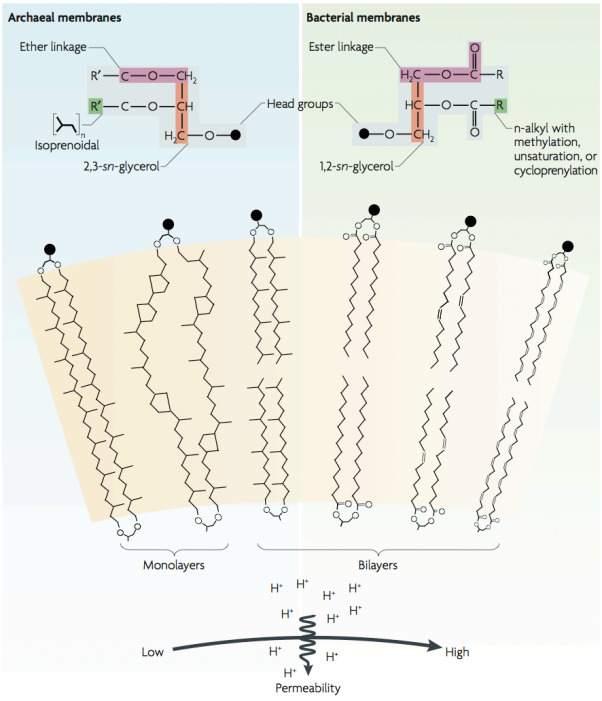


Figure 2 : Les membranes d'archées et de bactéries. Les membranes des eucaryotes sont similaires à celles des bactéries. D'après Valentine, 2007.

Les protéines informationnelles des archées ressemblent généralement plus aux protéines eucaryotes qu'aux protéines bactériennes. C'est typiquement le cas du complexe de réplication de l'ADN, qui ressemble fortement à celui des eucaryotes et qui n'a pas d'homologue chez les bactéries (Olsen and Woese, 1996). Les archées ont aussi, comme les eucaryotes, de la chromatine qui permet de compacter leur ADN (White and Bell, 2002) ; la régulation de ce système n'est pas encore totalement élucidée. Pour certaines archées, la présence d'histones, homologues à celles des eucaryotes, a été remarquée (Reeve et al., 1997).

En revanche, les protéines opérationnelles (par exemple impliquées dans le métabolisme) sont généralement de type bactérien. Les archées ressemblent également plus aux bactéries en termes d'organisation et de fonctionnement général de la cellule. Elles ne possèdent pas de noyau et les chromosomes sont circulaires, d'une taille comparable à ceux des bactéries ; les gènes sont organisés en opérons. Comme les bactéries, les archées ont des plasmides qu'elles peuvent transmettre horizontalement par conjugaison. Traduction et transcription sont couplées (French et al., 2007) et les séquences Shine-Dalgarno permettent, non-exclusivement, l'initiation de la traduction. Notons également que si l'ARN-polymerase est bien de type eucaryote (Werner, 2007), plusieurs co-facteurs de transcription sont de type bactérien (Bell and Jackson, 2001). Pour finir, les archées, comme les bactéries, possèdent des CRISPRs dans leur génome. Il s'agit de séquences courtes (≈ 20-30 bases) répétées en alternance avec des séquences variables, courtes également. Les parties variables sont souvent homologues à des séquences virales ou d'origine exogène. Chez les bactéries, il a été démontré que les CRISPRs permettent l'acquisition de la résistance aux virus qui infectent la souche considérée (Barrangou et al., 2007). Chez les archées, cela n'a pas été prouvé expérimentalement, mais on sait que les régions du génome correspondant aux CRISPRs sont transcrites (ex : Lillestol et al., 2006). Les CRISPRs étant plus fréquemment rencontrés chez les archées que chez les bactéries et occupant une fraction plus grande du génome (jusqu'à 1%), on peut penser que leur fonction est tout aussi importante chez les archées.

Le tableau dressé ci-dessus suggère que les archées utilisent des gènes de type eucaryote dans un contexte bactérien. Toutefois, elles présentent un degré de complexité supérieur à celui des bactéries, intermédiaire entre bactéries et eucaryotes. Par exemple, il existe chez certaines archées de multiples origines de réplication (Robinson and Bell, 2005), ce qui n'a encore jamais été observé chez une bactérie. La présence de chromatine concorde également avec une plus grande complexité.

Pour finir, certains mécanismes importants ne sont pas totalement décryptés chez les archées, par exemple ceux en lien avec la réponse aux stress, la maintenance du génome ou la division cellulaire. Les archées appartenant à deux phylums majeurs du domaine, Crenarchaeota et Euryarchaeota, présentent des différences très importantes sur certains de ces points (voir IV.1).

2.3 Diversité des archées

2.3.1 D'extrêmophiles et spécialisées, à mésophiles et ubiquitaires

Les archées présentent une grande diversité, en termes d'habitats, de morphologies et de systèmes métaboliques. La Figure 3 montre des exemples de morphologies de cellules d'archées.

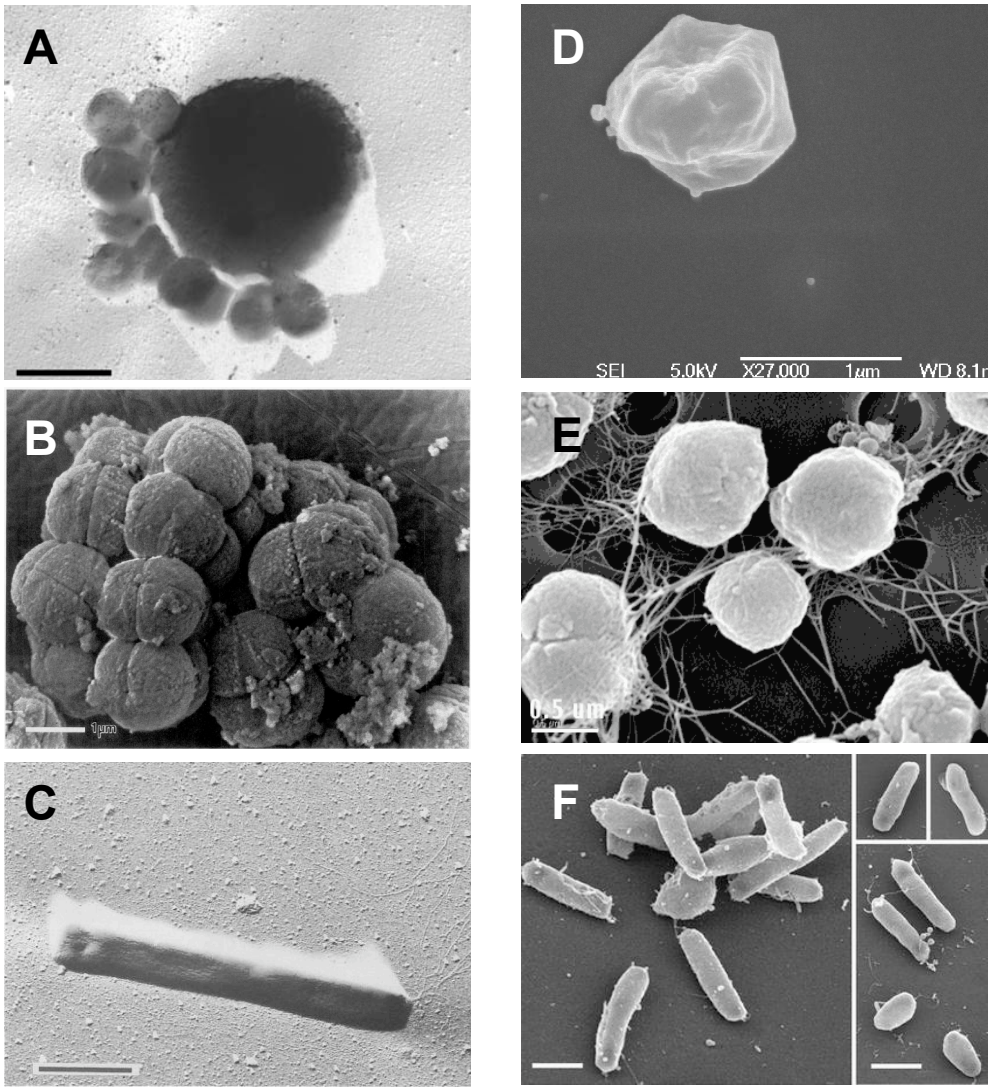


Figure 3 : Exemples de cellules d'archées aux morphologies variées. (A) cellule du genre *Ignicoccus*, et, attachées à la surface, cellules de Nanoarchaeum Equitans, barre = 1 µm (Huber et al., 2002). (B) cellules du genre *Halococcus*, barre = 1 µm (Denner et al., 1994). (C) cellules du genre *Pyrobaculum*, barre = 1 µm (Völkl et al., 1993). (D) cellule du genre *Acidianus*, barre = 1 µm. (E) cellules du genre *Methanococcus*, barre = 500 nm (http://em-lab.berkeley.edu/EML/images/sem.php?PHPSESSID=6e0f61ac47c40147a470df96afec633e) (F) cellules du genre *Halobacterium*, barres = 1 µm (Grüber et al., 2004).

Jusqu'à il y a une quinzaine d'années, les principaux phénotypes identifiés étaient limités aux archées halophiles extrêmes, aux méthanogènes, aux hyperthermophiles et acidothermophiles, correspondant donc à des environnements dits « extrêmes » ou à des niches spécialisées. Effectivement, les seuls organismes vivants connus capables de réaliser la méthanogenèse sont certains groupes d'archées. De plus, les archées prospèrent dans un grand nombre de milieux extrêmes, ce qui s'explique en grande partie par les propriétés de leurs membranes lipidiques (voir 2.2). Typiquement, les archées prospèrent dans les milieux chauds. Il s'agit des hyperthermophiles (température maximale de croissance ≥ 80°C, Stetter, 1996) et la Figure 4 illustre bien leur meilleure adaptation aux hautes températures par rapport aux bactéries. Le record de la plus haute température de croissance est détenu par une archée, *Pyrolobus fumarii* (113°C, Blöchl et al. 1997) ou peut-être même Strain 121 (121°C,

Cowen, 2004). Parmi les archées extrêmophiles, on distingue aussi les psychrophiles (température maximale de croissance $T \leq 15°C$), les halophiles (fortes concentrations de sels), les acidophiles (bas pH), les alkaliphiles (hauts pH), les piezophiles (fortes pressions), etc. (Rothschild and Mancinelli, 2001). Certaines archées combinent plusieurs de ces propriétés, telles les acidothermophiles. Notons, que la notion d'extrêmophile est principalement anthropomorphique. Tout ceci a contribué à donner aux archées une image quelque peu exotique, bien que des bactéries et des eucaryotes extrêmophiles existent également (Rothschild and Mancinelli, 2001).

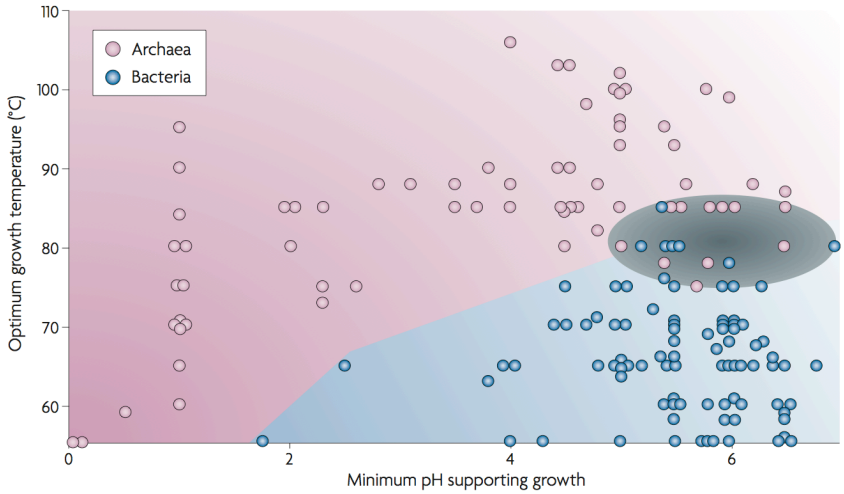


Figure 4 : Température et pH requis pour la croissance d'archées (cercles rose) et de bactéries (cercles bleus). Chaque cercle représente une espèce différente. Les archées méthanogènes n'ont pas été incluses. La zone grise correspond aux paramètres physico-chimiques convenant à la fois à des archées et des bactéries. D'après Valentine, 2007.

Mais les biotopes des archées ne sont pas restreints aux milieux extrêmes, comme cela a été révélé par des exploration plus récentes. Celles-ci, basées sur des méthodes culture-indépendantes, ont été initiées notamment par Norman Pace (ex : Stahl et al., 1984) et Edward DeLong (ex : Giovannoni et al., 1990). Les premières analyses étaient basées sur le séquençage des gènes de PSU-ARNr. Plus récemment encore, de nouvelles approches, dites de génomique environnementale, consistent à séquencer de longs fragments de l'ADN environnemental. Ceci permet de faire des hypothèses sur le métabolisme des microorganismes non-cultivables, en couplant l'analyse des marqueurs phylogénétiques à celle de gènes fonctionnels particuliers situés sur le même fragment d'ADN (revu par Schleper et al., 2005 ; Lopez-Garcia and Moreira, 2008). Toutes ces études ont révélé que les archées sont en réalité ubiquitaires, présentes en particulier dans les milieux mésophiles courants (sols, eaux douces, océans, tractus intestinal de l'homme et des ruminants, etc.) où elles représentent généralement autour de 10% des phylotypes d'ARNr (Robertson, 2005 ; Schleper et al., 2005). Elles sont une composante majeure du picoplancton marin, car elles pourraient représenter 30% des cellules procaryotes dans la zone mésopélagique de l'océan Pacifique (Karner et al., 2001), et de l'ordre de 20% dans l'océan Atlantique (Teira et al., 2006), et elles sont également présentes dans les sédiments marins (Teske and Sorensen,

2008). Or les océans représentent le plus grand biome terrestre. Enfin, fait important, les archées jouent probablement un grand rôle dans le cycle de l'azote (Valentine, 2002 ; Francis et al., 2007 ; Prosser and Nicol, 2008) et dans le cycle océanique du carbone (Herndl et al., 2005).

Les archées mésophiles et d'autres phylotypes récemment découverts suscitent donc un grand intérêt à l'heure actuelle. L'exploration de leur diversité et de leur distribution n'est pas encore achevée. Et un défi important est de comprendre plus précisément le rôle joué dans les écosystèmes par toutes ces archées, qui restent non-cultivables à de rares exceptions près (ex : Simon et al., 2005 ; Könneke et al., 2005 ; de la Torre et al., 2008). Enfin, des groupes entiers d'archées non-cultivables forment de nouveaux clades dont la classification fait débat (voir 2.3.2 ci-dessous).

2.3.2 Phylogénie et taxonomie des archées

Ci-après (Figure 5) est présenté un arbre phylogénétique des archées, schématique. Les archées cultivables sont principalement réparties dans les deux phylums établis par Carl Woese, Euryarchaeota et Crenarchaeota (Boone & Castenholz 2001). Les séquences des gènes de PSU-ARNr d'archées non-cultivables se sont accumulées, aussi représentent-elles maintenant une partie importante des phylotypes ; elles constituent parfois des clades entiers (voir Figure 5). De plus, certaines archées sont très particulières, comme *Nanoarchaeum equitans* (voir ci-dessous). Ceci a conduit, dans les quinze dernières années, à la proposition de nouveaux phylums dont la validité est en cours de débat. La diversité des archées n'est encore que partiellement connue (Schleper et al., 2005), et il est probable que des phylums supplémentaires émergeront à mesure que se poursuit l'exploration de cette diversité. En effet, deux phylums semblent peu pour un domaine du vivant présentant une très grande diversité biologique (Forterre et al., 2002) : à titre de comparaison, signalons que les bactéries comptent actuellement 12 phylums pour celles cultivées (Boone and Castenholz, 2001) et 23 à 35 si l'on inclut les non-cultivées (Hugenholtz et al., 1998; Hugenholtz 2002).

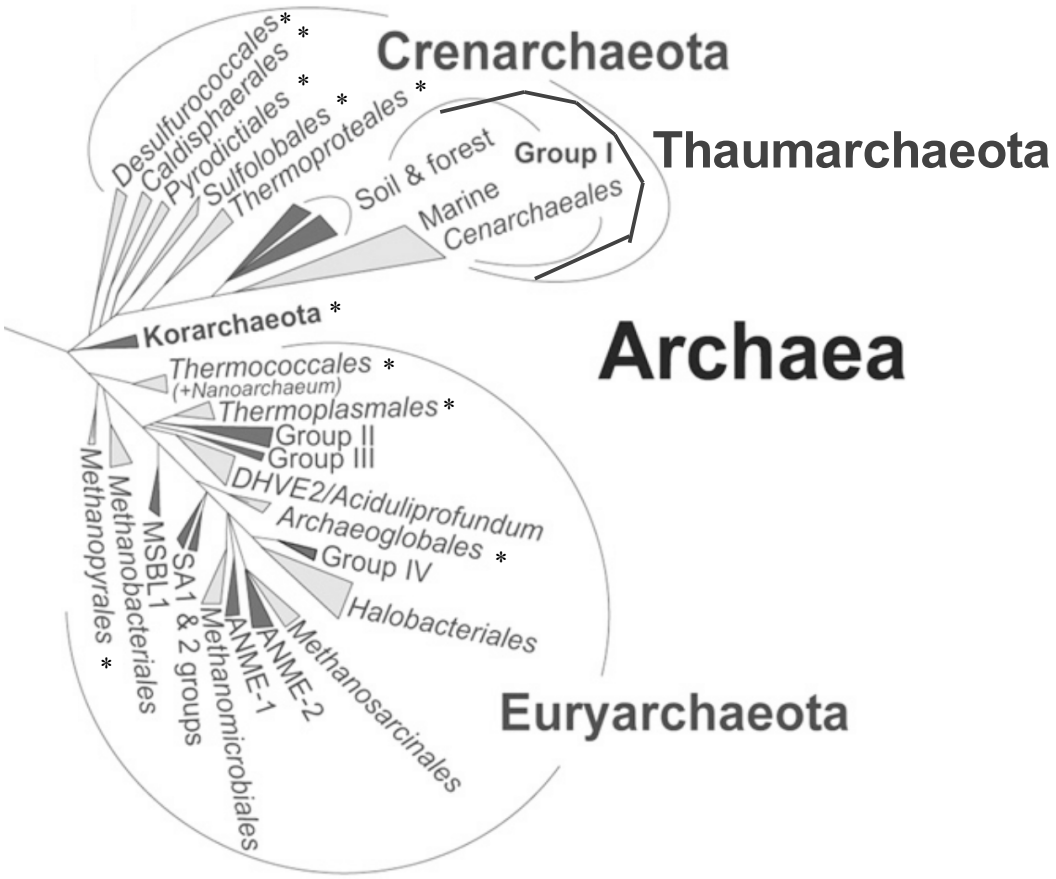


Figure 5 : Arbre phylogénétique schématique des archées cultivées et non cultivées, synthétisant les connaissances actuelles basées notamment sur l'utilisation des séquences des gènes de PSU-ARNr. Les triangles de couleur gris foncé correspondent aux branches contenant exclusivement des archées non-cultivées. Les triangles de couleur gris clair correspondent aux branches au sein desquelles se trouvent des archées cultivées. Les astérisques indiquent les branches contenant des organismes hyperthermophiles. D'après Lopez-Garcia and Moreira., 2008 ; Gribaldo and Brochier-Armanet, 2006 ; Brochier-Armanet et al., 2008.

Le phylum Crenarchaeota (voir Figure 5) est subdivisé en cinq ordres d'archées cultivables : Caldisphaerales, Desulfurococcales, Pyrodictiales, Sulfolobales, et Thermoproteales, qui appartiennent à une unique classe, Thermoprotei. Les crenarchées sont en effet relativement homogènes en termes d'habitat et de phénotype (voir 2.4.), étant toutes hyperthermophiles. Le phylum Euryarchaeota (voir Figure 5), plus diversifié, est subdivisé en huit classes (respectivement ordres) d'archées cultivables (Boone and Castenholz 2001) : Thermococci (Thermococcales), Methanopyri (Methanopyrales), Methanococci (Methanococcales, non-représenté dans la Figure 5), Methanobacteria (Methanobacteriales), Thermoplasmata (Thermoplasmatales), Archaeoglobi (Archaeaoglobales), Halobacteria (Halobacteriales) et Methanomicrobia . La classe Methanomicrobia inclut les ordres Methanomicrobiales et Methanosarcinales.

Les trois phylums proposés et en discussion sont Korarchaeota (Barns et al., 1996), Nanoarchaeota (Huber et al., 2002) et Thaumarchaeota (Brochier-Armanet et al., 2008).

Le phylum candidat Korarchaeota correspond à un groupe d'archées hyperthermophiles non-cultivables, la première ayant été identifiée dans les sources chaudes du parc Yellowstone. Ce phylum n'est pas encore établi de façon robuste, le clade n'étant pas

toujours clairement séparé des clades du phylum Crenarchaeota (Forterre et al., 2002). Toutefois, la séquence complète d'une archée représentant le groupe été obtenue récemment (Candidatus *Korarchaeum cryptofilum*), ce qui a permis d'en affiner la phylogénie (Elkins et al., 2008) : le phylum Korarchaeota correspondrait à une lignée profonde des archées, avec une affinité apparente pour le phylum Crenarchaeota.

Le phylum candidat Nanoarchaeota correspond à l'espèce *Nanoarchaeum equitans.* Cette archée hyperthermophile a été isolée de cheminées hydrothermales. Elle vit comme symbiote obligatoire à la surface d'une archée du genre *Ignicoccus*. Les cellules sont de très petite taille (400 nm de diamètre), et possèdent le plus petit génome cellulaire connu (0,5 Mb). Cependant, la position de *Nanoarchaeum equitans* dans l'arbre phylogénétique pourrait être un artefact lié à une évolution plus rapide, ce qui est typique des génomes ayant subi une réduction. Selon une analyse basée sur d'autres marqueurs moleculaires que les PSU-ARNr, les nanoarchées correspondraient en fait à une lignée du phylum Euryarchaeota et seraient apparentées aux archées de l'ordre Thermococcales (Brochier et al., 2005, voir Figure 5).

Enfin, le phylum candidat Thaumarchaeota (voir Figure 5) est composé de souches mésophiles initialement classées au sein des Crenarchaeota sur la base de l'analyse des PSU-ARNr. Néanmoins, une étude récente suggère que ces crenarchées mésophiles ne doivent pas être classées parmi les Crenarchaeota. En effet, leur position non-basale visible dans la Figure 5 serait un artefact. Les thaumarchées constitueraient plutôt un nouveau phylum basal des archées (Brochier-Armanet et al., 2008). Cette proposition résulte d'une étude du génome de *Cenarchaeum symbiosum*, la première « crenarchée mésophile » séquencée (voir Figure 5, Cenarchaeales).

2.4 Phénotype et habitat des principaux groupes d'archées

Cette question est revue en détail par Chaban et al., 2006, et plus synthétiquement par Schleper et al., 2005 et Robertson et al., 2005. Dans le texte qui suit, les archées sont groupées par phénotype et les archées non-cultivées ou cultivables depuis peu sont décrites à la fin de cette partie, dans un paragraphe dédié.

2.4.1 Méthanogènes

Les archées méthanogènes sont les seuls organismes capables de produire du méthane (CH_4). Elles sont polyphylétiques, correspondant à 5 ordres différents : Methanobacteriales, Methanomicrobiales, Methanococcales, Methanopyrales, Methanosarcinales. Elles présentent une très grande variété d'habitats, en ce qui concerne la température (0°C à 100°C), le pH (acide modéré à supérieur à 9) et les concentrations en sels (milieux doux, marin, hypersalé). Certains méthanogènes sont donc extrêmophiles. Le point commun est l'anaérobiose extrême des habitats, indispensable aux méthanogènes. Comme principaux types d'habitats, on peut citer les marécages et marais, rizières, sédiments d'eaux douces, digesteurs ; le rumen et le

tractus digestif de la plupart des animaux (dont l'homme et les arthropodes) ; et enfin les sources géothermiques et les cheminées hydrothermales.

Les substrats les plus courants pour la méthanogenèse sont CO_2/H_2, le formate et l'acétate, mais une variété d'autres substrats sont possibles (par exemple méthanol, éthanol, etc.). Les archées méthanogènes de l'ordre Methanosarcinales peuvent utiliser la plus grande variété de substrats (Liu and Whitman, 2008). Ce sont les seules à réaliser la méthanogenèse acétoclaste. Parmi elles, les archées du genre *Methanosaeta* utilisent exclusivement l'acétate, alors que celles du genre *Methanosarcina* peuvent utiliser de nombreux substrats (sept substrats différents listés dans Chaban et al., 2006).

Dans le contexte actuel de réchauffement climatique, les archées méthanogènes suscitent de l'intérêt car le méthane est un puissant gaz à effet de serre, mais peut également être une source d'énergie renouvelable (ex : biogaz).

Curieusement, alors que certaines archées méthanogènes vivent en association avec l'homme dans l'intestin, aucune archée pathogène n'est connue à ce jour.

2.4.2 Extrêmes halophiles

Les extrêmes halophiles sont des microogranismes qui croissent de façon optimale à des fortes concentrations de NaCl (≥ 3,4-5,1 mol/L soit 20%-30%). Les archées extrêmes halophiles constituent un groupe monophylétique d'euryarchées, l'ordre des Halobacteriales (si l'on excepte les quelques Methanosarcinales halophiles). Les habitats des membres de l'ordre Halobacteriales sont par exemple les lacs naturels salés (ex : Great Salt Lake), la Mer Morte, les tables saunantes (surfaces aménagées pour la cristallisation du sel), certains lacs sodés (ex : Lac Magadi au Kenya), les sols salins et les aliments salés. Certaines archées halophiles contiennent des pigments jaunes à rouges qui sont responsables de la couleur particulière de leur environnement (Figure 6, ex : Lac Magadi, saunantes, etc.). Des études récentes suggèrent que les archées halophiles peuvent croître à des concentrations de sels plus basses que ce que l'on pensait (≈ 7%), aussi leur habitat ne serait-il pas limité aux environnements hypersalés.

Figure 6 : Lac salé en Utah, aux Etats-Unis, coloré par la présence d'archées halophiles.

La plupart des Halobacteriales sont des chimioorganohétérotrophes aérobies. Elles utilisent alors les composés organiques produits par les producteurs primaires, tels l'algue *Dunaliella* (eucaryote, Chromalveolata Cryptophyceae, d'après Adl et al., 2005) qui produit du glycérol.

2.4.3 Thermophiles, hyperthermophiles et acidophiles

Les archées thermophiles, hyperthermophiles et acidophiles correspondent à un grand nombre d'ordres différents (Figure 5). Dans le phylum Euryarchaeota, il s'agit des ordres Thermococcales, Thermoplasmatales et Archaeoglobales. Chez les Crenarchaeota, il s'agit des ordres Thermoproteales, Sulfolobales, Desulfurococcales, Pyrodictiales et Caldisphaerales. Des membres des Methanococcales ainsi que *Methanopyrus kandleri*, unique représentant des Methanopyrales, sont également hyperthermophiles (mais traités dans le paragraphe 2.4.1 concernant les méthanogènes).

Les archées hyperthermophiles sont très hétérogènes en termes de métabolisme, types trophiques (etc.) ce qui est dû à la grande variété et variabilité des habitats. Ceux-ci sont par exemple les champs de solfatares et les sources chaudes autour des zones de volcanisme actif (ex : Parc de Yellowstone aux Etats-Unis, Kamchatka, Islande, etc.), les systèmes hydrothermaux marins profonds et peu profonds, le plancher océanique, et certaines installations industrielles, telles les drainages miniers acides ou les réservoirs de pétrole. Même dans une zone géothermique donnée, les paramètres physico-chimiques peuvent grandement varier entre des sites peu distants (ex : pH, éléments traces, Reysenbach and Shock, 2002), et donc une grande diversité microbienne y est attendue ou tout du moins la présence d'organismes pouvant résister à de grandes fluctuations. Enfin, il est supposé que les hyperthermophiles forment des réseaux trophiques complexes, avec des producteurs primaires et des consommateurs primaires, ce qui contribuerait aussi à la variété des types trophiques. Des exemples d'archées trouvées dans ces types d'habitats sont présentés dans le Tableau 2, et des exemples de réactions qu'elles utilisent pour l'obtention d'énergie sont montrés dans le Tableau 3.

Malgré cette diversité, on peut dire que les hyperthermophiles et beaucoup de thermophiles sont chimiotrophes, et non phototrophes, en lien avec la faible luminosité (par exemple dans les fonds marins). La majorité d'entre eux sont anaérobies stricts ou facultatifs, en lien avec les faibles concentrations en oxygène dans ces milieux. Enfin, les crénarchées cultivables sont assez homogènes en ce qui concerne le phénotype et l'habitat (hyperthermophiles métabolisant le soufre, généralement anaérobies facultatives ou strictes).

Habitat	Representative genera
Solfatara and hot springs	*Acidianus*, *Acidilobus*, *Aeropyrum*, *Archaeoglobus*, *Caldisphaera*, *Caldivirga*, *Desulfococcus*, *Ferroplasma*, *Metallosphaera*, *Picrophilus*, *Pyrobaculum*, *Sulfolobus*, *Thermocladium*, *Thermofilum*, *Thermoplasma*, *Thermoproteus*, *Vulcanisaeta*
Shallow hydrothermal systems	*Archaeoglobus*, *Ferroglobus*, *Palaeococcus*, *Pyrobaculum*, *Pyrococcus*, *Pyrodictium*, *Staphylothermus*, *Stetteria*, *Thermococcus*, *Thermodiscus*
Deep hydrothermal systems	*Archaeoglobus*, *Ferroglobus*, *Geoglobus*, *Ignicoccus*, *Nanoarchaeum*, *Pyrococcus*, *Pyrolobus*, *Thermococcus*
Subseafloor	*Pyrococcus*, *Thermococcus*
Bioreactors	*Ferroplasma*, *Metallosphaera*, *Sulfolobus*
Acid mine drainage	*Ferroplasma*, *Thermoplasma*
Acidic waters	*Desulfobacter*, *Ferroplasma*, *Thermoplasma*
Oil reservoirs	*Archaeoglobus*, *Desulfurococcus*, *Pyrococcus*, *Thermococcus*
Self-heating spoil heaps	*Metallosphaera*, *Sulfolobus*, *Thermoplasma*

Tableau 2 : Les habitats des archées hyperthermophiles, thermophiles et acidophiles. D'après Chaban et al., 2006.

Reaction	Representative genera
$2H_2 + O_2 \rightarrow 2H_2O$	*Acidianus*, *Metallosphaera*, *Pyrobaculum*, *Pyrolobus*, *Sulfolobus*
$2S^0 + 3O_2 + H_2O \rightarrow 2H_2SO_4$	*Acidianus*, *Metallosphaera*, *Sulfolobus*
$2FeS_2 + 7O_2 + 2H_2O \rightarrow 2FeSO_4 + 2H_2SO_4$	*Acidianus*, *Metallosphaera*, *Sulfolobus*
$H_2 + HNO_3 \rightarrow HNO_2 + H_2O$	*Ferroglobus*, *Pyrobaculum*, *Pyrolobus*
$4H_2 + HNO_3 \rightarrow NH_4OH + 2H_2O$	*Pyrolobus*
$2FeCO_3 + HNO_3 + 5H_2O \rightarrow 2Fe(OH)_3 + HNO_2 + 2H_2CO_3$	*Ferroglobus*
$H_2 + 6FeO(OH) \rightarrow 2Fe_3O_4 + 4H_2O$	*Pyrobaculum*
$4H_2 + H_2SO_4 \rightarrow H_2S + 4H_2O$	*Archaeoglobus*
H_2 or $2[H] + S^0 \rightarrow H_2S$	*Acidianus*, *Desulfurococcus*, *Ignicoccus*, *Pyrodictium*, *Pyrobaculum*, *Stetteria*, *Stygiolobus*, *Sulfurisphaera*, *Thermodiscus*, *Thermofilum*, *Thermoproteus*
$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O$	*Methanocaldococcus*, *Methanopyrus*, *Methanothermus*
Acetate + $8Fe(III) + 4H_2O \rightarrow 2HCO_3^- + 8Fe(II) + 9H^+$	*Ferroglobus*, *Geoglobus*
$H_2S + HNO_3 \rightarrow HNO_2 + S^0 + H_2O$	*Ferroglobus*
Arsenate + $H_2 \rightarrow$ arsenite + H_2O	*Pyrobaculum*
Benzoate, phenol, benzaldehyde, or other aromatic compounds + $Fe(III) \rightarrow HCO_3 + Fe(II) + H_2O + H^+$	*Ferroglobus*
Peptides → isovalerate, isobutyrate, butanol, CO_2, H_2, etc.	*Desulfurococcus*, *Hyperthermus*, *Pyrococcus*, *Pyrodictium*, *Staphylothermus*, *Sulfophobococcus*, *Thermococcus*, *Thermoproteus*, *Thermosphaera*
Cellobiose, maltose, or pyruvate → acetate + alanine + H_2 + CO_2	*Pyrococcus*

Tableau 3 : Les réactions productrices d'énergie chez les archées hyperthermophiles. D'après Chaban et al., 2006.

2.4.4 Archées non-cultivables ou cultivables depuis peu

Milieu marin et plancher océanique

Les archées marines sont réparties dans quatre clades (groupes I à IV) et ont été isolées, à quelques exceptions près (Barns et al., 1996), d'environnements mésophiles. Le groupe I constitue le phylum Thaumarchaeota (voir 2.3.2), et les autres groupes sont associés au phylum Euryarchaeota. Le groupe IV (Lopez-Garcia et al., 2001) est proche des Halobacteriales, ce qui n'est pas le cas des groupes II et III (voir Figure 5).

Les thaumarchées (Groupe I) sont très répandues, pourtant leur métabolisme et leur fonction ne sont pas élucidés. En 2005, la première archée de ce type, *Nitrosopumilus maritimus*, a pu être cultivée (Könneke et al., 2005) ; c'est une archée chimiolithoautotrophe qui réalise l'oxydation de l'ammoniac en nitrate en conditions aérobies. Cela serait également le cas de *Cenarchaeum symbiosum*, une autre thaumarchée, qui peut être maintenue en laboratoire en association avec son hôte (une éponge), et dont le génome a pu être séquencé par fragments (Hallam et al., 2006 ; Girguis et al., 2003 ; Preston et al., 1996). Ceci suggère que des archées nitrifiantes pourraient contribuer aux cycles marins du carbone et de l'azote. La présence ubiquitaire d'archées pouvant oxyder l'ammoniac a pu être confirmée (Francis et al., 2005).

Un groupe d'archées non-cultivable qui suscite actuellement beaucoup d'intérêt est celui des méthanotrophes. De grandes quantités de méthane sont présentes dans le plancher océanique, sous forme d'hydrates de gaz solides. Le méthane qui est libéré n'atteint pas les zones aérobies. Il est vraisemblablement oxydé en CO_2 par une association syntrophique (Figure 7) entre des archées oxydant le méthane en conditions anaérobies et des bactéries sulfato-réductrices (ex : *Desulfosarcina*). Ces archées méthanotrophes correspondent aux groupes ANME-1, ANME-2 et ANME-3, affiliés aux Methanosarcinales et Methanomicrobiales (ex : Figure 5), au sein du phylum Euryarchaeota. Une hypothèse actuelle, étayée par plusieurs données (ex : Hallam et al., 2004 ; Hallam et al., 2003), est que le processus d'oxydation du méthane repose sur une inversion de la voie métabolique de la méthanogenèse (Boetius et al., 2000). Ces découvertes sont d'autant plus intéressantes que les seuls organismes méthanotrophes connus jusqu'alors étaient des bactéries aérobies.

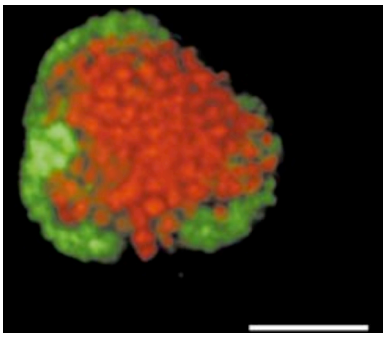


Figure 7 : Identification *in situ* de microorganismes. Agrégat d'archées réalisant l'oxydation anaérobie du méthane (rouge), et de bactéries sulfato-réductrices (vert). Il s'agit d'une association syntrophique entre archées et bactéries. D'après Boetius et al., 2000.

Autres milieux

Les archées semblent également nombreuses dans les écosystèmes d'eaux douces (méthanogènes, thaumarchées ou euryarchées des groupes II et III). Enfin, des archées non-cultivables ont été détectées dans tous les autres types de milieux aquatiques : milieux hydrothermaux terrestres (ex : phylum candidat Korarchaeota, voir 2.3.2) et marins, estuaires, bassins hypersalés anoxiques (comme dans la Mer Méditerranée, van der Wielen, 2005), lacs sodés. En 2005, la séquence du gène de PSU-ARNr d'une archée de phénotype hyperthermophile (origine, Pisciarelli en Italie, Kvist et al., 2005) s'est classée, du point de vue phylogénétique, au sein d'un cluster de thaumarchées. Ceci indique que les thaumarchées constituent un phylum séparé qui n'inclut pas seulement des archées mésophiles.

On trouve également fréquemment des archées dans les sols, où elles constituent cependant une fraction minoritaire des microorganismes. Une grande partie de ces archées

appartient au groupe I.1B (voir Figure 5) et pourraient être nitrifiantes (oxydation de l'ammoniac), comme une grande fraction des crenarchées en milieu océanique (voir ci-dessus). Les archées non-cultivables sont dans une variété d'autres milieux, comme la rhizosphère, les rumens, les tractus digestifs d'insectes, les digesteurs ou systèmes de traitement industriels anaérobies.

En conclusion, les archées sont présentes dans des milieux très divers et présentent de ce fait des métabolismes variés. De grands progrès sont à attendre de la culture des archées mésophiles ou appartenant à de nouveaux phylotypes, pour la meilleure compréhension de certains écosystèmes et du domaine Archaea. Cette étape de mise en culture a déjà été atteinte pour un nombre restreint d'entre elles (voir ci-dessus).

2.5 Questions évolutives en lien avec le domaine Archaea

2.5.1 La classification en trois domaines, une révolution

L'idée de classer les êtres vivants et les objets naturels est ancienne : des classifications existaient dès la période antique, telles celle d'Aristote (385-322 av. J.C.). Celui-ci décrit un continuum entre minéraux et animaux en passant par les végétaux. Puis il a existé différents types de classifications, selon les périodes, mais toutes reprenaient d'une façon ou d'une autre la dichotomie plante/animal. Au 16ème siècle, des classifications alphabétiques sont apparues, au départ pour les plantes (ex : Leonhart Fuchs, 1501-1566). Aux 17ème et 18ème siècles sont apparues des classifications artificielles, basées notamment sur un principe dichotomique appliqué à différents critères, conduisant à une structure hiérarchique. La classification proposée par Carl von Linné (1707-1778), qui introduisit la nomenclature binomiale, appartient à ce courant. Fin 18ème et au 19ème siècle émergea l'idée de classification naturelle : il s'agissait d'établir une classification qui reflète l'ordre naturel de la création divine. Aussi, la classification des micro-organismes, découverts au 17ème siècle, commença à poser problème : quelle place devaient-ils avoir dans la dichotomie existante plantes/animaux ? Cela aboutit à une crise de la classification au 19ème siècle. Haeckel, contemporain de Darwin, proposa en 1866 un arbre universel (Figure 8A) retraçant l'évolution du monde vivant. L'arbre comprenait une branche séparée pour les organismes unicellulaires (Protista) et avait à sa base les Moneres (futurs procaryotes). Mais il ne rencontra pas un grand succès et la dichotomie plante/animal se maintint. En 1937, Chatton introduisit et définit les termes procaryotes et eucaryotes. Le préfixe « pro » signifie « avant », et il était donc postulé qu'au cours de l'évolution, la cellule procaryote, plus simple, était apparue avant la cellule eucaryote. Il s'agit d'une vision gradiste de l'évolution. Ces travaux ne reçurent pas véritablement d'écho avant les années 60. En 1969, Whittaker proposa la division du monde vivant en 5 règnes, avec 3 niveaux d'organisation : à la base, les organismes unicellulaires sans noyau, dits Monera, puis les unicellulaires avec noyau, dits Protista, et enfin les organismes multicellulaires subdivisés en Plantae, Fungi, Animalia (voir

Figure 8B). La vision gradiste était toujours présente. La phylogénie moléculaire se développa dans la seconde moitié du 20ème siècle, aboutissant finalement à la classification du vivant en 3 domaines, Archaea, Bacteria, Eucarya.

Cette dernière classification du vivant est exceptionnelle dans la mesure où elle n'est ni dichotomique, ni gradiste. Les micro-organismes sont aussi évolués que les organismes multicellulaires et ne se trouvent donc pas à la base de la classification. C'est pourquoi Carl Woese et d'autres biologistes proposent d'abandonner le terme de procaryote qui sous-entend une antériorité non-fondée. Tous les êtres vivants partagent en revanche un unique ancêtre commun dont la nature exacte reste à déterminer. Cette classification relativise aussi beaucoup la place des eucaryotes (en particulier des plantes et des animaux) qui ne constituent qu'une fraction limitée de la diversité : deux des trois domaines sont constitués d'organismes quasi-exclusivement unicellulaires. Beaucoup de microbiologistes n'hésitent pas à qualifier la découverte des archées de révolution, et il faut signaler que la classification en trois domaines ne fut pas acceptée facilement.

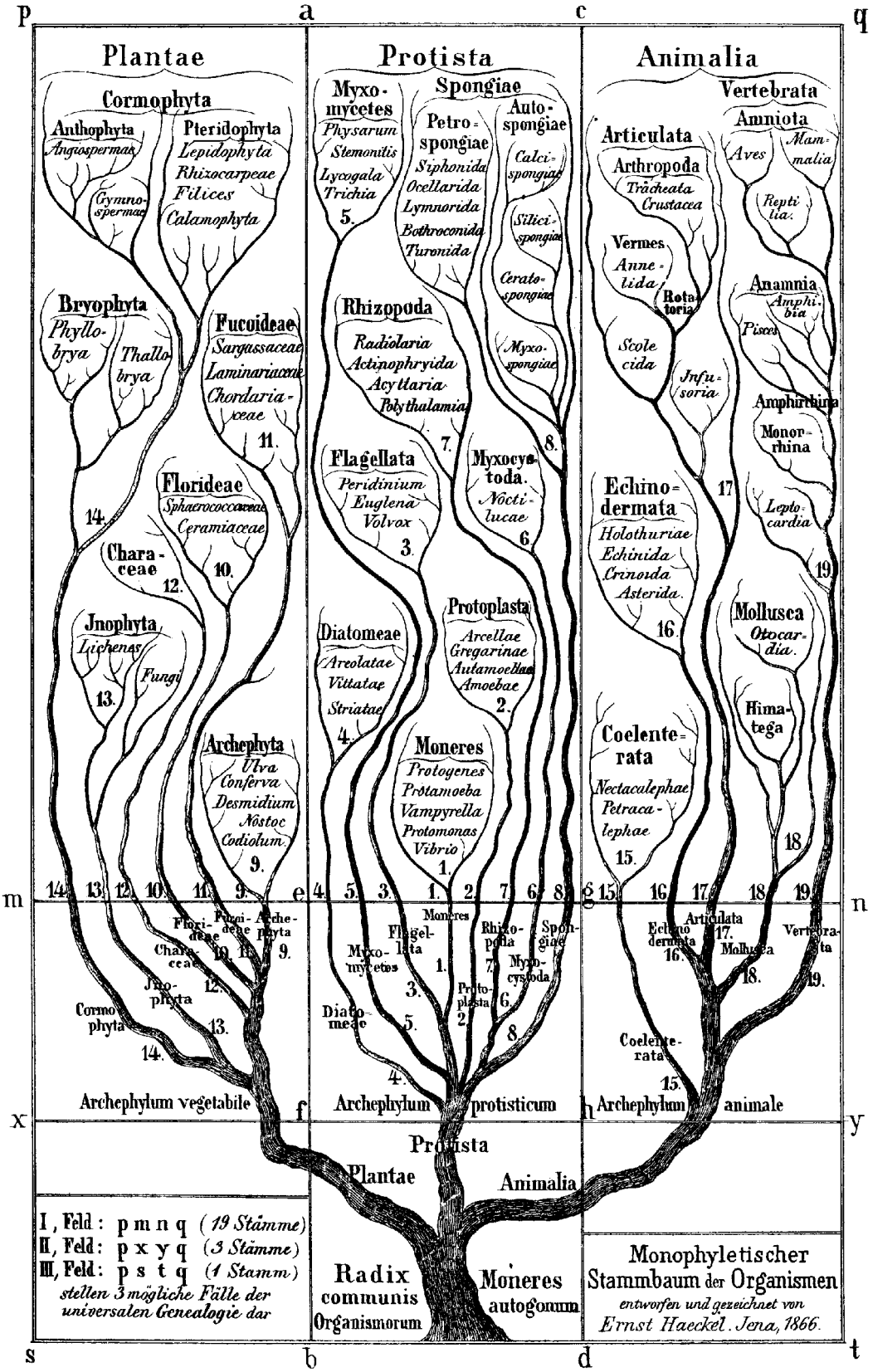


Figure 8 : (A) l'Arbre universel d'Haeckel (1866)

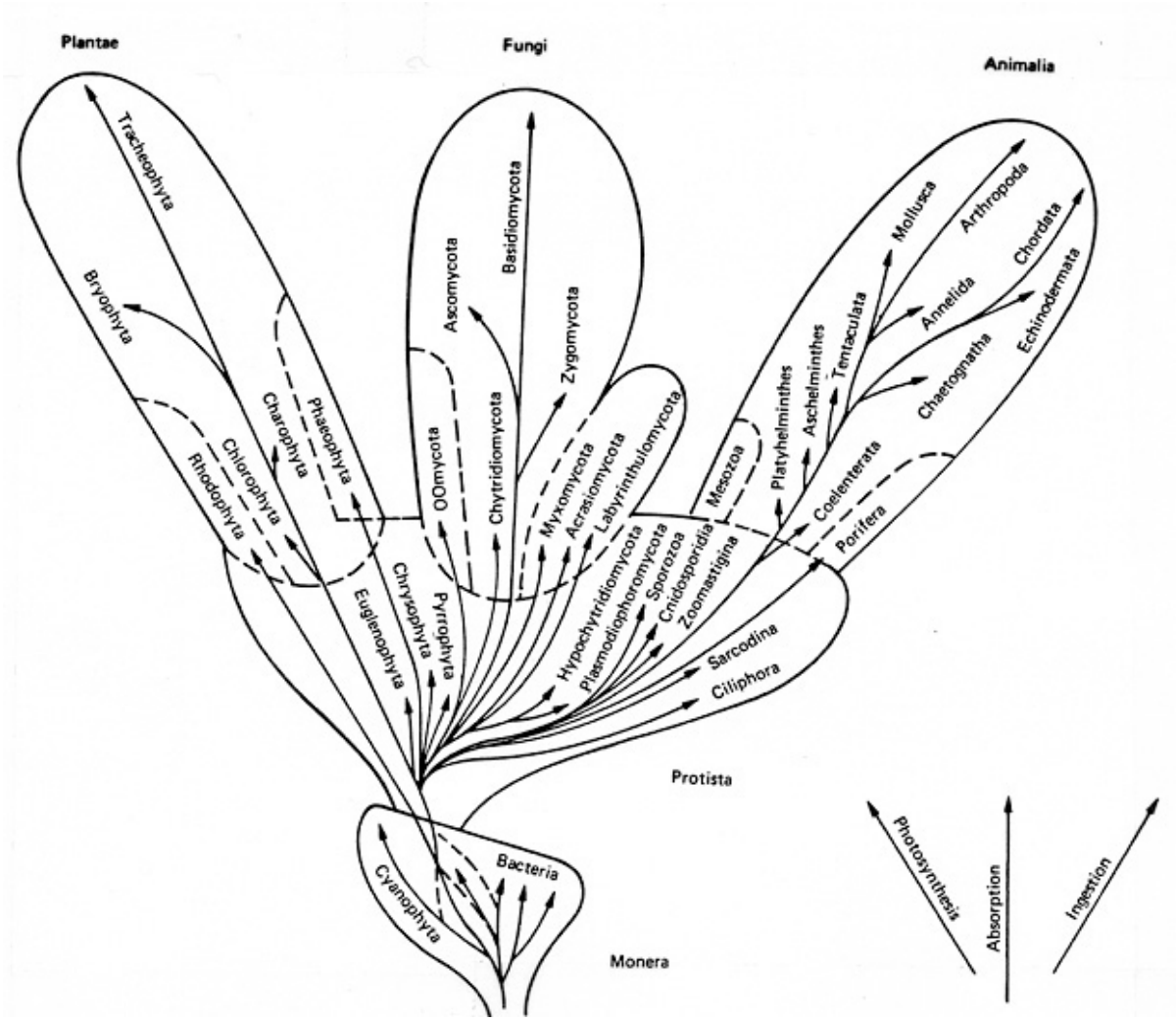


Figure 8 : (B) les 5 règnes de Whittaker (1969)

2.5.2 Les relations évolutives entre les trois domaines du vivant

Si la phylogénie moléculaire et les autres données disponibles soutiennent systématiquement l'existence de trois domaines séparés, les liens évolutifs entre les domaines sont très difficiles à établir et la question reste ouverte. Cette discussion est par exemple abordée dans Gribaldo and Brochier-Armanet, 2006.

Plusieurs arbres phylogénétiques racinés (arbres universels du vivant) publiés à la fin des années 80 et dans les années 90 placent systématiquement la racine chez les bactéries. Ceci signifie que les bactéries dériveraient directement de LUCA, tandis que les archées et les eucaryotes partageraient un ancêtre commun plus récent (schéma Figure 9A). En réalité, bien que toutes ces phylogénies donnent des résultats similaires, elles ne sont pas fiables à cause d'un phénomène de saturation moléculaire (Philippe and Forterre, 1999 ; Gribaldo and Philippe, 2002) : des substitutions multiples de bases sont survenues dans les séquences au cours du temps, induisant une perte de signal. Il n'y a donc pas de raison particulière de privilégier cette hypothèse par rapport à d'autres scénarios plausibles. D'autres possibilités sont une racine eucaryote ou archéenne. Il a également été proposé une origine chimérique

des eucaryotes par fusion d'une archée méthanogène et d'une gamma-protéo bactérie (Lopez-Garcia and Moreira, 2006, Figure 9B). Enfin, Carl Woese a suggéré qu'avant l'émergence des domaines existait un stade où les entités cellulaires étaient bien plus simples et échangeaient constamment du matériel génétique par transfert horizontal (hypothèse du progénote, Woese, 1998). L'émergence de domaines distincts n'aurait été possible qu'avec la complexification des cellules, les rendant réfractaires aux transferts horizontaux et provoquant un ralentissement de la vitesse d'évolution. L'ancêtre commun ne correspondrait donc pas à une lignée cellulaire telle qu'on la conçoit actuellement, et la phylogénie n'aurait pas de sens à la base de l'arbre.

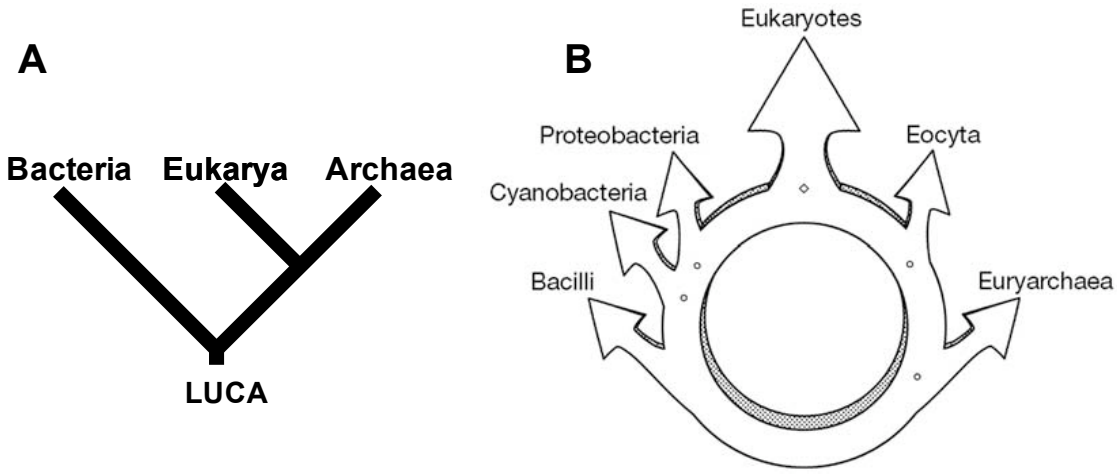


Figure 9: Exemples schématiques. (A) Arbre universel du vivant avec la racine chez les bactéries. (B) Hypothèse chimérique, d'après Rivera and Lake, 2004 (« Ring of Life »).

2.5.3 L'évolution des archées

Il est très difficile, au sein de chaque domaine, d'établir une histoire évolutive robuste. La phylogénie fonctionne relativement bien pour les archées, mais des points importants ne sont pas encore élucidés. L'évolution des archées est abordée en détail dans Gribaldo and Brochier-Armanet, 2006.

Un premier point concerne l'origine et l'évolution de la méthanogenèse. Dans Gribaldo and Brochier-Armanet, 2006, il est suggéré que la position basale de *Methanopyrus kandleri* (euryarchée méthanogène et hyperthermophile) pourrait être un artéfact. *Methanopyrus kandleri* serait en réalité proche des Methanobacteriales et Methanococcales. Il est également suggéré que la méthanogenèse serait apparue une unique fois, chez les euryarchées, juste après la divergence des Thermococcales. Selon ce scénario, le dernier ancêtre commun des archées n'était pas méthanogène, et d'autre part, la méthanogenèse aurait été perdue au moins 3 fois par la suite, chez les Thermoplasmatales, Archaeoglobales et Halobacteriales. La position basale de *Methanopyrus kandleri* pourrait être causée par le haut GC% de sa séquence d'ARN ribosomique (retrouvé chez tous les hyperthermophiles) et par une évolution rapide des sous-unités de son ARN-polymérase. D'autres scénarios évolutifs peuvent cependant être envisagés pour la méthanogénèse.

Le second point très important concerne le possible caractère hyperthermophile du dernier ancêtre commun des archées. Les branches correspondant aux hyperthermophiles occupent des positions basales dans les arbres phylogénétiques, que ce soit pour le phylum Crenarchaeota ou Euryarchaeota. D'autres données, comme la distribution de la réverse gyrase chez les archées, supportent également l'hypothèse d'un ancêtre commun hyperthermophile. Cependant, à cause du haut GC% des séquences d'ARNr chez les hyperthermophiles, il se pourrait que certaines positions homologues occupées par des G ou des C correspondent à une convergence et non à l'existence d'un ancêtre commun, introduisant un biais. En outre, la découverte du phylum Thaumarchaeota pourrait également remettre en question l'hyperthermophilie du dernier ancêtre des archées (bien que des thaumarchées hyperthermophiles existent).

L'obtention de nouvelles séquences de thaumarchées, de korarchées ou d'archées du groupe AAG (hyperthermophiles) pourrait permettre de trancher sur l'hyperthermophilie du dernier ancêtre des archées. S'il s'avérait que les archées sont initialement adaptées aux milieux chauds, ce sont alors les environnements mésophiles qui représenteraient, pour elles, une adaptation « extrême » (Gribaldo and Brochier-Armanet, 2006).

En conclusion, les archées posent des questions fondamentales sur l'évolution du vivant et ont une grande importance sur le plan écologique. Découvertes récemment, elles sont encore relativement moins bien caractérisées et comprises que les bactéries et les eucaryotes. Elles sont ubiquitaires mais c'est plus souvent dans les milieux dits extrêmes qu'elles sont dominantes, en particulier dans les environnements chauds ; de plus, les archées mésophiles restent difficilement cultivables, car elles sont anaérobies strictes ou de métabolisme inconnu. Du côté des crenarchées, les organismes modèles sont des hyperthermophiles, et c'est sur des crenarchées hyperthermophiles de l'ordre des Sulfolobales que la partie expérimentale de la thèse a été réalisée, ainsi que sur leurs virus.

3
Virosphère et virus d'archées

Mon travail de thèse, bien que transversal, est centré autour des virus d'archées (3.3), et plus particulièrement des virus de crenarchées hyperthermophiles. Ceux-ci constituent un groupe tout-à-fait particulier, aussi je les décrirai plus en détail (3.3.2), après quelques rappels généraux sur la découverte des virus (3.1) et sur leur écologie et diversité (3.2). Pour finir, je présenterai des questions évolutives (3.4) soulevées par la distribution des morphotypes et familles de virus à travers les trois domaines du vivant.

3.1 La découverte des virus et quelques grandes phases de recherche

3.1.1 Découverte des virus infectant des hôtes eucaryotes

La découverte des virus date de la fin du 19ème siècle. Le premier virus découvert était un virus de plante à ARN, le virus de la mosaïque du tabac (TMV). En 1892, Dmitrii Ivanowski, un scientifique russe, démontra que la sève de feuilles de tabac souffrant de la mosaïque était encore infectieuse après une étape de filtration permettant d'éliminer les bactéries (Takanami, 2006). Il pensa que la maladie pouvait donc être causée par une toxine bactérienne. En 1898, le scientifique néerlandais Martinus Beijenrinck prouva que la sève contenait un agent pathogène se multipliant dans la plante. Il nomma ce phénomène « contagium vivum fluidum » et employa le terme « virus », jetant les toutes premières bases de la définition moderne de virus, développée plus avant par André Lwoff (1957) (Norrby, 2008). TMV devint un virus modèle et fut caractérisé très en détail (Scholthof, 2004). Par ailleurs, après la découverte de 1898, des virus furent rapidement identifiés chez les animaux et les êtres humains. Les pans de la virologie qui se développèrent à partir de ces virus (plantes, animaux, hommes) concernèrent beaucoup l'étude des pathologies qu'ils causent.

3.1.2 Découverte des virus infectant des hôtes bactériens

Côté bactéries, les virus furent découverts au début du 20ème siècle, très certainement indépendamment par Frederick Twort (1915), et par Félix d'Hérelle (1910 ou 1917) (Duckworth, 1976). Félix d'Hérelle poursuivit et approfondit ses recherches sur les virus de bactéries, et développa de nombreuses méthodes, fondant ainsi un nouveau champ de la biologie. Il nomma ces virus « bactériophages », ce qui signifie « mangeurs de bactéries », avant de réaliser qu'il s'agissait d'agents de même nature que les virus de plantes et d'animaux. Dans les années qui suivirent la découverte, l'attention fut portée sur la nature de ces agents infectieux : s'agissait-il d'agents purement chimiques, comme des enzymes, ou étaient-ils vivants ? En effet, ce n'est pas avant 1940 que purent être obtenues les premières images de phages, grâce à l'invention et au développement de la microscopie électronique (Pennazio, 2006). L'idée d'utiliser les phages pour combattre les infections d'origine

bactérienne émergea toutefois immédiatement après leur découverte et se développa : on parle de thérapie phagique ou phagothérapie. Elle fut appliquée par d'Hérelle lui-même en 1919 (Sulakvelidze et al., 2001) et son laboratoire réalisa également des productions commerciales de phages. Toute cette période correspond au « premier âge des phages » (Mann, 2005). Avec le développement des antibiotiques, la phagothérapie fut abandonnée dans les pays de l'Ouest mais se maintint dans les pays d'Europe de l'Est et de l'ancienne Union soviétique.

A partir des années 40, l'étude des phages devint un pan de la biologie moléculaire, avec les travaux célèbres de Max Delbrück, Salvador Luria, Alfred Hershey et leurs collègues (Groupe du Phage). Ils choisirent les virus de bactéries comme modèle, en raison de leur petit génome et de leur cycle de reproduction très court. Ils s'intéressèrent à des questions fondamentales, tels les mécanismes d'attaque, de multiplication et de lyse. Ces travaux contribuèrent grandement au développement de la biologie moléculaire et de la génétique bactérienne. Ce fut le « second âge du phage ».

Le « troisième âge du phage » a commencé dans le milieu des années 90. Un nouveau champ de recherche s'est ouvert, suite à la découverte de l'abondance des virus dans les écosystèmes, typiquement dans les océans ou dans les sols (Wommack and Colwell, 2000 ; Srinivasiah et al., 2008). Ainsi, le nombre de virus à ADN double-brin sur la planète pourrait être de l'ordre de 10^{31} et les virus seraient environ 10 fois plus nombreux que les cellules procaryotes (Whitman et al., 1998 ; Suttle, 2007 ; Srinivasiah et al., 2008) . Suite à cette découverte, la diversité des virus dans l'environnement ainsi que leur rôle au sein des écosystèmes ont fait l'objet de nombreuses recherches (3.2). Pour finir, notons que la thérapie phagique connaît actuellement un regain d'intérêt dans les pays de l'Ouest, suite à l'apparition de souches bactériennes multi-résistantes aux antibiotiques et à la très grande difficulté de mettre au point de nouveaux antibiotiques dans des délais appropriés (Debarbieux, 2008 ; Brüssow, 2005 ; Dublanchet and Fruciano, 2008).

3.1.3 Et les virus d'archées ?

Naturellement, les virus d'archées ne pouvaient pas être étudiés en tant que tels avant l'identification des archées en 1977. Quelques virus, infectant des espèces du genre *Halobacterium* (ex : Torsvik and Dundas, 1974) étaient néanmoins connus. Les virus et plasmides d'archées furent étudiés spécifiquement à partir des années 80, notamment sous l'impulsion de Wolfram Zillig. L'un des buts était de stimuler la mise au point d'outils de génétique pour les archées, alors récemment identifiées. Les premiers virus isolés infectaient des euryarchées, plus précisément des halophiles (revue dans Dyall-Smith et al., 2003) et des méthanogènes (ex : Jordan et al., 1989 ; Pfister et al., 1989). Ces virus-ci appartiennent aux familles Myoviridae ou Siphoviridae (type tête-queue), ressemblant donc beaucoup aux virus de bactéries.

Une image très différente émergea pour les virus de crenarchées. En 1981 furent décrits pour la première fois des virus infectant une archée hyperthermophile, *Thermoproteus tenax*, et tous étaient de morphotype filamenteux (Janekovic et al., 1983). En 1984, des

particules virales fusiformes furent produites par des cellules de *Sulfolobus shibatae* après induction aux UV (Martin et al., 1984). Le caractère infectieux de ces virions fut démontré en 1992 (virus SSV1, Schleper et al., 1992). A partir du milieu des années 90, une série de nouveaux virus de crenarchées hyperthermophiles furent isolés. Les travaux démontrèrent la grande particularité de ce groupe de virus (3.3.2) en révélant des morphotypes nouveaux, complexes et très divers (Prangishvili and Garrett, 2005).

3.2 Ecologie et diversité des virus

Comme indiqué ci-dessus (3.1.2), la grande abondance des virus dans l'environnement est une découverte récente. Les virus sont sans aucun doute les entités les plus nombreuses sur la planète, mais en raison de leur petite taille, ils représenteraient environ 5% de la biomasse procaryotique (Suttle, 2007). Il faut toutefois mentionner que certaines méthodes d'observations comportent des biais (Weinbauer, 2004) qui pourraient conduire à surestimer le nombre de particules virales : la cytométrie en flux ou la microscopie à épifluorescence ne permettent pas de discriminer entre des vésicules contenant de l'ADN ou de l'ADN adsorbé à un débris, et des particules virales ; elles ne permettent pas non plus de faire la différence entre un très gros virus et une très petite cellule. Néanmoins les différentes méthodes ont été comparées et ces biais ne doivent pas affecter la nature du résultat.

Dans les océans, les virus tueraient chaque jour approximativement 20% de l'ensemble de la biomasse (Suttle, 2007), et doivent donc jouer un rôle important dans les réseaux trophiques en contribuant à la production de matière organique dissoute (Figure 10). Ils pourraient également être impliqués dans les cycles biogéochimiques (Fuhrman, 1999 ; Suttle, 2007).

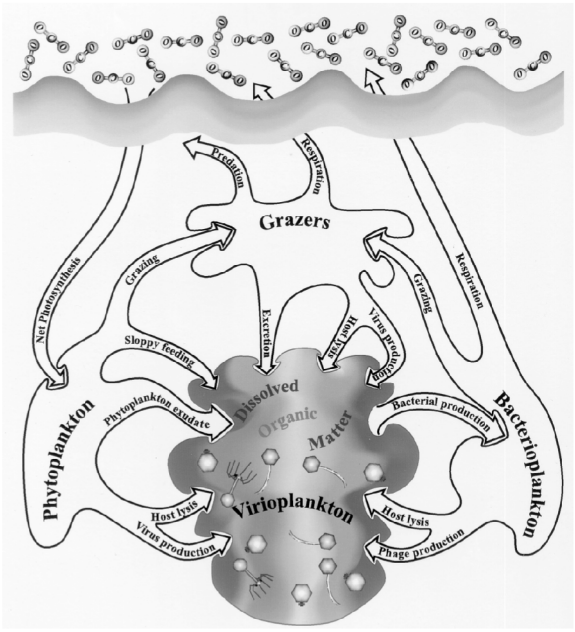


Figure 10 : Virus et boucle microbienne. Schéma indiquant le rôle possible de l'infection et la lyse virale dans la production de matières organiques dissoutes, dans les écosystèmes aquatiques. D'après Wommack and Colwell, 2000.

Suite à cette constatation et comme pour les microoganismes cellulaires, une démarche de caractérisation de la diversité virale a été entreprise. Cependant, il n'y a pas d'équivalent de l'ARN ribosomique chez les virus, c'est-à-dire de gènes ou de structures conservés et ubiquitaires permettant d'évaluer simplement et systématiquement la diversité. La tâche est donc plus complexe.

Un premier type d'approche consiste à séquencer plusieurs phages d'une même famille (ex : Comeau et al., 2007 pour les phages de la famille T4 ; Kwan et al., 2005 pour les phages de *Staphylococcus aureus* ; Hatfull et al., 2008 pour les mycobactériophages). Ces études ont révélé le mosaïcisme des phages (Hendrix, 2003; Desplats and Krisch, 2003) : les gènes sont organisés en modules qui s'échangent facilement d'un virus à l'autre (transfert horizontal), par des mécanismes de recombinaison encore peu compris. Certains modules sont conservés au sein d'une même famille de phages, et d'autres contiennent au contraire des gènes très peu conservés. Ces derniers constituent un immense réservoir de diversité génétique qui comprend beaucoup d'ORFans (Yin and Fischer, 2008). L'origine des gènes non-conservés et leurs fonctions au cours du cycle viral sont encore peu connues, bien que des hypothèses aient été formulées (Comeau et al., 2008).

Un second type d'approche est la métagénomique, qui consiste à séquencer systématiquement ce qui est présent dans un échantillon environnemental. Parfois, seuls quelques gènes, présents chez de nombreux virus, ont été ciblés (Breitbart et al., 2004a; Comeau and Krisch, 2008). Ou bien, c'est l'ensemble du matériel génétique viral qui a été séquencé. La métagénomique des virus a déjà été appliquée à des environnements très variés (revue dans Comeau et al., 2008) et elle semble démontrer la présence de nombreux gènes du métabolisme cellulaire dans les génomes viraux (Comeau et al., 2008). Mya Breitbart et Forest Rohwer effectuent de nombreuses études de métagénomique environnementale. Ils ont en particulier analysé des sources chaudes terrestres, où l'on s'attend à trouver – entre autres – des archées hyperthermophiles et leurs virus (Breitbart et al., 2004b). Un point intéressant est qu'ils ont observé relativement moins de virus dans les milieux chauds que dans d'autres types de milieux. Les concentrations rapportées dans l'article cité sont de l'ordre de 10^6/mL alors qu'elles sont de l'ordre de 10^7/mL dans les océans et encore supérieures dans les sols (Srinivasiah et al., 2008).

Un troisième type d'approche consiste à observer les échantillons en microscopie électronique, ce qui permet d'observer les morphotypes mais limite les possibilités d'analyses. On peut alors ensuite utiliser des méthodes culture-dépendantes, en passant par des étapes d'enrichissement, pour tenter d'isoler des systèmes hôtes-virus. Cela introduit sûrement un fort biais sélectif, mais permet l'étude approfondie de virus en laboratoire. C'est une des approches qui est beaucoup développée pour les virus de crenarchées hyperthermophiles.

3.3 Les virus d'archées

Les virus d'archées sont très divers, par leurs morphotypes (Figure 11) et par leur contenu génétique. En effet, ils sont répartis dans 9 familles virales différentes et certains virus ne sont pas encore classés (exemples dans le Tableau 4). Ceci est d'autant plus frappant qu'une quarantaine de virus d'archées seulement ont été observés en microscopie électronique, contre plus de 5000 virus de bactéries (Ackermann, 2007). Au total, une trentaine de virus ont été isolés et caractérisés plus en détail (Tableau 4). Tous contiennent de l'ADN double brin et la plupart ont été séquencés.

En fait, ce sont les virus infectant les crenarchées qui concentrent l'essentiel de la diversité et des nouvelles familles virales, avec une série de morphotypes uniques et exceptionnels. Mais le tableau de la diversité virale chez les archées est encore inachevé puisque l'accent a surtout été mis sur les hôtes des ordres Sulfolobales et Halobacteriales. Des clades entiers (y compris d'archées cultivables) n'ont pas encore de virus connus, ou très peu. Dans le Tableau 4, on peut voir qu'un seul virus a été isolé pour les Thermococcales et les Methanococcales, et qu'aucun virus n'a été isolé pour les Desulfurococcales, Thermoplasmatales, Methanosarcinales et Methanomicrobiales.

3.3.1 Les virus d'euryarchées

La plupart des virus d'euryarchées isolés sont des virus tête-queue très similaires aux bactériophages (Figure 11L). Outre le morphotype commun, ils présentent des caractéristiques similaires de mosaïcisme génomique (Tang et al., 2002) et ont de plus un génome linéaire. Par la suite, différents virus fusiformes ont été identifiés (hôtes appartenant aux ordres Halobacteriales et Thermococcales, Tableau 4), un virus icosaédrique (Figure 11K, ordre Halobacteriales) et un virus oblong (ordre Methanococcales). Les virus d'Halobacteriales sont très abondants dans les lacs salés (Porter et al., 2007) et parmi eux, certains sont tempérés. Pour les autres, l'interaction avec l'hôte n'est pas encore bien comprise. Il est possible que les trois catégories classiques d'interaction hôte-virus (lytique ou virulent, tempéré, chronique) soient trop schématiques pour décrire parfaitement les interactions observées.

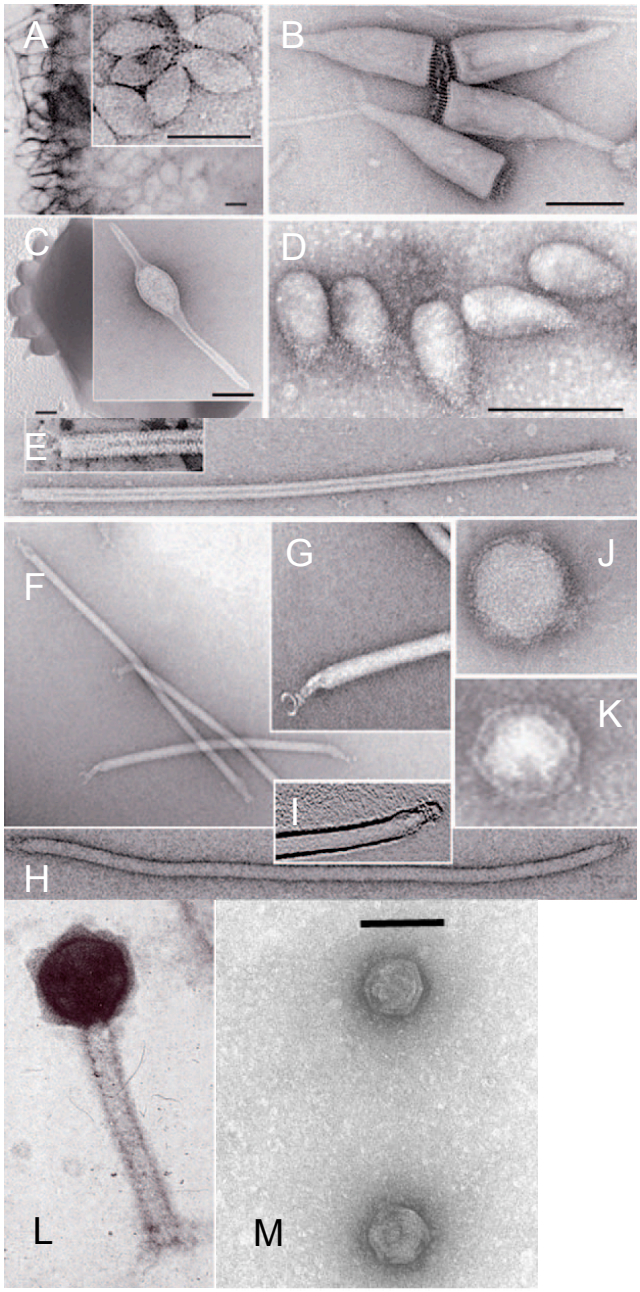


Figure 11 : Exemples représentatifs de morphotypes de virus d'archées. K, L : virus d'euryarchées. Tous les autres : virus de crenarchées. A : *Sulfolobus* spindle-shaped virus 1 (SSV1), famille Fuselloviridae. B : *Acidianus* bottle-shaped virus (ABV), famille Ampullaviridae. C : *Acidianus* two-tailed virus (ATV), famille Bicaudaviridae. D : *Sulfolobus neozealandicus* droplet-shaped virus (SNDV), famille Guttaviridae. E : *Sulfolobus islandicus* rod-shaped virus 1 (SIRV1), famille Rudiviridae. F-G : *Acidianus* filamentous virus 1 (AFV1), Lipothrixviridae. H-I : *Acidianus* filamentous virus 2 (AFV2), famille Lipothrixviridae. J : *Pyrobaculum* spherical virus (PSV), famille Globuloviridae. K : *Haloarcula hispanica* virus (SH1), non-classé. L : *Halobacterium halobium* virus ΦH, famille Myoviridae. M : *Sulfolobus* icosaedral turreted virus (STIV), non-classé. A-L : d'après Prangishvili et al., 2006a. M : http://serc.carleton.edu/microbelife/yellowstone/

Ordre	Famille	Type de morphotype	Génome	Nb	Séquence	Intégré	Enveloppe	Noms
Euryarchaeota								
Halobacteriales	Myoviridae	Icosaédrique/Tête-queue	L, ≈ 59	2	o	o	n	ΦH ΦCh1
	Non-classé HF1-HF2	Icosaédrique/Tête-queue	L, ≈ 76	2	o	n	n	HF1 HF2
	Non-classé SH1	Icosaédrique/Sphérique	L, ≈ 31	1	o	n	o	SH1
	Fuselloviridae, Salterprovirus	Fusiforme	C, ≈ 15	2	o	n	o?	His1 His2
Methanobacteriales	Siphoviridae	Icosaédrique/Tête-queue	L, ≈ 26-30	2	o	o?	n	ΨM1 ΨM2
Methanococcales	Non-classé VLP	Exceptionnel/Oblong	C, ≈ 23	1	n	o	o	-
Thermococcales	Non-classé PAV1	Fusiforme	C, ≈ 18	1	o	n	o	PAV1
Crenarchaeota								
Sulfolobales	Lipothrixviridae-β	Linéaire/Filament	L, ≈ 40	6	o	n	o	SIFV AFV3 AFV6 AFV7 AFV8 AFV9
	Lipothrixviridae-γ	Linéaire/Filament	L, ≈ 21	1	o	n	o	AFV1
	Lipothrixviridae-δ	Linéaire/Filament	L, ≈ 32	1	o	n	o	AFV2
	Rudiviridae	Linéaire/Baguette	L, ≈ 25-35	4	o	n	n	SIRV1 SIRV2 ARV1 SRV
	Non-classé STIV	Icosaédrique	C, ≈ 18	1	o	n	n	STIV
	Fuselloviridae/Fusellovirus	Fusiforme	C, ≈ 15	5	o	o	o	SSV1 SSV2 SSV4 SSV-RH SSV-K1
	Guttaviridae	Exceptionnel/Goutte	C, ≈ 20	1	n	-	-	SNDV
	Bicaudaviridae*	Exceptionnel/≈ Fusiforme	C, ≈ 60	1	o	o	?	ATV
	Non-classé STSV1	≈ Fusiforme	C, ≈ 75	1	o	n	o	STSV1
	Ampullaviridae*	Exceptionnel/Bouteille	L, ≈ 24	1	o	n?	o	ABV
Thermoproteales	Globuloviridae*	Sphérique	L, ≈ 21-28	2	o	n	o	PSV TTSV1
	Lipothrixviridae-β	Linéaire/Filament	-	2	n	-	o	TTV2 TTV3
	Lipothrixviridae-α	Linéaire/Filament	L, ≈ 16	1	o	o?	o	TTV1

Tableau 4 : La diversité des virus d'archées : principales caractéristiques des virus d'archées isolés et décrits dans la littérature. Les virus sont classés selon le phylum puis l'ordre de l'hôte, et enfin selon la famille virale. Colonne « Génome » : indique si les génomes sont circulaires (C) ou linéaires (L), et leur taille en kb. Colonne « Nb » : indique le nombre de virus entrant dans la même catégorie (même famille et hôte du même ordre). Colonne « Séquence » : indique si la séquence de l'un des virus au moins est disponible (o) ou non (n). Colonne « Enveloppe » : indique si les virus sont enveloppés (o) ou non (n), quelle que soit la nature de l'enveloppe. La présence d'une astérisque à côté de noms de familles virales indique que le processus d'approbation par le ICTV est encore en cours. D'après Prangishvili et al., 2006a, et Peng, 2008.

3.3.2 Les virus de crenarchées

Côté crenarchées, seuls des virus de crenarchées hyperthermophiles ont été isolés jusqu'à présents. Ils ont été obtenus après des étapes d'enrichissement, à partir d'échantillons environnementaux.

Malgré un nombre de travaux encore relativement restreint et une étape initiale d'enrichissement très probablement sélective, un grand nombre de morphotypes différents et uniques a été obtenu (Ackermann, 2007, Tableau 4). Les morphotypes qui semblent dominer sont les formes linéaires (filaments, baguettes) et fusiformes (voir Tableau 4). Les virus de morphotype filamenteux sont très nombreux et ont été classés en plusieurs sous-familles. *A contrario*, les morphotypes plus rares, et très surprenants, sont par exemple celui d'ABV (forme de bouteille, Figure 11B) et celui d'ATV (Figure 11C). Si le corps, chez ATV, est

« classiquement » fusiforme, les virions présentent la particularité de développer deux queues lorsqu'ils sont en dehors de l'hôte, sous certaines conditions de température (Prangishvili et al., 2006c). Le stade de développement extracellulaire d'un virus est sans précédent, tous domaines confondus. Par contre, le morphotype tête-queue est absent chez les crenarchées, alors qu'il est ultra-majoritaire chez les bactéries (95% des morphotypes observés, Ackermann, 2007), et courant chez les euryarchées.

Le contenu génétique des virus de crénarchées est également unique. En effet, la plupart des gènes ne présentent aucune homologie avec les séquences disponibles dans les bases de données (Figure 12A). Les quelques gènes présentant une homologie sont souvent partagés par d'autres virus de crenarchées hyperthermophiles ou par l'hôte (Figure 12, A et B). Les génomes de certains virus, comme ABV (Ampullaviridae) et le duo PSV-TTSV1 (Globuloviridae) ne présentent même strictement aucune homologie.

Tout ceci contribue à faire des virus de crenarchées hyperthermophiles un groupe très particulier, et sur la base de leur unicité de morphotypes et de contenu génétique, ils ont été classés en sept nouvelles familles. Quatre familles on été approuvées par l'International Committee on Taxonomy of Viruses (ICTV). Il s'agit des familles Fuselloviridae, Lipothrixviridae, Rudiviridae et Guttaviridae. Pour trois autres, Ampullaviridae, Bicaudaviridae, et Globuloviridae, le processus d'approbation est en cours. Enfin, certains virus attendent encore d'être classés, comme le virus icosaédrique STIV (Figure 11M), et le virus fusiforme PAV1.

Un point commun à la plupart de tous ces virus est la présence d'appendices sur la particule virale. Ceux-ci doivent permettre l'adsorption des particules virales à l'hôte, comme l'indique parfois l'observation de particules adsorbées à des pili via ces appendices. Des exemples sont les trois petites fibres aux extrémités des rudivirus (Figure 11E), les diverses extrémités des lipothrixvirus (pinces chez AFV1, Figure 11G, structures plus compactes ressemblant à une brosse chez AFV2, Figure 11I, etc.), la fibre présente à une seule extrémité des fusellovirus (Figure 11A, insert), ou encore les petites tours sur la capside de STIV (Figure 11M). Par ailleurs, la plupart des virions ont une enveloppe (Tableau 4), dont la nature n'est pas toujours clairement établie, en particulier concernant la présence possible de lipides (Prangishvili, communication personnelle).

Les génomes font 13,7 à 75,3 kpb et comprennent 31 à 74 cadres ouverts de lecture (ORFs). Ils sont linéaires, notamment lorsque le morphotype l'est aussi (Lipothriviridae, Rudiviridae, Ampullaviridae, Globuloviridae) ou circulaires (tous les autres) (Tableau 4). Les extrémités des génomes linéaires comportent des régions inversées répétées, allant de quelques bases pour AFV1 à quelques kb pour SIRV2, et des motifs répétés. Les régions terminales interviennent *a priori* lors de la réplication du génome, ce qui est courant pour les génomes linéaires (ex : Peng et al., 2001). Pour la plupart des virus, l'intégration du génome viral au chromosome de l'hôte n'a pas été observée (Tableau 4). Elle a lieu chez les fusellovirus (SSV1, etc.) et chez ATV ; ceux-ci ont des génomes circulaires qui codent pour une intégrase. Chez SSV1, l'intégrase est étudiée en détail (ex : Letzelter et al., 2004) et c'est bien elle qui permet l'intégration au génome de l'hôte (Muskhelishvili et al., 1993).

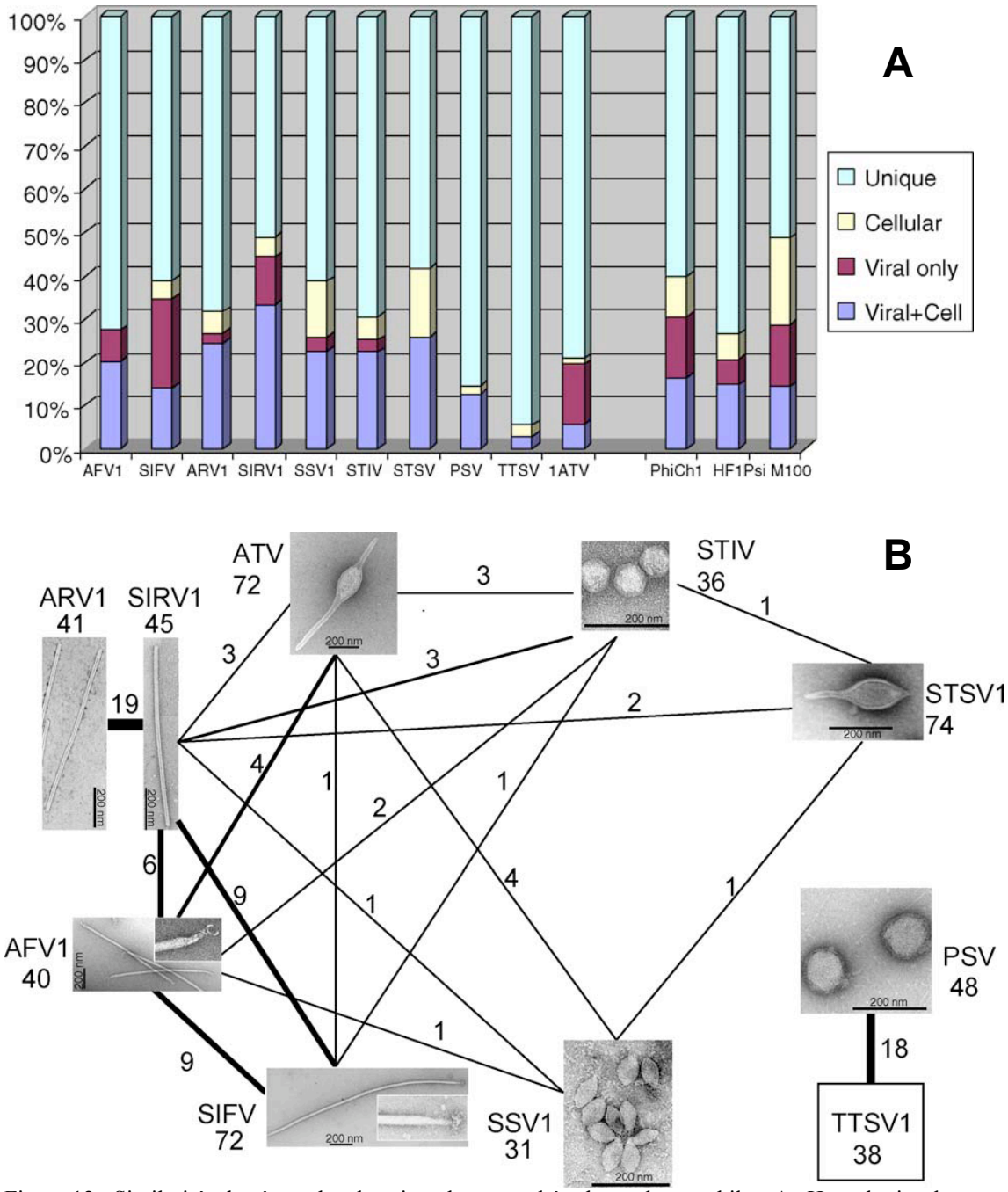


Figure 12 : Similarités de gènes chez les virus de crenarchées hyperthermophiles. A : Homologies des gènes ; quelques virus d'euryarchées ont été inclus (les 3 derniers à droite). L'axe Y correspond à un pourcentage de gènes au sein du génome. Les homologies entre virus très proches (exemple : SIRV1, ARV1) n'ont pas été prises en compte. B : Les gènes orthologues au sein des virus d'archées. D'après Prangishvili et al., 2006b.

Concernant les relations avec l'hôte, différents cas existent. Pour la plupart des virus, le génome ne s'intègre pas au chromosome de l'hôte et aucune lyse cellulaire n'est observée. Il a donc été postulé que les particules virales sont produites et émises en continu sans que la cellule ne soit endommagée, et que celle-ci continue à se diviser, quoique plus lentement (Prangishvili and Garrett, 2005). Il s'agirait ainsi d'une relation avec l'hôte similaire à celle que développent les bactériophages filamenteux (ex : M13). Dans le cas des crenarchées

hyperthermophiles, cette interaction aiderait les virus à se maintenir dans l'environnement agressif, en leur garantissant un refuge intracellulaire (Prangishvili and Garrett, 2005) et elle a été dénommée état porteur (« carrier state ») (Zillig et al., 1994). Mais l'état porteur n'est pas encore bien caractérisé. Comme pour les virus d'Halobacteriales, un flou persiste sur la nature exacte des relations hôtes-virus. Quelques virus de crenarchées développent des interactions différentes de l'état porteur. Il s'agit des fusellovirus, combinant lysogénie et état porteur ; d'ATV combinant lysogénie et phase lytique ; et enfin de STIV et TTV1, purement lytiques (Prangishvili et al., 2006b ; Ortmann et al., 2008).

L'exploration de la diversité des virus de crenarchées hyperthermophiles se poursuit, par l'utilisation d'approches métagénomiques (Snyder et al., 2007 ; Schoenfeld et al., 2008, et voir fin de la partie 3.2) et par l'isolement de nouveaux virus (ex : Xiang et al., 2005). Concernant les virus déjà isolés, les recherches ont pris plusieurs directions (revue dans Lawrence et al., 2009). SSV1, premier virus de crenarchée isolé (voir 3.3) et l'un des rares à présenter un caractère lysogène, est l'un des mieux caractérisés ; c'est un virus modèle, utilisé notamment pour l'étude des phénomènes d'intégration et d'induction (voir ci-dessus). SIRV2 et STIV sont également beaucoup étudiés. D'une manière générale, les aspects structuraux sont beaucoup abordés, que ce soit par l'analyse structurale des virions par des techniques de cryo-microscopie électronique et de reconstruction (ex : Vestergaard et al., 2008a) ou par l'identification des protéines composant les virions (ex : Reiter et al., 1987a). Des projets importants de génomique structurale sont également en cours, dans l'espoir d'obtenir des indications sur les fonctions des protéines ne présentant aucune homologie avec des systèmes connus. Par ailleurs, l'expression des gènes viraux a été analysée pour SIRV2 (Kessler et al., 2004), SSV1 (Reiter et al., 1987b ; Fröls et al., 2007), et STIV (Ortmann et al., 2008) avec une approche transcriptomique pour les études de 2007 et 2008 concernant respectivement SSV1 et STIV. Des études sur les mécanismes moléculaires de la transcription et la réplication de certains virus sont en cours. Enfin, les études *in silico* se développent, allant de la génomique comparative (Prangishvili et al., 2006b) à des études spécifiques et ciblées, par exemple sur les CRISPRs (ex : Vestergaard et al., 2008b) ou sur l'usage des cystéines et des ponts disulfures (ex : Menon et al., 2008).

3.4 Virus de crenarchées hyperthermophiles et évolution

Selon certaines hypothèses, les virus, d'une manière générale, auraient joué un grand rôle dans les transitions évolutives majeures. Ce sujet dépasse le cadre de la thèse et ne sera pas décrit en détail. Mentionnons simplement que les virus pourraient être à l'origine du passage d'ARN à ADN et pourraient également être à l'origine de l'apparition du noyau chez les eucaryotes (ex : Forterre, 2006a et 2006b). Ceci s'ajoute au fait qu'ils constituent, au sein de chaque domaine, le plus grand réservoir de gènes pouvant être transmis horizontalement aux hôtes ou à d'autres virus (ex : Filée et al., 2002).

La découverte des virus infectant les crenarchées hyperthermophiles pose des questions spécifiques. Elles sont exposées et discutées dans Prangishvili et al., 2006a, et les connaissances actuelles permettent déjà de proposer des scénarios hypothétiques.

Les virus à ADN double-brin des trois domaines du vivant forment des ensembles bien distincts, puisque chaque famille virale est présente dans un seul des trois domaines à la fois (Figure 13). L'unique exception concerne les virus tête-queue, avec les familles Myoviridae et Siphoviridae, retrouvées à la fois chez les bactéries et les euryarchées.

Par ailleurs, des capsides icosaédriques sont présentes dans les trois domaines, même si elles correspondent à des familles virales différentes (Figure 13) et que les génomes associés ne présentent pas d'homologie de séquence. Des études spécifiques (Bamford et al., 2005), basées sur les structures des capsides et des protéines les composant, suggèrent qu'il ne s'agit pas de convergence et que les virus des trois domaines pourraient avoir évolué à partir d'un ensemble d'entités et de gènes antérieurs à LUCA (« Last Universal Common Ancestor »).

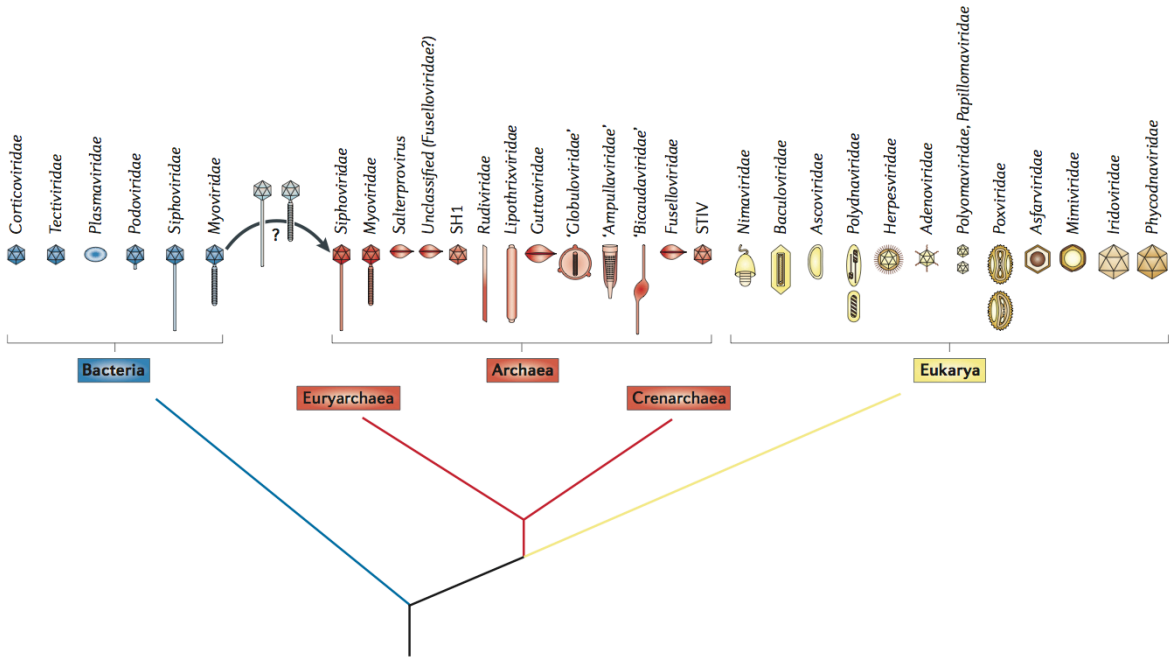


Figure 13 : Les virus à ADN double-brin dans les 3 domaines du vivant. La flèche indique les transferts horizontaux possibles. D'après Prangishvili et al., 2006a.

Ceci a conduit à l'hypothèse selon laquelle les trois virosphères distinctes auraient été sélectionnées à partir d'un pool ancien de gènes, en concomitance avec l'émergence des domaines. Les trois virosphères seraient désormais réfractaires aux échanges horizontaux inter-virosphère. La présence de virus tête-queue chez les euryarchées constituerait l'unique exception et correspondrait à un transfert plus récent, de la virosphère bactérienne vers celle des archées (Prangishvili et al., 2006a). Ces hypothèses sont résumées dans la Figure 14.

Toutefois, la plupart des nouveaux morphotypes de virus d'archées correspondent à des milieux chauds (≥ 80°C), et l'on ne peut donc pas totalement exclure une spécificité qui refléterait une adaptation aux hautes températures, et ne concernerait pas l'ensemble des archées.

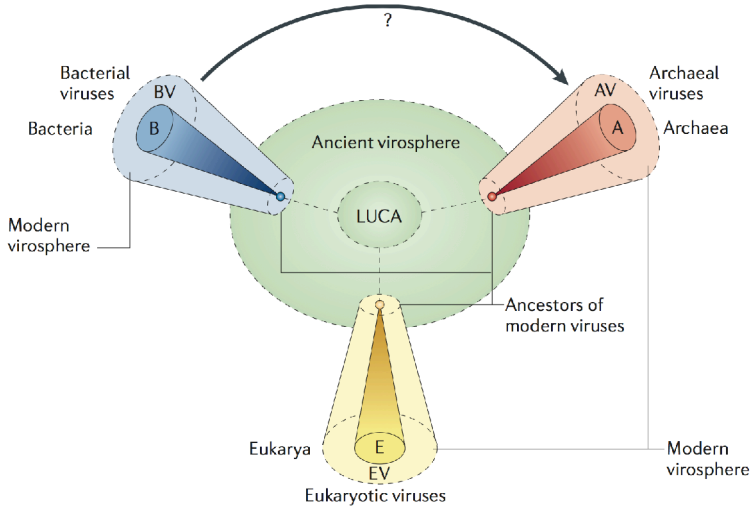


Figure 14 : L'origine des virus. La flèche représente un possible transfert horizontal. D'après Prangishvili et al., 2006a.

La poursuite de l'exploration de la virosphère des archées pourrait apporter des éléments de réponses à ces questions sur l'évolution.

4

Présentation du travail réalisé

Mon travail de thèse, transversal, comporte trois parties très différentes et indépendantes.

La première partie est expérimentale et concerne l'étude du virus SIRV2. Elle a principalement consisté à clarifier la nature de l'interaction du virus avec son hôte, celle-ci n'étant pas précisément décrite. Le virus s'est avéré être virulent, ce qui était inattendu. Secondairement, les particules virales ont été mieux caractérisées (stabilité, protéines de structures). Un projet de coévolution expérimentale basé sur SIRV2 avait été initié, mais il a été progressivement abandonné face aux difficultés théoriques et pratiques rencontrées et n'est pas présenté ici.

La seconde partie, expérimentale également, a consisté à enrichir des échantillons environnementaux afin d'isoler et de caractériser de nouveaux virus. Un virus et son hôte ont été obtenus. Le virus, de morphotype filamenteux, appartient à une famille déjà connue (Lipothrixviridae), et son étude a permis d'approfondir les connaissances sur cette famille virale (identification de nouvelles protéines structurales par exemple). Cette partie contribue à la description de la diversité virale intra-famille.

La dernière partie est une étude bioinformatique des motifs exceptionnels courts (4 à 8 pb) dans les génomes d'archées et de leurs éléments extra-chromosomiques. Ce type d'approche peut contribuer à l'identification de motifs fonctionnels non-codants et n'est pas encore très utilisé pour les archées. Nous avons tiré partie de l'accumulation des séquences disponibles pour produire des données de références et pour tenter de dégager des pistes intéressantes. Dans ce cadre, nous avons également examiné la composition des séquences d'éléments extra-chromosomiques et observé que les séquences forment des groupes cohérents, reflétant l'ordre auquel appartiennent leurs hôtes, et reflétant également les familles d'appartenance des virus et des plasmides.

Tous ces travaux ont impliqué plusieurs personnes, au sein de BMGE et en collaboration avec des plate-formes de l'Institut Pasteur, avec d'autres laboratoires en France et à l'étranger, ce qui illustre bien le caractère collectif de la démarche scientifique.

Sulfolobus islandicus rod-shaped virus 2 et interaction hôte-virus chez les crenarchées

1
Introduction

Au cours de ma thèse, j'ai beaucoup travaillé sur le rudivirus *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2). La manipulation du virus est bien maîtrisée au laboratoire BMGE, puisque le virus a été isolé par David Prangishvili (BMGE, Institut Pasteur) et ses collaborateurs, lorsque David travaillait au Max-Plank-Institut. Le but principal de mon projet de thèse était de caractériser l'interaction hôte-virus. Aussi, Olivier Tenaillon (Écologie et Évolution des Microorganismes, UMR INSERM - Université Paris 7, site Xavier Bichat) a apporté son expertise théorique et expérimentale sur les interactions hôte-virus et l'évolution.

Pour commencer, je décrirai l'état de l'art sur les interactions hôte-virus chez les crenarchées (section 2), puis le modèle biologique *Sulfolobus islandicus* – SIRV2 (section 3). Après une brève présentation du travail réalisé (section 4), les résultats sont inclus, sous la forme de trois articles rédigés en anglais.

Interaction hôte-virus chez les crenarchées : état de l'art

Les paragraphes suivants constituent un approfondissement de notions sur les virus de crenarchées, abordées dans l'introduction (I.3.3.2).

Les virus de crenarchées isolés et caractérisés jusqu'à présent infectent des hôtes appartenant aux ordres Thermoproteales (environ 5 virus) et Sulfolobales (une quinzaine de virus). Il s'agit donc uniquement d'hôtes hyperthermophiles. Concernant l'ordre Thermoproteales, les hôtes appartiennent aux genres *Thermoproteus* et *Pyrobaculum*, et sont neutrophiles et anérobies exclusifs. Concernant l'ordre Sulfolobales, les hôtes appartiennent aux genres *Sulfolobus* et *Acidianus*, et sont acidophiles. Les espèces du genre *Sulfolobus* sont aérobies exclusives, tandis que celles du genre *Acidianus* sont anaérobies facultatives, et poussent en présence de soufre. Enfin, tous ces virus ont été isolés à partir d'échantillons environnementaux prélevés en de nombreux points du globe (ex : États-Unis, Italie, Islande, Açores, Russie, Japon, Indonésie, Nouvelle-Zélande), et tous issus de sources géothermiques. Concernant la caractérisation des virus, l'accent a beaucoup été mis jusqu'à présent sur les aspects génomiques, moléculaires et structuraux.

2.1 L'état porteur, la relation hôte-virus majoritaire chez les crenarchées

Les relations hôte-virus sont encore assez peu caractérisées chez les crenarchées. A quelques exceptions près, on pense que les virus établissent une relation chronique avec la cellule-hôte, se maintenant dans un état porteur (« carrier state » en Anglais) (voir I.3.3.2).

Dans l'état porteur (Prangishvili and Garrett, 2005 ; Prangishvili et al., 2006b), le virus ne provoque pas de lyse cellulaire. Il y a un équilibre entre réplication virale et émission de particules virales infectieuses d'une part, et réplication et division cellulaire d'autre part. Le génome viral persiste donc dans la cellule pendant de nombreuses générations, sans toutefois être intégré au génome cellulaire. En outre, la production de virions n'est pas stimulée par une exposition à différents stress, comme les UV ou l'ajout de mitomycine C. L'état porteur diffère donc bien d'un état lysogène. Enfin, la croissance cellulaire est ralentie par l'infection. Cette dernière propriété permet, si l'hôte peut former un tapis cellulaire, d'observer des halos révélant le retard de croissance des cellules infectées, et éventuellement de déterminer le titre des suspensions de virus.

Pour les virus de bactéries, les relations hôte-virus prédominantes sont la lyse et la lysogénie (Abedon, 2006). C'est également souvent le cas pour les virus d'euryarchées de type tête-queue (Siphoviridae, Myoviridae). Aussi, il a été suggéré que la singularité de l'interaction hôte-virus observée chez les crenarchées hyperthermophiles correspondrait à une stratégie de survie du virus, bénéfique dans les environnements géothermiques. L'état porteur permettrait aux virions d'éviter un contact direct et prolongé avec le milieu extérieur agressif (Prangishvili and Garrett, 2005).

2.2 Des preuves expérimentales encore insuffisantes

Mais les données expérimentales sur la base desquelles les virus ont été qualifiés de chroniques sont encore limitées et ne constituent pas une démonstration solide. Dans la plupart des cas, le virus a été considéré comme chronique si :

- aucune chute de la DO n'était observée suite à l'infection
- il n'y avait pas de débris cellulaires dans les cultures infectées.

Deux autres arguments parfois mentionnés sont la possibilité d'isoler des colonies produisant le virus, et l'éventuelle capacité du virus à se maintenir dans une culture au cours de dilutions sérielles. Des exemples de telles observations, couvrant l'ensemble des familles virales de crenarchées, sont détaillés dans le Tableau 5.

Famille	Genre de l'hôte	Nom	Colonie	Absence de débris	DO ne chûtant pas	Retard de croissance	Halo sur tapis	« Coévolution »	Intégration	Inductible	Réf
Etat_porteur											
Ampullaviridae	Acidianus	ABV	-	+	+	+		?	-?	?	Häring et al., 2005
Globuloviridae*	Pyrobaculum	PSV	-	+	+	-		+ (plusieurs mois)	?	?	Häring et al., 2004
Globuloviridae*	Thermoproteus	TTSV1	+	?	-	+		?	-	?	Ahn et al., 2006
Guttaviridae	Sulfolobus	SNDV	+	+	+(/-)	+	+(si concentré)	- (disparition progressive)	-	?	Arnold et al., 2000a
Lipothrixviridae-γ	Acidianus	AFV1	+	+	+	+		+ (2 mois)	-?	?	Bettsteter et al., 2003
Lipothrixviridae-β	Sulfolobus	SIFV	+	+	+	+	+	- (disparition progressive)	-	?	Arnold et al., 2000b
Rudiviridae	Acidianus	ARV1	+(200 col)	+	+	+		?	-?	?	Vestergaard et al., 2005
Rudiviridae	Sulfolobus	SIRV2	+	+	+	+	+	-	-	-	Prangishvili et al., 1999
Rudiviridae	Sulfolobus	SIRV1	+	+	+	+	+	-	-	-	Prangishvili et al., 1999 Zillig et al., 1994
Non-classé	Sulfolobus	STSV1	?	+	+	+	+	+ (10 transferts)	-	?	Xiang et al., 2005
Etat_porteur et lysogénie											
Fuselloviridae	Sulfolobus	SSV1	+	+	+	+	+	+ (24 jours)	+	+	Martin et al., 1984 Schleper et al., 1992 Stedman et al., 2003
Fuselloviridae	Sulfolobus	SSV2	+	+	+	+	+	?	+	+	Stedman et al., 2003 Contursi et al., 2006
Lyse et lysogénie											
Bicaudaviridae*	Acidianus	ATV	+	-	-			?	+	+	Prangishvili et al., 2006c
Lyse											
Lipothrixviridae-α	Thermoproteus	TTV1	+	-	-				?	?	Janekovic et al., 1983
Non-classé	Sulfolobus	STIV	?	?	-		?	?	-	-	Ortmann et al., 2008

Tableau 5 : Données expérimentales sur l'interaction hôte-virus chez les crenarchées hyperthermophiles. « Colonie » : possibilité d'isoler des colonies de cellules produisant le virus. « Absence de débris » : absence de débris cellulaires visibles dans les cultures infectées. '200 col' indique que 200 colonies ont du être analysées. « DO ne chûtant pas » : absence de chûte de la DO dans les cultures infectées. « Retard de croissance » : croissance ralentie des cultures infectées. « Halo sur tapis » : observation de zones de retard de croissance causées par le virus, sur un tapis cellulaire. « Coévolution » : capacité du virus à se maintenir dans une culture lors de dilutions et incubations successives. « Intégration » : intégration du génome viral dans le génome de l'hôte. « Inductible » : production de particules virales induite par exposition aux UV ou ajout de mitomycine C. « Réf » : références bibliograhiques. Un signe « + » signifie que l'assertion est exacte, un signe « - » qu'elle ne correspond pas aux observations expérimentales. La coloration grise signifie que l'assertion n'est pas pertinente (par exemple, si l'hôte ne forme pas de tapis cellulaire, la présence de halo ne peut pas être testée). Un signe « ? » signifie que l'information n'est pas disponible dans la littérature.

Ces diverses observations sont généralement insuffisantes pour conclure avec certitude sur le mode d'interaction hôte-virus. Par exemple, à l'inverse de ce à quoi l'on pourrait s'attendre, un virus lytique peut coévoluer pendant de très longues périodes avec son

hôte (ex : *Pseudomonas fluorescens* et le bactériophage Φ2, Buckling and Rainey, 2002). Au contraire, des cellules peuvent devenir totalement résistantes à un virus chronique, si celui ci a un fort impact négatif sur la croissance des cellules (ex : *Escherichia coli* et les bactériophages filamenteux M13, f1, fd, Bull et al., 1991). Par ailleurs, l'observation d'une augmentation de DO dans une culture infectée peut correspondre à la croissance de mutants résistants. Tous les cas sont donc possibles.

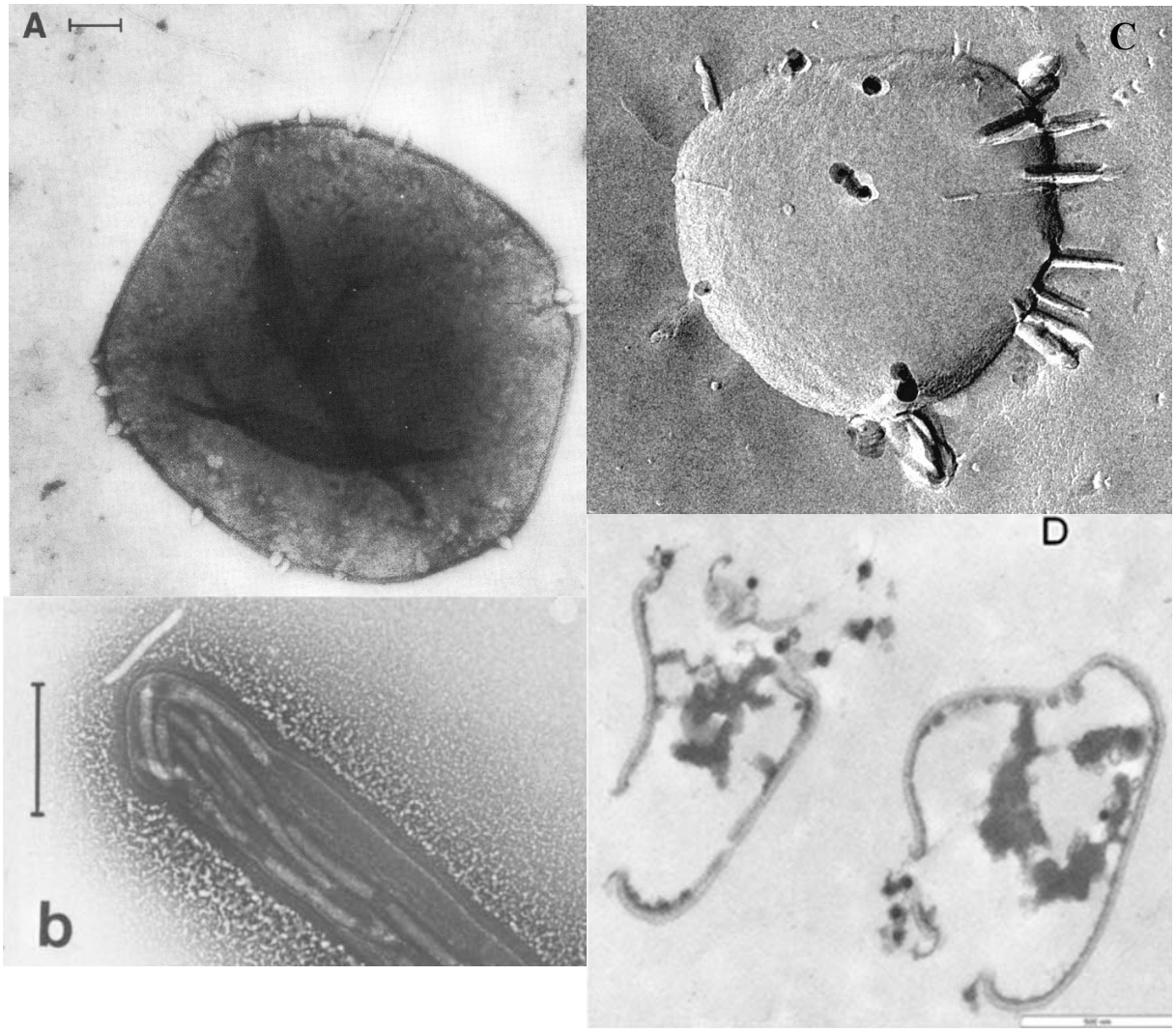


Figure 15 : Cellules de crenarchées infectées par des virus. (A) : Cellule de *Sulfolobus shibatae* infectée par SSV1. D'après Martin et al., 1984. (B) : Virions de TTV1 à l'intérieur d'une cellule de *Thermoproteus tenax*. D'après Janekovic et al., 1983. (C) : Cellule de *Sulfolobus islandicus* infectée par SIRV1. D'après Vestergaard et al., 2005. (D) : Section de cellules lysées de *Sulfolobus solfataricus* infectées par STIV. D'après Ortmann et al., 2008.

Parfois, les cellules infectées, intactes, ont été observées en microscopie électronique en transmission. Il est alors possible d'apercevoir les virions assemblés à l'intérieur de la cellule (exemples dans la Figure 15). Mais les cellules ont rarement été observées en coupe (Figure 15D), donc les informations sur le processus de sortie des virions sont limitées.

Les virus les mieux caractérisés du point de vue de l'interaction avec l'hôte sont SSV1 et SSV2 d'une part, STIV d'autre part. Concernant STIV (Ortmann et al., 2008), l'observation de sections de cellules issues de cultures infectées a démontré l'effet lytique du virus (Figure 15 D). Concernant SSV1 et SSV2, les phénomènes d'intégration et d'induction du virus ont été beaucoup étudiés. Dans les cellules infectées, le génome viral est présent sous la forme d'un provirus intégré au chromosome de l'hôte et sous forme épisomale (Martin et al., 1984 ; Schleper et al., 1992). Le site d'intégration est connu (Schleper et al., 1992 ; Muskhelishvili et al., 1993), et une particularité de SSV1 est que le gène codant pour l'intégrase (famille tyrosine recombinase) est clivé lors de l'intégration (Palm et al., 1991). Par ailleurs, il est prouvé que la production de particules virales est induite par une exposition aux rayons ultra-violets (Martin et al., 1984) et que l'électroporation de l'ADN viral dans les cellules conduit à la production de particules infectieuses (Schleper et al., 1992). Récemment, il a été démontré que l'intégrase n'est pas essentielle (Clore and Stedman, 2007), c'est-à-dire que des particules virales infectieuses sont encore produites lorsque le gène viral codant pour l'intégrase est supprimé. Enfin, chez SSV2, le degré d'induction de la réplication a été comparé dans différentes souches de *Sulfolobus*, incluant la souche naturelle (Contursi et al., 2006). Mais concernant le processus de sortie des virions, il n'y a pas eu d'études spécifiques. Il a été conclu que le virus n'est pas lytique, sur la base d'observations expérimentales similaires à celles indiquées ci-dessus (Tableau 5).

En conclusion, beaucoup d'aspects de la relation hôte-virus chez les crenarchées doivent encore être caractérisés. Quelques paramètres généraux du cycle infectieux ont déjà été déterminés dans certains cas, comme, par exemple, la période de latence et la période d'éclipse de SIRV1 et SIRV2 (Prangishvili et al., 1999). Mais la nature exacte de l'interaction n'est pas toujours claire, et les différentes étapes du cycles infectieux ne sont pratiquement pas connues (mécanismes d'adsorption, d'assemblage, de sortie des virions).

Pour étudier l'interaction hôte-virus, nous avons choisi le virus SIRV2, un rudivirus. Ce virus est relativement simple à manipuler et compte parmi les virus les mieux caractérisés. En outre, ce virus n'est pas lysogène, ce qui permet de ne pas confondre les effets liés à la lysogénie et les effets liés au cycle productif. SIRV2 infecte un hôte du genre *Sulfolobus*, le genre le mieux caractérisé chez les archées. Le modèle biologique est présenté en détail ci-après (section 3).

3
Le modèle biologique

Des souches de *Sulfolobus islandicus* ont été utilisées comme hôte pour SIRV2. Nous présenterons donc le genre *Sulfolobus* en mentionnant les informations plus spécifiques à *Sulfolobus islandicus* lorsqu'elles sont disponibles. Puis nous décrirons le virus SIRV2 et la famille Rudiviridae.

3.1 Le genre *Sulfolobus*

En raison de la chronologie de la découverte des archées (I.2.1 et I.2.3.1) et parce que les archées hyperthermophiles sont, contre toute attente, plus aisément cultivables que leurs sœurs mésophiles, plusieurs organismes modèles pour l'étude des archées sont des hyperthermophiles. Pour les crenarchées, ce sont des espèces du genre *Sulfolobus* (ordre des Sulfolobales), toutes hyperthermophiles, acidophiles, et aérobies strictes, qui sont actuellement utilisées.

3.1.1 Principales caractéristiques

Sulfolobus est l'un des six genres de l'ordre Sulfolobales. Les autres genres sont *Acidianus*, *Metallosphaera*, *Stygiolobus*, *Sulfurococcus*, *Sulfurisphaera* (Itoh, 2003). Les cellules de *Sulfolobus* sont coccoïdes, très irrégulières. Leur diamètre est de l'ordre de 0,5 à 2 µm et elles se divisent symétriquement. La paroi cellulaire est composée de la membrane lipidique et d'un S-layer (Gram négatif). Elles possèdent des pili et des flagelles, et produisent des vésicules par un processus de bourgeonnement encore peu caractérisé (Ellen et al., 2009).

Tous les organismes du genre *Sulfolobus* sont aérobies strictes, et ont une croissance optimale aux alentours d'une température de 80°C et d'un pH de 2-3 (hyperthermophiles et acidothermophiles). Elles maintiennent un pH intracellulaire d'environ 5,5 (Lübben and Schäfer, 1989). Dans leur habitat naturel, les sources chaudes terrestres de type solfatares, les espèces de *Sulfolobus* sont chimiolithoautotrophes : elles utilisent les composés soufrés comme source d'énergie en produisant de l'acide sulfurique, et le CO_2 est leur source de carbone. Elles peuvent aussi croître en mode hétérotrophe, tirant leur énergie et leur carbone de différents composés organiques. Elles sont présentes dans de nombreuses sources chaudes terrestres du globe et ont par exemple été isolées d'Islande, d'Italie, du Japon, de la Nouvelle-Zélande, de la Russie et des Etats-Unis.

Trois d'entre elles ont été séquencées intégralement, *Sulfolobus acidocaldarius*, *Sulfolobus solfataricus* et *Sulfolobus tokodaii*. Des projets de séquençage sont en cours, par exemple pour 9 souches de *Sulfolobus islandicus*. Les génomes sont circulaires. Pour les trois organismes intégralement séquencés, la taille de génome varie de 2,1 à 3,0 Mpb, avec approximativement 2200 à 3000 protéines codées, et un GC% compris entre 33% et 37%.

Les génomes sont riches en éléments mobiles (Brügger et al., 2004), en particulier celui de *Sulfolobus solfataricus* (Redder and Garrett, 2006). C'est *Sulfolobus acidocaldarius* qui présente le génome le plus stable. Dans certains génomes, les CRISPRs sont très représentés : c'est le cas chez *Sulfolobus solfataricus* et *Sulfolobus tokodaii* (Peng et al., 2003 ; Lillestol et al., 2006). Les taux de mutations ont pu être estimés pour plusieurs espèces et souches (ex : Berkner and Lipps, 2008a), en particulier pour des souches de *Sulfolobus islandicus* qui semblent d'un même niveau de stabilité que *Sulfolobus acidocaldarius*.

Les espèces de *Sulfolobus* sont associées à un très grand nombre d'éléments extra-chromosomiques (virus et plasmides conjugatifs ou cryptiques). Pour *Sulfolobus islandicus*, ils s'agit de : SIFV (lipothrixvirus), SIRV1 et SIRV2 (rudivirus), SSV2, SSV4 et SSV5 (fusellovirus) pour les virus (I.3.3.2), pHEN7, pRN1, pRN2, pXZ1 pour les plasmides cryptiques, pKEF9, pSOG1, pSOG2, pARN3, pARN4, pING1 pour les plasmides conjugatifs, et pSSVx, hybride entre un plasmide et un fusellovirus.

Enfin, mentionnons que certaines souches de *Sulfolobus islandicus* produisent des toxines protéiques appelées sulfolobicines auxquelles sont susceptibles d'autres souches de la même espèce, ainsi que *Sulfolobus solfataricus* et *Sulfolobus shibatae* (Prangishvili et al., 2000).

3.1.2 *Sulfolobus*, organisme modèle

Sulfolobus est progressivement devenu l'organisme modèle pour les crenarchées et pour l'étude de mécanismes fondamentaux chez les archées (réplication, transcription, traduction, voies métaboliques, par exemple). Les espèces les plus utilisées jusqu'à présent sont *Sulfolobus solfataricus* et *Sulfolobus acidocaldarius*. De nombreux avantages justifient le choix du genre *Sulfolobus*.

Sulfolobus peut-être cultivé en « mode » hétérotrophe, en conditions aérobies, ce qui est relativement simple, surtout par rapport à ce qu'exigent d'autres archées. Les cellules poussent par exemple sur du tryptone, éventuellement en combinaison avec de l'extrait de levure, ou sur divers sucres. En conditions de laboratoire, le temps de doublement varie de trois à une douzaine d'heures, ce qui reste raisonnable. Des densités cellulaires satisfaisantes sont également atteintes dans les cultures (DO_{600nm} pouvant dans certains cas dépasser 1,0). Il est d'ailleurs rapporté que certaines souches utilisées en laboratoire ont perdu la capacité à croître en chimiolithoautotrophie (*Sulfolobus solfataricus* et *Sulfolobus acidocaldarius*, Prangishvili, communication personnelle). Par ailleurs, les cellules poussent aussi bien en milieu liquide que sur milieu solide, sur une matrice à base de gomme gellane. Il est possible d'obtenir des colonies et des tapis cellulaires. Les cellules peuvent être congelées, en employant du DMSO ou du glycérol.

Un dernier avantage, très important est le grand nombre d'éléments extra-chromosomiques, comme mentionné au 3.1.1, ce qui permet d'envisager le développement d'outils de génétique.

Naturellement, quelques points faibles sont aussi apparus. C'est tout d'abord l'instabilité génétique de certaines espèces ou souches liée à la présence d'éléments mobiles dans les génomes (3.1.1). Par ailleurs, la mise au point d'outils de génétique s'est avérée très complexe, alors qu'elle fut plus simple pour les Halobacteriales. Les difficultés sont liées par exemple à la présence de systèmes de restriction-modification dans les cellules, à la spécificité des souches (difficulté de mettre au point de systèmes généralistes fonctionnant pour beaucoup d'espèces de *Sulfolobus*), à des rendements de transformation plutôt bas et à la difficulté d'obtenir de bons marqueurs sélectifs. Parfois, les systèmes mis aux points ne pouvaient être reproduits dans d'autres laboratoires, à cause d'un trop haut niveau d'expertise requis ou de la sensibilité du système aux petites fluctuations qui surviennent inévitablement entre laboratoires et expérimentateurs (Berkner and Lipps, 2008b).

Après plus d'une dizaine d'années de travail sur ce sujet, un nombre conséquent de systèmes sont disponibles et se développent à un rythme croissant (revu dans Berkner and Lipps, 2008b). Pour *Sulfolobus solfataricus*, des vecteurs navette (« shuttle vectors ») et des systèmes de knock-out existent. Pour *Sulfolobus acidocaldarius*, en revanche, aucun système de knock-out n'est disponible, ce qui pose problème. Concernant *Sulfolobus islandicus*, un article paru très récemment, (She et al., 2009), décrit la mise au point et l'utilisation d'une gamme d'outils de génétique qui apparaît comme très prometteuse : par exemple, un système de knock-out permettant l'étude de gènes essentiels a été développé. A l'avenir, il pourrait être intéressant d'obtenir des protéines rapporteuses fluorescentes, qui ne sont pas encore disponibles pour *Sulfolobus*.

3.1.3 Exemples d'avancées importantes réalisées grâce à *Sulfolobus*

Sulfolobus a été étudié de manière extensive pour les mécanismes de réplication et de transcription, les systèmes métaboliques et le cycle cellulaire. De nombreuses études biochimiques et structurales de ses protéines ont été réalisées. Plusieurs puces à ADN sont disponibles et des études transcriptomiques globales sont décrites (ex : Götz et al., 2007, réponse aux UV). Enfin, des projets protéomiques ont démarré dans plusieurs laboratoires (ex : Barry et al., 2006).

Sans être exhaustif, soulignons qu'une ARN-polymérase de *Sulfolobus* fut isolée dès 1979 par Wolfram Zillig (Zillig et al., 1979). Ceci permit de remarquer la ressemblance avec les systèmes eucaryotes (ARNPII) et de poursuivre l'étude de la transcription chez les archées. Concernant la réplication, *Sulfolobus* a également beaucoup été utilisé, par exemple pour les études structurales et biochimiques de protéines de la machinerie de réplication. Là encore, les protéines sont homologues à celles des eucaryotes (revu pour *Sulfolobus solfataricus* par Duggin and Bell, 2006). Si la première origine de réplication fut identifiée chez *Pyrococcus abyssi* (Myllykallio et al., 2000 ; Matsunaga et al., 2001), c'est chez *Sulfolobus solfataricus* que furent identifiées de multiples origines de réplication (Robinson et al., 2004 ; Lundgren et al., 2004). Il a été suggéré que *Sulfolobus* pourrait constituer un bon modèle simplifié pour l'étude de certains aspects de la réplication « de type eucaryote » (ex : Kelman and Kelman, 2004, Lundgren et al., 2004), indiquant que l'intérêt de *Sulfolobus*

comme modèle dépasse parfois le cadre des archées. On peut par exemple penser à l'étude de la terminaison de la réplication, en présence de multiples origines de réplication. Il n'existe toutefois pas encore de système de réplication *in vitro* pour les archées alors que la transcription *in vitro* est au point.

S'agissant du cycle cellulaire, *Sulfolobus* est l'archée la mieux caractérisée. En cette matière, il existe des différences majeures entre crenarchées et euryarchées. Les archées utilisent des protéines nombreuses et variées pour moduler la compaction de leur ADN au cours du cycle cellulaire. Alors que les euryarchées ont, entre autres, des histones homologues à celles de eucaryotes, les crenarchées n'en possèdent pas. Les protéines constituant la chromatine des crenarchées les plus étudiées viennent de *Sulfolobus* (Barry and Bell, 2006). Il s'agit d'une protéine spécifique au genre, Sul7d, et d'une protéine trouvée dans un grand nombre d'euryarchées et de crenarchées, Alba. Par ailleurs, alors que les euryarchées utilisent un système similaire à celui des bactéries pour la division cellulaire (mécanisme basé sur FtsZ), les crenarchées ont un système différent. Jusqu'à très récemment, aucun gène et aucune protéine n'étaient connus. On sait désormais qu'un opéron de trois gènes, cdv, intervient au cours de la division, et que les protéines Cdv polymérisent entre les deux nucléoïdes ségrégeant, puis forment une structure plus petite pendant la constriction. Ce système de division, unique, a été mis en évidence sur *Sulfolobus acidocaldarius* (Lindas et al., 2008). Beaucoup reste encore à découvrir sur les cycles cellulaires des archées et leur régulation.

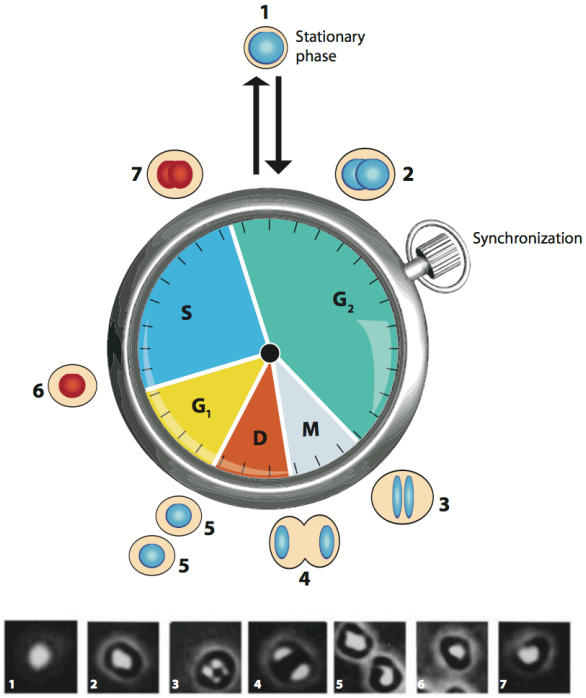


Figure 16 : Représentation schématique du cycle cellulaire chez *Sulfolobus*. D'après Bernander, 2007.

Le cycle cellulaire de *Sulfolobus* comporte trois phases principales (Figure 16, Bernander, 2007). Il s'agit de la phase pré-réplicative G1 (< 5% du temps de doublement), de la phase de réplication S (≈ 33%), de la phase post-réplicative (60-65%). Pendant la majeure partie de la phase post-réplicative, les deux génomes ne sont pas ségrégés (stade G2). La ségrégation et la constriction ont lieu respectivement pendant les stades courts M et D de la phase post-réplicative. La majorité des cellules dans une culture contient donc deux génomes. Dans les cultures en phase stationnaire, les cellules sont bloquées au stade G2, ce qui fournit l'un des moyens, simple, de synchroniser les cellules. Récemment, il a été démontré que les cycles cellulaires sont similaires chez des crenarchées représentatives des ordres Sulfolobales, Desulfurococcales et Thermoproteales (Lungdren et al., 2008).

3.2 *Sulfolobus islandicus* rod-shaped virus 2, SIRV2

3.2.1 Principales caractéristiques

SIRV2 appartient à la famille Rudiviridae (I..3.3.2) et a été isolé de sources géothermiques situées en Islande (Prangishvili et al., 1999). Les autres virus de cette famille sont SIRV1, ARV1 et SRV, isolés respectivement d'Islande (Zillig et al., 1994), d'Italie (Vestergaard et al., 2005) et des Açores (Vestergaard et al., 2008b). SIRV2 et SIRV1 ont des génomes présentant de fortes homologies de séquences (Peng et al., 2001). Des particules virales de morphotype identique à celui des rudivirus sont fréquemment rencontrées dans les environnements géothermiques (I..3.3.2).

Comme pour tous les rudivirus, les virions de SIRV2 sont en forme de baguette (23 nm x 900 nm) avec trois petites fibres à chaque extrémité (I..3.3.2). Les virions sont en fait des tubes creux (diamètre de la cavité intérieure, 6 nm) car il s'agit de nucléoprotéines enroulées en hélice, un mode d'empaquetage du génome qui rappelle le TMV (I.3.1.1). Les particules virales sont extrêmement robustes (Prangishvili et al., 1999).

Comme les autres rudivirus, SIRV2 possède un génome d'ADN double-brin linéaire (35502 kpb) dont les extrémités, liées de façon covalente, sont des séquences inversées répétées (1628 bp). Les séquences inversées répétées contiennent des répétitions directes multiples. Le génome a un bas contenu en GC, de 25% (Peng et al., 2001). Il comprend 54 ORFs, et les deux premières (et en conséquence les deux dernières) sont incluses dans les séquences terminales inversées répétées. Une dizaine d'ORFs a des homologues dans les génomes de tous les rudivirus connus (Figure 17). Ce sont généralement des ORFs de grande taille ou de taille moyenne, situées dans la partie centrale du génome, tandis que les gènes courts, situés vers les extrémités, sont non-conservés. Peu de fonctions ont été identifiées : une protéine structurale (Prangishvili et al., 1999), une dUTPase (Prangishvili et al., 1998), une Holliday Junction Resolvase (Birkenbihl et al., 2001). Mais certaines ORFs ont des fonctions putatives (ex : glycosyl transférase) ou codent pour des domaines connus (ex : RHH). Enfin, des insertions-délétions de fragments de 12 pb (ou multiples de 12 pb) ont été remarquées dans plusieurs génomes de rudivirus, mais pas chez SIRV2 (Vestergaard et al., 2008b). Le rôle possible de ces insertions-délétions n'est pas évident.

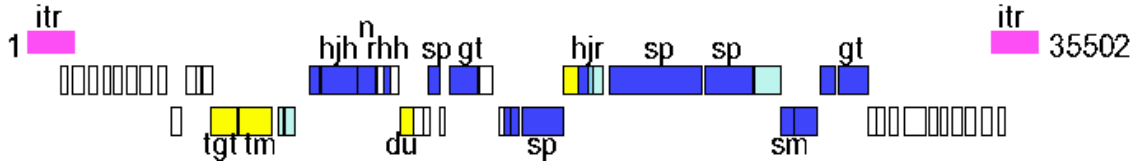


Figure 17 : Carte des ORFs (cadres ouverts de lecture) dans le génome de SIRV2. Les chiffres indiquent la taille du génome. Rose, itr : régions terminales inversées répétées. Bleu foncé : gènes très conservés, ayant au moins un homologue dans chaque génome de rudivirus séquencé. Bleu clair : gènes conservés avec un moindre degré d'homologie, ayant un homologue dans chaque génome de rudivirirus séquencé. Jaune : gènes conservés dont des homologues sont présents dans les autres génomes de rudivirus à l'exception d'ARV1. Fonctions prouvées ou putatives : tARN guanine transglycosylase (tjt) ; transmembrane (tm) ; Holliday junction hélicase (hjh) ; nucléase (n) ; "ribbon-helix-helix" protéine (rhh) ; dUTPase (du) ; protéine structurale (sp) ; Holliday junction résolvase (hjr) ; S-adenosylméthionine-dépendente méthyltransférase (sm) ; glycosyl transférase (gt). Inspiré de Vestergaard et al., 2008b.

SIRV2 infecte certaines souches de *Sulfolobus islandicus*, comme LAL14/1, REN2H1 et peut-être HVE10/4. SIRV2 est stable dans l'hôte, il s'y maintient *a priori* dans un état porteur (2.1 et I.3.3.2). Cependant, comme précisé ci-dessus (2.2), ce type d'interaction n'est pas bien caractérisé.

3.2.2 Vers un virus modèle ?

Les études sur les virus de crenarchées sont encore à un stade précoce et il est difficile d'avoir une vision claire des lignes de recherche qui émergeront. Néanmoins, SIRV2 semble bien placé pour devenir un virus modèle aux côtés de SSV1, modèle pour les virus lysogènes.

D'un point de vue pratique, l'hôte du genre *Sulfolobus* est lui-même un organisme modèle et les conditions de culture sont relativement simples (3.1.2). Le particules virales peuvent être obtenues avec un bon rendement (en comparaison avec d'autres virus d'archées hyperthermophiles). Les concentrations en particules virales peuvent être estimées par titration sur tapis cellulaire (plages de retard de croissance cellulaire). Enfin, les particules peuvent être conservées à 4°C dans un tampon de Tris-acétate pH 6, ce pendant plusieurs années.

De plus, la famille des rudivirus est bien homogène, en termes de morphotypes et de contenu génétique, avec un set de gènes conservés assez conséquent. C'est beaucoup moins le cas chez les lipothrivirus, qui sont fréquemment observés dans les enrichissements mais sont beaucoup plus variables (extrémités des particules virales, sous-familles, moins de gènes conservés, voir I.3.3.2). La séquence intégrale du génome de SIRV2 est disponible.

Par rapport aux autres rudivirus, SIRV2 est certainement le meilleur candidat. SIRV1 infecte des hôtes de l'espèce *Sulfolobus islandicus*, les mêmes souches que SIRV2 à peu de choses près. Cependant, SIRV1 était à l'origine composé d'un mélange de variants, un nouveau variant dominant étant sélectionné lorsque le virus infectait une nouvelle population de cellules (Peng et al., 2004). Quelques variants purs ont pu être isolés, mais l'existence initiale de tant de variants au sein d'une même culture ne laisse pas présager d'un système très stable. Par ailleurs, ARV1 infecte un hôte du genre *Acidianus*, moins bien caractérisé et

légèrement moins facilement cultivable (soufre requis, densités cellulaires plus faibles, ne produit pas de tapis cellulaire). Et SRV infecte un hôte du genre *Stygiolobus*, également moins bien caractérisé.

Enfin, plusieurs études ont déjà été réalisées sur SIRV2. Quelques paramètres du cycle d'infection ont été évalués, comme la période d'éclipse et la période de latence (Prangishvili et al., 1999).

Concernant les aspects moléculaires, outre la caractérisation de la dUTPase et de l'Hollyday Junction Resolvase mentionnées ci-dessus, la carte de transcription des gènes viraux au cours d'une infection a été établie par des expériences d'hybridation Northern (Kessler et al., 2004). Cette étude a mis en évidence un schéma de transcription très uniforme. La seule exception concerne la protéine structurale principale, associée à un ARNm tout d'abord polycistronique, puis monocistronique à l'approche de l'éclipse.

Récemment, un système de transcription *in vitro* a permis de démontrer l'activation de la transcription à partir de promoteurs viraux, par la protéine régulatrice Sta1 codée dans le génome de l'hôte (Kessler et al., 2006). Sta1 est codée dans le génome de *Sulfolobus solfataricus*, et elle augmente l'expression du gène paralogue de radA en cas de lésion de l'ADN (Abella et al., 2007).

Enfin, des intermédiaires de réplication de SIRV2, des concatémères tête-tête, ont été observés, ainsi qu'une coupure spécifique de l'un des brins d'ADN dans les régions terminales (Peng et al., 2001). L'Holliday Hunction Resolvase intervient très probablement pour la conversion des concatémères en monomères (Birkenbih et al., 2001). Tout ceci a permis d'émettre des hypothèses sur les mécanismes de réplication du génome viral, et de postuler un rôle important pour les extrémités du génome : il pourrait s'agir d'une réplication unidirectionnelle par déplacement de brin, initiée par auto-amorçage au niveau de la coupure spécifique de l'un des brins d'ADN (Peng et al., 2001).

L'adoption définitive de SIRV2 comme modèle biologique sera certainement conditionnée par les possibilités de manipulation génétique du génome viral. Pour l'instant, aucun outil spécifique n'est disponible. En particulier, l'électroporation de l'ADN génomique de SIRV2 dans les cellules hôtes n'a jamais mené, jusqu'à présent, à la production de virions ; cela n'a d'ailleurs jamais fonctionné pour les virus de crenarchées hyperthermophiles au génome linéaire (Prangishvili, communication personnelle), alors que cela fonctionne bien pour SSV1, fusellovirus au génome circulaire et déjà virus modèle (Schleper et al., 1992).

En conclusion de cette partie sur le modèle biologique utilisé, *Sulfolobus* est et restera probablement un modèle pour l'étude des crenarchées hyperthermophiles, car les derniers obstacles pour le développement d'outils génétiques ne paraissent pas insurmontable. Le développement récent d'outils assez complets pour *Sulfolobus islandicus* (3.1.2), une espèce qui a l'air suffisamment stable, est extrêmement encourageant. Le choix de l'espèce *Sulfolobus islandicus* est également judicieux du fait que les projets de séquençage de plusieurs souches devraient aboutir dans les prochaines années. Quant à SIRV2, il est l'un des virus de crenarchées hyperthermophiles les mieux caractérisés actuellement, et il est relativement simple à utiliser d'un point de vue expérimental.

4

Présentation de l'étude

Les résultats sont présentés sous la forme de trois articles. Le premier article correspond au projet principal, sur l'interaction hôte-virus. Dans les deux article suivants, mes travaux constituent une contribution limitée, et concernent la caractérisation des particules virales de SIRV2.

Le premier article, soumis, concerne donc l'interaction hôte-virus chez SIRV2. Nous démontrons que SIRV2 est un virus virulent et observons un nouveau mécanisme de lyse. Patrick Forterre a contribué aux réflexions sur le projet et sur les virus en tant qu'organismes. Nos collaborateurs Erik Karlsson, Karin Ekefjärd et Rolf Bernander (Department of Molecular Evolution, Uppsala University, Suède), ont réalisé les analyses par cytométrie en flux. Nous avons également collaboré avec Marie-Christine Prevost (Institut Pasteur, plate-forme de microscopie ultrastructurale) pour l'observation des cellules infectées par microscopie électronique en transmission.

Le second article, dans le domaine des bionanosciences (Steintmetz et al., 2008), a pour but général de démontrer le potentiel des virions de SIRV2 pour des applications en nanotechnologies. Pour cela, les possibilités de marquage sélectif et localisé des particules virales ont été explorées. Cette caractérisation chimique des particules virales a été réalisée par Nicole Steinmetz et Marianne Manchester (Scripps Research Institute, USA), ainsi que Kim Findlay, George Lomonossoff et Dave Evans (John Innes Centre, UK). Ma contribution a été l'étude de la stabilité des virions dans différentes conditions (température ambiante, différents solvants) et par plusieurs techniques (titration sur tapis cellulaire, spectrophotométrie, microscopie électronique en transmission). En effet, la stabilité des virions est un atout important si l'on souhaite les manipuler pour des applications en nanotechnologies. A la suite des données supplémentaires de cet article, des résultats non-publiés sur la stabilité des virions à haute température (80°C) ont été inclus. Il s'agit cette fois de conditions proches de celles caractérisant l'environnement naturel.

Le troisième article (Vestergaard et al., 2008b) présente la description d'un nouveau rudivirus, SRV et rassemble différents travaux portant sur la famille Rudiviridae. Il s'agit notamment d'études de génomique comparative entre les rudivirus, et d'une étude *in silico* en lien avec les CRISPRs, systèmes putatifs de défense anti-virale chez les archées (I.2.2). Le séquençage et l'analyse des séquences ont été réalisées par Gisle Vestergaard, Roger Garrett et leurs collaborateurs (Danish Archaea Center, Danemark). Les parties expérimentales sur SRV et ARV1 ont été réalisées par les collaborateurs de David Prangishvili à l'université de Regensburg et au Max-Plank-Institut (biochimie). Pour cet article, avec l'aide de Pascal Lenormand (Institut Pasteur, plate-forme protéomique), j'ai identifié trois nouvelles protéines structurales de SIRV2. Les protéines constituant les virions ont été séparées par électrophorèse sur gel SDS-PAGE, puis analysées par la technique MALDI-TOF.

A Unique Virus Release Mechanism in Archaea

Ariane Bize[1], Erik A. Karlsson[2], Karin Ekefjärd[2], Marie-Christine Prevost[3], Patrick Forterre[1], Olivier Tenaillon[4], Rolf Bernander[2], David Prangishvili[1,*]

[1]Institut Pasteur, Biologie Moléculaire du Gène chez les Extrêmophiles, 25 rue du Docteur Roux, 75724 Paris cedex 15, France

[2]Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Norbyvägen 18C, SE-752 36 Uppsala, Sweden

[3]Institut Pasteur, Plate-Forme de Microscopie Ultrastructurale, 25 rue du Docteur Roux, 75724 Paris cedex 15, France

[4]INSERM U722, Faculté de médecine Xavier Bichat, Université Paris 7, 16 rue Henri Huchard, 75018 Paris, France

[*]E-mail: david.prangishvili@pasteur.fr, Phone: +33 (0)144 38 91 19, Fax: +33 (0) 145 68 88 34

Summary

We demonstrate that release of virions of the archaeal rod-shaped virus SIRV2 from host cells occurs through a novel biological mechanism. Large virus-induced protrusions, VIPs, presumably encoded by the virus, rupture the cell envelope at several positions, thereby creating large openings through which viral particles may escape the cell. We also demonstrate that massive degradation of the host chromosome occurs during virus infection and that virus assembly takes place in the cytoplasm. Furthermore, intracellular viral DNA is visualized by flow cytometry for the first time. The results show that SIRV2 does not persist in an intracellular carrier state, as previously believed, but is a lytic virus that transforms transiently *Sulfolobus* cells into elaborated viral factories.

Introduction

As for all organisms, cells belonging to the Archaea domain are infected by specific viruses. More than 30 archaeal viruses have been isolated and sequenced so far, all of which contain double-stranded DNA (dsDNA) as the genetic material (Prangishvili et al., 2006a). Among these, about 20 infect hyperthermophilic hosts belonging to the Crenarchaeota phylum. The morphological and genetic diversity and uniqueness of these viruses are such that they have been classified into seven new viral families (Prangishvili et al., 2008). The knowledge on the biology of this exceptional group of viruses is still limited, despite their importance for comprehension of the origin and evolution of DNA viruses and their role in early cellular evolution (Prangishvili et al., 2006a).

An important feature of crenarchaeal viruses is the presumed dominance of the carrier state, in which a non-lytic relationship is established with the host (Prangishvili and Garrett, 2005). It allows both the transmission of the viral genetic material between cell generations and the continuous release of virions at a low rate. Only three crenarchaeal viruses have been described as lytic (Prangishvili et al., 2006b; Ortmann et al., 2008; Janekovic et al., 1983), and these are considered exceptions, representing less frequently encountered virus families (Snyder et al., 2003; Prangishvili et al., 2006a). The carrier state lifestyle may thus be considered a characteristic feature of the Crenarchaeota, while in the Euryarchaeota phylum, and in Bacteria, lysis and lysogeny are dominant (Calendar and Abedon, 2006). Why this particular trait would be advantageous to crenarchaeal viruses is not clear, although it is possible that it could provide a durable intracellular refuge for the viruses under the harsh physico-chemical conditions encountered in their habitats ($T \approx 60$-$90°C$, $pH \approx 3.0 – 6.0$) (Prangishvili and Garrett, 2005).

The data available for classification of crenarchaeal viruses as chronic is, however, limited. The conclusions are often based on the lack of optical density (OD) reduction and the absence of cellular debris in infected cultures, and on the possibility to isolate colonies producing infectious virions (e.g. Schleper et al., 1992; Vestergaard et al., 2005). While these observations provide indirect evidence, studies specifically addressing host-virus interactions are scarce. Induction of replication was studied in detail for the lysogenic *Sulfolobus* spindle-shaped virus 2 (Contursi et al., 2006), and the transcriptional patterns of 3 different viruses have been determined (Frols et al., 2007; Kessler et al., 2004; Ortmann et al., 2008), but the results do not directly address the carrier state interaction. The difficulties are also due to the unique genetic content, with very few gene functions that can be predicted by *in silico* analyzes. To circumvent this, approaches such as studies of virion structure (e.g. Prangishvili et al., 2006b; Rice et al., 2004; Vestergaard et al., 2008) and determination of the three-dimensional structures of encoded proteins (e.g. Keller et al., 2007; Larson et al., 2007) have been used. Despite these efforts, however, the number of genes with known roles remains limited: e.g. for 54 putative genes of *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2), one of the best characterized crenarchaeal virus, functions of only 6 genes have been confirmed (Birkenbihl et al, 2001; Prangishvili et al., 1998; Vestergaard et al., 2008).

To better understand the nature of host-virus interaction in crenarchaea, we selected and studied the virus SIRV2 and its host *Sulfolobus islandicus* (*S. islandicus*). The virus

belongs to the Rudiviridae (Prangishvili et al., 1999), a common crenarchaeal virus family (Prangishvili et al., 1999; Rice et al., 2001; Verstergaard et al., 2005; Vestergaard et al., 2008; Zilig et al., 1994), and contains a linear 35.5 kb dsDNA genome; SIRV2 has been described as a carrier state, non-lysogenic virus (Prangishvili et al., 1999). The host belongs to *Sulfolobus*, a well-characterized crenarchaeal genus (Bernander, 2007; Duggin et al., 2006), and from which also other viruses are known (Prangishvili et al., 2008). Using this model system, we describe detailed *in vivo* effects of the virus on its host and, unexpectedly, demonstrate that SIRV2 is a cytocidal, lytic virus. Remarkably, an entirely novel virus release mechanism was encountered during the characterization, involving pyramidal structures that allow viral escape, by causing a local disruption of the cell envelope. In addition, intracellular viral DNA was visualized by flow cytometry for the first time, and the technique was also used to demonstrate chromosome degradation in infected cells. SIRV2 therefore kills *Sulfolobus* cells and transforms them transiently into very efficient and complex viral factories. The results also show that lytic viruses are more frequent in extreme hot environments than previously believed.

Results

Growth kinetics of SIRV2-infected cultures

OD and CFU values from uninfected and infected (m.o.i. about 7) cultures were monitored over time. The effects of the virus were visible already 1.5 h after addition (Figure 1, arrows). Whereas uninfected cultures pursued normal growth with a generation time around 13 h, the OD in infected cultures remained constant for about 60 h (Figures 1A and 1C), after which growth at normal rate resumed (Figure 1C). During this time period, the CFU values of uninfected controls remained constant or increased slightly. In contrast, the CFU values decreased dramatically in infected cultures, resulting in an about 1000-fold reduction at 6 h post-infection (Figure 1B, 10.5 h). The CFU values also revealed growth of a minor cell population in infected cultures starting at early time points (Figure 1D, from 15 h). This was not detectable in the OD measurements (Figure 1C), due to the low concentration of this cell population at early time points. Thus, infection by SIRV2 had a pronounced effect on the cultures, preventing growth of a majority of the cells.

To exclude the possibility that the results were linked to the high m.o.i. employed, or to the specific growth conditions, similar experiments were performed at low m.o.i. (about 10^{-3}), at different temperatures (70°C; 75°C; 78°C), different pH (3.0; 3.5), different medium richness (standard medium or 5-fold less rich medium), and with different host strains (*S. islandicus* strains KVEM10H3, HVE10/4 and LAL14/10). No significant differences were observed (data not shown), indicating that the effects occurred independently of these parameters.

Characterization of the cell population growing in the presence of SIRV2

The cell population growing in the presence of SIRV2 could consist of virus-resistant cells, or of cells containing SIRV2 in a carrier state. To discriminate between these possibilities, 23 colonies were isolated from each of 3 infected cultures (Inf1, Inf2, Inf3) and from an uninfected control culture, resulting in a total of 92 clones.

To assess the production of infectious virions, supernatants from minicultures were spotted on a fresh cell lawn. No growth inhibition zones were detected (data not shown), indicating that none of the 92 clones produced a detectable amount of virus during growth, disfavoring the carrier state hypothesis. The sensitivity of the same 92 clones to SIRV2 infection was tested by growth in minicultures in the presence of SIRV2 added at a low m.o.i. (10^{-3}). In late exponential phase, the OD was measured and the virus concentration in the supernatants assessed by a spot test. As expected, the control clones were highly sensitive to SIRV2 infection: growth was strongly inhibited by the virus (Figure 2A) and the spot tests were positive (Figure 2B, clearings in the left column), each clone thus having produced a high amount of infectious virions. In contrast, the clones from the Inf1-3 cultures displayed no growth reduction (Figure 2A) and the spot tests did not reveal an increase in virus concentration in the supernatants (Figure 2B, 3 right columns). Thus, the cells able to grow in the presence of SIRV2 were totally resistant to re-infection and did not produce new infectious virions. The high initial proportion of resistant cells strongly suggests that, in addition to random mutations, specific mechanisms could have been involved in their generation, such as CRISPRs-related mechanisms (Barrangou et al., 2007).

To determine whether the resistance could be linked to the presence of the SIRV2 genome within the cells, a series of PCR reactions with primer pairs distributed along the SIRV2 genome were performed on 6 clones from the infected cultures. All results were negative (data not shown), and the resistance could consequently not be correlated to the presence of the SIRV2 genome. This is consistent with the fact that the SIRV2 genome never has been observed to integrate into the host chromosome (Prangishvili et al., 1999), and does not encode a DNA integrase (Peng et al., 2001). The results also excluded the possibility that the SIRV2 genome is present in a non-integrated carrier state in the cells.

Flow cytometry analysis of infected cells over a time course

The cell size and intracellular DNA content in uninfected and SIRV2-infected cultures (m.o.i. about 10) over a time course were monitored by flow cytometry (Figures 3, Supplemental Data).

The relative lengths of the *S. islandicus* cell cycle periods in the control cultures were found to be similar to those of other *Sulfolobus* species (Bernander and Poplawski, 1997; Bernander, 2007), with about 2% of the generation time occupied by the pre-replicative phase, 30% by the replicative phase and 68% by the post-replicative phase (Figure S2). The genome size was estimated to 2.6 Mb (Figure S1). The cell size (Figure 3A, left column) progressively decreased when the cultures approached stationary phase. In the infected cultures (Figure 3B, left column), a cell size increase initially occurred in part of the cell population, evident as an extension of the distribution towards the right (6 h - 8 h). Subsequently, the average cell size gradually decreased over time.

The DNA content distributions of the control cultures (Figure 3A, right column) were typical for exponentially growing *Sulfolobus* cells (Bernander and Poplawski, 1997), with a majority of the cells containing 2 chromosomes. In the infected cultures (Figure 3B, right column), massive degradation of the chromosome DNA began at 0.5 h post-infection and, at 12 h, a majority of the cell population contained no detectable intracellular DNA. The proportion of cells containing 1 – 2 genome equivalents decreased over time, concomitant with an increase in the proportion of cells with a very low DNA content (<<1 genome equivalent). The SIRV2 latent period has previously been estimated to 6 h (Prangishvili et al., 1999) and thus, chromosome degradation occurred before virus release. It is likely that viral DNA tightly packaged into virus particles is poorly stained (below); in this case, cells with degraded chromosome DNA still could contain viral genomes.

Interestingly, the populations of chromosome-less cells and cells containing DNA were clearly separated and well defined (Figures 3B, right column, 3C and 3D). Thus, for a given infected cell, chromosome degradation must occur within a brief time interval. If otherwise, a broad range of DNA contents would be present in the distributions, progressively moving towards low values and lacking distinct peaks corresponding to complete chromosomes.

In parallel to chromosome degradation, an increase in the total DNA content occurred in part of the cell population, evident as an extension of the 2-genome equivalents peak towards the right (Figure 3B, right column, before 9 h). This corresponded to newly synthesized viral DNA (below), and the increase was estimated to about 0.5 genome equivalents on average (Figure 3E, arrow), at 3 h after infection. Assuming that the cells still

contained two intact chromosomes, the SIRV2 DNA represented 1.3 Mb/cell, or equivalent to 37 viral genomes. However, if chromosome degradation already had been initiated, or if part of the viral DNA already had been packaged into viral particles and thereby become refractory to staining, the number of viral genomes would be underestimated (see below).

In addition to the first ever detection of intracellular virus DNA by flow cytometry, the results demonstrate that infection by SIRV2 causes massive degradation of the host chromosome in virtually all infected cells during the first 12 h of infection, excluding the possibility that SIRV2 genomes are vertically transmitted between cell generations.

Links between the virus infection cycle and the kinetics of host chromosome degradation

To discriminate between host chromosome and viral DNA, uninfected and infected cultures (m.o.i. about 15) were monitored by dot blot hybridizations, in addition to flow cytometry. In an uninfected control culture, the percentage of chromosome-less cells did not exceed 5% (Figure 4A), and tended to decrease over time. In infected cultures, chromosome-less cells were already evident during the first hours: at 5 h, the percentage was about 40%, confirming that degradation occurred before virion release, and at 11 h, >80% of the cells were chromosome-less. Subsequent degradation occurred at a lower rate and finally reached 97%, confirming that genome degradation occurred in most cells.

The intracellular amount of SIRV2 DNA (Figures 4B and 4D) increased gradually and reached a maximum around 8 h, followed by a rapid decrease up until 14 h. The initial increase presumably corresponded to viral DNA replication, and the decrease to virus release, indicating a latent period of about 9 h. Thus, a single round of infection occurred in the cultures at the high m.o.i. employed. To relate viral DNA production to the dynamics of chromosome degradation, the percentage of DNA-less cells appearing between successive time points was superimposed (Figure 4B). A first small peak of degradation was visible at 0.5 h post-infection. The major peak occurred at 11 h, in the middle of the virus release period, indicating that genome degradation mainly occurred towards the end of the infection cycle. The use of a 16S rRNA gene probe combined with similar data analysis confirmed the chromosome degradation observed by flow cytometry (Figures 4C and 4E).

In conclusion, a single round of infection occurred in cultures when a high m.o.i. is used, and the SIRV2 latent period was determined to 9 h under the conditions employed. Host chromosome degradation mainly occurred late during the infection cycle.

Kinetics of cell mortality after viral infection

Infected cells might remain metabolically active for some time despite a degraded genome. To determine when cell death exactly occurs, the membrane potential-sensitive dye $DiBac(4)_3$ was used. The dye penetrates depolarized cells and binds to intracellular proteins and lipids, causing enhanced fluorescence.

The ratio of fluorescence/OD in stained culture samples of constant volume was measured over time in uninfected and infected (m.o.i. about 3) cultures. The high ratios obtained at the initial time points in both controls and infected cultures (Figure 5) were due to the low initial OD, and not to a high proportion of depolarized cells (Figure S3). In infected cultures only, a moderate peak was observed at 11 h (Figure 5), corresponding to

the middle of the virus release period as assessed by hybridization (Figure 5). Surprisingly, the higher ratios in infected cultures were transitory, despite the persistence of chromosome-less cells as shown by flow cytometry. It is possible that intracellular macromolecules were degraded shortly after virion release, such that the fluorescence signal emitted by dead cells decreased strongly.

The results suggest that the cells died concomitant with virus release, and that damage was not restricted to chromosome degradation. Thus, SIRV2 is a lytic virus that kills the host cell during the process of virus production and release.

One – step growth curve

A one-step growth experiment was performed to confirm the latent period and estimate the burst size. A low m.o.i. was employed, such that the effect of infection by a single virus could be monitored. Virus release was shown to begin at around 8 - 10 h post-infection (Figure 6), consistent with all other data. Although the precision of the PFU assay is limited, it clearly showed that virus production was significantly reduced after 13 – 15 h. The burst size was estimated to 30±10 viruses per cell, consistent with the estimates of SIRV2 intracellular DNA by flow cytometry (above). Thus, the one-step growth curve corresponded well to a lytic SIRV2 life cycle.

Observation of novel cellular ultrastructures induced by SIRV2 infection

Ultrathin sections of samples fixed at various times after infection (m.o.i. about 1) were analyzed by transmission electron microscopy (TEM). The cells were fixed at 10 h (just before virus release), 13 h (middle of release period) and 26 h (several hours after release). Uninfected cells in mid-exponential growth phase were used as control.

The morphology of uninfected cells was as expected for Sulfolobales species, with a cell envelope consisting of the lipid membrane and an S-layer (Figure 1A). They however presented a distorted shape, which may have been caused by the fixation and dehydration procedure. At 10 h post-infection, dense aggregates of virions were visible within the infected cells (Figure 7F), showing that assembly occurred in the cytoplasm. Up to 3 densely packed aggregates, together containing up to about 150 virions, were detected in the cell sections and occupied a high fraction of the intracellular volume. Remarkably, large angular protrusions associated with a local absence of S-layer were reproducibly observed at the cell envelope (Figure7G and 7H); several such structures were usually visible per cell section (Figure 7G, arrows), and were never observed in uninfected cells. The base of the virus-induced protrusions (VIPs) was polygonal, as shown by a section in the perpendicular plane (Figure 7I), and the overall structure must, thus, have been pyramidal. The VIPs often contained localized regions producing a denser staining, at the tip of the pyramidal structure (Figure 7G). At 13 h post-infection, both virion-containing unlysed cells (Figure 7F) and lysed, empty cells were present (Figure 7B). The lysed cells only contained virion remains (Figure 7B, arrow a) and collapsed cellular material appearing as dark aggregates. Also, a more regular oval shape than in intact cells was observed, indicating relaxation of the envelope. Furthermore, large (about 200 nm diameter) openings were observed in the membrane, up to 4 per cell section, delimited by C-shaped structures (Figure 7B, unlabeled arrows). Apart from the openings, the cell envelope appeared to be intact, with both the S-

layer and the membrane visible (Figure 7E, respectively arrows b and c). After 26 h, almost all cells were empty (Figure 7B). Noticeably, well-preserved C-shaped structures (Figure 7C), as well as polygonal rings (Figure 7D), that were detached from the cell envelope, could be detected. These observations suggest that the VIPs gave rise to the openings in the cell envelope for virion extrusion, and that specific modifications of the cell envelope were involved in the formation of these structures and in subsequent perforation of the envelope.

Given the high number of virions observed in the cytoplasm, the flow cytometry and one-step growth experiments (above) appeared to have resulted in an underestimation of the burst size. It is possible that viral DNA becomes refractory to staining when coated by protein and densely packed into the virion, reducing the fluorescence signals in the flow cytometry analysis. In the one-step experiments, virions may still have been aggregated after release, and it is also possible that a proportion of the virions could be uninfectious. Furthermore, the efficiency of the SIRV2 plaque assay is unknown.

The TEM analyzes demonstrated that SIRV2 virions were assembled in the cytoplasm, and confirmed that cell lysis and death occurred at the end of the infection cycle. Since ongoing virus release could not be detected, this must have occurred within a brief time interval, and it is likely that the intracellular macromolecular content simultaneously was degraded. Remarkably, the TEM studies also showed that SIRV2 infection induces specific cellular structures of pyramidal shape, VIPs, pointing outwards from the cell envelope. Later in the infection cycle, the VIPs appeared to give rise to openings delimited by rings of C-shaped section that allowed virion release.

Discussion

We report the first detailed cellular study of the infection cycle of a crenarchaeal virus. We demonstrate that SIRV2 is lytic and that the virions are assembled in the cytoplasm and released through well-defined openings in the cell envelope. Remarkably, these are produced through angular protrusions at the membrane level, VIPs, induced by the virus, which leave behind ring structures of C-shaped section in the envelope of lysed cells. We also visualize intracellular viral DNA by flow cytometry for the first time, and show that host chromosome DNA is completely degraded during the viral infection cycle. Together, the results demonstrate that the cells die as the consequence of specific and unique mechanisms orchestrated by the virus, rather than from general deleterious effects of the infection.

Recently, Claverie (2006) suggested that emphasis should be put on viral factories, and not the virion, in defining viruses as organisms. However, until now, viral factories were only detected in viruses infecting eukaryotic cells (e.g. Fontana et al., 2008). The data presented here show that SIRV2 transforms the whole cellular compartment into an elaborated viral factory, in agreement with the idea that viruses should be considered as a specific type of living organisms (Raoult and Forterre, 2008).

A unique virus release mechanism

To our knowledge, the virus release mechanism identified here is unprecedented in virus biology. In lytic bacteriophages, the two main lysis strategies rely on the direct degradation of peptidoglycan, e.g. with the holin-endolysin system (Wang et al., 2000), or on the inhibition of cell wall synthesis (Bernhardt et al., 2002). Both strategies result in complete cell disruption, and do not involve a modification of the cell envelope in several localized regions, as reported here. In eukaryotes, modification of membrane structures occurs with certain RNA and DNA viruses (Miller and Krijnse-Locker, 2008). However, this concerns intracellular membranes (endoplasmic reticulum, Golgi complex) and is possibly linked to viral replication rather than release (Miller and Krijnse-Locker, 2008). Involvement of the VIPs in cell lysis is thus a unique feature, and specific structural modifications of the cell envelope are most likely implicated in their formation.

Prediction of functions to be encoded in SIRV2 genome

The increased knowledge about the SIRV2 infection cycle generates new hypotheses concerning functions encoded within the genome. Since the virus is not present in a carrier state, no specific mechanisms aiming at preserving host cell viability and integrity are expected. In particular, the mechanisms for virion assembly may be quite simple. It is likely that a set of SIRV2 genes must control the formation of the VIPs and the generation of the openings through which the virions are released. The genes might either directly code for the proteins involved, or modulate host-encoded mechanisms. The timing of virus release must also be controlled by virus-encoded functions, such that lysis does not occur until the virions have been assembled, as for any lytic virus. Also, since host chromosome degradation accompanies the infection cycle, a virus-encoded mechanism may actively trigger the degradation, thereby both neutralizing the host and increasing the supply of nucleotides for

viral replication. Such mechanisms have previously been observed in bacteriophages of the T family (Souther et al., 1972; Weed and Cohen, 1951).

Lytic properties and morphotypes, a comparison with bacteriophages

For SIRV2, virion assembly in the cytoplasm is in accordance with the ability of the major coat protein to self-assemble in the absence of an energy source (Vestergaard et al., 2008). Consequently, complete virions have to cross the cell envelope during release, which may require lysis. The existence of a lytic crenarchaeal virus of linear morphotype contrasts to the correlation between morphotype and lifestyle in bacterial viruses. In these, the correspondence between a linear morphotype and a chronic infection mode is systematic (*Inoviridae* familiy, Calendar and Abedon, 2006), and cell integrity is preserved by assembling virions at the membrane and progressively extruding these. In contrast, phages with more compact nucleic acid packaging structures (head-and-tail, spherical, etc.) exhibit lytic properties, unless they are enveloped (Calendar and Abedon, 2006). It is, thus; possible that difficulties in releasing particles of large diameter while preserving cell envelope integrity favor lytic life cycles. Surprisingly, for SIRV2, the linear morphotype is not exploited to conserve the host cell, supporting a model in which an intracellular refuge is not essential for viruses in hyperthermophilic environments.

Lytic viruses in extreme hot environments

Currently, four viruses representative of different crenarchaeal virus families have been described as lytic: SIRV2 (*Rudiviridae*), *Acidianus* two-tailed virus (*Bicaudaviridae*, Prangishvili et al., 2006b), *Thermoproteus* tenax virus 1 (*Lipothrixviridae*, Janekovic et al., 1983), and *Sulfolobus* turreted icosahedral virus (unclassified, Ortmann et al., 2008). The fact that these diverse viruses are lytic is significant and indicates that this mode of infection is not as rare in hyperthermophilic environments as previously assumed. At the methods level, it should be noted that empty Sulfolobales cells still give rise to light scatter, and that the flow cytometry light scatter parameter and OD measurements, consequently, not are reliable indicators for deducing lytic properties. It is, thus, possible that lytic properties have been overlooked in other hyperthermophilic viruses.

At the evolutionary level, the findings imply that virus particles can persist in extreme ecosystems for long enough to find a new host. For SIRV2, the virions are well adapted to harsh conditions, being extremely stable in various solvents (Steinmetz et al., 2008; Zillig et al., 1994), and almost as stable at 80°C (unpublished data) as viruses infecting mesophilic bacteria are at 37°C (De Paepe and Taddei, 2006). Also, SIRV2 virions stored under mild conditions are exceptionally stable (Steinmetz et al., 2008; Zillig et al., 1994). Geothermal environments are extremely heterogeneous, due to a variety of gradients, dynamic movements and changes over time, and viruses could be trapped and preserved for long time periods in different environmental refuges in the absence of potential hosts. Finally, the fact that virus particles apparently are able to travel across the globe (Breitbart et al., 2004; Sano et al., 2004; Snyder et al., 2007) also suggests that they are robust to variable environmental conditions and should be stable over evolutionary times in a given location.

Experimental procedures

Strains, virus, and cultures

S. islandicus LAL14/1 was used. The strains *S. islandicus* KVEM10H3 and HVE10/4 were used in addition for one specific experiment. Virus stocks were prepared by polyethylene glycol precipitation of the virus from the culture supernatants, followed by concentration and purification on Cesium chloride density gradients, as described in Bettstetter et al., 2003.

In most cases, the cells were grown in shaken 50 mL flasks at 78°C, pH 3.0, in rich medium as previously described (Zillig et al., 1994). For the growth of minicultures, 24-well flat-bottomed microplates (Corning-Costar) were used. The microplates were incubated at 78°C and shaken at 150 rpm (Infors dry incubator) in a humid atmosphere. Colonies were obtained on Gellan gum plates as described in Zillig et al., 1994. To infect cultures, the appropriate volume of virus solution was dialyzed against medium or water on 0.025 μm or 0.05 μm MF membrane filters (Millipore) and directly added to the liquid cultures during the early exponential phase (OD_{600nm} between 0.09 and 0.25). For the time-course experiments (growth kinetics, flow cytometry, dot blot hybridization, $DiBac_4(3)$ staining), all conditions were tested in triplicates. Six identical 50 mL cultures were launched by dilution of a same pre-culture. After overnight growth, SIRV2 was added to 3 of them at the appropriate m.o.i..

Titrations, OD and fluorescence measurements

To determine CFU titers, culture samples were submitted to serial dilutions and 5 μL of each dilution were spotted on plates. After incubation, the colonies were counted in the last or last two positive spots.

To determine the PFU titers in cultures, culture supernatants, or virus stocks, the same method was used, except that 5 μL of each dilution were spotted on a fresh cell lawn. When required, the cells were removed by centrifugation before spotting. The cell lawns were prepared as described in Prangishvili et al., 1999, using a soft Gellan gum overlay. After incubation, single plaques were counted in the last or last two positive spots.

ODs were measured in 96-well round-bottomed culture microplates (TPP) in a Multiskan Ascent microplate photometer (Thermo LabSystems) at 595 nm using 200 μL of culture. For the experiment with $DiBac(4)_3$ only, ODs were measured at 600 nm in the Infinite M200 microplate reader (Tecan), as it was suitable for both OD and fluorescence measurements. The values acquired with the 2 apparatus cannot directly be compared.

For the fluorescence measurements, the excitation wavelength was 490 nm, the emission wavelength 550 nm, and the manual gain was set to 120.

Primers and PCRs

The 16S rRNA gene of *S. islandicus* LAL14/1 was amplified using primers 8aF and 1512uR (Eder et al., 1999). These primers were used both for sequencing and for generating a probe for the dot blot hybridization experiments (see below).

Six primer pairs specific for SIRV2 genome were used (each time, the 5' position in SIRV2 genome is indicated). The pair [5'-ACATGAAAAGTTAGAGAGATACAAACG

(3872); 5'-TGGTTACCACTAGCTTCGCTAC (4086)] was used to generate a probe for the dot blot hybridization experiments (see below).

The five primer pairs:
[5´-GAGACGCAAAAGCAAGTCAAA (1707); 5´-TCCACTTCCCATATCTCCGA (2221)],
[5´-AAAACAAATTGGAATAACGCC (10071); 5´-CTGCAATTTCATGTGGATCA (10616)],
[5´-GAAATGCGGAAAGCGAGTTA (22972); 5´-TGGAGTTCTTGTGGAATCTGG (23482)],
[5´-ATCTCCTGGCAAAATGGGA (27018); 5´-GCGGGCTTTATTGGAACTG (27479)],
[5´-CTTCCATCTTGAACCCCTGA (32635); 5´-GAACTTGCGGAATTTTACGGA (33183)],
were used for the detection, by PCR, of SIRV2 genome potentially present inside isolated clones. The 5 primer pairs were used for each clone. As a positive control, 0.5 μL of SIRV2 stock were used as template with the 5 primer pairs. As a negative control, some clones isolated from uninfected *S. islandicus* LAL14/1 cultures were each time included and used with the 5 primer pairs. Finally, as a control for suitable amount and quality of the template, the 16S rRNA primer pair was systematically used on all tested clones.

Controls for the non contamination of the strain
(Supplemental Text)

Flow cytometry

Sampling and flow cytometry were performed as described in Bernander and Poplawski, 1997: the cells were fixed in 70% (v/v) ethanol and the intracellular DNA was stained with mithramycin A and ethidium bromide. Samples were analyzed in an A40 Analyzer (Apogee, 25 mW solid-state laser, 405 nm wavelength). *S. islandicus* cell cycle was characterized preliminarily to the study of infected cultures (Figures S1 and S2).

For the study of infected cultures, a high m.o.i. was used (~10-15) to obtain as synchronous an infection as possible. It was at the risk of observing effects of a modified or enhanced nature compared to a m.o.i closer to 1. At each time point, OD_{595nm} was measured and CFU titers were determined to control that the usual growth pattern was obtained.

The cell populations were identified based on the cell size distributions, DNA content distributions and 2D diagrams of cell size and DNA content. The data were gated, and several contours tested, to ensure the robustness of the analysis and of the identified cell populations. The proportion of empty cells over time was computed by gating the 2D-diagrams, similar to what is shown in Figure 3D. In Figure 4A, the total percentage of chromosome-less cells in the culture is shown. In Figure 4B, for the curve related to chromosome-less cells, the difference between the values at time points T and T-1 is plotted instead of simply the value at time point T, reflecting more the production than the accumulation of empty cells.

Dot blot hybridization

Cells were washed once in cold medium, pelleted by low-speed centrifugation and stored at -20°C until further use. Cell pellets were re-suspended in Tris-acetate pH 6.0 pre-cooled at 4°C. The suspension volume was adjusted for cell concentration to be roughly constant in all samples, on the basis of OD measurements. 4 μL of each sample were spotted on Hybond-N+ nylon membranes (Amersham Biosciences). The membranes were further prepared as for colony hybridization (Sambrook et al., 1989).

The probes were generated by PCR. A ~240 bp SIRV2 specific probe and a ~1300 bp 16S rRNA gene specific probe were used (see above). The choice of the 16S rRNA gene was motivated by the absence of the host genome sequence. The probes were [^{32}P]-end-labelled with EasyTide [α-^{32}P]-dATP (PerkinElmer) using a random-primed DNA labeling Kit (Roche Applied Science), according to manufacturer's instructions.

All hybridization steps were performed at 65°C in pre-warmed solutions. After a minimum of 2 h prehybridization followed by overnight hybridization, both performed in Church Buffer (7% SDS (w/v), 0.5M sodium phosphate pH 7.2, 1mM EDTA), membranes were washed 2 times 15 min in a solution of 2X SSC- 0.1% SDS, and 2 times 15 min in a solution of 0.5X SSC - 0.1% SDS.

Membranes were exposed on a GP Phosphor Screen (Amersham Biosciences). The screen was scanned in a Molecular Dynamics Storm 860 (Amersham Biosciences). The images were analyzed with the ImageQuantTL software (Amersham Biosciences). After contrast and brightness adjustment, the radioactivity of each spot on the membranes was quantified, using the background removal option (local average). The images of Figure 4D and 4E were processed with ImageQuantTL for contrast and brightness adjustment and with ImageJ software (http://rsbweb.nih.gov/ij/) for background removal, using the "sliding paraboloid" function.

DiBac(4)$_3$ staining

DiBac(4)$_3$ was selected based on Beck and Huber, 1997 and purchased from Molecular Probes. It was verified by fluorescence microscopy (video inverted microscope Leica DM IRE2 HC FLUO, Metamorph software) that the growth medium was suitable for the probe, not causing any background nor suppressing the cell fluorescence; it was also verified that fluorescence from dead cells was significantly stronger than that of living cells, and that unstained cells were not auto-fluorescent.

Afterwards, the changes in cell mortality were assessed by measuring the ratio fluorescence/OD$_{600nm}$ of constant-volume culture samples. Culture samples of 1.6 mL volume were allowed to cool down to room temperature for 10 min. 1 μL of DiBac(4)$_3$ in ethanol (0.5 mg/mL) was added to each and the samples were incubated for 5 min in the dark. 200 μL of each sample were transferred into wells of a 96-well round-bottomed culture microplate (TPP) for OD and fluorescence measurements (see above). The ratio fluorescence/OD$_{600nm}$ is strongly OD-dependant (Figure S3); but for a given OD, the ratio is higher for higher rates of dead cells, so that the method is convenient.

One-step growth curve

The virus was added to 1 mL *S. islandicus* LAL14/10 culture in early exponential phase at a low m.o.i. (10^{-3}). After a 1.5 h-2 h incubation at 78°C for adsorption of almost all

viruses, the preparation was diluted in 200 mL pre-warmed medium to prevent the occurrence of new adsorptions in the course of the experiment. The obtained diluted culture was further incubated at 78°C. PFU titers of samples were regularly determined.

The burst size was estimated by averaging the PFU titers from time points 0 h-9 h (before release) and from time points 13 h-17 h (after virus release), and then by dividing the latter value by the former one.

Transmission Electron Microscopy

Cells were pelleted by low speed centrifugation. The cell pellet was fixed overnight at 4°C with 2.5% (w/v) glutaraldehyde in 20 mM Tris-acetate pH 6 buffer, post-fixed for 1 h in 1% (w/v) OsO_4, and dehydrated in a graded series of ethanol solutions (25% to 100%). The cells were embedded in an epoxy resin and polymerized at 60°C for 48 hours. Ultrathin sections (~60 nm) were cut on a Leica Ultratuc UCT microtome and deposited on carbon-coated copper grids. They were stained for 30 min with 2% (w/v) uranyl acetate and for 5 min with 2.5% (w/v) lead citrate.

The grids were examined under a JEOL JEM-1010 transmission electron microscope (Tokyo, Japan) operated at 80 kV. Images were recorded using an Eloise Keen View camera and the Analysis Pro software version 3.1 (Eloïse SARL, Roissy, France).

Acknowledgements

We are indebted to Dr Soizick Lucas-Staat for great help in virus preparation, to Léa Lepelletier for assistance in optical microscopy and to Professor Ryland F. Young for enlightening discussions. This work was supported by the Swedish Research Council and by a strong research environment grant (Uppsala Microbiomics Center; www.microbiomics.se) from the Swedish Research Council Formas.

References

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Beck, P., and Huber, R. (1997). Detection of cell viability in cultures of hyperthermophiles. FEMS Microbiology Letters *48*, 11-14.

Bernander, R. (2007). The cell cycle of Sulfolobus. Mol Microbiol *66*, 557-562.

Bernander, R., and Poplawski, A. (1997). Cell cycle characteristics of thermophilic archaea. J Bacteriol *179*, 4963-4969.

Bernhardt, T. G., Wang, I. N., Struck, D. K., and Young, R. (2002). Breaking free: "protein antibiotics" and phage lysis. Res Microbiol *153*, 493-501.

Bettstetter, M., Peng, X., Garrett, R. A., and Prangishvili, D. (2003). AFV1, a novel virus infecting hyperthermophilic archaea of the genus acidianus. Virology *315*, 68-79.

Birkenbihl, R. P., Neef, K., Prangishvili, D., and Kemper, B. (2001). Holliday junction resolving enzymes of archaeal viruses SIRV1 and SIRV2. J Mol Biol *309*, 1067-1076.

Breitbart, M., Miyake, J. H., and Rohwer, F. (2004). Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol Lett *236*, 249-256.

Calendar, R., and Abedon, S. T., eds. (2006). The Bacteriophages, 2d ed (New York: Oxford University Press).

Claverie, J. M. (2006). Viruses take center stage in cellular evolution. Genome Biol *7*, 110.

Contursi, P., Jensen, S., Aucelli, T., Rossi, M., Bartolucci, S., and She, Q. (2006). Characterization of the Sulfolobus host-SSV2 virus interaction. Extremophiles *10*, 615-627.

De Paepe, M., and Taddei, F. (2006). Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. PLoS Biol *4*, e193.

Duggin, I. G., and Bell, S. D. (2006). The chromosome replication machinery of the archaeon Sulfolobus solfataricus. J Biol Chem *281*, 15029-15032.

Eder, W., Ludwig, W., and Huber, R. (1999). Novel 16S rRNA gene sequences retrieved from highly saline brine sediments of kebrit deep, red Sea. Arch Microbiol *172*, 213-218.

Fontana, J., Lopez-Montero, N., Elliott, R. M., Fernandez, J. J., and Risco, C. (2008). The unique architecture of Bunyamwera virus factories around the Golgi complex. Cell Microbiol *10*, 2012-2028.

Frols, S., Gordon, P. M., Panlilio, M. A., Schleper, C., and Sensen, C. W. (2007). Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. Virology *365*, 48-59.

Janekovic, D., Wunderl, S., Holz, I., Zillig, W., Gierl, A., and Neumann, H. (1983). TTV1, TTV2 and TTV3, a family of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaebacterium *Thermoproteus tenax*. Mol Gen Genet *192*, 39-45.

Keller, J., Leulliot, N., Cambillau, C., Campanacci, V., Porciero, S., Prangishvilli, D., Forterre, P., Cortez, D., Quevillon-Cheruel, S., and van Tilbeurgh, H. (2007). Crystal structure of AFV3-109, a highly conserved protein from crenarchaeal viruses. Virol J *4*, 12.

Kessler, A., Brinkman, A. B., van der Oost, J., and Prangishvili, D. (2004). Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon sulfolobus. J Bacteriol *186*, 7745-7753.

Larson, E. T., Eilers, B. J., Reiter, D., Ortmann, A. C., Young, M. J., and Lawrence, C. M. (2007). A new DNA binding protein highly conserved in diverse crenarchaeal viruses. Virology *363*, 387-396.

Miller, S., and Krijnse-Locker, J. (2008). Modification of intracellular membrane structures for virus replication. Nat Rev Microbiol *6*, 363-374.

Ortmann, A. C., Brumfield, S. K., Walther, J., McInnerney, K., Brouns, S. J., van de Werken, H. J., Bothner, B., Douglas, T., van de Oost, J., and Young, M. J. (2008). Transcriptome analysis of infection of the archaeon Sulfolobus solfataricus with Sulfolobus turreted icosahedral virus. J Virol *82*, 4874-4883.

Peng, X., Blum, H., She, Q., Mallok, S., Brugger, K., Garrett, R. A., Zillig, W., and Prangishvili, D. (2001). Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. Virology *291*, 226-234.

Prangishvili, D., Arnold, H. P., Gotz, D., Ziese, U., Holz, I., Kristjansson, J. K., and Zillig, W. (1999). A novel virus family, the Rudiviridae: Structure, virus-host interactions and genome variability of the sulfolobus viruses SIRV1 and SIRV2. Genetics *152*, 1387-1396.

Prangishvili, D., Basta, T., and Garrett, R. A. (2008). Crenarchaeal viruses: morphotypes and genomes, In Encyclopedia of Virology, M. B. W. J., and V. R. M. H. V., eds. (Oxford: Elsevier), pp. 587-595.

Prangishvili, D., Forterre, P., and Garrett, R. A. (2006a). Viruses of the Archaea: a unifying view. Nat Rev Microbiol *4*, 837-848.

Prangishvili, D., and Garrett, R. A. (2005). Viruses of hyperthermophilic Crenarchaea. Trends Microbiol *13*, 535-542.

Prangishvili, D., Klenk, H. P., Jakobs, G., Schmiechen, A., Hanselmann, C., Holz, I., and Zillig, W. (1998). Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. J Biol Chem *273*, 6024-6029.

Prangishvili, D., Vestergaard, G., Haring, M., Aramayo, R., Basta, T., Rachel, R., and Garrett, R. A. (2006b). Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. J Mol Biol *359*, 1203-1216.

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

Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., McDermott, T., and Young, M. J. (2001). Viruses from extreme thermal environments. Proc Natl Acad Sci U S A *98*, 13341-13345.

Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., Johnson, J. E., Douglas, T., and Young, M. (2004). The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proc Natl Acad Sci U S A *101*, 7716-7720.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: A Laboratory Manual, 2d ed (New York: Cold Spring Harbor Laboratory Press).

Sano, E., Carlson, S., Wegley, L., and Rohwer, F. (2004). Movement of viruses between biomes. Appl Environ Microbiol *70*, 5842-5846.

Schleper, C., Kubo, K., and Zillig, W. (1992). The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. Proc Natl Acad Sci U S A *89*, 7645-7649.

Snyder, J. C., Stedman, K., Rice, G., Wiedenheft, B., Spuhler, J., and Young, M. J. (2003). Viruses of hyperthermophilic Archaea. Res Microbiol *154*, 474-482.

Snyder, J. C., Wiedenheft, B., Lavin, M., Roberto, F. F., Spuhler, J., Ortmann, A. C., Douglas, T., and Young, M. (2007). Virus movement maintains local virus population diversity. Proc Natl Acad Sci U S A *104*, 19102-19107.

Souther, A., Bruner, R., and Elliott, J. (1972). Degradation of Escherichia coli chromosome after infection by bacteriophage T4: role of bacteriophage gene D2a. J Virol *10*, 979-984.

Steinmetz, N. F., Bize, A., Findlay, K. C., Lomonossoff, G. P., Manchester, M., Evans, D. J., and Prangishvili, D. (2008). Site-specific and spatially controlled addressability of a new viral nanobuilding block: *Sulfolobus islandicus* rod-shaped virus 2. Adv Funct Mater *18*, 1-9.

Vestergaard, G., Haring, M., Peng, X., Rachel, R., Garrett, R. A., and Prangishvili, D. (2005). A novel rudivirus, ARV1, of the hyperthermophilic archaeal genus Acidianus. Virology *336*, 83-92.

Vestergaard, G., Shah, S. A., Bize, A., Reitberger, W., Reuter, M., Phan, H., Briegel, A., Rachel, R., Garrett, R. A., and Prangishvili, D. (2008). Stygiolobus rod-shaped virus and the interplay of crenarchaeal rudiviruses with the CRISPR antiviral system. J Bacteriol *190*, 6837-6845.

Wang, I. N., Smith, D. L., and Young, R. (2000). Holins: the protein clocks of bacteriophage infections. Annu Rev Microbiol *54*, 799-825.

Weed, L. L., and Cohen, S. S. (1951). The utilization of host pyrimidines in the synthesis of bacterial viruses. J Biol Chem *192*, 693-700.

Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., Lanzendörfer, M., and Kristjanson, J. K. (1994). Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. System Appl Microbiol *16*, 609-628.

Figures

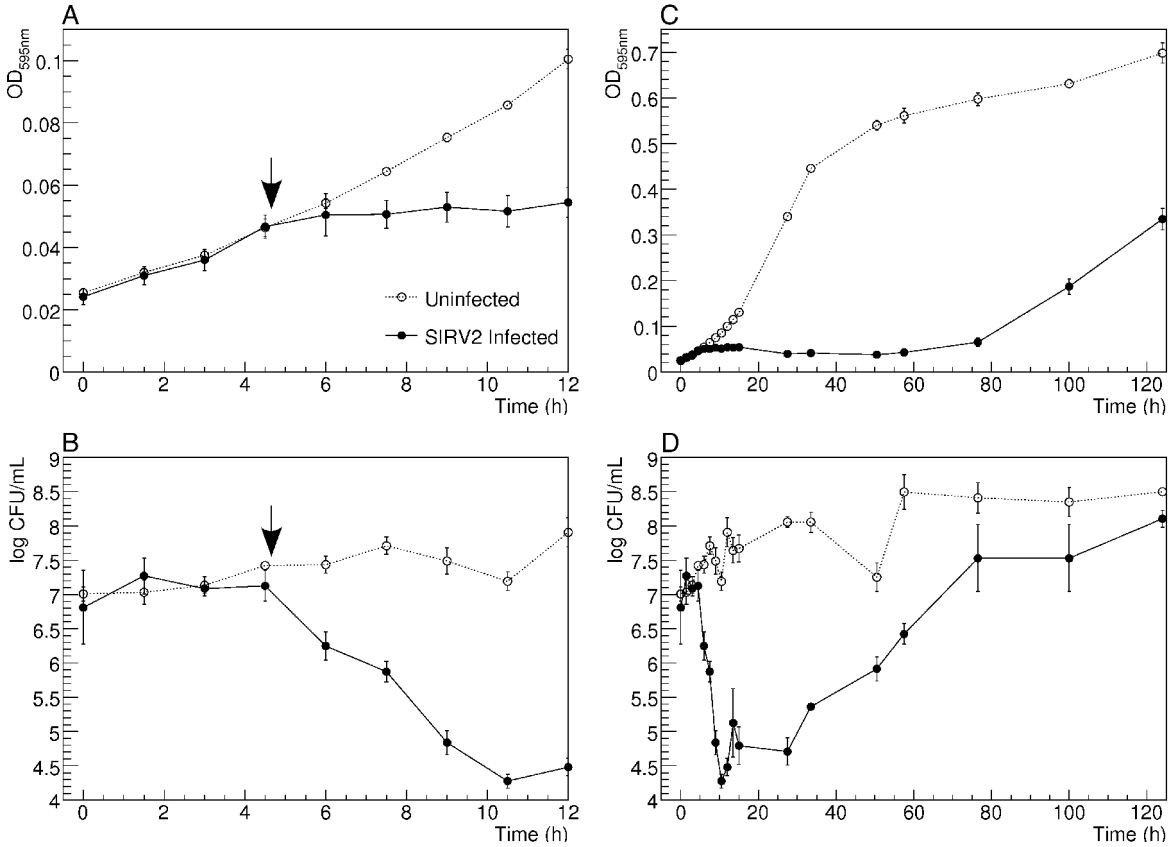


Figure 1. Impact of SIRV2 infection on the growth kinetics of *S. islandicus* cultures

Cultures infected at a m.o.i. of about 7 (filled circles, continuous line) and uninfected cultures (empty circles, dotted lines), were launched in triplicates. Averages of the replicates +/- one standard deviation (SD) are shown. The vertical arrow in (A) and (B) corresponds to virus addition, at 4.5 h.

(A): OD_{595nm}, detail of the first hours. (B): Log transformation of the CFU titers, detail of the first hours. (C): OD_{595nm} over the entire time-course. (D): Log transformation of the CFU titers over the entire time-course.

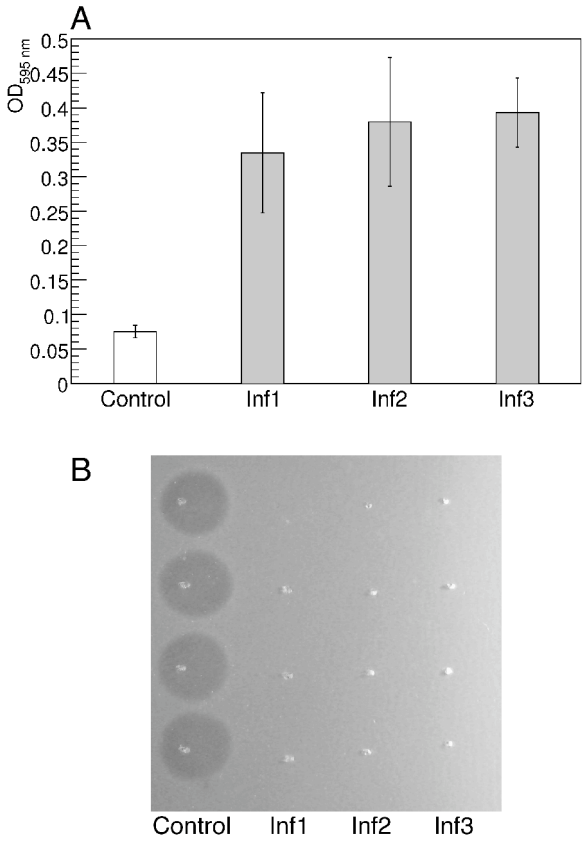
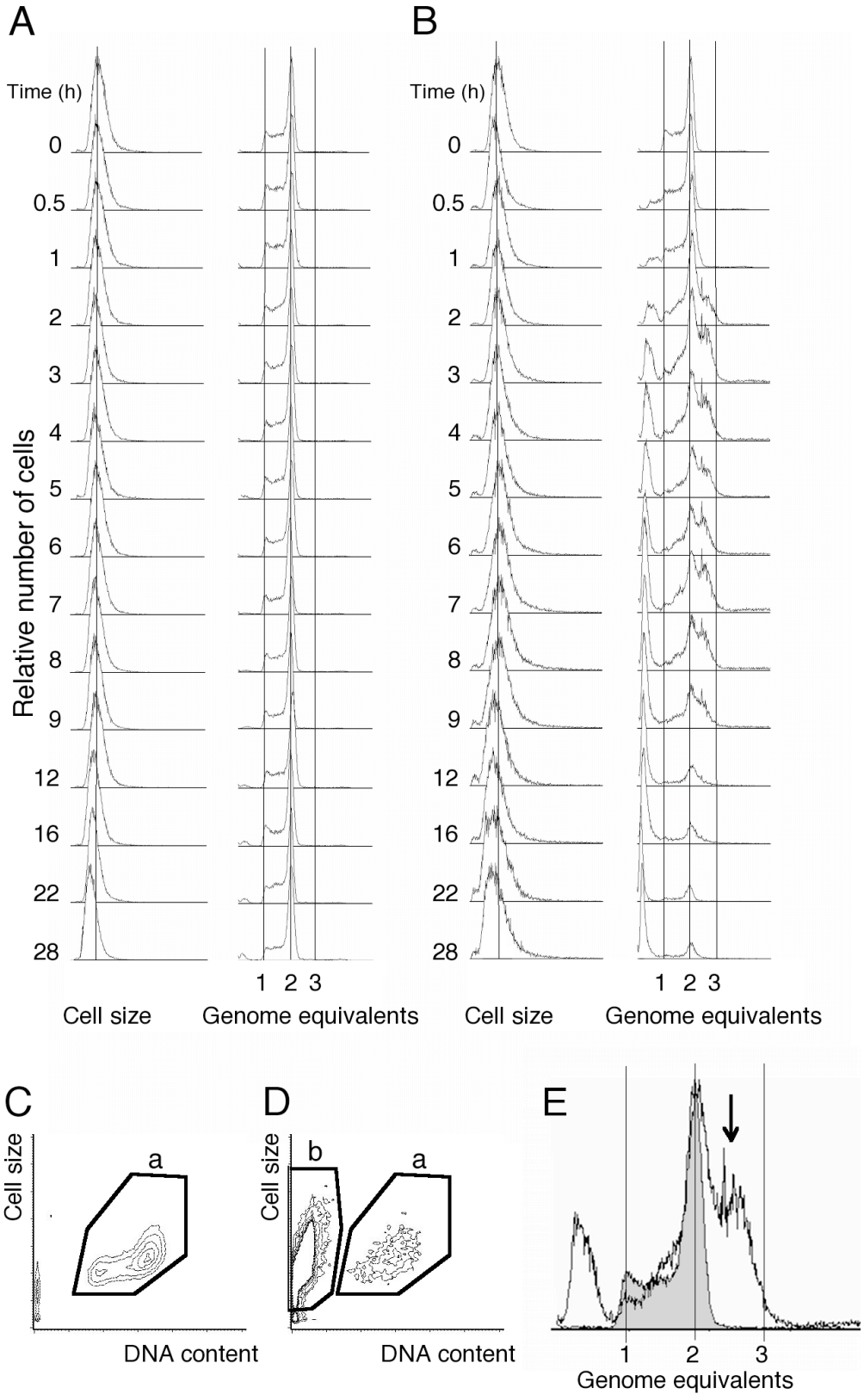


Figure 2. Sensitivity to SIRV2 infection of clones isolated from infected cultures
To obtain the clones, 23 colonies were isolated from an uninfected culture (control) and from each of 3 heavily infected cultures (Inf1, Inf2, Inf3), resulting in a total of 92 clones.
(A): OD_{595nm} measurements after a 3-days growth of the clones in the presence of SIRV2 added at a m.o.i. of about 10^{-3}. Averages corresponding to the 23 clones +/- one SD are shown.
(B): Spot test on a lawn of *S. islandicus* LAL14/10 cells of 20-fold diluted supernatants obtained after growth of the clones in presence of SIRV2 (see (A)). A clearing indicates the presence of infectious virions in the supernatant. The results for 4 out of 23 clones are shown for each culture; identical results were obtained for all clones.

Figure 3. Flow cytometry time-course analysis of *S. islandicus* cells infected by SIRV2
(A): Cell size and DNA content distributions from an uninfected culture.
(B): Cell size and DNA content distributions from a culture infected with SIRV2 (m.o.i. about 10). The virus was added at time point 0 h.
(C): Two-parameter distribution (cell size vs. DNA content) from an uninfected culture in early exponential phase. Contour ‘a’: intact cells (populations containing between 1 and 2 genome equivalents).
(D): Two-parameter distribution from a culture 11 hours after infection by SIRV2. Contour ‘a’: intact cells (populations containing between 1 and 2 genome equivalents). Contour ‘b’: DNA-less cells.
(E): Visualization of intracellular SIRV2 DNA by flow cytometry at 3 h post-infection.
The DNA content distribution from an infected *S. islandicus* culture is shown against the distribution from an uninfected culture (translucent grey). The arrow indicates the average amount of DNA in the infected cells.


A
B
Time (h)
0
0.5
1
2
3
4
5
6
7
8
9
12
16
22
28
Relative number of cells
1 2 3
Cell size
Genome equivalents
C
D
E
Cell size
a
b
DNA content
Genome equivalents

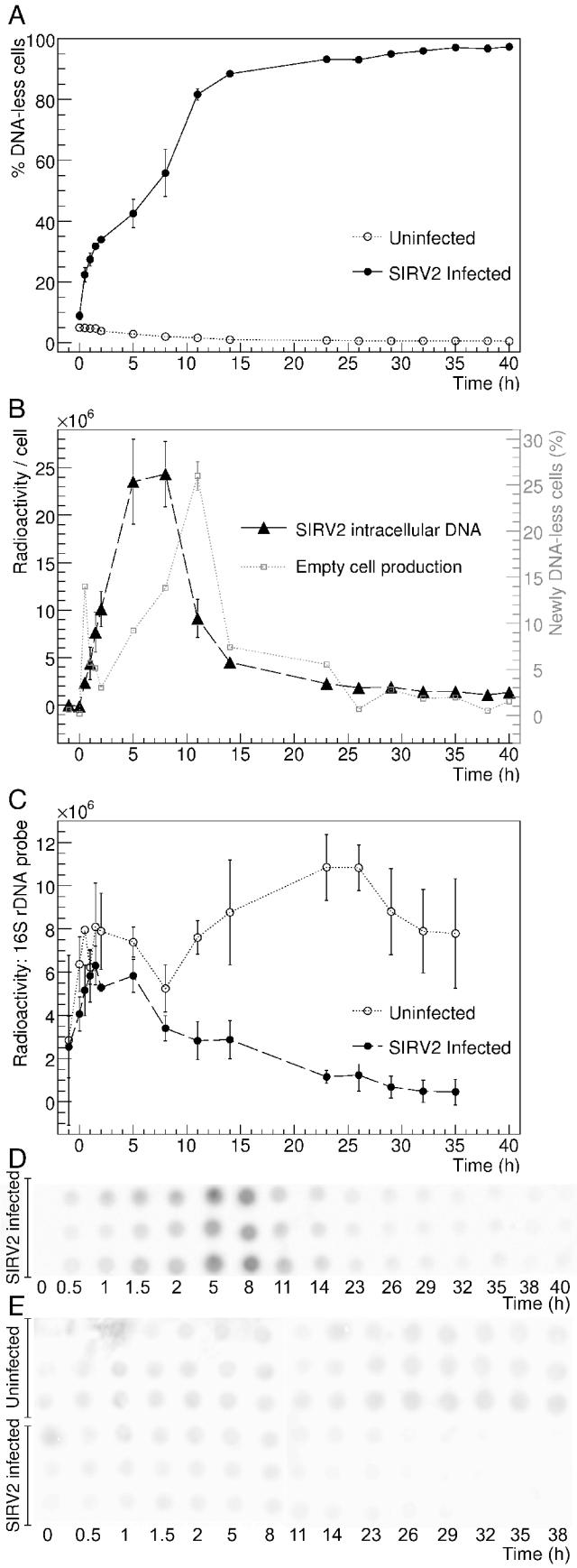

A
% DNA-less cells
Uninfected
SIRV2 Infected
Time (h)
B
×10^6
Radioactivity / cell
SIRV2 intracellular DNA
Empty cell production
Newly DNA-less cells (%)
Time (h)
C
×10^6
Radioactivity: 16S rDNA probe
Uninfected
SIRV2 Infected
Time (h)
D
SIRV2 infected
0 0.5 1 1.5 2 5 8 11 14 23 26 29 32 35 38 40
Time (h)
E
Uninfected
SIRV2 infected
0 0.5 1 1.5 2 5 8 11 14 23 26 29 32 35 38
Time (h)

Figure 4. Links between the kinetics of host chromosome degradation and the SIRV2 infection cycle
Infected cultures and uninfected cultures were launched in triplicates. SIRV2 was added in infected replicates (m.o.i. about 15) at time point 0 h. Averages of the 3 replicates +/- one SD are shown in (A), (B) and (C).
(A) Percentage of DNA-less cells in uninfected and infected cultures. Values were obtained by gating flow cytometry data as illustrated in Figure 3D.
(B) Radioactivity/cell (filled triangles, discontinuous line, left axis) indicative of SIRV2 intracellular DNA in infected cultures, over a time-course. Values in arbitrary units were obtained by quantifying the hybridization signal from each spot in the image shown in (D). The percentage of DNA-less cells appearing between 2 successive time points (empty circles, dotted line, right axis) was also plotted, based on the data from (A).
(C) Radioactivity/cell indicative of intracellular 16S rDNA amounts in uninfected cultures (empty circles, dotted line) or infected cultures (filled circles, discontinuous line). Values in arbitrary units were obtained by quantifying the hybridization signal from each spot in the image shown in (E).
(D) Autoradiogram of hybridization with a SIRV2 specific probe of cell spots from infected cultures. Each spot corresponds to the same approximate number of cells, based on OD measurements. The time course corresponds to horizontal lines, and the 3 replicates are shown vertically for each time point.
(E): Autoradiogram of hybridization with a 16S rRNA gene-specific probe of cell spots from uninfected and infected cultures. See (D) for additional explanations.

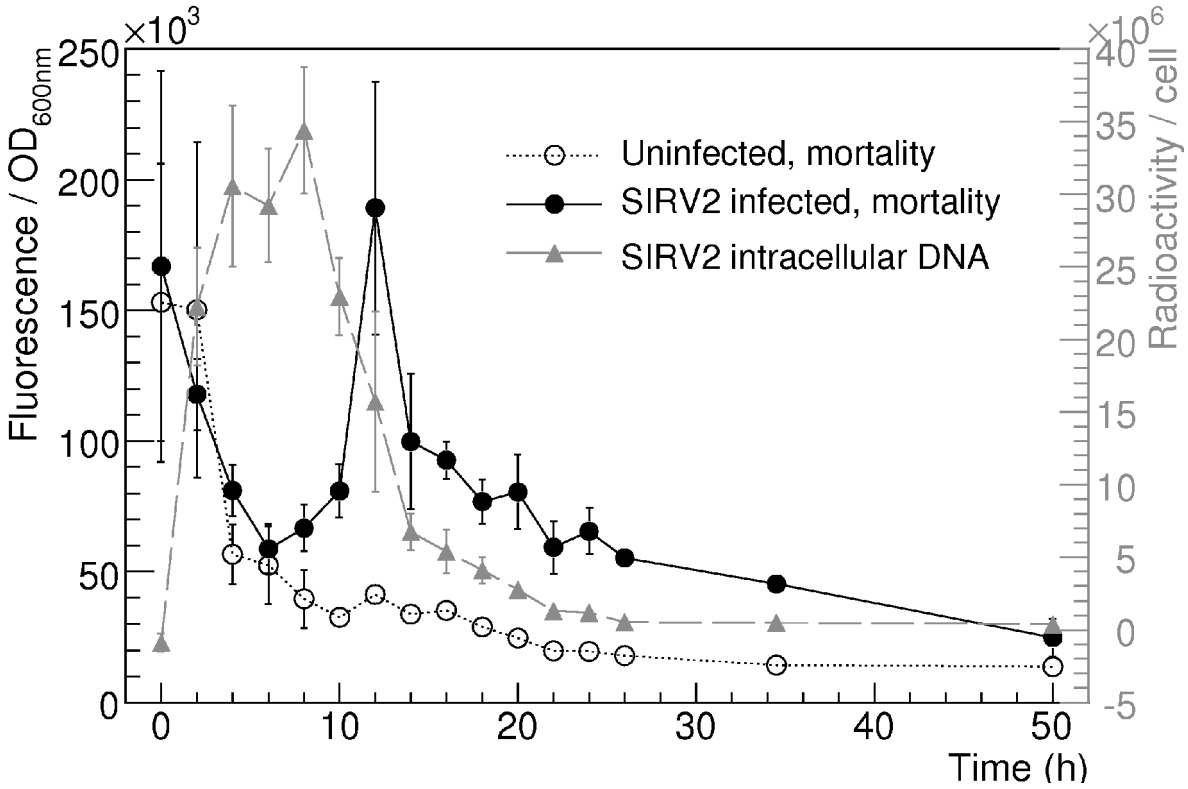


Figure 5. Impact of SIRV2 infection on cell mortality
The ratios of fluorescence/OD_{600nm} for infected (filled circles, continuous line) and uninfected (empty circles, dotted line) cultures reflect the changes in cell mortality. For the infected culture, the radioactivity/cell (filled triangles, discontinuous line, right axis) indicative of SIRV2 intracellular DNA amount was also plotted. Infected (m.o.i. about 5) and uninfected cultures were launched in triplicates, and averages of the replicates +/- one SD are shown. In infected replicates, the virus was added at time point 0 h.

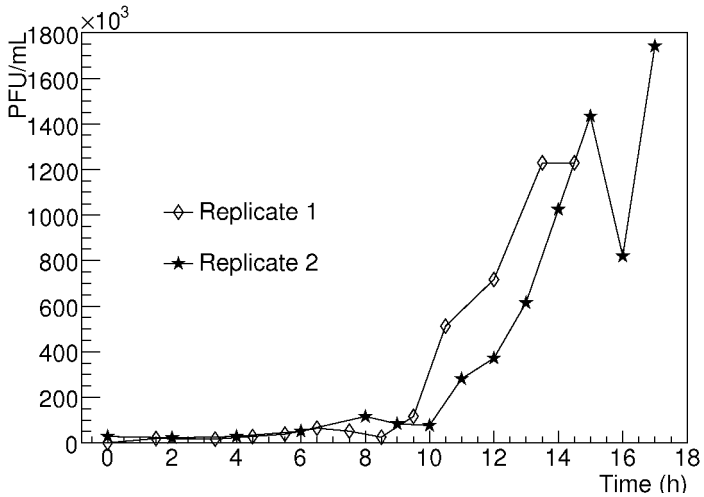


Figure 6. One-step growth curve of SIRV2 infection of *S. islandicus* cells

The virus was added at a m.o.i. of about 10^{-3}, and the experiment repeated twice. PFU = Plaque Forming Units.

Figure 7. Specific structures detected by TEM on thin sections from infected cells

(A): Uninfected cell.

(B): Lysed infected cell. Unlabeled arrows: openings in the cell membrane. Arrow a: example of virion remains.

(C): Section of a detached polygonal ring delimiting the openings, presenting a C-shaped aspect (see (D)).

(D): Polygonal detached ring, corresponding to the structure delimiting the openings.

(E): Detail of membrane openings in lysed cells (arrows a). The S-layer (arrow b) and the membrane (arrow c) are still visible in the remaining cell envelope.

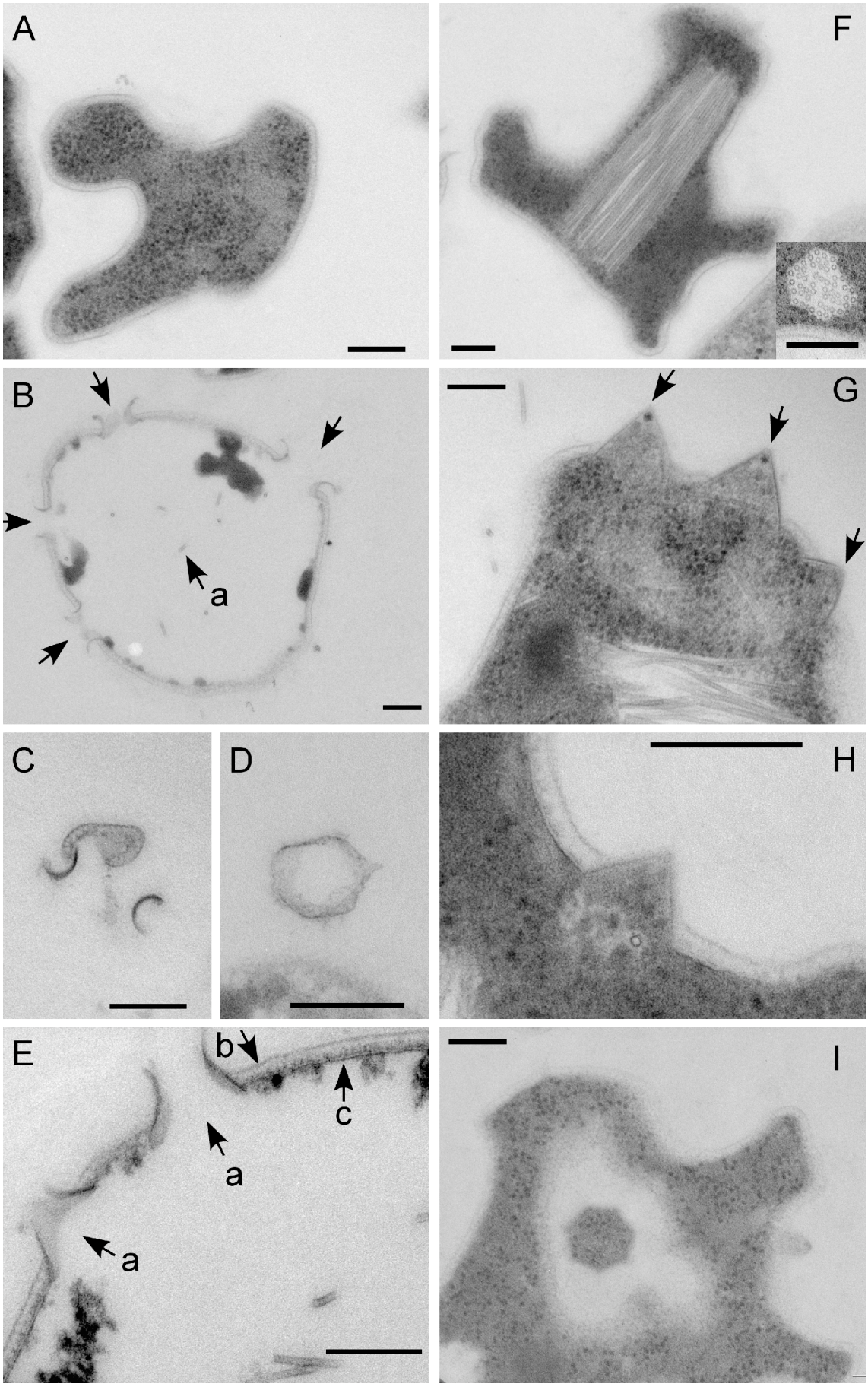
(F): Aggregate of virus particles in the cytoplasm of an infected cell. Longitudinal section plane in the main picture, perpendicular section plane in the insert.

(G): Infected cell presenting VIPs, indicated by arrows.

(H): Detail of VIP in an infected cell.

(I): Infected cell presenting a VIP sectioned in a perpendicular plane, demonstrating the polygonal base. The S-layer appears as a pale crystalline surface.

Bars = 200 nm.


A
F
B
a
G
C
D
H
E
b
c
a
a
I

Supporting Materials

Supplemental Text

Controls for the non contamination of the strain

Specific controls were designed to verify that the *S. islandicus* LAL14/1 strain stock was not contaminated by a distinct SIRV2 insensitive strain. It could have explained the growth of resistant cells in *S. islandicus* LAL14/1 cultures infected by SIRV2.

The 16S rRNA gene of a SIRV2 resistant clone isolated from an infected culture was sequenced and verified to be 100% identical to the sequence obtained from an uninfected culture, over the 1372 amplified bp (data not shown).

The EcoRI restriction patterns of genomes of 4 resistant clones isolated from infected cultures were verified to be identical to the pattern obtained using some uninfected *S. islandicus* LAL14/10 cells; as a reference, other closely related *Sulfolobus* strains were included and always produced distinguishable patterns (data not shown).

As a last control, the reference strain *S. islandicus* LAL14/1 was re-purified: 4 clones were randomly selected after 5 successive streaking steps. The OD_{595nm} of SIRV2 infected minicultures of the clones was monitored over time. It was verified that a growth was detectable after a few days (data not shown), indicating that re-purifying the strain had not eliminated the presence/appearance of SIRV2 resistant cells.

From these 3 controls, it was concluded that the resistant cells are variants of *S. islandicus* LAL14/1 strain.

Supplemental Figures

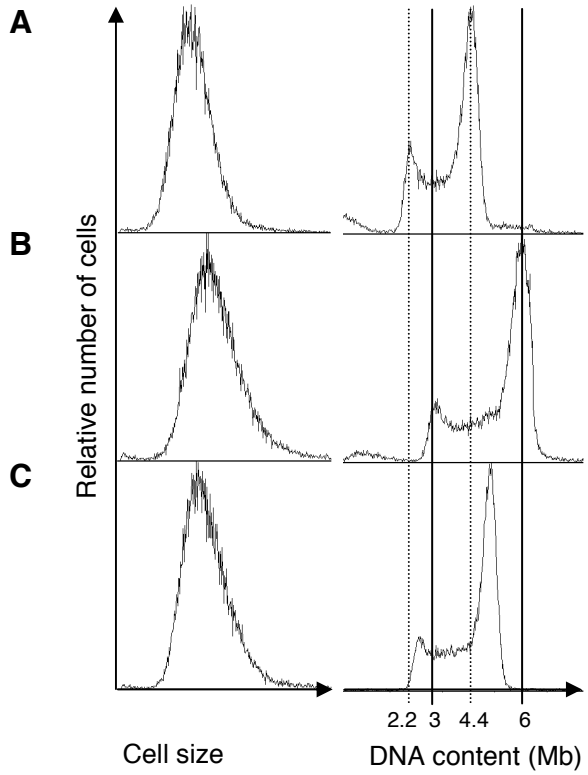


Figure S1. Flow cytometry distributions of DNA content and cell size for *S. acidocaldarius* (A), *S. solfataricus* (B), and *S. islandicus* LAL14/1 (C).
The vertical lines in the DNA content distributions correspond to the one- and two-chromosome peaks for *S. acidocaldarius* (dotted lines) and *S. solfataricus* (continuous lines). Cell cycles from *S. solfataricus* and *S. acidocaldarius* (Bernander and Poplawski, 1997) and their genome sizes (Chen et al., 2005; She et al., 2001) are known from previous studies. Assuming similar GC% and chromatin organisation among *Sulfolobus* species and by comparison of fluorescence signal strengths, the genome size of *S. islandicus* strain LAL14/1 was estimated to be 2.6Mb.

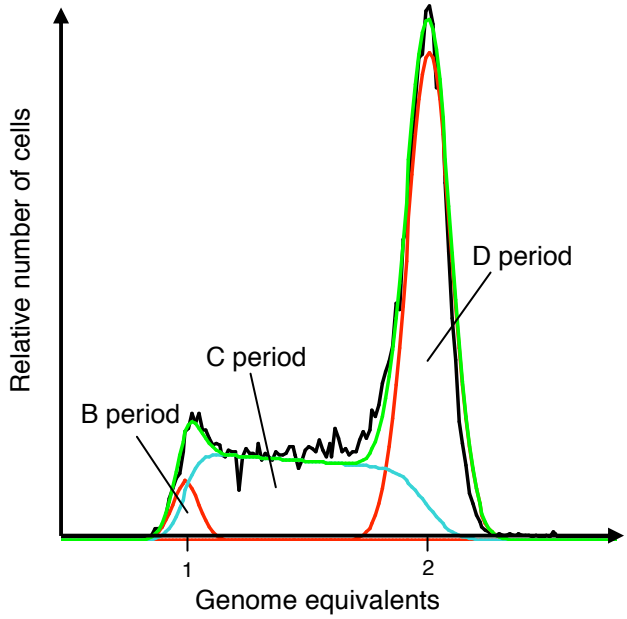


Figure S2. Simulation of the relative lengths of *S. islandicus* cell cycle periods.
The relative lengths of the different cell cycle periods were estimated based on DNA content distributions using the software Winflow (Flemming G. Hansen, The Technical University of Denmark, Lyngby, Denmark). The exponential age distribution was taken into account using formula 5A (Wold et al., 1994). The green line constitutes the combined simulated DNA distribution generated by the software.
B period: G_1, pre-replicative
C period: S, replicative
D period: G_2 + M, post-replicative

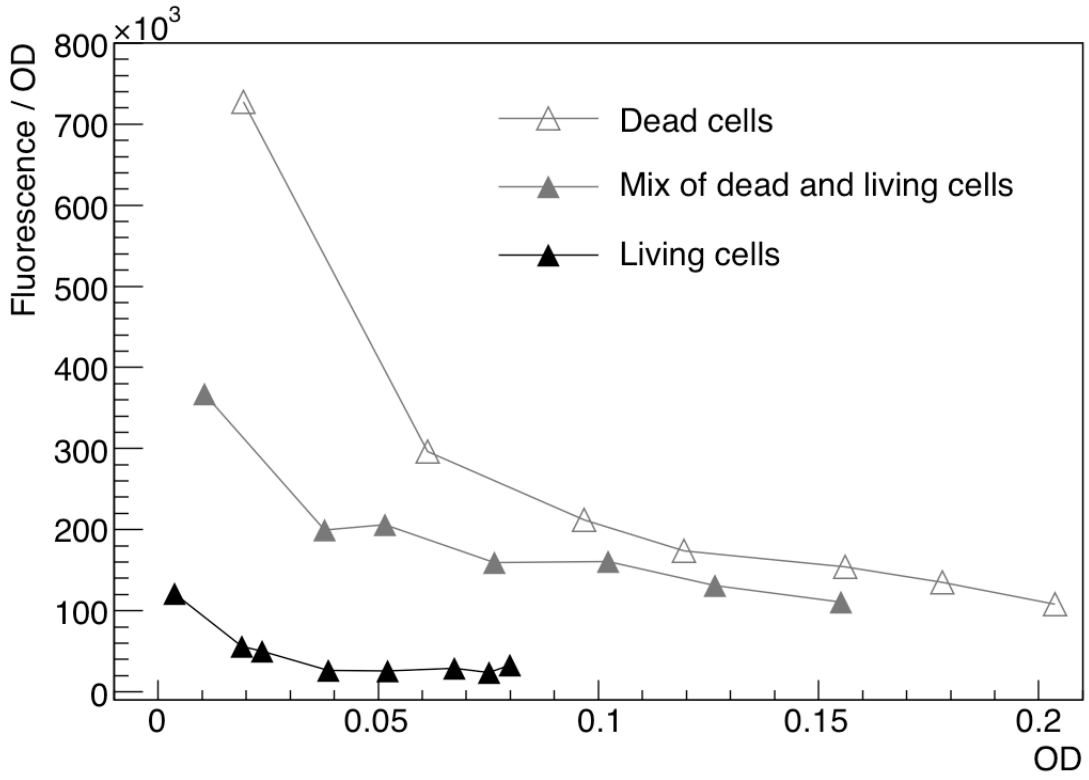


Figure S3. Impact of the OD of cultures on the ratio Fluorescence/OD measured after incubation with DiBac(4)$_3$ dye.

Cultures containing living cells ($OD_{600nm} \approx 0.08$), dead cells ($OD_{600nm} \approx 0.20$), and a mix of the two ($OD_{600nm} \approx 0.16$) were submitted to dilutions in order to generate a range of ODs.

The ratio fluorescence/OD_{600nm} was measured for each type of culture and each dilution after incubation with the membrane potential sensitive DiBac(4)$_3$ dye.

The ratios were much higher for low OD values. However, at a given OD and as expected, the ratio was always higher for a higher rate of dead cells.

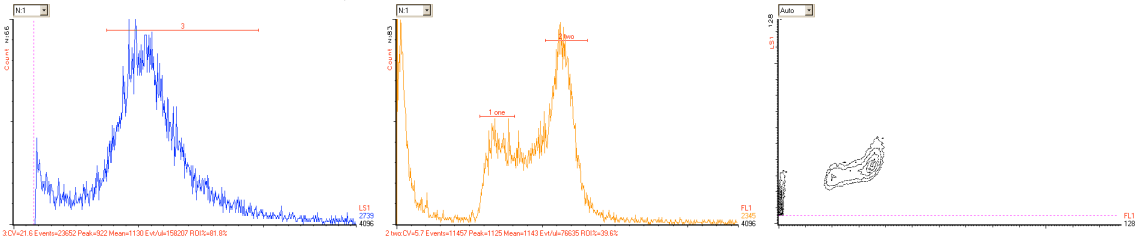
The dead cells were obtained by keeping a culture at room temperature for 1 month.

The dead and living cells were mixed in a ratio 50:50 (v:v).

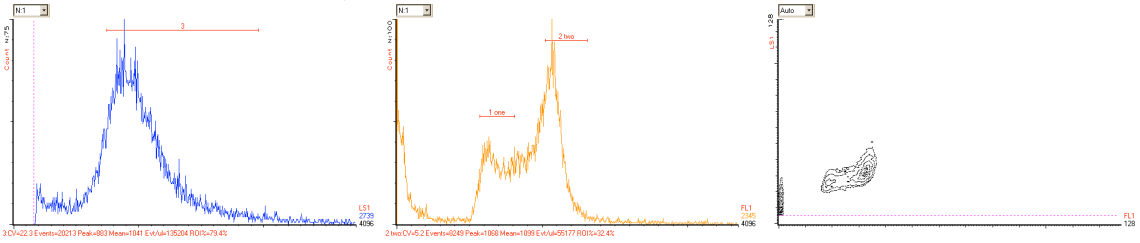
Supplemental Data

Flow cytometry analysis of a control uninfected culture of *S. islandicus* LAL14/10 strain, over a time-course.
Left column: cell size distribution
Middle column: DNA content distribution
Right column: 2D diagrams of cell size (Y-axis) against DNA content (X-axis)

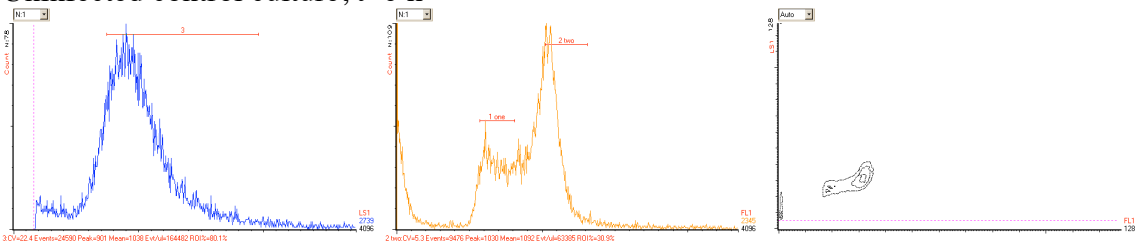
Uninfected control culture, t=-2 h



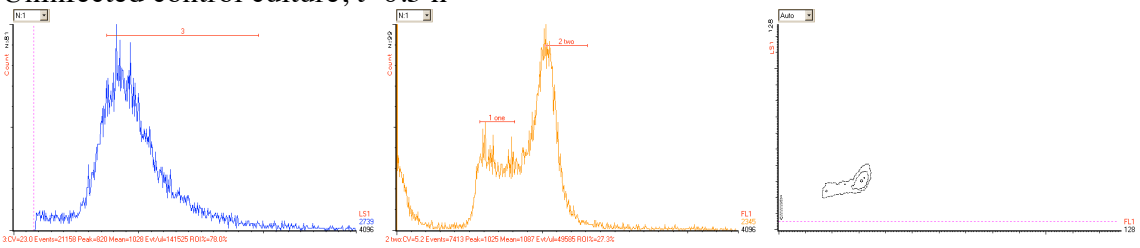
Uninfected control culture, t=-1 h



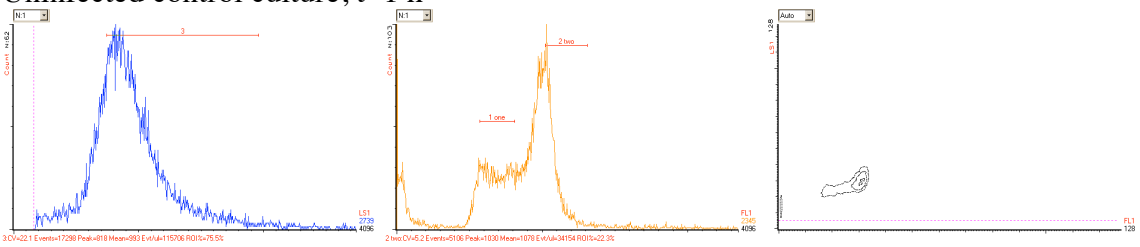
Uninfected control culture, t=0 h



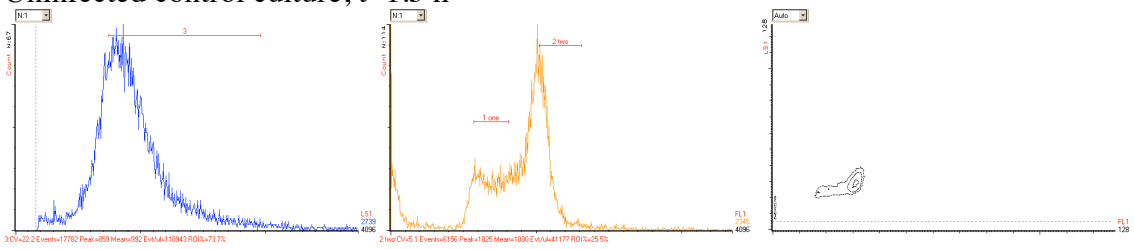
Uninfected control culture, t=0.5 h



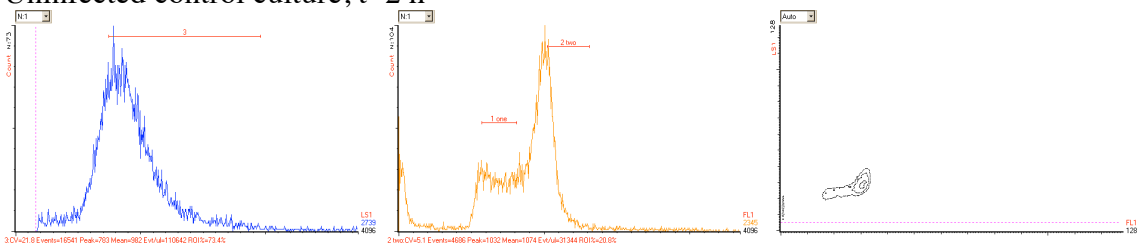
Uninfected control culture, t=1 h



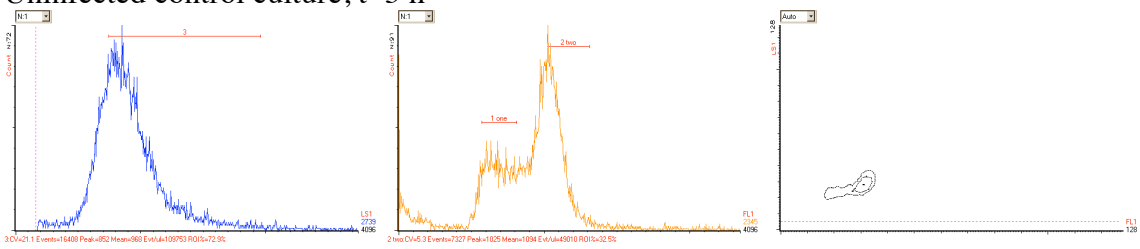
Uninfected control culture, t=1.5 h



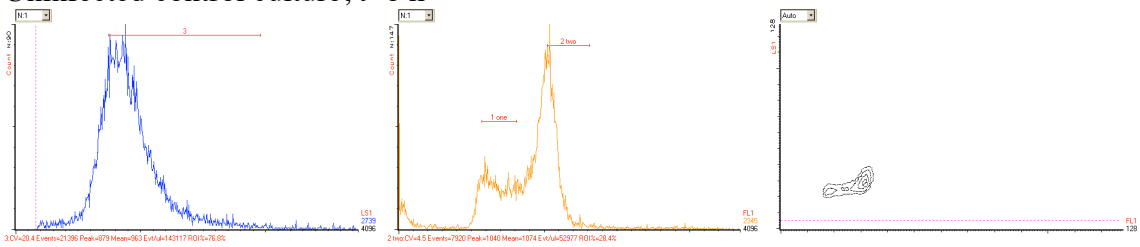
Uninfected control culture, t=2 h



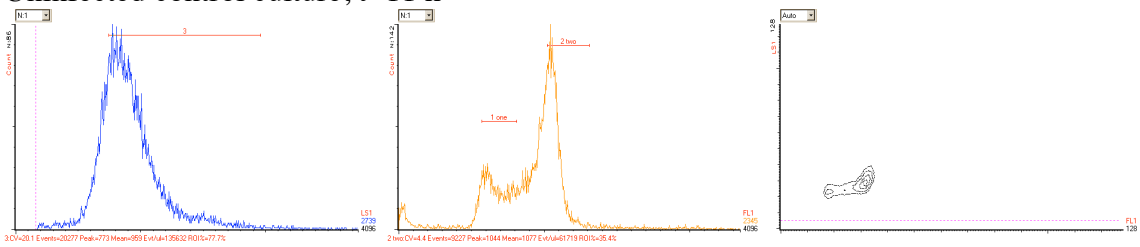
Uninfected control culture, t=5 h



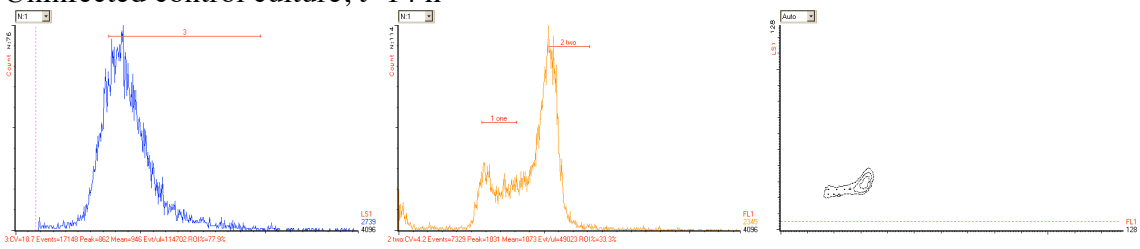
Uninfected control culture, t=8 h



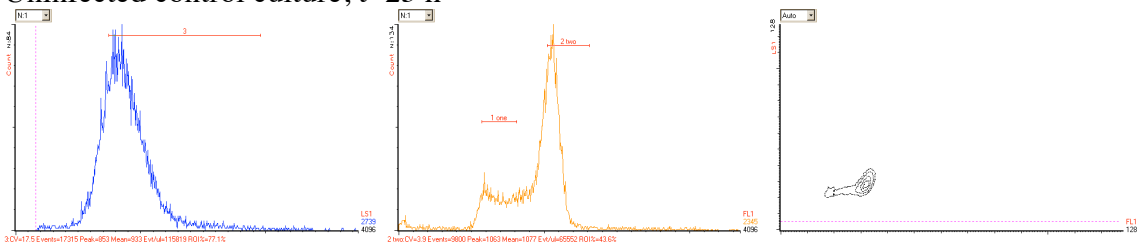
Uninfected control culture, t=11 h



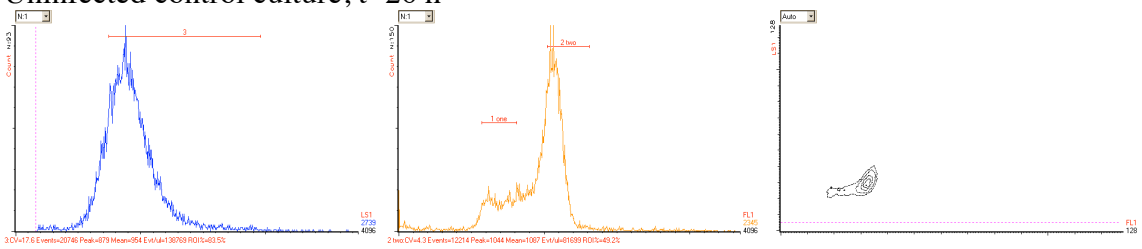
Uninfected control culture, t=14 h



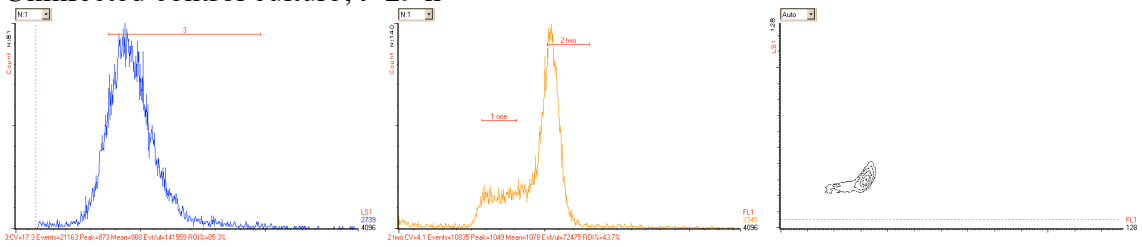
Uninfected control culture, t=23 h



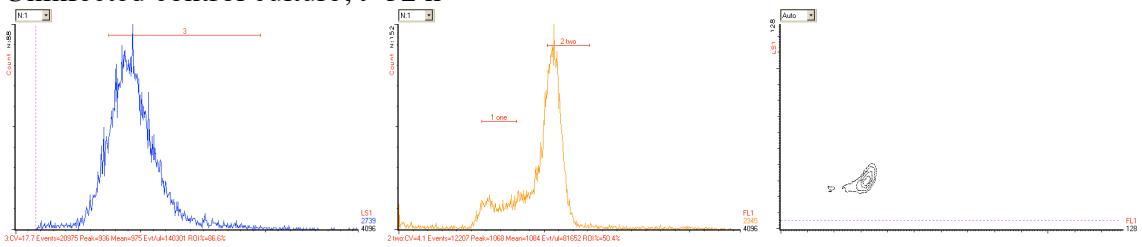
Uninfected control culture, t=26 h



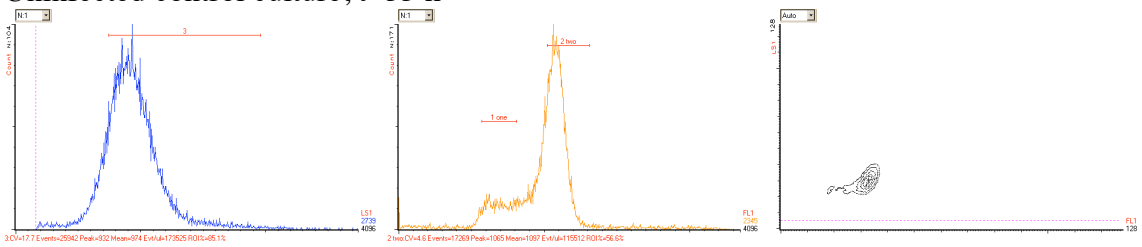
Uninfected control culture, t=29 h



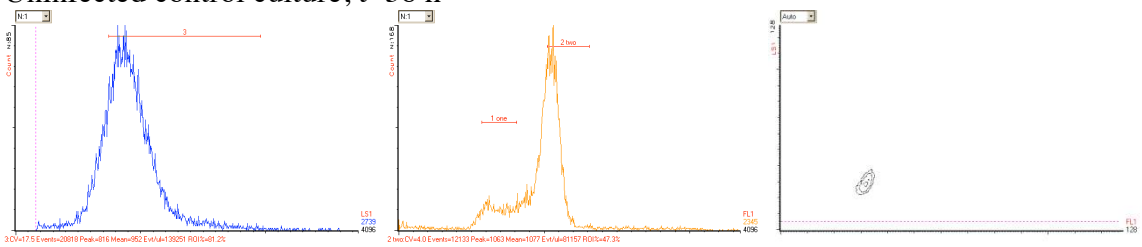
Uninfected control culture, t=32 h



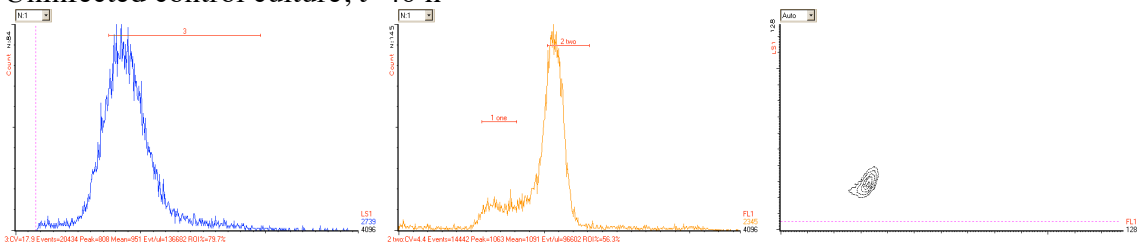
Uninfected control culture, t=35 h



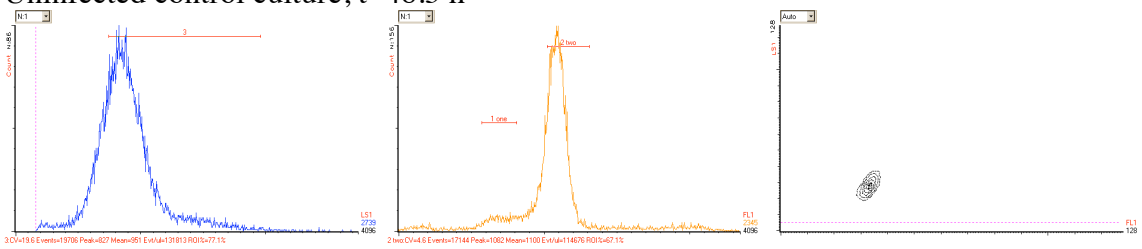
Uninfected control culture, t=38 h



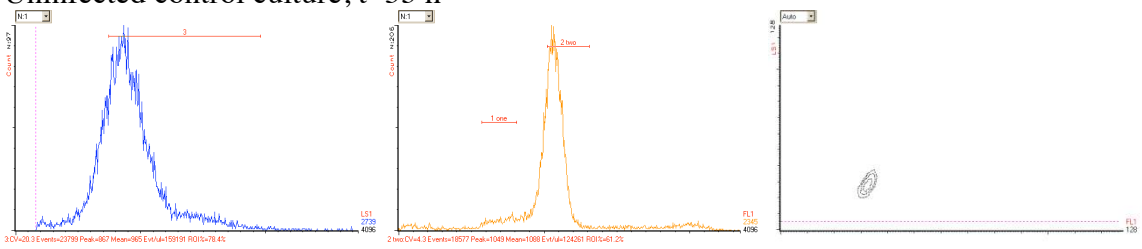
Uninfected control culture, t=40 h



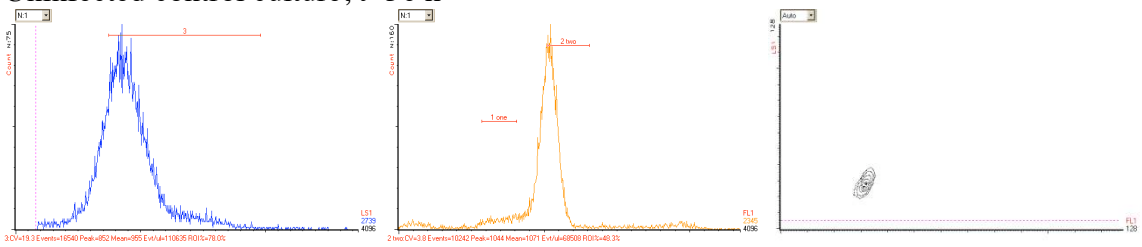
Uninfected control culture, t=48.5 h



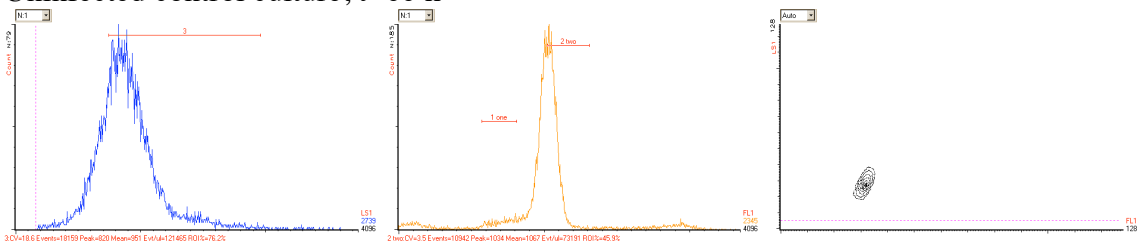
Uninfected control culture, t=53 h



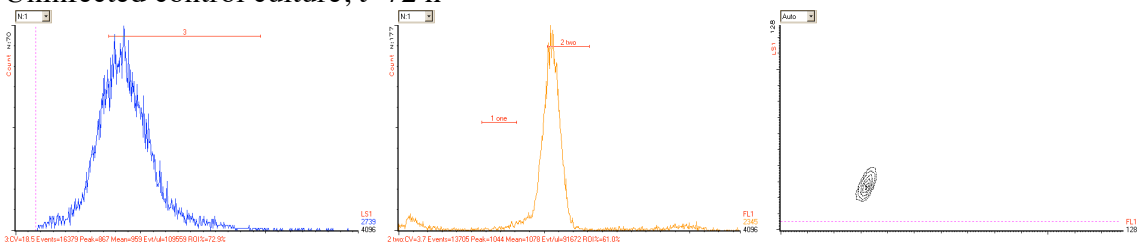
Uninfected control culture, t=56 h



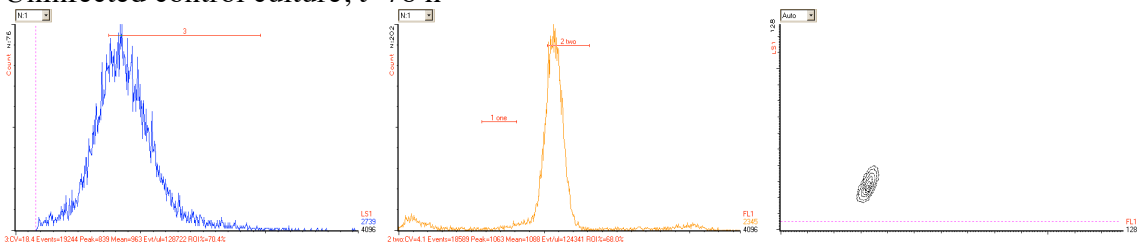
Uninfected control culture, t=60 h



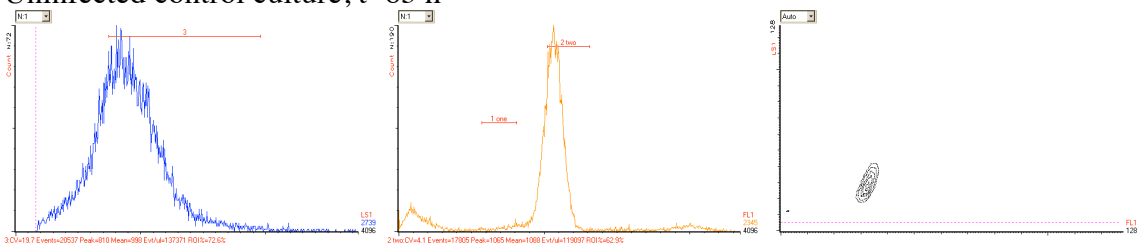
Uninfected control culture, t=72 h



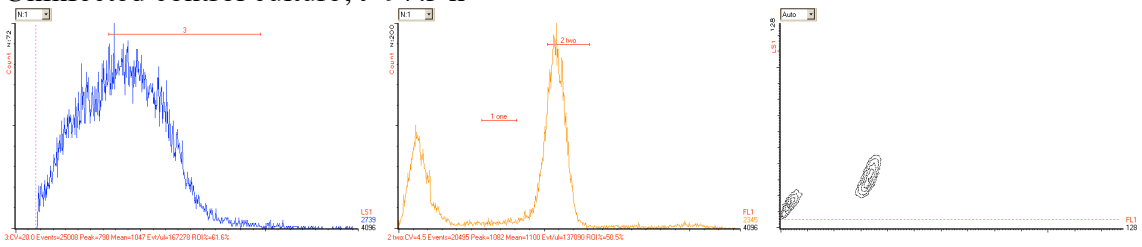
Uninfected control culture, t=78 h



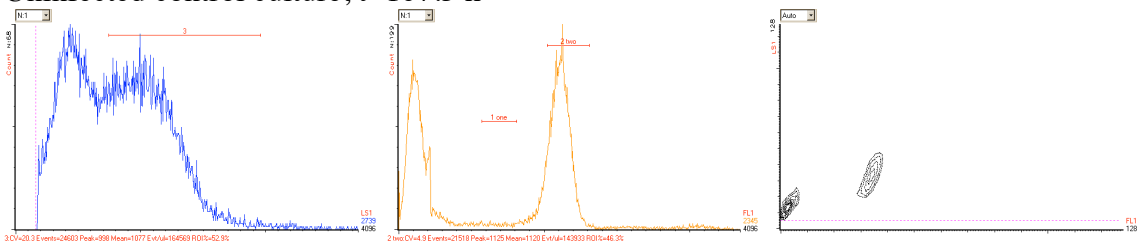
Uninfected control culture, t=83 h



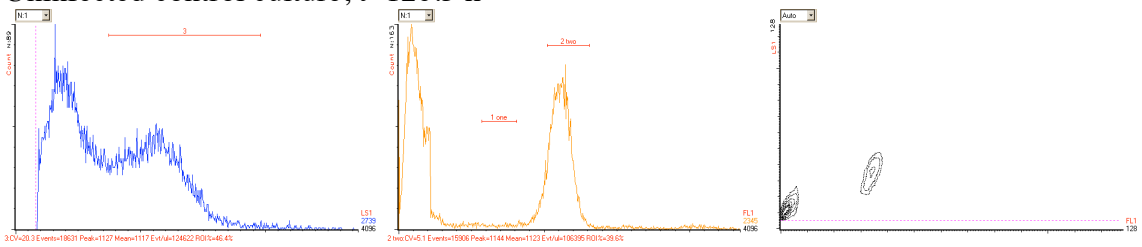
Uninfected control culture, t=97.5 h



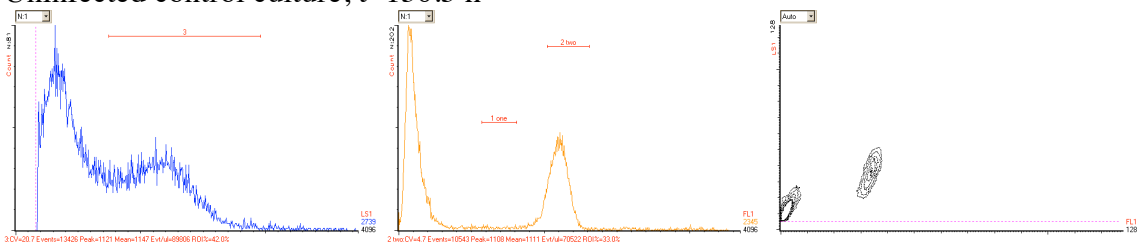
Uninfected control culture, t=107.5 h



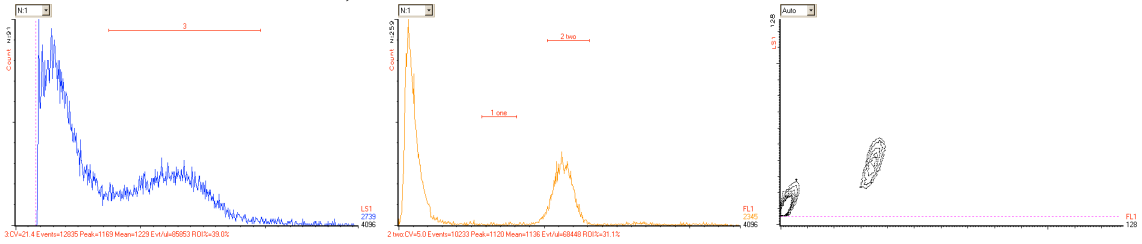
Uninfected control culture, t=120.5 h



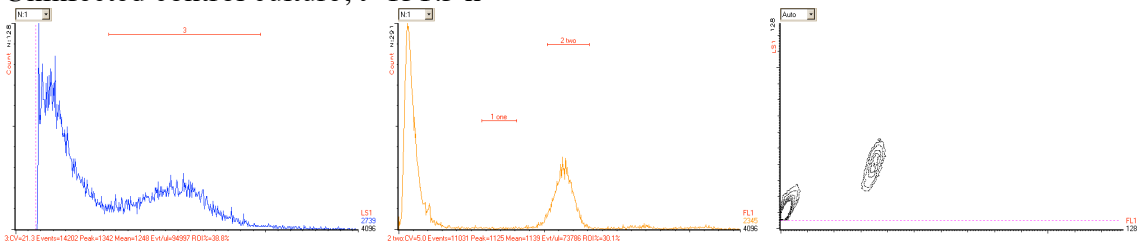
Uninfected control culture, t=130.5 h



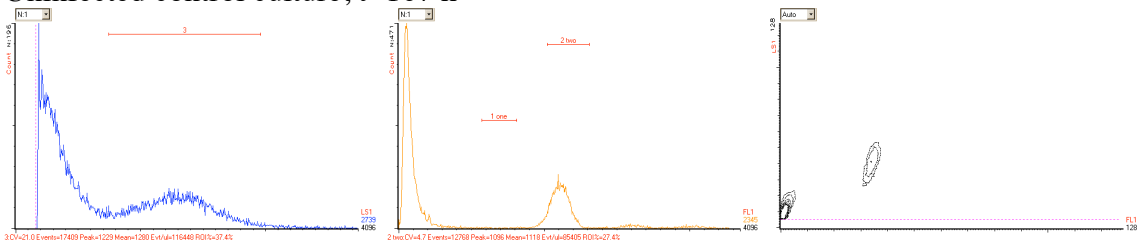
Uninfected control culture, t=144.5 h



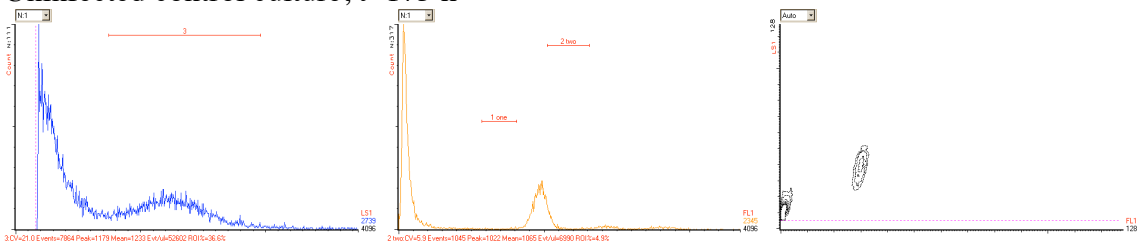
Uninfected control culture, t=151.5 h



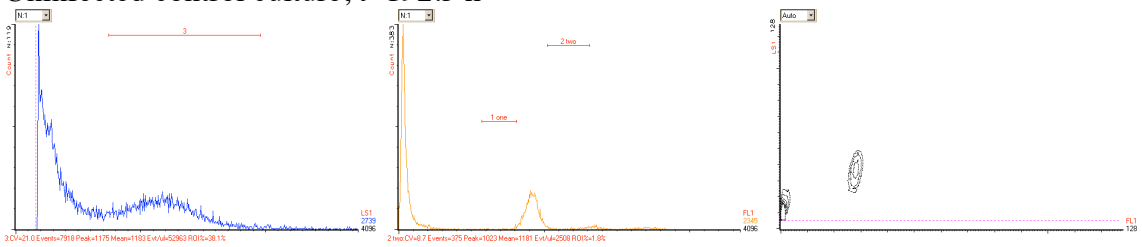
Uninfected control culture, t=167 h



Uninfected control culture, t=175 h

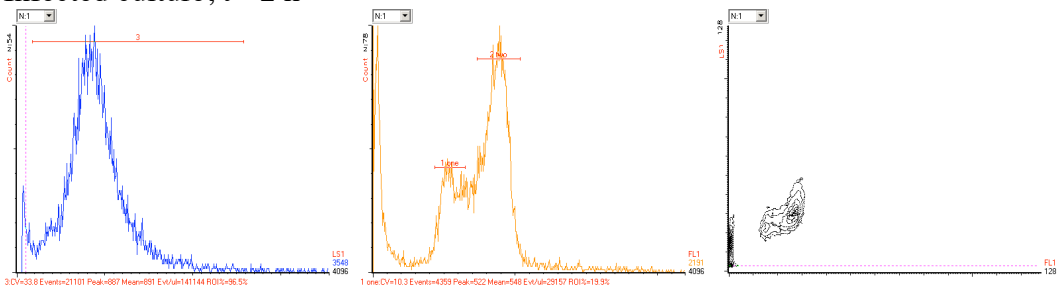


Uninfected control culture, t=192.5 h

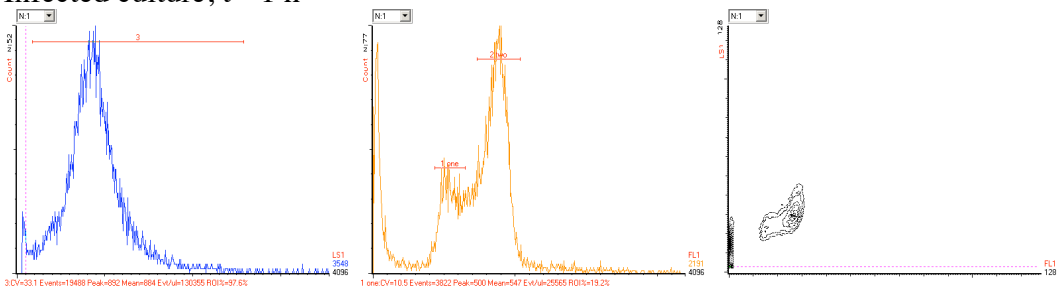


Flow cytometry analysis of a SIRV2 infected culture of *S. islandicus* LAL14/10 strain, over a time-course.
Left column: cell size distribution
Middle column: DNA content distribution
Right column: 2D diagrams of cell size (Y-axis) against DNA content (X-axis)
The growth of the resistant cells was detectable at late stages of the time course (from 97.5h).

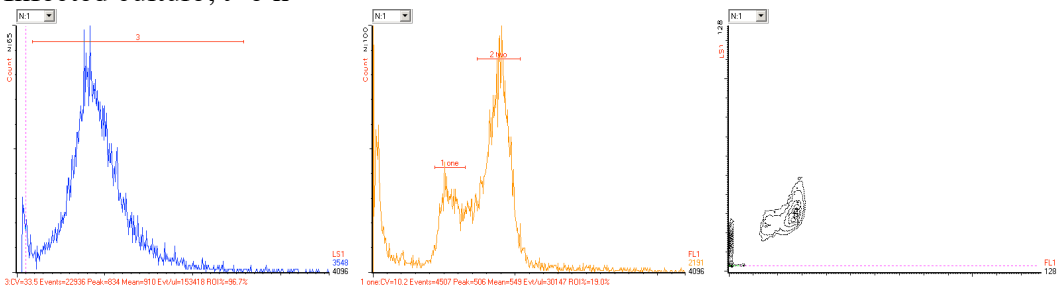
Infected culture, t=-2 h



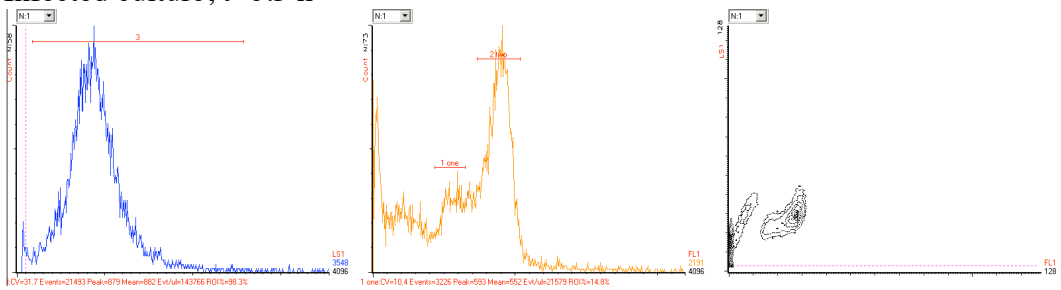
Infected culture, t=-1 h



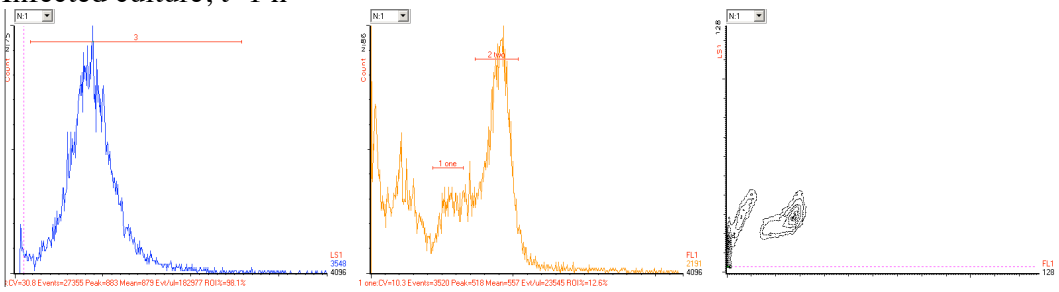
Infected culture, t=0 h



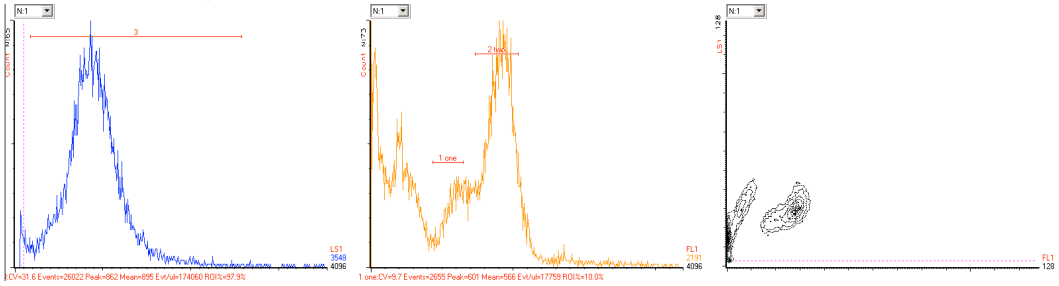
Infected culture, t=0.5 h



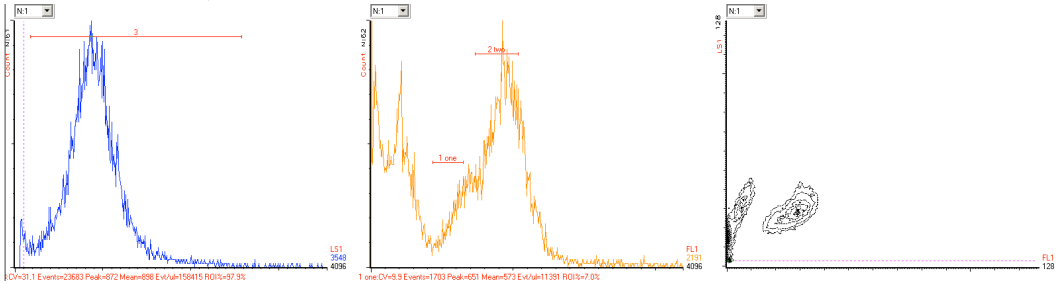
Infected culture, t=1 h



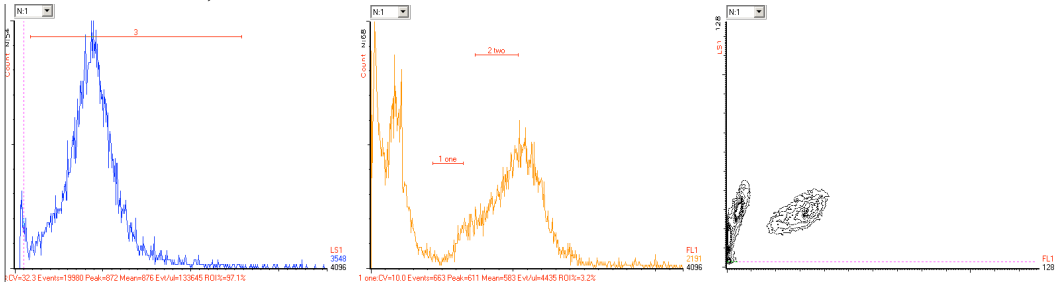
Infected culture, t=1.5 h



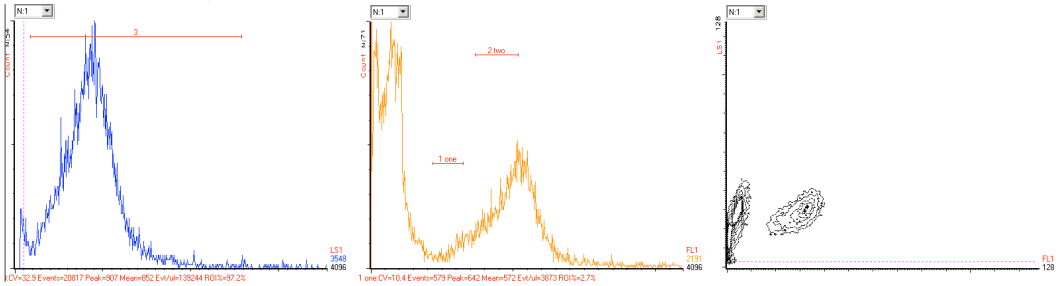
Infected culture, t=2 h



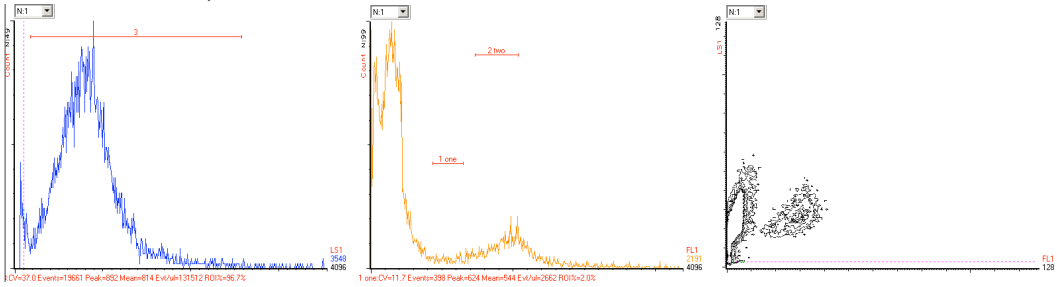
Infected culture, t=5 h



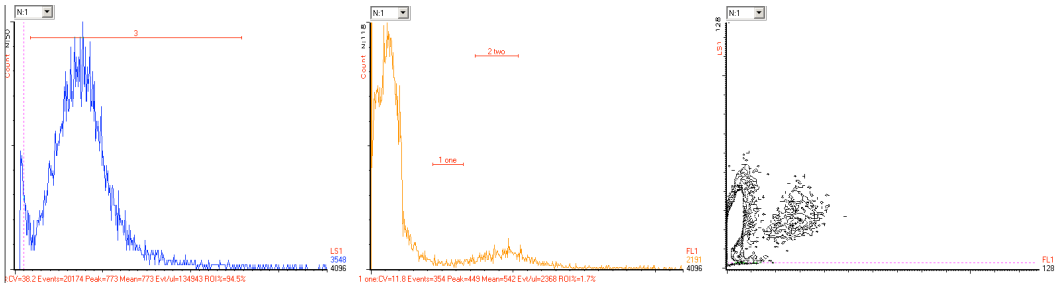
Infected culture, t=8 h



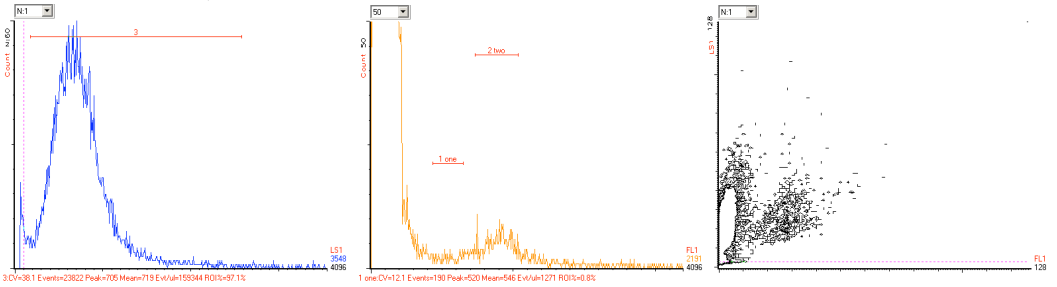
Infected culture, t=11 h



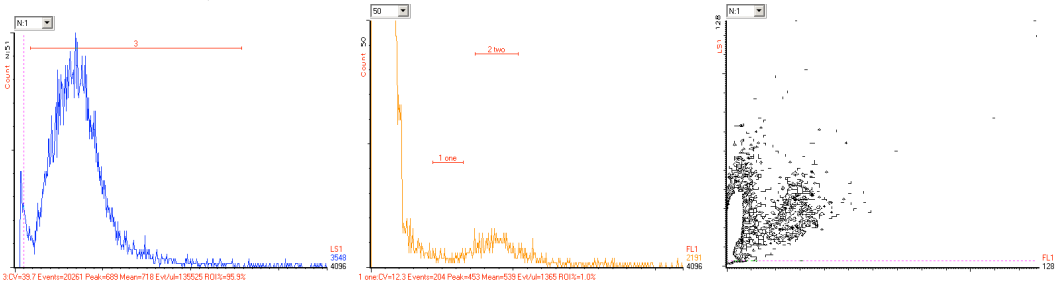
Infected culture, t=14 h



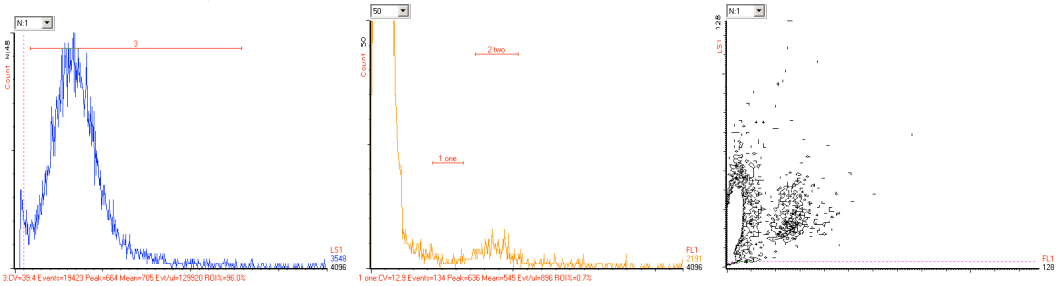
Infected culture, t=23 h



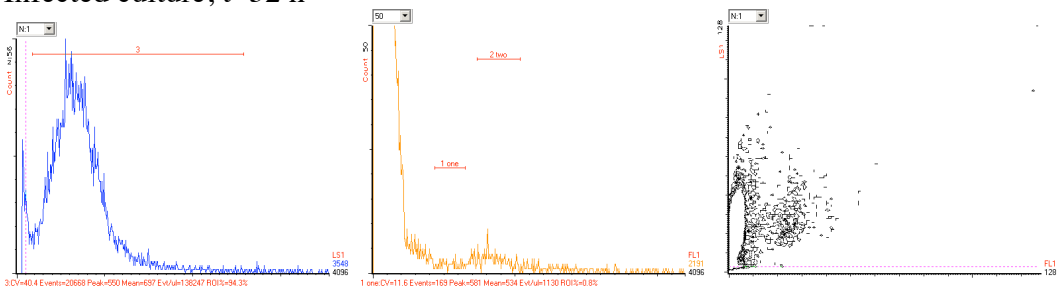
Infected culture, t=26 h



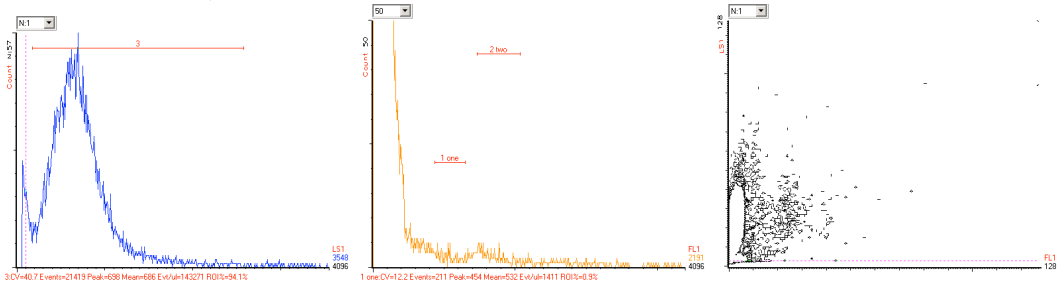
Infected culture, t=29 h



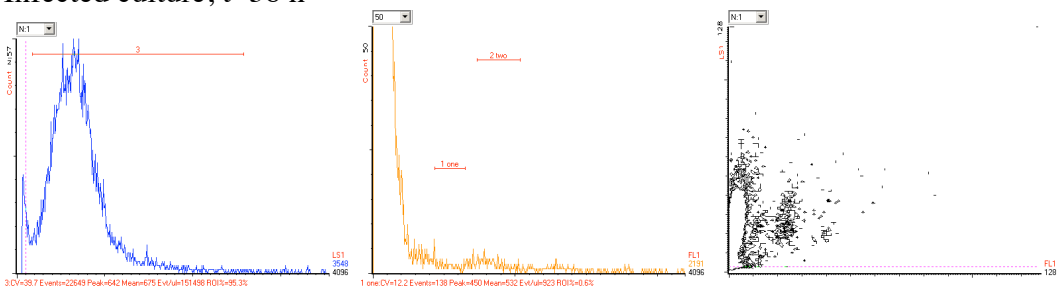
Infected culture, t=32 h



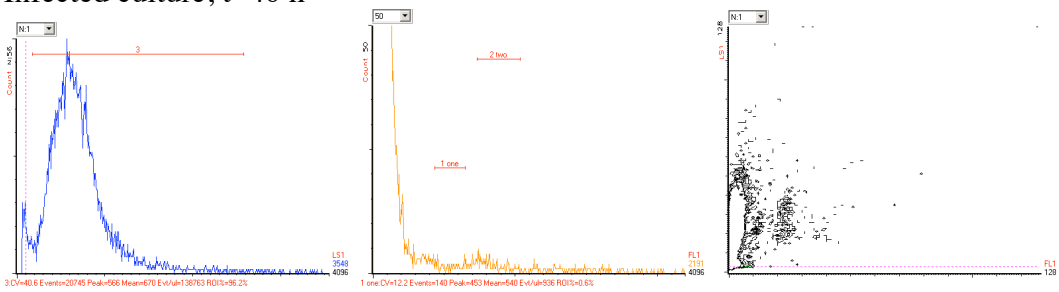
Infected culture, t=35 h



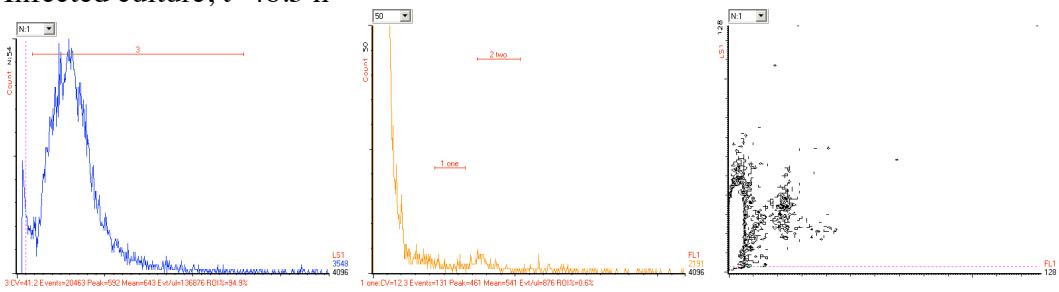
Infected culture, t=38 h



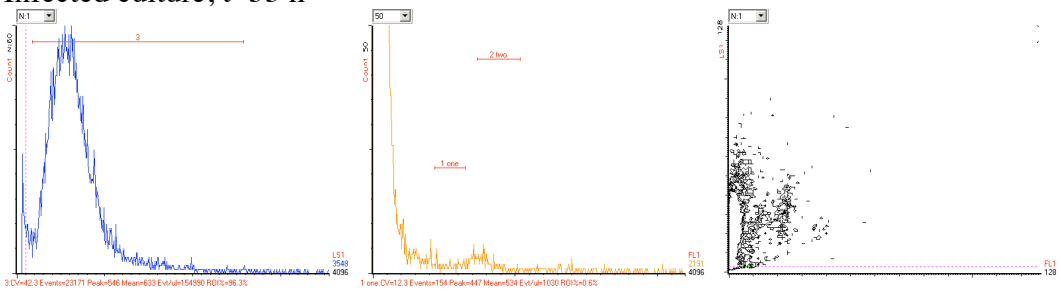
Infected culture, t=40 h



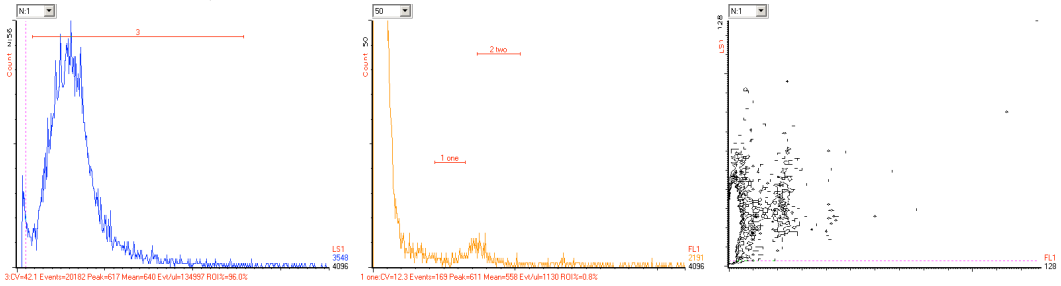
Infected culture, t=48.5 h



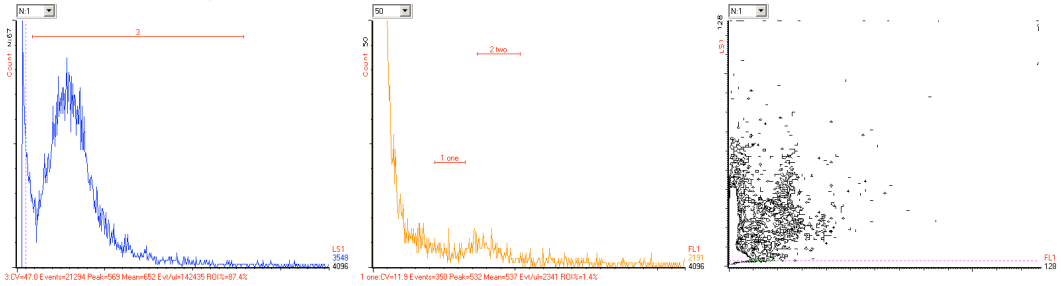
Infected culture, t=53 h



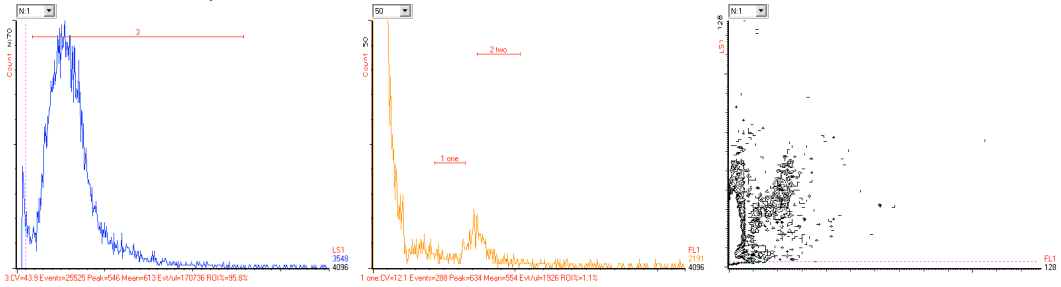
Infected culture, t=56 h



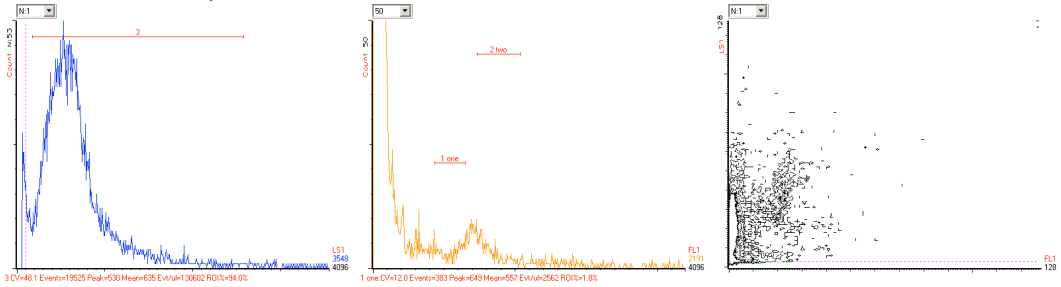
Infected culture, t=60 h



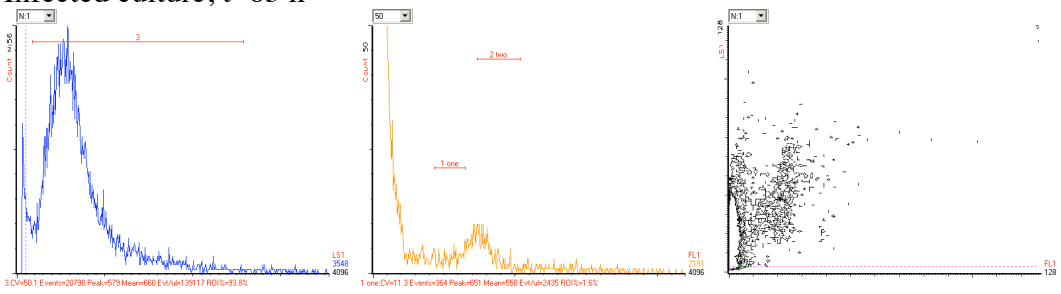
Infected culture, t=72 h



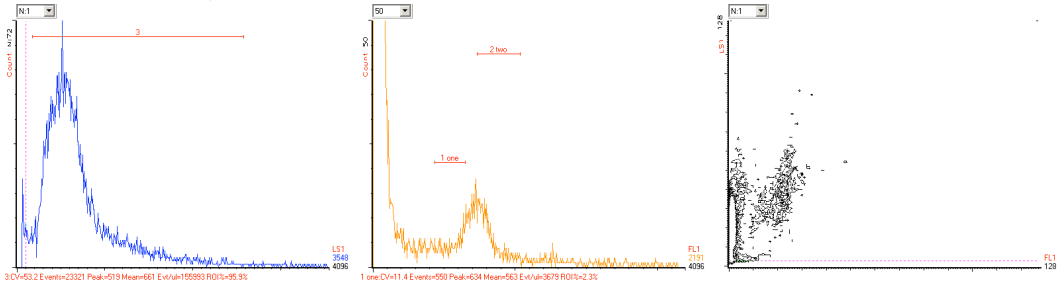
Infected culture, t=78 h



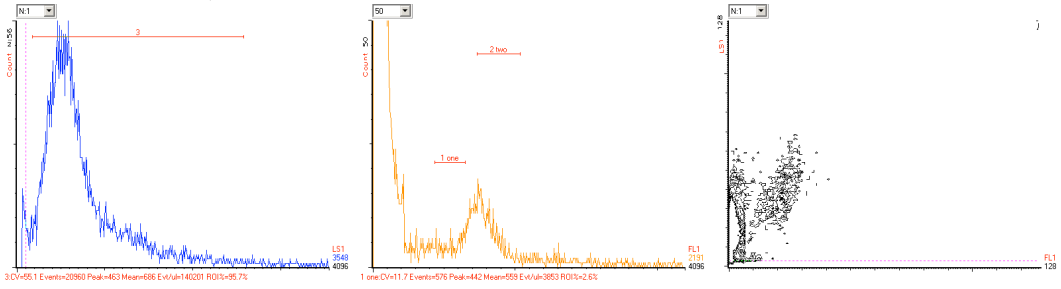
Infected culture, t=83 h



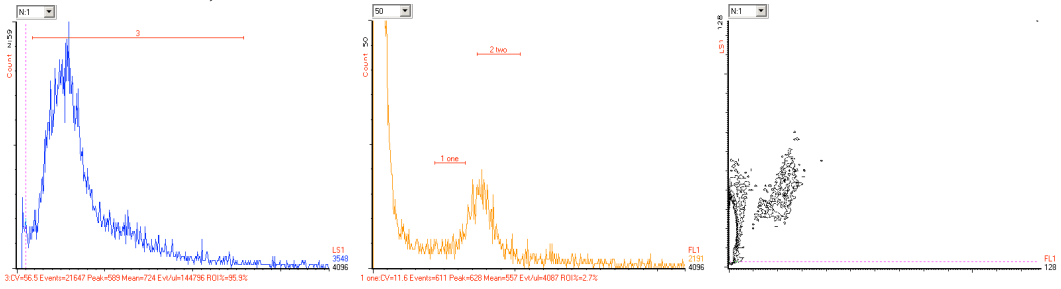
Infected culture, t=97.5 h



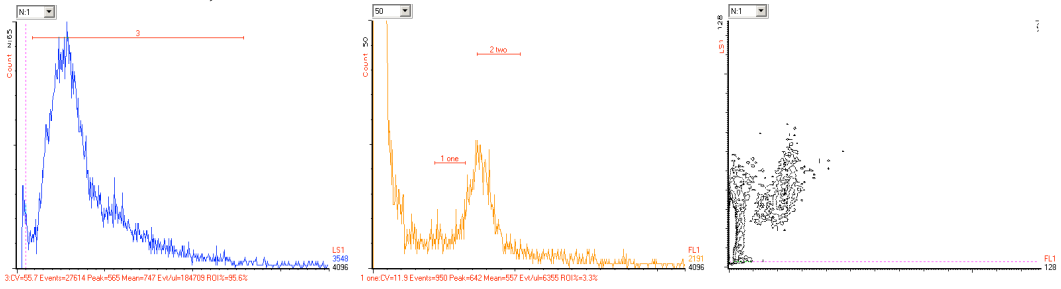
Infected culture, t=107.5 h



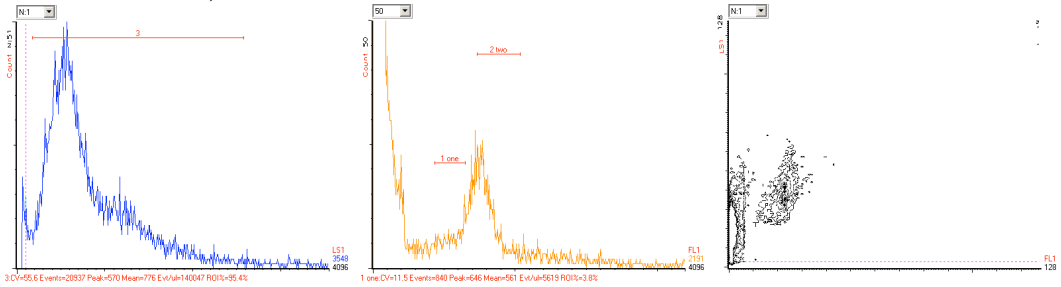
Infected culture, t=120.5 h



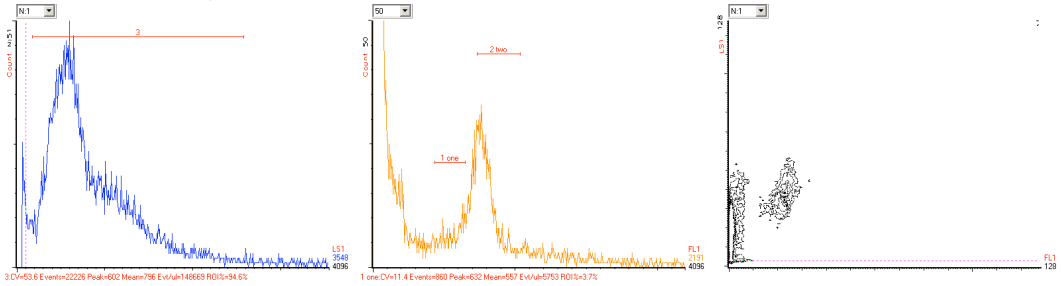
Infected culture, t=130.5 h



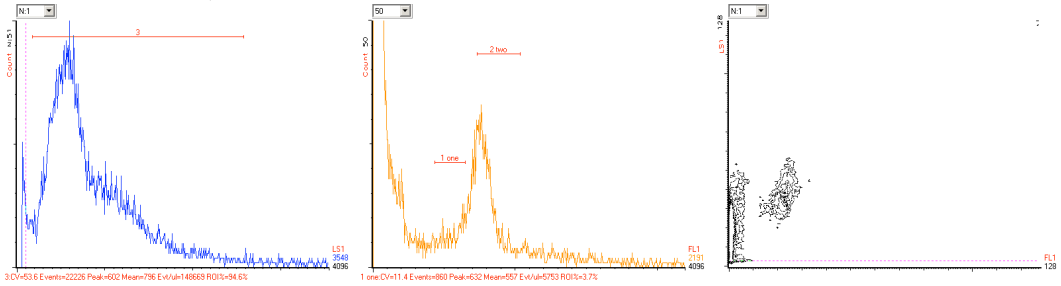
Infected culture, t=144.5 h



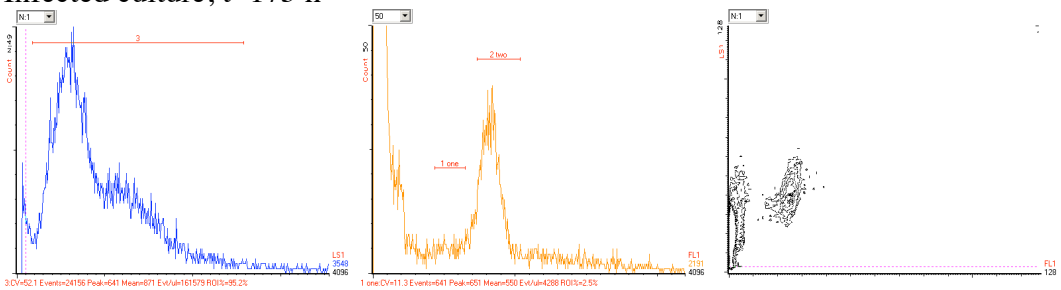
Infected culture, t=151.5 h



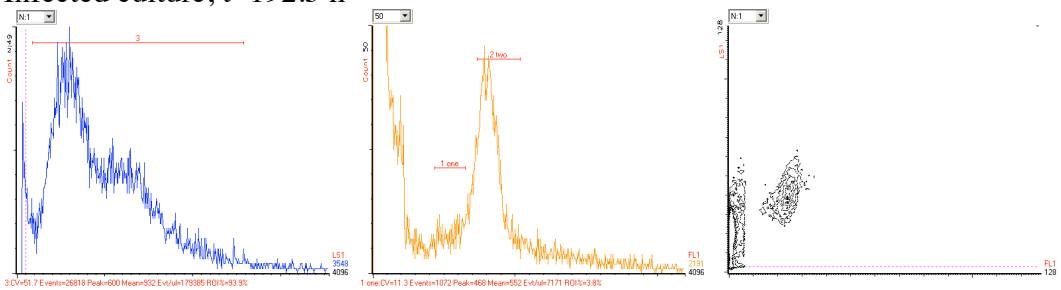
Infected culture, t=167 h



Infected culture, t=175 h



Infected culture, t=192.5 h



Supplemental References

Bernander, R., and Poplawski, A. (1997). Cell cycle characteristics of thermophilic archaea. J Bacteriol *179*, 4963-4969.

Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H. P., and Garrett, R. A. (2005). The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota. J Bacteriol 187, 4992-4999.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., *et al*. (2001). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc Natl Acad Sci U S A *98*, 7835-7840.

Wold, S., Skarstad, K., Steen, H. B., Stokke, T., and Boye, E. (1994). The initiation mass for DNA replication in Escherichia coli K-12 is dependent on growth rate. Embo J *13*, 2097-2102.

ADVANCED FUNCTIONAL MATERIALS

FULL PAPER

DOI: 10.1002/adfm.200800711

Site-specific and Spatially Controlled Addressability of a New Viral Nanobuilding Block: *Sulfolobus islandicus* Rod-shaped Virus 2**

By *Nicole F. Steinmetz,** *Ariane Bize, Kim C. Findlay, George P. Lomonossoff, Marianne Manchester, David J. Evans,* and *David Prangishvili*

Nanotechnology seeks to mimic what nature has achieved: self-assembly at the nanometer scale. Viral nanoparticles (VNPs) provide natural examples of self-assembled architectures with unique structural and chemical properties. Here, the utilization of an archaeal virus, *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2), as a template for site-selective and spatially controlled bioconjugation is described. SIRV2 is a virus of a hyperthermophilic and acidophilic host, the archaeon *S. islandicus* growing optimally at 80 °C and pH 3, and is thus, by its nature, an extremely stable VNP. The stability of SIRV2 in different solvent/water mixtures is monitored, and it is found that in other, non-natural harsh conditions the VNPs also remained intact. Further, the question of whether the particles offer attachment sites allowing for selective chemical modification and decoration with functional ligands using biotin as a probe is addressed. It is found that carboxylate-, carbohydrate- and amine-selective chemistries are applicable and various biotinylated SIRV2 formulations can be fabricated. Depending on the chemistry and hence attachment site used, the display of the biotin labels can be spatially controlled at the virus body and at the ends. Labeling studies also provide novel insights into the structural properties of SIRV2, indicating that the major coat protein (CP) forms the virus body while the minor CP is located in the tail fibers at the end of the particles. Overall, SIRV2 represents an extremely stable and structurally interesting VNP with the potential for novel nanobiotechnological applications.

1. Introduction

A major theme in nanotechnology is the design and fabrication of highly organized materials and devices at the nanometer scale. Physically and chemically tailoring materials at the nanoscale is challenging with synthetic materials, because of polydispersity and poorly controlled chemistry. Examples of synthetic nanomaterials include nanocrystals, quantum dots, dendrimers, and polymer vesicles.[1] Although functionalization methods have been developed which allow manipulation of these materials, it remains challenging to precisely tune these materials in a controllable fashion.

In recent years a set of naturally occurring bionanoparticles, including virus capsids and protein cages, have been studied and utilized as templates and building blocks for nanosciences/technology. Virus capsids and protein cages have received particular attention in nanotechnology because they are naturally self-assembled with atomic precision. The main advantages of viral nanoparticles (VNPs) are their nanometer-range size, their propensity to self-assemble into monodisperse nanoparticles of discrete shape and size, their high degree of symmetry and polyvalency, the relative ease of producing large quantities of material, their exceptional stability, robustness, biocompatibility, and bioavailability. Last but not least, the particles are composed of programmable units, which can be modified by either genetic modification or chemical bioconjugation methods.[2]

The development of VNPs for nanotechnology applications has fueled the search for novel VNPs with natural structural or chemical features that can be utilized. For example, the pH-controlled and metal ion-dependent reversible permeability of *Cowpea chlorotic mottle virus* has led to its use as a constrained chemical reaction vessel,[3] icosahedral particles such as MS2 and *Cowpea mosaic virus* have been utilized for potential applications ranging from electronics to medicine,[4] and the rod-shaped architectures of *Tobacco mosaic virus* (TMV) and

[*] Dr. N. F. Steinmetz, Prof. M. Manchester
Department of Cell Biology, The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
E-mail: nicoles@scripps.edu
A. Bize, Dr. D. Prangishvili
Department of Microbiology, Institut Pasteur
25-28 Rue Du Dr. Roux, 75724 Paris Cedex 15 (France)
K. C. Findlay
Department of Cell and Developmental Biology, John Innes Centre
Colney Lane, Norwich, NR4 7UH (UK)
Prof. G. P. Lomonossoff, Dr. D. J. Evans
Department of Biological Chemistry, John Innes Centre
Colney Lane, Norwich, NR4 7UH (UK)

[**] This work was funded by EU grant Marie Curie Early Stage Training CT-2004-504273 (NFS), the Biotechnology and Biological Sciences Research Council (KCF, GPL, DJE), and NIH CA112075 (MM). Dr. Tamara Basta (Institut Pasteur) is thanked for helpful discussions regarding studying the stability of SIRV2. Transmission electron micrographs and UV/Visible spectroscopy graphs of SIRV2 stored in different solvent/water mixtures and sonicated SIRV2 particles are given in the Supporting Information. Supporting Information is available online from Wiley InterScience or from the author.

WILEY InterScience DISCOVER SOMETHING GREAT
© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ADVANCED FUNCTIONAL MATERIALS

M13 with their helical arrays of coat proteins (CPs) have yielded templates for batteries and nanowires.[5,6]

The development of VNPs for nanotechnology applications often requires that they withstand harsh conditions during either fabrication or use. For example, bioconjugation techniques mostly require the organic solvent dimethyl sulfoxide (DMSO) in concentrations of at least 20% by volume.[7] Biomedical applications often require acid stability for use in vivo. Finally, for electronics applications temperature stability of materials is required. To this end, viruses that have been isolated from geothermal environments and which display a wide range of unique morphotypes appear especially attractive.[8,9]

Here, we studied *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2) as a candidate VNP. SIRV2 is an archaeal virus from the family *Rudiviridae*, comprising non-enveloped, rod-shaped viruses with linear double-stranded DNA genomes.[10,11] SIRV2 infects strains of the hyperthermophilic archaeal genus *Sulfolobus* isolated from a solfataric (a volcanic area that gives off sulfurous gases and steam) hot, acidic spring in Iceland with temperature 88 °C and pH 2.5. Due to its hyperthermostable and acid-resistant nature, SIRV2 is potentially an extremely stable VNP, and is therefore expected to be a useful additional building block for materials scientists.

The rod-shaped SIRV2 particles have dimensions of 23 nm × 900 nm. Visualized by transmission electron microscopy (TEM; Fig. 1), the virion body, which encapsidates the DNA genome, has a central channel with a diameter of approximately 6 nm. The terminal structures of the viral capsid are unique for rod-shaped viruses: each end of the viral particle appears to be filled by a plug about 48 nm in length and 6 nm in diameter that fills the terminal portion of the cavity (Fig. 1). In addition, three tail fibers of about 28 nm in length are found at either end of the VNPs. It is currently unknown whether the tail fibers are attached to the plugs or to the virus body.

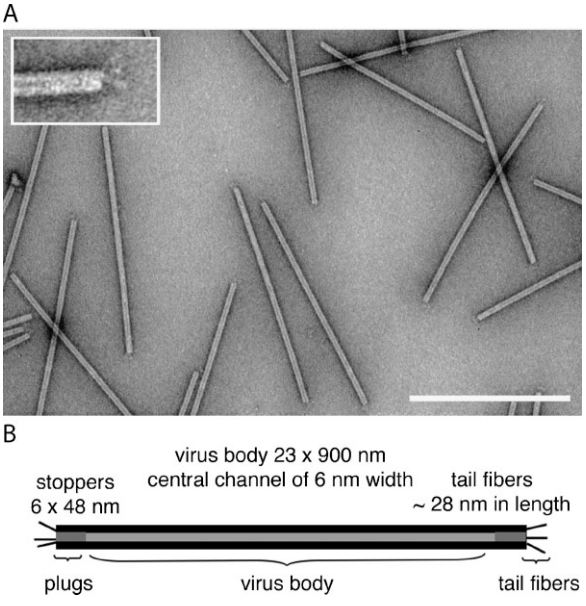


Figure 1. Structure of SIRV2 particles. A). TEM of negatively stained SIRV2 particles. Inset shows a high resolution image of the end structure of the capsid. The scale bar is 500 nm. B). Schematic depiction of SIRV2 particles.

Two structural SIRV2 proteins have been identified: a major CP of about 20 kDa and a minor CP of about 100 kDa. High-resolution TEM showed that the virus body is formed by the major CP and that the subunits are arranged in a helical conformation with a periodicity of 4.3 nm.[10] The major CP is highly basic with a theoretical pI of approximately 11.0, suggesting that the major CP binds with the negatively charged DNA and forms the helical array of the virus body.[10] The level of minor CP was found to be less than 1% of the major CP, leading to the hypothesis that the minor CP is associated with the end structure of the particles and is present in either the tail fibers or the plugs. Neither the structures of the individual SIRV2 CPs nor the assembled particle have been solved to date.

Due to their highly thermostable and acid-resistant nature, SIRV2 particles are promising candidates as platforms for functionalization for technological and/or medical applications. In addition, the structural properties are interesting, particularly the fact that the particles are comprised of two different CPs and that the ends are different from the virus body. These physical and structural properties differentiate SIRV2 from other rod-shaped VNPs such as M13 and TMV. TMV is a rather stiff rod with dimensions of 18 nm × 300 nm, and the capsid consists of a single type of CP subunit. Selective end-modification of TMV has been achieved using a controlled disassembly protocol, to expose the encapsidated genomic RNA molecules that were then subsequently selectively functionalized.[12] However, problems with this strategy derive from the inherent chemical and enzymatic instability of the RNA molecule. The filamentous coliphage M13 has dimensions of 6.5 nm × 880 nm in length and is comprised by a major CP and four different minor CPs that are found at the ends of the particles.[13] M13 thus offers, like SIRV2, ends that are different from the virus body. However, M13, in contrast to SIRV2, is a flexible filament. This property can be an advantage or disadvantage depending on the application. In terms of controlled self-assembly of arrays it is likely that less flexible particles can be aligned with greater ease and precision. Overall the SIRV2 virion presents a unique particle, and no comparable VNP building block has been studied so far.

In this paper, we have studied the possibility of chemically addressing SIRV2 particles. We first tested and monitored the stability of SIRV2 under conditions that are of relevance in terms of chemically programming the particles. Second, we asked whether the particles offer attachment sites allowing selective chemical modification. To do this, the particles were probed with functionalized biotin molecules using carboxylate-, carbohydrate-, and amine-selective chemistries. We also investigated whether the display of ligands can be spatially controlled, i.e., whether the minor and major CP could be selectively labeled, in order to provide further information about the role of the CPs in the architecture of the particle.

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

2. Results and Discussion

2.1. Stability of SIRV2 in Different Solvents

SIRV2 is a hyperthermophile and acidophile naturally growing at 88 °C and at pH values of ≤2.5. Therefore, it is reasonable to assume that its particles are stable at high temperatures and extremely low pH values. To assess the tolerance of particles toward non-natural harsh conditions, the stability of the particles in different solvents was monitored by TEM to observe particle architecture and intactness, and UV/Vis spectroscopy to quantify the absorbance of the particles over time. In addition, the maintenance of infectivity for *S. islandicus* was assessed by plaque assay. For solvents, DMSO and ethanol at 20 and 50% by volume were chosen. DMSO is used in many bioconjugation methods with reactive dyes and/or linker molecules. Ethanol is a suitable solvent for mineralization reactions such as formation of silica on the particle surface.

SIRV2 particles were dissolved in ethanol/water and DMSO/water mixtures. Tris-acetate buffer was used as a reference, since from studying and handling the virus we have observed that the particles are stable under these conditions. The stability was monitored over 6 days at room temperature. For comparison, non-intact SIRV2 particles were obtained by sonication of a concentrated virus suspension. TEM imaging, UV/Vis spectroscopy and plaque assays confirmed that sonicated SIRV2 particles were non-intact, highly disrupted, and non-infectious (Supporting Information). Data obtained using the sonicated, non-intact SIRV2 particles also demonstrated that the methods chosen, TEM, UV/Vis spectroscopy, and plaque assays, are feasible techniques to determine whether SIRV2 is intact or not.

2.1.1. Stability in Ethanol

TEM studies (Supporting Information) confirmed that the particles remained intact for at least 6 days in ethanol/water mixtures containing 20 and 50% ethanol by volume, at room temperature.

The structural features were well preserved. Monitoring the stability using UV/Visible spectroscopy (Fig. 2 and Supporting Information) was also consistent with the particles remaining stable, the concentration (proportional to the absorbance at 260 nm) being consistent over the monitored time period. The minor fluctuations seen in Figure 2 are probably attributable to evaporation of solvents and/or pipetting variation.

Infectivity measured by plaque assays confirmed that SIRV2 stored in 20% v/v ethanol/water mixtures remained infectious, indicating stability of the particles under these conditions. In contrast, the ability to infect their host was altered when the particles were stored in an ethanol/water mixture containing 50% v/v ethanol (Fig. 3). The data indicate that the particles remained infectious for a time period of two days. After this time point the infectivity dropped reproducibly, indicating that ethanol at this concentration impaired the ability of SIRV2 to infect cells; however, it did not eliminate infectivity completely. Also, it is possible that the solvent interfered with the host strain.

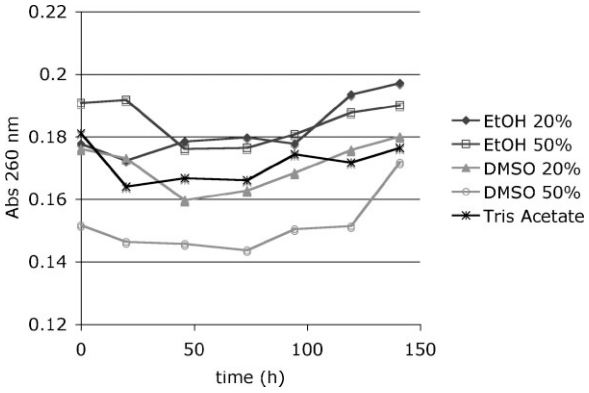


Figure 2. Stability of SIRV2 in ethanol/water and DMSO/water mixtures monitored by UV/Vis spectroscopy. The absorbance at 260 nm versus time is shown; SIRV2 was monitored in 20 and 50% v/v ethanol/water and DMSO/water, respectively. Buffer = 20 mM Tris-acetate buffer pH 6.0. SIRV2 in buffer is stable and was also monitored as a reference.

Overall, the stability of SIRV2 particles in ethanol/water mixtures was demonstrated, for at least two days, even at higher concentration. Ethanol is a useful solvent when working with silica. It is used either for mineralization reactions or for activating and functionalizing glass surfaces for immobilization of biomolecules using silanol chemistry. For example, mineralization of TMV has been achieved by adding tetraethoxysilane and/or aminopropylethoxysilane to a suspension of TMV in an acidified (pH 2.5) ethanol/water mixture.[14] It is expected that the more stable SIRV2 particles would offer a feasible platform to serve as a template for such mineralization reactions. In a pilot study it was shown that SIRV2 can indeed be used as a template for the nucleation of silica (Steinmetz, unpublished).

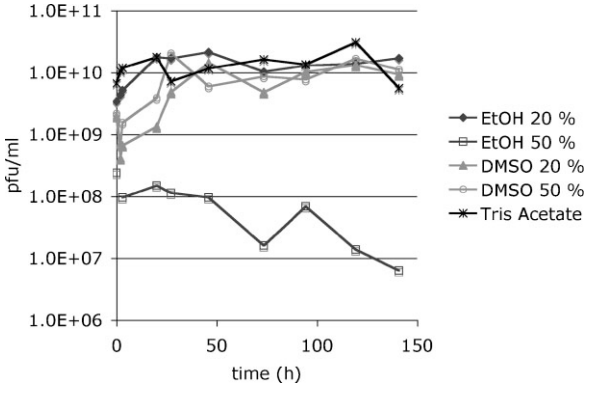


Figure 3. The infectivity of SIRV2 after storage in 20 and 50% v/v ethanol/water and DMSO/water, respectively, was monitored using plaque assays. Buffer = 20 mM Tris-acetate buffer pH 6.0. SIRV2 in buffer was also monitored as a reference. The number of pfu per milliliter was determined and is shown as a function of time in hours.

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ADVANCED FUNCTIONAL MATERIALS

2.1.2. Stability in DMSO

TEM imaging (Supporting Information), UV/Vis spectroscopy (Fig. 2 and Supporting Information), and plaque assays (Fig. 3) were all consistent with the particles remaining intact in DMSO/water mixtures containing up to 50% DMSO v/v for at least 6 days. Stability in DMSO is important, as DMSO is used as a solvent for many bioconjugation methods. Succinimide-activated and maleimide-activated labels and linkers are typically DMSO-soluble. Labeling reactions carried out later in this study were performed in DMSO/water mixtures using a concentration of 20% DMSO by volume.

2.2. Addressability of SIRV2

The atomic structure of SIRV2 particles has not been solved; therefore, it is unknown which amino acids are present on the solvent-exposed surface of the viral capsids. However, the amino acid sequences of the major CP as well as the minor CP are known.[15] Functional groups that can be easily modified with activated selective chemical modifiers are thiols derived from cysteines, amines from lysine side chains, and carboxylates from aspartic and glutamic acid. Furthermore, carbohydrates from glycosylated proteins can also serve as a suitable target for bioconjugation methods.[7]

SIRV2 does not contain any cysteine residues in either the minor or the major CP. However, a large number of carboxylate-containing amino acids are present: 41 aspartic acids and 81 glutamic acids on the minor CP, and 1 each of aspartic and glutamic acid on the major CP (Genebank minor CP amino acid sequence NP_666572, major CP sequence NP_666560). Also a high number of amines derived from lysines are found: 73 lysines are on the minor CP and six on the major CP.

Three different chemistries were employed in order to test the addressability and utility of SIRV2 as a platform for the selective display of functional molecules. For a proof-of-concept study we chose biotin as a functional molecule for selective attachment. Biotin is a small and stable molecule, and various activated-biotin probes are available allowing facile bioconjugation. Also, detection of biotin can be easily carried out using either labeled streptavidin probes or conjugated anti-biotin antibodies. After chemical derivatization, biotinlyated SIRV2 samples were characterized using Western blot analysis probing with alkaline phosphatase-conjugated streptavidin and by TEM imaging using immunogold-staining by anti-biotin antibodies.

2.2.1. Addressable Carboxylates on SIRV2

To test whether carboxylates derived from aspartic and glutamic acids on SIRV2 were addressable and to test whether functional molecules could be installed, SIRV2 particles were reacted with biotin-hydrazide in the presence of the carboxylate-selective coupling agent 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). EDC activates carboxylates by the formation of an *O*-acylisourea ester intermediate. This unstable intermediate reacts readily with primary amines, in this case derived from biotin-hydrazide, which results in the formation of a stable covalent amide bond (Fig. 4).

The coupling reaction was performed as described in Section 4, followed by purification using size-exclusion spin columns. Western blot characterization probing with streptavidin-alkaline phosphatase, showed that the biotin labels were attached to both the minor and major CP of SIRV2 (Fig. 5A). This is as expected since both CP subunits contain carboxylates derived from aspartic and glutamic acid.

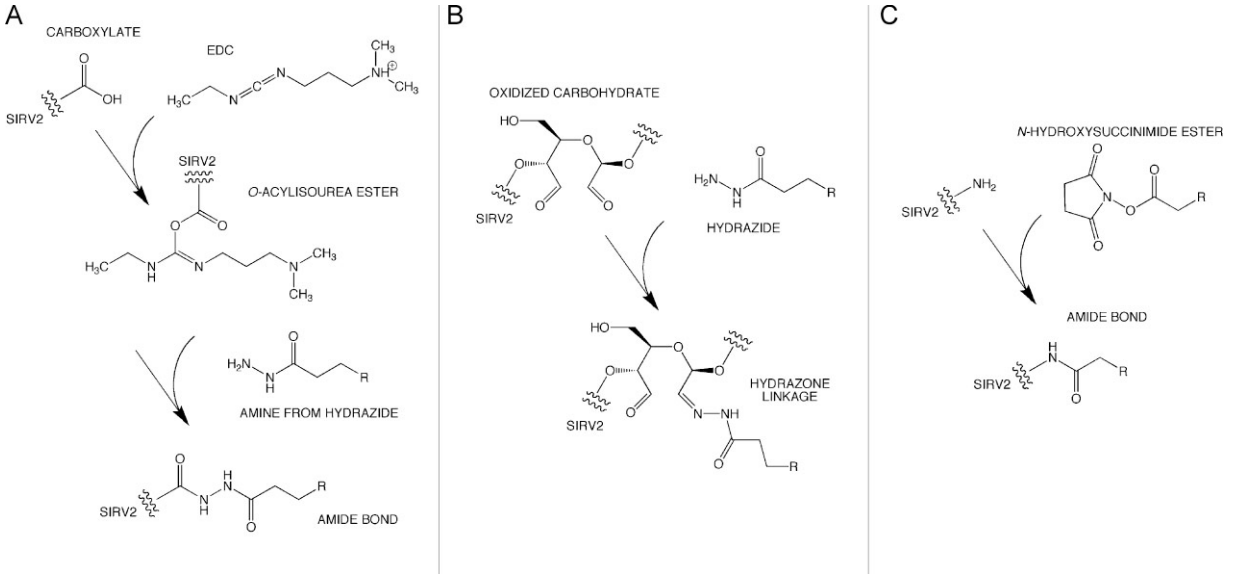


Figure 4. Chemical bioconjugation methods applied to SIRV2. A) Addressing carboxylates, chemical modification is carried out activating the carboxylate with EDC followed by reaction with a primary amine to give an amide bond. B) Carbohydrates on SIRV2 were oxidized and reacted with a hydrazide resulting in a stable hydrazone linkage. C) Utilizing primary amines on SIRV2, reaction is achieved by using a *N*-hydroxysuccinimde ester. Coupling results in the formation of a stable amide bond.

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

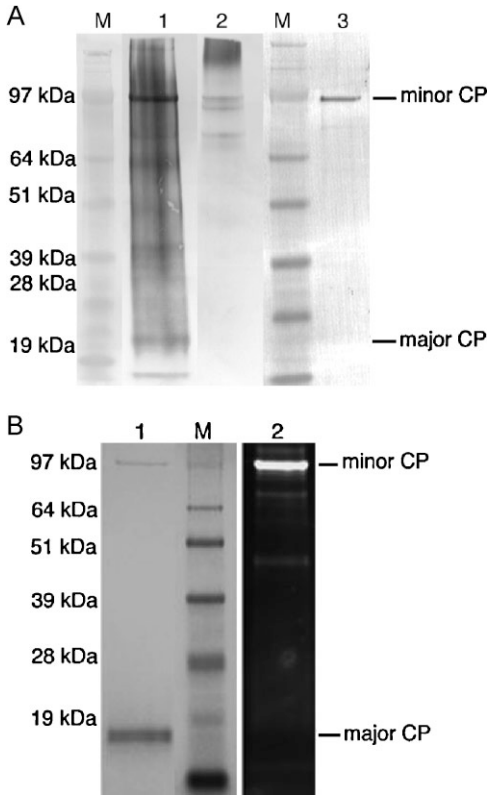


Figure 5. A) Western blot characterization of biotinylated SIRV2 particles. Blots were probed with alkaline phosphatase-conjugated streptavidin. M = SeeBlue Plus2 prestained protein marker (Invitrogen), 1, Biotinylated SIRV2 particles labeled using carboxylate-selective chemistry, 2, Biotinylated SIRV2 particles labeled at carbohydrates, 3, Biotinylated SIRV2 particles labeled using amine-selective bioconjugation techniques. The major CP is about 20 kDa and the minor CP of about 100 kDa in size. B) SDS-PAGE of AF568-labeled SIRV2 using amine-selective bioconjugation methods. 1, Gel visualized after Coomassie Blue Staining, the minor and major CP are detected. M = SeeBlue Plus2 prestained protein marker (Invitrogen), 2, Visualization of the gel under UV light, only the minor CP is detectable.

In addition to the CP bands, further bands of various molecular weights were detected. These additional bands can be explained by non-SIRV2 impurities found in the virus preparation. We identified some of these bands as being vesicles and pili that are derived from the host archaea. Purification of SIRV2 is challenging and we are currently working on the development of an improved purification protocol in order to avoid impurities being present in the virus preparation. This will be an important requirement for use of chemically tailored SIRV2 for applications in materials science and nanotechnology.

The major CP forms the virus body, and we hypothesized that the minor CP is associated with the end structures of the VNP, either the plugs or the tail fibers. Immunogold-staining using gold-labeled anti-biotin antibodies and TEM imaging showed that biotin labels were present in both the virus body and the end structures (Fig. 6). The presence of background staining also reflects that the virus preparation was not entirely free from cellular contaminants. The data from Western blot analysis and immunogold-staining studies are thus in good agreement. Western blot studies confirmed decoration of the major and minor CP with biotin, and immungold-staining indicates that the gold labels were attached to both the virus body and the end of the virion. The tail fibers are about 28 nm in length and this is consistent with the relative position of the gold-labeled anti-biotin antibody relative to the end structures of SIRV2 (Fig. 6). This implies that the minor CP is in the tail fibers.

2.2.2. Addressable Carbohydrates

In a second experiment, the addressability of SIRV2 using carbohydrate-selective chemistries was examined. These studies are also expected to give clues about whether SIRV2 is glycosylated. Prior to bioconjugation carbohydrates were oxidized using mild oxidation conditions and the reagent sodium meta-periodate. This treatment leads to the oxidation of primary hydroxyl groups present in carbohydrates into aldehyde groups. The aldehyde groups are reactive toward hydrazide conjugates and undergo a facile coupling reaction, which results in the formation of covalent hydrazone linkage (Fig. 4).

Biotin-hydrazide was used for the coupling reaction, after which the product was purified via size-exclusion spin columns. Similar to the results obtained with labeling of SIRV2 at carboxylates, characterization using Western blot analysis confirmed that the biotin labels were attached to both the minor and major CP (Fig. 5A). The minor CP was detectable at its expected molecular weight of about 100 kDa. The major CP was retained at the top of the gel and did not migrate through the polyacrylamide matrix. It is possible that the highly glycosylated major CP was not fully derivatized with biotin, resulting in the availability of non-conjugated oxidized and activated carbohydrates that may react with the polyacrylamide matrix, thus preventing migration of the major CP through the gel matrix.

Full-decoration of SIRV2 by carboxylate labeling was also evident from immunogold-labeling TEM studies (Fig. 6). TEM imaging showed that gold labels derived from the anti-biotin antibody conjugates were attached to both the virus body and the end structures. In good agreement to labeling studies using carboxylate-selective bioconjugation, labels at the end structures were indicated to be associated with the tail fibers rather than with the plugs. Labeling using carbohydrate-chemistry was more intense when compared to carboxylate-selective labeling of SIRV2 (Fig. 6E). A reasonable explanation is that the carbohydrates on the surface of SIRV2 are shielding other surface-exposed amino acids, thus the carboxylates are not as solvent-exposed and accessible, and therefore do not react as readily as carbohydrates.

2.2.3. Addressable Amines

Labeling of SIRV2 using amine-selective chemistry was achieved using a succinimide-activated biotin as well as a fluorescent dye (AlexaFluor dye 568). Labeling reactions were carried out at pH 7.0; at this pH, amines derived from lysine side chains only are reactive toward the succinimide moiety (Fig. 4). After the labeling reaction, labeled particles were

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ADVANCED FUNCTIONAL MATERIALS

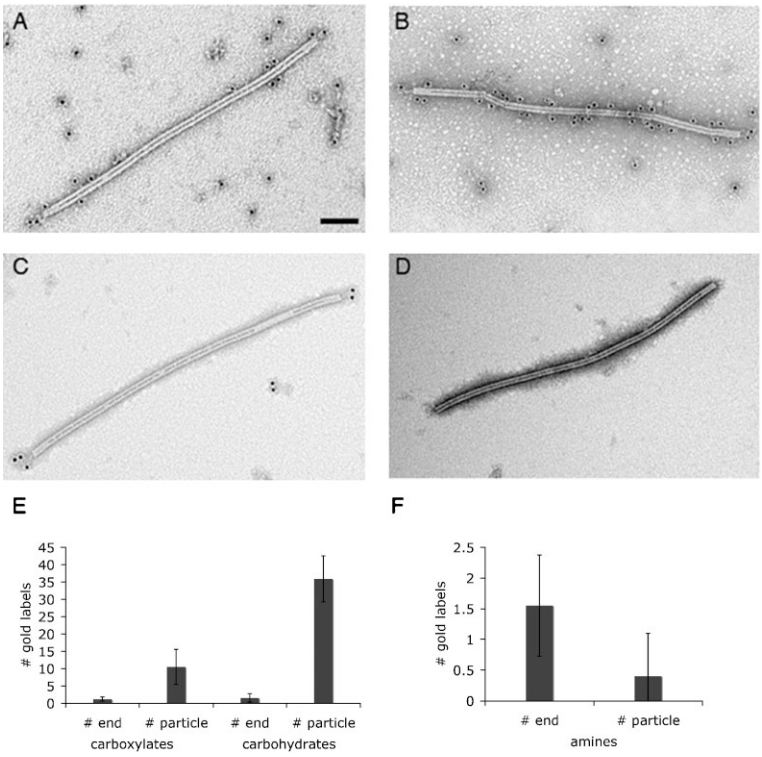


Figure 6. Immunogold staining of biotinylated SIRV2 particles using gold-labeled anti-biotin antibodies and TEM; A) biotinylated SIRV2 particles labeled using carboxylate-selective chemistry, B) biotinylated SIRV2 particles labeled at carbohydrates, C) biotinylated SIRV2 particles labeled using amine-selective bioconjugation techniques and D) non-modified SIRV2 was used as a control. E and F) Ten particles each were further analyzed and the gold labels per particle at the virus body and the virus end structures (tail fibers), respectively, counted. Note difference in scale of (E) and (F). F) Statistical analysis of differences between virus body and end labeling was performed using Student's two-tailed *t*-test (Microsoft Excel) where $p = 0.0078$.

purified from the reaction mixture using a size-exclusion spin column prior to further characterization.

In contrast to carboxylate- or carbohydrate-labeling, Western blot characterization of biotinylated SIRV2 indicated that amine-specific labeling occurred exclusively at the minor CP (Fig. 5A). This was consistent with results obtained with fluorescent-labeled SIRV2 particles as shown by visualization of the CPs on SDS gels under UV light (Fig. 5B, lane 2). Again the fluorescent dyes were exclusively attached to minor CP. Coomassie staining of the gels confirmed that both the minor and major CP were present at their expected position (Fig. 5B). Since, the minor CP contains a high number of lysine residues (73 lysines), decoration of SIRV2 at the minor CP using succinimide-activated biotin probes was anticipated. The major CP also contains lysine residues (6 lysines). However, Western blot studies of biotinylated particles as well as the characterization of dye-labeled SIRV2 indicate that in intact particles those lysines are not addressable. It is possible that the lysines are not reactive because they are not on the exterior solvent-exposed surface of the capsid. Another possible explanation is that the carbohydrates are shielding the lysines, thus preventing chemical bioconjugation.

Immunogold-staining and TEM imaging showed that the biotin labels coupled via lysines were attached exclusively to the end structures of SIRV2 (Fig. 6), selectively labeling the tail fibers. SIRV2 contains three tail fibers at each end of the particle, and the particles were labeled either with one, two, or three gold labels per virus end. This was also true for the tail fibers of particles labeled using carboxylate- and carbohydrate-labeling chemistries. The relative position of the gold-conjugated anti-biotin antibodies relative to the end structure of SIRV2, together with the finding that biotin labels were attached exclusively to the minor CP (see Western blot data) indicate that labeling is specific to the minor CP and that the minor CP forms the tail fibers. This is a highly interesting finding and demonstrates that bioconjugation techniques are not only useful in order to install functionalities, but also labeling techniques provide a useful tool in order to gain structural information.

3. Conclusions

In this paper, we showed that SIRV2 particles are stable in two different solvent/water mixtures that are of relevance for bioconjugation and mineralization reactions. Second, we showed that SIRV2 particles offer attachment sites allowing selective chemical modification. Both the major CP and minor CP possessed reactive carboxylates and carbohydrates as confirmed by Western blot and TEM. Interestingly, amine reactivity showed that the minor CP could be selectively labeled, and this labeling reaction targeted the ends of the particles only. These results suggest that SIRV2 particles have novel physical and chemical properties that provide routes to installing various functional molecules at different positions on the VNP. SIRV2 particles can thus be precisely programmed in a chemically selective and spatially controlled manner.

SIRV2 replicates in hyperthermophile and acidophile hosts, growing optimally at 80 °C and pH 3. Therefore, temperature stability and resistance to extremely low pH values is naturally given. In addition to its inherently stable nature, we have shown that the particles also withstand harsh solvent/water mixtures. SIRV2 remained intact and infectious in DMSO in concentrations up to 50% by volume for at least 6 days. Stability was also demonstrated in ethanol/water mixtures. At a concentration of 20% v/v ethanol no differences to particles stored in buffer were apparent for the monitored time period of 6 days. When particles were stored in a mixture containing 50% ethanol by volume, the ability to infect their host was altered after 2 days storage, after which the infectivity was reduced, but not eliminated. UV/Vis and TEM studies were also consistent with the particles remaining intact. This indicates that at this high ethanol concentration subtle changes occur that lead to

 © 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

impaired infectivity. This may be regarded as a minor issue when considering utilization of the particles for applications in materials science or nanotechnology.

Labeling studies of SIRV2 not only confirmed that the particles offer various and multiple attachment sites but also it was found that the particles may be decorated at different locations on the viral capsid (end vs. body). The studies provide novel insights into the structural organization of the particles and the location of the minor CP. Chemical bioconjugation studies revealed that biotin labels were successfully installed at carboxylates, carbohydrates, and amines, although labeling was different in each case. Covalent attachment of biotins at carboxylates and carbohydrates yielded particles displaying biotin moieties throughout the virion including virus body and virus end. Decoration of carbohydrates resulted in a more intense labeling when compared to particles labeled at carboxylates. This phenomenon might be explained by the glycosylation status of the particle itself. The carbohydrate shell will be the most solvent-exposed feature on the exterior surface of the virion and is thus expected to be most easily addressable. Carboxylates might be shielded by the carbohydrates, which may explain why these groups are less reactive.

The most interesting results were gained from particles labeled with biotin selectively at amines. In stark contrast to labeling carboxylates and carbohydrates, labeling of amines resulted in particles where the end structures, the tail fibers, were exclusively labeled; biotin molecules were attached exclusively to the minor CP. This not only demonstrates that functional molecules can be attached to SIRV2 in a spatially controlled manner using different chemistries, but also provides novel insights into the structural organization of the particles. It was hypothesized that the minor CP, which only accounts for less than 1% of the major CP, forms parts of the end structures of the virion, however its composition within the tail assembly had not been proven. Our results suggest that this minor CP is the structural component of the virus ends and that it forms the tail fibers. Further studies to confirm the data obtained from Western blots and immunogold-staining and TEM imaging will include mass spectrometry studies of labeled and non-labeled SIRV2. This is also expected to give detailed information about how many and which amino acids were chemically derivatized.

The fact that native SIRV2 offers three different chemical attachment sites is unique for rod-shaped VNPs. By comparison, native TMV particles only offer a single addressable tyrosine side chain per CP subunit on the solvent-exposed exterior surface.[16] Chemical bioconjugation methods using wild-type TMV are thus rather limited.

Besides offering multiple and various chemical attachment sites, the differential amine reactivity of the body and the ends of SIRV2 allows spatial control and is unique for rod-shaped VNPs. By comparison, the TMV rod is formed by a single CP subunit; thus native virions only offer a single addressable tyrosine side chain per CP subunit on the solvent-exposed exterior surface[16] and bioconjugation to CPs with spatial control is not achievable. In contrast to the rigid virions of TMV and SIRV2, filamentous virions of M13 are more complex and flexible. M13 is an asymmetric particle with two different ends that can be joined to form a circle, highlighting the flexibility of the particles.[17] Several virions may also be joined end-to-end to form lines or tetrapods.[18] Bioconjugation of M13 particles via chemical means has not been reported, instead binding of ligands to engineered peptides displayed on the CPs is more frequently employed.[6,19]

The symmetry, rigidity, and straightforward selective bioconjugation of SIRV2 ends suggest the utility of the particle for applications where identical functionalization of both ends is required. For example, SIRV2 is expected to be a suitable nanobuilding block and template for array formation using layer-by-layer assembly protocols. Since, virus body and ends can be selectively labeled it is expected that arrays with different physical properties can be fabricated, by for example aligning particles body-to-body versus self-assembly end-to-end. The latter might result in a porous structure that could then be further functionalized by attaching molecules to the virus body. Neither TMV nor M13 is expected to be applicable for such strategies; the TMV capsid does not offer addressable ends, and the flexibility of the M13 filament may preclude precise deposition and structural control. In addition, higher-order SIRV2 structures built by end-to-end joining to form rectangles or larger circles may be feasible. Similarly the combination of flexible M13 and rigid SIRV2 rods may be complementary for applications such as liquid crystal assembly, nanoscale templating, and electronic applications.

In summary, we have demonstrated the versatility of SIRV2 as a platform for chemical derivatization with potential use in materials science and nanotechnology. SIRV2, isolated from acidic geothermal environments, is an exceptionally robust bionanoparticle, stable at extremes of temperature and acidity as well as in diverse non-natural harsh conditions. Chemical conjugation of SIRV2 can be controlled selectively and spatially by the type of chemistry used, and the overall suitability of SIRV2 for use in materials science was demonstrated. SIRV2 may thus be regarded as a new, extremely stable, and structurally unique nanobuilding block. The work presented here is only the starting point to the next exciting approaches that may lead to new developments in nanobiotechnology at the virus/chemistry interface.

4. Experimental

Virus Growth and Purification: SIRV2 was propagated in *S. islandicus* strain LAL14/1. Viral particles were collected from 6 liters culture by polyethylene glycol (molecular weight 6000 Da) precipitation and purified by ultracentrifugation in a cesium chloride gradient, as described previously [20].

Stability of SIRV2: A highly concentrated suspension of SIRV2 in 20 mM Tris-actetate buffer pH 6.0 (in the following referred to as TA-buffer) and 0.45 mg mL^{-1} cesium chloride was dialyzed against MilliQ water using a MF-Millipore membrane (0.025 μm). Thirty microliter of the dialyzed suspension were added to 670 μL of ethanol/water and DMSO/water mixtures in proportions of 1:4 and 1:1. As a reference 30 μL SIRV2 suspension was added to 670 μL TA-buffer. Triplicates of each of the five different solvent conditions were stored at room

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

temperature and monitored using TEM, UV/Vis spectroscopy, and plaque assays at various time points over a time period of 6 days.

UV/Vis Spectroscopy: For each sample, the absorbance between 250 and 300 nm was measured in a 105.202-QS quartz cell (Hellma) using a UV-1700 PharmaSpec spectrophotometer (Shimadzu).

Plaque Assay: A soft layer of 0.3 g L^{-1} Gelrite gel (Duchefa Biochemie BV, The Netherlands) containing approximately 2×10^7 host cells (*S. islandicus* strain LAL14/1)/mL was layered over a 0.8 g L^{-1} Gelrite supporting gel. Serial dilutions ranging from 26 to 26×9^7 of the samples were prepared in 96-well plates. Five microliter droplets of each dilution were spotted on the polymerized soft layer. The plates were incubated at 80 °C for 48 h in order to determine the number of plaque forming units (pfu) in each sample.

Chemical Bioconjugation: Carboxylates derived from aspartic and glutamic acid on SIRV2 were labeled using biotin-LC-hydrazide (Pierce) in combination with EDC (Sigma). SIRV2 was dialyzed against 0.1 M MES buffer pH 5.5 using centrifugal devices with a cut-off of 10 kDa (Millipore). Biotin-hydrazide was dissolved in DMSO and added to SIRV2 to give a final concentration of 1.25 mM biotin-hydrazide and 20% v/v DMSO; biotin was used in a molar excess of approximately 10 000 biotin moieties per SIRV2. SIRV2 was used at a concentration of about 0.5–1 mg mL^{-1}. A final concentration of 6.5 mM EDC dissolved in MilliQ water was added in order to activate carboxylates. The reaction was allowed to proceed overnight at room temperature. Samples were purified prior to further characterization using centrifugal devices with a cut-off of 10 kDa. The columns were washed several times to make sure that all non-reacted chemicals were washed away. Biotinylated SIRV2 was resuspended in 10 mM potassium phosphate buffer pH 7.0 and stored at 4 °C until further characterization.

Labeling of carbohydrates was performed by oxidizing primary hydroxyls derived from carbohydrates to give aldehydes followed by reaction with biotin-LC-hydrazide (Pierce). An equal volume of ice cold sodium meta-periodate (Sigma) solution in 0.1 M sodium acetate buffer pH 5.5 and SIRV2 in 0.1 M sodium acetate buffer pH 5.5 were mixed on ice to give a final concentration of 10 mM sodium meta-periodate. The reaction vessel was protected from light and the mixture incubated for 30 min on ice. Excess of sodium meta-periodate was removed by dialysis against phosphate-buffered saline (PBS) pH 7.2 made from tablets (Sigma) using 10 kDa cut-off columns. A final concentration of 50–100 mM biotin-hydrazide dissolved in DMSO was added and the reaction allowed to proceed overnight at room temperature; biotin was used in a molar excess of approximately 10 000 biotin moieties per SIRV2. SIRV2 was used at a concentration of about 0.5–1 mg mL^{-1}. The mixture was purified using 10 kDa-cut-off spin columns (Millipore) as described above. Biotinylated SIRV2 was resuspended in 10 mM potassium phosphate buffer pH 7.0 and stored at 4 °C until further characterization.

Biotinylation at amines derived from lysine side chains in the SIRV2 CPs was carried out using sulfo-*N*-succinimidyl-6-biotinamidohexanoate sodium salt (Toronto Research Chemicals). The biotin moiety was dissolved in DMSO and added in a 10 000-fold molar excess to SIRV2 in 10 mM potassium phosphate buffer pH 7.0. The final DMSO concentration was 20% by volume. In order to allow the reaction to complete the mixture was allowed to react at 4 °C overnight. The mixture was purified using 10 kDa-cut-off spin columns (Millipore) as described above. Biotinylated SIRV2 was resuspended in 10 mM potassium phosphate buffer pH 7.0 and stored at 4 °C until further characterization.

SIRV2 was also labeled at amines using the amine-selective fluorescent dye AlexaFluor568-*N*-hydroxysuccinimide (Invitrogen). The bioconjugation protocol was as described for using sulfo-*N*-succinimidyl-6-biotinamidohexanoate sodium salt (see above).

Western Blot Analysis: To detect the biotin labels biotinylated SIRV2 samples (10–20 μg) were separated on 4–12% NuPage precast gels using MOPS buffer (Invitrogen). After separation of the CPs, the proteins were transferred onto nitrocellulose membranes (Millipore). After blocking with Tris-buffered saline (TBS) pH 7.6 containing 0.05% v/v Tween20 (Sigma) and 5% w/v dry milk, detection was carried out using alkaline phosphatase-conjugated streptavidin (1:1000 in TBS + Tween + milk). After 2 h incubation at room temperature with shaking the blot was washed several times with TBS + Tween20 prior to detection of alkaline phosphatase activity using BCIP/NBT liquid substrate system (Sigma).

Denaturing SDS-PAGE: Fluorescent dye-labeled SIRV2 samples (20–50 μg) were separated on 10% NuPage precast gels using MOPS buffer (Invitrogen). After separation of the viral CPs, the proteins were first visualized under UV/light and second stained with Coomassie blue.

Immungold-staining and TEM Imaging: Samples were diluted in PBS and allowed to settle on pyroxylin- and carbon-coated gold TEM grids for 20 min. Excess virus was washed off by rinsing briefly in PBS. Grids were placed in blocking buffer (1% v/v BSA-c/PBS/0.1% v/v Tween20) containing acetylated bovine serum albumin (BSA-c; Aurion, Wageningen, The Netherlands) for 30 min, then equilibrated briefly in incubation buffer (0.1% BSA-c/PBS/0.1% Tween20) before being incubated for 1 h with a goat anti-biotin antibody conjugated to 10 nm gold (British BioCell International, Agar Scientific Ltd., Stansted, UK) diluted 1:5 in incubation buffer. Grids were washed 4 times over 20 min in drops of PBS/0.01% Tween20 followed by two washes over 5 min in PBS and three washes over 10 min in distilled water. Finally, grids were negatively stained in either 2% w/v aqueous uranyl acetate or 2% w/v aqueous ammonium molybdate. Samples were imaged in a Jeol 1200 EX TEM (Jeol UK, Welwyn Garden City, UK) at 80 KV and TIFF files were stored using a Deben AMT 1.3k camera (Deben UK Ltd., Bury St. Edmunds, UK).

Received: May 24, 2008

[1] a) Z. Liu, C. Davis, W. Cai, L. He, X. Chen, H. Dai, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1410. b) M. E. Akerman, W. C. W. Chan, P. Laakkonen, S. N. Bhatia, E. Ruoslahti, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12617. c) J. Lim, E. E. Simanek, *Org. Lett.* **2008**, *10*, 201. d) S. K. Lai, K. Hida, C. Chen, J. Hanes, *J. Controlled Release* **2008**, *125*, 107. e) Y. Hu, T. Liwin, A. R. Nagaraja, B. Kwong, J. Katz, N. Watson, D. J. Irvine, *Nano Lett.* **2007**, *7*, 3056. f) O. Mertins, M. B. Cardoso, A. R. Pohlmann, N. P. Da Silveira, *J. Nanosci. Nanotechnol.* **2006**, *6*, 2425. g) H. S. Kang, J. E. Park, Y. J. Lee, I. S. Chang, Y. I. Chung, G. Tae, *J. Nanosci. Nanotechnol.* **2007**, *7*, 3944.

[2] a) N. F. Steinmetz, in, *Tomorrow's Chemistry Today*, (Ed: B. Pignataro), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany **2008**, pp. 215–236. b) N. F. Steinmetz, D. J. Evans, *Org. Biomol. Chem.* **2007**, *5*, 2891. c) P. Singh, M. J. Gonzalez, M. Manchester, *Drug Dev. Res.* **2006**, *67*, 23. d) M. Manchester, P. Singh, *Adv. Drug Delivery Rev.* **2006**, *58*, 1505. e) C. E. Flynn, S.-W. Lee, B. R. Peelle, A. M. Belcher, *Acta Mater.* **2003**, *51*, 5867. f) T. Douglas, M. Young, *Science* **2006**, *312*, 873.

[3] a) T. Douglas, M. Young, *Adv. Mater.* **1999**, *11*, 679. b) T. Douglas, M. Young, *Nature* **1998**, *393*, 152. c) T. Douglas, E. Strable, D. Willits, *Adv. Mater.* **2002**, *14*, 415.

[4] a) Q. Wang, T. Lin, J. E. Johnson, M. G. Finn, *Chem. Biol.* **2002**, *9*, 813. b) Q. Wang, E. Kaltgrad, T. Lin, J. E. Johnson, M. G. Finn, *Chem. Biol.* **2002**, *9*, 805. c) Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, M. G. Finn, *J. Am. Chem. Soc.* **2003**, *125*, 3192. d) N. F. Steinmetz, G. P. Lomonossoff, D. J. Evans, *Small* **2006**, *2*, 530. e) N. F. Steinmetz, E. Bock, R. P. Richter, J. P. Spatz, G. P. Lomonossoff, D. J. Evans, *Biomacromolecules* **2008**, *9*, 456. f) P. Singh, D. Prasuhn, R. M. Yeh, G. Destito, C. S. Rae, K. Osborn, M. G. Finn, M. Manchester, *J. Controlled Release* **2007**, *120*, 41. g) K. S. Raja, Q. Wang, M. J. Gonzalez, M. Manchester, J. E. Johnson, M. G. Finn, *Biomacromo-*

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

lecules **2003**, *3*, 472. h) K. S. Raja, Q. Wang, M. G. Finn, *ChemBioChem* **2003**, *4*, 1348. i) N. G. Portney, K. Singh, S. Chaudhary, G. Destito, A. Schneemann, M. Manchester, M. Ozkan, *Langmuir* **2005**, *21*, 2098. j) J. D. Lewis, G. Destito, A. Zijlstra, M. J. Gonzalez, J. P. Quigley, M. Manchester, H. Stuhlmann, *Nat. Med.* **2006**, *12*, 354. k) J. M. Hooker, E. W. Kovacs, M. B. Francis, *J. Am. Chem. Soc.* **2004**, *126*, 3718. l) C. L. Cheung, S. W. Chung, A. Chatterji, T. Lin, J. E. Johnson, S. Hok, J. Perkins, J. J. De Yoreo, *J. Am. Chem. Soc.* **2006**, *128*, 10801. m) A. Chatterji, W. Ochoa, L. Shamieh, S. P. Salakian, S. M. Wong, G. Clinton, P. Ghosh, T. Lin, J. E. Johnson, *Bioconjugate Chem.* **2004**, *15*, 807. n) E. A. Anderson, S. Isaacman, D. S. Peabody, E. Y. Wang, J. W. Canary, K. Kirshenbaum, *Nano Lett.* **2006**, *6*, 1160.

[5] R. J Tseng,, C. Tsai, L. Ma, J. Ouyang, C. S. Ozkan, Y. Yang, *Nat. Nanotechnol.* **2006**, *1*, 72.

[6] K. T. Nam, D. W. Kim, P. J. Yoo, C. Y. Chiang, N. Meethong, P. T. Hammond, Y. M. Chiang, A. M. Belcher, *Science* **2006**, *312*, 885.

[7] G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, Elsevier, London **1996**.

[8] D. Prangishvili, P. Forterre, R. A. Garrett, *Nat. Rev. Microbiol.* **2006**, *4*, 837.

[9] T. Basta, D. Prangishvili, in, *Thermophiles: Biology and Technology at High Temperatures*, (Eds: F. Robb, G. Antranikian, D. Grogan, A. Driessen), CRC Press, Boca Raton **2008**, pp. 225–236.

[10] D. Prangishvili, H. P. Arnold, D. Gotz, U. Ziese, I. Holz, J. K. Kristjansson, W. Zillig, *Genetics* **1999**, *152*, 1387.

[11] G. Vestergaard, M. Haring, X. Peng, R. Rachel, R. A. Garrett, D. Prangishvili, *Virology* **2005**, *336*, 83.

[12] H. Yi, G. W. Rubloff, J. N. Culver, *Langmuir* **2007**, *23*, 2663.

[13] J. W. Kehoe, B. K. Kay, *Chem. Rev.* **2005**, *105*, 4056.

[14] a) W. Shenton, T. Douglas, M. Young, G. Stubbs, S. Mann, *Adv. Mater.* **1999**, *11*, 253. b) C. E. Fowler, W. Shenton, G. Stubbs, S. Mann, *Adv. Mater.* **2001**, *13*, 1266.

[15] X. Peng, H. Blum, Q. She, S. Mallok, K. Brugger, R. A. Garrett, W. Zillig, D. Prangishvili, *Virology* **2001**, *291*, 226.

[16] T. L. Schlick, Z. Ding, E. W. Kovacs, M. B. Francis, *J. Am. Chem. Soc.* **2005**, *127*, 3718.

[17] K. T. Nam, B. R. Peele, S.-W. Lee, A. M. Belcher, *Nano Lett.* **2004**, *4*, 23.

[18] Y. Huang, C.-Y. Chiang, S. K. Lee, Y. Gao, E. L. Hu, J. De Yoreo, A. M. Belcher, *Nano Lett.* **2005**, *5*, 1429.

[19] a) C. Mao, C. Flynn, A. Hayhurst, R. Sweeney, J. Qi, G. Georgiou, B. Iverson, A. M. Belcher, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6946. b) S.-W. Lee, S. K. Lee, A. M. Belcher, *Adv. Mater.* **2003**, *9*, 689.

[20] W. Zillig, A. Kletzin, C. Schleper, I. Holz, D. Janekovic, J. Hain, M. Lanzendoerfer, J. K. Kristjansson, *Syst. Appl. Microbiol.* **1994**, *16*, 609.

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Submitted to ADVANCED FUNCTIONAL MATERIALS

Supporting information

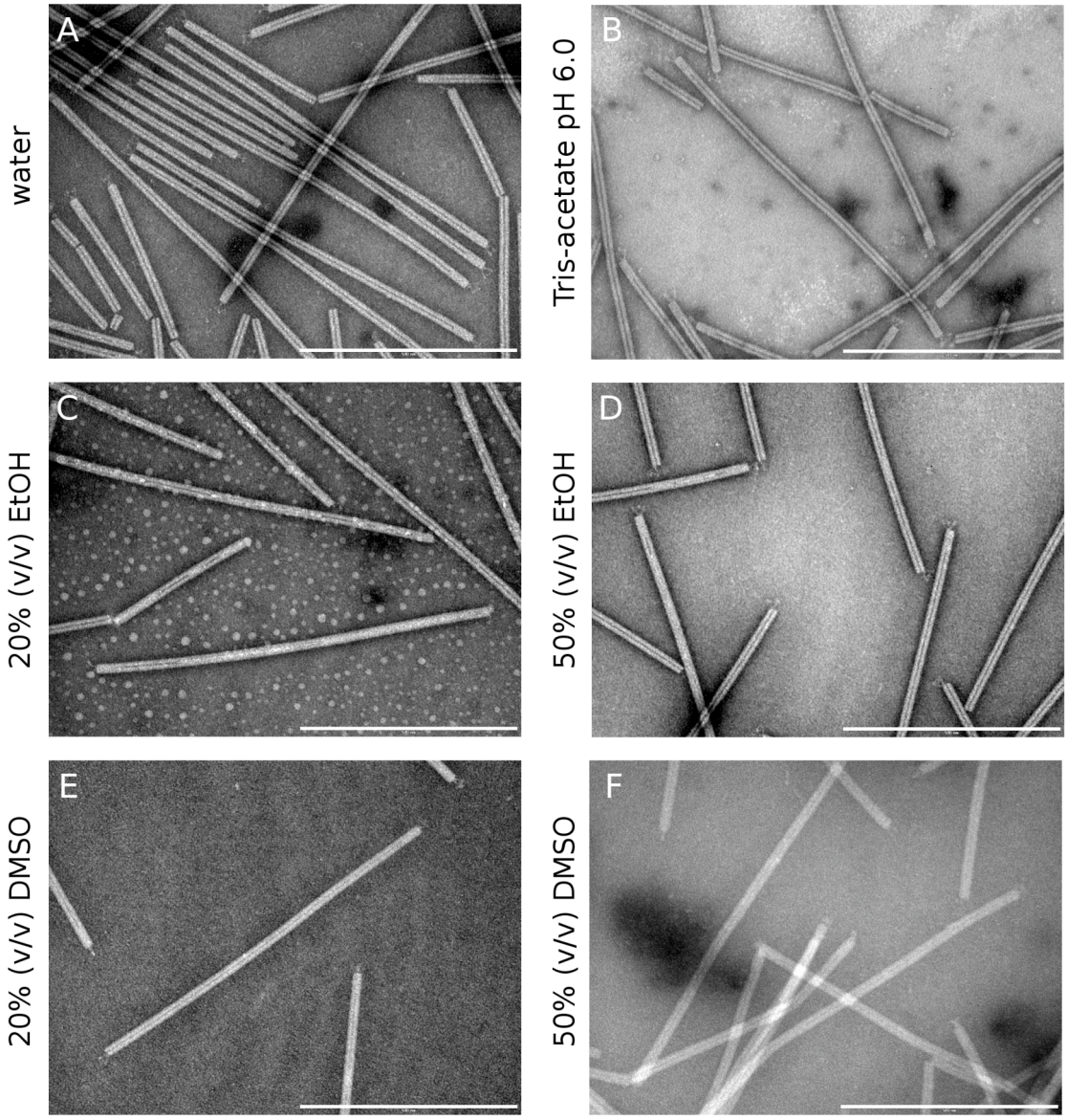
Site-specific and spatially controlled addressability of a new viral nanobuilding bock: *Sulfolobus islandicus* rod-shaped virus 2**

Nicole F. Steinmetz,[1] Ariane Bize,[2] Kim C. Findlay,[3] George P. Lomonossoff,[4] Marianne Manchester,[1] David J. Evans,[4] David Prangishvili[2]*

1. Transmission electron microscopy (TEM) of SIRV2 in solvent:water mixtures

The integrity of SIRV2 particles in different solvent:water mixtures was studied using TEM imaging. Ethanol and DMSO at concentration of 20 % and 50 % by volume were used. Particles were stored in these mixtures for six days at room temperature prior to TEM analysis. SIRV2 in 20 mM Tris-acetate buffer pH 6.0 was also monitored as a reference. TEM imaging confirmed the intactness of the particles under all conditions after six days. The structures were in all cases well preserved (Supporting Figure 1).

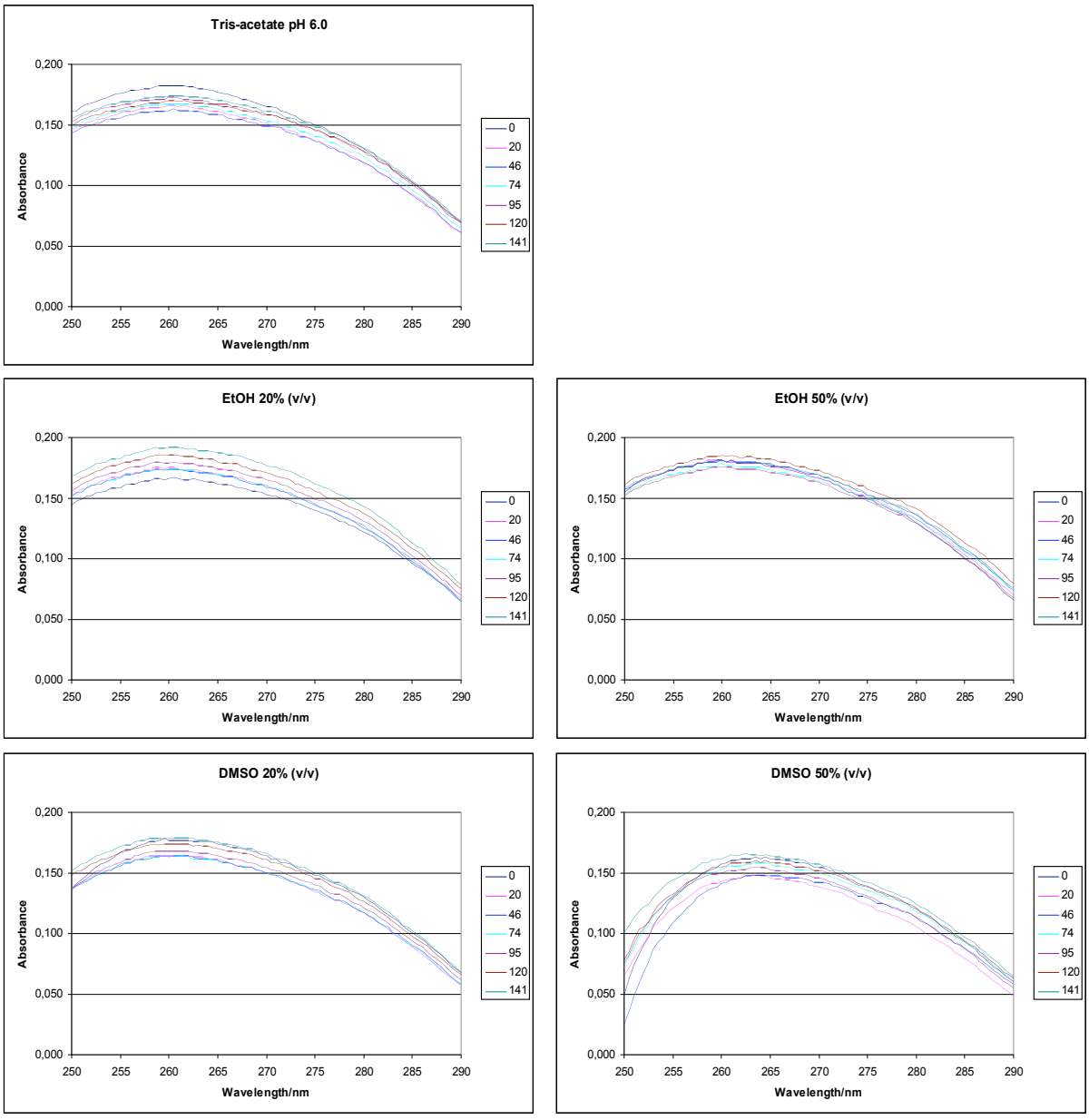
Experimental Section: Samples were deposited on 400 mesh carbon-coated copper grids (Ted Pella) and negatively stained with 2% (w/v) uranyl acetate, pH 4.5. They were observed in a JEOL 1200 EXII transmission electron microscope (Tokyo, Japan) at 120 kV accelerating voltage. Images were digitally recorded using an Eloise Mega View III Pro camera and the Analysis Pro software version 3.1 (Eloïse SARL, Roissy, France).



Supporting Figure 1. Transmission electron micrographs of negatively stained SIRV2 particles after storage for six days at room temperature in various solvents. A = SIRV2 in MilliQ water, B = 10 mM Tris-acetate buffer pH 6.0, C = 20 % (v/v) ethanol:water mixture, D= 50 % (v/v) ethanol:water mixture, E = 20 % (v/v) DMSO:water, and F = 50 % (v/v) DMSO:water (F). The scale bar is 500 nm.

2. UV/visible spectroscopy of SIRV2 in solvent:water mixtures

The stability of SIRV2 in 20 % (v/v) and 50 % (v/v) ethanol:water and DMSO:water mixtures was monitored over a time period of six days. Spectra were recorded at various time points; the spectra were consistent for all samples at all time points (Supporting Figure 2).



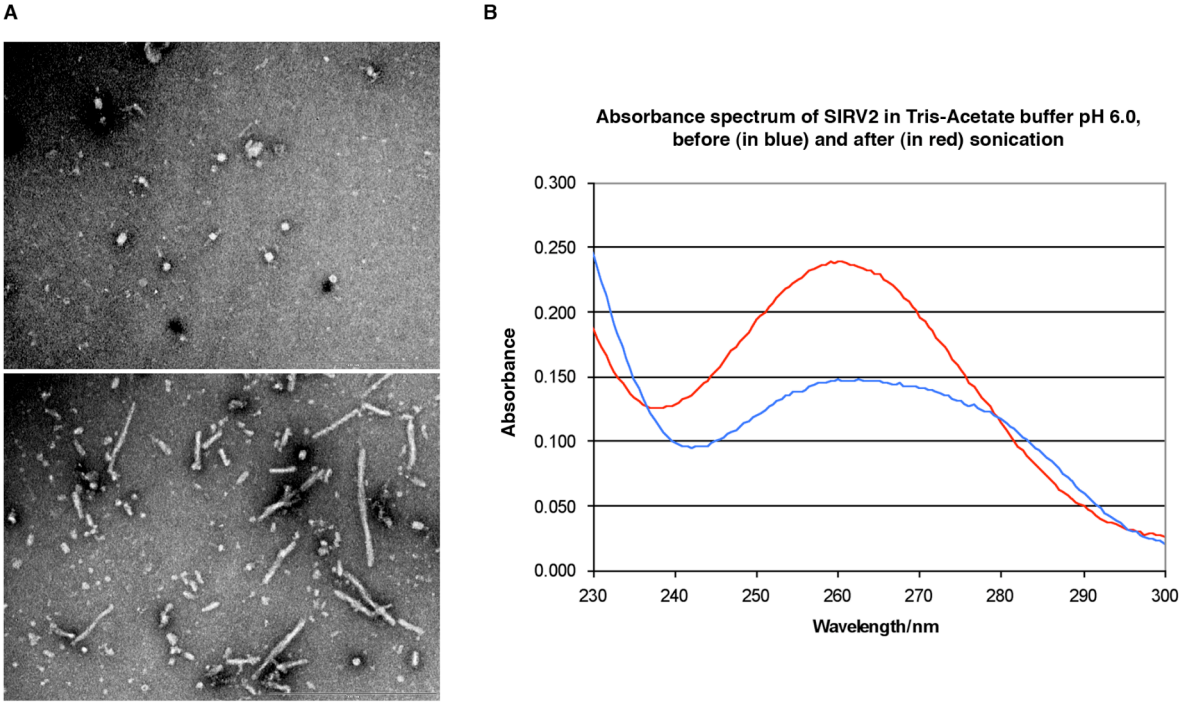
Supporting Figure 2. UV/visible spectra of SIRV2 particles stored for six days at room temperature in various solvents. Measurements were taken at t = 0 h, 20 h, 46 h. 74 h, 95, 120, 141 h after initial dilution of SIRV2 particles in the respective solvent:water mixture. The stability was monitored in 20 % and 50 % (v/v) ethanol:water and DMSO:water, respectively. SIRV2 in 20 mM Tris-acetate buffer pH 6.0 was also monitored as a reference.

3. Non-intact SIRV2 particles obtained by sonication

For comparison, non-intact SIRV2 particles were prepared by sonication. A concentrated virus suspension (500 μl) on ice was sonicated four times for 20 sec with an amplitude of about 12 μm. Samples were then analyzed for structural integrity and infectivity using TEM, UV/visible spectroscopy, and plaque assays.

TEM imaging showed disrupted, structurally non-intact, broken particles (Supporting Figure 3A). The UV/visible spectrum of intact SIRV2 *versus* non-intact (sonicated) SIRV2 particles is different; the spectrum of non-intact particles is shifted towards the shorter wavelength. The ratio of A260:A280 of intact particles lies at 1.26, values found for non-intact particles are higher and lie at 2.10 (Supporting Figure 3B). In addition, the PFU titers before and after sonication were determined. PFU titer before sonication, i.e. intact SIRV2 particles, was at the order of 1.8 $x10^{10}$ PFU/mL. After sonication the infectivity dropped significantly to 3.6 $x10^{4}$ PFU/mL.

All together, it was shown that sonicated SIRV2 particles were non-intact and less infectious. This also demonstrates the feasibility of the techniques used to determine whether SIRV2 is intact or not.



Supporting Figure 3. Assessment of the structural integrity of SIRV2 particles after sonication using TEM (panel A) and UV/visible spectroscopy (panel B). A. TEM of SIRV2 after sonication. The particles were negatively stained with 2% (w/v) uranyl acetate. The scale bar is 500 nm. B. UV/visible spectroscopy of intact SIRV2 (in blue) and non-intact SIRV2 samples after sonication (in red).

Stabilité des particules virales de SIRV2 à haute température

Afin d'estimer la stabilité des particules virales de SIRV2 en environnement chaud, différentes solutions contenant une même concentration initiale de virions ont été maintenues à 80°C. Un contrôle a été réalisé à température ambiante, en utilisant le tampon de stockage habituel des virus (Tris-acétate, pH 6,0). Le titre en particules infectieuses des différentes préparations a été déterminé au cours du temps (titration sur tapis cellulaire, comme indiqué dans la section Matériels et Méthodes de l'article précédent).

A 80°C, les conditions suivantes ont été testées : eau (« H2O ») ; solutions de sels identiques à celle utilisée pour le milieu de culture, à pH acide (« Sels pH 3,5 ») ou non-ajusté (« Sels) » ; Tris-acétate pH 6,0 (« TA ») ; milieu habituel de culture de pH compris entre 3,3 et 3,5 (« Milieu 3,3-3,5 »). Les solutions de sels ont été utilisées car le milieu de culture riche est instable à 80°C, à long terme.

Les résultats sont présentés dans le Tableau 6 et la Figure 18.

	TA RT	TA	Milieu 3,3-3,5	Sels pH 3,5	Sels	H2O
Demi-vie/Jour	240,64	0,50	0,81	1,24	2,82	0,38
R^2	0,08	0,99	0,98	0,89	0,88	0,96
Nb points	30	7	8	15	19	3

Tableau 6 : Stabilité des particules virales de SIRV2 dans différentes solutions. La demi-vie a été calculée grâce aux paramètres d'une régression exponentielle effectuée sur les données expérimentales (Figure 18). Nb points : indique le nombre de points expérimentaux disponibles pour la régression. TA RT : Tris-acétate pH 6,0 à température ambiante (Contrôle). Dans tous les autres cas, la température était de 80°C.

Aucune décroissance significative du titre n'a été observée pour le témoin après plus de 250 jours (Tableau 6 et Figure 18A). A 80°C, les demi-vies sont de l'ordre de 1 jour (Tableau 6). Les sels contribuent à une meilleure stabilité des particules, tandis que le pH acide est défavorable (Tableau 6 et Figure 18B). Ainsi, à 80°C, c'est dans l'eau que les particules sont les moins stables, dans les sels à pH non-ajusté qu'elles sont le plus stables.

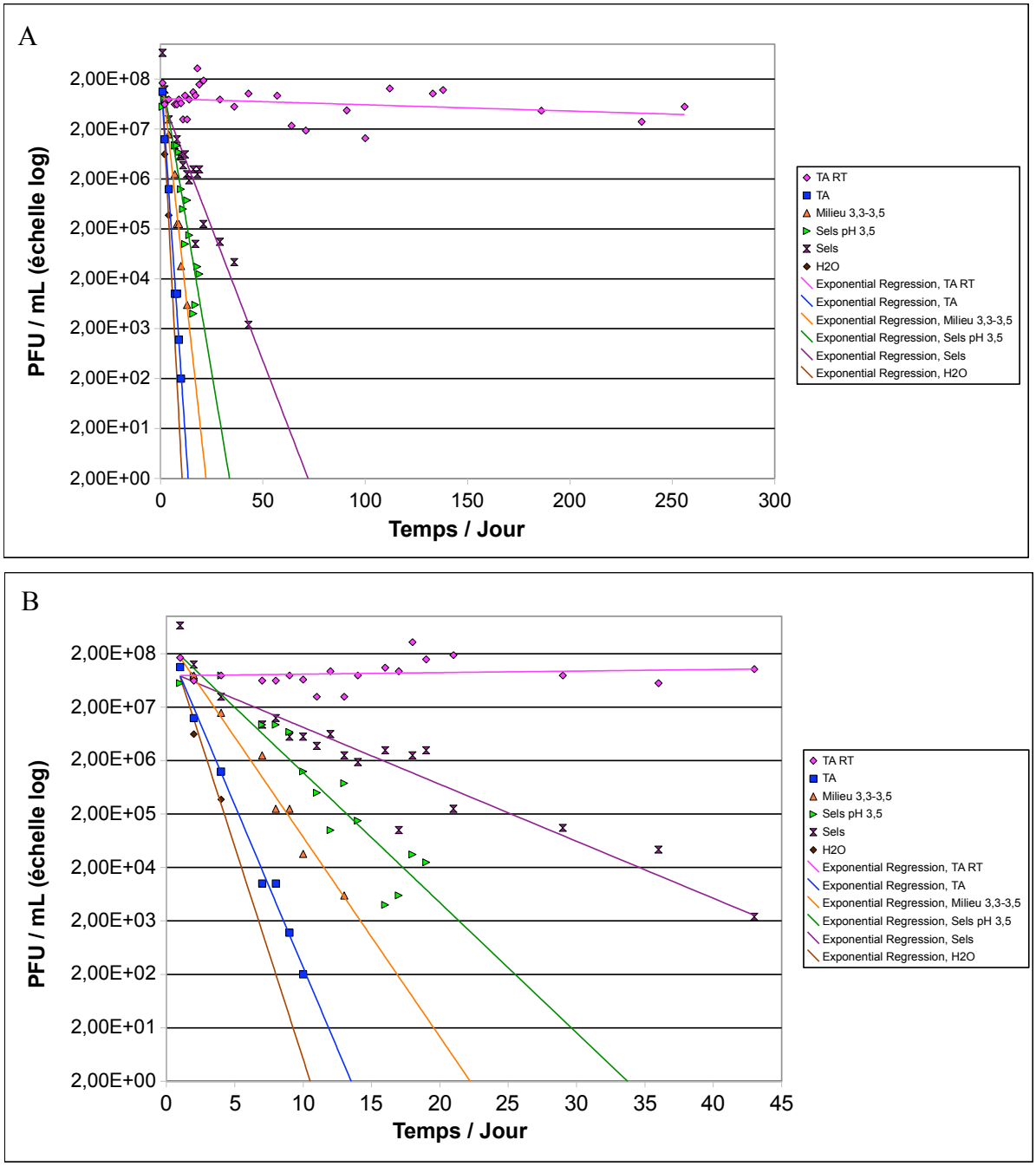


Figure 18 : Décroissance du titre de différentes préparations de virus en fonction du temps. (A) Totalité de l'expérience. (B) : Détail des 45 premiers jours de l'expérience. PFU/mL : unités virales infectieuses par millilitre. TA RT : Contrôle, dans Tris-acétate pH 6,0 à température ambiante. Toutes les autres préparations ont été maintenues à 80°C. TA : Tris-acétate pH 6,0.

JOURNAL OF BACTERIOLOGY, Oct. 2008, p. 6837–6845 Vol. 190, No. 20
0021-9193/08/$08.00+0 doi:10.1128/JB.00795-08
Copyright © 2008, American Society for Microbiology. All Rights Reserved.

Stygiolobus Rod-Shaped Virus and the Interplay of Crenarchaeal Rudiviruses with the CRISPR Antiviral System▽†

Gisle Vestergaard,[1] Shiraz A. Shah,[1] Ariane Bize,[2] Werner Reitberger,[3] Monika Reuter,[3] Hien Phan,[1] Ariane Briegel,[4] Reinhard Rachel,[3] Roger A. Garrett,[1] and David Prangishvili[2]*

Danish Archaea Centre and Centre for Comparative Genomics, Department of Biology, Copenhagen University, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark[1]; Molecular Biology of the Gene in Extremophiles Unit, Institut Pasteur, rue Dr. Roux 25, 75724 Paris Cedex 15, France[2]; Department of Microbiology, University of Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany[3]; and Max-Planck-Institut of Biochemistry, Molecular Structural Biology, Am Klopferspitz 21, D-82152 Martinsried, Germany[4]

Received 6 June 2008/Accepted 11 August 2008

A newly characterized archaeal rudivirus *Stygiolobus* rod-shaped virus (SRV), which infects a hyperthermophilic *Stygiolobus* species, was isolated from a hot spring in the Azores, Portugal. Its virions are rod-shaped, 702 (± 50) by 22 (± 3) nm in size, and nonenveloped and carry three tail fibers at each terminus. The linear double-stranded DNA genome contains 28,096 bp and an inverted terminal repeat of 1,030 bp. The SRV shows morphological and genomic similarities to the other characterized rudiviruses *Sulfolobus* rod-shaped virus 1 (SIRV1), SIRV2, and *Acidianus* rod-shaped virus 1, isolated from hot acidic springs of Iceland and Italy. The single major rudiviral structural protein is shown to generate long tubular structures in vitro of similar dimensions to those of the virion, and we estimate that the virion constitutes a single, superhelical, double-stranded DNA embedded into such a protein structure. Three additional minor conserved structural proteins are also identified. Ubiquitous rudiviral proteins with assigned functions include glycosyl transferases and a *S*-adenosylmethionine-dependent methyltransferase, as well as a Holliday junction resolvase, a transcriptionally coupled helicase and nuclease implicated in DNA replication. Analysis of matches between known crenarchaeal chromosomal CRISPR spacer sequences, implicated in a viral defense system, and rudiviral genomes revealed that about 10% of the 3,042 unique acidothermophile spacers yield significant matches to rudiviral genomes, with a bias to highly conserved protein genes, consistent with the widespread presence of rudiviruses in hot acidophilic environments. We propose that the 12-bp indels which are commonly found in conserved rudiviral protein genes may be generated as a reaction to the presence of the host CRISPR defense system.

Viruses of the hyperthermophilic crenarchaea are extremely diverse in their morphotypes and in the properties of their double-stranded DNA (dsDNA) genomes (reviewed in references 19 and 23). Moreover, some of the virion morphotypes are unique for dsDNA viruses from any domain of life. Many of these viruses have been classified into seven new families that include rod-shaped rudiviruses, filamentous lipothrixviruses, spindle-shaped fuselloviruses, and a bottle-shaped ampullavirus (reviewed in reference 24). The bicaudavirus *Acidianus* two-tailed virus (ATV) exhibits an exceptional two-tailed morphology and the unique viral property of developing long tail-like appendages independently of the host cell (11). Crenarchaeal viral research is still at an early stage of development, and insights into basic molecular processes, including infection, replication, packaging, and virus-host interactions, are limited. One of the main reasons for this lies in the high proportion of predicted genes with unknown functions (25).

At present, viruses of the family *Rudiviridae* are the most promising for detailed studies because they can be obtained in reasonable yields, and there are already some insights into their mechanisms of replication, transcriptional regulation, and host cell adaptation (4, 12, 13, 20, 21). To date, three rudiviruses have been characterized, all from the order *Sulfolobales*: the closely related *Sulfolobus* rod-shaped virus 1 (SIRV1), and SIRV2, isolated on Iceland, which infect strains of *Sulfolobus islandicus* (20, 22), and *Acidianus* rod-shaped virus 1 (ARV1), isolated at Pozzuoli, Italy, which propagates in *Acidianus* strains (34). Moreover, rudivirus-like morphotypes and partial rudiviral genome sequences have been detected in environmental samples collected from both acidic and neutrophilic hot aquatic sites (27, 29, 32).

All rudiviral genomes carry linear dsDNA genomes with long inverted terminal repeats (ITRs) ending in covalently closed hairpin structures with 5′-to-3′ linkages (4, 20). The terminal structure is important for replication, which presumably is initiated by site-specific single-strand nicking within the ITR, with the subsequent formation of head-to-head and tail-to-tail intermediates, and the conversion of genomic concatemers into monomers by a virus-encoded Holliday junction resolvase (20). This basic replication mechanism appears to be similar to that used by the eukaryal poxviruses, *Chlorella* virus and African swine fever virus, although there is no clear similarity between the se-

* Corresponding author. Mailing address: Molecular Biology of the Gene in Extremophiles Unit, Institut Pasteur, rue Dr. Roux 25, 75724 Paris Cedex 15, France. Phone: 33-(0)144-38-9119. Fax: 33-(0)145-68-8834. E-mail: prangish@pasteur.fr.

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

▽ Published ahead of print on 22 August 2008.

Downloaded from jb.asm.org at INSTITUT PASTEUR-Médiathèque on October 6, 2008

quences of the implicated archaeal and eukaryal proteins (20, 25).

The transcriptional patterns of rudiviruses SIRV1 and SIRV2 are relatively simple, with few temporal expression differences. An exception is the gene encoding the major structural protein that binds to DNA and, at an early stage of infection, is expressed as a polycistronic mRNA but appears as a single gene transcript close to the eclipse period (12). It has also been shown that rudiviral transcription can be activated by a *Sulfolobus* host-encoded protein, Sta1, that interacts specifically with TATA-like promoter motifs in the viral genome (13).

For SIRV1, a detailed study of the mechanism of adaptation to foreign hosts was conducted. Upon passage of the virus through closely related *S. islandicus* strains, complex changes were detected that were concentrated within six genomic regions (21, 22). These changes included insertions, deletions, gene duplications, inversions, and transpositions, as well as changes in gene sizes that often involved the insertion or deletion of what appeared to be "12-bp elements." It was concluded that the virus generated a complex mixture of variants, one or more of which were preferentially propagated when the virus entered a new host (21).

Here we describe a novel rudivirus, *Stygiolobus* rod-shaped virus (SRV), isolated from the Azores, Portugal, a location geographically distant from the locations of the other characterized rudiviruses (20, 34). SRV shows sufficient differences from the other rudiviruses, both morphologically and genomically, to warrant its classification as a novel species. The structural and genomic properties of the rudiviruses are compared and contrasted, and new data on the conserved virion structural proteins are presented. Different rudiviruses were selected for these studies on the basis of the virion or protein yields that were obtained. Moreover, matches between the spacer regions of the crenarchaeal chromosomal CRISPR repeat clusters, which have been implicated in a viral defense system (18) involving processed RNA transcribed from one DNA strand (reviewed in references 16 and 17), and the rudiviral genomes are analyzed and their significance, and possible relationships to the 12-bp indels, are considered.

MATERIALS AND METHODS

Enrichment culture, isolation of viral hosts, and virus purification. An environmental sample was taken from a hot acidic spring (93°C, pH 2) in the Furnas Basin on Saõ Miguel Island, the Azores, Portugal. The aerobic enrichment culture was established from the environmental sample and maintained at 80°C under conditions described previously for cultivation of members of the *Sulfolobales* (35). Single strains were isolated by plating on Gelrite (Kelco, San Diego, CA) containing colloidal sulfur (35) and grown in the medium of the enrichment culture. Cell-free supernatants of cultures were analyzed by transmission electron microscopy for the presence of virus particles.

SRV was isolated from the growth culture of its host strain *Stygiolobus* sp., which was colony purified as described above. After cells were grown to the late exponential phase and harvested by low-speed centrifugation (Sorvall GS3 rotor) (4,500 rpm), virions were precipitated from the supernatant by adding NaCl (1 M) and polyethylene glycol 6000 (10% [wt/vol]) and maintaining the mixture at 4°C overnight. They were purified further by CsCl gradient centrifugation (34).

Transmission electron microscopy. Samples were deposited on carbon-coated copper grids, negatively stained with 2% uranyl acetate (pH 4.5), and examined in a CM12 transmission electron microscope (FEI, Eindhoven, The Netherlands) operated at 120 keV. The magnification was calibrated using catalase crystals negatively stained with uranyl acetate (28). Images were digitally recorded using a slow-scan charge-coupled-device camera connected to a PC running TVIPS software (TVIPS GmbH, Gauting, Germany). To some samples, 0.1% sodium dodecyl sulfate (SDS) was added, and those samples were maintained at 22°C for 30 min in order to study the stability of the virion particles. Electron tomography of intact, negatively stained virions was performed as described previously (10, 26). Visualization of the three-dimensional (3D) data was performed using Amira software (Visage Imaging, Fürth, Germany).

Protein analyses. Proteins of SRV were separated in 13.5% SDS–polyacrylamide gels (14) and stained with Coomassie brilliant blue R-250 (Serva, Heidelberg, Germany). N-terminal protein sequences were determined by Edman degradation using a Procise 492 protein sequencer (Applied Biosystems, Foster City, CA).

SIRV2 proteins were separated in 4 to 12% SDS–polyacrylamide NuPAGE gradient gels by the use of MES (morpholineethanesulfonic acid) buffer (both from Invitrogen, Paisley, United Kingdom). The gels were stained with Sypro Ruby (Invitrogen). Protein bands were analyzed by peptide mass fingerprinting with matrix-assisted laser desorption ionization–time of flight mass spectrometry using a Voyager DE-STR biospectrometry workstation (Applied Biosystems, Framingham, MA) as described earlier (26). The analysis was performed in conjunction with the proteomic platform at the Pasteur Institute.

Cloning and heterologous expression of ARV1-ORF134b and purification of the recombinant protein and its self-assembly. ARV1-ORF134b was amplified from purified viral DNA with primers ARV1ORF134F (GGAATTCCATATG ATGGCGAAAGGACACACACC) and ARV1ORF134R (GGAATTCTCGA GACTTACGTATCCGTTAGGAC). The PCR product was purified (PCR purification kit; Roche, Mannheim, Germany) and cloned into pET30a expression vector (Novagen, Madison, WI) between restriction sites for EcoRI and XbaI. The protein was expressed overnight at 20°C in the *Escherichia coli* Rosetta(DE3)pLysS strain. Protein expression was controlled by SDS-polyacrylamide gel electrophoresis analysis and by performing a Western blot analysis using anti-His-tag-specific antibodies (Novagen). The native protein was purified on a Ni^{2+}-nitrilotriacetic acid (Ni^{2+}-NTA)-agarose column (Novagen) with elution buffers containing 50 to 500 mM imidazole. The accuracy of its sequence was confirmed. Self-assembly of the recombinant protein into filamentous structures was performed at 75°C and pH 3.5 and observed by electron microscopy.

Preparation of cellular and viral DNA and DNA sequencing. DNA was extracted from *Stygiolobus azoricus* cells as described previously (2), and the 16S rRNA gene was amplified by PCR using primers 8aF and 1512 uR (6) and sequenced.

Viral DNA was obtained by disrupting SRV particles with 1% SDS for 1 h at room temperature and extraction with phenol-chloroform (9). A shotgun library was prepared by sonicating viral DNA to generate fragments of 2 to 4 kb and cloning these into the SmaI site of the pUC18 vector. DNA was purified from single colonies by the use of a Biorobot 8000 workstation (Qiagen, Westburg, Germany) and sequenced in MegaBACE 1000 sequenators (Amersham Biotech, Amersham, United Kingdom). The viral sequence was assembled using Sequencher 4.2 software (Gene Code, Ann Arbor, MI). PCR primers for gap closing and resolving sequence ambiguities were designed using Primers for Mac, version 1.0. Sequence alignments were obtained using MUSCLE software (7). Open reading frames (ORFs) were defined with the help of ARTEMIS software (30) and investigated in searches using the EMBL and GenBank (1), 3D-Jury (8), and SMART (15) databases. Genome maps were generated and compared using Mutagen software, version 4.0 (5).

Bioinformatical matching of crenarchaeal CRISPR spacers to rudiviral genomes. CRISPRs were predicted for each of the 14 publicly available crenarchaeal genomes in GenBank (NC_000854 [*Aeropyrum pernix* K1], NC_002754 [*Sulfolobus solfataricus* P2], NC_003106 [*Sulfolobus tokodaii* strain 7], NC_003364 [*Pyrobaculum aerophilum* strain IM2], NC_007181 [*Sulfolobus acidocaldarius* DSM 639], NC_008698 [*Thermofilum pendens* Hrk5], NC_008701 [*Pyrobaculum islandicum* DSM 4184], NC_008818 [*Hyperthermus butylicus* DSM 5456], NC_009033 [*Staphylothermus marinus* F1], NC_009073 [*Pyrobaculum calidifontis* JCM 11548], NC_009376 [*Pyrobaculum arsenaticum* DSM 13514], NC_009440 [*Metallosphaera sedula* DSM 5348], NC_009676 [*Cenarchaeum symbiosum*], and NC_009776 [*Ignicoccus hospitalis* KIN4/I]). In addition, the six sequenced repeat clusters from *Sulfolobus solfataricus* P1 (16) were added to the data set as well as CRISPRs from five incomplete *Sulfolobus islandicus* genomes publicly available through the Joint Genome Institute (http://genome.jgi.doe.gov/mic_asmb.html) and unpublished genome sequences of *Sulfolobus islandicus* HVE10/4 and *Acidianus brierleyi* from the Copenhagen laboratory. The repeat cluster sequences were found using publicly available software (3, 7).

All predictions were curated manually. The orientation of each repeat cluster was inferred from the repeat sequence and by locating the low-complexity flanking sequence that generally resides immediately upstream from the cluster and contains the transcriptional leader (16). All unique spacer sequences of the

Downloaded from jb.asm.org at INSTITUT PASTEUR-Médiathèque on October 6, 2008

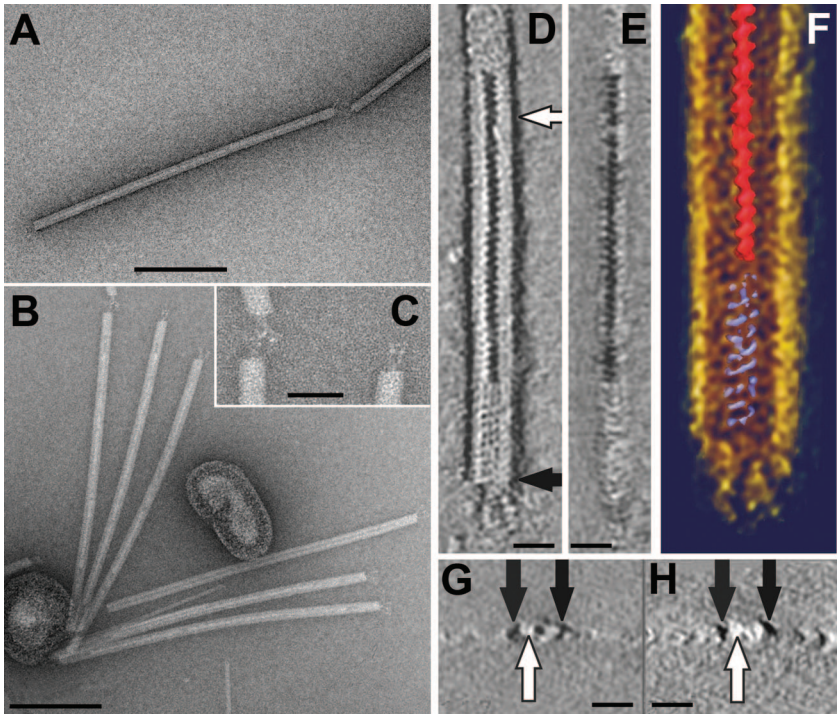


FIG. 1. Electron micrographs of SRV virions negatively stained with 3% uranyl acetate. (A) A full virion particle, with a discontinuous central line along the virion. (B) Six virions attached to liposome-like structures. (C) Enlargement of a portion of panel B displaying the terminal fibers. (D to H) Electron tomography images of an SRV virion. (D) Horizontal *x-y* slice (0.7 nm) showing the accumulated stain in the central part of the virion (white arrow). (E) Vertical *y-z* slice (0.7 nm) through the 3D data set of the reconstructed part of an SRV particle. (F) Visualization of the 3D data set using Amira software. (G and H) Vertical *x-z* slice (0.7 nm) through the tomogram showing that the virion particles are embedded in negative stain and that accumulated stain visible in panel D is absent from the plug (black arrows). Bars, 200 nm (A and B); 50 nm (C); 20 nm (D, E, G, and H).

repeat clusters, corresponding to the processed spacer transcript sequence (16), were aligned to the complete nucleotide sequences on each strand of all four rudiviral genomes (SRV [accession no. FM164764], SIRV1 [AJ414696], SIRV2 [AJ344259], and ARV1 [AJ875026]) by use of Paralign, an MMX-optimized implementation of the Smith-Watermann algorithm (31). Moreover, assuming that the spacer DNA can be incorporated into the oriented CRISPRs in either direction, we also translated the two strands of the spacer DNA into all the reading frames, yielding six amino acid sequences per spacer. Reading frames containing stop codons (ca. 50%) were omitted to make the subsequent search more specific. Each translation was aligned against the amino acid sequences of all the annotated ORFs in each of the four rudiviral genomes. Significant e-value cutoffs were determined for both the nucleotide and amino acid sequence searches using the genome sequence of *Saccharomyces cerevisiae* as a negative control (data not shown).

RESULTS

SRV isolation and structure. The virus-producing strain was colony purified from an enrichment culture established from a sample collected from an acidic hot spring in the Azores (see Materials and Methods). Its 16S rRNA sequence represented the genus *Stygiolobus* of the *Sulfolobales* crenarchaeal order and was closely related to that of *Stygiolobus azoricus*. However, it differs from *S. azoricus*, the type species of the genus, in its capacity to grow aerobically, and a description of the new species is in preparation. The virus particles produced constituted flexible rods 702 (± 50) by 22 (± 3) nm in size, with three short fibers at each terminus (Fig. 1A to C; Table 1). A Fourier analysis of the virion (not shown) revealed the presence of regular features with a periodicity of $(4.2\ \text{nm})^{-1}$, which probably reflect a helical subunit arrangement. This feature is also seen in the tomographic data set (Fig. 1D to H), which revealed more structural details. The helical arrangement in the virion core occurs in two different configurations. In the central region, a zigzag structure with dark contrast, probably arising from uranyl acetate staining, is surrounded by a protein shell (Fig. 1D and E). In contrast, in the terminal plug, which is

TABLE 1. Properties of the rudiviruses

Rudivirus	Origin	Virion length (nm)	Genome size (bp)	Total no. of ORFs	G+C (%)	ITR length (bp)	Reference
SRV	Azores	702	28,097	37	29.3	1,030	
ARV1	Pozzuoli	610	24,655	41	39.1	1,365	34
SIRV1	Iceland	830	32,308	45	25.3	2,032	20
SIRV2	Iceland	900	35,498	54	25.2	1,626	20

Downloaded from jb.asm.org at INSTITUT PASTEUR-Médiathèque on October 6, 2008

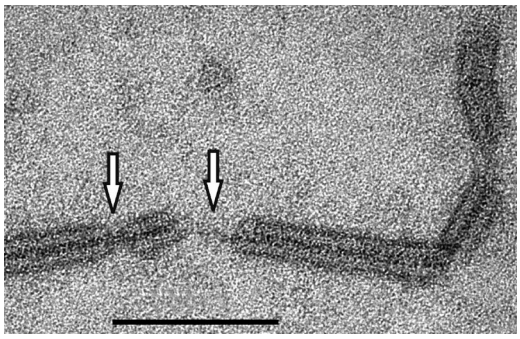


FIG. 2. Electron micrograph of a portion of an SRV virion after treatment with 0.1% SDS for 30 min (see Materials and Methods). White arrows indicate DNA or DNA-protein fibers lacking the protein core. Bar, 100 nm.

about 50 nm in length, a helically arranged protein mass, with no obvious uranyl acetate inclusions, is seen (Fig. 1D to F). The three terminal fibers, anchored in the plug-like structure, appear to be built up of multiple subunits ordered in a linear array (Fig. 1D). The side view of the reconstructed virion particle (Fig. 1E), as well as cross-sections of the negatively stained virions obtained from the tomograms (Fig. 1G and H), shows that the virion particles are embedded in negative stain (Fig. 1G and H) and partially collapsed due to staining and air drying; the height of the particles was about half of the apparent diameter. Nevertheless, the accumulated central stain is clearly visible in the cross-section (Fig. 1G) of the central part of the virion (Fig. 1D), while this feature is absent from the plug (Fig. 1G and H). The rod-shaped morphology of SRV, with a regular helical core and tail fibers, is characteristic of rudiviruses.

To investigate further the fine structure of the virion, virion particles were incubated in buffer containing 0.1% SDS for 30 min at 22°C. Most of the virion remained undisturbed, with the particles showing the same diameter as native virions and the densely stained, helical core. However, in local regions the protein shell had dissociated (Fig. 2) and a fine fiber with a diameter of 3 to 4 nm that constituted either naked DNA or a DNA-protein complex was visible.

Self-assembly of the major coat protein. The major rudiviral structural protein is highly conserved in sequence and is glycosylated (20, 22, 32a, 34). In order to study its possible self-assembly properties, the ARV1 protein (ORF134b [34]) was expressed heterologously in *E. coli* (see Materials and Methods) and a His-tagged protein was purified to homogeneity on an Ni^{2+}-NTA-agarose column. The protein was shown by transmission electron microscopy to self-assemble to produce filamentous structures of uniform widths and different lengths (Fig. 3). The optimal conditions for the assembly, 75°C and pH 3, were close to those of the natural environment, and no additional energy source was required for this process. The transmission electron microscopy analysis revealed that the filaments had structural parameters similar to those of the native virions, with a diameter of 21 ($\pm$ 3) nm and a periodicity of $(4.2\ \text{nm})^{-1}$. Thus, the data suggest that the single major coat protein alone can generate the body of the virion.

Minor rudiviral virion proteins. To date, the major coat protein is the only rudiviral structural protein to have been characterized. Given the closely similar structures of the different rudiviruses, we attempted to identify minor structural proteins for the SIRV2 virus, which can be produced in high yields. Protein components of SIRV2 virions, separated on a polyacrylamide gel, yielded six distinct major bands (Fig. 4), and all except D2, which is the strongest band and corresponds to ORF134 (gp26), were analyzed by mass spectrometry. Their identities were as follows: band A contained ORF1070 (gp38), band B contained ORF488 (gp33), and band C contained ORF564 (gp39), while bands D1 and D3 both contained ORF134 (gp26), probably in a glycosylated or, in the case of D3, a proteolytically degraded form. Thus, three additional SIRV2 structural proteins were identified, each highly conserved in sequence in all rudiviruses (Table 2).

SRV genome content. A shotgun library of the viral genome was prepared, sequenced, and assembled (see Materials and Methods) to yield an approximately 10-fold coverage of a 26-kb contig. Since 1 to 2 kb of terminal sequence is always absent from shotgun libraries of linear viral genomes, these additional sequences were generated by primer walking using viral DNA, or using PCR products obtained therefrom, until subsequent rounds of walking yielded no further sequence.

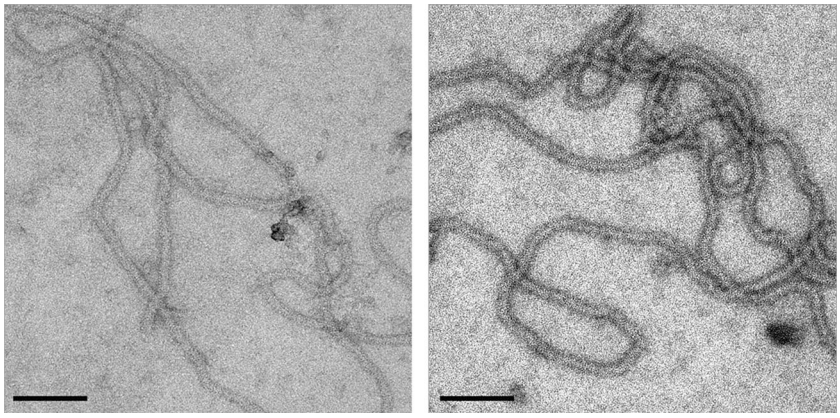


FIG. 3. Electron micrograph images of the self-assembled major coat protein of ORF134 from ARV1 after negative staining with 3% uranyl acetate. Bar, 100 nm.

Downloaded from jb.asm.org at INSTITUT PASTEUR-Médiathèque on October 6, 2008

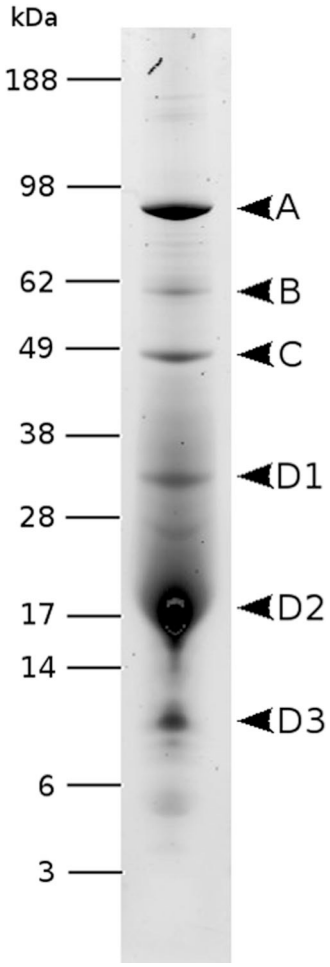


FIG. 4. SIRV2 virion proteins separated by SDS-polyacrylamide gel electrophoresis and stained with Sypro Ruby. Molecular masses of protein standards are indicated in kilodaltons on the left.

The total sequence obtained was 28,096 bp, with a G+C content of 29% and an ITR of about 1,030 bp (Table 1). An EcoRI restriction digest yielded fragments consistent with the genome size (data not shown).

Thirty-seven ORFs were predicted for which start codons were assigned on the basis of the upstream locations of TATA-like and transcription factor B-responsive element (BRE) promoter motifs and/or Shine-Dalgarno motifs. Details of the putative genes and operon structures are presented in Table S1 in the supplemental material, and a comparative genome map of SRV and rudiviruses SIRV1 and ARV1 is presented in Fig. 5; the genome map of SIRV2, which is closely similar to that of SIRV1, is not included (12, 20). SRV differs from the other rudiviruses in that fewer ORFs are organized in operons, and it has a lower level of gene order conservation (Fig. 5). Moreover, whereas for the other rudiviruses TATA-like motifs are often directly preceded by a conserved GTC triplet (12, 20, 34), in SRV the ensuing triplet sequence was GTA for 10 of the 30 putative TATA-like motifs (see Table S1 in the supplemental material).

Homologs of 17 SRV ORFs are present in all rudiviruses, and a further 10 SRV ORFs are conserved in some rudiviruses. Each virus type carries a few genes which are unique, and these are generally clustered near the ends of the linear genomes and yield no matches to genes in public sequence databases. In SRV, these are ORF145, -116a, -109, -59, -108, -97b, and -92 (left to right in Fig. 5). Although for SIRV1 and SIRV2 some of these nonconserved ORFs have been shown to be transcribed (12), further work is necessary to establish whether they are all protein-coding genes. Some of the proteins carry predicted structural motifs, and putative functions could be assigned to some of the conserved ORFs on the basis of public database searches; most of these are encoded in other crenarchaeal genomes (Table 2).

The host-encoded transcriptional regulator Sta1, a winged helix-turn-helix protein, was shown to bind to some SIRV1 promoters, including those of ORF134 and ORF399, and to enhance their transcription (13). A similar regulation may occur also for SRV, since the promoter regions of the homologs of ORF134 and ORF399 contain putative Sta1 binding sites. In contrast, in ARV1 only the ORF134 homolog is present in an operon for which the first ORF is a putative transcriptional regulator, and its promoter does not carry Sta1 binding motifs.

Genomic features. Sequence heterogeneities and other exceptional properties were detected in the SRV genome and in other rudiviral genomes that are described below.

(i) ITRs. For SRV, the 1,030-bp ITR is perfect, except for a 36-bp insert at positions 799 to 834 at the left end and inverted tetramer sequences (AAAA [positions 425 to 428] and TTTT [positions 27672 to 27669]). It shows little sequence similarity to ITRs of the other rudiviruses, except for the 21-bp sequence (AATTTAGGAATTTAGGAATTT) located at the terminus that is predicted to be a Holliday junction resolvase binding site occurring in all sequenced rudiviruses (34). The ITRs of SRV and SIRV1 and -2 carry four to five degenerate copies of this direct sequence repeat, while that of ARV1 carries multiple degenerate copies of other diverse repeats of similar sizes.

(ii) Genome heterogeneity in SRV and 12-bp indels. Sequence heterogeneities were detected in the SRV genome, within the 10-fold sequence coverage, and mutations were localized to groups of subpopulations, including one 180-bp deletion between positions 11896 and 12077 in two out of six clones. Moreover, a 48-bp insertion was observed in one variant (out of 18 clones) precisely at the C terminus of ORF533 (position 20285) that generated a third copy of a 16-amino-acid direct repeat. Some changes corresponding to 12-bp indels were also apparent in overlapping clones, and they are indicated in Table 3 together with those observed earlier for SIRV1 (4, 20, 21). Moreover, sequence comparison of highly conserved ORFs present in the four rudiviral genomes revealed several additional 12-bp indels. The locations of all the identified indels which occur in conserved rudiviral genes or sites corresponding to SRV ORF75, -104, -138, -163, -168, -197, -199, -286, -294, -419, -440, -464, -533, and -1059 (Fig. 5) are indicated in the SIRV1 genome map in Fig. 6.

Rudiviral matches to CRISPRs. The availability of four separate rudiviral genome sequences provided a basis for analyzing the frequency and distribution of the matches of CRISPR spacer sequences to the viral genomes. Therefore, we analyzed the repeat clusters of each of the available crenarchaeal genomes in the public EMBL/GenBank and JGI sequence data-

Downloaded from jb.asm.org at INSTITUT PASTEUR-Médiathèque on October 6, 2008

TABLE 2. Rudiviral proteins with predicted functions

SRV ORF category	Rudiviral homolog(s)	Predicted function or description	Analysis tool	E-value or score	Other crenarchaeal virus(es)
Structural proteins					
ORF134	All	Structural protein			
ORF464	All	Structural protein			
ORF581	All	Structural protein			
ORF1059	All	Structural protein			
Transcriptional regulators					
ORF58	All	RHH-1	SMART	2.0e-08	Many
ORF95	None	"Winged helix" repressor DNA binding domain	3D-Jury	64.00	None
Translational regulator					
ORF294	SIRV1 and -2	tRNA-guanine transglycosylase	3D-Jury	167.57	STSV1
DNA replication					
ORF440	All	RuvB Holliday junction helicase (Lon ATPase)	3D-Jury	53.71	AFV1, AFV2
ORF116c	All	Holliday junction resolvase (archaeal)	SMART	2.4e-45	
ORF199	All	Nuclease	3D-Jury	63.86	AFV1, SIFV
DNA metabolism					
ORF168	SIRV1 and -2	dUTPase	SMART	1.5e-12	STSV1
ORF257	ARV1	Thymidylate synthase (Thy1)	SMART	7.9e-46	STSV1
ORF159	All	*S*-adenosylmethionine-dependent methyltransferase	3D-Jury	73.67	SIFV
Glycosylation					
ORF335	All	Glycosyl transferase group 1	SMART	6.7e-09	
ORF355	All	Glycosyl transferase	SMART	5.1e-04	
Other					
ORF419	SIRV1 and -2	11 transmembrane regions	TMHMM		

bases and in our own unpublished genomes (see Materials and Methods). Fourteen complete genomes and 8 partial genomes were analyzed. In total, 82 repeat clusters from complete genomes and 44 clusters, some incomplete, from partial genomes yielded 4,283 spacer sequences. Subsequently, 278 sequences that are shared between *S. solfataricus* strains P1 and P2 (16) were omitted from the data set, yielding a total of 4,005 spacer sequences.

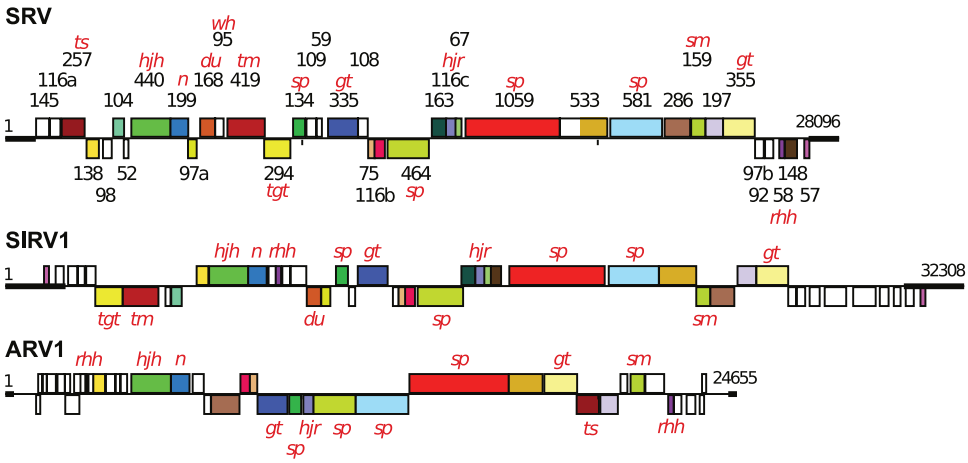


FIG. 5. Genome maps of SRV, SIRV1, and ARV1 showing the predicted ORFs and the ITRs (bold lines). SRV ORFs are identified by their amino acid lengths. Homologous genes shared between the rudiviruses are color-coded. Genes above the horizontal line are transcribed from left to right, and those below the line are transcribed in the opposite direction. Predicted functions or structural characteristics of the gene products are indicated as follows: sp, structural protein; rhh, ribbon-helix-helix protein; wh, winged helix protein; tm, transmembrane; tgt, tRNA guanine transglycosylase; hjh; Holliday junction helicase; hjr, Holliday junction resolvase; n, nuclease; du, dUTPase; ts, thymidylate synthase; sm, *S*-adenosylmethionine-dependent methyltransferase; gt, glycosyl transferase.

Downloaded from jb.asm.org at INSTITUT PASTEUR-Médiathèque on October 6, 2008

TABLE 3. Occurrence of the 12-bp indels in overlapping rudiviral clone libraries

ORF or ITR	No. of +12-bp clones	No. of −12-bp clones	Sequence	Genome position (reference)
SRV ORF58	5	1	AATTAAATTATG	26079–26068
SRV ORF95	8	8	TTTTGAATTATG	7112–7101
SIRV1 ORF335	7	3	AACATTCATTAA	Variant (21)
SIRV1 ORF562	1	4	ATACAAATTTCA	Variant (21)
SIRV1-ITR	10	29	TTTAGCAGTTCA	(20)

In the first analysis, each of the 4,005 spacer sequences was compared to the four rudiviral genomes at the nucleotide level. In total, 158 spacers yielded 268 rudiviral matches. The latter number exceeds the former because (i) some spacers match to more than one locus within repeat sequences of a given virus and (ii) some spacers match to more than one virus. Second, the analysis was performed at the protein level (see Materials and Methods). This analysis revealed 148 additional matching spacers and a further 427 rudiviral genome matches exclusively at the protein level. (An additional 105 matching spacer sequences from the latter analysis that overlapped, partially or completely, with 158 of those detected within rudiviral ORFs at the nucleotide level were not counted.) Only 6 of the 14 completed crenarchaeal genomes carried spacers yielding matches to rudiviral genomes, and they are listed, together with the results for the partial genomes, in Table 4. These results reinforced the choice of criteria employed for determining the significance of sequence matches (see Materials and Methods).

The locations of the spacer sequence matches are superimposed on the genome map of SIRV1 in Fig. 6. The matches are not evenly distributed along the genome; some genes have no matches, while others carry up to 18. Although there is no strict correlation between the level of gene sequence conservation and the number of matching spacers, the five most conserved genes, ORF440, ORF1059, ORF134, ORF355, and ORF581, exhibit the highest number of matches (18, 15, 14, 14, and 13, respectively) (Fig. 3 and 6).

DISCUSSION

The morphological and genomic data for SRV and the other characterized rudiviruses are summarized in Table 1. The conservation of their morphologies and genomic properties contrasts with that of other crenarchaeal viruses and, in particular, with that of the filamentous lipothrixviruses, which exhibit a variety of surface, envelope, and tail structures and much more heterogeneous genomes (24).

The virion length of SRV, 702 (± 50) nm, shows the same direct proportionality to genome size (28 kb) as those for the other rudivirus virions, which range in length from 610 (± 50) nm (ARV1) to 900 (± 50) nm (SIRV2) (Table 1). A superhelical core, with a pitch of 4.3 nm and a width of 20 nm, terminates in 45-nm-long nonhelical "plugs," and it correlates with the internal structure observed earlier in electron micrographs of SIRV1 (22). In order to determine whether a single superhelical DNA can span the SRV virion length, we applied the following formulae to estimate the sizes and length of the superhelical DNA:

$$L_{\text{turn}} = \sqrt{p^2 + c^2}$$

where L_{turn} represents the arc length of a turn, p represents the pitch, and c represents the cylinder circumference, and

$$L_{\text{total}} = t \times L_{\text{turn}}$$

where t represents the number of turns and L_{total} represents the arc length of entire helix.

Calculations using structural parameters for B-form DNA yielded a genome size of 26 kbp without, and 30 kbp with, terminal "plugs". The estimated width (20 nm) is an upper-limit estimate. A reciprocal calculation, with a 28-kbp genome, yields a diameter of 21.2 nm without, and 18.5 nm with, the "plugs." Given that the major rudiviral coat protein is capable of self-assembly into filamentous structures similar in width to the native virion (Fig. 4), it is likely that the rod-shaped body consists of a single superhelical DNA embedded within this filamentous protein structure. Thus, the three newly identified minor structural proteins probably contribute to conserved terminal features of the virion; consistent with this, the largest

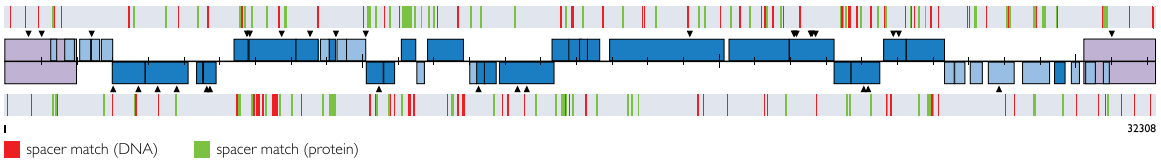


FIG. 6. CRISPR spacer sequence matches for SIRV1 are superimposed on the SIRV1 genome map. Protein-coding regions translated from left to right are shown above the line, and those translated from right to left are shown below the line. Highly conserved coding genes are presented in dark blue, while less-conserved or nonconserved genes are in light blue. The inverted terminal repeat is shaded in violet. Matches to spacers are shown as vertical lines and are color-coded as indicated. Matches to the upper DNA strand are placed above the genome, and those to the lower strand are located below the genome. The red vertical lines correspond to the nucleotide sequence matches, and the green vertical lines correspond to matching amino acid sequences, after translation of the spacer sequences from both DNA strands. In total, there were 106 matches to SIRV1 at the nucleotide level, some of them occurring more than once, and an additional 127 matches to SIRV1 ORFs at the amino acid level. The black arrowheads indicate the positions of the 12-bp indels that occur in one or more conserved rudiviral genes.

Downloaded from jb.asm.org at INSTITUT PASTEUR-Médiathèque on October 6, 2008

TABLE 4. Number of CRISPR spacer sequences from complete and partial crenarchaeal genomes which match rudiviral genomes[a]

Strain	No. of matching sequences at the indicated level[b]		Accession no., reference, or source
	Nucleotide	Amino acid	
Complete genomes			
S. solfataricus P2	22 (14)	31 (18)	NC_002754
S. tokodaii 7	9	14	NC_003106
M. sedula	5	15	NC_009440
S. acidocaldarius	5	9	NC_007181
S. marinus F1	2	1	NC_009033
H. butylicus	0	1	NC_008818
Incomplete genomes			
S. solfataricus P1	20 (14)	30 (18)	16
S. islandicus (5 strains)	39/12/4/2/1	26/7/2/2/0	See text
S. islandicus HVE10/4	36	11	Unpublished
A. brierleyi	15	14	Unpublished

[a] All the acidothermophilic organisms from the family *Sulfolobaceae* have spacers matching those of the rudiviral genomes. However, the neutrophilic hyperthermophiles *S. marinus* and *H. butylicus* produced very few matches. Matches at the amino acid sequence level that overlapped with those at the nucleotide sequence level were excluded from the data.

[b] Numbers in parentheses in columns 2 and 3 indicate the number of matches that arose from spacers shared by *S. solfataricus* strains P1 and P2 (16).

structural protein (corresponding to SRV ORF1059) was localized within the virion tail fibers of SIRV2 by studying functional groups by the use of bioconjugation (Steinmetz et al., submitted).

We still have limited insight into functional roles of rudivirus-encoded proteins (Table 2). The glycosyl transferases have been implicated in the glycosylation of the structural proteins (34). Moreover, a few proteins have been linked to viral replication. Two of these, ORF440 and ORF199, lie within an operon and are conserved in phylogenetically diverse lipothrixviruses (33). The former yielded significant matches to RuvB, the helicase facilitating branch migration during Holliday junction resolution, while ORF199 yielded the best matches to nucleases, including Holliday junction resolvases (Table 2). Thus, they are likely to facilitate rudiviral replication, which, in SIRV1, involves site-specific nicking within the ITR, formation of head-to-head and tail-to-tail intermediates, and conversion of genomic concatemers to monomers by a Holliday junction resolvase (ORF116c) (20). In addition, SRV encodes a dUTPase and a thymidylate synthase, both of which are involved in thymidylate synthesis, whereas the other rudiviruses encode only one of these enzymes, both of which are considered helpful in maintaining of a low dUTP/dTTP ratio and thus in minimizing detrimental effects of misincorporating uracil into DNA. Two putative transcriptional regulators have been identified, together with the putative tRNA transglycosylase encoded by SRV, which has homologs in SIRV1 and -2 and in other crenarchaeal viruses (Table 2) and is distantly related to a tRNA-guanine transglycosylase implicated in archeosine formation.

The two approaches employed to analyze CRISPR spacers matching the four rudiviral genomes demonstrated that about 10% of the 3,042 unique acidothermophile spacers yielded positive matches. Employing alignments at the amino acid level considerably increased the number of positive matches detected, because nucleotide sequences diverge more rapidly. Thus, the genomes of SRV and SIRV1 share almost no (~4%) similarity at the DNA level, whereas most homologous proteins show, on average, 47% sequence identity or similarity. When studying the distribution of the spacer matches in the rudiviral genomes, some trends are evident. First, there is no significant bias with regard to the DNA strand carrying the matching sequence. In SIRV1, for example, 122 matches occur on one strand and 111 on the other (Fig. 6). This is consistent with our assumption that the incorporation of viral or plasmid DNA into the orientated CRISPRs is nondirectional. Second, in accordance with earlier analyses (16), for matches to coding regions, there is no significant bias to matches occurring in a sense or antisense direction. Thus, for SIRV1, 39% of the matches are in the sense direction whereas 54% are antisense—the remaining 7% constitute nucleotide matches to non-protein-coding regions (Fig. 6). Third, when the latter nucleotide sequence-based matches are considered, the proportion of matches which occur in intergenic regions, as opposed to those occurring in protein-coding regions, is not significantly different from the overall coding percentage of the virus. For SIRV1 19% of the nucleotide matches fall within intergenic regions, whereas 20% of the genome is non-protein-coding. Finally, some genes have many matches whereas others have none at all. Five genes have 13 or more matches in SIRV1; these genes correspond to SRV ORF440, ORF1059, ORF134, ORF355, and ORF581. Apart from being conserved in each rudivirus, their gene products have important structural or functional roles (Table 2).

The results pose an important question as to how the host distinguishes between more important and less important genes when adding the spacers to its CRISPRs. Possibly, although the de novo addition of spacers may well be an unbiased process with respect to both viral genome position and direction, the selective advantage provided by some spacers would result in a population being enriched in hosts with CRISPRs carrying spacers targeting crucial viral genes.

The 12-bp viral indels were originally shown to occur commonly in SIRV1 variants that arose as a result of passage of an SIRV1 isolate through different closely related *S. islandicus* strains from Iceland, and it was inferred that this unusual activity reflected adaptation of the rudivirus to the different hosts (21). The positions of the 12-bp indels that have been identified in conserved rudiviral protein genes are shown together with the CRISPR spacer matches on the SIRV1 genome map in Fig. 6. Many of the sites are very close or overlap. This raises the possibility that lengthening or shortening of conserved protein genes by 12 bp could be a mechanism to overcome the host CRISPR defense system.

We conclude that the rudiviruses are excellent models for studying details of viral life cycles and virus-host interactions in crenarchaea. These viruses appear to be much more conserved in their morphologies and genomes than, for example, the equally ubiquitous lipothrixviruses. Moreover, they are relatively stably maintained in their hosts and can be isolated in reasonable yields for experimental studies.

Downloaded from jb.asm.org at INSTITUT PASTEUR-Médiathèque on October 6, 2008

ACKNOWLEDGMENTS

We are grateful to Georg Fuchs for providing the environmental sample from Saõ Miguel Island, the Azores.

The research in Copenhagen was supported by grants from the Danish Natural Science Research Council, the Danish National Research Foundation, and Copenhagen University. The research in Paris was partly supported by grant NT05-2_41674 from Agence Nationale de Recherche (Programme Blanc).

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Bettstetter, M., X. Peng, R. A. Garrett, and D. Prangishvili.** 2003. AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus*. Virology **315:**68–79.
3. **Bland, C., T. L. Ramsey, F. Sabree, M. Lowe, K. Brown, N. C. Kyrpides, and P. Hugenholtz.** 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics **8:**209.
4. **Blum, H., W. Zillig, S. Mallok, H. Domdey, and D. Prangishvili.** 2001. The genome of the archaeal virus SIRV1 has features in common with genomes of eukaryal viruses. Virology **281:**6–9.
5. **Brügger, K., P. Redder, and M. Skovgaard.** 2003. MUTAGEN: multi-user tool for annotating genomes. Bioinformatics **19:**2480–2481.
6. **Eder, W., W. Ludwig, and R. Huber.** 1999. Novel 16S rRNA gene sequences retrieved from highly saline brine sediments of Kebrit Deep, Red Sea. Arch. Microbiol. **172:**213–218.
7. **Edgar, R. C.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5:**113.
8. **Ginalski, K., A. Elofsson, D. Fischer, and L. Rychlewski.** 2003. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics **19:**1015–1018.
9. **Häring, M., X. Peng, K. Brügger, R. Rachel, K. O. Stetter, R. A. Garrett, and D. Prangishvili.** 2004. Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera *Pyrobaculum* and *Thermoproteus*: a novel virus family, the Globuloviridae. Virology **323:**233–242.
10. **Häring, M., G. Vestergaard, K. Brügger, R. Rachel, R. A. Garrett, and D. Prangishvili.** 2005. Structure and genome organization of AFV2, a novel archaeal lipothrixvirus with unusual terminal and core structures. J. Bacteriol. **187:**3855–3858.
11. **Häring, M., G. Vestergaard, R. Rachel, L. Chen, R. A. Garrett, and D. Prangishvili.** 2005. Virology: independent virus development outside a host. Nature **436:**1101–1102.
12. **Kessler, A., A. B. Brinkman, J. van der Oost, and D. Prangishvili.** 2004. Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon *Sulfolobus*. J. Bacteriol. **186**:7745–7753.
13. **Kessler, A., G. Sezonov, J. I. Guijarro, N. Desnoues, T. Rose, M. Delepierre, S. D. Bell, and D. Prangishvili.** 2006. A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters. Nucleic Acids Res. **34:** 4837–4845.
14. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680–685.
15. **Letunic, I., R. R. Copley, B. Pils, S. Pinkert, J. Schultz, and P. Bork.** 2006. SMART 5: domains in the context of genomes and networks. Nucleic Acids Res. **34:**D257–D260.
16. **Lillestøl, R. K., P. Redder, R. A. Garrett, and K. Brügger.** 2006. A putative viral defence mechanism in archaeal cells. Archaea **2:**59–72.
17. **Makarova, K. S., N. V. Grishin, S. A. Shabalina, Y. I. Wolf, and E. V. Koonin.** 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct **1:**7.
18. **Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez, and E. Soria.** 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. **60:**174–182.
19. **Ortmann, A. C., B. Wiedenheft, T. Douglas, and M. Young.** 2006. Hot crenarchaeal viruses reveal deep evolutionary connections. Nat. Rev. Microbiol. **4:**520–528.
20. **Peng, X., H. Blum, Q. She, S. Mallok, K. Brügger, R. A. Garrett, W. Zillig, and D. Prangishvili.** 2001. Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. Virology **291:**226–234.
21. **Peng, X., A. Kessler, H. Phan, R. A. Garrett, and D. Prangishvili.** 2004. Multiple variants of the archaeal DNA rudivirus SIRV1 in a single host and a novel mechanism of genomic variation. Mol. Microbiol. **54:**366–375.
22. **Prangishvili, D., H. P. Arnold, D. Gotz, U. Ziese, I. Holz, J. K. Kristjansson, and W. Zillig.** 1999. A novel virus family, the Rudiviridae: structure, virus-host interactions and genome variability of the *Sulfolobus* viruses SIRV1 and SIRV2. Genetics **152:**1387–1396.
23. **Prangishvili, D., P. Forterre, and R. A. Garrett.** 2006. Viruses of the Archaea: a unifying view. Nat. Rev. Microbiol. **4:**837–848.
24. **Prangishvili, D., and R. A. Garrett.** 2005. Viruses of hyperthermophilic Crenarchaea. Trends Microbiol. **13:**535–542.
25. **Prangishvili, D., R. A. Garrett, and E. V. Koonin.** 2006. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res. **117:**52–67.
26. **Prangishvili, D., G. Vestergaard, M. Häring, R. Aramayo, T. Basta, R. Rachel, and R. A. Garrett.** 2006. Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. J. Mol. Biol. **359:**1203–1216.
27. **Rachel, R., M. Bettstetter, B. P. Hedlund, M. Häring, A. Kessler, K. O. Stetter, and D. Prangishvili.** 2002. Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. Arch. Virol. **147:**2419–2429.
28. **Reilin, A.** 1998. Preparation of catalase crystals. University of Illinois at Urbana-Champaign, Urbana, IL. http://www.itg.uiuc.edu/publications/techreports/98-009.
29. **Rice, G., K. Stedman, J. Snyder, B. Wiedenheft, D. Willits, S. Brumfield, T. McDermott, and M. J. Young.** 2001. Viruses from extreme thermal environments. Proc. Natl. Acad. Sci. USA **98:**13341–13345.
30. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. ARTEMIS: sequence visualization and annotation. Bioinformatics **16:**944–945.
31. **Sæbø, P. E., S. M. Andersen, J. Myrseth, J. K. Lærdahl, and T. Rognes.** 2005. PARALIGN: rapid and sensitive sequence similarity searches powered by parallel computing technology. Nucleic Acids Res. **33:**W535–W539.
32. **Snyder, J. C., B. Wiedenheft, M. Lavin, F. F. Roberto, J. Spuhler, A. C. Ortmann, T. Douglas, and M. Young.** 2007. Virus movement maintains local virus population diversity. Proc. Natl. Acad. Sci. USA **104:**19102–19107.

32a. **Steinmetz, N. F., A Bize, K. C. Findlay, G. P. Lomonossoff, M. Manchester, D. J. Evans, and D. Prangishvili.** Site-specific and spatially controlled addressability of a new viral nanobuilding block: *Sulfolobus islandicus* rod-shaped virus 2. Adv. Funct. Mat., in press.

33. **Vestergaard, G., R. Aramayo, T. Basta, M. Häring, X. Peng, K. Brügger, L. Chen, R. Rachel, N. Boisset, R. A. Garrett, and D. Prangishvili.** 2008. Structure of the *Acidianus* filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. J. Virol. **82:**371–381.
34. **Vestergaard, G., M. Häring, X. Peng, R. Rachel, R. A. Garrett, and D. Prangishvili.** 2005. A novel rudivirus, ARV1, of the hyperthermophilic archaeal genus *Acidianus*. Virology **336:**83–92.
35. **Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, H. Hain, M. Lanzendörfer, and J. K. Kristjansson.** 1994. Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. System. Appl. Microbiol. **16:**609–628.

Downloaded from jb.asm.org at INSTITUT PASTEUR-Médiathèque on October 6, 2008

Supplementary TABLE S1. Properties of predicted ORFs and operons of SRV

ORF	TATA-like motif	S-D motif	terminator (+/-)	operon (+)
Forward strand				
145	TTAAATAT(-35)	GGGGGA(-6)	-	
116a	AATAATTA(-29)	GGTGA(-7)	+	
257	TTAAATAA(-24)		+	
104	TTATATAT(-22)		+	
440	TTTTTAAA(-24)		-	+
199*		G**G**GG**T**TGA(-7)	+	
168	TAAAATAT**GTA**(-19)		-	+
95		GGGTG(-5)	+	
419	AAATTTAA(-29)		+	
134	AATATATA(-31)		+	
109	TATTTT(-24)		+	
59	AATATT(-26)		+	
335	ATTTTAAT**GTA** (-20)		-	
108	ATTATATA (-21)		+	
163	TTAATAAT**GTA**(-21)		-	+
116c	AAAATAA(-27)	GAGG(-4)	-	
67*		GAAGTG	+	
1059	TTTATTAT(-29)		-	+
533	ATTAAATT(-23)	GAGGT(-11)	+	
581	TTATTTAT**GTA**(-23)		+	
286	ATTTATTT(-25)		+	
159	TTTTTTAT**GTA**(-19)		-	+
197		AGGTGA(-7)	-	
355		GAGGG(-8)	+	
Reverse strand				
57	ATTTTAAA (-38)	GGGGA(-8)	-	
147	ATTTAAATAT(-40)	AGGTGA(-6)	-	+
58		GGAGGT(-10)	+	
92	TATTTTTT(-25)		+	
97b	TTAAATAT**GTA**(-18)		+	
464	ATTATTTT(-22)		+	
116b	TTTATTTT(-21)		-	+
75		GTGGGG(-8)	+	
294	ATAAATTT**GTA**(-21)		+	
97a	TTAATATT**GTA**(-19)		+	
52	TTAAATAT**GTA**(-22)		+	
98	ATTTAAAA(-27)		+	
138	TTAAATAT**GTA**(-32)	GAGGCG(-7)	+	

Isolement et caractérisation d'un virus filamenteux et de son hôte, à partir d'échantillons environnementaux de sources géothermiques de la péninsule de Kamchatka

1

Présentation de l'étude

La recherche de nouveaux virus à partir d'échantillons environnementaux de diverses provenances est au coeur des activités du laboratoire BMGE. Dans le cadre d'une collaboration avec le « Danish Archaea Center », avec l'Institut de Minéralogie et de Physique des Milieux Condensés, de l'Université Pierre et Marie Curie et avec l'équipe du « Laboratory of Hyperthermophilic Microbial Communities », à Moscou, nous avons travaillé sur des échantillons environnementaux prélevés dans les sources géothermiques de la péninsule de Kamchatka.

L'équipe russe a réalisé des enrichissement en conditions anaérobies.

A l'Institut Pasteur, nous avons réalisé des enrichissement en conditions aérobies, favorables à la croissance d'espèces des genres *Sulfolobus* et *Acidianus*, hyperthermophiles appartenant à l'ordre Sulfolobales, phylum Crenarchaeota. Un virus filamenteux, AFV9 (famille β-lipothrixviridae), a été isolé et caractérisé. Soizick Lucas-Staat a réalisé les enrichissements et découvert le virus. Ensemble, nous avons produit le virus en quantité suffisante pour permettre son séquençage, isolé le système hôte-virus, observé les virions en microscopie électronique en transmission, identifié de nouvelles protéines structurales et construit l'arbre phylogénétique basé sur les séquences des PSU-ARNr, permettant de mieux situer l'hôte. La plate-forme protéomique de l'Institut Pasteur a contribué à l'identification des protéines structurales (Pascal Lenormand).

Le séquençage du génome d'AFV9, l'analyse de la séquence obtenue et l'étude des fragments terminaux du génome viral ont été réalisés par l'équipe danoise. Les études de cryo-microscopie électronique ont été réalisées par l'équipe de l'Institut de Minéralogie.

Ces travaux ont fait l'objet d'une publication (Bize et al., 2008), insérée ci-après. Les informations présentées dans les parties I et II sont suffisantes pour comprendre l'article. Signalons simplement que les virions des lipothrixvirus présentent des extrémités variées (I.3.3.2), c'est pourquoi celles des virions d'AFV9 sont spécifiquement décrites dans l'article. Par ailleurs, le génome d'AFV9 est linéaire, comme celui des autres lipothrixvirus. On s'attend donc à ce que ses extrémités soient protégées, soit par la présence d'un protéine, soit par une liaison covalente des deux brins de l'ADN. C'est pourquoi les extrémités ont été spécifiquement étudiées. Signalons enfin que deux articles antérieurs présentent des études du même genre sur les β-lipothrixvirus SIFV (Arnold et al., 2000b) et AFV3, AFV6, AFV7, AFV8 (Vestergaard et al., 2008a).

INSTITUT PASTEUR

Research in Microbiology 159 (2008) 358–366

www.elsevier.com/locate/resmic

Viruses in acidic geothermal environments of the Kamchatka Peninsula

Ariane Bize [a], Xu Peng [b], Maria Prokofeva [c], Kirsty MacLellan [d,e], Soizick Lucas [a], Patrick Forterre [a], Roger A. Garrett [b], Elizaveta A. Bonch-Osmolovskaya [c], David Prangishvili [a,*]

[a] *Molecular Biology of the Gene in Extremophiles Unit, Institut Pasteur, rue du Dr. Roux 25, 75724 Paris Cedex 15, France*
[b] *Danish Archaea Centre, Department of Molecular Biology, Biocenter, Copenhagen University, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark*
[c] *Laboratory of Hyperthermophilic Microbial Communities, Winogradsky Institute of Microbiology RAS, Prospekt 60, Let Oktyabrya 7, Bldg. 2, 117312 Moscow, Russia*
[d] *Institut de Minéralogie et de Physique des Milieux Condensés, Université Pierre et Marie Curie, IMPMC-UMR7590, F-75005 Paris, France*
[e] *CNRS, IMPMC-UMR7590, F-75005 Paris, France*

Received 4 March 2008; accepted 16 April 2008
Available online 8 May 2008

Abstract

Screening for viruses in samples taken from acidic hot springs of Kamchatka (Russia) revealed a collection of morphotypes, including linear, spherical and complex fusiform shapes, which show partial similarity to those found in acidic geothermal environments in other geographical locations. One of the viruses, *Acidianus* filamentous virus 9, AFV9, was isolated and its structure and genome were studied in detail.
© 2008 Elsevier Masson SAS. All rights reserved.

Keywords: Virus; Archaea; Thermophilic; Acidophilic

1. Introduction

Our knowledge of viral diversity in diverse ecological niches is extremely limited and insufficient considering that viruses are the most abundant biological entities on our planet [31]. The available data reveal a relatively uniform landscape of viral morphotypes in freshwater, estuarine and marine systems at moderate and low temperatures, which are dominated by head-tail bacteriophages (reviewed in refs. [1,27,37]). However, a radically different picture of viral diversity emerges in acidic geothermal environments, at temperatures above 80 °C. In samples taken from terrestrial acidic hot springs in Iceland [38], in Yellowstone National Park, USA [27,28] and in Southern Italy [12], head-tail bacteriophages have never been detected. Instead, novel and complex virion morphotypes were observed, some of which have never been reported before for any virus.

Here we report on a survey of virus-like particles (VLPs) in samples collected from acidic hot springs of the hydrothermal fields of the Kamchatka Peninsula in Eastern Russia. Moreover, a novel filamentous virus AFV9, which infects strains of the hyperthermophilic archaeal genus *Acidianus*, was isolated from one of the samples and characterized.

2. Materials and methods

2.1. Cultivation of thermoacidophilic hosts

Sample 1 was collected from a hot acidic lake (82 °C, pH 4) in a hydrothermal field near the Moutnovski volcano of the Kamchatka Peninsula. A total of 1 mL of the sample was used to establish a 20 ml anaerobic enrichment culture at 82 °C in medium containing 2 g l^{-1} of yeast extract, pH 4 [26].

Sample 2 was taken from a hot spring on the slope of a sulfur-rich hill on the western margin of the East Thermal Field of Uzon Caldera, Kamchatka Peninsula. An aerobic enrichment

* Corresponding author. Tel.: +33 (0)144 389 119.
E-mail address: prangish@pasteur.fr (D. Prangishvili).

0923-2508/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.resmic.2008.04.009

culture was established, under conditions favorable for the growth of members of the order Sulfolobales [38]. 50 ml medium containing 2 g l^{-1} of tryptone and 1 g l^{-1} of yeast extract, pH 3.5, was inoculated with 1 ml of the sample and incubated at 75 °C with shaking.

2.2. Isolation of viruses

After removal of cells from the enrichment culture (4500 rpm, 20 min, Sorvall GS3 rotor), virions were precipitated by adjusting the cell-free supernatant to 1 M NaCl and 10% (w/v) polyethylene glycol 6000. Virions were pelleted by centrifugation (12,000 rpm, 30 min, Sorvall RC5C Plus, GSA rotor), resuspended in 20 mM Tris-acetate, pH 6 and purified by centrifugation in a CsCl buoyant density gradient (38,000 rpm, 44 h, Beckman Coulter Optima LE-80K ultracentrifuge, SW41 rotor).

2.3. Electron microscopy

For sample 1, the material was treated with 2% (w/v) phosphotungstic acid (pH 3.5) on Formvar (Serva, Heidelberg, Germany) and analyzed in a JEM-100B electron microscope (Jeol, Tokyo, Japan) operated at 60 kV. Particles in sample 2 and the purified AFV9 virions were negatively stained with 2% (w/v) uranyl acetate (pH 4.5) and examined using a JEOL 1200 EXII TEM operated at 120 keV or a JEOL 2100F TEM operated at 200 keV. Cryoelectron microscopy (cryoEM) was performed as described earlier [35]. Images were recorded at nominal magnification of 40,000 for negatively stained samples, and at 50,000 for cryosamples using a CCD 4K camera (Gatan Inc., USA).

2.4. Image processing

Digital images of negatively stained virions were pre-processed to invert the image densities. The contrast of both the negatively stained and cryoEM images were normalized from 0 to 1 and the subsequent images were confined to 128^2 boxes with the program EMAN [17]. The boxed particles, consisting of 1057 and 223 images, respectively, for the negatively stained and the cryoEM data sets, were aligned and classified in EMAN. The five most coherent classes, consisting of 222 images, were selected for a separate principle component analysis (PCA), which was performed in SPIDER on the aligned and classified cryoEM data [9]. The resulting classes from PCA were averaged and the most coherent, as judged visually, is presented here. The final averaged classes consisted of 220 and 21 images, respectively.

2.5. 16S rDNA analysis

16S rRNA genes were amplified by PCR using primers 8aF and 1512uR [8]. The 16S rRNA phylogenetic tree was constructed with 16S rDNA sequences of 43 crenarchaeal strains. The alignment was performed with ClustalX [32] and the positions were selected using GBlocks [7]. The tree was computed on 1234 positions with the maximum likelihood method using PhyML [10] (bootstrap value of 500).

2.6. Protein analysis

Proteins were separated in 4–12% SDS-polyacrylamide gradient gels Nupage using MOPS buffer (Invitrogen, Paisley, UK). The gels were stained either with PageBlue (Fermentas, Lithuania) or SYPRO Ruby (Invitrogen). Protein bands were analyzed by peptide mass fingerprinting (PMF) with MALDI-TOF mass spectrometry using a Voyager DE-STR (Applied Biosystems, Framingham, MA, USA) as described earlier [24]. The analysis was performed by the Proteomic Platform at the Pasteur Institute.

2.7. DNA sequencing and sequence analysis

A shotgun library was prepared by sonicating AFV9 DNA to produce fragments of 2–4 kb. After end-filling, the fragments were cloned into the SmaI site of the pUC18 vector. DNA clones were isolated using a Model 8000 Biobot (Qiagen, Westburg, Germany) and sequenced in MegaBACE 1000 Sequenators (Amersham Biotech, Amersham, UK). Genomic termini were sequenced by primer-walking either on the isolated viral DNA or on DNA amplified with the GenomiPhi kit (Amersham Biotech). The viral genome sequence was assembled using Sequencher 4.5 (http://www.genecodes.com) and analyzed using Artemis (http://www.sanger.ac.uk/software/artemis/). Gene sequence searches were made in the *Sulfolobus* database (http://www.sulfolobus.org/cbin/mutagen.pl) and in GenBank/EMBL (http://www.ncbi.nlm.nih.gov/blast/) and sequence motifs were analyzed using MEME (http://meme.sdsc.edu).

In order to determine the terminal sequences of the linear genome, the 3′-OH at the right end was tailed with poly dA using terminal deoxynucleotidyl transferase as described previously [18]. A 45-cycle PCR step was performed to amplify the poly-dA-tailed end using a 30-mer oligo dT and a primer annealing to the right terminal region of AFV9. The PCR product generated was cloned with the Zero Blunt TOPO PCR cloning kit (Invitrogen) and two independent clones were sequenced.

2.8. Nucleotide sequence accession number

The sequence of AFV9 genome has been submitted to GenBank under accession number EU545650.

3. Results

3.1. Putative virus-host systems in enrichment cultures

3.1.1. Enrichment culture from sample I

Sample 1 was used to establish an anaerobic enrichment culture at 82 °C, pH 4. Cells growing in the primary enrichment were morphologically diverse. Dominant were coccoid cells of species of the crenarchaeal genus *Acidolobus* as

determined by 16S rDNA analysis [25,26]. In addition, rod-shaped cells of unidentified species were present in the culture.

Examination of enrichment by electron microscopy revealed several types of virus-like particles (VLPs) (Fig. 1). Because no VLPs could be observed in the environmental sample, these particles apparently were produced as a result of host growth and they consistently re-appeared after inoculating fresh medium with the same environmental sample. However, VLPs were present only in primary enrichment and they disappeared after subsequent transfers. Induction of VLPs in the ''cured'' culture did not occur on treatment of the cells with mytomycin C or upon UV irradiation. It was inferred that the viral hosts could not be stably cultured, possibly because of lysis and/or the onset of species resistance to virus infection. The morphotypes of VLPs observed in the enrichment culture reinforced the view that fusiform and linear viruses dominate in hot environments (reviewed in ref. [22]).

Fusiform particles (270–310 nm × 50–130 nm) of different types were observed, mainly differing in the tail length at a pointed end of the virion (0–620 nm) (Fig. 1a–c). Sometimes, sickle-shaped terminal structures were visible which were probably involved in adsorption. The structure does not match fusiform virions of known viruses, but rather, resembles in its shape the claw-like ends of the filamentous lipothrixvirus AFV1 of *Acidianus* which clamp onto viral receptors located on the host cell pili [4]. However, none of the cells in the analyzed enrichment possessed any appendages, and the ''sickles'' apparently mediate adsorption onto host receptors in an unknown and novel manner.

The tail length polymorphism of the virions does not necessarily reflect the presence of diverse viral species. A broad range of observed tail lengths (0–620 nm) may indicate the intrinsic pleomorphic nature of a single (or a few) species in the culture. For some VLPs, the tail is completely absent and the ''sickle'' is directly attached to one of the pointed ends of the virion (Fig. 1a), which suggests that its formation does not require the presence of a tail. A tail may develop later from the pointed end of the virion, as described for virus ATV [14,24]. The observations do not exclude the possibility that the process can be extracellular, as for ATV [14,24]. In Fig. 1d, it is unclear whether one is observing virion extrusion or adsorption.

Linear particles, 450 nm × 20 nm, were observed (Fig. 1e) and their terminal structures resemble the ends of the filamentous betalipothrixvirus SIFV, which are tapered and carry six thin flexible filaments [3]. Linear particles of similar length were also observed associated with rod-shaped host cells. In Fig. 1f, the particles can be seen through the hexagonal envelope of the partially lysed rod-shaped cell. Based on the envelope characteristics, the cell could be assigned to a crenarchaeon, probably of the order Thermoproteales [16]. Spherical particles were observed exclusively inside coccoid producer cells but not in the culture medium (Fig. 1g). They appear uniform in their diameter and may represent icosahedral virions.

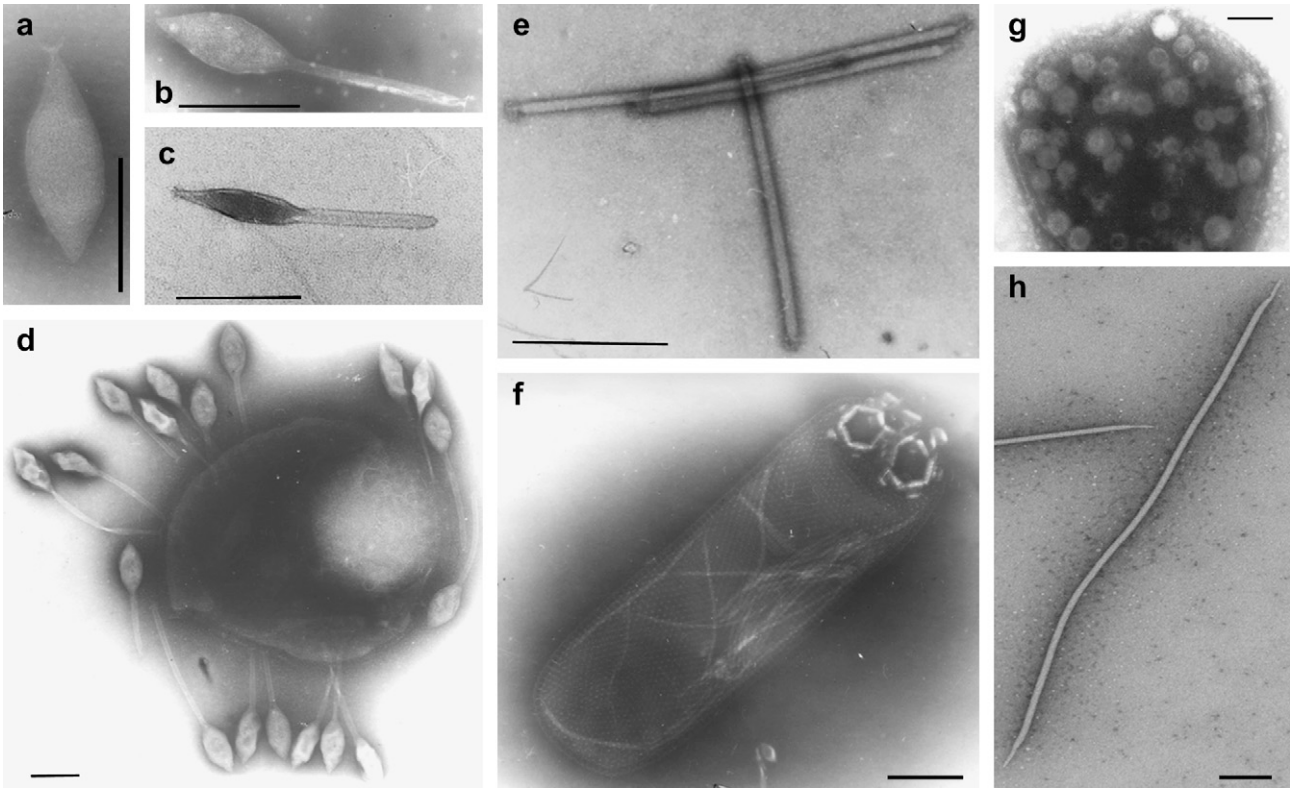


Fig. 1. Electron micrographs of virus-like particles observed in enrichment cultures. (a–g) Particles in the enrichment culture from sample 1. (h) Particles in the enrichment culture from sample 1. Samples were negatively stained with 2% phosphotungstic acid (a–g), or with 3% uranyl acetate. Bars, 200 nm.

3.1.2. Enrichment culture from sample 2

Only one type of filamentous VLP was observed in the enrichment cultures derived from sample 2, grown aerobically at 75 °C, pH 3.5 (Fig. 1h). The morphologies of putative hosts were also very limited in this culture, with only similarly sized cells of irregular coccoid shape being observed.

The VLPs were collected and purified from the culture supernatant and double-stranded (ds) DNA was extracted. Sequencing confirmed that a single virus species was replicating in the culture (see below), as suggested by TEM. The virus was named *Acidianus* filamentous virus 9, AFV9, on the basis of its morphology and host range (see below), following a succession of previously described filamentous viruses of *Acidianus* [4,13,35].

In order to purify the virus-host system, serial dilutions of the enrichment culture were spread on Gelrite plates. Colonies appeared after incubation for 6 days at 80 °C and colony hybridization was used to identify virus-producer strains [29]. The probe was a [^{32}P]-end-labelled 492 bp fragment of AFV9 DNA amplified by PCR using primers GGTTACCA AACAAGGCGAAA and GAAATTGTTTGCGTGCGTTA. Of the 170 colonies analyzed, about 90% were from AFV9-infected cells. Viral replication in three randomly selected colony-purified strains was confirmed by TEM analysis of liquid cultures. Filamentous VLPs observed in these three cultures were indistinguishable in their morphology and size from AFV9. Moreover, 16S rDNA sequences from the three strains were identical and representative of the archaeal genus *Acidianus*; one of them, named "*Acidianus uzoniensis*" (Fig. 2), was used for subsequent production of AFV9. DNA of the virus produced by "*A. uzoniensis*" and that from the primary enrichment yielded identical fragments after digestion with different restriction endonucleases (data not shown).

3.2. Host range of AFV9

The host range of AFV9 was examined by adding purified virions to exponentially growing cultures of potential hosts that included strains of the genera *Acidianus* and *Sulfolobus*: "*Acidianus hospitalis*" W1 (host of AFV1 [4]), "*Acidianus convivator*" AA9 (host of AFV3, AFV6, AFV7 and AFV6 [35]) and "*Sulfolobus islandicus*" HVE10/4 (host of SIFV, [3]). Virus propagation in the cultures was analyzed by TEM 48 h postinfection. Whereas *A. hospitalis* and *S. islandicus* were not infected by AFV9, replication of the virus was clearly observed in the infected culture of "*A. convivator*", a strain isolated from a hot spring in Pozzuoli, Italy [12], and the result confirmed the viral nature of AFV9. During a cycle of productive infection with AFV9, no cell debris was formed nor was there a decrease in cell density; this indicated that virus production had not caused lysis of the host cells.

3.3. Virion structure of AFV9

Filamentous virions of AFV9 (2000 ± 10 nm × 21 ± 0.5 nm) (Fig. 1h), taper at each terminus and have a special terminal structure to which a thin 50 nm long filament is attached (Fig. 3b,c). The shape of the terminal structure is poorly resolved in negatively stained images, but it resembles a T-bar or claw in cryoEM images (Fig. 3b,c). The virions appear to be enveloped and in partially disrupted particles the envelope is removed to reveal the core. The latter is shown in more detail in Fig. 3d, where one particle is completely devoid of an envelope; a small region of free dsDNA is also visible. The core is 15 (±1) nm in width. Both the enveloped and un-enveloped virions, when prepared by rapid freezing, reveal a striated pattern, perpendicular to the viral axis.

The envelope structure was visualized by negative staining and image processing (see Section 2). The 2D average shows a tube-like structure, 21 nm in diameter (Fig. 4a), with high-density globular regions distinguishable across the surface (Fig. 4a, arrows). CryoEM images were analyzed similarly to produce a 2D averaged image consisting of 21 aligned images. The 2D average had the same overall parameters as the negative stain 2D average, with a diameter of 21 nm, but

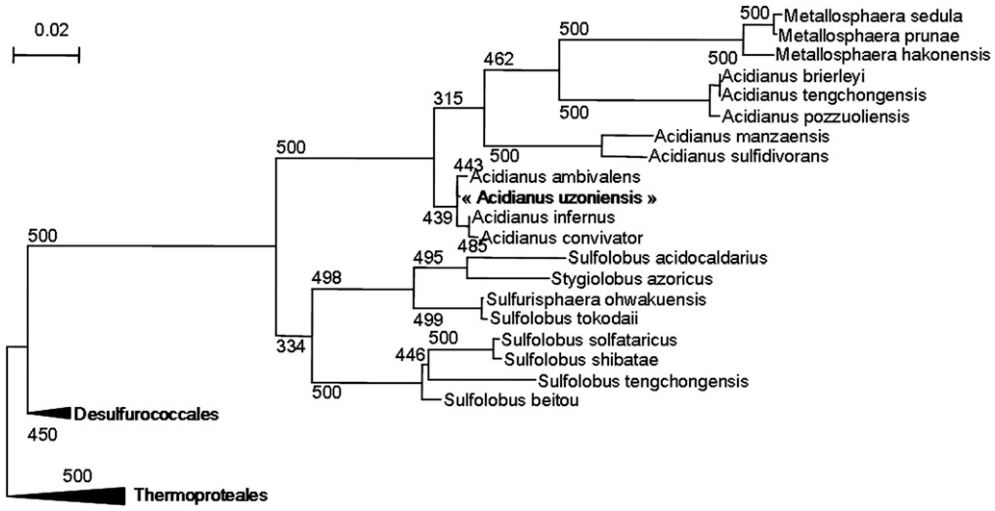


Fig. 2. Phylogenetic tree of Sulfolobales species based on 16S rDNA sequences. The AFV9 host is shown in bold characters.

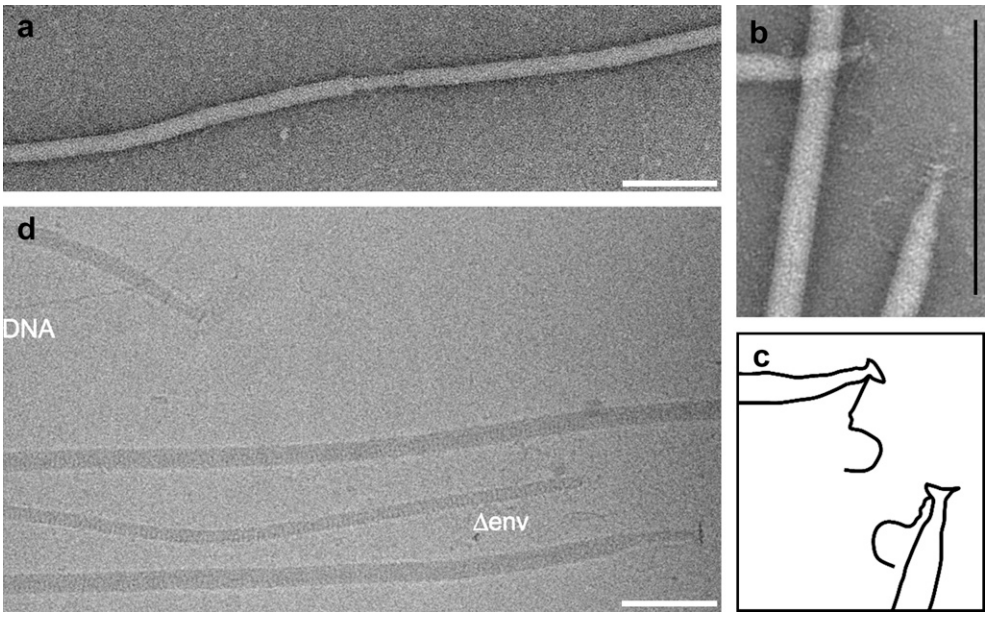


Fig. 3. Electron micrographs of portions of AFV9 virions. (a) The central region of the virion is devoid of envelope. (b) The virion termini consisting of a T-bar or claw structure and a single thin filament; (c) line diagram to aid the location of the single thin filaments attached to the virion termini in b. (d) Portions of three enveloped virions, one un-enveloped virion and a small region of DNA. (a, b) Negative staining with 2% uranyl acetate; (d) cryoelectron micrograph. Bars, 100 nm.

showed more structural features. A cylindrical outer wall of 3.4 nm occurs at the virion periphery, which corresponds to the envelope. The most coherent class average image reveals that the internal structure is closely similar to that of AFV3, with the central lumen filled with ''ovoid masses'' arranged in a zig-zag pattern [35]. These ovoid structures are 3.2 ($\pm$0.6) $\times$ 5.3 ($\pm$0.3) nm and the center to center vertical distance between adjacent ovoids is 5.0 nm. Principle component analysis (see Section 2), a classification technique which sorts the data into groups of least variance, did not reveal any variation between the arrangements of the ovoid structures in the filament lumen. The overall resolution of the averaged images is insufficient to visualize the dsDNA within the virus or to localize it with respect to the ovoid masses.

3.4. Terminal structures of the linear genome

When AFV9 DNA was digested with restriction enzymes and then extracted with phenol, the terminal fragments (<1600 bp) were exclusively absent from the aqueous layer (Fig. 5a) and they could only be recovered from the phenol-water interface (data not shown). This suggested that some protein remained strongly bound to the DNA termini despite the phenol treatment.

In order to test whether the two strands of the AFV9 DNA were linked at their termini, the viral DNA was digested with HindIII and SpeI and the resulting terminal fragments of 1576 bp and 618 bp, as well as an internal control fragment of 933 bp, were extracted from a 0.9% low-melting agarose

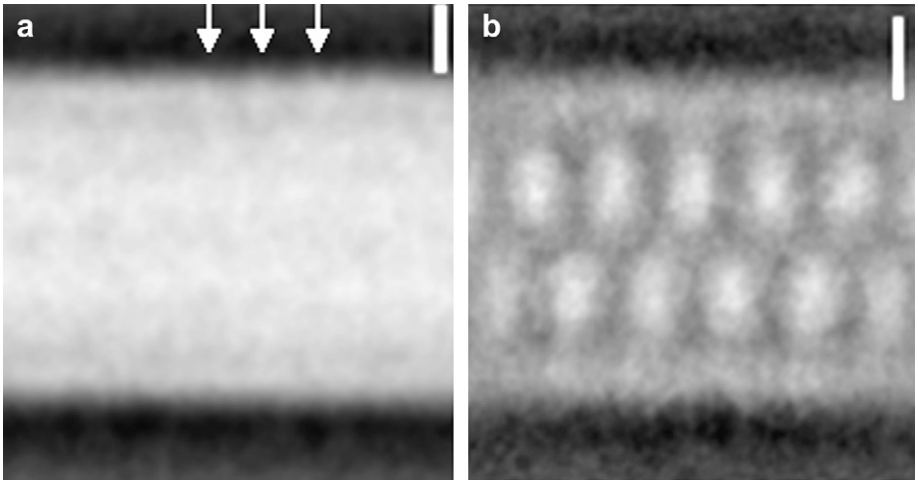


Fig. 4. Two-dimensional maps of AFV9 virions. (a) Averaged 2-D map generated from 220 negative staining images showing envelope structure at low resolution. (b) Average 2-D map generated from 21 cryoEM images revealing a cylindrical envelope surrounding two internal rows of ovoid masses. Images were windowed, aligned, classified, and averaged. Bars, 5 nm.

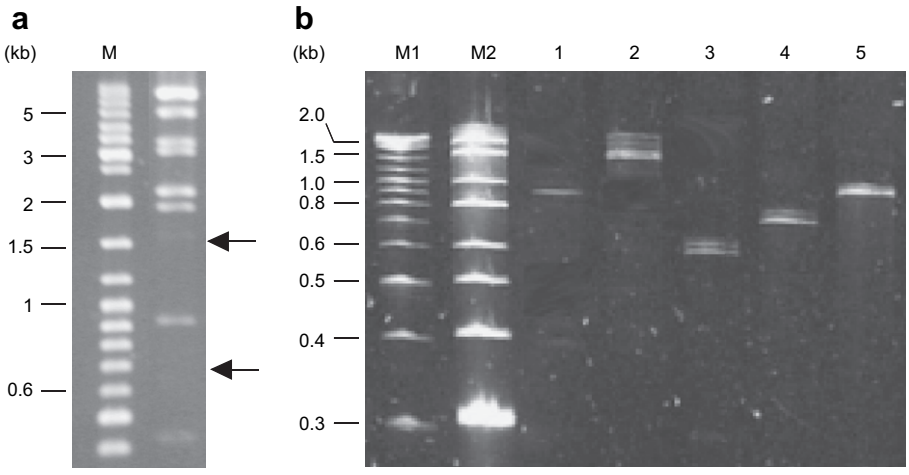


Fig. 5. Electrophoretic separation of fragments of viral DNA produced by restriction endoncleases HindIII and SpeI. (a) Fragments of AFV9 DNA were isolated from the aqueous phase after phenol extraction and separated in a 1% agarose gel; the positions where the terminal fragments should migrate are indicated by arrows; M indicates the size markers. (b) Denatured terminal fragments of AFV9 and SIFV DNA were subjected to electrophoresis in denaturing polyacrylamide gel. DNA fragments were loaded as follows: lane 1, internal AFV9 fragment of 933 bp; lanes 2 and 3, terminal AFV9 fragments of 1576 bp and 618 bp, respectively; lanes 4 and 5, terminal fragments from SIFV DNA of 759 bp and 999 bp, respectively; M1 and M2, DNA and RNA size ladders, with fragment sizes indicated on the left.

gel. The fragments were then heated in 47% formamide at 95 °C for 5 min and immediately loaded onto a denaturing 4% polyacrylamide gel containing 8 M urea run at 40–50 °C. The result, shown in Fig. 5b, demonstrates that the two strands of each fragment had separated, thereby establishing that they were non-covalently linked. Nevertheless, for the terminal fragments, but not the internal fragment, which yielded a single sharp band, there was some band heterogeneity which we infer arose from residually bound protein or protein fragments.

To confirm this result, a similar experiment was performed with DNA from the filamentous betalipothrixvirus SIFV. It was digested with ClaI and the end fragments of 759 bp and 999 bp were separated and analyzed as described above. They yielded heterogeneous single-stranded DNA bands (Fig. 5b), consistent with the two strands being non-covalently linked but with at least one strand being strongly associated with protein(s).

3.5. Genome map

The viral genome was sequenced with approximately a 5-fold coverage (see Section 2) and local sequence ambiguities were resolved by sequencing PCR fragments amplified from appropriate genomic regions. Termini of the linear genome were absent from the clone library and, therefore, in order to determine the terminal sequences, poly dA was ligated to the terminal 3′-OH groups of the DNA strands. Telomere PCR with subsequent cloning and sequencing of the PCR product was performed at the right end of AFV9 (see Section 2) and primer walking was performed at the left end. Approximately 300 bp of sequence obtained from each end was compatible with the presence of a perfect inverted terminal repeat (ITR).

The genome consisted of 41,172 bp and a map was constructed carrying a total of 74 predicted protein-coding genes (≥49 aa), and a single pseudogene, where the ORFs were labelled according to their amino acid lengths (Fig. 6). AUG start codons predominated (69), whereas stop codons were mainly UAA (34) or UGA (29). Homologous genes shared with other betalipothrixviral genomes are color-coded. Thirty six genes, mainly clustered in the left half of the genome, were common to all the sequenced betalipothrixviruses from the genera *Acidianus* and *Sulfolobus*, while another 26 genes were homologous to some betalipothrixviruses and they were located mainly in the right half of the genome or at the extremities. Fourteen genes appeared to be exclusive to

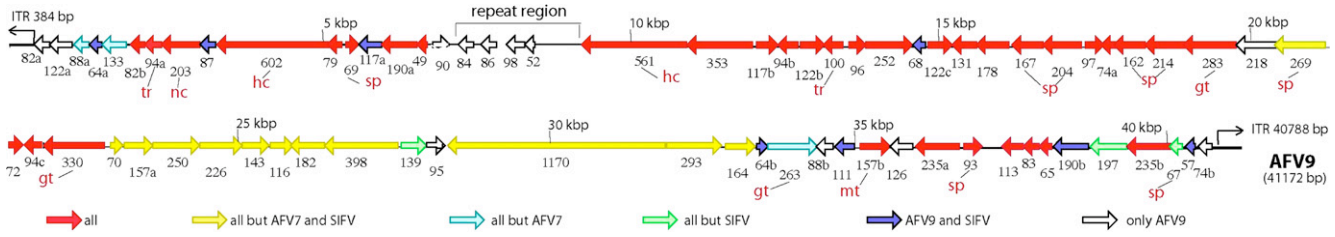


Fig. 6. Genome map for AFV9. Putative genes with homologs in filamentous viruses AFV3, AFV6, AFV7, AFV, and/or SIFV are color-coded as indicated in the figure. The ORFs with putative functions are indicated with red letters: tr – transcription regulators, nc – nucleases, hc – helicase, sp – structural protein, gt – glycosyl transferase, mt – methyl transferase.

AFV9 and showed no significant sequence matches in GenBank, and they tended to occur in small clusters (Fig. 6). One of these clusters containing five genes in the region 6820–8460 bp was less compactly organized than the remainder of the genome, suggesting that it may have another, possibly non-viral, origin. The genome was very efficiently packed with several small overlaps between pairs of genes. Exceptional was the region 7500–9100 bp which carried a low complexity A + T-rich sequence rich in repeats with few protein coding genes (Fig. 6).

The genome is bordered by a 384 bp inverted terminal repeat (ITR). Significantly, terminal 42 bp 5′-CGAGAGAGAAACGTTATAAGACGTGTTAGACACGGATACGTC-3′ shows only three mismatches from the terminal sequence obtained for the *Sulfolobus* betalipothrixvirus SIFV [3,19]. This high level of sequence conservation between viruses from different genera suggests that the terminal region is very important functionally. However, the remainder of the terminal repeat sequence is more divergent between the two viruses. Whereas SIFV exhibits seven tandem repeats of the 22 bp sequence (underlined above), adjoining the terminus, AFV9 exhibits multiple, degenerate and irregularly-spaced direct repeats of the sequence ACGTCTTATAAGA within the terminal 200 bp, which show partial similarity to the first part of the underlined 22 bp sequence above. Moreover, the ITR of AFV9 exhibits a 41 bp perfect direct repeat in the sequence ranges 246–285 and 318–357.

The predicted transcriptional pattern of AFV9 is quite complex both directionally and at a regulatory level. Only 32 of the putative 74 genes carry Shine-Dalgarno (S-D) motifs for which the sequence motif GGTG predominates, while the remainder exhibit TATA-like promoter motifs centered approximately 25 bp upstream from the start codon and the latter should yield leaderless transcripts, as is common for crenarchaea [33,34]. Those genes lacking an S-D motif are generally single genes or first genes of putative operons. Nevertheless, for 14 of the genes which are physically linked there is both an S-D motif and an upstream TATA-like motif about 25 bp upstream, suggesting that either polycistronic or monocistronic mRNAs can be produced from these genes. Termination sites, constituting T-rich polypyrimidine motifs, occur downstream from most single genes and putative operons but there are a few exceptions suggesting that transcriptional readthrough occurs at some sites.

3.6. Structural proteins and predicted protein functions

Proteins extracted from AFV9 virus were resolved electrophoretically into two major bands and approximately ten minor bands (Fig. 7) and their tryptic digests were analyzed by mass spectrometry. A total of eight protein products were identified corresponding to ORF67 (band F), ORF69 (band F), ORF93 (band E), ORF162 (band D), ORF167 (band C), ORF204 (band B), ORF214 (not shown), ORF269 (band A), with respective molecular masses of 7.9, 8.2, 10.5, 18.9, 19.0, 22.6, 24.3, and 30.4 kDa.

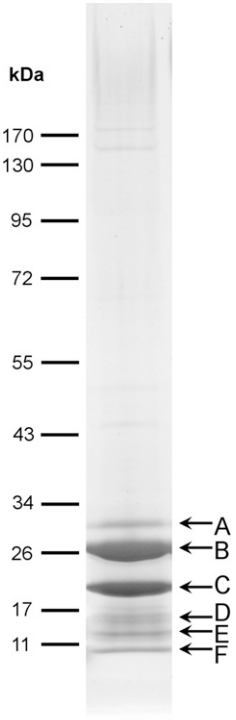


Fig. 7. SDS-gel electrophoresis of protein constituents of the AFV9 virion stained with SYPRO Ruby. Molecular masses of protein standards are indicated in kD.

Proteins from ORF167 and ORF204 are homologous to the major structural proteins of the betalipothrixviruses SIFV and AFV3 [3,35]. Among the six minor structural proteins identified, four also have homologs in all sequenced betalipothrixviruses: ORF69, ORF93, ORF162 and ORF214 [35]. The other two, from ORF67 and ORF269, are shared by all betalipothrixviruses except AFV7 and SIFV (Fig. 6).

Structural proteins of the betalipothrixvirus SIFV were analyzed in a similar way (data not shown). In addition to the two known major structural proteins [3], two minor structural proteins, SIFV0067 and SIFV0072, were identified. The latter is homologous to the AFV9 structural protein encoded by ORF93 and has homologs in all betalipothrixviruses. The AFV9 homolog of the former protein, from ORF235b, was not detected among structural proteins of AFV9, although is encoded on its genome.

Searches against public sequence databases revealed matches to proteins including a Holliday junction DNA helicase (ORF602), another DNA helicase (ORF561), a nuclease (ORF117b), a DNA primase (ORF122c) and resolvase (ORF126). In addition, significant matches were obtained to a SAM-dependent methyl transferase (ORF126), three glycosyl transferases (ORF263, ORF283, and ORF330), a protein phosphatase (ORF139) and two transcriptional regulators (ORF94 and ORF100). Most of these are homologous to proteins encoded in other lipothrixviruses. However, few AFV9 proteins apart from the ubiquitous glycosyl transferases are encoded in other families of archaeal viruses, the exceptions being ORF64a, ORF64b, ORF122b, and ORF157b which are present in some rudiviral genomes.

4. Discussion

Virus particles are generally not observed in untreated samples from hot acidic, extreme thermal environments, by TEM because of their low concentrations and, therefore, substantial particle concentration or the establishment of enrichment cultures is required (reviewed in ref. [21]), Enrichments are commonly performed at temperatures and pH values close to those of the natural environment, under conditions favoring growth of members of the hyperthermophilic crenarchaeal order Sulfolobales, which are abundant in such environments [15]. Thus, overlapping subsets of unique morphotypes of dsDNA viruses of the Sulfolobales were observed in samples from diverse hot acidic springs of Iceland, Italy and the USA (reviewed in refs. [20,22,23]). Here, we report the results of the first screening for viral morphotypes in hot, acidic environments in Eastern Russia, from which until now only the fuselloviruses SSV-K1 had been isolated [36]. In addition to Sulfolobales, members of the other two orders of Crenarchaeota, the Desulfurococcales and Thermoproteales, were potential hosts for the viruses. The results are consistent with previous observations in that linear, spherical and complex fusiform particles predominated in analyzed samples [6]. In order to estimate the degree of similarity with viruses from other geographical locations, one filamentous virus, named *Acidianus* filamentous virus 9 (AFV9), was studied in detail. Along with autochthonous strains of *Acidianus*, the virus was shown to infect and replicate in ''*Acidianus convivator*'', a strain isolated from a hot spring in Italy [12]. Morphological features and genomic properties supported the virus being assigned to the *Betalipothrixvirus* genus of the family *Lipothrixviridae* [35].

A detailed study of the virion structure revealed its high similarity to virions of previously described betalipothrixviruses AFV3, AFV6, and AFV8 of *Acidianus* from Italy [35] and SIFV of *Sulfolobus* from Iceland [3]. The only unique feature of AFV9 is the long terminal filament which may facilitate cellular adsorption. As for AFV3, image processing of the negatively stained virion revealed regularly arranged globular subunits on the virion surface, while cryoEM showed highly dense ovoid masses in the virion lumen which probably represent protein assemblages of similar size to those in AFV3 [35].

Studies on AFV9 have provided an understanding of previously uncharacterized features of the betalipothrixviruses including the nature of the ends of the linear genomes. Earlier studies on SIFV indicated that the DNA termini were inaccessible to both 5′- and 3′-exonucleases as well as terminal transferases and T4 polynucleotide kinase, but no evidence was found for proteins being covalently bound at the termini [3]. Here we demonstrate that there is no covalent linkage between the terminal DNA strands, as occurs in rudiviruses [5,19], but we do provide strong evidence for proteins being covalently bound at each terminus.

Another region of the AFV9 genome, of special interest, is shared with all the other betalipothrixviruses. It constitutes an A + T-rich region, in the size range 1.2–1.5 kb, lying between genome positions 8000–10,000 (Fig. 6). The sequence is of relatively low complexity with few or no protein-coding genes. All the regions are rich in direct and inverted repeats; AFV9, for example, carries two copies of the inverted repeat sequence TTTTTTGTTTGAAAACTTTTTTGCGCCGT and the closely related SIFV carries duplicated inverted repeat of the sequence GTTTGAATACTATTTTGCCCCCTT; moreover, the AFV9 and SIFV regions share about 180 bp with about 90% sequence identity. A corresponding region, rich in direct and inverted repeats, with relatively low sequence complexity, is also present at a similar position in the deltalipothrixvirus AFV2 (8500–9620) and in the gammalipothrixvirus AFV1 (positions 7770–9010). The region is generally bordered on one side by a putative DEXDc DNA helicase and is therefore, a strong candidate for the origin of replication. The region could also serve as an initiation site for virion assembly, given the presence of many potential protein-binding sites.

Earlier members of the spindle-shaped *Fuselloviridae* [36] and the spherical *Globuloviridae* [2,11] with highly similar morphology and genome content were isolated from hot springs of different continents. The lipothrixvirus AFV9 and its relatives from Italy [35] and Iceland [3] provide further examples of viruses with highly similar properties occurring in geographically distant acidic hot springs. These examples probably reflect a high level of virus or virus-host mobility rather than a low rate of virus evolution imposed by specific and extreme environmental conditions. This view is reinforced by the extensive viral movement between both moderate environments, and within hot acidic springs, observed in Yellowstone National Park USA [30].

Acknowledgments

We are grateful to Tamara Basta for providing information on the structural proteins of SIFV, Diego Cortez for his help in constructing the phylogenetic tree, Reinhard Rachel for discussions on the structure of host cells and Nadezhda Kostrikina for her help with electron microscopy of anaerobic enrichments from Kamchatka. Hien Phan and Sarah Hansen helped with the genome analyses. X.P. was supported by a Steno Stipend from the Danish Natural Science Research Council who also supported the research. The research was partly supported by grant NT05-2_41674 from the Agence Nationale de Recherche, France (ANR, programme blanc) and by RAS Programs ''Molecular and Cell Biology'' and ''Origin and Evolution of Biosphere''.

References

[1] Ackermann, H.W. (2007) 5500 Phages examined in the electron microscope. Arch. Virol. 152, 227–243.

[2] Ahn, D.G., Kim, S.I., Rhee, J.K., Kim, K.P., Pan, J.G., Oh, J.W. (2006) TTSV1, a new virus-like particle isolated from the hyperthermophilic crenarchaeote *Thermoproteus tenax*. Virology 351, 280–290.

[3] Arnold, H.P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., et al. (2000) A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon Sulfolobus. Virology 267, 252–266.

[4] Bettstetter, M., Peng, X., Garrett, R.A., Prangishvili, D. (2003) AFV1, a novel virus infecting hyperthermophilic archaea of the genus acidianus. Virology 315, 68–79.
[5] Blum, H., Zillig, W., Mallok, S., Domdey, H., Prangishvili, D. (2001) The genome of the archaeal virus SIRV1 has features in common with genomes of eukaryal viruses. Virology 281, 6–9.
[6] Bonch-Osmolovskaya, E.A., Miroshnichenko, M.L., Slobodkin, A.I., Sokolova, T.G., Karpov, G.A., Kostrikina, N.A., et al. (1999) Biodiversity of anaerobic litotrophic prokaryotes in terrestrial hot springs of Kamchatka. Microbiologya (Russia) 68, 398–406.
[7] Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.
[8] Eder, W., Ludwig, W., Huber, R. (1999) Novel 16S rRNA gene sequences retrieved from highly saline brine sediments of kebrit deep, red Sea. Arch. Microbiol. 172, 213–218.
[9] Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., et al. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. 116, 190–199.
[10] Guindon, S., Lethiec, F., Duroux, P., Gascuel, O. (2005) PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33, W557–559.
[11] Haring, M., Peng, X., Brugger, K., Rachel, R., Stetter, K.O., Garrett, R.A., et al. (2004) Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera Pyrobaculum and Thermoproteus: a novel virus family, the Globuloviridae. Virology 323, 233–242.
[12] Haring, M., Rachel, R., Peng, X., Garrett, R.A., Prangishvili, D. (2005) Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, Acidianus bottle-shaped virus, from a new family, the Ampullaviridae. J. Virol. 79, 9904–9911.
[13] Haring, M., Vestergaard, G., Brugger, K., Rachel, R., Garrett, R.A., Prangishvili, D. (2005) Structure and genome organization of AFV2, a novel archaeal lipothrixvirus with unusual terminal and core structures. J. Bacteriol. 187, 3855–3858.
[14] Haring, M., Vestergaard, G., Rachel, R., Chen, L., Garrett, R.A., Prangishvili, D. (2005) Virology: independent virus development outside a host. Nature 436, 1101–1102.
[15] Huber, H., Prangishvili, D. (2006). In: M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt (Eds.), The prokaryotes (pp. 23–51). New York: Springer.
[16] König, H., Rachel, R., Claus, H. (2007). In: R. Cavicchioli (Ed.), Archaea: molecular and cellular biology (pp. 315–340). ASM Press.
[17] Ludtke, S.J., Baldwin, P.R., Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. 128, 82–97.
[18] Peng, X., Basta, T., Haring, M., Garrett, R.A., Prangishvili, D. (2007) Genome of the Acidianus bottle-shaped virus and insights into the replication and packaging mechanisms. Virology 364, 237–243.
[19] Peng, X., Blum, H., She, Q., Mallok, S., Brugger, K., Garrett, R.A., et al. (2001) Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. Virology 291, 226–234.
[20] Prangishvili, D. (2003) Evolutionary insights from studies on viruses of hyperthermophilic archaea. Res. Microbiol. 154, 289–294.
[21] Prangishvili, D. (2006). In: F.A. Rainey, & A. Ohren (Eds.), Methods in microbiology, extremophiles vol. 35 (pp. 329–345). Academic Press.
[22] Prangishvili, D., Forterre, P., Garrett, R.A. (2006) Viruses of the Archaea: a unifying view. Nat. Rev. Microbiol. 4, 837–848.
[23] Prangishvili, D., Garrett, R.A. (2005) Viruses of hyperthermophilic Crenarchaea. Trends Microbiol. 13, 535–542.
[24] Prangishvili, D., Vestergaard, G., Haring, M., Aramayo, R., Basta, T., Rachel, R., et al. (2006) Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. J. Mol. Biol. 359, 1203–1216.
[25] Prokofeva, M.I., Miroshnichenko, M.L., Kostrikina, N.A., Chernyh, N.A., Kuznetsov, B.B., Tourova, T.P., et al. (2000) *Acidilobus aceticus* gen. nov., sp. nov., a novel anaerobic thermoacidophilic archaeon from continental hot vents in Kamchatka. Int. J. Syst. Evol. Microbiol. 50(Pt 6), 2001–2008.
[26] Prokofeva, M.I., Kublanov, I.V., Nercessian, O., Tourova, T.P., Kolganova, T.V., Lebedinsky, A.V., et al. (2005) Cultivated anaerobic acidophilic/acidotolerant thermophiles from terrestrial and deep-sea hydrothermal habitats. Extremophiles 9, 437–448.
[27] Rachel, R., Bettstetter, M., Hedlund, B.P., Haring, M., Kessler, A., Stetter, K.O., et al. (2002) Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. Arch. Virol. 147, 2419–2429.
[28] Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., et al. (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proc. Natl. Acad. Sci. U. S. A. 101, 7716–7720.
[29] Sambrook, J., Russell, D.W. (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
[30] Snyder, J.C., Wiedenheft, B., Lavin, M., Roberto, F.F., Spuhler, J., Ortmann, A.C., et al. (2007) Virus movement maintains local virus population diversity. Proc. Natl. Acad. Sci. U. S. A. 104, 19102–19107.
[31] Suttle, C.A. (2005) Viruses in the sea. Nature 437, 356–361.
[32] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
[33] Tolstrup, N., Sensen, C.W., Garrett, R.A., Clausen, I.G. (2000) Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. Extremophiles 4, 175–179.
[34] Torarinsson, E., Klenk, H.P., Garrett, R.A. (2005) Divergent transcriptional and translational signals in Archaea. Environ. Microbiol. 7, 47–54.
[35] Vestergaard, G., Aramayo, R., Basta, T., Haring, M., Peng, X., Brugger, K., et al. (2008) Structure of the acidianus filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. J. Virol. 82, 371–381.
[36] Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A.K., Zoeller, L., et al. (2004) Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. J. Virol. 78, 1954–1961.
[37] Wommack, K.E., Colwell, R.R. (2000) Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64, 69–114.
[38] Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., et al. (1994) Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. Syst. Appl. Microbiol. 16, 609–628.

Motifs courts exceptionnels dans les génomes des archées et de leurs éléments extra-chromosomiques

1

Présentation de l'étude

L'accumulation des séquences de génomes complets permet, de mieux en mieux, d'analyser leurs propriétés globales. Notamment, l'identification de motifs présentant une fréquence ou une répartition inattendue dans les génomes a constitué une approche fructueuse. Pour ce faire, des méthodes statistiques robustes sont disponibles (ex : Robin et al., 2005). Ces approches ont par exemple permis d'identifier de bons candidats pour des fonctions liées à la terminaison de la réplication (Hendrickson and Lawrence, 2006). Couplées à une validation expérimentale, elles ont de plus contribué à l'identification de motifs fonctionnels non-codants (voir Tableau 7), par exemple les motifs Chi de *Staphylococcus aureus* et plusieurs espèces de *Streptococcus* (Halpern et al., 2007), le motif matS chez *Escherichia coli* (Mercier et al., 2008), et les motifs KOPS chez *Escherichia coli* également (Bigot et al., 2005 ; Levy et al., 2005). Il s'agissait là d'études ciblées. Dans un esprit différent, des études systématiques et aussi exhaustives que possible des motifs dans des génomes ont également été réalisées (Rocha et al., 1998 pour le génome de *Bacillus subtilis* ; Leung et al., 1996 pour les génomes d'Herpesvirus ; Elhai, 2001 pour le génome d'*Escherichia coli*).

1.1 Motifs fonctionnels non-codants et cycle cellulaire

Dans quels types de fonctions sont impliqués les motifs fonctionnels non-codants ? Plusieurs d'entre eux interviennent pour le bon déroulement des différentes étapes du cycle cellulaire : réplication de l'ADN, division, etc.

Motif	Protéines associées	Intervenant dans ...	Chez les autres bactéries ?	Réf.
DnaA boxes	DnaA	Initiation de la réplication	Autres bactéries dont *Bacillus subtilis*	a
Ter	Tus	Blocage de la fourche de réplication dans une seule direction	Autres bactéries dont *Bacillus subtilis* (Ter/RTP)	b
dif	XerCD	Monomérisation et décaténation du chromosome	Autres bactéries dont *Bacillus subtilis*, *Xanthomonas campestris*	c
KOPS	FtsK	Positionnement des chromosomes avant décaténation et division	*Bacillus subtilis* (SRS6/FtsK)	d
Chi	recBCD	Réparation de l'ADN par recombinaison notamment lors de la réplication	Autres bactéries dont *Bacillus subtilis* *Haemophilus influenzae* *Lactococcus lactis* *Staphylococcus aureus* espèces de *Streptococcus*	e
matS	matP	Organisation de la région Ter en macrodomaine	? ≈ *Bacillus subtilis* Spo0J-parS RacA-ram	f

Tableau 7 : Exemples de motifs fonctionnels non-codants chez *Escherichia coli*, et leur présence chez les autres bactéries. a : Yoshikawa and Ogasawara, 1991. b : Neylon et al., 2005. c : Blakely and Sherratt, 1994. d : Bigot et al., 2005 ; Levy et al., 2005. e : Smith et al., 1981. f : Mercier et al., 2008.

Les bactéries les mieux caractérisées de ce point de vue-là sont *Escherichia coli* (modèle pour les bactéries à Gram négatif) et *Bacillus subtilis* (modèle pour les bactéries à Gram postitif). Des exemples de motifs fonctionnels non-codant chez *Escherichia coli* sont présentés dans le Tableau 7 (voir également Duggin et al., 2008).

Chez les eucaryotes, les mécanismes sont généralement plus complexes et plus variables au sein du domaine. Certains sont pour l'instant moins compris, comme ceux de la terminaison de la réplication.

Chez les archées, encore très peu de mécanismes et de motifs de ce type sont connus. Comme pour les eucaryotes, le déroulement de la terminaison de la réplication n'est pas bien compris. En revanche, pour l'initiation, des origines de réplications ont été identifiées, ainsi que les protéines initiatrices (homologues des protéines Orc1/Cdc6, présentes chez les eucaryotes) et les motifs d'ADN qu'elles reconnaissent (ORB, « Origin recognition boxes ») (Robinson et al., 2004). Les séquences ORB ont été retrouvées tant chez les crenarchées que chez les euryarchées (Robinson et al., 2004), suggérant un mécanisme d'initiation globalement conservé pour l'ensemble du domaine des archées.

Certaines crenarchées, en particulier celles du genre *Sulfolobus*, pourraient constituer un très bon modèle simplifié pour l'étude des mécanismes de réplication chez les eucaryotes (Lundgren et al., 2004). Comme chez toutes les archées, le complexe de réplication est de type eucaryote. Mais de plus, la vitesse de polymérisation de l'ADN est comparable à celle des eucaryotes, et les fragments d'Okazaki sont également courts. La présence de multiples origines de réplication et de chromatine est une ressemblance supplémentaire. L'identification de nouveaux mécanismes chez ces crenarchées pourrait donc permettre de proposer des hypothèses pour les eucaryotes et de voir jusqu'où va la ressemblance.

Chez les euryarchées, mieux connaître les mécanismes qui régulent le cycle cellulaire, incluant une étape de ségrégation, permettrait de voir comment les protéines de la réplication, de type eucaryote, s'accordent avec les protéines de la division cellulaire de type bactérien (ex : protéine FtsZ, Bernander, 2003).

Un des moyens de progresser sur ces questions peut être l'identification de motifs exceptionnels dans les génomes d'archées.

1.2 Motifs palindromiques

Un autre type de motifs non-codants, souvent sous-représentés dans les génomes de bactéries et de leurs phages, sont les palindromes courts (4 et 6 pb). Leur évitement s'explique en partie par l'existence des systèmes de restriction-modification de type II (Kobayashi et al., 1999 ; Gelfand and Koonin, 1997 ; Rocha et al., 2001). Ces systèmes sont fréquemment sujets aux transferts horizontaux. Ils permettent de cliver l'ADN au niveau du motif reconnu, lorsque celui-ci n'est pas protégé par la modification appropriée. L'ADN non-modifié est potentiellement d'origine exogène, aussi s'agirait-il d'un système de défense. Mais d'un point de vue évolutif, ces systèmes pourraient également se comporter comme des

éléments égoïstes (Rocha et al., 2001). Leur rôle et leur évolution chez les bactéries fait donc l'objet d'un débat. L'évitement de certains motifs palindromiques s'explique aussi par l'existence de certains systèmes de réparation de l'ADN, comme Vsr chez *Escherichia coli* (Bhagwat et McClelland, 1992). L'évitement de palindromes courts n'est pas limité aux bactéries, puisqu'il a également été décrit chez la levure (Lisnic et al., 2005). Enfin, les palindromes plus longs ou imparfaits pourraient être impliqués dans la régulation de divers processus cellulaires (Lisnic et al., 2005).

Aussi, une étude des motifs exceptionnels dans les génomes d'archées peut contribuer à l'identification de palindromes importants. Quelques systèmes de restriction ou de modification ont déjà été identifiés chez certaines archées, mais ils sont encore peu nombreux (voir 3.3.1).

Des génomes d'archées ont déjà été inclus dans plusieurs travaux (ex : 6 génomes d'archées sur un total de 26 dans Rocha et al., 2001 ; 1 génome sur un total de 2 dans Gelfand and Koonin, 1997). Mais *a priori*, aucune étude systématique des motifs exceptionnels ne leur a été spécifiquement dédiée. Par ailleurs, le nombre de séquences d'archées disponibles dans les bases de données devient conséquent et laisse entrevoir des possibilités intéressantes.

1.3 Approche développée

Encadrée par Sophie Schbath (Unité Mathématique, Informatique et Génome, INRA), et Patrick Forterre (BMGE, Institut Pasteur), et également avec Diego Cortez (BMGE, Institut Pasteur), nous avons analysé les motifs exceptionnels dans les génomes complets d'archées et de leurs éléments extra-chromosomiques (plasmides, virus). Nous avons utilisé le logiciel RMES (Hoebeke and Schbath, 2006 ; http://genome.jouy.inra.fr/ssb/rmes). Ont été incluses les séquences de 41 génomes d'archées, de 40 génomes de virus ou provirus, et 44 séquences de plasmides. Les mots de taille 4 à 8 ont été considérés.

Le travail comporte quatre parties. Dans la première partie, seules les séquences des génomes d'éléments extra-chromosomiques ont été considérées. Leur composition en mots a été étudiée dans le but de repérer les séquences homogènes et de les concaténer afin d'obtenir des séquences plus longues pour l'analyse des motifs exceptionnels (voir section 2.2). Cette première analyse a de plus débouché sur des observations intéressantes concernant la composition des séquences d'éléments extra-chromosomiques par rapport à celle de leurs hôtes. Dans un second temps, nous avons étudié en détail les motifs palindromiques, dont l'intérêt a été exposé à la section 1.2. Ensuite, nous nous sommes intéressés à une autre famille de motifs particuliers, les motifs symétriques, ce qui constitue une démarche nouvelle. Enfin, nous nous sommes concentrés sur les motifs très exceptionnels, en particuliers ceux qui sont exceptionnellement fréquents, car potentiellement impliqués dans des fonctions de maintenance du génome similaires à celles présentées en section 1.1.

Les tableaux de données produits lors de l'analyse sont regroupés à la fin de cette partie bioinformatique, pour permettre une comparaison plus aisée (Tableaux A à G), tandis que les tableaux plus synthétiques sont inclus dans le texte (Tableaux 8 à 11).

2
Méthodes

2.1 Séquences utilisées

Les génomes complets utilisés sont bien représentatifs de la diversité des archées cultivables, tandis que pour les éléments extra-chromosomiques, la plupart des séquences correspondent aux ordres Sulfolobales, Thermococcales et Halobacteriales (Tableau 8).

Ordre	**Hôtes**	**Plasmides**	**Virus et provirus**	**Total pour les éléments extra-chromosomiques**
Desulfurococcales	4			
Sulfolobales	4	21	24	45
Thermoproteales	5		3	3
Nitrosopumilales	1			
Methanococcales	3			
Methanosarcinales	5			
Methanomicrobiales	4			
Methanobacteriales	2			
Methanopyrales	1			
Archaeoglobales	1			
Halobacteriales	3	13	7	20
Thermoplasmatales	3			
Thermococcales	4	10	6	16
Nanoarchaeota	1			
Total	41	44	40	84

Tableau 8 : Répartition des séquences utilisées pour l'étude (en nombre de séquences). Caractères gras : ordres pour lesquels des séquences d'éléments extra-chromosomiques ont été inclus.

Les séquences de 41 génomes d'archées ont été obtenues à partir du site KEGG (http://www.genome.jp/kegg/catalog/org_list.html) :
Methanococcus maripaludis, *Methanococcus aeolicus*, *Methanococcus vannielii*, *Methanosarcina acetivorans*, *Methanosarcina barkeri*, *Methanosarcina mazei*, *Methanococcoides burtonii*, *Methanosaeta thermophila*, *Methanospirillum hungatei*, *Methanocorpusculum labreanum*, *Methanoculleus marisnigri*, *Candidatus Methanoregula boonei*, *Methanobacterium thermoautotrophicum*, *Methanosphaera stadtmanae*, *Methanopyrus kandleri*, *Archaeoglobus fulgidus*, *Halobacterium* sp. NRC-1, *Haloquadratum walsbyi*, *Natronomonas pharaonis*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, *Picrophilus torridus*, *Pyrococcus horikoshii*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *Thermococcus kodakaraensis*, *Aeropyrum pernix*, *Staphylothermus*

marinus, *Ignicoccus hospitalis*, *Hyperthermus butylicus*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Sulfolobus acidocaldarius*, *Metallosphaera sedula*, *Pyrobaculum aerophilum*, *Pyrobaculum islandicum*, *Pyrobaculum calidifontis*, *Pyrobaculum arsenaticum*, *Thermofilum pendens*, *Nitrosopumilus maritimus*, *Nanoarchaeum equitans*.

Les séquences de 40 génomes de virus ou provirus d'archées ont pour la plupart été obtenues à partir du site internet de NCBI. Il s'agit de :
Sulfolobus islandicus rod-shaped virus 2 (SIRV2), *Sulfolobus islandicus* rod-shaped virus 1, variant XX (SIRV1XX), *Sulfolobus islandicus* rod-shaped virus 1, variant VIII (SIRV1VIII), *Stygiolobus* rod-shaped virus (SRV), *Acidianus* rod-shaped virus 1 (ARV1), *Acidianus* filamentous virus 1 (AFV1), *Acidianus* filamentous virus 2 (AFV2), *Acidianus* filamentous virus 3 (AFV3), *Acidianus* filamentous virus 6 (AFV6), *Acidianus* filamentous virus 7 (AFV7), *Acidianus* filamentous virus 8 (AFV8), *Acidianus* filamentous virus 9 (AFV9), *Sulfolobus islandicus* filamentous virus (SIFV), *Acidianus* bottle-shaped virus (ABV), *Acidianus* two-tailed virus (ATV), *Sulfolobus tengchongensis* spindle-shaped virus 1 (STSV1), *Sulfolobus* spindle-shaped virus 1 (SSV1), *Sulfolobus* spindle-shaped virus 2 (SSV2), *Sulfolobus* spindle-shaped virus 4 (SSV4), *Sulfolobus* spindle-shaped virus 5 (SSV5), *Sulfolobus* spindle-shaped virus Ragged Hills (SSVrh), *Sulfolobus* spindle-shaped virus Kamchatka 1 (SSVk1), *Thermoproteus tenax* virus 1 (TTV1), *Sulfolobus* turreted icosahedral virus (STIV), *Pyrobaculum* spherical virus (PSV), *Thermoproteus tenax* spherical virus 1 (TTSV1), *Pyrococcus abyssi* virus 1 (PAV1), Halovirus HF1, Halovirus HF2, *Natrialba* phage ΦCh1, Archaeal BJ1, *Haloarcula* phage SH1, His1 virus, His2 virus, *Pyrococcus horikoshii* provirus 1 (PHV1), *Thermococcus kodakarensis* provirus 1 (TKV1), *Thermococcus kodakarensis* provirus 2 (TKV2), *Thermococcus kodakarensis* prophage 3 provirus TKV3), *Thermococcus kodakarensis* provirus 4 (TKV4).

Les 5 dernières séquences mentionnées correspondent à des provirus récemment intégrés au génome de l'hôte, et identifiés par Diego Cortez dans les séquences complètes des génomes des hôtes. La séquence de *Sulfolobus* spindle-shaped virus 6 (SSV6), non-publiée, a été gracieusement fournie par Peter Redder, David Prangishvili (BMGE), et Xu Peng, Roger Garrett (« Danish Archaea Center », Copenhagen). Les séquences des génomes de virus infectant des archées de l'ordre des Methanobacteriales, n'ont pas été incluses dans l'étude car en nombre trop faible (ψM2 et ψM100).

Les séquences de 44 plasmides d'archées ont pour la plupart été obtenues à partir du site internet de NCBI. Voici la liste, avec l'hôte indiqué entre parenthèses :
pTIK4 (*Sulfolobus neozealandicus*), pHEN7 (*Sulfolobus islandicus*), pTAU4 (*Sulfolobus neozealandicus*), pSSVi (*Sulfolobus solfataricus*), pRN1 (*Sulfolobus islandicus*), pRN2 (*Sulfolobus islandicus*), pXZ1 (*Sulfolobus islandicus*), pDL10 (*Acidianus ambivalens*), pORA1 (*Sulfolobus neozealandicus*), pSSVx (*Sulfolobus islandicus*), pSOG2 (*Sulfolobus islandicus*), pHVE14 (*Sulfolobus islandicus*), pING1 (*Sulfolobus islandicus*), pARN3 (*Sulfolobus islandicus*), pKEF9 (*Sulfolobus islandicus*), pARN4 (*Sulfolobus islandicus*), pNOB8 (*Sulfolobus* sp. NOB8H2), pAH1 (*Acidianus ambivalens*), pSOG1 (*Sulfolobus*

islandicus), pTC (*Sulfolobus tengchongensis*), pIT3 (*Sulfolobus solfataricus*), pGT5 (*Pyrococcus abyssi*), pRT1 (*Pyrococcus* sp. JT1), pTN1 (*Thermococcus nautilus*), pL47 (*Haloquadratum walsbyi*), pNG700 (*Haloarcula marismortui*), pNG600 (*Haloarcula marismortui*), pNG500 (*Haloarcula marismortui*), pNRC200 (*Halobacterium* sp. NRC-1), pNRC100 (*Halobacterium* sp. NRC-1), pL131 (*Natronomonas pharaonis*), pNG400 (*Haloarcula marismortui*), pNG300 (*Haloarcula marismortui*), pNG200 (*Haloarcula marismortui*), pNG100 (*Haloarcula marismortui*), pL23 (*Natronomonas pharaonis*).

La séquence du chromosome ChrII de *Haloarcula marismortui* a également été incluse et traitée comme un plasmide. Les séquences des plasmides suivants, non-publiées, nous ont gracieusement été fournies par Matthieu Gonnet et Daniel Prieur (Laboratoire de Microbiologie des Environnements Extrêmes, Bretagne), ainsi que Nicolas Soler et Patrick Forterre (BMGE Orsay) : pP12-1 (*Pyrococcus* sp. 12/1), pAMT11 (*Thermococcus kodakarensis* ou *gammatolerans*), pIRI33 (*Thermococcus gammatolerans*), pTN2 (*Thermococcus nautilus*), pCIR10 (*Thermococcus gammatolerans*), pT26-2 (*Thermococcus* sp. 26/2), pIRI48 (*Thermococcus gammatolerans*). La séquence de pMB1 (*Methanosarcina barkeri*) n'a pas été incluse, car seule représentante des éléments extra-chromosomiques de Methanosarcinales, un nombre insuffisant.

2.2 Analyse en composante principale (ACP) et clustering hiérarchique (CH)

Ces deux méthodes ont été utilisées pour examiner l'homogénéité des séquences d'éléments extra-chromosomiques, puis décider des séquences à concaténer. En effet, l'analyse statistique des motifs exceptionnels est d'autant plus riche que les séquences sont longues (voir 2.3). Pour chaque taille de mot comprise entre 2 et 6, un tableau comportant, pour chaque virus et plasmide, la fréquence de tous les mots, a été préparé.

Pour l'ACP, la fonction dudi.pca du logiciel R a été utilisée. Les données n'ont été ni centrées, ni réduites, puisque les valeurs étaient du même ordre de grandeur, par définition comprises entre 0 et 1. Pour le CH, la fonction hclust du logiciel R a été utilisée. Trois distances différentes, (« single », « ward », « complete ») ont été testées. Elles ont donné des résultats légèrement différents, mais ces différences pouvaient être interprétées.

2.3 Détection des motifs exceptionnels

Un motif est considéré comme exceptionnel dans une séquence donnée s'il possède un comptage exceptionnel dans cette séquence. Par comptage exceptionnel, on entend un comptage significativement plus grand ou plus petit que le comptage auquel on pourrait s'attendre dans d'autres séquences ressemblant à la séquence donnée. La démarche classique (Robin et al., 2005) consiste à considérer des séquences aléatoires ayant en moyenne la même composition en mots de taille 1 à $m+1$ que la séquence observée (modèle de chaîne de Markov d'ordre m, noté Mm) et de comparer le comptage observé d'un mot à son

comptage attendu sous ce modèle Mm. Plus précisément, cette comparaison se fait *via* le calcul d'une probabilité critique (« p-value » en anglais) définie comme la probabilité qu'une séquence aléatoire (modèle Mm) possède autant d'occurrences du mot considéré que la séquence observée.

Le logiciel RMES (Hoebeke and Schbath, 2006) a été utilisé pour le calcul des p-values ou plus exactement des scores d'exceptionnalité qui s'obtiennent de façon bijective à partir des p-values. Ainsi, un score positif (resp. négatif) correspondra à un mot sur-représenté (resp. sous-représenté) ; un score positif (resp. négatif) très grand en valeur absolue correspondra à un mot exceptionnellement fréquent (resp. rare). Ces scores d'exceptionnalité seront souvent visualisés sous la forme de nuage de points grâce à l'application java rmesplot.jar.

Puisque le modèle Mm s'ajuste sur la composition des mots de taille 1 à $m+1$ de la séquence analysée, on pourra s'intéresser à rechercher des mots exceptionnels de taille $L>m+1$. Réciproquement, si l'on souhaite s'intéresser à des mots de taille L, on pourra considérer les modèles d'ordre 0 à $L-2$. Ce dernier modèle, d'ordre $L-2$, est dit maximal pour l'étude des mots de taille L. En effet, c'est le plus gros (au sens de l'information dont il tient compte sur la séquence observée) que l'on puisse utiliser pour évaluer si un mot de taille L a un comptage attendu ou non (la réponse serait systématiquement oui dans des modèles d'ordre supérieurs). L'avantage de considérer le modèle maximal pour chaque longueur L de mot est d'identifier des mots exceptionnels de taille L dont le comptage (très fort ou très faible) ne peut s'expliquer par les comptages de leurs sous-mots.

Néanmoins, l'ordre *m* d'un modèle Mm que l'on peut correctement ajuster sur une séquence de taille *n* dépend de cette taille n. En effet, le nombre de paramètres à estimer pour le modèle Mm est 3×4^m ; il est classiquement raisonnable d'exiger 100 fois plus de données que de paramètres, donc on imposera que $n\geq 300\times 4^m$. Typiquement, pour $m=2$, il faudra $n\geq 4800$ tandis que pour $m=4$, il faudra $n\geq 76800$. Ainsi, étant donnée la séquence analysée (et donc sa longueur n), on déterminera le plus gros modèle raisonnable à utiliser.

En résumé, si nous disposons d'une séquence de longueur n =100 000 pbs, alors nous pourrons considérer les modèles *M0* à *M4* ; nous étudierons donc les mots de taille 6 dans le modèle *M4*, les mots de taille 5 dans *M3*, les mots de taille 4 dans *M2* etc. Dans la suite, nous nous focaliserons sur les mots de taille 4 à 8.

Enfin, le calcul de la p-value nécessite de connaître la loi du comptage d'un mot dans une chaîne de Markov. RMES utilise deux approximations de cette loi : une loi gaussienne (Schbath, 1997) ou une loi de Poisson composée (Schbath, 1995) selon que le comptage attendu du mot est suffisamment grand (gaussienne) ou non (Poisson composée). Il n'y a pas

de règle théorique pour établir le domaine de validité exact de ces deux approximations, aussi nous utilisons le critère empirique (courant) suivant :

$$\text{si } \frac{n}{4^L} > 400 \quad ,$$

alors on utilise l'approximation gaussienne ; sinon on utilise l'approximation de Poisson composée. Noter que la quantité :

$$\frac{n}{4^L}$$

est le comptage attendu d'un mot de taille L dans le modèle qui suppose que tous les mots de taille L sont équiprobables.

2.4 Exceptionnalité d'ensembles de mots dans les séquences

Afin de déterminer si les mots d'un type particulier (ex : ensemble des motifs palindromiques ou symétriques d'une taille donnée) sont généralement évités ou préférés dans les séquences, nous avons compté le nombre de ces motifs présentant un score négatif, et ceux présentant un score positif. Nous avons ensuite calculé la p-value associée à cette répartition des signes des scores. La méthode est décrite plus en détail ci-dessous.

Soit E la famille des mot que l'on souhaite étudier, et n_{total} le nombre de mots que contient cette famille. Si les mots de E ne sont pas spécialement exceptionnels, c'est-à-dire ni évités, ni préférés, chacun d'entre eux a autant de chances d'avoir un score d'exceptionnalité positif que négatif (probabilité ½) pour une séquence considérée. Dans ce cas, la variable aléatoire du nombre de mots de E présentant un score positif dans une séquence suit une loi binomiale de paramètres (n_{total} , ½). Il en est de même pour la variable aléatoire du nombre de mots de E présentant un score négatif

On peut donc calculer la p-value associée au nombre de mots de E observés dans la séquence étudiée et présentant un score positif, c'est-à-dire la probabilité qu'il y ait au moins autant de mots de E avec un score positif, selon la loi binomiale décrite ci-dessus. Et on peut faire de même pour les mots de E présentant un score négatif.

2.5 Significativité de l'évitement du mot CCTGGG dans certaines régions des génomes de *Pyrococcus*

Dans la séquence des génomes d'espèces du genre *Pyroccocus*, le motif CCTGGG est absent de régions de taille allant jusqu'à environ 30 000 pb, alors qu'on le trouve en moyenne environ toutes les 1 000 pb. Pour déterminer si cette absence du mot est significative, il faut estimer la probabilité qu'aucune occurrence du motif ne soit trouvée dans une région de taille supérieure ou égale à celle observée (p-value). Comme le mot CCTGGG n'est pas

recouvrant, une bonne approximation pour la p-value est :

$$e^{\frac{-N_{total} \times L_{région}}{L_{séquence}}},$$

où N_{total} est le nombre total de motifs CCTGGG comptés dans la séquence, $L_{région}$ est la longueur de la région du génome ne comprenant pas le motif, et $L_{séquence}$ est la longueur totale de la séquence.

Cette formule découle du fait que l'on considère que les occurrences du mot CCTGGG arrivent le long de la séquence selon un processus de Poisson homogène d'intensité estimée par $\frac{N_{total}}{L_{séquence}}$.

2.6 Position privilégiée des mots par rapport aux gènes

Pour déterminer si certains motifs exceptionnels occupent une position privilégiée par rapport au début des gènes, la méthode FADO, décrite dans Gusto and Schbath, 2005, a été utilisée.

3.1 Composition en mots courts des séquences de virus et de plasmides

La composition en mots des séquences d'éléments extra-chromosomique a été examinée dans le but de concaténer les séquences homogènes, et ainsi avoir des séquences plus longues pour la recherche de mots exceptionnels (Voir section 2.3). Afin d'identifier les groupes de séquences homogènes, nous nous sommes basés sur la composition en mots de taille 2 à 6, et nous avons utilisé deux méthodes pour visualiser les résultats : l'analyse en composante principale (ACP) et le clustering hiérarchique (CH) (voir section 2.2).

3.1.1 Observations générales

Les deux méthodes utilisées (ACP, CH) ont produit des résultats cohérents (Figures 19 et 20), *grosso modo* indépendants de la taille de mot considérée. Ils étaient toutefois meilleurs, pour le CH, avec une taille de mot supérieure ou égale à 3. Pour l'ACP, Figure 19, les points proches du demi-axe des X de la partie gauche sont riches en mots contenant des A et des T. Les points proches du demi-axe des X situé dans la partie droite sont riches en mots contenant des G et des C. L'axe des Y ne peut pas être décrit en termes aussi simples.

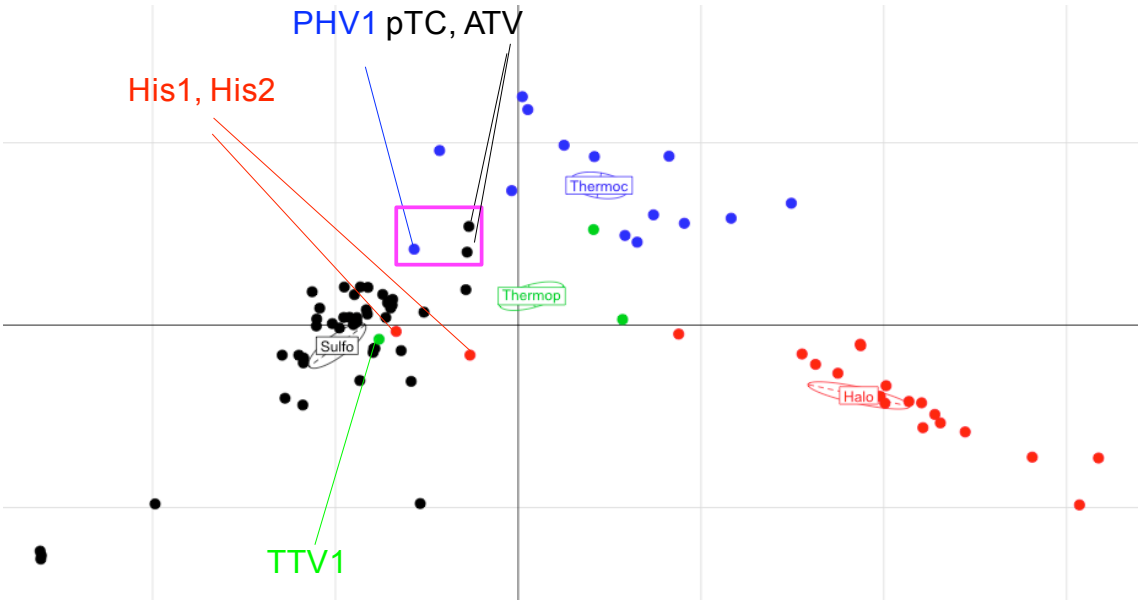


Figure 19 : ACP de la composition en mots de 4 lettres des séquences d'éléments extra-chromosomiques. Projection sur les deux premiers axes. Axe des X : 61% de la variance. Axe des Y : 9% de la variance. d = 0,02. Noir : virus et plasmides infectant des hôtes de l'ordre Sulfolobales. Bleu : virus et plasmides infectant des hôtes de l'ordre Thermococcales. Vert : virus et plasmides infectant des hôtes de l'ordre Thermoproteales. Rouge : virus et plasmides infectant des hôtes de l'ordre Halobacteriales. Cadre magenta : virus et plasmide situés dans une zone frontière.

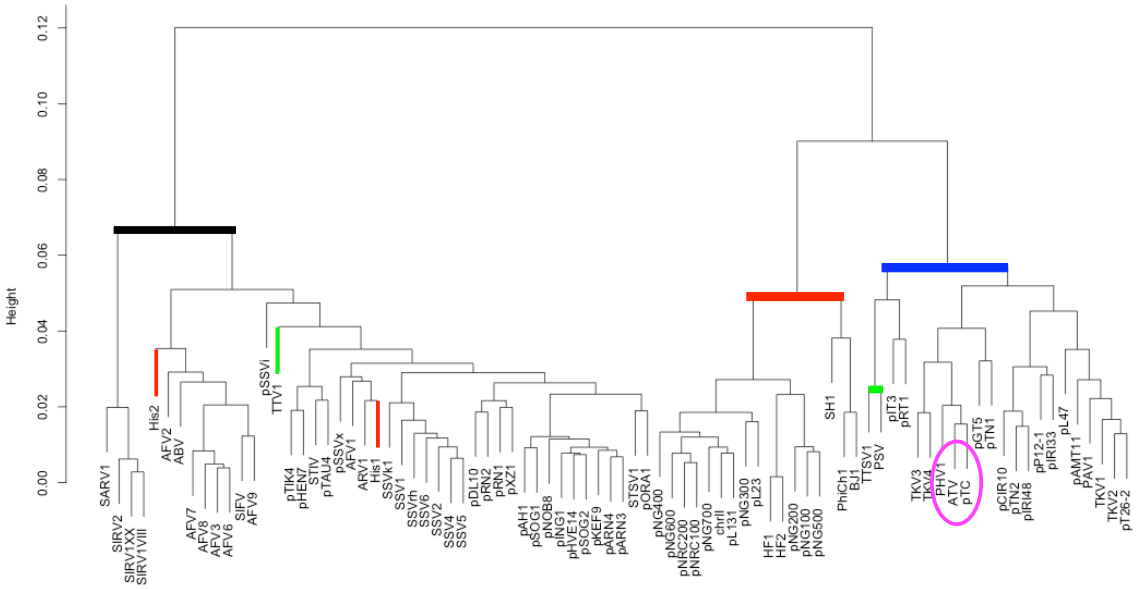


Figure 20 : Analyse par clustering hiérarchique de la composition en mots de 4 lettres des séquences d'éléments extra-chromosomiques. Distance utilisée : « complete ». Trait noir épaissi : cluster dont les virus et plasmides infectent en majorité des hôtes de l'ordre Sulfolobales. Bleu : de même, mais pour l'ordre Thermococcales. Vert : de même, mais pour l'ordre Thermoproteales. Rouge : de même, mais pour l'ordre Halobacteriales. Magenta : cluster de virus et plasmide dont la position n'est pas très stable et fluctue selon la taille des mots étudiés.

3.1.2 Interprétation et concaténation des séquences

Les séquences forment des groupes qui reflètent très bien les ordres auxquels appartiennent les hôtes des virus et des plasmides. On retrouve donc trois groupes correspondant bien aux virus et plasmides de Sulfolobales, Thermococcales et Halobacteriales. Quelques exceptions existent. Les virus d'Halobacteriales His1 et His2 sont systématiquement placés au milieu du groupe correspondant aux virus et plasmides de Sulfolobales, ceci avec les deux méthodes utilisées. Le provirus PHV, dont l'hôte appartient à l'ordre Thermococcales, le plasmide pTC (qui a un hôte Sulfolobales) et le virus ATV (qui infecte un hôte Sulfolobales) semblent être dans une zone frontière en terme de composition (Figure 19). L'analyse par CH les place donc tantôt dans le groupe correspondant aux éléments extra-chromosomiques de Sulfolobales, tantôt dans celui correspondant aux élements extra-chromosomiques de Thermococcales (ex : Figure 20). Enfin, les séquences de virus infectant des hôtes de l'ordre Thermoproteales (PSV, TTV1, TTSV1) sont trop peu nombreuses pour donner un résultat interprétable. TTV1 est toujours placé au milieu du groupe correspondant aux virus et plasmides de Sulfolobales. PSV et TTSV1 semblent plus proches du groupe des virus et plasmides de Thermococcales et forment toujours une paire dans l'analyse par CH, indiquant une composition de séquence très proche. Ils partagent d'ailleurs 18 gènes homologues (Prangishvili et al., 2006b) et appartiennent à la famille (en attente d'approbation) Globuloviridae.

Pour les Sulfolobales, les séquences d'éléments extra-chromosomiques sont suffisamment nombreuses (voir Tableau 8) pour identifier cinq sous-groupes. Les sous-

groupes reflètent non plus l'ordre d'appartenance de l'hôte du plasmide ou du virus considéré, mais la famille d'appartenance du virus (Rudiviridae, Lipothrixviridae, Fuselloviridae) et le type de plasmide (conjugatif ou cryptique) (voir Figure 21).

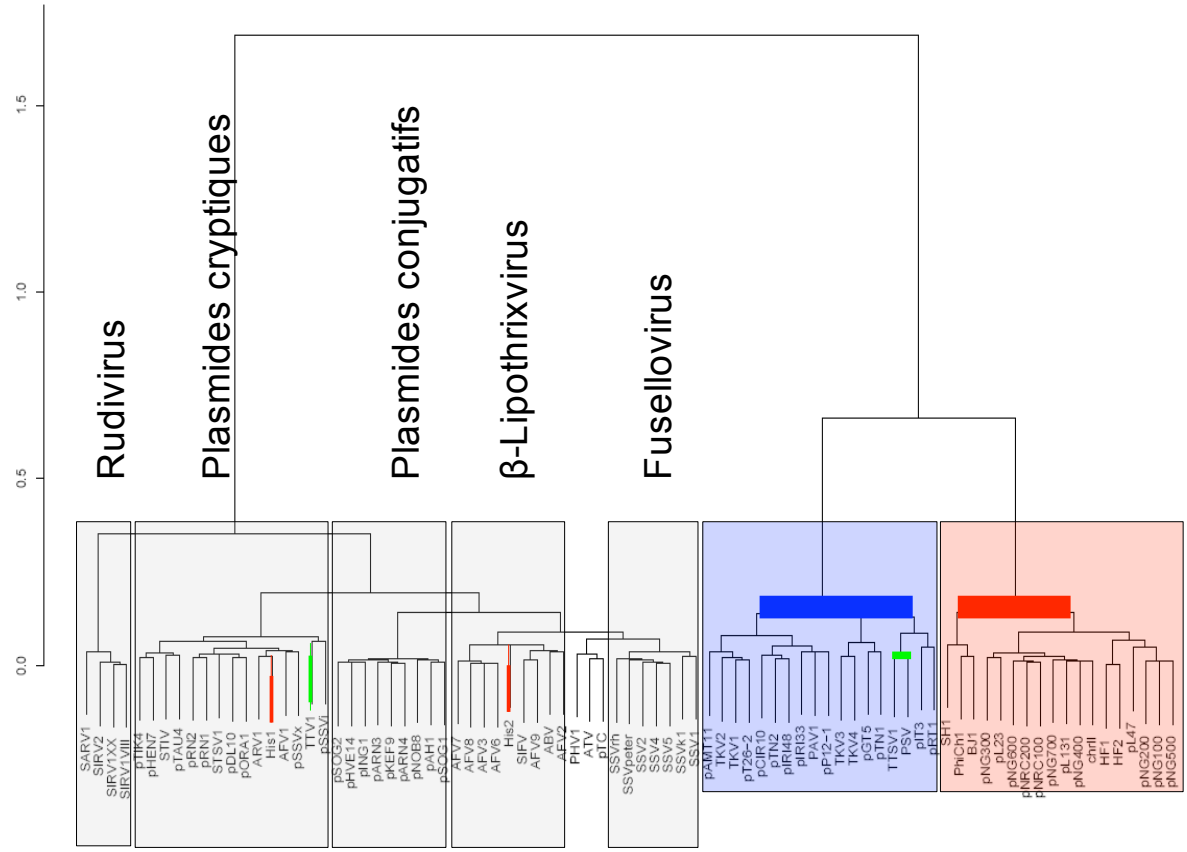


Figure 21 : Analyse par clustering hiérarchique de la composition en mots de 4 lettres des séquences d'éléments extra-chromosomiques. Distance utilisée : « ward ». Couleur noire : clusters dont les virus et plasmides infectent en majorité des hôtes de l'ordre Sulfolobales. Bleu : de même, mais pour l'ordre Thermococcales. Vert : de même, mais pour l'ordre Thermoproteales. Rouge : de même, mais pour l'ordre Halobacteriales. Les clusters d'éléments extra-chromosomiques de Sulfolobales reflétant des familles virales et des types particuliers de plasmides sont indiqués par des rectangles gris, avec le nom correspondant.

Après un examen attentif des résultats obtenus par les deux méthodes pour les mots de différentes tailles, les séquences ont été concaténées, comme indiqué dans les tableaux suivants (Tableaux 9 et 10). L'homogénéité de composition étant le premier critère retenu, toutes les séquences disponibles n'ont pas nécessairement été incluses (ex : ATV et ARV1 ne font partie d'aucun sous-groupe de virus de Sulfolobales). *A contrario*, certaines séquences ont été incluses dans le groupe correspondant aux virus et plasmides Sulfolobales, bien que cela contredise la taxonomie de l'hôte infecté par le virus ou plasmide considéré (ex : virus His1, His2, provirus TTV1 dans le Tableau 9). La séquence concaténée des virus et plasmides de Sulfolobales (SULFO_EXTRA) est la plus AT-riche, tandis que celle correspondant aux virus et plasmides d'Halobacteriales (HALO_EXTRA) l'est le moins. Parmi les sous-groupes de virus et plasmides de Sulfolobales, on peut remarquer celui correspondant à une partie des rudivirus (RUDI), particulièrement AT-riche. Les séquences

des génomes de pTC et STSV1 ont également été concaténées sous le nom TENGCHONG (données non-présentées dans les Tableaux 8 et 9), et incluses dans l'étude.

	Sulfolobales (SULFO_EXTRA)	Thermococcales (THERMOC_EXTRA)	Halobacteriales (HALO_EXTRA)
Nombre de séquences	47	14	18
Taille obtenue/pb	1 140 647	187 437	2 186 274
%A / G / C / T	32 / 18 / 18 / 32	26 / 26 / 22 / 25	21 / 29 / 29 / 21
% AT / GC	64 / 36	51 / 49	42 / 58
Noms des séquences concaténées	SIRV2, SIRV1XX, SIRV1VIII, SRV, ARV1 AFV1, AFV2, AFV3, AFV6, AFV7, AFV8, AFV9, SIFV SSV1, SSV2, SSV4, SSV5, SSV6, SSVrh, SSVk1 STIV, STSV1, ABV, ATV His1, His2 TTV1 pSOG1, pSOG2, pNOB8, pKEF9, pAH1, pHVE14, pING1, pARN3, pARN4 pTIK4, pHEN7, pTAU4, pSSVi, pRN1, pRN2, pXZ1, pDL10, pORA1, pSSVx pTC	PAV1, PSV, TTSV1 pIT3, pAMT11, pP12-1, pGT5, pRT1, pIRI33, pTN1, pTN2, pCIR10, pT26-2, pIRI48	HF1, HF2, Φch1, BJ1, SH1 ChrII, pNG700, pNG600, pNG500, pNG400, pNG300, pNG200, pNG100 pNRC200, pNRC100 pL131, pL47, pL23

Tableau 9 : Séquences concaténées, reflétant les ordres auxquels appartiennent les hôtes des virus et des plasmides. Les noms soulignés indiquent les virus qui infectent un hôte d'un autre ordre que le groupe auquel ils ont été incorporés. En lettres capitales, le nom donné aux séquences concaténées.

	Plasmides conjugatifs (CONJUG)	Plasmides cryptiques (CRYPTIC)	Lipothrixvirus (LIPOTHRIX)	Fusellovirus (FUSELLO)	Rudivirus (RUDI)
Nombre de séquences	9	10	6	7	4
Taille obtenue/pb	266 428	76 680	236 324	110 274	129 499
%A/G/C/T	32/18/18/31	34/20/17/29	31/17/19/33	31/20/19/31	38/13/13/36
% AT/GC	63 / 37	63 / 37	64 / 36	61 / 39	74 / 26
Noms des séquences concaténées	pSOG1, pSOG2, pAH1, pARN3, pARN4, pING1, pNOB8, pKEF9 pHVE14	pRN1, pRN2, pDL10, pORA1, pXZ1, pTIK4, pHEN7, pTAU4, pSSVi, pSSVx	AFV3, AFV6, AFV7, AFV8, AFV9, SIFV	SSV1, SSV2, SSV4, SSV5, SSVrh, SSV6, SSVk1	SIRV2, SIRV1XX, SIRV1VIII, SRV

Tableau 10 : Séquences concaténées, reflétant les familles auxquels appartiennent les virus et plasmides de Sulfolobales. En lettres capitales, le nom donné aux séquences concaténées.

3.1.3 Remarques sur l'évolution des virus et plasmides

Le fait que les virus et plasmides de Sulfolobales soient groupés par famille (plasmides conjugatifs, fusellovirus, etc.) suggère que l'origine évolutive commune de ces éléments extra-chromosomiques est prépondérante par rapport à l'adaptation à une espèce ou un genre d'hôte particulier. Ceci concerne, bien sûr, la composition en mots des séquences. Prenons l'exemple de *Sulfolobus islandicus*. Cette espèce est infectée par des rudivirus (SIRV1, SIRV2), par un lipothrixvirus (SIFV), possède des plasmides conjugatifs (pSOG1, pARN4, etc.) et cryptiques (pHEN7, pRN2, etc.). Mais ces divers éléments ne forment pas un groupe homogène et sont bien répartis dans les groupes correspondant aux différentes familles de virus et plasmides (Figure 21). Ceci est encourageant quant à l'utilisation des signatures génomiques pour la classification de séquences d'origine inconnue (ORFans, projets de métagénomiques).

Par ailleurs, les seuls virus halophiles à ne pas se placer dans le groupe attendu sont His1 et His2. Ils sont systématiquement placés dans le groupe correspondant aux virus et plasmides de Sulfolobales. Ce sont aussi les seuls virus fusiformes, tous les autres étant de type tête-queue ou ayant une capside icosaédrique. Compte-tenu de la remarque exposée au paragraphe précédent, on pourrait émettre l'hypothèse que les ancêtres d'His1 et His2 infectaient des hôtes de l'ordre des Sulfolobales, et qu'il s'agit, en quelques sortes, d'un transfert vers les Halobacteriales. Ce possible transfert serait alors postérieur à la divergence entre crenarchées et euryarchées. Un autre virus fusiforme, PAV1, est lui toujours placé dans le groupe prévisible des virus et plasmides de Thermococcales, ce qui indique probablement une coévolution ancienne avec les hôtes de cet ordre.

3.2 Analyse des palindromes

3.2.1 Exceptionnalité des palindromes correspondant à des systèmes de restriction-modification connus

Afin d'estimer si les palindromes exceptionnellement évités peuvent correspondre à des systèmes de restriction-modification, comme chez les bactéries, nous avons noté les rangs d'exceptionnalité des palindromes correspondant aux quelques systèmes déjà connus chez les archées. Le rang 1 correspond au mot le plus exceptionnellement rare du génome, tandis que le rang le plus élevé correspond au mot le plus exceptionnellement fréquent dans la séquence. Le rang a été noté pour le génome de l'organisme qui avait été étudié, ou, si la séquence n'était pas disponible, pour le génome de l'organisme le plus proche. Les résultats sont synthétisés dans le Tableau 11. Il nous a de plus semblé intéressant de signaler l'existence du système de restriction associé au motif GCAGC/GCTGC chez les bactéries du genre *Thermus*, qui sont hyperthermophiles (Tableau 11). En effet, il y a eu de nombreux transferts horizontaux entre archées hypethermophiles et bactéries hyperthermophiles (Aravind et al., 1998) et peut-être existe-t-il des systèmes de restriction-modification communs.

Motif	Organisme et référence biblio.	Séquence utilisée ici	Rang(s)	Nb d'occurences
CCWGG	*Pyrococcus* sp. strain GI-H Morgan et al., 1998.	*P. abyssi* *P. furiosus* *P. horikosii*	506/591 895/885 944/857	2257/2283 1981/1943 2166/2111
GGCC	*Methanobacterium thermoformicicum* (plasmides) Nölling and de Vos, 1992b.	*M. thermoautotrophicum*	177	8745
GGCC	*Sulfolobus acidocaldarius* Grogan, 2003.	***S. acidocaldarius***	1	570
GTAC	*Pyrococcus abyssi* Watanabe et al., 2006.	***P. abyssi***	246	4530
CTAG	*Methanobacterium thermoformicicum* Nölling and de Vos, 1992a.	*M. thermoautotrophicum*	6	612
GATC	ΦCh1 (virus) Baranyi et al., 2000.	**ΦCh1**	36	645
GCWGC	*Thermus* genus (bacteria) Welch and Williams, 1995.			

Tableau 11 : Rang d'exceptionnalité de palindromes correspondant à des systèmes de restriction-modification (ou de méthylation) déjà connus chez les archées. Le rang 1 correspond au mot le plus évité. Il existe 256 mots différents de 4 lettres et 1024 mots de 5 lettres. Caractères gras : signalent que la séquence utilisée pour l'étude correspond exactement à l'organisme pour lequel le système a été caractérisé. Motif souligné : activité de méthylation uniquement, pas d'activité nucléase observée. W : base A ou T.

Ces résultats confirment que les palindromes courts associés à des systèmes de restriction-modification peuvent être très évités. C'est en effet le cas pour CTAG chez *Methanobacterium thermoautotrophicum* (Tableau 11, rang 6 sur 256), bien que cette espèce ne soit pas celle qui ait été utilisée pour la caractérisation du système de restriction-modification associé au motif CTAG chez *Methanobacterium*. Cela est particulièrement le cas pour GGCC chez *Sulfolobus acidocaldarius* (Tableau 11, rang 1 sur 256 et Figure 22, axe des Y) et avait déjà été signalé (Chen et al., 2005). Mais ce n'est pas systématique, et le motif GTAC est au contraire sur-représenté chez *Pyrococcus abyssi* (Tableau 11, rang 246 sur 256). Pour résumer, repérer les palindromes exceptionnels dans les génomes d'archées peut fournir de bons candidats de motifs de restriction, mais la liste obtenue ne sera pas exhaustive. En outre, il faut garder en mémoire que les palindromes interviennent également dans d'autres types de fonctions (voir section 1.2).

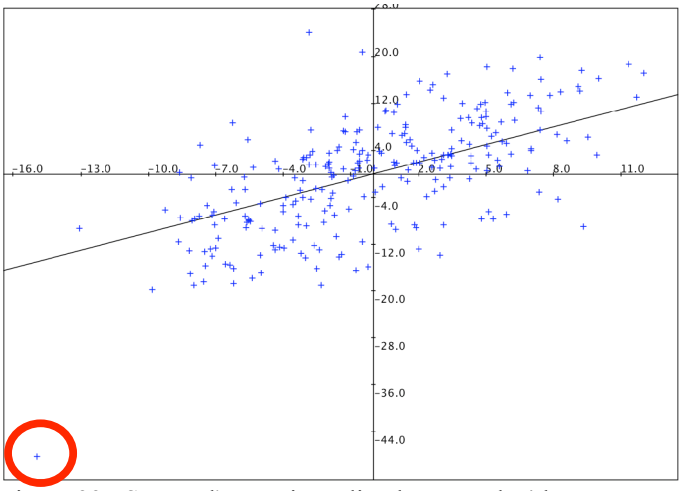


Figure 22 : Scores d'exceptionnalité des mots de 4 lettres pour les séquences concaténées LIPOTHRIX (axe des X) et pour *Sulfolobus acidocaldarius* (axe des Y). Entouré en rouge : le point correspondant au mot GGCC.

3.2.2 Rangs de palindromes et pseudo-palindromes de tailles 4 à 6 dans les séquences

La brève analyse précédente confirmant le fait que certains palindromes peuvent être très exceptionnels dans les séquences des génomes d'archées, nous avons systématiquement noté le rang de certains palindromes dans toutes les séquences des génomes d'hôte cellulaire, et dans les séquences concaténées d'éléments extra-chromosomiques. Les motifs palindromiques de taille 4, au nombre de 16, ont tous pu être examinés. Pour les mots plus longs, seuls les motifs apparaissant comme les plus exceptionnels ont été étudiés. Les tableaux contenant les données (Tableaux A, B, et C) se trouvent à la fin de cette partie (IV). Comme précédemment, le rang 1 correspond au mot le plus exceptionnellement rare du génome, tandis que le rang le plus haut correspond au mot le plus exceptionnellement fréquent dans la séquence.

De cette analyse, il ressort tout d'abord que, pour une séquence donnée, tous les palindromes ne sont pas exceptionnels. Seul un petit nombre d'entre eux sont très exceptionnels, de l'ordre de 2 à 6 par séquence, pour une longueur de mots donnée. Deux exemples sont montrés dans les Figures 23 et 24. Comme on pouvait s'y attendre, les motifs qui sont exceptionnels sont généralement évités. Mais il arrive que certains soient exceptionnellement fréquents : un exemple très marqué est le pseudo-palindrome CTGAG/CTCAC dans les séquences génomiques d'hôtes appartenant à l'ordre Thermoproteales (Tableau B). D'ailleurs, un même motif palindromique peut être exceptionnellement rare dans une séquence et exceptionnellement fréquent dans une autre. Un bon exemple est le couple CCTGG/CCAGG, qui correspond aux mots les plus évités dans les séquences concaténées CRYPTIC (Tableau B, Sulfolobales, rangs 1023 et 1024), et qui d'autre part correspond aux mots les plus sur-représentés dans la séquence génomique de

Methanosarcina acetivorans (Tableau B, Methanosarcinales, rangs 1 et 2). Ceci suggère que la plupart des motifs ont un certain niveau de spécificité, et qu'ils ne sont pas nécessairement reliés aux mêmes fonctions biologiques dans des organismes différents.

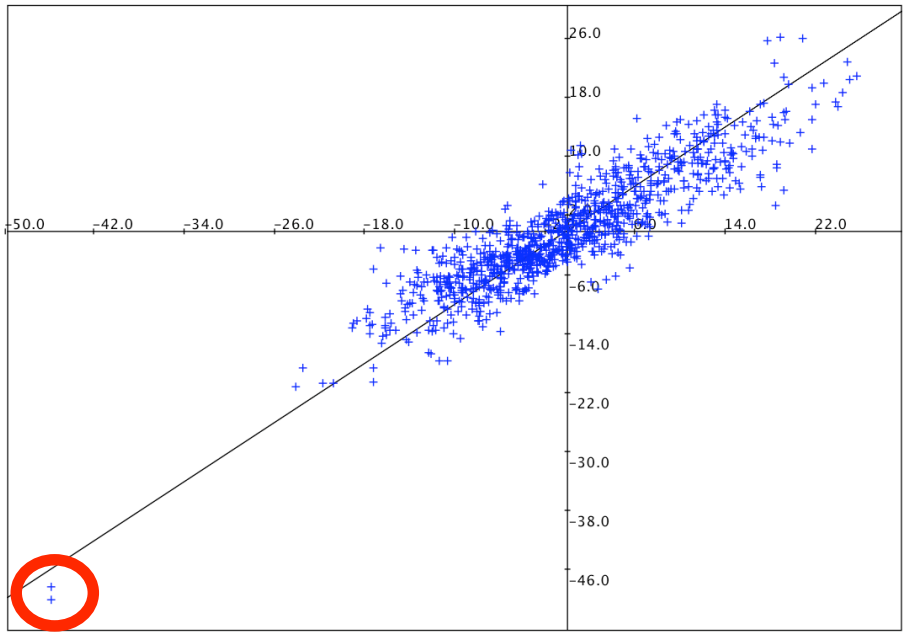


Figure 23 : Scores d'exceptionnalité des mots de 5 lettres chez *Pyrobaculum aerophilum* (axe des X) et *Pyrobaculum arsenaticum* (axe des Y). Entourés en rouge : points correspondant au couple de pseudo-palindromes GCTGC/GCAGC.

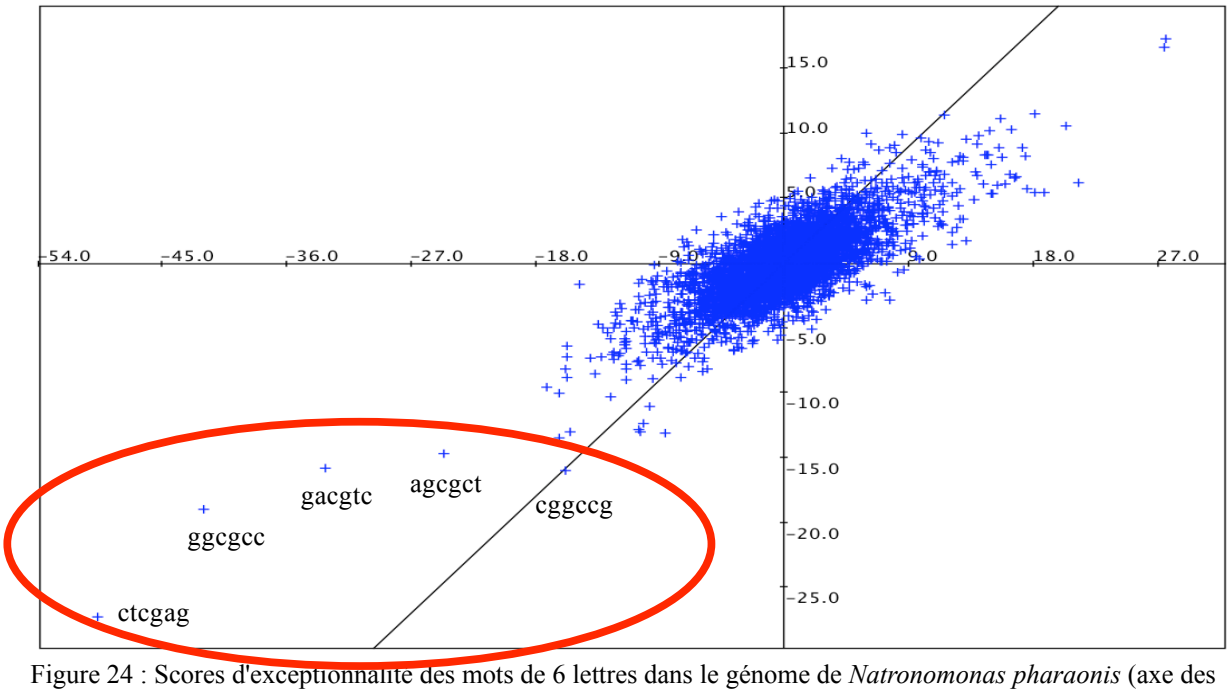


Figure 24 : Scores d'exceptionnalité des mots de 6 lettres dans le génome de *Natronomonas pharaonis* (axe des X) et les séquences concaténées HALO_EXTRA (axe des Y). Entourés en rouge : points correspondant à des palindromes très évités.

Il semble toutefois que les motifs présentent généralement les mêmes caractéristiques d'exceptionnalité dans les séquences correspondant à un même ordre phylogénétique (Thermoplasmatales, Methanococcales, etc.). En effet, les résultats sont généralement homogènes au sein d'un même ordre, y compris pour les séquences concaténées d'éléments extra-chromosomiques. On entend par là que ce sont les mêmes motifs très exceptionnels qui sont retrouvés dans ces séquences. Un exemple est visible dans la Figure 25, ou l'on voit que le mot GATC est le mot le plus évité à la fois dans le génome d'*Haloquadratum walsbyi* et dans les séquences concaténées HALO_EXTRA. C'est même l'ensemble des scores d'exceptionnalité qui apparaissent comme bien corrélés. La Figure 24, citée précédemment, est aussi un bon exemple pour HALO_EXTRA, cette fois avec les mots de taille 6. Globalement, c'est pour des ordres du phylum Crenarchaeota que l'homogénéité est la moins forte (Sulfolobales, Desulfurococcales). L'homogénéité est très bonne pour les ordres Methanosarcinales, Halobacteriales, et Thermoproteales.

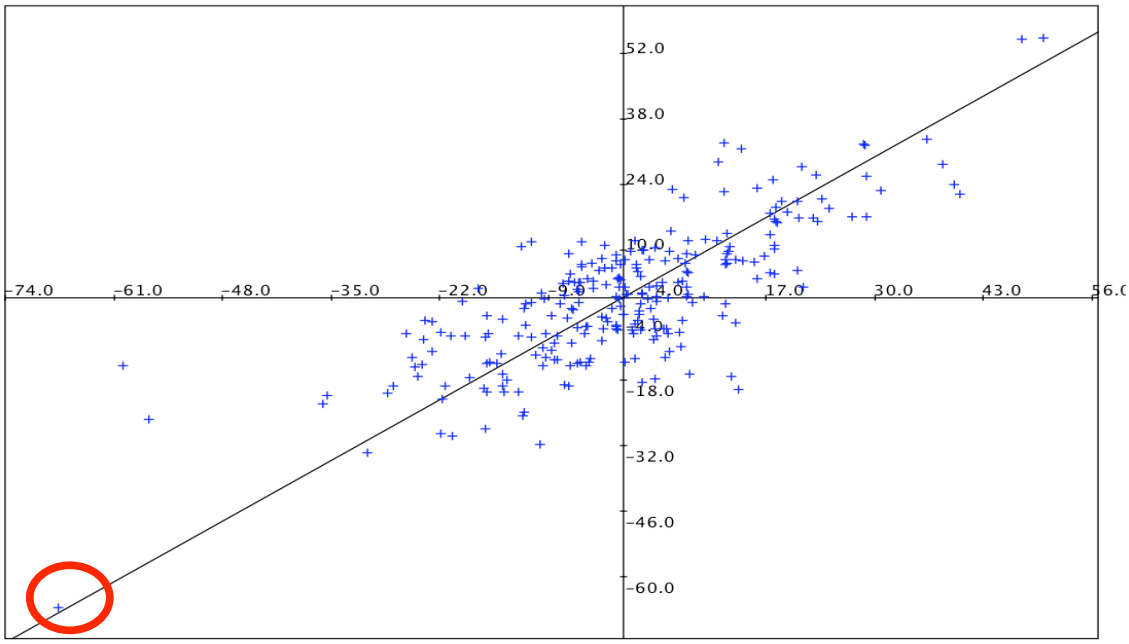


Figure 25 : Scores d'exceptionnalité des mots de 4 lettres dans les séquences d'*Haloquadratum walsbyi* (axe des X) et les séquences concaténées HALO_EXTRA (axe des Y). Entouré en rouge : point correspondant au palindrome GATC.

Un point intéressant est que pour les motifs pseudo-palindromiques de taille 5, quelques motifs sont exceptionnellement rares pour la plupart des séquences rattachées au phylum Euryarchaeota. Il s'agit notamment des 4 motifs GGTCC, GGACC, CCTGG et CCAGG (Tableau B). D'autres motifs, différents, sont exceptionnellement évités dans la plupart des séquences rattachées au phylum Crenarchaeota. Il s'agit notamment du couple GCAGC/GCTGC. Peut-être ces motifs sont-ils liés à des fonctions biologiques importantes conservées au sein de chaque phylum.

ORDRE	**Toujours SOUS-représentés**	**Souvent SOUS-représentés**	**Souvent SUR-représentés**	**Toujours SUR-représentés**
Desulfuroccocales	cctagg	tata cgatcg, (cgatcg)	(ttgcaa)	
Sulfolobales	gcgc gatatc	ggcc, gtac gcagc-gctgc, gcggc-gccgc, tttaa-ttaaa, taatta, cctagg, caattg		
SULFO_EXTRA		ttaa, ggcc, gcgc, gatc gcagc-gctgc, gcggc-gccgc taatta, gatatc	(cgtcg)	
Thermoproteales	**ggcc** **gcagc-gctgc**	cgcg, gcgc gcggc-gccgc, (gagtc) cacgtg, gacgtc	ttaa (cgacg-cgtcg) tatata, (ttgcaa)	ctgag-ctcag
Halobacteriales	ggcc, catg, gatc **ctcgag,** cgatcg	cgcg, agct, ctag gactc-gagtc, ctgag-ctcag, cctgg-ccagg gggccc, cacgtg, agatct, tcgcga, gacgtc	(gatatc)	
HALO_EXTRA	cgcg, agct, catg, ctag, gatc gcagc-gctgc, ggtcc-ggacc, cctgg-ccagg, gactc-gagtc, ctcag-ctgag cccggg, gggccc, cacgtg, **ctcgag**, gagctc, agatct, tcgcga, gacgtc, cgatcg, ggatcc		gatatc	
Methanococcales	tata, gatc, ctag cctgg-ccagg taatta, ctgcag, gatatc	cgcg, gtac ggtcc-ggacc ctcgag, gacgtc	atat, tgca ttaaa-tttaa	caattg
Methanomicrobiales	ggtcc-ggacc ctgag-ctcag gggccc, agatct, cgatcg	ggcc, ctag cctgg-ccagg, gactc-gagtc cccggg, ctcgag, ctgcag	ccgg	gatatc, ttgcaa
Methanosarcinales	ctag ggtcc-ggacc ctcgag, cgatcg	(tata), ggcc cctgg-ccagg cacgtg, gagctc, agatct, ctgcag, ggatcc	(ccgg) (gcagc-gctgc) caattg, tatata, gatatc, (ttgcaa)	
Thermococcales	ggtcc-ggacc, gactc-gagtc gatatc, cgatcg	tata, gcgc, **gatc** gcagc-gctgc gacgtc	(cgacg-cgtcg) (ttgcaa)	
THERMOC_EXTRA	atat, tata, catg, ctag, gatc gcagc-gctgc, gccgc, gagtc ctcgag, cccggg		(ctgag) tatata, taatta, (ttgcaa)	
Thermoplasmatales	**atat**, ctag agatct, cgatcg, ggatcc	tata, cgcg ggtcc-ggacc, cctgg-ccagg	gatc (cccggg)	

Tableau 12 : Palindromes et pseudo-palindromes exceptionnels de 4 à 6 lettres, dans les séquences d'archées et de leurs éléments extra-chromosomiques, selon leur ordre phylogénétique d'appartenance.

Finalement, un très grand nombre de palindromes très exceptionnels ont été identifiés, et il est impossible de les énumérer tous. Nous avons tenté de regrouper l'information dans un unique tableau, plus synthétique, ci-dessus (Tableau 12). Lorsque l'on compare au Tableau 11, dans lequel seuls les systèmes connus de restriction-modification sont indiqués, on voit que notre travail d'analyse ouvre de nombreuses pistes. Tout d'abord, des motifs déjà connus pour une espèce, ponctuellement, (ex : GGCC ou GATC), s'avèrent être exceptionnels dans un grand nombre d'autres séquences. En outre, des motifs totalement nouveaux ont été identifiés. Notamment, il semble que pas ou peu de fonctions associées à des motifs palindromiques de taille 6 soient encore connues chez les archées. Des motifs exceptionnels de taille 4 à 6 ont déjà été identifiés pour 6 génomes d'archées (Rocha et al., 2001). Le fait d'avoir ici plusieurs génomes pour chaque ordre permet en plus de voir à quel point les phénomènes d'exceptionnalité sont conservés. Lorsque c'est le cas, cela renforce la conviction que les observations ne sont pas le produit du hasard et qu'il doit bien exister des mécanismes biologiques expliquant l'exceptionnalité des différents motifs. Notre analyse concerne de plus un bien plus grand nombre d'ordres phylogénétiques.

Enfin, le fait que dans de nombreux cas, un même palindrome soit exceptionnel dans la plupart des séquences d'un même groupe phylogénétique, pourrait indiquer que certains palindromes sont reliés à des fonctions biologiques plus essentielles que la restriction-modification. Par exemple, chez *Escherichia coli*, comme brièvement mentionné en introduction, le système de réparation Vsr (« Very short patch repair ») corrige les mésappariements GT en GC, ce qui semble provoquer l'évitement de certains mots contenant un T, comme CTAG (Bhagwat and McClelland, 1992). Il n'est pas exclu que des mécanismes similaires existent chez les archées. De plus, la sur-représentation de certains palindromes est intrigante car inhabituelle, et renforce l'idée que les palindromes pourraient être reliés à une variété de fonctions biologiques.

3.2.3 Évitement global des palindromes

Après nous être intéressés aux motifs palindromiques individuellement, nous les avons considéré de manière globale : pour chaque séquence et pour chaque taille de mot, nous avons compté le nombre de motifs palindromiques présentant un score négatif (de même avec les scores positifs). Nous avons ensuite déterminé si cette répartition des signes des scores était attendue ou non (voir Méthodes, section 2.4). Cela permet par exemple d'estimer si les motifs palindromiques sont globalement évités, sans que, nécessairement, chacun d'entre eux soit exceptionnellement rare (ie n'ait un score très bas). Les mots palindromiques de taille 4 à 8 ont été considérés (lorsque cela était possible, voir Méthodes, section 2.3)

Les résultats sont synthétisés dans le Tableau F, à la fin du chapitre. On voit que les palindromes de 4 et 6 lettres sont globalement évités. L'exemple des mots de 4 lettres dans les séquences concaténées CONJUG et LIPOTHRIX est présenté dans la Figure 26 (A et B).

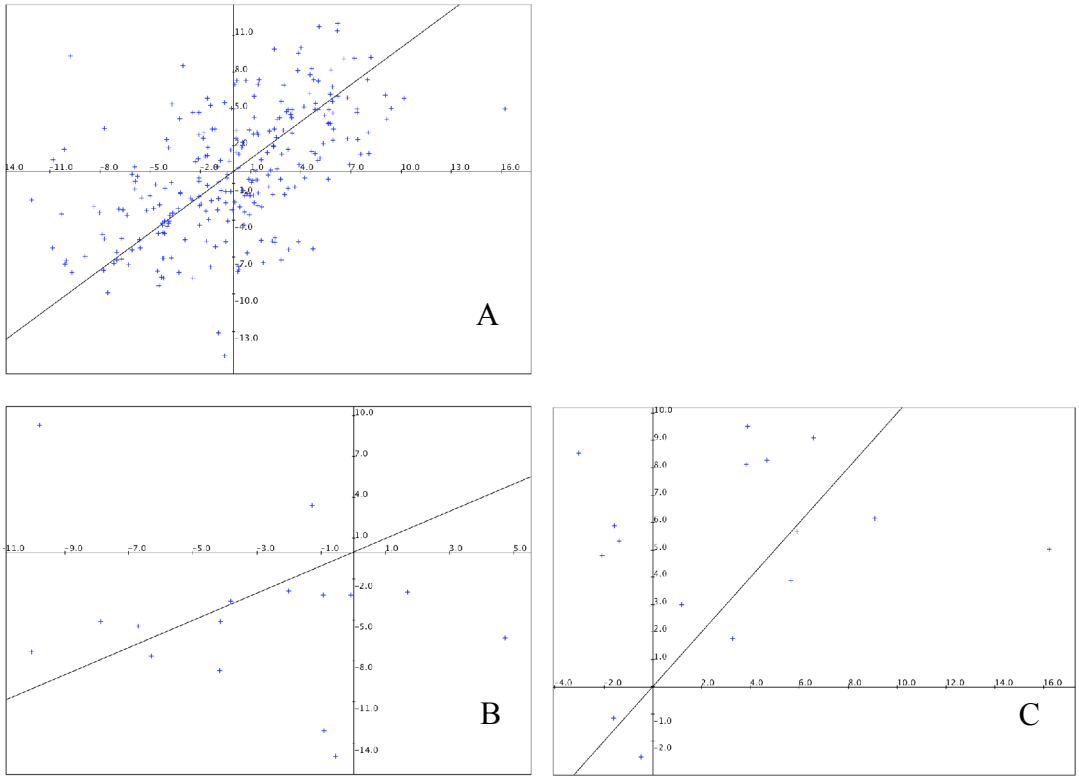


Figure 26 : Scores d'exceptionnalité des mots de taille 4 dans les séquences concaténées CONJUG (axes des X), et LIPOTHRIX (axe des Y). A : tous les mots. B : palindromes uniquement. C : mots symétriques uniquement.

Ce n'est toutefois pas systématique, ou bien ce n'est pas toujours aussi marqué. Étrangement, dans la séquence de *Methanopyrus kandleri*, c'est même le contraire qui est observé (Figure 27).

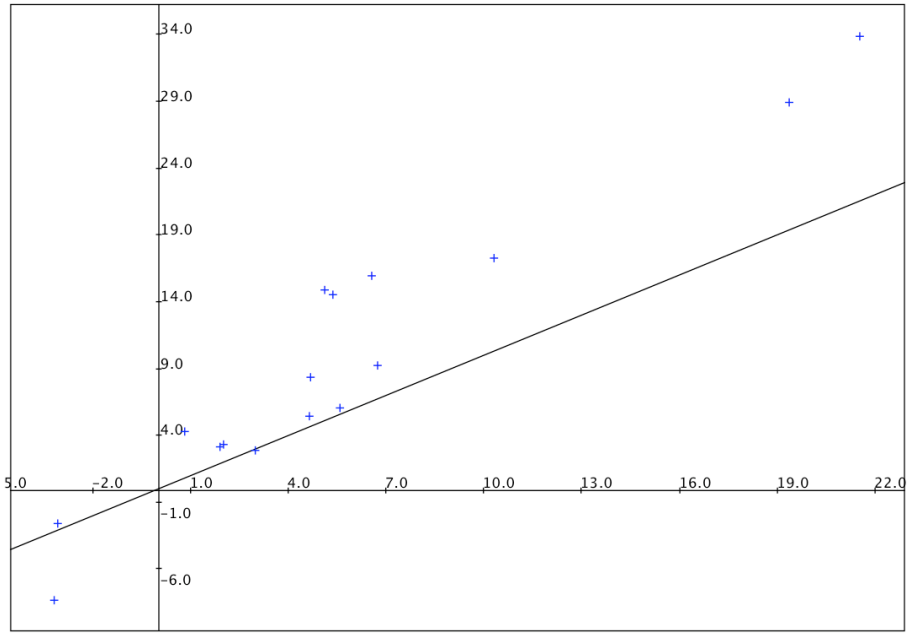


Figure 27 : Scores d'exceptionnalité des palindromes de 4 lettres dans le génome de *Methanopyrus kandleri*. Axe des X : « core genome », correspondant à un ensemble de gènes conservés concaténés. Axe des Y : génome complet.

Une observation intéressante est que les pseudo-palindromes de taille 7 sont tantôt globalement évités (Tableau F : séquences d'hôtes du genre *Pyrococcus*, certaines séquences de Methanosarcinales, etc.), tantôt globalement sur-représentés (Tableau F : Thermoproteales, certaines Halobacteriales, etc). Enfin, pour les palindromes de 8 lettres, il y a plutôt une sur-représentation, mais pas de manière très marquée.

En conclusion, les palindromes de 4 à 6 lettres sont globalement évités chez les archées. Ce phénomène a déjà été mis en évidence chez les bactéries (voir Introduction, section 1). Dans le cas des archées, ce n'est cependant pas aussi systématique et marqué. Les palindromes de 7 et 8 lettres sont même parfois sur-représentés. Il serait intéressant de déterminer l'origine de ces différences entre archées et bactéries.

3.3 Analyse des mots symétriques

Considérant que d'autres types de motifs réguliers pouvaient avoir de l'importance dans les génomes, nous avons examiné le cas des mots symétriques (exemple de mot symétrique : GATAG). Nous avons utilisé la même approche que celle décrite dans la section 3.2 (et Méthodes, section 2.4), c'est-à-dire que nous avons regardé s'ils sont globalement exceptionnels dans les génomes. Les mots symétriques de taille 4 à 8 ont été considérés (lorsque cela était possible, voir Méthodes, section 2.3). Les résultats sont exposés dans le Tableau G, placé à la fin de chapitre.

Les mots symétriques de taille 4 et 7 sont globalement très sur-représentés. Les mots de 8 lettres sont également globalement sur-représentés, mais le résultat est moins marqué. Enfin, les mots de 5 et 6 lettres sont globalement évités. Des exemples sont montrés dans les Figures 28, 29 et 26 (A et C).

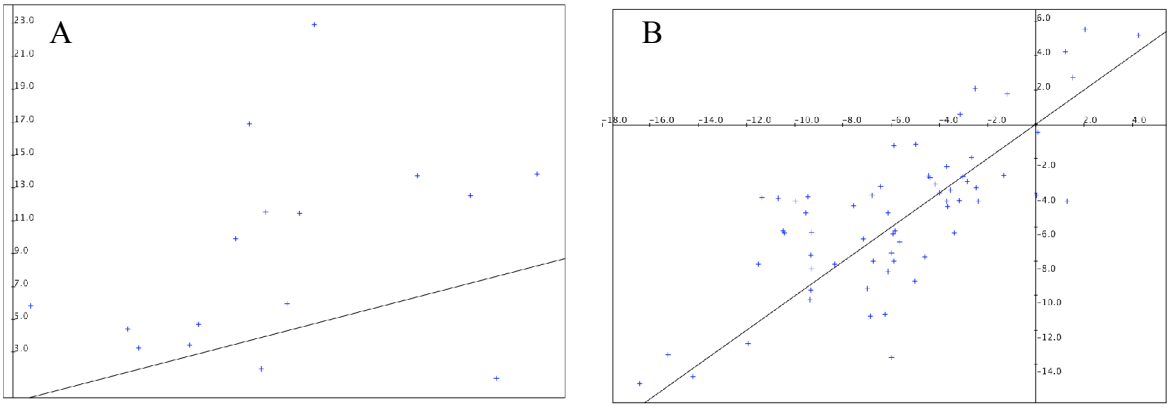


Figure 28 : Exemples d'exceptionnalité globale des mots symétriques. A : Scores des mots symétriques de 4 lettres, obtenus pour les séquences concaténées THERMOC_EXTRA (axe des X) et SULFO_EXTRA (axe des Y). B : Scores des mots symétriques de 5 lettres, obtenus pour les séquences de *Pyrobaculum aerophilum* (axe des X) et *Pyrobaculum arsenaticum* (axe des Y) ; voir Figure 23 pour le nuage correspondant à tous les mots de 5 lettres.

Les résultats sont particulièrement marqués pour les mots de 4 lettres : pour 11 séquences parmi toutes celles étudiées, la totalité des mots symétriques a un score positif (exemples Tableau G : *Staphylothermus marinus*, SULFO_EXTRA, *Pyrobaculum calidifontis*, *Methanococcus aeolicus*, etc. ; voir aussi Figure 28A). La comparaison des Tableaux F et G suggère même que la sur-représentation des mots symétriques de 4 lettres est plus forte que l'évitement des palindromes de 4 lettres. Par ailleurs, pour les mots symétriques de 7 lettres, une p-value inférieure à 10^{-18} est atteinte pour 10 séquences (ce qui apparaît comme 0,00E+0 dans le Tableau G), incluant par exemple celles d'*Aeropyrum pernix*, *Methanoculleus marisnigri*, *Methanosarcina thermophila*. Un exemple est présenté dans la Figure 29.

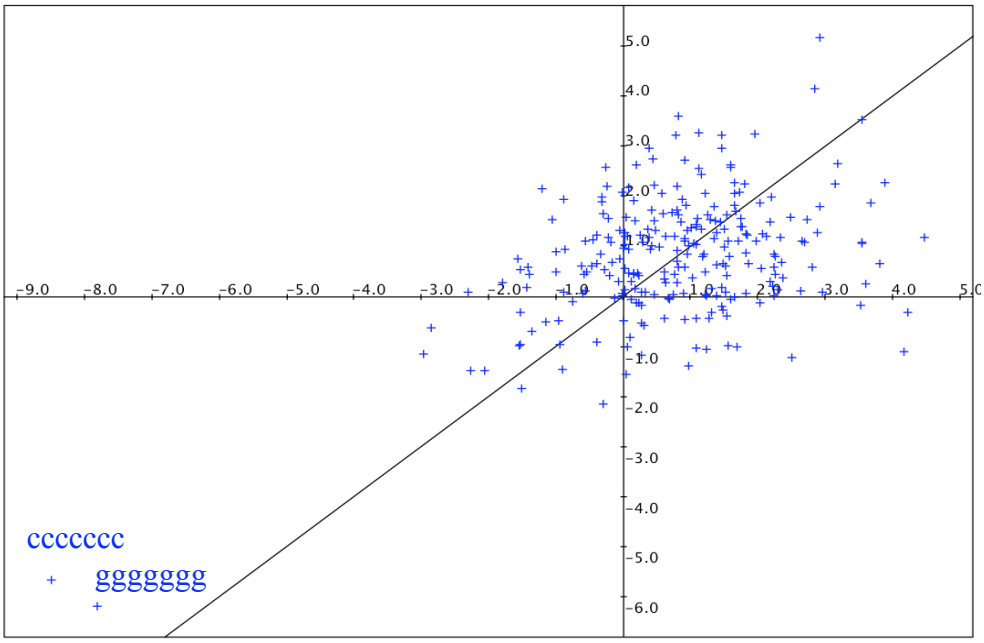


Figure 29 : Scores d'exceptionnalité des mots symétriques de 7 lettres chez *Ignicoccus hospitalis* (axe des X) et *Aeropyrum pernix* (axe des Y).

En conclusion, les mots symétriques ne doivent pas être anodins dans les génomes, puisqu'ils sont souvent globalement sur-représentés ou sous-représentés, ceci en fonction de leur taille. Selon nos connaissance, c'est la première fois que l'exceptionnalité des mots symétriques est spécifiquement étudiée dans des génomes. Pour l'instant, nous ne sommes pas en mesure d'expliquer les observations. Il se pourrait que les motifs symétriques soient impliqués dans des fonctions biologiques (régulation, etc.) ou bien que ces motifs confèrent, localement, des propriétés physico-chimiques particulières à l'ADN. Il serait intéressant de comparer avec les bactéries et les eucaryotes, pour voir s'il s'agit d'un phénomène général.

3.4 Analyse d'autres motifs très exceptionnels

3.4.1 A la recherche de motifs impliqués dans des fonctions générales, de type maintenance du génome

Les rangs de motifs exceptionnels supplémentaires, repérés lors d'un criblage, ont été notés. Les résultats sont présentés dans les Tableau D et E, à la fin de ce chapitre. De nouveau, les résultats sont relativement homogènes au sein de chaque ordre. Le motif TAGTAC est souvent évité dans les séquences de crenarchées et d'euryarchées. La paire GAAG/CTTC est fréquemment sur-représentée. Cette dernière observation pourrait correspondre à un phénomène général, car la paire GAAG/CTTC est également sur-représentée dans le génome d'*Escherichia coli* (Elhai, 2001). Il reste néanmoins difficile de repérer des motifs fonctionnels importants sur le plan biologique sans concevoir une approche spécifique.

D'après ce que l'on observe chez les bactéries, les motifs fonctionnels non-codant intervenant dans la maintenance du génome sont assez spécifiques du genre d'appartenance de l'organisme (ex : Halpern et al., 2007). Afin d'identifier des motifs candidats, nous nous sommes donc concentrés sur les genres comprenant plus de 3 génomes séquencés. Pour chaque genre, nous avons noté les 4 motifs les plus sur-représentés et communs à tous les génomes, et de même avec les motifs évités. Ont donc été retenus les genres suivants : *Methanococcus*, *Methanosarcina*, *Pyrococcus*, *Pyrobaculum* et *Sulfolobus*. Nous avons également inclus les espèces de l'ordre Halobacteriales, même si elles n'appartiennent pas toutes à un même genre. Les résultats sont présentés dans le Tableau 13.

Comme on pouvait s'y attendre, beaucoup de palindromes sont présents parmi les mots sous-représentés, et quelques motifs symétriques sont présents, tantôt évités, tantôt préférés, selon leur taille (voir section 3.3). Par ailleurs, de nombreuses répétitions d'une même lettre semblent évitées, et d'autres motifs sont apparus comme des candidats intéressants, notamment les motifs de taille supérieure ou égale à 5 lettres. Nous avons donc examinés quelques motifs plus spécifiquement.

Afin d'avoir des indices sur la fonction possible des motifs les plus intéressants (en gras dans le Tableau 13), nous avons utilisé la méthode FADO (voir Méthodes, section 2.6) pour déterminer si ces motifs occupent une place préférentielle par rapport au début des gènes. Mais les résultats n'étaient pas significatifs. Nous nous sommes ensuite basés sur la distribution des mots dans les génomes.

	SUR-représentés					SOUS-représentés				
	L4	L5	L6	L7	L8	L4	L5	L6	L7	L8
Methanococcus	*catc*	*ctttt*	*attttt*	*caaaatc*	acgccgcg	*taag*	*aattt*	ctgcag	*taaaaag*	*tttttttt*
(*maripaludis*, *aeolicus*, *vannielii*)	*gatg*	*aaaag*	*aaaaat*	*gattttg*	acgcgacg	*ctta*	*aaatt*	gaattc	*cttttta*	*aaaaaaaa*
	gaag	taagt	ctcttg	tttttta	cacgcgcg	tata	*taaga*	*tattta*	*taaattg*	*atttttta*
	cttc	gattt	cttttg	gaaaaag	agtcgcgc	gatc	*tctta*	*taatta*	*caattta*	*taaaaaat*
								gatttc		
nb	15	32	25	58	410	15	22	48	15	50
Methanosarcina	*cctg*	*ccaga*	*attttt*	***caagttc***	*ttttttg*	*ccaa*	*ggtcc*	*aaaaaa*	caacttg	*aaaaaaaa*
(acetivorans, *barkeri*, *mazei*)	*cagg*	*tctgg*	*ttttta*	***gaacttg***	*caaaaaaa*	*ttgg*	*ggacc*	*ttttt*	cttttta	*tttttttt*
	ttga	*tcagg*	cttttt	*aaaaaag*	*aaaaaaat*	ggcc	*ccagg*	aaaatt	caagcaa	caaaaaat
	atag	*cctga*	**ttggag**	*cttttt*	*attttttt*	ctag	*cctgg*	ggatcc	ttttttt	gttccata
							ttttt			
nb	7	8	10	10	100	7	6	8	8	150
Halobacteriales	*ggcg*	gcgtt	*ctcgaa*	gaccccg	ggactgct	ggcc	*caact*	ctcgag	*cctcgga*	*caccgcct*
(*halobacterium*, *h.walsbyi*, *n.pharaonis*)	*cgcc*	agttc	*ttcgag*	gagcgtg	gccgcgct	cgga	*agttg*	ggcgcc	*tccgagg*	*aggcggtg*
	catc	ggact	tccttg	aggcgtt	tcgggggtc	tgcg	cagct	agcgct	cgcgagt	ggcgcgtc
	gatg	gcgct	tccgtg	ctcgggg	agcgcgtc	gatc	ggcgc	tcgagc	tcggctg	ctccgcgc
nb	15	72	100	218	480	20	61	74	350	600
Pyrococcus	*ctgg*	***tcagc***	***cccagg***	gttcttg	cattttca	*tccg*	tctta	*aaaaaa*	tagcttg	taatagaa
(*horikoshii*, *abyssi*, *furiosus*)	*ccag*	***gctga***	***cctggg***	cgaattc	cacctcac	*cgga*	gcttc	*ttttt*	tagctcg	tagctaat
	gccg	gaaaa	taatag	accaggc	aataagac	*ctta*	gactc	*tccagg*	cggccta	taagaggc
	catc	taggg	atatac	gagcttg	tagtctta	*taag*	gaatc	*cctgga*	cccaggc	ctcctctc
nb	10	14	25	100	300	17	24	15	45	700
Sulfolobus	*gctg*	***catca***	*attttt*	gaaatag	*ggtagacc*	*attc*	*ggatg*	*aaaaaa*	*aaaaaaa*	gcttctta
(*solfataricus*, *acidocaldarius*, *tokodaii*)	*cagc*	***tgatg***	*aaaaat*	caaattc	*ggtatacc*	*gaat*	*catcc*	*ttttt*	*ttttttt*	taaggttt
	gaag	***ggatt***	aaatta	caatatc	ggttctcc	ctta	gaaac	taatta	taggatg	taagattt
	cttc	***aatcc***	taaatc	attttt	ctttcaat	gctc	aatca	gatatc	taaatag	tttatttg
	gttg									
nb	19	10	40	65	500	11	21	20	25	700
Pyrobaculum	*gccg*	aaggc	*aggggg*	***tgagaga***	*cccccccc*	cgcg	*gctgc*	*aaaaaa*	*ggggggg*	*taggcgct*
(*aerophilum*, *islandicum*, *calidifontis*, *arsenaticum*)	*cggc*	tctgc	*ccccct*	***tctctca***	*gggggggg*	ggcc	*gcagc*	*ttttt*	*ccccccc*	*agcgccta*
	gagg	gagga	gttttt	tctctaa	aacggcac	gctc	*gcctc*	*cccccc*	*cctctca*	gcctaggc
	cctc	gagct	**agccgc**	taagaga	tagactac	tccg	*gaggc*	*gggggg*	*tgagagg*	gcctctac
	cgcc									
nb	7	14	20	80	900	14	8	20	80	750

Tableau 13 : Motifs les plus exceptionnels (selon leur rang) communs à plusieurs espèces d'un même genre ou d'un même ordre.
nb : pour les mots sur-représentés (resp. évités), correspond approximativement au nombre de mots parcourus avant d'en obtenir 4 qui soient communs à toutes les séquences de la catégorie considérée, ceci en partant du mot ayant le score le plus élevé (resp. bas).
Fond vert : nb faible, peu de mots ont du être examinés avant de trouver 4 mots communs. Fond jaune pâle : nb élevé, beaucoup de mots ont du être examinés avant de trouver 4 mots communs. Ces couleurs sont indicatives, et les valeurs seuil choisies, indicatives également, dépendent de la taille des mots considérés.
Caractères bleus : palindromes. Oranges : mots symétriques. Gris : répétitions de une lettre, avec éventuellement une position occupée par une lettre différente. En italique : motifs dont l'inverse-complémentaire est également présent dans le tableau. Caractères gras : motifs considérés comme les plus prometteurs.

3.4.2 Répétitions d'une même lettre

D'après le Tableau 13, beaucoup de motifs évités sont des répétitions d'une même lettre (ex : AAAAAA, GGGGGGG, etc., en caractères gris dans le tableau). Une raison pourrait être que les répétitions constituent des points chauds mutationnels : elles favorisent les décalages du cadre de lecture lors de la réplication (Maki, 2002). Une exception notable est le genre *Pyrobaculum,* car les mots GGGGGGG et CCCCCCC sont sur-représentés dans les séquences de ce genre (Tableau 13).

A contrario, beaucoup de motifs sur-représentés sont des répétitions d'une même lettre finissant ou commençant par une lettre différente (ex : AGGGGG, AAAAAT, CTTTTTT, etc., en caractères gris dans le tableau). Il pourrait s'agir d'une compensation de l'évitement des répétitions strictes de la même lettre. En effet, d'un point de vue statistique, le

fort évitement d'un mot est nécessairement compensé par la sur-représentation globale de l'ensemble des mots identiques à une lettre près. Ceci a également été observé chez *Bacillus subtilis* (Rocha et al., 1998), et la même explication a été suggérée.

L'analyse des séquences concaténées d'éléments-extra-chromosomiques les plus longues suggère que le phénomène d'évitement des répétitions de une lettre y est bien moins marqué (données non présentées). Il pourrait donc s'agir d'une spécificité des chromosomes. Une analyse de la distribution du mot AAAAAA dans le génome de *Sulfolobus acidocaldarius* montre que les régions du génome où se trouvent les origines de réplications sont plus riches en ce motif que le reste du génome (Figure 30, A et B), suggérant un lien avec la réplication.

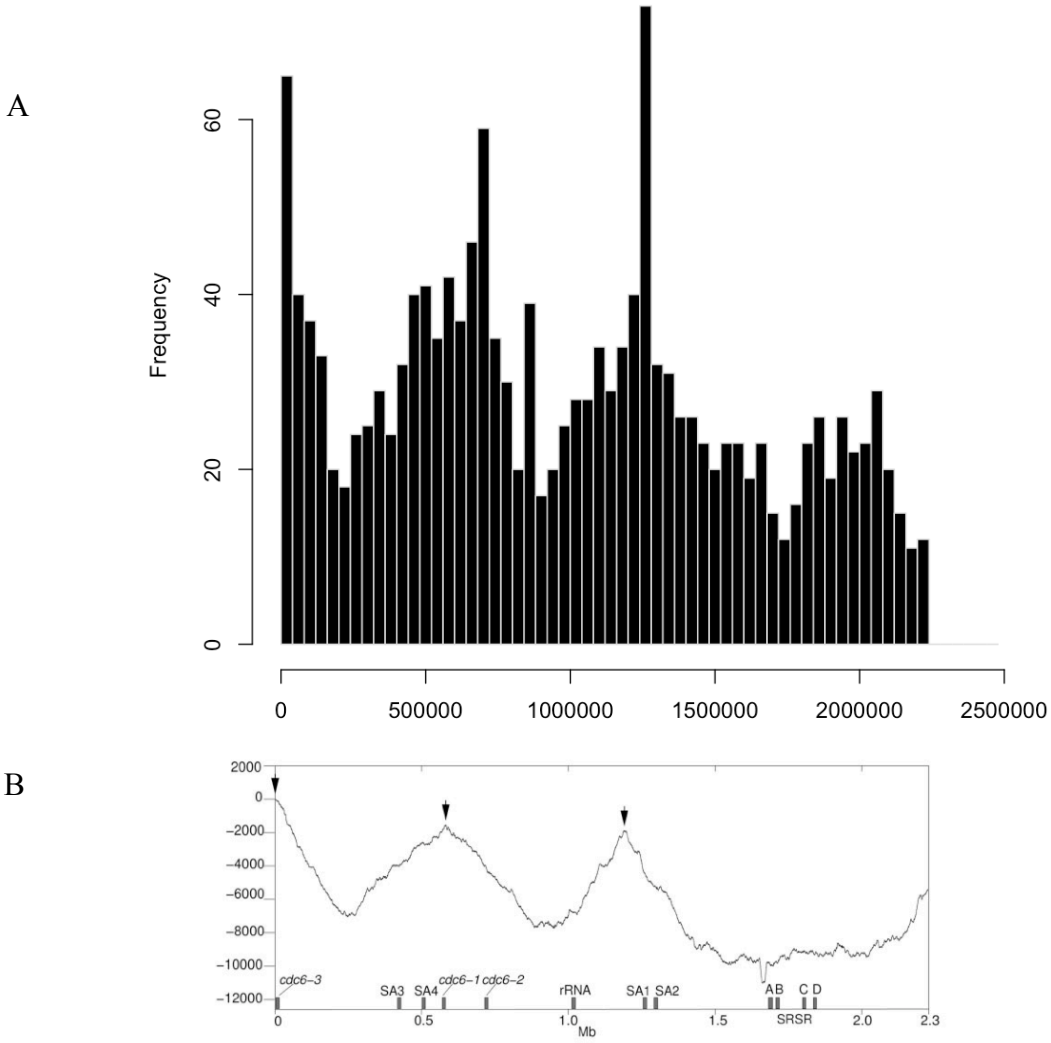


Figure 30 : Le motif AAAAAA dans le génome de *Sulfolobus acidocaldarius.* (A) : Distribution du motif, sur le brin séquencé. Une barre de l'histogramme correspond à un intervalle de 40000 pb (intervalles non-recouvrants). (B) Courbe-Z (composante Y) du génome de *Sulfolobus acidocaldarius*. Les 3 origines de réplications (flèches) et la position des gènes cdc6 sont indiquées. D'après Chen et al., 2005.

3.4.3 AATCC et les CRISPRs dans les génomes de *Sulfolobus*

L'analyse de la distribution de AATCC dans les génomes a permis de constater, dans le cas de *Sulfolobus solfataricus* et *Sulfolobus tokodaii*, que ce motif est présent dans les parties répétées de certains CRISPRs. En effet, les distributions présentent un ou quelques pics très localisés. Les distribution sur l'un des brins pour *Sulfolobus solfataricus* est montrée dans la Figure 31. Sur ce brin, le motif AATCC est présent dans une région de CRISPRs comprenant 102 répétitions (http://crispr.u-psud.fr/crispr/). Le motif AATCC est également présent dans les séquences répétées de CRISPRs chez *Sulfolobus acidocaldarius*, mais les régions correspondantes ne comptent pas plus de 10 répétitions, et ce n'est donc pas détectable sur la distribution. Le motif n'est pas présent dans les séquences répétées de CRISPRs d'espèces du genre *Pyrococcus*. La sur-représentation de AATCC ne s'explique donc pas uniquement par la présence de CRISPRs.

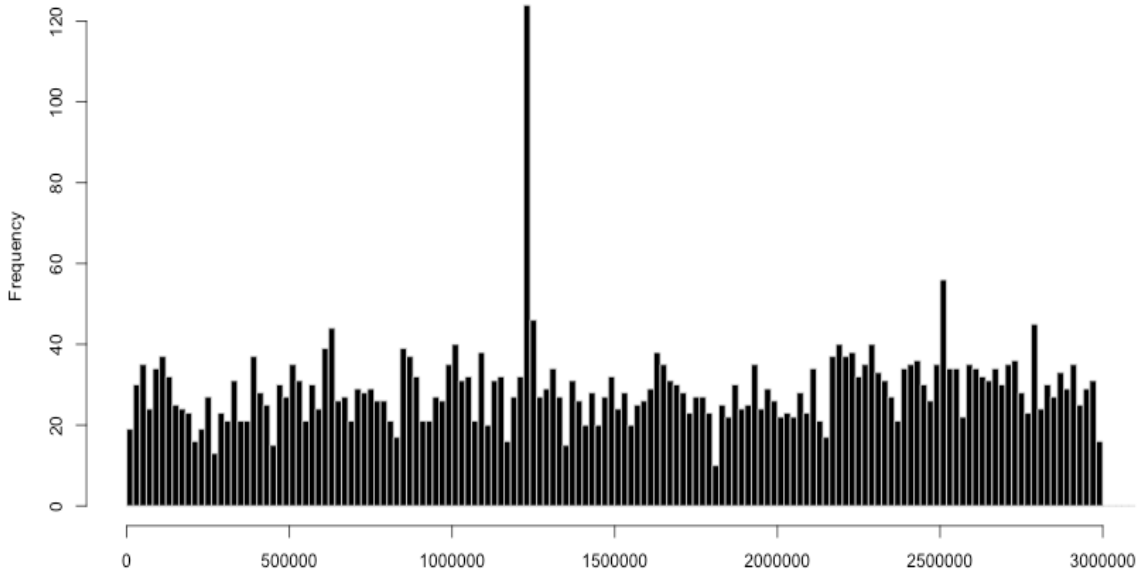


Figure 31 : Distribution du mot AATCC dans le génome de *Sulfolobus solfataricus*, sur le brin séquencé. Chaque barre de l'histogramme correspond à un intervalle de 20 000 pb (intervalles non-recouvrants). Nombre total d'occurrences sur le brin analysé : 4364.

3.4.4 CCTGGG/CCCAGG dans les génomes de *Pyrococcus*

Les mots CCTGGG/CCCAGG, dans les génomes d'espèces du genre *Pyrococcus*, ont donné des résultats intéressants. Tout d'abord, mentionnons qu'il s'agit d'un couple de mots extrêmement sur-représentés dans ces génomes (Figure 32).

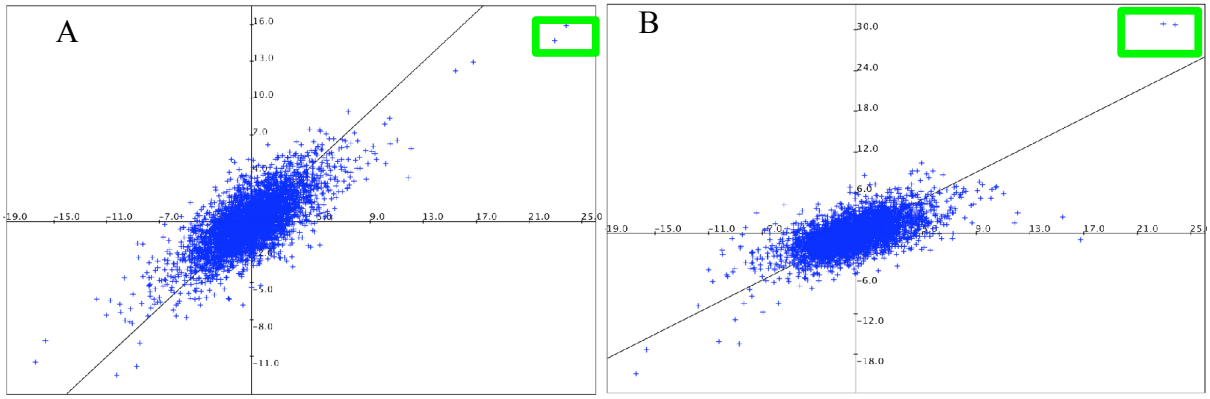


Figure 32 : Scores des mots de 6 lettres dans les génomes d'espèces du genre *Pyrococcus*. A : *Pyrococcus abyssi* (axe des X) et *Pyrococcus horikoshii* (axe des Y). B : *Pyrococcus abyssi* (axe des X) et *Pyrococcus furiosus* (axe des Y). Encadrés en vert : les points correspondant au couple de mots CCTGGG/CCCAGG.

L'analyse des distributions a montré que, bien que le mot soit sur-représenté, il est absent dans certaines régions des génomes, et cette absence est significative (voir Méthodes, section 2.5). Les résultats sont synthétisés dans le Tableau 14. Les zones d'évitement pourraient correspondre aux régions de terminaison de la réplication, mais on ne peut pas en être certain. En effet, la localisation du terminus n'est pas encore précisément connue.

	Génome	Ori	Ter	Brin non-séquencé					Brin séquencé				
				Nb	Dmoy	Ecart	p-valeur	Position	Nb	Dmoy	Ecart	p-valeur	Position
Pyrococcus abyssi	1765118	121000	700000-120000	1172	1506	13995	9,21E-005	130581	1193	1480	13477	1,11E-004	1609275
Pyrococcus furiosus	1908256	16000	700000-120000	1190	1604	19902	4,07E-006	789600	1175	1624	18623	1,05E-005	717294
											16095	4,97E-005	770645
Pyroccus horikoshii	1738505	109000	700000-120000	938	1853	29388	1,30E-007	1056455	886	1962	24672	3,46E-006	1086182
											22166	1,24E-005	375813
											13591	9,82E-004	434522

Tableau 14 : Régions d'évitement du mot CCTGGG dans les génomes d'espèces de *Pyrococcus*. Génome : taille du génome/pb. Ori : position approximative de l'origine de réplication. Ter : position très approximative du terminus de réplication. Nb : nombre d'occurences du mot sur le brin considéré. Dmoy : distance moyenne entre deux motifs CCTGGG, en pb. Ecart : distance entre deux motifs CCTGGG dans la région d'évitement considérée, en pb. p-value : significativité de l'évitement (voir section 2.5). Position : position du début de la zone d'évitement dans le génome, en pb.

La Figure 33 indique la distance qui sépare deux motifs successifs sur le brin non-séquencé de *Pyrococcus horikoshii*, tout le long du brin. Elle montre bien la principale zone d'évitement du mot, et correspond à l'exemple le plus clair.

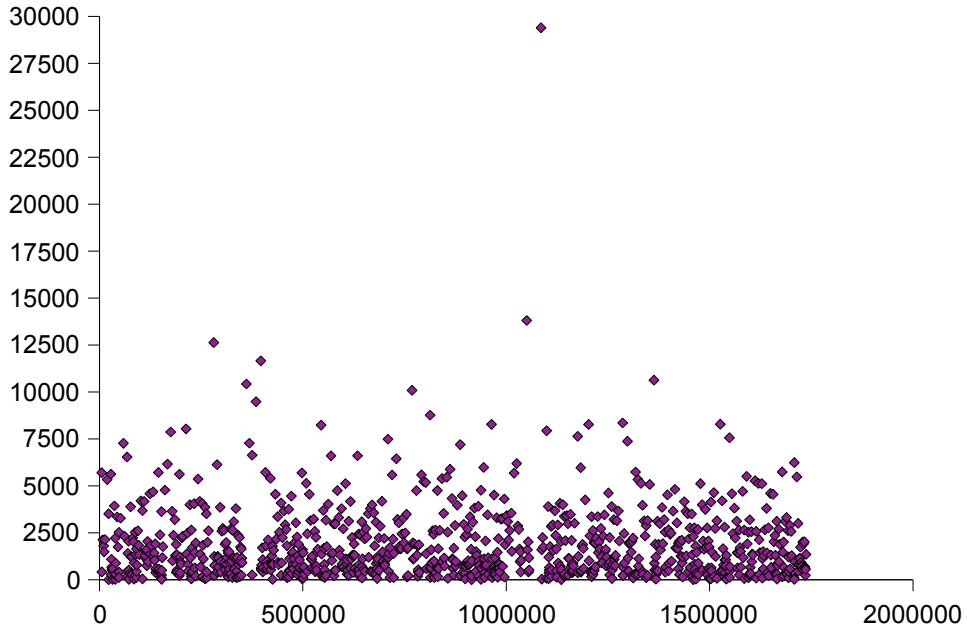


Figure 33 : Distance entre deux motifs successifs CCTGGG chez *Pyrococcus horikoshii*, sur le brin séquencé. Axe des X : position dans le génome (pb). Axe des Y : distance séparant le motif du motif précédent (pb).

En conclusion, CCTGGG/CCCAGG pourrait être un motif fonctionnel non-codant. L'évitement de ce mot dans certaines régions du génome pourrait être lié à sa fonction, ou pourrait refléter l'existence de transferts horizontaux récents. Toutefois, si les travaux de Zivanovic et al., 2002 montrent que des réarrangements nombreux sont survenus dans la région de la terminaison de réplication, l'existence de transferts horizontaux n'y est pas mentionnée.

4
Conclusions

Nous avons réalisé une des premières études systématiques des motifs exceptionnels dans les génomes d'archées et de leurs éléments extra-chromosomiques. Nous avons identifié de nombreux motifs palindromiques très exceptionnels, évités dans la grande majorité des cas. Or, on sait par des études réalisées sur les génomes bactériens (voir Introduction, section 1) que ce phénomène peut s'expliquer par l'existence de différents mécanismes biologiques, incluant les systèmes de restriction-modification de type II et le système de réparation Vsr. Nous avons ici retrouvé l'exceptionnel évitement de motifs de restriction déjà connus, notamment celui de GGCC chez *Sulfolobus acidocaldarius,* et proposons de nombreux candidats supplémentaires, potentiellement liés aux deux types de fonctions biologiques cités ci-dessus, ou, pourquoi pas, à d'autres fonctions. Par ailleurs, il est apparu que certains motifs palindromiques sont au contraire exceptionnellement fréquents chez les archées, et il serait intéressant d'en déterminer la cause.

Par ailleurs, nous avons remarqué le comportement particulier des mots symétriques, tantôt très sur-représentés et tantôt très évités selon leur taille, et ceci de façon très nette. Il ne nous semble pas que ce phénomène ait été décrit précédemment, et il serait intéressant de tester la généralité de cette observation sur des séquences correspondant aux deux autres domaines du vivant.

Enfin, nous avons réalisé une étude plus détaillée sur quelques motifs particulièrement intéressants, ce qui a permis d'observer des distributions irrégulières de ces mots dans les génomes (sections 3.4.2 à 3.4.4) et de les lier, soit à l'existence des CRISPRs (AATCC chez Sulfolobus), soit plutôt, à la position des origines ou terminus de réplication (CCTGGG/CCAGGG chez *Pyrococcus*, répétitions de la lettre A chez *Sulfolobus acidocaldarius*).

Une interprétation plus précise reste difficile car les connaissances sur les archées sont encore limitées sur tous ces aspects. Aussi, il pourrait être intéressant, à l'avenir, de cibler une question plus spécifique et de se limiter à certains organismes, afin de procéder à des études encore plus détaillées (ex : observation des différences entre brin direct et brin précoce, utilisation de séquences desquelles les gènes non-conservés ont été exclus, etc.), et éventuellement de coupler l'analyse statistique des séquences à un travail expérimental.

SEQUENCES	aatt	ttaa	atat	tata	ccgg	ggcc	cgcg	gcgc	acgt	agct	catg	ctag	gatc	gtac	tcga	tgca
Thaumarchaeota																
Nitrosopumilales																
n.maritimus	238	5	234	4	132	93	173	146	44	148	14	199	1	6	102	225
Crenarchaeaota																
Desulfurococcales																
s.marinus	1	199	34	2	171	78	157	48	153	42	120	219	254	6	178	91
a.pernix	122	217	13	70	145	3	72	12	144	14	151	73	31	77	249	197
h.butylicus	58	211	8	13	217	5	225	68	120	72	66	67	17	18	229	220
i.hospitalis	152	141	95	7	247	157	146	112	234	250	156	10	88	186	154	151
Sulfolobales																
s.solfataricus	65	49	132	20	239	12	72	17	139	163	33	66	23	24	118	194
s.tokodaii	85	20	151	27	179	24	99	14	97	111	21	46	11	15	142	159
m.sedula	164	210	97	26	67	113	138	14	149	236	256	245	86	63	112	133
s.acidocaldarius	102	29	152	6	130	1	42	20	80	84	37	46	36	7	135	212
SULFO_EXTRA	169	4	146	37	135	12	22	2	60	130	24	71	7	26	140	133
CONJUG	102	7	120	26	98	108	103	13	51	83	59	8	50	22	165	212
CRYPTIC	179	34	207	49	30	2	123	1	12	98	106	47	16	124	158	162
LIPOTHRIX	70	20	72	16	188	1	2	47	48	79	61	250	5	45	76	32
FUSELLO	127	8	80	97	172	24	31	1	27	68	23	4	2	21	187	164
RUDI	256	21	253	174	111	125	147	9	85	249	43	192	83	75	144	251
TENGCHONG	59	15	125	134	167	74	23	82	212	131	1	128	26	17	158	44
Thermoproteales																
p.calidifontis	206	244	187	122	6	1	5	14	221	240	163	29	88	140	174	183
p.aerophilum	190	248	154	136	30	1	2	13	197	226	198	45	193	54	64	173
p.arsenaticum	200	227	176	104	53	1	2	25	226	231	172	46	127	121	157	178
t.pendens	131	254	96	46	214	1	233	145	184	86	147	10	45	99	246	177
p.islandicum	221	232	206	136	70	1	2	13	195	246	140	44	81	20	162	211
Euryarchaeota																
Archaeoglobales																
a.fulgidus	127	122	31	86	56	7	5	6	134	182	11	2	1	10	79	126
Halobacteriales																
n.pharaonis	62	155	158	49	159	4	3	31	27	10	2	6	1	26	214	78
h.walsbyi	32	56	109	14	103	2	30	35	26	3	6	55	1	16	180	205
halobacterium	171	188	194	176	160	3	20	111	99	109	25	6	13	125	231	170
HALO_EXTRA	107	140	211	63	174	33	6	44	27	7	2	3	1	96	194	119
Methanobacteriales																
m.thermoautotrophicum	201	237	3	38	200	177	85	140	187	189	51	6	7	154	230	252
m.stadtmanae	4	157	136	1	196	198	211	156	182	56	77	68	7	87	137	96
Methanococcales																
m.vannielii	90	224	252	15	87	222	14	56	125	201	98	24	1	28	173	256
m.aeolicus	12	198	161	3	11	10	73	20	2	242	18	7	1	6	17	57
m.maripaludis	140	223	251	12	55	210	29	94	115	204	96	13	3	21	201	254
Methanomicrobiales																
m.hungatei	182	63	7	1	256	24	80	45	68	9	88	31	6	38	112	143
m.boonei	186	139	141	92	256	63	87	180	109	82	187	21	243	221	249	33
m.labreanum	61	43	35	26	182	2	32	4	23	142	38	6	210	5	179	66
m.marisnigri	212	187	189	160	254	6	12	57	149	11	112	5	128	16	231	150
Methanopyrales																
m.kandleri	230	214	183	121	169	186	147	144	237	255	146	69	154	218	256	162
Methanosarcinales																
m.mazei	123	67	17	28	209	1	119	197	81	86	185	4	27	48	210	152
m.burtonii	129	87	39	4	249	20	72	61	170	177	7	3	139	108	220	80
m.thermophila	157	121	1	14	254	32	63	161	143	214	62	11	45	184	236	127
m.acetivorans	176	75	32	25	188	3	62	212	93	112	130	4	31	45	195	151
m.barkeri	137	54	50	40	156	5	91	187	84	168	127	3	28	71	159	184
Thermococcales																
p.abyssi	184	162	10	8	95	218	94	12	171	241	73	26	1	246	87	225
p.horikoshii	219	143	32	35	96	256	91	17	163	216	50	38	40	160	100	251
p.furiosus	193	74	26	10	41	220	99	23	162	245	16	11	1	50	96	222
t.kodakaraensis	97	197	5	8	22	125	123	67	132	221	31	2	1	175	54	145
THERMOC_EXTRA	56	34	22	10	37	75	198	53	163	156	23	2	1	74	164	139
Thermoplasmatales																
t.acidophilum	234	86	1	4	217	164	10	83	68	139	60	12	254	209	181	89
p.torridus	13	192	1	80	244	65	72	123	104	219	122	19	147	173	200	197
t.volcanium	141	160	1	4	153	81	6	55	87	175	7	19	252	131	217	135
Nanoarchaeota																
n.equitans	7	21	8	22	204	161	47	57	15	82	18	1	31	34	41	36

Tableau A : Rangs d'exceptionnalité des palindromes de 4 lettres dans les différentes séquences. Le rang 1 correspond au mot le plus évité dans le génome. Rouge : rangs 1-12 (très évité). Orange : rangs 13-25 (évité). Jaune-vert : rangs 232-244 (sur-représenté). Vert : rangs 245-256 (très sur-représenté). Il existe 256 mots de 4 lettres.

SEQUENCES	gcagc	gctgc	cgacg	cgtcg	ggtcc	ggacc	cctgg	ccagg	gcggc	gccgc	gggcc	ggccc	tttaa	ttaaa	ctgag	ctcag	gactc	gagtc
Thaumarchaeota																		
Nitrosopumilales																		
n.maritimus	979	983	575	537	951	966	105	126	785	467	683	698	44	29	48	28	868	823
Crenarchaeaota																		
Desulfurococcales																		
a.pernix	642	443	830	866	791	877	507	548	808	707	102	175	606	514	836	864	291	177
h.butylicus	540	610	887	926	405	247	1020	1018	831	800	560	689	266	218	861	682	183	190
i.hospitalis	34	83	738	735	4	5	107	217	446	294	**1**	3	203	238	797	642	24	15
s.marinus	55	40	825	626	619	813	131	204	362	497	927	867	344	448	922	906	557	540
Sulfolobales																		
m.sedula	16	25	988	993	12	9	164	96	2	**1**	49	22	6	3	662	600	342	306
s.acidocaldarius	27	49	468	213	5	26	55	67	52	85	372	461	82	54	641	764	305	105
s.solfataricus	128	88	353	394	206	143	712	803	112	46	263	408	41	38	179	329	202	138
s.tokodaii	52	40	531	468	500	369	946	959	260	232	494	571	57	43	145	165	179	100
SULFO_EXTRA	46	58	678	950	168	180	973	918	2	10	388	329	42	96	320	466	93	23
CONJUG	904	705	446	818	340	95	1021	1020	170	116	77	673	38	156	601	433	12	6
CRYPTIC	1004	717	883	970	576	529	1023	1024	**1**	2	786	231	171	87	147	491	597	418
LIPOTHRIX	**1**	2	714	993	21	6	322	18	3	4	642	640	162	104	261	119	164	24
FUSELLO	11	28	448	533	460	772	445	918	3	4	741	49	112	183	671	559	502	73
RUDI	11	28	424	596	89	123	722	438	10	5	1004	598	349	287	205	274	249	393
TENGCHONG	446	827	845	709	449	831	529	336	674	950	507	145	48	127	198	806	39	617
Thermoproteales																		
p.aerophilum	2	**1**	852	860	249	223	520	644	10	14	304	233	717	657	949	929	140	131
p.arsenaticum	**1**	2	919	913	214	208	371	494	47	32	607	572	657	648	948	968	108	85
p.calidifontis	2	**1**	963	964	181	229	645	594	53	50	262	289	547	643	998	995	134	79
p.islandicum	3	4	910	927	57	87	865	777	19	11	555	755	679	740	981	984	71	58
t.pendens	**1**	2	980	965	81	172	195	193	321	392	365	541	491	277	967	950	29	26
Euryarchaeota																		
Archaeoglobales																		
a.fulgidus	18	31	463	430	13	9	2	**1**	10	15	87	79	237	179	73	89	28	20
Halobacteriales																		
h.walsbyi	312	177	504	701	92	167	10	28	285	462	161	145	346	314	370	335	169	231
halobacterium	48	45	872	849	7	5	589	761	40	31	15	14	418	463	66	47	21	17
n.pharaonis	16	33	462	508	312	209	**1**	2	309	288	129	105	442	399	15	18	11	13
HALO_EXTRA	3	6	968	910	13	12	**1**	2	279	166	436	334	530	284	33	22	10	5
Methanobacteriales																		
m.stadtmanae	919	978	857	853	519	617	182	125	466	546	291	310	958	927	536	419	274	240
m.thermoautotrophicum	981	964	928	978	397	349	622	550	771	799	834	806	783	710	252	260	405	454
Methanococcales																		
m.aeolicus	63	24	371	504	46	45	9	17	120	123	543	567	882	800	488	524	133	140
m.maripaludis	248	308	186	216	302	287	13	17	423	421	465	562	973	992	151	326	305	282
m.vannielii	779	599	101	127	19	17	25	29	315	178	626	488	1004	1012	185	177	293	458
Methanomicrobiales																		
m.boonei	315	358	142	110	6	5	8	7	551	673	183	162	682	641	17	10	53	57
m.hungatei	982	980	776	668	50	15	102	126	573	483	168	236	978	973	**1**	2	13	14
m.labreanum	60	86	381	361	17	15	26	8	219	430	690	485	453	349	54	49	253	73
m.marisnigri	181	214	253	243	2	**1**	131	97	679	571	22	29	653	703	14	18	21	30
Methanopyrales																		
m.kandleri	225	322	944	960	31	16	229	196	792	896	48	30	771	641	318	314	25	14
Methanosarcinales																		
m.acetivorans	890	874	436	363	3	4	**1**	2	533	627	406	314	254	266	43	57	333	246
m.barkeri	968	965	656	589	5	4	6	3	710	902	508	384	132	215	183	204	694	546
m.burtonii	693	674	12	14	6	7	**1**	2	382	348	116	157	494	641	39	29	262	321
m.mazei	868	849	146	230	4	3	2	**1**	854	766	357	221	174	358	103	91	641	502
m.thermophila	787	732	274	215	22	7	288	350	267	193	26	30	625	472	783	761	240	98
Thermococcales																		
p.abyssi	36	5	943	930	2	**1**	491	507	189	181	326	396	484	311	698	677	7	11
p.furiosus	18	21	921	948	8	15	885	895	244	393	200	94	146	134	598	599	5	11
p.horikoshii	93	139	990	904	32	85	857	944	219	225	413	210	318	427	574	716	23	37
t.kodakaraensis	3	4	813	873	2	**1**	8	9	40	61	75	34	487	396	84	119	12	17
THERMOC_EXTRA	14	6	691	463	745	717	180	574	350	20	611	654	347	283	940	422	65	9
Thermoplasmatales																		
p.torridus	938	954	806	832	104	99	324	246	828	819	612	602	558	452	465	406	689	575
t.acidophilum	191	164	522	586	3	**1**	6	12	313	219	54	34	493	467	223	186	211	84
t.volcanium	649	407	118	231	6	11	12	14	299	611	875	701	80	45	709	720	37	40
Nanoarchaeota																		
n.equitans	381	357	373	628	83	75	536	820	749	276	812	784	74	45	542	637	60	158

Tableau B : Rangs d'exceptionnalité de certains pseudo-palindromes de 5 lettres dans les diverses séquences. Le rang 1 correspond au mot le plus évité. Rouge : rangs 1-51, mots très évités. Orange : rangs 52-102, mots évités. Vert-jaune : rangs 923-973, mots sur-représentés. Vert : rangs 974-1024, mots très sur-représentés. Il existe 1024 mots de 5 lettres.

SEQUENCES	tatata	taatta	cccggg	gggccc	cacgtg	ctcgag	gagctc	agatct	tcgcga	ctgcag	gacgtc	gatatc	cgatcg	cctagg	ggatcc	ttgcaa	caattg
Thaumarchaeota																	
Nitrosopumilales																	
n.maritimus	966	1	2770	1723	785	2328	1700	1068	1009	335	737	21	2288	1257	117	941	4089
Crenarchaeaota																	
Desulfurococcales																	
a.pernix	2001	2482	2461	1727	499	2079	51	131	2173	3082	1246	143	75	272	2803	4052	802
h.butylicus	109	2864	11	2578	88	19	1	627	265	507	2054	124	216	16	1650	3962	357
i.hospitalis	1504	2947	863	3879	305	2879	3476	2473	3526	1500	936	991	1254	30	3184	924	310
s.marinus	1	28	892	2597	1579	1108	3075	2033	1591	818	1601	494	375	240	30	3371	504
Sulfolobales																	
m.sedula	1538	1688	1015	1378	3	898	3673	228	2280	1507	1633	269	14	59	2292	3227	82
s.acidocaldarius	53	3	2307	3837	754	338	470	1055	72	976	455	15	690	547	834	1924	1696
s.solfataricus	2139	4	1960	1089	4078	1264	2044	916	2057	1904	1901	3	1644	70	22	2599	29
s.tokodaii	395	3	1164	2132	3576	1232	536	3061	2704	3379	709	4	1155	38	57	1118	239
SULFO_EXTRA	3242	9	596	14	3466	1804	960	2961	3186	3656	1577	132	3888	1410	4083	2265	173
CONJUG	1283	136	1274	48	539	103	2326	3278	1263	2863	2017	302	4002	1115	4084	2014	1867
LIPOTHRIX	2140	610	2254	2493	2967	1616	762	398	3046	1505	2656	1942	1099	1285	3012	1883	36
FUSELLO	933	1405	149	118	4086	3553	77	2622	625	2771	1667	35	2484	240	3656	848	1429
RUDI	3977	2	1114	728	1204	1538	119	1254	1467	68	133	1163	3626	3993	2481	1885	19
TENGCHONG	2602	179	150	2412	1073	1329	979	2250	2793	2105	1371	107	3475	188	388	3862	751
Thermoproteales																	
p.aerophilum	4043	3548	764	2037	68	717	260	2818	2393	484	7	1161	230	1080	396	3894	2423
p.arsenaticum	3885	3791	908	1923	8	12	295	814	1897	58	11	2488	79	23	2215	3597	3236
p.calidifontis	3755	3843	213	3020	6	726	179	1816	1001	1248	4	784	205	7	1937	2782	1874
p.islandicum	3646	3412	2798	2743	1246	863	611	2081	1176	1477	28	503	1768	3530	2118	3841	2437
t.pendens	1726	3241	38	3496	70	1	3167	584	1602	1433	3360	2081	1033	4	3615	3466	2307
Euryarchaeota																	
Archaeoglobales																	
a.fulgidus	3216	1665	24	883	730	1	2	644	808	94	364	740	638	719	1002	3771	516
Halobacteriales																	
h.walsbyi	3204	3251	2304	2187	2797	2	791	607	3183	1617	571	3096	123	1220	3227	98	4092
halobacterium	2754	3194	449	125	25	1	560	104	8	161	6	3595	5	1588	97	1476	2552
n.pharaonis	3812	3587	56	12	61	1	23	182	10	1031	3	4016	107	1909	1697	683	2191
HALO_EXTRA	3489	2433	9	18	11	1	43	50	64	854	4	4075	7	1365	147	492	1645
Methanobacteriales																	
m.stadtmanae	219	1	1897	1786	757	3389	1052	190	2198	3153	609	4	1781	386	57	1882	4032
m.thermoautotrophicum	1516	3686	1314	409	3327	2370	3943	16	772	461	2466	1189	265	2360	2412	3760	3681
Methanococcales																	
m.aeolicus	93	35	2069	3507	2182	733	4096	1635	2777	4	2836	34	3782	994	1209	96	4050
m.maripaludis	3979	4	863	1209	2244	77	2252	3329	1173	5	39	247	1923	787	494	2729	4090
m.vannielii	3678	3	1311	1283	2884	139	1727	1626	1039	5	86	10	2804	1849	627	1265	4094
Methanomicrobiales																	
m.boonei	2863	2305	36	379	41	1	2014	8	5	4	555	3909	2	1291	536	4096	3570
m.hungatei	3304	351	1	39	3493	769	460	38	1798	47	925	3834	92	1323	2	4031	3680
m.labreanum	1621	2723	2491	150	2108	28	48	1	507	1386	1713	4039	57	1719	1780	4093	3262
m.marisnigri	4090	2666	1	67	271	2	650	10	8	173	748	4086	3	1550	4	4094	2229
Methanopyrales																	
m.kandleri	3781	1660	432	1801	22	28	3275	21	435	1970	3787	4041	12	794	3970	1987	1001
Methanosarcinales																	
m.acetivorans	3986	2690	37	238	247	12	419	250	1525	14	2995	4035	160	757	3	4022	3973
m.barkeri	4089	1737	134	302	181	37	252	462	1343	3	732	4053	99	395	4	3906	4051
m.burtonii	3835	1469	3668	1376	2176	146	15	1	2695	2	183	2612	3	2918	33	73	4060
m.mazei	3765	2603	77	1348	365	14	84	125	3730	3	2793	3751	235	1759	4	3878	4060
m.thermophila	3211	2309	4000	94	98	16	180	1	1799	678	829	2358	36	2325	1350	3162	2180
Thermococcales																	
p.abyssi	1677	3299	3731	2664	2543	793	4012	675	3101	1549	372	297	216	89	779	3241	1370
p.furiosus	1708	312	3620	379	1684	42	519	1134	3405	2403	296	230	219	1956	75	3629	1098
p.horikoshii	896	2379	1960	827	2481	471	2412	211	3025	1235	550	84	33	2463	129	3670	542
t.kodakaraensis	1724	1616	845	19	2702	1	3	28	2106	842	199	257	36	359	1222	4049	1387
THERMOC_EXTRA	3971	3981	3267	21	94	1	2782	250	3514	1221	154	938	2570	287	230	3872	748
Thermoplasmatales																	
p.torridus	513	713	3989	1353	3418	1598	244	7	1720	3566	2730	69	38	1590	65	2408	4075
t.acidophilum	1056	3842	3566	1138	310	3017	3201	1	1137	2440	3725	2456	2	1450	6	4058	3151
t.volcanium	227	2465	3145	352	2122	1381	3775	8	2214	459	1507	551	1	568	2	3818	2698
Nanoarchaeota																	
n.equitans	1153	15	2250	1399	1700	399	6	2847	2367	4080	3644	449	1235	2276	69	3272	972

Tableau C : Rangs d'exceptionnalité de certains palindromes de 6 lettres dans les diverses séquences. Le rang 1 correspond au mot le plus évité. Rouge : rangs 1-204, mots très évités. Orange : rang 205-409, mots évités. Vert-jaune : rangs 3688-3892, mots sur-représentés. Vert : rangs 3893-4096, mots très sur-représentés. Il existe 4096 mots de 5 lettres.

SEQUENCES	cagc	cttc	gaag	ggcg	cgcc	gatg	catc	aata	tatt	atag	ctat	ctta	taag	gatt	aatc	tccg	cgga	gagc	gctc	tttg	caaa	ttgg	ccaa	cctg	cagg	gaat	attc
Thaumarchaeota																											
Nitrosopumilales																											
n.maritimus	228	237	229	236	245	252	256	141	161	29	19	3	2	96	53	46	48	153	170	186	166	78	67	231	235	144	123
Crenarchaeota																											
Desulfurococcales																											
a.pernix	248	252	253	251	254	161	165	172	164	179	184	16	19	142	134	7	10	2	1	50	61	5	9	137	149	74	71
h.butylicus	254	250	249	238	241	176	181	204	200	182	173	15	14	95	89	19	21	2	1	54	49	4	3	16	12	88	82
i.hospitalis	255	239	236	226	215	142	149	227	230	221	231	4	3	41	31	45	44	114	120	29	40	80	82	2	1	16	17
s.marinus	228	234	248	240	225	149	154	255	256	90	95	67	52	24	26	29	41	36	27	5	3	20	33	192	175	11	21
Sulfolobales																											
m.sedula	199	231	202	123	136	84	75	193	207	216	208	5	10	35	47	25	30	33	39	7	3	9	12	184	187	4	6
s.acidocaldarius	244	251	252	241	235	228	230	224	220	127	134	8	15	5	2	73	51	14	11	68	60	24	38	158	151	4	3
s.solfataricus	255	245	250	191	180	252	249	237	240	131	135	8	7	9	15	35	47	4	6	16	13	21	25	2	1	3	5
s.tokodaii	255	253	254	210	212	249	252	223	217	114	154	3	4	25	22	34	33	5	6	47	41	18	16	2	1	8	7
SULFO_EXTRA	185	254	256	252	253	245	232	164	144	98	83	25	8	6	11	29	16	14	15	88	90	10	13	32	35	1	5
CRYPTIC	147	245	252	256	254	115	61	70	89	67	82	69	27	59	178	20	40	51	28	204	253	14	25	238	197	57	104
CONJUG	240	252	256	243	220	245	237	226	217	178	153	65	27	2	17	35	49	18	34	186	130	6	10	11	24	1	14
FUSELLO	121	200	250	218	223	253	231	133	116	209	128	136	30	98	72	38	25	17	52	247	220	49	44	9	7	36	157
LIPOTHRIX	172	229	214	248	255	204	228	210	209	46	83	9	30	34	3	90	25	18	33	21	13	17	27	78	71	87	12
RUDI	45	214	238	245	241	212	184	8	4	17	16	19	3	6	2	10	7	105	44	50	49	112	27	181	182	1	5
TENGCHONG	239	250	254	204	249	247	253	215	135	52	61	9	18	60	25	22	81	2	12	73	138	48	13	32	8	3	16
Thermoproteales																											
p.aerophilum	244	235	231	254	253	188	177	184	186	172	178	19	23	72	75	14	15	7	6	11	12	3	5	4	8	47	51
p.arsenaticum	241	248	246	255	256	207	211	209	195	105	112	50	48	97	81	10	9	7	8	11	12	6	5	3	4	55	44
p.calidifontis	239	246	245	249	250	186	195	160	162	128	132	32	36	141	148	8	11	15	13	40	39	7	4	3	2	75	66
p.islandicum	225	247	248	255	256	204	201	154	146	128	129	22	18	117	103	9	8	10	6	14	15	17	19	31	35	32	29
t.pendens	154	250	248	242	243	149	150	164	172	165	169	19	11	98	100	6	9	23	24	28	13	5	2	219	225	53	49
Euryarchaeota																											
Archaeoglobales																											
a.fulgidus	246	248	247	229	230	256	255	95	99	240	235	20	24	236	241	66	88	3	4	28	22	14	16	8	9	40	41
Halobacteriales																											
h.walsbyi	234	224	230	251	252	255	256	227	229	81	79	11	12	188	170	23	18	8	7	52	49	70	73	140	132	44	40
halobacterium	241	183	181	255	256	248	245	223	226	121	131	103	113	102	117	1	2	32	24	119	120	30	34	62	79	44	40
n.pharaonis	241	218	220	254	253	255	256	154	156	145	128	29	28	239	235	5	7	15	13	100	95	39	44	16	14	45	43
HALO_EXTRA	231	230	233	254	248	255	256	207	212	156	114	32	29	172	188	4	5	20	13	122	117	23	22	47	43	14	21
Methanobacteriales																											
m.stadtmanae	225	254	253	229	230	249	250	256	255	102	99	2	3	28	31	61	64	174	165	6	5	22	18	200	214	19	20
m.thermoautotrophicum	251	222	215	225	216	245	244	173	178	256	255	23	20	57	83	11	10	18	19	15	12	2	1	32	33	120	105
Methanococcales																											
m.aeolicus	133	255	256	230	229	250	248	249	247	152	142	5	4	98	74	96	100	244	246	22	23	80	92	150	168	9	8
m.maripaludis	111	255	256	238	232	252	250	217	220	38	53	2	1	22	28	58	50	187	163	6	7	104	135	229	230	23	20
m.vannielii	132	251	255	247	242	253	244	221	219	85	70	2	3	23	36	47	48	162	182	4	5	131	135	204	196	19	18
Methanomicrobiales																											
m.boonei	160	179	184	100	104	120	128	237	233	224	219	26	22	7	8	3	4	198	201	50	52	6	5	206	216	91	96
m.hungatei	136	241	239	193	196	245	248	240	235	233	244	4	3	21	13	17	25	149	159	43	47	14	11	252	255	50	56
m.labreanum	199	246	245	253	249	223	215	250	252	176	172	29	28	13	12	3	1	79	87	205	203	118	146	140	139	44	55
m.marisnigri	131	240	237	253	252	221	228	226	229	106	90	35	40	32	30	4	3	203	201	158	161	8	7	223	217	31	27
Methanopyrales																											
m.kandleri	217	239	233	221	236	185	182	229	223	184	178	31	41	22	24	1	2	143	149	93	89	103	110	152	151	4	3
Methanosarcinales																											
m.acetivorans	70	246	245	184	191	230	231	213	215	253	254	14	17	41	36	105	111	134	147	22	23	1	2	256	255	13	15
m.barkeri	39	246	243	155	164	238	242	211	214	253	248	8	7	55	52	85	69	151	135	22	18	2	1	255	256	15	13
m.burtonii	150	187	207	225	222	253	255	254	256	245	246	21	19	1	2	65	64	132	130	38	41	11	9	219	212	14	16
m.mazei	37	243	248	184	191	228	224	234	222	253	254	14	11	61	45	89	120	165	170	18	20	3	2	256	255	16	15
m.thermophila	141	252	251	243	245	223	224	240	230	238	248	7	15	20	18	56	42	174	158	181	151	5	4	242	244	137	112
Thermococcales																											
p.abyssi	202	244	243	133	131	249	252	164	166	251	253	9	6	72	60	5	3	53	57	70	67	63	50	161	165	59	54
p.furiosus	236	256	255	158	133	251	252	189	165	219	212	2	3	35	22	7	17	61	67	65	59	24	28	108	121	27	38
p.horikoshii	214	242	245	80	79	246	247	185	202	243	248	11	7	2	5	8	4	22	21	73	72	52	58	180	165	41	57
t.kodakaraensis	185	254	253	202	209	252	251	198	190	248	249	11	9	154	143	6	7	136	158	65	60	14	13	186	161	40	44
THERMOC_EXTRA	213	219	232	207	182	225	253	206	175	220	176	49	68	209	204	52	47	107	69	127	51	54	129	25	63	8	46
Thermoplasmatales																											
p.torridus	69	61	58	199	194	232	229	251	252	255	256	3	2	21	22	29	36	174	166	6	5	20	25	241	233	100	135
t.acidophilum	219	228	237	184	153	90	109	251	250	256	255	16	14	3	2	75	88	19	22	20	24	8	7	199	188	30	31
t.volcanium	228	244	236	222	218	248	249	253	254	255	256	10	9	3	2	93	70	18	17	33	14	11	12	140	138	13	16
Nanoarchaeota																											
n.equitans	156	245	251	155	130	248	252	250	254	247	253	4	3	60	69	29	32	51	44	206	216	88	91	135	127	46	53

Tableau D : Rangs d'exceptionnalité de divers motifs de 4 lettres dans les séquences. Le rang 1 correspond au mot le plus évité dans le génome. Rouge : rangs 1-12 (très évité). Orange : rangs 13-25 (évité). Jaune-vert : rangs 232-244 (sur-représenté). Vert : rangs 245-256 (très sur-représenté). Il existe 256 mots de 4 lettres.

SEQUENCES	aattc	ggcgc	ataat	taata	gagct	agctc	ccaga	tctgg	cattg	caatg		tagtac	gtcgtc
Thaumarchaeota													
Nitrosopumilales													
n.maritimus	871	515	1001	2	222	190	747	691	89	119		3185	1595
Crenarchaeota													
Desulfurococcales													
a.pernix	651	288	1007	360	855	925	843	832	578	504		46	737
h.butylicus	676	311	936	94	640	703	612	646	576	653		90	1170
i.hospitalis	717	56	988	277	1001	999	523	742	247	551		7	575
s.marinus	844	687	931	2	158	261	647	606	460	380		1431	2726
Sulfolobales													
m.sedula	813	478	716	196	685	777	622	566	783	800		43	410
s.acidocaldarius	982	635	593	83	666	744	901	678	152	167		532	55
s.solfataricus	1011	469	815	421	653	786	749	763	26	33		34	155
s.tokodaii	1013	650	506	86	712	646	829	868	9	14		616	2847
SULFO_EXTRA	914	300	978	512	962	956	507	158	47	5		582	1126
CRYPTIC	565	103	590	432	793	938	132	63	5	72			
CONJUG	910	271	1009	451	929	983	68	45	618	47		200	880
FUSELLO	977	496	968	349	1011	731	765	283	63	25		837	735
LIPOTHRIX	580	208	341	604	845	992	996	507	252	9		2131	2442
RUDI	950	628	899	655	696	526	311	348	3	2		1374	3071
TENGCHONG	948	640	963	809	876	832	373	274	37	15		3679	370
Thermoproteales													
p.aerophilum	989	40	297	561	1024	1023	66	72	830	813		32	12
p.arsenaticum	842	70	541	534	1019	1017	22	25	822	901		7	96
p.calidifontis	808	80	646	473	1018	1020	19	16	855	844		37	84
p.islandicum	980	13	535	231	1018	1009	407	329	757	728		40	46
t.pendens	535	85	932	354	985	997	176	225	824	855		29	1386
Euryarchaeota													
Archaeoglobales													
a.fulgidus	693	101	917	446	867	940	1017	1018	879	695		54	179
Halobacteriales													
h.walsbyi	414	48	322	798	375	675	718	710	93	117		494	2083
halobacterium	466	1	545	483	946	945	656	607	679	695		116	3817
n.pharaonis	647	4	514	609	910	821	896	909	933	914		140	3592
HALO_EXTRA	557	4	638	471	695	414	963	978	780	703		29	3644
Methanobacteriales													
m.stadtmanae	1024	699	863	42	408	728	655	710	331	280		461	2228
m.thermoautotrophicum	906	671	959	926	948	993	339	396	1017	1012		99	1631
Methanococcales													
m.aeolicus	1018	447	969	49	1023	1024	899	826	573	499		2813	491
m.maripaludis	910	330	1012	18	788	705	887	948	593	542		409	1472
m.vannielii	976	473	1013	34	661	536	854	857	248	270		1352	1671
Methanomicrobiales													
m.boonei	719	791	880	449	933	903	982	968	927	926		30	823
m.hungatei	242	402	837	154	679	602	644	686	578	521		1538	2146
m.labreanum	271	668	966	614	682	504	988	950	900	961		233	1302
m.marisnigri	431	68	400	455	481	583	963	900	515	440		26	1082
Methanopyrales													
m.kandleri	569	24	918	659	991	1011	599	588	477	395		13	2061
Methanosarcinales													
m.acetivorans	899	626	764	94	823	885	1023	1024	936	967		45	1198
m.barkeri	731	609	790	199	843	650	1024	1023	982	874		89	878
m.burtonii	803	336	964	259	990	968	1000	1013	1009	1018		34	422
m.mazei	813	500	820	124	852	817	1022	1024	949	836		28	1587
m.thermophila	102	261	229	793	989	983	658	903	661	456		178	1131
Thermococcales													
p.abyssi	947	354	969	645	780	834	792	740	928	981		13	2225
p.furiosus	992	385	898	56	886	826	762	665	694	664		64	306
p.horikoshii	964	390	986	613	949	922	618	565	791	874		6	1108
t.kodakaraensis	700	43	868	342	990	960	969	981	891	887		23	1229
THERMOC_EXTRA	1012	340	811	370	993	1019	597	571	912	439		2	252
Thermoplasmatales													
p.torridus	979	364	736	1008	972	970	735	645	1023	1024		13	1490
t.acidophilum	664	134	943	894	859	843	963	953	1014	1017		75	2644
t.volcanium	508	339	937	305	664	631	1003	981	1022	1012		807	1475
Nanoarchaeota													
n.equitans	1024	564	558	674	308	325	529	566	267	77		2062	1072

Tableau E : Rangs d'exceptionnalité de divers motifs de 5 et 6 lettres dans les séquences. Le rang 1 correspond au mot le plus évité dans le génome. Rouge : rangs 1-51 pour les mots de 5 lettres, rangs 1-204 pour les mots de 6 lettres (mots très évités). Orange : rangs 52-102 pour les mots de 5 lettres, rangs 205-409 pour les mots de 6 lettres (évité). Vert-jaune : rangs 923-973 pour les mots de 5 lettres, rangs 3688-3892 pour les mots de 6 lettres (sur-représenté). Vert : rangs 974-1024 pour les mots de 5 lettres, rangs 3893-4096 pour les mots de 6 lettres (très sur-représenté). Il existe 1024 mots de 5 lettres et 4096 mots de 6 lettres.

	L4 (16)		L5 (64)		L6 (64)		L7 (256)		L8 (256)	
SEQUENCES	**Nb_NEG**	**p-value**	**Nb_NEG**	**p-value**	**Nb_NEG**	**p-value**	**Nb_POS**	**p-value**	**Nb_POS**	**p-value**
Thaumarchaeota										
Nitrosopumilales										
n.maritimus	9	4,02E-01	35	2,66E-01	46	3,09E-04	133	2,87E-01	164	4,00E-06
Crenarchaeota										
Desulfurococcales										
a.pernix	10	2,27E-01	22	9,96E-01	34	3,54E-01	139	9,46E-02	132	3,31E-01
h.butylicus	11	1,05E-01	23	9,92E-01	35	2,66E-01	131	3,77E-01	124	7,13E-01
i.hospitalis	5	9,62E-01	42	8,43E-03	36	1,91E-01	154	6,94E-04	126	6,23E-01
s.marinus	9	4,02E-01	37	1,30E-01	31	6,46E-01	147	1,03E-02	137	1,44E-01
Sulfolobales										
m.sedula	9	4,02E-01	41	1,64E-02	39	5,17E-02	114	9,65E-01	152	1,62E-03
s.acidocaldarius	12	3,84E-02	46	3,09E-04	46	3,09E-04	114	9,65E-01	137	1,44E-01
s.solfataricus	11	1,05E-01	43	4,07E-03	45	7,81E-04	122	7,92E-01	135	2,08E-01
s.tokodaii	12	3,84E-02	42	8,43E-03	43	4,07E-03	124	7,13E-01	147	1,03E-02
SULFO_EXTRA	10	2,27E-01	41	1,64E-02	29	8,09E-01	126	6,23E-01		
CRYPTIC	10	2,27E-01	28	8,70E-01						
CONJUG	14	2,09E-03	35	2,66E-01	32	5,50E-01				
FUSELLO	13	1,06E-02	46	3,09E-04	34	3,54E-01				
LIPOTHRIX	14	2,09E-03	43	4,07E-03	29	8,09E-01				
RUDI	8	5,98E-01	35	2,66E-01	30	7,34E-01				
TENGCHONG	12	3,84E-02	32	5,50E-01	37	1,30E-01				
Thermoproteales										
p.aerophilum	7	7,73E-01	40	3,00E-02	40	3,00E-02	166	1,18E-06	124	7,13E-01
p.arsenaticum	8	5,98E-01	43	4,07E-03	41	1,64E-02	170	8,34E-08	147	1,03E-02
p.calidifontis	7	7,73E-01	40	3,00E-02	42	8,43E-03	157	1,74E-04	136	1,74E-01
p.islandicum	9	4,02E-01	35	2,66E-01	35	2,66E-01	171	4,13E-08	136	1,74E-01
t.pendens	8	5,98E-01	31	6,46E-01	30	7,34E-01	175	2,08E-09	107	9,96E-01
Euryarchaeota										
Archaeoglobales										
a.fulgidus	14	2,09E-03	45	7,81E-04	51	9,40E-07	117	9,25E-01	133	2,87E-01
Halobacteriales										
h.walsbyi	14	2,09E-03	48	3,87E-05	40	3,00E-02	128	5,25E-01	127	5,74E-01
halobacterium	9	4,02E-01	40	3,00E-02	44	1,84E-03	171	4,13E-08	124	7,13E-01
n.pharaonis	12	3,84E-02	51	9,40E-07	47	1,13E-04	157	1,74E-04	130	4,26E-01
HALO_EXTRA	12	3,84E-02	42	8,43E-03	53	5,03E-08	150	3,54E-03	149	5,13E-03
Methanobacteriales										
m.stadtmanae	10	2,27E-01	37	1,30E-01	51	9,40E-07	124	7,13E-01	163	7,19E-06
m.thermoautotrophicum	6	8,95E-01	20	9,99E-01	36	1,91E-01	119	8,83E-01	144	2,62E-02
Methanococcales										
m.aeolicus	13	1,06E-02	43	4,07E-03	41	1,64E-02	137	1,44E-01	176	9,43E-10
m.maripaludis	10	2,27E-01	44	1,84E-03	43	4,07E-03	135	2,08E-01	155	4,45E-04
m.vannielii	10	2,27E-01	41	1,64E-02	46	3,09E-04	148	7,31E-03	128	5,25E-01
Methanomicrobiales										
m.boonei	7	7,73E-01	47	1,13E-04	45	7,81E-04	142	4,57E-02	132	3,31E-01
m.hungatei	13	1,06E-02	39	5,17E-02	42	8,43E-03	120	8,56E-01	138	1,17E-01
m.labreanum	12	3,84E-02	42	8,43E-03	48	3,87E-05	145	1,95E-02	150	3,54E-03
m.marisnigri	8	5,98E-01	42	8,43E-03	50	3,53E-06	164	4,00E-06	124	7,13E-01
Methanopyrales										
m.kandleri	2	1,00E+00	32	5,50E-01	37	1,30E-01	160	3,80E-05	141	5,90E-02
Methanosarcinales										
m.acetivorans	10	2,27E-01	43	4,07E-03	47	1,13E-04	104	9,99E-01	141	5,90E-02
m.barkeri	9	4,02E-01	34	3,54E-01	46	3,09E-04	106	9,98E-01	164	4,00E-06
m.burtonii	11	1,05E-01	38	8,43E-02	42	8,43E-03	111	9,86E-01	155	4,45E-04
m.mazei	11	1,05E-01	42	8,43E-03	46	3,09E-04	115	9,54E-01	137	1,44E-01
m.thermophila	9	4,02E-01	34	3,54E-01	36	1,91E-01	159	6,42E-05	136	1,74E-01
Thermococcales										
p.abyssi	9	4,02E-01	38	8,43E-02	39	5,17E-02	102	1,00E+00	121	8,26E-01
p.furiosus	11	1,05E-01	33	4,50E-01	39	5,17E-02	90	1,00E+00	123	7,54E-01
p.horikoshii	9	4,02E-01	35	2,66E-01	36	1,91E-01	108	9,95E-01	122	7,92E-01
t.kodakaraensis	11	1,05E-01	40	3,00E-02	53	5,03E-08	141	5,90E-02	134	2,46E-01
THERMOC_EXTRA	11	1,05E-01	36	1,91E-01	38	8,43E-02				
Thermoplasmatales										
p.torridus	7	7,73E-01	18	1,00E+00	39	5,17E-02	118	9,05E-01	147	1,03E-02
t.acidophilum	9	4,02E-01	39	5,17E-02	36	1,91E-01	120	8,56E-01	121	8,26E-01
t.volcanium	8	5,98E-01	41	1,64E-02	39	5,17E-02	106	9,98E-01	146	1,43E-02
Nanoarchaeota										
n.equitans	14	2,09E-03	41	1,64E-02	31	6,46E-01	141	5,90E-02		

Tableau F : Exceptionnalité globale des palindromes et pseudo-palindromes de 5 à 8 lettres dans les différentes séquences. L4 à L8 : longueur des mots considérés et, entre parenthèses, nombre de palindromes de cette longueur. Nb_POS : nombre de palindromes ayant un score positif. Nb_NEG : nombre de palindromes ayant un score négatif. p-value : p-value associée à la répartition des scores selon leur signe (voir section 2.4). Rouge : très évité (p-value ≤ 0,01 dans le cas où c'est Nb_NEG qui a été considéré). Orange : évité 0,01 ≤ p-value ≤ 0,05. Vert-jaune : sur-représenté (0,01 ≤ p-value ≤ 0,05 dans le cas où c'est N_POS qui a été considéré). Vert : très sur-représenté (p-value ≤ 0,01).

	L4 (16)		L5 (64)		L6 (64)		L7 (256)		L8 (256)	
SEQUENCES	Nb_POS	p-value	Nb_NEG	p-value	Nb_NEG	p-value	Nb_POS	p-value	Nb_POS	p-value
Thaumarchaeota										
Nitrosopumilales										
n.maritimus	16	1,53E-05	38	8,43E-02	29	8,09E-001	173	9,59E-09	158	1,07E-04
Crenarchaeota										
Desulfurococcales										
a.pernix	14	2,09E-03	49	1,22E-05	37	1,30E-001	205	0,00E+00	155	4,45E-04
h.butylicus	16	1,53E-05	59	4,50E-13	38	8,43E-002	185	3,39E-13	144	2,62E-02
i.hospitalis	15	2,59E-04	48	3,87E-05	33	4,50E-001	194	0,00E+00	136	1,74E-01
s.marinus	16	1,53E-05	34	3,54E-01	39	5,17E-002	140	7,52E-02	135	2,08E-01
Sulfolobales										
m.sedula	8	5,98E-01	45	7,81E-04	30	7,34E-001	177	4,20E-10	147	1,03E-02
s.acidocaldarius	12	3,84E-02	46	3,09E-04	29	8,09E-001	180	3,34E-11	147	1,03E-02
s.solfataricus	12	3,84E-02	52	2,28E-07	25	9,70E-001	163	7,19E-06	137	1,44E-01
s.tokodaii	14	2,09E-03	47	1,13E-04	36	1,91E-001	192	2,22E-16	137	1,44E-01
SULFO_EXTRA	16	1,53E-05	24	9,84E-01	35	2,66E-001	193	1,11E-16		
CRYPTIC	13	1,06E-02	25	9,70E-01						
CONJUG	10	2,27E-01	35	2,66E-01	33	4,50E-001				
FUSELLO	11	1,05E-01	30	7,34E-01	34	3,54E-001				
LIPOTHRIX	14	2,09E-03	29	8,09E-01	26	9,48E-001				
RUDI	12	3,84E-02	28	8,70E-01	34	3,54E-001				
TENGCHONG	16	1,53E-05	29	8,09E-01	26	9,48E-001				
Thermoproteales										
p.aerophilum	15	2,59E-04	57	3,82E-11	40	3,00E-02	170	8,34E-08	145	1,95E-02
p.arsenaticum	16	1,53E-05	57	3,82E-11	39	5,17E-002	193	1,11E-16	153	1,07E-03
p.calidifontis	16	1,53E-05	54	9,98E-09	50	3,53E-06	194	0,00E+00	152	1,62E-03
p.islandicum	12	3,84E-02	48	3,87E-05	39	5,17E-002	177	4,20E-10	126	6,23E-01
t.pendens	15	2,59E-04	58	4,51E-12	49	1,22E-05	189	6,11E-15	135	2,08E-01
Euryarchaeota										
Archaeoglobales										
a.fulgidus	14	2,09E-03	36	1,91E-01	40	3,00E-02	162	1,27E-05	158	1,07E-04
Halobacteriales										
h.walsbyi	13	1,06E-02	19	1,00E+00	44	1,84E-03	213	0,00E+00	136	1,74E-01
halobacterium	16	1,53E-05	29	8,09E-01	41	1,64E-02	189	6,11E-15	108	9,95E-01
n.pharaonis	14	2,09E-03	47	1,13E-04	44	1,84E-03	180	3,34E-11	112	9,81E-01
HALO_EXTRA	15	2,59E-04	33	4,50E-01	43	4,07E-03	197	0,00E+00	116	9,41E-01
Methanobacteriales										
m.stadtmanae	16	1,53E-05	21	9,98E-01	28	8,70E-01	196	0,00E+00	163	7,19E-06
m.thermoautotrophicum	12	3,84E-02	53	5,03E-08	30	7,34E-01	164	4,00E-06	145	1,95E-02
Methanococcales										
m.aeolicus	16	1,53E-05	33	4,50E-01	35	2,66E-01	167	6,22E-07	162	1,27E-05
m.maripaludis	10	2,27E-01	38	8,43E-02	45	7,81E-04	131	3,77E-01	153	1,07E-03
m.vannielii	10	2,27E-01	43	4,07E-03	45	7,81E-04	144	2,62E-02	143	3,49E-02
Methanomicrobiales										
m.boonei	13	1,06E-02	50	3,53E-06	30	7,34E-01	184	8,82E-13	120	8,56E-01
m.hungatei	11	1,05E-01	39	5,17E-02	34	3,54E-01	171	4,13E-08	133	2,87E-01
m.labreanum	13	1,06E-02	44	1,84E-03	26	9,48E-01	177	4,20E-10	123	7,54E-01
m.marisnigri	14	2,09E-03	46	3,09E-04	29	8,09E-01	221	0,00E+00	114	9,65E-01
Methanopyrales										
m.kandleri	12	3,84E-02	48	3,87E-05	46	3,09E-04	183	2,25E-12	134	2,46E-01
Methanosarcinales										
m.acetivorans	10	2,27E-01	38	8,43E-02	34	3,54E-01	208	0,00E+00	141	5,90E-02
m.barkeri	10	2,27E-01	36	1,91E-01	34	3,54E-01	210	0,00E+00	168	3,24E-07
m.burtonii	15	2,59E-04	40	3,00E-02	41	1,64E-02	190	2,11E-15	135	2,08E-01
m.mazei	10	2,27E-01	33	4,50E-01	39	5,17E-02	219	0,00E+00	152	1,62E-03
m.thermophila	14	2,09E-03	44	1,84E-03	39	5,17E-02	194	0,00E+00	146	1,43E-02
Thermococcales										
p.abyssi	12	3,84E-02	58	4,51E-12	30	7,34E-01	182	5,63E-12	135	2,08E-01
p.furiosus	10	2,27E-01	48	3,87E-05	35	2,66E-01	157	1,74E-04	142	4,57E-02
p.horikoshii	10	2,27E-01	51	9,40E-07	30	7,34E-01	170	8,34E-08	136	1,74E-01
t.kodakaraensis	12	3,84E-02	56	2,78E-10	46	3,09E-04	191	7,77E-16	127	5,74E-01
THERMOC_EXTRA	16	1,53E-05	39	5,17E-02	37	1,30E-01				
Thermoplasmatales										
p.torridus	12	3,84E-02	31	6,46E-01	39	5,17E-02	138	1,17E-01	160	3,80E-05
t.acidophilum	10	2,27E-01	47	1,13E-04	34	3,54E-01	185	3,39E-13	152	1,62E-03
t.volcanium	8	5,98E-01	39	5,17E-02	27	9,16E-01	178	1,84E-10	138	1,17E-01
Nanoarchaeota										
n.equitans	10	2,27E-01	27	9,16E-01	35	2,66E-01	165	2,19E-06		

Tableau G : Exceptionnalité globale des mots symétriques de 5 à 8 lettres dans les différentes séquences. L4 à L8 : longueur des mots considérés et nombre de mots symétriques de cette longueur. Nb_POS : nombre de mots symétriques ayant un score positif. Nb_NEG : nombre de mots symétriques ayant un score négatif. p-value : p-value associée à la répartition des scores selon leur signe (voir section 2.4). Rouge : très évité (p-value ≤ 0,01 dans le cas où c'est Nb_NEG qui a été considéré). Orange : évité 0,01 ≤ p-value ≤ 0,05. Vert-jaune : sur-représenté (0,01 ≤ p-value ≤ 0,05 dans le cas où c'est N_POS qui a été considéré). Vert : très sur-représenté (p-value ≤ 0,01).

Discussion

1

Résumé des résultats obtenus

1.1 Virus SIRV2

1.1.1 Interaction hôte-virus

Mes travaux de thèse ont contribué à une meilleure caractérisation du cycle infectieux du virus SIRV2. Ils ont été effectués en collaboration avec l'équipe « Archaea Group » de l'université d'Uppsala en Suède, et avec la plate-forme de microscopie électronique de l'Institut Pasteur. J'ai été encadrée par Olivier Tenaillon (Ecologie et Evolution des Microorganismes, UMR Inserm – Paris 7) et David Prangishvili (BMGE, Institut Pasteur). Les travaux ont démontré que SIRV2 est un virus virulent, contrairement à ce qui avait été suggéré précédemment. En effet, SIRV2 était qualifié de virus chronique dans la littérature (état porteur, Prangishvili et al., 1999). La lyse des cellules survient 10 à 12 heures après l'infection. Les virions sont assemblés avant la phase de lyse, dans le cytoplasme, où ils forment des agrégats denses et structurés ressemblant à des fagots d'allumettes.

Lors de la lyse, les virions quittent la cellule par des ouvertures dans la paroi cellulaire. Ces ouvertures sont induites par l'infection car elles ne sont pas observées dans les cellules non-infectées. En outre, les ouvertures ont un contour net, consistant en un anneau polygonal. La section de l'anneau a une forme de C. L'anneau est parfois observé détaché de la cellule, et semble donc correspondre à une structure spécifique relativement rigide. Les ouvertures doivent être nombreuses puisqu'on en compte jusqu'à quatre par section de cellule. Hormis ces ouvertures, la paroi cellulaire est intacte et stable, même longtemps après la lyse. En revanche, le matériel intracellulaire est dégradé. Les cellules lysées s'apparentent donc à des coquilles vides. La paroi cellulaire seule provoque toujours de la diffusion lumineuse, si bien que la lyse ne peut pas être détectée par une simple mesure de densité optique des cultures.

Les ouvertures sont créées par un mécanisme qui repose sur la formation de structures spéciales au niveau de la membrane. Les structures sont saillantes et pointues, de forme pyramidale. Nous les avons baptisées « VIPs » (« Viral Induced Protrusions »). Elles sont visibles sur les cellules n'ayant pas encore lysé et n'ont pas été observées sur les cellules non-infectées. *A priori*, chaque VIP, en éclatant, donne lieu à une ouverture. La paroi cellulaire au niveau des VIPs est modifiée. En effet, il n'y a pas de S-layer. De plus, on observe un léger épaississement de la membrane, mais cette dernière observation requiert confirmation car la fixation et la coloration utilisées en microscopie électronique lors de ce projet ne sont pas adaptées aux études structurales détaillées.

Le virus étant lytique, le chromosome de l'hôte est dégradé. Avec la multiplicité d'infection élevée employée dans mon travail, l'apparition de cellules sans chromosome se produit dès 30 minutes après l'infection. La proportion de cellules sans chromosome augmente progressivement au cours du temps, si bien que pour au moins 40% d'entre elles, la dégradation a lieu avant la phase de lyse. Dans des phases assez précoces du cycle infectieux, la majorité des cellules présentent une

quantité accrue d'ADN total, ce qui doit refléter, au moins en partie, la réplication du génome viral. Ceci a été détecté par cytométrie en flux. C'est *a priori* la première fois que de l'ADN viral intracellulaire peut être détecté, et sa quantité estimée, par cette méthode.

1.1.2 Données sur la particule virale de SIRV2

J'ai également travaillé sur les propriétés physiques et biochimiques des virions de SIRV2.

D'une part, j'ai évalué la stabilité des particules virales dans différentes conditions proches des caractéristiques de l'habitat (température de 80°C, pH ≈ 3,0). La stabilité a été estimée en titrant les particules infectieuses au cours du temps. Dans ces conditions, la demi-vie des virions est de l'ordre de 1 jours. Avec un pH neutre et toutes choses égales par ailleurs (température, solution), la demi-vie est de l'ordre de 3 jours. A température ambiante, dans le tampon habituel de stockage des virions, aucun déclin significatif n'a été observé après plus de 200 jours. Pour fixer les idées, le phage le moins robuste (R17, Leviviridae), dans l'étude de De Paepe and Taddei, 2006, a une demi-vie de 1,33 jours à 37°C ; le plus stable (PRD1, Tectiviridae) a une demi-vie de 18,73 jours ; et des valeurs usuelles sont de 2 à 10 jours. Les résultats obtenus ici indiquent donc que les particules virales sont plutôt bien adaptées à leur environnement très chaud. Elles sont en conséquence ultra-stables en conditions mésophiles.

D'autre part, j'ai évalué la stabilité des particules dans différents mélange d'eau et d'éthanol ou d'eau et de DMSO (Steinmetz et al., 2008). Ces conditions, artificielles, peuvent intervenir lors de l'utilisation de particules virales pour des applications en nanotechnologies (étapes de bioconjugaison, de minéralisation, etc.). La stabilité a été suivie sur une semaine, en combinant titration des particules infectieuses, détermination du spectre d'absorption dans la gamme 200-300 nm, et observation de l'état des virions par microscopie électronique en transmission. A cette échelle de temps, aucune dégradation des particules n'a été observée. Par contre, dans l'éthanol 50%, les particules perdent progressivement leur caractère infectieux. Peut-être y a-t-il un léger effet fixateur, qui empêche les modifications conformationnelles souvent requises pour l'infection d'un l'hôte (fixation à un récepteur, etc.).

Enfin, nous avons identifié trois protéines structurales de SIRV2 (Vestergaard et al., 2008b), portant le total de protéines structurales connues à quatre. La particule est donc composée d'une protéine majoritaire, qui était déjà connue. Associée à l'ADN (nucléoprotéine), elle constitue le corps hélicoïdal du virion. Une seconde protéine, correspondant au plus long gène du génome de SIRV2, est minoritaire et a été localisée aux extrémités du virion, dans les fibres (Steinmetz et al., 2008). Les deux dernières protéines sont très minoritaires et leur rôle/position dans la structure de la particule n'est pas connu.

1.2 Enrichissement d'échantillons environnementaux de sources chaudes

Afin d'isoler de nouveaux virus, des étapes d'enrichissement à partir d'échantillons de sources chaudes collectés dans la péninsule russe de Kamchatka ont été réalisées. Nous avons testé des conditions favorables à la croissance d'espèces des genres *Acidianus* et *Sulfolobus* (ordre des Sulfolobales). Il s'agissait d'un milieu riche, acide (pH ≈ 3), avec ou sans soufre, et de conditions aérobies, avec un température de 80°C.

Un seul morphotype de virus, filamenteux, a été observé par microscopie électronique en transmission. Ce morphotype n'est pas nouveau mais nous avons décidé de poursuivre la caractérisation. Le virus, nommé AFV9, appartient à la famille β-Lipothrixviridae, et infecte un hôte du genre *Acidianus*. Les structures terminales du virus sont uniques : elles consistent en une partie aplatie et légèrement dentelée, perpendiculaire à la direction du filament. Il y est accroché un filament très fin en forme de crochet. Il semble que les structures terminales des lipothrixvirus varient souvent d'un virus à l'autre. Par ailleurs, huit protéines structurales différentes ont été identifiées. Parmi elles, six ont des homologues chez tous les autres β-lipothrixvirus séquencés, et parmi ces six, deux protéines sont majoritaires et homologues aux protéines structurales principales identifiées chez SIFV et AFV3. Les deux protéines restantes ont des homologues chez tous les β-lipothrixvirus à l'exception de SIFV et AFV7.

Ce travail a fait l'objet d'une publication (Bize et al., 2008), qui comprend des données supplémentaires sur AFV9 et sur les enrichissements en conditions anaérobies, produites par nos collaborateurs.

Les virus de Sulfolobales sont les mieux caractérisés. Même s'il est utile de caractériser la diversité à l'intérieur d'une famille de virus, ou que de nouveaux morphotypes peuvent encore être découverts, il pourrait aussi être porteur désormais de tenter l'isolement de virus infectant des hôtes d'ordres moins caractérisés (Desulfurococcales, divers ordres de méthanogènes, etc., voir Partie I, 3.3 et Tableau 4).

1.3 Motifs courts exceptionnels dans les génomes d'archées et de leurs éléments extra-chromosomiques

Un moyen de détecter les motifs fonctionnels non-codant dans les génomes est l'analyse statistique des séquences, permettant de détecter les motifs exceptionnels. Pour les archées, encore peu de motifs fonctionnels non-codants sont connus. Par ailleurs, un nombre conséquent de génomes d'archées, de leurs virus et de leurs plasmides sont aujourd'hui disponibles dans les bases de données. En collaboration avec Sophie Schbath, de l'Unité Mathématique, Informatique & Génome de l'INRA Jouy-en-Josas, nous avons recherché des motifs exceptionnels dans une quarantaine de génomes d'archées, une quarantaine de génomes de virus ou provirus et une quarantaine de séquences de plasmides. Les mots de taille 4 à 8 ont été étudiés, avec le logiciel RMES (Hoebeke and Schbath, http://genome.jouy.inra.fr/ssb/rmes).

Au préalable, la composition des séquences de virus et plasmides en mots courts a été analysée, afin de concaténer les séquences homogènes. Ceci dans le but de produire une analyse statistique plus riche. Pour constituer les groupes, deux méthodes différentes on été utilisées, le clustering hiérachique et l'analyse en composante principale. Outre le but méthodologique, cette étape a montré que les groupes formés reflètent clairement l'ordre auquel appartiennent les hôtes des virus ou plasmides (Thermococcales, Halobacteriales, Sulfolobales). Quelques exceptions existent mais ne sont pas nombreuses. Pour les éléments extra-chromosomiques de Sulfolobales, dont les séquences sont plus nombreuses, des sous-groupes ont même pu être établis, qui reflétaient non plus la phylogénie des hôtes, mais plutôt les familles virales (Rudiviridae, Lipothrixviridae, etc.) et le type de plasmide (conjugatif, cryptique). Cette dernière observation est encourageante quant à l'utilisation des signatures pour identifier la nature ou l'origine de séquences non-identifiées. Des exemples d'applications possibles sont la métagénomique, ou l'étude des ORFans, les cadres de lecture orphelins.

Concernant les motifs exceptionnels, deux familles de motifs ont d'abord été examinées, les motifs palindromiques et pseudo-palindromiques d'une part, et les motifs symétriques d'autre part.

Les motifs palindromiques sont évités dans les génomes de bactéries, par exemple chez *Escherichia coli*, et l'évitement s'explique en partie par l'existence de systèmes de restriction-modification ou de méthylation et dans certains cas par l'existence de systèmes de réparation.

Nous avons montré que les motifs palindromiques (ou pseudo-palindromiques) de taille 4 à 6 sont évités dans les génomes d'archées et de leurs éléments extra-chromosomiques. En nous basant sur les quelques systèmes de restriction-modification/méthylation déjà identifiés chez les archées, nous avons estimé que cet évitement peut sûrement, dans certains cas, être lié à la présence de systèmes de restriction-modification dans les génomes. Par ailleurs, nous avons remarqué que les motifs évités sont souvent les mêmes dans les génomes correspondant à un même ordre (ex : Halobacteriales), ceci y compris pour les génomes d'éléments extra-chromosomiques. Enfin, si les palindromes sont globalement évités dans les génomes d'archées, ce phénomène semble moins marqué que pour les bactéries. En effet, dans chaque séquence, seuls quelques palindromes sont exceptionnellement évités. En outre, les mots pseudo-palindromiques de taille 7 seraient même plutôt, en moyenne, sur-représentés.

Concernant les motifs symétriques, il nous semble que ceux-ci n'ont encore jamais été spécifiquement étudiés, ni chez les archées, ni pour d'autres types d'organismes. Ici, nous avons observé une sur-représentation globale des mots symétriques de taille 4 et 7, et une sous-représentation globale des mots symétriques de taille 5. Dans une bien moindre mesure, les mots symétriques de taille 6 étaient évités et ceux de taille 8 sur-représentés. Pour l'instant, nous ne sommes pas en mesure de proposer d'hypothèses précises pour expliquer ce phénomène.

Pour finir, nous nous sommes intéressés aux mots très exceptionnels, sur-représentés, de taille 5 à 7, retrouvés chez toutes les espèces d'un même genre. Ces mots pourraient être de bons candidats pour une implication dans des fonctions générales, de type maintenance du génome, régulation de la transcription, etc.

Un certain nombre de candidats ont été identifiés, et des analyses supplémentaires ont été

réalisées pour les motifs les plus intéressants. Sophie Schbath a par exemple regardé s'ils avaient une position préférentielle par rapport aux gènes (méthode FADO décrite dans Gusto and Schbath, 2005) mais les résultats n'ont rien montré de significatif. La distribution des motifs dans les génomes a également été examinée. Ceci a montré que AATCC correspond aux séquences répétées de CRISPRs chez *Sulfolobus solfataricus* et *Sulfolobus tokodaii*. Par ailleurs, le couple de mots CCTGGG/CCCAGG, extrêmement sur-représenté chez *Pyrococcus*, semble évité en deux endroits du génome, en particulier chez *Pyrococcus horikoshii*.

En conclusion, les travaux ont bien démontré l'intérêt d'étudier les motifs exceptionnels chez les archées car de nombreux candidats ont pu être identifiés. Mais l'interprétation est difficile en raison des connaissances encore limitées sur les archées. Il est pourtant nécessaire de s'atteler à la tâche. Des allers-retours entre approches expérimentales et *in silico* pourraient être fructueuses. Des projets plus ciblés sur certains mots ou certains organismes pourraient également être utiles, permettant des analyses plus poussées et l'utilisation de critères supplémentaires (ex : différence de composition en motifs entre brin direct et indirect, soit « leading » et « lagging »).

2

Des pistes pour poursuivre sur le virus SIRV2

2.1 Mise au point d'outils de génétique

La mise au point d'outils de génétiques pour le virus SIRV2 serait très utile, tant pour le projet bionanotechnologies que pour l'étude de l'interaction hôte virus. Cela contribuerait de plus à conférer définitivement à SIRV2 le statut de virus modèle.

Plusieurs approches sont possibles. Tout d'abord, la protéine majoritaire de ARV1, un autre rudivirus, a déjà été exprimée dans une souche d'*Escherichia coli* puis purifiée (protéine recombinante). Elle possède la propriété de s'auto-assembler en l'absence de toute source d'énergie (Vesteergard et al., 2008b). Cette propriété peut être très utile pour le projet nanotechnologies, et l'ADN recombinant, directement, pourrait être modifié en fonction des besoins. Ceci nécessite de réévaluer les propriétés du tube de protéines obtenu, car celles-ci ne seront plus glycosylées comme dans leur hôte naturel, ce qui peut affecter les propriétés physiques et chimiques (stabilité, rigidité, fonctions exposées).

Pour l'étude de l'interaction hôte-virus, il peut être envisagé de cloner des gènes viraux d'intérêt sur un plasmide pour transformer les cellules hôtes. L'avantage de cette méthode est qu'elle permettrait d'étudier l'effet de gènes isolés, et qu'elle devrait être réalisable avec les outils qui existent aujourd'hui.

Pour les deux projets, il pourrait être utile de pouvoir modifier et travailler sur l'ensemble du génome viral. Jusqu'à présent, l'électroporation du génome viral dans les cellules hôtes n'a pas permis d'obtenir de particules infectieuses, ni pour SIRV2, ni pour aucun virus de crenarchées de génome linéaire (Prangishvili, communication personnelle). Cette méthode fonctionne en revanche pour le fusellovirus SSV1, qui a un génome circulaire. Cependant, il est possible que l'échec vienne non pas de la linéarité du génome mais du fait que les protocoles testés aient été adaptés à un virus chronique. Le fait que SIRV2 soit lytique modifie la donne, et des expériences d'électroporation pourraient être de nouveau tentées.

2.2 Mise au point d'un nouveau protocole de purification du virus

L'une des exigences pour l'utilisation d'un virus dans le domaine des nanotechnologies est l'obtention de préparations très pures. Le protocole actuel, basé sur une précipitation au PEG (polyéthylène-glycol) a permis d'obtenir un préparation suffisamment propre pour les premiers essais. Néanmoins, à plus long terme, ce protocole risque d'être insuffisant, car quelques vésicules et débris cellulaires sont encore présents, en très faible concentration, dans les préparations de virus. Des techniques de chromatographie pourraient par exemple être testées (Smrekar et al., 2008, bactériophage T4).

2.3 Etude du cycle infectieux

Le travail de thèse a permis de mesurer de façon robuste la période de latence, et, un peu plus approximativement, la taille de la progénie virale (« burst size »). Outre ces paramètres du cycle infectieux, un premier aperçu des mécanismes de lyse a été obtenu. Enfin, il a été observé que les virions sont assemblés dans le cytoplasme, avant la lyse. Mais beaucoup reste à découvrir sur le cycle infectieux de SIRV2.

Pour achever une description générale de l'interaction hôte-virus, il pourrait être utile de déterminer la sensibilité des cellules au virus en fonction de la phase de croissance (exponentielle, stationnaire, etc.).

Pour mieux caractériser le cycle infectieux, il serait intéressant d'observer des sections de cellules infectées à divers points de temps du cycle infectieux, et pas seulement dans les phases tardives. Cela devrait permettre d'estimer plus précisément la période d'éclipse, et peut-être d'identifier différentes étapes dans le cycle infectieux ou de mieux caractériser le mode de production des virions. Par exemple, cela pourrait permettre de voir comment se forment les agrégats de virions, si la taille des agrégats augmente progressivement au cours du cycle infectieux, ou si les agrégats apparaissent brusquement. Afin d'éviter les artefacts liés à la fixation et la coloration des cellules, les points de temps identifiés comme intéressants pourraient être analysés en ayant recours à la cryo-microscopie électronique.

Une autre question est de comprendre ce qui conduit à la formation d'agrégats de virions. S'agit-il d'un phénomène purement physique, lié par exemple à des contraintes d'espace ? Ou bien est-ce lié à des mécanismes biologiques, comme l'existence de foyers de réplication, et le couplage de la réplication à l'assemblage des nucléoprotéines constituant le virion ? *A priori*, toutes ces hypothèses sont envisageables.

Par ailleurs, le nombre de tubules observé dans certaines cellules était très supérieur au nombre de génomes viraux intracellulaires estimé par cytométrie en flux, et très supérieur également à la taille de la progénie virale estimée par titration sur tapis cellulaire (courbe de lyse). Plusieurs explications sont possibles. Certains des tubules observés pourraient ne pas contenir d'ADN, et refléter, par exemple, l'existence d'un phénomène de stockage des protéines structurales. Une fraction des tubules ne correspondrait alors pas à des virions infectieux. Une seconde possibilité est au contraire que la taille de progénie ait été sous-estimée, si les virions sont encore partiellement agrégés après être sortis des cellules (cf titration sur tapis cellulaire). Des analyses par PCR quantitative, à partir de culots de cellules infectées et de surnageants de cultures, pourraient aider à distinguer entre ces différentes possibilités.

Concernant les mécanismes d'entrée dans la cellule, pratiquement rien n'est connu chez les virus de crenarchées hyperthermophiles, et cette question pourrait être étudiée sur SIRV2, notamment en utilisant les cellules résistantes qui émergent rapidement, comme j'ai pu l'observer.

Enfin, une question à creuser est celle de la dégradation des chromosomes de l'hôte. Nous avons observé qu'elle survient bien avant la phase de lyse dans une grande partie des cellules, mais il faudrait vérifier que ce phénomène n'a pas été induit par l'utilisation d'une multiplicité d'infection élevée, destinée à synchroniser l'infection. Dans cette optique, il serait très utile de pouvoir synchroniser l'infection par un autre moyen, permettant de travailler avec différentes m.o.i. tout en

conservant une synchronisation satisfaisante. Si le phénomène de dégradation était confirmé, il serait alors intéressant de déterminer quand et comment elle survient, et quel est son rôle au cours du cycle infectieux (simple neutralisation de l'hôte et/ou source de nucléotides).

Après avoir obtenu des informations plus détaillées sur le cycle infectieux, des études transcriptomiques avec des points de temps bien ciblés pourraient être refaites. En effet, les premières études de transcription réalisées sur SIRV2 (Kessler et al., 2004) se basaient sur une période de latence de six heures. Les points de temps dans cet article sont 30 minutes, 1, 2, 3 et 4 heures après l'infection. Or nous avons observé ici une période de latence de plus de neuf heures. Il pourrait donc être intéressant d'analyser des points de temps plus tardifs. *A contrario*, les phases précoces du cycle infectieux pourraient faire l'objet d'un échantillonnage plus fréquent, avec des points de temps plus rapprochés. Nous avons observé que les quantités d'ADN viral intracellulaire commencent à augmenter très peu de temps après l'infection (voir Partie II), et la moitié de la quantité maximale est atteinte environ deux heures après l'infection. Ceci semble indiquer que nombre de mécanismes permettant au virus de se multiplier doivent être mis en place lors de phases très précoces du cycle infectieux.

2.4 Aspects moléculaires

Bien sûr, une question importante est la compréhension des mécanismes moléculaires impliqués dans la formation des VIPs.

D'après une publication récente (Fulton et al., 2009), des structures pyramidales sont également observées sur les cellules de *Sulfolobus solfataricus* infectées par le virus icosaédrique lytique STIV. Les images obtenues par microscopie électronique en transmission (Young, données non-publiées) permettent de penser qu'il s'agit d'un phénomène très similaire à celui observé pour SIRV2. C'est une donnée importante, car cela prouve que les VIPs ne sont pas un phénomène marginal.

Seules quatre protéines de SIRV2 ont des homologues chez STIV. Il s'agit de SIRV2gp49, SIRV2gp21 et SIRV2gp37-SIRV2gp29 (ces deux dernières protéines étant elles-mêmes homologues entre elles) et elles pourraient donc intervenir, directement ou indirectement, dans la formation des VIPs. On ne peut toutefois pas exclure que les protéines impliquées respectivement chez SIRV2 et STIV ne présentent aucune homologie.

La structure de deux homologues de SIRV2gp37-SIRV2gp29 est décrite dans la littérature. Il s'agit de AFV3gp16 (Keller et al., 2007) et de B116 chez STIV (Larson et al., 2007). Dans les deux publications, il est suggéré que cette protéine peut se lier à l'ADN. Elle se lierait au grand sillon de l'ADN double-brin (Larson et al., 2007). Ces hypothèses sont étayées par des études biochimiques. SIRV2gp37 a des homologues chez tous les rudivirus séquencés, mais les scores sont bas pour ARV1gp13 (score de 45) et SRV_ORF108 (score de 51). SIRV2gp29 est moins conservée.

SIRV2gp21 est homologue à une protéine de SRV (*Stygiolobus* rod-shaped virus), codée par l'ORF58, qui est un facteur de transcription putatif de la super-famille RHH (« ribbon-helix-helix »)

(Vestergaard et al., 2008b). Des protéines homologues de SIRV2gp21 sont présentes chez tous les rudivirus séquencés, avec le score le plus faible pour ARV1_ORF59b (score de 56).

Concernant SIRV2gp49, peu d'information est disponible. Elle a des homologues chez les autres rudivirus à l'exception d'ARV1. En outre, le score est faible pour SRV_ORF92 (score de 41). Elle aurait un domaine putatif conservé appelé PHA01757.

Il se peut aussi que les principales protéines impliquées dans la formation des VIPs soient codées dans le génome de l'hôte, et que le virus ne fasse que détourner et manipuler une fonction existante dans la cellule. Un piste intéressante est celle des trois protéines Cdv (Lindas et al., 2008). On sait depuis peu que ces protéines interviennent lors de la division cellulaire, et de la ségrégation des génomes chez les crenarchées. En effet, les crenarchées ne possèdent pas de protéines FtsZ, protéines qui sont présentes chez les euryarchées et les bactéries et forment l'anneau de constriction. Les crénarchées ont donc un autre système, qui comprend les protéines Cdv identifiées récemment chez *Sulfolobus acidocaldarius* et présentes chez toutes les crenarchées.

Chez *Sulfolobus acidocaldarius*, les protéines Cdv forment des structures allongées, qui s'étendent sur toute la largeur de la cellule, et sont situées entre les deux génomes ségrégeant. Au moment de la constriction, les structures deviennent plus petite. CdvA et CdvB colocalisent. D'autre part, l'opéron cdv est fortement induit chez *Sulfolobus solfataricus*, lorsque les cellules sont infectées par STIV (Otmann et al., 2008). Enfin, les protéines CdvB et CdvC présentent une homologie avec le complexe eucaryote ESCRT-III. Ce complexe fait partie de la machinerie endosomale d'adressage des protéines, qui est requise pour des événements de bourgeonnement (« budding ») intervenant dans plusieurs processus biologiques importants. Les protéines d'ESCRT-III pourraient de plus former une structure en anneau (discussion de Lindas et al., 2008 ; Hanson et al., 2008). Enfin, le complexe ESCRT est impliqué dans le processus de bourgeonnement de HIV type 1 (discussion de Lindas et al., 2008 ; Morita and Sundquist, 2004).

Lindas et al. après une discussion plus détaillée, émettent l'hypothèse que les deux systèmes (Cdv, ESCRT-III) pourraient avoir une origine évolutive commune, et que les protéines Cdv pourraient être recrutées lors du bourgeonnement de STIV.

Si le processus de lyse induit par STIV chez *Sulfolobus solfataricus* est effectivement identique à celui induit par SIRV2, il ne s'agit pas à proprement parler d'un bourgeonnement. Il y a néanmoins apparition de structures saillantes impliquant des modifications de la paroi cellulaire. Une hypothèse plausible serait donc que les protéines Cdv sont recrutées pour former les VIPs lors de l'infection par SIRV2 ou STIV. Le phénomène de localisation observé lors de la division cellulaire serait dans ce cas perdu. Des anticorps contre les protéines Cdv ayant été utilisés par Lindas et al., cette hypothèse doit pouvoir être testée. Si elle était juste, ce mécanisme de lyse serait spécifique des crenarchées puisque les euryarchées n'ont pas l'opéron cdv.

Une autre manière d'obtenir de l'information sur les VIPs serait d'isoler ces structures pour analyser leur composition en lipides et protéines. Une possibilité serait de travailler sur les cellules lysées, donc vides, ce qui permettrait d'éliminer une grande partie du matériel intracellulaire. Les anneaux délimitant les ouvertures présents sur les cellules lysées semblent se détacher facilement (voir Partie II). Un processus qui permette de les séparer sans trop les dégrader puis des les

concentrer est donc peut-être envisageable. Ceci permettrait une analyse restreinte à un anneau polygonal au lieu de la pyramide entière. Mais cela pourrait déjà donner des indications.

En conclusion, les pistes sont nombreuses, il s'agit d'identifier les plus prometteuses et les plus simples à mettre en œuvre.

3

Interaction hôte-virus chez les crenarchées

3.1 La sortie des virus, un problème mécanique

L'émission de particules virales dans le milieu extra-cellulaire est *grosso modo* un problème mécanique : il faut traverser la paroi cellulaire de l'hôte. Deux possibilités existent, soit la cellule hôte est détruite (lyse), soit elle est préservée.

Pour la lyse, chez les bactéries, trois mécanismes sont décrits. Le plus répandu est le système holine-endolysine, avec des enzymes qui permettent la dégradation du peptidoglycane. Un second mécanisme repose sur une protéine qui bloque la synthèse de la paroi cellulaire (ex : chez ΦX174). Un troisième mécanisme a été décrit récemment pour le phage marin PM2, de la famille Corticoviridae (Krupovic et al., 2007) et ressemble, dans son principe, au système holine-endolysine.

Lorsqu'au contraire, la cellule est préservée, deux stratégies différentes existent. La particule virale peut être assemblée à l'intérieur de la cellule, puis émise par un processus de bourgeonnement qui préserve l'intégrité de la paroi cellulaire. Dans ce cas, le virus est enveloppé, emportant lors du bourgeonnement une fraction de la membrane cellulaire, éventuellement modifiée. C'est le cas des virus de bactérie de la famille Plasmaviridae (L2). Ou bien, la particule virale peut être assemblée au fur et à mesure qu'elle est produite, par le biais d'un dispositif ancré dans la membrane. C'est le cas des virus de bactéries de la famille Inoviridae (ex: M13).

Cependant, l'existence d'une membrane externe n'est pas nécessairement associée à un processus de bourgeonnement. Par exemple, pour le bacteriophage à ARN Φ6, de la famille Cystoviridae, la membrane est ajoutée à l'intérieur du cytoplasme, et le virus est lytique (Mindich et al., 1976).

(Données sur les divers bactériophages obtenues également sur le site http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/index.htm)

3.2 Les observations effectuées sur le virus SIRV2 ne peuvent être directement généralisées

Que peut-on penser de l'interaction hôte-virus chez les virus de crenarchées ? Le mécanisme observé sur SIRV2 au cours de la thèse, et très probablement similaire chez STIV, est totalement nouveau. Du fait que la lyse ne provoque pas nécessairement une chute de densité optique chez les crenarchées, comme le montre le cas de SIRV2, il se peut que les propriétés lytiques d'autres virus soient passées inaperçues.

Mais une particularité distingue SIRV2 et STIV des autres virus d'archées : c'est l'absence d'enveloppe à l'extérieur du virion (pour STIV, une enveloppe lipidique se trouve à l'intérieur de la particule et non autour, Rice et al., 2004). Un processus de bourgeonnement permettant une interaction chronique n'est donc pas exclu pour les autres virus. Il faudrait regarder au cas par cas.

D'une manière générale, il me semble que les critères qui, le plus, suggèrent une interaction

chronique, sont la persistance du virus dans une culture sur de longues durées (sans ajout de nouvelles cellules hôtes) et la possibilité d'isoler des colonies produisant du virus, à condition qu'il y ait eu plusieurs étapes de purification de chaque colonie. Dans le cas de SIRV2, par exemple, l'isolement de colonies produisant le virus n'était pas clairement mentionné ; les hôtes originaux ne produisaient plus de virus après quelques dilutions sérielles et étaient de plus devenus résistants (Prangishvili et al., 1999). Cela s'accordait bien avec l'hypothèse d'un virus lytique. Mais encore une fois, ce type d'observation ne constitue pas une preuve, puisque tous les cas ont déjà été observés. Par exemple, des virus lytiques peuvent parfois coévoluer sur de longues périodes avec un hôte, de même que que des cellules peuvent devenir totalement résistante à un virus chronique et empêcher toute coévolution (voir II.2.2).

Il serait intéressant d'obtenir une réponse claire pour les virus des familles Fuselloviridae et Lipothrixviridae, fréquemment observés dans les sources chaudes. En ce qui concerne les fusellovirus, l'intégrase de SSV1 n'est pas essentielle (Clore and Stedman, 2007). Le virus sans intégrase pourrait donc être utilisé pour étudier la question. Dans l'article de Clore and Stedman, 2007, il est rapporté que le virus, en l'absence d'intégrase, a persisté dans une culture pendant 24 jours, soit environ 70 générations de l'hôte. Ceci penche en faveur d'une relation chronique, mais il faudrait le démontrer clairement. Si le virus est chronique, le morphotype se prêterait *a priori* mieux à une production par bourgeonnement qu'à un assemblage couplé à l'extrusion. Dans le cas des lipothrixvirus, l'ensemble des données disponible ne donne pas d'indications claires. Chez AFV1 comme chez SIFV, des colonies produisant le virus ont pu être isolées (Bettstetter et al., 2003 ; Arnold et al., 2000b). Pour AFV1, la production de virus était toujours observée après plus de deux mois de culture continue, tandis que chez SIFV, le virus a été perdu après plusieurs dilutions sérielles. Si certains de ces lipothrixvirus étaient chroniques, le morphotype, linéaire, pourrait être compatible avec différents modes d'extrusion. Un mécanisme « à la M13 » serait possible. Un mécanisme de bourgeonnement, où le virus emporterait une partie de la membrane cellulaire, est également envisageable. Enfin, le virus étant linéaire et fin (24 nm de largeur, Arnold et al., 2000b), on pourrait également imaginer un processus de translocation de la particule, déjà assemblée, à travers la membrane, mais cela paraît plus compliqué. En outre, peut-être différents virus d'une même famille ont-ils des effets différents sur leur hôte. Par exemple, SIFV infecte un hôte *Sulfolobus* et AFV1 un hôte *Acidianus*, et peut-être la relation hôte-virus est-elle différente dans les deux cas.

3.3 Virulence, environnement et évolution

Une hypothèse étayée par plusieurs études est que la virulence d'un parasite bien adapté à son environnement est positivement corrélée à sa durabilité dans le milieu externe (Walther and Ewald, 2004). En effet, le fait d'être durable rend moins cruciale la rencontre rapide d'un nouvel hôte. Le cas extrême est celui où le parasite, très robuste, a évolué vers une virulence maximale, et provoque la mort de son hôte. Cette hypothèse est connue sous le nom de « malédiction du pharaon » (Gandon, 1998). Ce nom fait référence à l'un des scénarios proposés pour expliquer la mort mystérieuse de Lord Carnarvon et d'autres membres de l'équipe ayant exhumé la momie du

pharaon Toutânkhamon. Selon ce scénario très hypothétique, la mort aurait pu être causée par un agent pathogène présent depuis plus de 3000 ans dans le tombeau.

Les sources chaudes constituant un milieu très agressif sur le plan physico-chimique, on aurait pu s'attendre à ce que le virus SIRV2, à l'inverse, ait évolué vers une virulence faible et ménage son hôte. Or c'est tout le contraire qui est observé, puisque le virus est lytique (et n'est pas lysogène).

La stabilité des particules virales, mesurée en laboratoire pendant ma thèse, pourrait apporter une explication. En effet, les résultats montrent que les particules sont plutôt bien adaptées aux hautes températures et restent infectieuses pendant plusieurs jours en milieu chaud. En outre, comme elles sont très stables en conditions mésophiles, et que les environnements géothermiques sont très hétérogènes, on peut imaginer qu'une fraction des particules virales, piégées dans des niches locales plus clémentes, persistent pendant de très longues périodes dans l'environnement.

Ainsi, malgré des conditions en apparence défavorables, les virions doivent être stables à très stables dans leur environnement, expliquant que la virulence de SIRV2 ait évolué vers des valeurs extrêmes.

Il pourrait être intéressant de déterminer la stabilité et le mode d'interaction d'autres virus de crenarchées hyperthermophiles, afin de voir si la corrélation entre virulence et stabilité est respectée. En outre, cela pourrait donner des indications sur les modes de vie principaux de ces virus en milieu naturel.

Conclusion

Conclusion

Virus d'archées hyperthermophiles et nouveaux mécanismes biologiques

Les virus d'archées hyperthermophiles constituent un groupe de virus uniques, par leur contenu génétique, et par la diversité et la nouveauté de leurs morphotypes. On peut donc penser que se trouve là un gisement pour la découverte de nouveaux mécanismes biologiques. Dans l'optique de découvrir ceux-ci et de les comprendre, l'unicité génétique constitue à la fois un intérêt et une difficulté, puisque très peu de fonctions peuvent être prédites par des analyses *in silico*.

Le travail réalisé au cours de ma thèse a contribué à mettre en évidence un nouveau mécanisme de lyse virale, chez le virus SIRV2 et son hôte hyperthermophile *Sulfolobus islandicus.* Différentes techniques ont été utilisées, principalement la cytométrie en flux, l'hybridation dot blot ADN-ADN et la microscopie électronique en transmission. Le mécanisme observé repose sur la formation de structures pyramidales saillantes en divers points de l'enveloppe cellulaire, (Figure 34C). En fin de cycle infectieux, chaque pyramide s'ouvre, *a priori* par le sommet, donnant naissance à une ouverture bien délimitée permettant la sortie des virions. Au niveau des pyramides, l'enveloppe cellulaire est modifiée, car il n'y a plus de S-layer, et il est possible que des modifications supplémentaires interviennent. L'observation de ce mécanisme de lyse tend à confirmer l'idée que des mécanismes biologiques encore inconnus sont à découvrir chez les virus d'archées hyperthermophiles.

Par ailleurs, le cycle infectieux de SIRV2, incluant le processus de lyse, semble assez complexe. On peut en effet noter la dégradation des chromosomes de l'hôte avec la forte m.o.i employée, qui survient en partie avant la sortie des virions ; la formation d'agrégats de virions bien ordonnés (Figure 34A et 34B) ; et l'induction des structures pyramidales complexes que sont les VIPs, impliquant une modification locale de la membrane cellulaire en de multiples points.

Cette complexité rappelle un peu celle observée pour les virus d'eucaryotes. En effet, les virus d'eucaryotes à ARN modifient souvent les membranes de l'hôte, en général celles des voies de sécrétion, car c'est là qu'ils ancrent leur complexe de réplication et établissent les sites d'assemblage des particules virales (Miller and Krijnse-Locker, 2008). La modification de la structure des membranes des hôtes n'est en revanche pas observée, selon nos connaissances, pour les virus de bactéries. Par ailleurs, un nombre conséquent de virus d'eucaryotes appartenant à des familles différentes, contenant de l'ARN ou de l'ADN, mettent en place de véritables usines à l'intérieur des cellules (« virus factory », ex : Novoa et al., 2005 ; Fontana et al., 2008). Elles permettent une production très efficace des particules virales. Les données accumulées tendent à prouver qu'il s'agit d'un processus actif, impliquant le recrutement de certaines organelles et la synthèse de nouvelles structures. A l'inverse, d'autres organelles et la plupart des protéines de l'hôte sont exclues de l'usine à virus. De très belles images pour le Mimivirus, un virus géant, sont visibles dans Suzan-Monti et al., 2007. Donc certains virus d'eucaryotes s'approprient la cellule hôte et la réorganisent à leur profit. Il semble de plus que cela soit un phénomène général, car observé pour un grand nombre de virus très différents.

Chez les archées et les bactéries, les usines à virus n'ont pas été observées. Mais la problématique n'est pas la même car les cellules hôtes sont de bien plus petite taille et ne sont pas compartimentées en diverses organelles. Les observations de cellules de *Sulfolobus* infectées par SIRV2 réalisées pendant la thèse suggèrent néanmoins que les cellules sont métamorphosées (paragraphe précédent et Figure 34A et 34C). En outre, les agrégats de particules virales occupent un tel espace dans la cellule (Figure 34A et 34B) que celle-ci ne semble plus avoir qu'une seule fonction, la production de virions. C'est donc la cellule entière qui constituerait, chez *Sulfolobus*, une usine à virus.

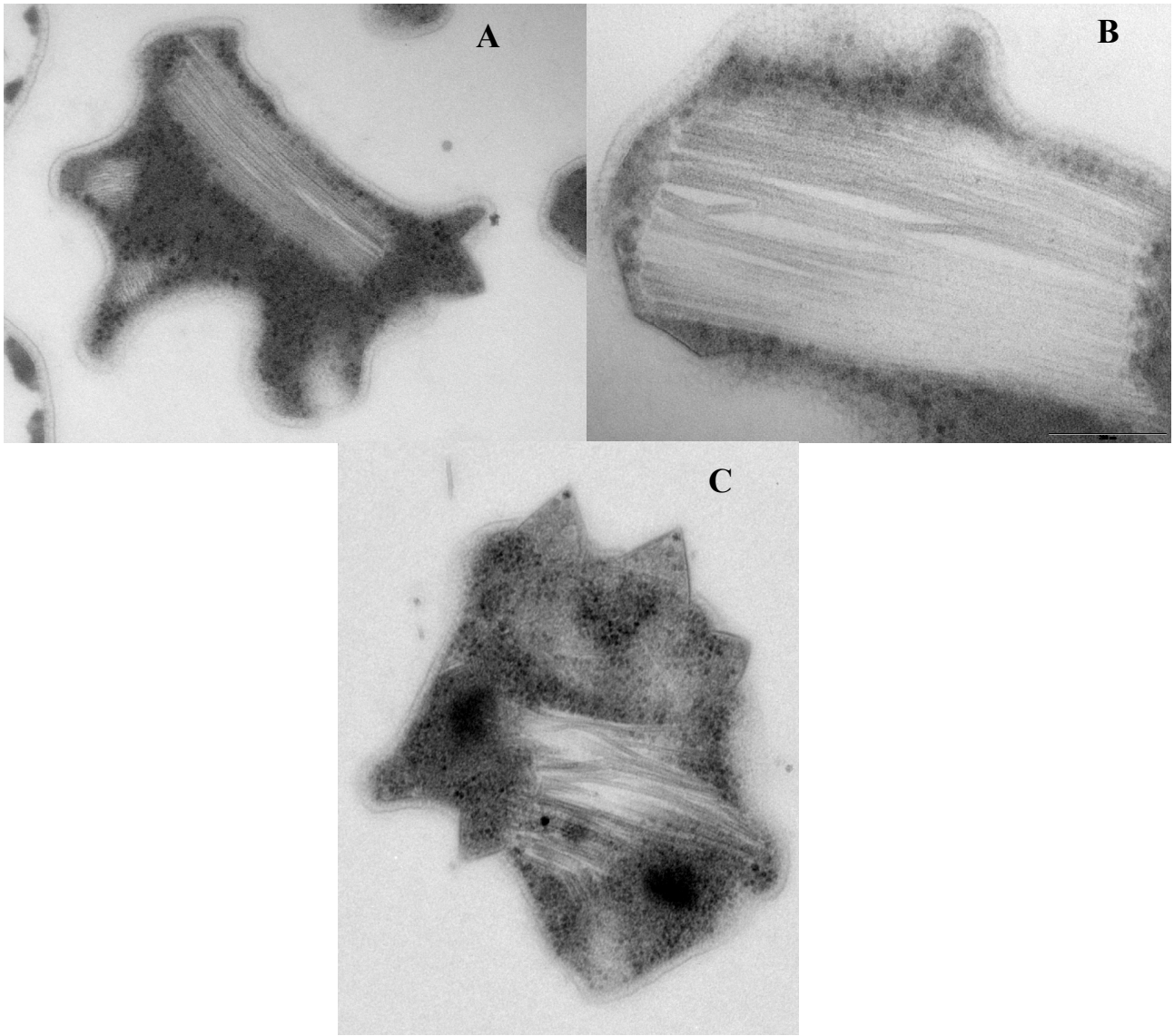


Figure 34 : Sections de cellules de *Sulfolobus islandicus* infectées par SIRV2, en fin de cycle infectieux.

En conclusion, le degré de complexité du cycle infectieux du virus SIRV2 est cohérent avec la notion, abordée en introduction (I.2.2) que les archées présentent un degré de complexité intermédiaire entre celui des bactéries et des eucaryotes.

Bibliographie

Abedon, S.T. (2006) *The Bacteriophages, 2d ed.*. Abedon, S.T. and Calendar, R. (Eds). Oxford University Press, Oxford.

Abella, M., Rodriguez, S., Paytubi, S., Campoy, S., White, M.F., and Barbe, J. (2007) The Sulfolobus solfataricus radA paralogue sso0777 is DNA damage inducible and positively regulated by the Sta1 protein. *Nucleic Acids Res* **35**: 6788-6797.

Ackermann, H.W. (2007) 5500 Phages examined in the electron microscope. *Arch Virol* **152**: 227-243.

Adl, S.M., Simpson, A.G., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., McCourt, R.M., Mendoza, L., Moestrup, O., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A.V., Spiegel, F.W., and Taylor, M.F. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* **52**: 399-451.

Ahn, D.G., Kim, S.I., Rhee, J.K., Kim, K.P., Pan, J.G., and Oh, J.W. (2006) TTSV1, a new virus-like particle isolated from the hyperthermophilic crenarchaeote Thermoproteus tenax. *Virology* **351**: 280-290.

Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R., and Koonin, E.V. (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet* **14**: 442-444.

Arnold, H.P., Ziese, U., and Zillig, W. (2000a) SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon Sulfolobus. *Virology* **272**: 409-416.

Arnold, H.P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J.F., Kristjanson, J.K., Klenk, H.P., Nelson, K.E., and Fraser, C.M. (2000b) A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon Sulfolobus. *Virology* **267**: 252-266.

Bamford, D.H., Grimes, J.M., and Stuart, D.I. (2005) What does structure tell us about virus evolution? *Curr Opin Struct Biol* **15**: 655-663.

Baranyi, U., Klein, R., Lubitz, W., Kruger, D.H., and Witte, A. (2000) The archaeal halophilic virus-encoded Dam-like methyltransferase M. phiCh1-I methylates adenine residues and complements dam mutants in the low salt environment of Escherichia coli. *Mol Microbiol* **35**: 1168-1179.

Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A* **93**: 9188-9193.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

Barry, E.R., and Bell, S.D. (2006) DNA replication in the archaea. *Microbiol Mol Biol Rev* **70**: 876-887.

Barry, R.C., Young, M.J., Stedman, K.M., and Dratz, E.A. (2006) Proteomic mapping of the hyperthermophilic and acidophilic archaeon Sulfolobus solfataricus P2. *Electrophoresis* **27**: 2970-2983.

Bell, S.D., and Jackson, S.P. (2001) Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* **4**: 208-213.

Berkner, S., and Lipps, G. (2008a) Mutation and reversion frequencies of different Sulfolobus species and strains. *Extremophiles* **12**: 263-270.

Berkner, S., and Lipps, G. (2008b) Genetic tools for Sulfolobus spp.: vectors and first applications. *Arch Microbiol* **190**: 217-230.

Bernander, R. (2003) The archaeal cell cycle: current issues. *Mol Microbiol* **48**: 599-604.

Bernander, R. (2007) The cell cycle of Sulfolobus. *Mol Microbiol* **66**: 557-562.

Bettstetter, M., Peng, X., Garrett, R.A., and Prangishvili, D. (2003) AFV1, a novel virus infecting hyperthermophilic archaea of the genus acidianus. *Virology* **315**: 68-79.

Bhagwat, A.S., and McClelland, M. (1992) DNA mismatch correction by Very Short Patch repair may have altered the abundance of oligonucleotides in the E. coli genome. *Nucleic Acids Res* **20**: 1663-1668.

Bigot, S., Saleh, O.A., Lesterlin, C., Pages, C., El Karoui, M., Dennis, C., Grigoriev, M., Allemand, J.F., Barre, F.X., and Cornet, F. (2005) KOPS: DNA motifs that control E. coli chromosome segregation by orienting the FtsK translocase. *Embo J* **24**: 3770-3780.

Birkenbihl, R.P., Neef, K., Prangishvili, D., and Kemper, B. (2001) Holliday junction resolving enzymes of archaeal viruses SIRV1 and SIRV2. *J Mol Biol* **309**: 1067-1076.

Bize, A., Peng, X., Prokofeva, M., Maclellan, K., Lucas, S., Forterre, P., Garrett, R.A., Bonch-Osmolovskaya, E.A., and Prangishvili, D. (2008) Viruses in acidic geothermal environments of the Kamchatka Peninsula. *Res Microbiol* **159**: 358-366.

Blakely, G.W., and Sherratt, D.J. (1994) Interactions of the site-specific recombinases XerC and XerD with the recombination site dif. *Nucleic Acids Res* **22**: 5613-20.

Blochl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H.W., and Stetter, K.O. (1997) Pyrolobus fumarii, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1**: 14-21.

Boetius, A., Ravenschlag, K., Schubert, C.J., Rickert, D., Widdel, F., Gieseke, A., Amann, R., Jorgensen, B.B., Witte, U., and Pfannkuche, O. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**: 623-626.

Boone, D.R. and Castenholz, R.W. (2001) *Bergey's manual of systematic bacteriology*. Springer, New York.

Breitbart, M., Miyake, J.H., and Rohwer, F. (2004a) Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett* **236**: 249-256.

Breitbart, M., Wegley, L., Leeds, S., Schoenfeld, T., and Rohwer, F. (2004b) Phage community dynamics in hot springs. *Appl Environ Microbiol* **70**: 1633-1640.

Briffotaux, J. (2008) Maintenance génomique chez l'Archaea hyperthermophile Pyrococcus abyssi : découverte de nouvelles interactions physiques et caractérisation fonctionnelle. *Thèse, Microbiologie, Bretagne Occidentale*. 200 p.

Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F., and Forterre, P. (2005) Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol* **6**: R42.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6**: 245-252.

Brugger, K., Torarinsson, E., Redder, P., Chen, L., and Garrett, R.A. (2004) Shuffling of Sulfolobus genomes by autonomous and non-autonomous mobile elements. *Biochem Soc Trans* **32**: 179-183.

Brussow, H. (2005) Phage therapy: the Escherichia coli experience. *Microbiology* **151**: 2133-2140.

Buckling, A., and Rainey, P.B. (2002) Antagonistic coevolution between a bacterium and a bacteriophage. *Proc Biol Sci* **269**: 931-936.

Bull, J.J., Molineux, I.J., and Rice, W.R. (1991) Selection of benevolence in a host-parasite system. *Evolution*. **45**: 875-882.

Chaban, B., Ng, S.Y., and Jarrell, K.F. (2006) Archaeal habitats--from the extreme to the ordinary. *Can J Microbiol* **52**: 73-116.

Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H.P., and Garrett, R.A. (2005) The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota. *J Bacteriol* **187**: 4992-4999.

Clore, A.J., and Stedman, K.M. (2007) The SSV1 viral integrase is not essential. *Virology* **361**: 103-111.

Comeau, A.M., Bertrand, C., Letarov, A., Tetart, F., and Krisch, H.M. (2007) Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology* **362**: 384-396.

Comeau, A.M., Hatfull, G.F., Krisch, H.M., Lindell, D., Mann, N.H., and Prangishvili, D. (2008) Exploring the prokaryotic virosphere. *Res Microbiol* **159**: 306-313.

Comeau, A.M., and Krisch, H.M. (2008) The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. *Mol Biol Evol* **25**: 1321-1332.

Contursi, P., Jensen, S., Aucelli, T., Rossi, M., Bartolucci, S., and She, Q. (2006) Characterization of the Sulfolobus host-SSV2 virus interaction. *Extremophiles* **10**: 615-627.

Cowen, D.A. (2004) The upper temperature of life--where do we draw the line? *Trends Microbiol* **12**: 58-60.

de la Torre, J.R., Walker, C.B., Ingalls, A.E., Konneke, M., and Stahl, D.A. (2008) Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol* **10**: 810-818.

De Paepe, M., and Taddei, F. (2006) Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biol* **4**: e193.

Debarbieux, L. (2008) [Experimental phage therapy in the beginning of the 21st century]. *Med Mal Infect* **38**: 421-425.

Denner, E.B.M., McGenity, T.J., Busse, H-J., Grant, W.D., Wanner, G., and Stan-Lotter, H. (1994) Halococcus salifodinae sp. nov., an Archaeal Isolate from an Austrian Salt Mine. *Int. J. Syst. Bacteriol.* **44**: 774-780.

Desplats, C., and Krisch, H.M. (2003) The diversity and evolution of the T4-type bacteriophages. *Res Microbiol* **154**: 259-267.

Dublanchet, A., and Fruciano, E. (2008) [A short history of phage therapy]. *Med Mal Infect* **38**: 415-420.

Duckworth, D.H. (1976) "Who discovered bacteriophage?" *Bacteriol Rev* **40**: 793-802.

Duggin, I.G., and Bell, S.D. (2006) The chromosome replication machinery of the archaeon Sulfolobus solfataricus. *J Biol Chem* **281**: 15029-15032.

Duggin, I.G., Wake, R.G., Bell, S.D., and Hill, T.M. (2008) The replication fork trap and termination of chromosome replication. *Mol Microbiol* **70**: 1323-1333.

Dyall-Smith, M., Tang, S.L., and Bath, C. (2003) Haloarchaeal viruses: how diverse are they? *Res Microbiol* **154**: 309-313.

Elhai, J. (2001) Determination of bias in the relative abundance of oligonucleotides in DNA sequences. *J Comput Biol* **8**: 151-175.

Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L., Hedlund, B.P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E.V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M., and Stetter, K.O. (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci U S A* **105**: 8102-8107.

Ellen, A.F., Albers, S.V., Huibers, W., Pitcher, A., Hobel, C.F., Schwarz, H., Folea, M., Schouten, S., Boekema, E.J., Poolman, B., and Driessen, A.J. (2009) Proteomic analysis of secreted membrane vesicles of archaeal Sulfolobus species reveals the presence of endosome sorting complex components. *Extremophiles* **13**: 67-79.

Filee, J., Forterre, P., Sen-Lin, T., and Laurent, J. (2002) Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *J Mol Evol* **54**: 763-773.

Fontana, J., Lopez-Montero, N., Elliott, R.M., Fernandez, J.J., and Risco, C. (2008) The unique architecture of Bunyamwera virus factories around the Golgi complex. *Cell Microbiol* **10**: 2012-2028.

Forterre, P., Brochier, C., and Philippe, H. (2002) Evolution of the Archaea. *Theor Popul Biol* **61**: 409-422.

Forterre, P. (2006a) The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res* **117**: 5-16.

Forterre, P. (2006b) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc Natl Acad Sci U S A* **103**: 3669-3674.

Fox, G.E., Magrum, L.J., Balch, W.E., Wolfe, R.S., and Woese, C.R. (1977) Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc Natl Acad Sci U S A* **74**: 4537-4541.

Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N., and Woese, C.R. (1980) The phylogeny of prokaryotes. *Science* **209**: 457-463.

Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., and Oakley, B.B. (2005) Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci U S A* **102**: 14683-14688.

Francis, C.A., Beman, J.M., and Kuypers, M.M. (2007) New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. *Isme J* **1**: 19-27.

French, S.L., Santangelo, T.J., Beyer, A.L., and Reeve, J.N. (2007) Transcription and translation are coupled in Archaea. *Mol Biol Evol* **24**: 893-895.

Frols, S., Gordon, P.M., Panlilio, M.A., Schleper, C., and Sensen, C.W. (2007) Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. *Virology* **365**: 48-59.

Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* **399**: 541-548.

Fulton, J., Bothner, B., Lawrence, M., Johnson, J.E., Douglas, T., and Young, M. (2009) Genetics, biochemistry and structure of the archaeal virus STIV. *Biochem Soc Trans* **37**: 114-117.

Gandon, S. (1998) The curse of the pharaoh hypothesis. *Proc Biol Sci* **265**: 1545-1552.

Gelfand, M.S., and Koonin, E.V. (1997) Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes. *Nucleic Acids Res* **25**: 2430-2439.

Giovannoni, S.J., DeLong, E.F., Schmidt, T.M., and Pace, N.R. (1990) Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. *Appl Environ Microbiol* **56**: 2572-2575.

Girguis, P.R., Orphan, V.J., Hallam, S.J., and DeLong, E.F. (2003) Growth and methane oxidation rates of anaerobic methanotrophic archaea in a continuous-flow bioreactor. *Appl Environ Microbiol* **69**: 5472-5482.

Gotz, D., Paytubi, S., Munro, S., Lundgren, M., Bernander, R., and White, M.F. (2007) Responses of hyperthermophilic crenarchaea to UV irradiation. *Genome Biol* **8**: R220.

Gribaldo, S., and Philippe, H. (2002) Ancient phylogenetic relationships. *Theor Popul Biol* **61**: 391-408.

Gribaldo, S., and Brochier-Armanet, C. (2006) The origin and evolution of Archaea: a state of the art. *Philos Trans R Soc Lond B Biol Sci* **361**: 1007-1022.

Grogan, D.W. (2003) Cytosine methylation by the SuaI restriction-modification system: implications for genetic fidelity in a hyperthermophilic archaeon. *J Bacteriol* **185**: 4657-4661.

Gruber, C., Legat, A., Pfaffenhuemer, M., Radax, C., Weidler, G., Busse, H.J., and Stan-Lotter, H. (2004) Halobacterium noricense sp. nov., an archaeal isolate from a bore core of an alpine Permian salt deposit, classification of Halobacterium sp. NRC-1 as a strain of H. salinarum and emended description of H. salinarum. *Extremophiles* **8**: 431-439.

Gusto, G., and Schbath, S. (2005) FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model. *Stat Appl Genet Mol Biol* **4**: Article24.

Hallam, S.J., Girguis, P.R., Preston, C.M., Richardson, P.M., and DeLong, E.F. (2003) Identification of methyl coenzyme M reductase A (mcrA) genes associated with methane-oxidizing archaea. *Appl Environ Microbiol* **69**: 5483-5491.

Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., and DeLong, E.F. (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**: 1457-1462.

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., de la Torre, J., Richardson, P.M., and DeLong, E.F. (2006) Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *Proc Natl Acad Sci U S A* **103**: 18296-18301.

Halpern, D., Chiapello, H., Schbath, S., Robin, S., Hennequet-Antier, C., Gruss, A., and El Karoui, M. (2007) Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. *PLoS Genet* **3**: 1614-1621.

Hanson, P.I., Roth, R., Lin, Y., and Heuser, J.E. (2008) Plasma membrane deformation by circular arrays of ESCRT-III protein filaments. *J Cell Biol* **180**: 389-402.

Haring, M., Peng, X., Brugger, K., Rachel, R., Stetter, K.O., Garrett, R.A., and Prangishvili, D. (2004) Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera Pyrobaculum and Thermoproteus: a novel virus family, the Globuloviridae. *Virology* **323**: 233-242.

Haring, M., Rachel, R., Peng, X., Garrett, R.A., and Prangishvili, D. (2005) Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, Acidianus bottle-shaped virus, from a new family, the Ampullaviridae. *J Virol* **79**: 9904-9911.

Hatfull, G.F., Cresawn, S.G., and Hendrix, R.W. (2008) Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. *Res Microbiol* **159**: 332-339.

Hendrickson, H., and Lawrence, J.G. (2006) Selection for chromosome architecture in bacteria. *J Mol Evol* **62**: 615-629.

Hendrix, R.W. (2003) Bacteriophage genomics. *Curr Opin Microbiol* **6**: 506-511.

Herndl, G.J., Reinthaler, T., Teira, E., van Aken, H., Veth, C., Pernthaler, A., and Pernthaler, J. (2005) Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Appl Environ Microbiol* **71**: 2303-2309.

Hoebeke, M. and Schbath, S. (2006) R'MES: Finding Exceptional Motifs, version 3. *User guide*. http://genome.jouy.inra.fr/ssb/rmes.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., and Stetter, K.O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**: 63-67.

Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**: 4765-4774.

Hugenholtz, P. (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3**: REVIEWS0003.

Itoh, T. (2003) Taxonomy of nonmethanogenic hyperthermophilic and related thermophilic archaea. *J Biosci Bioeng* **96**: 203-212.

Janekovic, D., Wunderl, S., Holz, I., Zillig, W., Gierl, A. and Neumann, H. (1983) TTV1, TTV2 and TTV3, a family of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaebacterium Thermoproteus tenax. *Mol. Gen. Genet*. **192**: 39-45.

Jordan, M., Meile, L., and Leisinger, T. (1989) Organization of Methanobacterium thermoautotrophicum bacteriophage psi M1 DNA. *Mol Gen Genet* **220**: 161-164.

Kandler, O., and Konig, H. (1998) Cell wall polymers in Archaea (Archaebacteria). *Cell Mol Life Sci* **54**: 305-308.

Karner, M.B., DeLong, E.F., and Karl, D.M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**: 507-510.

Keller, J., Leulliot, N., Cambillau, C., Campanacci, V., Porciero, S., Prangishvilli, D., Forterre, P., Cortez, D., Quevillon-Cheruel, S., and van Tilbeurgh, H. (2007) Crystal structure of AFV3-109, a highly conserved protein from crenarchaeal viruses. *Virol J* **4**: 12.

Kelman, L.M., and Kelman, Z. (2004) Multiple origins of replication in archaea. *Trends Microbiol* **12**: 399-401.

Kessler, A., Brinkman, A.B., van der Oost, J., and Prangishvili, D. (2004) Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon sulfolobus. *J Bacteriol* **186**: 7745-7753.

Kessler, A., Sezonov, G., Guijarro, J.I., Desnoues, N., Rose, T., Delepierre, M., Bell, S.D., and Prangishvili, D. (2006) A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters. *Nucleic Acids Res* **34**: 4837-4845.

Kobayashi, I., Nobusato, A., Kobayashi-Takahashi, N., and Uchiyama, I. (1999) Shaping the genome--restriction-modification systems as mobile genetic elements. *Curr Opin Genet Dev* **9**: 649-656.

Konneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543-546.

Krupovic, M., Daugelavicius, R., and Bamford, D.H. (2007) A novel lysis system in PM2, a lipid-containing marine double-stranded DNA bacteriophage. *Mol Microbiol* **64**: 1635-1648.

Kvist, T., Mengewein, A., Manzei, S., Ahring, B.K., and Westermann, P. (2005) Diversity of thermophilic and non-thermophilic Crenarchaeota at 80 degrees C. *FEMS Microbiol Lett* **244**: 61-68.

Kwan, T., Liu, J., DuBow, M., Gros, P., and Pelletier, J. (2005) The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. *Proc Natl Acad Sci U S A* **102**: 5174-5179.

Larson, E.T., Eilers, B.J., Reiter, D., Ortmann, A.C., Young, M.J., and Lawrence, C.M. (2007) A new DNA binding protein highly conserved in diverse crenarchaeal viruses. *Virology* **363**: 387-396.

Lawrence, C.M., Menon, S., Eilers, B.J., Bothner, B., Khayat, R., Douglas, T., and Young, M.J. (2009) Structural and functional studies of archaeal viruses. *J Biol Chem*.

Letzelter, C., Duguet, M., and Serre, M.C. (2004) Mutational analysis of the archaeal tyrosine recombinase SSV1 integrase suggests a mechanism of DNA cleavage in trans. *J Biol Chem* **279**: 28936-28944.

Leung, M.Y., Marsh, G.M., and Speed, T.P. (1996) Over- and underrepresentation of short DNA words in herpesvirus genomes. *J Comput Biol* **3**: 345-360.

Levy, O., Ptacin, J.L., Pease, P.J., Gore, J., Eisen, M.B., Bustamante, C., and Cozzarelli, N.R. (2005) Identification of oligonucleotide sequences that direct the movement of the Escherichia coli FtsK translocase. *Proc Natl Acad Sci U S A* **102**: 17618-17623.

Lillestol, R.K., Redder, P., Garrett, R.A., and Brugger, K. (2006) A putative viral defence mechanism in archaeal cells. *Archaea* **2**: 59-72.

Lindas, A.C., Karlsson, E.A., Lindgren, M.T., Ettema, T.J., and Bernander, R. (2008) A unique cell division machinery in the Archaea. *Proc Natl Acad Sci U S A* **105**: 18942-18946.

Lisnic, B., Svetec, I.K., Saric, H., Nikolic, I., and Zgaga, Z. (2005) Palindrome content of the yeast Saccharomyces cerevisiae genome. *Curr Genet* **47**: 289-297.

Liu, Y., and Whitman, W.B. (2008) Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann N Y Acad Sci* **1125**: 171-189.

Lopez-Garcia, P., Moreira, D., Lopez-Lopez, A., and Rodriguez-Valera, F. (2001) A novel haloarchaeal-related lineage is widely distributed in deep oceanic regions. *Environ Microbiol* **3**: 72-78.

Lopez-Garcia, P., and Moreira, D. (2006) Selective forces for the origin of the eukaryotic nucleus. *Bioessays* **28**: 525-533.

Lopez-Garcia, P., and Moreira, D. (2008) Tracking microbial biodiversity through molecular and genomic ecology. *Res Microbiol* **159**: 67-73.

Lubben, M., and Schafer, G. (1989) Chemiosmotic energy conversion of the archaebacterial thermoacidophile Sulfolobus acidocaldarius: oxidative phosphorylation and the presence of an F0-related N,N'-dicyclohexylcarbodiimide-binding proteolipid. *J Bacteriol* **171**: 6106-6116.

Lundgren, M., Andersson, A., Chen, L., Nilsson, P., and Bernander, R. (2004) Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* **101**: 7046-7051.

Lundgren, M., Malandrin, L., Eriksson, S., Huber, H., and Bernander, R. (2008) Cell cycle characteristics of crenarchaeota: unity among diversity. *J Bacteriol* **190**: 5362-5367.

Magrum, L.J., Luehrsen, K.R., and Woese, C.R. (1978) Are extreme halophiles actually "bacteria"? *J Mol Evol* **11**: 1-8.

Maki, H. (2002) Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annu Rev Genet* **36**: 279-303.

Mann, N.H. (2005) The third age of phage. *PLoS Biol* **3**: e182.

Martin, A., Yeats, S., Janekovic, D., Reiter, W.D., Aicher, W., and Zillig, W. (1984) SAV 1, a temperate u.v.-inducible DNA virus-like particle from the archaebacterium Sulfolobus acidocaldarius isolate B12. *Embo J* **3**: 2165-2168.

Matsunaga, F., Forterre, P., Ishino, Y., and Myllykallio, H. (2001) In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci U S A* **98**: 11152-11157.

Menon, S.K., Maaty, W.S., Corn, G.J., Kwok, S.C., Eilers, B.J., Kraft, P., Gillitzer, E., Young, M.J., Bothner, B., and Lawrence, C.M. (2008) Cysteine usage in Sulfolobus spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. *Virology* **376**: 270-278.

Mercier, R., Petit, M.A., Schbath, S., Robin, S., El Karoui, M., Boccard, F., and Espeli, O. (2008) The MatP/matS site-specific system organizes the terminus region of the E. coli chromosome into a macrodomain. *Cell* **135**: 475-485.

Miller, S., and Krijnse-Locker, J. (2008) Modification of intracellular membrane structures for virus replication. *Nat Rev Microbiol* **6**: 363-374.

Mindich, L., Sinclair, J.F., and Cohen, J. (1976) The morphogenesis of bacteriophage phi6: particles formed by nonsense mutants. *Virology* **75**: 224-231.

Morgan, R., Xiao, J., and Xu, S. (1998) Characterization of an extremely thermostable restriction enzyme, PspGI, from a Pyrococcus strain and cloning of the PspGI restriction-modification system in Escherichia coli. *Appl Environ Microbiol* **64**: 3669-3673.

Morita, E., and Sundquist, W.I. (2004) Retrovirus budding. *Annu Rev Cell Dev Biol* **20**: 395-425.

Muskhelishvili, G., Palm, P., and Zillig, W. (1993) SSV1-encoded site-specific recombination system in Sulfolobus shibatae. *Mol Gen Genet* **237**: 334-342.

Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H., and Forterre, P. (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288**: 2212-2215.

Neylon, C., Kralicek, A.V., Hill, T.M., and Dixon, N.E. (2005) Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiol Mol Biol Rev* **69**: 501-526.

Nolling, J., and de Vos, W.M. (1992a) Identification of the CTAG-recognizing restriction-modification systems MthZI and MthFI from Methanobacterium thermoformicicum and characterization of the plasmid-encoded mthZIM gene. *Nucleic Acids Res* **20**: 5047-5052.

Nolling, J., and de Vos, W.M. (1992b) Characterization of the archaeal, plasmid-encoded type II restriction-modification system MthTI from Methanobacterium thermoformicicum THF: homology to the bacterial NgoPII system from Neisseria gonorrhoeae. *J Bacteriol* **174**: 5719-5726.

Norrby, E. (2008) Nobel Prizes and the emerging virus concept. *Arch Virol* **153**: 1109-1123.

Novoa, R.R., Calderita, G., Arranz, R., Fontana, J., Granzow, H., and Risco, C. (2005) Virus factories: associations of cell organelles for viral replication and morphogenesis. *Biol Cell* **97**: 147-172.

Olsen, G.J., and Woese, C.R. (1996) Lessons from an Archaeal genome: what are we learning from Methanococcus jannaschii? *Trends Genet* **12**: 377-379.

Ortmann, A.C., Brumfield, S.K., Walther, J., McInnerney, K., Brouns, S.J., van de Werken, H.J., Bothner, B., Douglas, T., van de Oost, J., and Young, M.J. (2008) Transcriptome analysis of infection of the archaeon Sulfolobus solfataricus with Sulfolobus turreted icosahedral virus. *J Virol* **82**: 4874-4883.

Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.D., and Zillig, W. (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium Sulfolobus shibatae. *Virology* **185**: 242-250.

Peng, X. (2008) Evidence for the horizontal transfer of an integrase gene from a fusellovirus to a pRN-like plasmid within a single strain of Sulfolobus and the implications for plasmid survival. *Microbiology* **154**: 383-391.

Peng, X., Blum, H., She, Q., Mallok, S., Brugger, K., Garrett, R.A., Zillig, W., and Prangishvili, D. (2001) Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. *Virology* **291**: 226-234.

Peng, X., Brugger, K., Shen, B., Chen, L., She, Q., and Garrett, R.A. (2003) Genus-specific protein binding to the large clusters of DNA repeats (short regularly spaced repeats) present in Sulfolobus genomes. *J Bacteriol* **185**: 2410-2417.

Peng, X., Kessler, A., Phan, H., Garrett, R.A., and Prangishvili, D. (2004) Multiple variants of the archaeal DNA rudivirus SIRV1 in a single host and a novel mechanism of genomic variation. *Mol Microbiol* **54**: 366-375.

Pennazio, S. (2006) The origin of phage virology. *Riv Biol* **99**: 103-129.

Pfister, P., Wasserfallen, A., Stettler, R., and Leisinger, T. (1998) Molecular analysis of Methanobacterium phage psiM2. *Mol Microbiol* **30**: 233-244.

Philippe, H., and Forterre, P. (1999) The rooting of the universal tree of life is not reliable. *J Mol Evol* **49**: 509-523.

Porter, K., Russ, B.E., and Dyall-Smith, M.L. (2007) Virus-host interactions in salt lakes. *Curr Opin Microbiol* **10**: 418-424.

Prangishvili, D., Klenk, H.P., Jakobs, G., Schmiechen, A., Hanselmann, C., Holz, I., and Zillig, W. (1998) Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. *J Biol Chem* **273**: 6024-6029.

Prangishvili, D., Arnold, H.P., Gotz, D., Ziese, U., Holz, I., Kristjansson, J.K., and Zillig, W. (1999) A novel virus family, the Rudiviridae: Structure, virus-host interactions and genome variability of the sulfolobus viruses SIRV1 and SIRV2. *Genetics* **152**: 1387-1396.

Prangishvili, D., Holz, I., Stieger, E., Nickell, S., Kristjansson, J.K., and Zillig, W. (2000) Sulfolobicins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus Sulfolobus. *J Bacteriol* **182**: 2985-2988.

Prangishvili, D., and Garrett, R.A. (2005) Viruses of hyperthermophilic Crenarchaea. *Trends Microbiol* **13**: 535-542.

Prangishvili, D., Forterre, P., and Garrett, R.A. (2006a) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4**: 837-848.

Prangishvili, D., Garrett, R.A., and Koonin, E.V. (2006b) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res* **117**: 52-67.

Prangishvili, D., Vestergaard, G., Haring, M., Aramayo, R., Basta, T., Rachel, R., and Garrett, R.A. (2006c) Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *J Mol Biol* **359**: 1203-1216.

Preston, C.M., Wu, K.Y., Molinski, T.F., and DeLong, E.F. (1996) A psychrophilic crenarchaeon inhabits a marine sponge: Cenarchaeum symbiosum gen. nov., sp. nov. *Proc Natl Acad Sci U S A* **93**: 6241-6246.

Prosser, J.I., and Nicol, G.W. (2008) Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environ Microbiol* **10**: 2931-2941.

Redder, P., and Garrett, R.A. (2006) Mutations and rearrangements in the genome of Sulfolobus solfataricus P2. *J Bacteriol* **188**: 4198-4206.

Reeve, J.N., Sandman, K., and Daniels, C.J. (1997) Archaeal histones, nucleosomes, and transcription initiation. *Cell* **89**: 999-1002.

Reiter, W-D., Palm, P., Henschen, A., Lottspeich, F., Zillig, W., and Grampp, B. (1987a) Identification and characterization of the genes encoding three structural proteins of the Sulfolobus virus-like particle SSV1. *Mol. Gen. Genet.* **206**: 144-153.

Reiter, W-D., Palm, P, Yeats, S., and Zillig, W. (1987b) Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the Sulfolobus virus-like particle SSV1. *Mol Gen Genet* **209**: 270-275

Reysenbach, A.L., and Shock, E. (2002) Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296**: 1077-1082.

Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., Johnson, J.E., Douglas, T., and Young, M. (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc Natl Acad Sci U S A* **101**: 7716-7720.

Rivera, M.C., and Lake, J.A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**: 152-155.

Robertson, C.E., Harris, J.K., Spear, J.R., and Pace, N.R. (2005) Phylogenetic diversity and ecology of environmental Archaea. *Curr Opin Microbiol* **8**: 638-642.

Robin, S., Rodolphe, F., and Schbath, S. (2005) *DNA, Worlds and Models*. Cambridge University Press, New York.

Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R., and Bell, S.D. (2004) Identification of two origins of replication in the single chromosome of the archaeon Sulfolobus solfataricus. *Cell* **116**: 25-38.

Robinson, N.P., and Bell, S.D. (2005) Origins of DNA replication in the three domains of life. *Febs J* **272**: 3757-3766.

Rocha, E.P., Viari, A., and Danchin, A. (1998) Oligonucleotide bias in Bacillus subtilis: general trends and taxonomic comparisons. *Nucleic Acids Res* **26**: 2971-2980.

Rocha, E.P., Danchin, A., and Viari, A. (2001) Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Res* **11**: 946-958.

Rothschild, L.J., and Mancinelli, R.L. (2001) Life in extreme environments. *Nature* **409**: 1092-1101.

Schbath, S. (1995) Compound Poisson approximation of word counts in DNA sequences. *ESAIM: Probability and Statistics*. **1**: 1-16.

Schbath, S. (1997) An efficient statistic to detect over- and under- represented words in DNA sequences. *J. Comp. Biol.* **4**: 189-192.

Schleper, C., Kubo, K., and Zillig, W. (1992) The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc Natl Acad Sci U S A* **89**: 7645-7649.

Schleper, C., Jurgens, G., and Jonuscheit, M. (2005) Genomic studies of uncultivated archaea. *Nat Rev Microbiol* **3**: 479-488.

Schoenfeld, T., Patterson, M., Richardson, P.M., Wommack, K.E., Young, M., and Mead, D. (2008) Assembly of viral metagenomes from yellowstone hot springs. *Appl Environ Microbiol* **74**: 4164-4174.

Scholthof, K.B. (2004) Tobacco mosaic virus: a model system for plant biology. *Annu Rev Phytopathol* **42**: 13-34.

She, Q., Zhang, C., Deng, L., Peng, N., Chen, Z., and Liang, Y.X. (2009) Genetic analyses in the hyperthermophilic archaeon Sulfolobus islandicus. *Biochem Soc Trans* **37**: 92-96.

Simon, H.M., Jahn, C.E., Bergerud, L.T., Sliwinski, M.K., Weimer, P.J., Willis, D.K., and Goodman, R.M. (2005) Cultivation of mesophilic soil crenarchaeotes in enrichment cultures from plant roots. *Appl Environ Microbiol* **71**: 4751-4760.

Smith, G.R., Kunes, S.M., Schultz, D.W., Taylor, A., and Triman, K.L. (1981) Structure of chi hotspots of generalized recombination. *Cell* **24**: 429-436.

Smrekar, F., Ciringer, M., Peterka, M., Podgornik, A., and Strancar, A. (2008) Purification and concentration of bacteriophage T4 using monolithic chromatographic supports. *J Chromatogr B Analyt Technol Biomed Life Sci* **861**: 177-180.

Snyder, J.C., Wiedenheft, B., Lavin, M., Roberto, F.F., Spuhler, J., Ortmann, A.C., Douglas, T., and Young, M. (2007) Virus movement maintains local virus population diversity. *Proc Natl Acad Sci U S A* **104**: 19102-19107.

Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., and Wommack, K.E. (2008) Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res Microbiol* **159**: 349-357.

Stahl, D.A., Lane, D.J., Olsen, G.J., and Pace, N.R. (1984) Analysis of Hydrothermal Vent-Associated Symbionts by Ribosomal RNA Sequences. *Science* **224**: 409-411.

Stedman, K.M., She, Q., Phan, H., Arnold, H.P., Holz, I., Garrett, R.A., and Zillig, W. (2003) Relationships between fuselloviruses infecting the extremely thermophilic archaeon Sulfolobus: SSV1 and SSV2. *Res Microbiol* **154**: 295-302.

Steinmetz, N.F., Bize, A., Findlay, K.C., Lomonossoff, G.P., Manchester, M., Evans, D.J., and Prangishvili, D. (2008) Site-specific and spatially controlled addressability of a new viral nanobuilding block: *Sulfolobus islandicus* rod-shaped virus 2. *Adv Funct Mater* **18**: 1-9.

Stetter, K.O., Lauerer, G., Thomm, M., and Neuner, A. (1987) Isolation of Extremely Thermophilic Sulfate Reducers: Evidence for a Novel Branch of Archaebacteria. *Science* **236**: 822-824.

Stetter, K.O. (1996) Hyperthermophiles in the history of life. *Ciba Found Symp* **202**: 1-10; discussion 11-18.

Sulakvelidze, A., Alavidze, Z., and Morris, J.G., Jr. (2001) Bacteriophage therapy. *Antimicrob Agents Chemother* **45**: 649-659.

Suttle, C.A. (2007) Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801-812.

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

Takanami, Y. (2006) Contributions of plant virus research to the biological sciences. *J. Gen. Plant. Pathol.* **72**: 393-395.

Tang, S.L., Nuttall, S., Ngui, K., Fisher, C., Lopez, P., and Dyall-Smith, M. (2002) HF2: a double-stranded DNA tailed haloarchaeal virus with a mosaic genome. *Mol Microbiol* **44**: 283-296.

Teira, E., Lebaron P., van Aken, H., and Herndl, G.J. (2006) Distribution and activity of Bacteria and Archaea in the deep water masses of the North Atlantic. *Limnol Oceanogr* **51**: 2131-2144.

Teske, A., and Sorensen, K.B. (2008) Uncultured archaea in deep marine subsurface sediments: have we caught them all? *Isme J* **2**: 3-18.

Torsvik, T., and Dundas, I.D. (1974) Bacteriophage of Halobacterium salinarium. *Nature* **248**: 680-681.

Valentine, D.L. (2002) Biogeochemistry and microbial ecology of methane oxidation in anoxic environments: a review. *Antonie Van Leeuwenhoek* **81**: 271-282.

Valentine, D.L. (2007) Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat Rev Microbiol* **5**: 316-323.

van der Wielen, P.W., Bolhuis, H., Borin, S., Daffonchio, D., Corselli, C., Giuliano, L., D'Auria, G., de Lange, G.J., Huebner, A., Varnavas, S.P., Thomson, J., Tamburini, C., Marty, D., McGenity, T.J., and Timmis, K.N. (2005) The enigma of prokaryotic life in deep hypersaline anoxic basins. *Science* **307**: 121-123.

Vestergaard, G., Haring, M., Peng, X., Rachel, R., Garrett, R.A., and Prangishvili, D. (2005) A novel rudivirus, ARV1, of the hyperthermophilic archaeal genus Acidianus. *Virology* **336**: 83-92.

Vestergaard, G., Aramayo, R., Basta, T., Haring, M., Peng, X., Brugger, K., Chen, L., Rachel, R., Boisset, N., Garrett, R.A., and Prangishvili, D. (2008a) Structure of the acidianus filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. *J Virol* **82**: 371-381.

Vestergaard, G., Shah, S.A., Bize, A., Reitberger, W., Reuter, M., Phan, H., Briegel, A., Rachel, R., Garrett, R.A., and Prangishvili, D. (2008b) Stygiolobus rod-shaped virus and the interplay of crenarchaeal rudiviruses with the CRISPR antiviral system. *J Bacteriol* **190**: 6837-6845.

Volkl, P., Huber, R., Drobner, E., Rachel, R., Burggraf, S., Trincone, A., and Stetter, K.O. (1993) Pyrobaculum aerophilum sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. *Appl Environ Microbiol* **59**: 2918-2926.

Walther, B.A., and Ewald, P.W. (2004) Pathogen survival in the external environment and the evolution of virulence. *Biol Rev Camb Philos Soc* **79**: 849-869.

Watanabe, M., Yuzawa, H., Handa, N., and Kobayashi, I. (2006) Hyperthermophilic DNA methyltransferase M.PabI from the archaeon Pyrococcus abyssi. *Appl Environ Microbiol* **72**: 5367-5375.

Weinbauer, M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**: 127-181.

Welch, S.G., and Williams, R.A. (1995) Taq52 I, a novel and thermostable type II restriction endonuclease from the genus Thermus, recognising the pentanucleotide sequence GC(A or T)GC and cleaving DNA between the first and second bases of the recognition sequence: G decreased or reduced C(A or T)GC. *Nucleic Acids Res* **23**: 4573-4575.

Werner, F. (2007) Structure and function of archaeal RNA polymerases. *Mol Microbiol* **65**: 1395-1404.

White, M.F., and Bell, S.D. (2002) Holding it together: chromatin in the Archaea. *Trends Genet* **18**: 621-626.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**: 6578-6583.

Woese, C.R., and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**: 5088-5090.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**: 4576-4579.

Woese, C. (1998) The universal ancestor. *Proc Natl Acad Sci U S A* **95**: 6854-6859.

Wommack, K.E., and Colwell, R.R. (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**: 69-114.

Xiang, X., Chen, L., Huang, X., Luo, Y., She, Q., and Huang, L. (2005) Sulfolobus tengchongensis spindle-shaped virus STSV1: virus-host interactions and genomic features. *J Virol* **79**: 8677-8686.

Yin, Y., and Fischer, D. (2008) Identification and investigation of ORFans in the viral world. *BMC Genomics* **9**: 24.

Yoshikawa, H., and Ogasawara, N. (1991) Structure and function of DnaA and the DnaA-box in eubacteria: evolutionary relationships of bacterial replication origins. *Mol Microbiol* **5**: 2589-2597.

Zillig, W., Stetter, K.O., and Janekovic, D. (1979) DNA-dependent RNA polymerase from the archaebacterium Sulfolobus acidocaldarius. *Eur J Biochem* **96**: 597-604.

Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., Lanzendörfer, M., and Kristjanson, J.K. (1994) Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. *System. Appl. Microbiol*. **16**: 609-628.

Zivanovic, Y., Lopez, P., Philippe, H., and Forterre, P. (2002) Pyrococcus genome comparison evidences chromosome shuffling-driven evolution. *Nucleic Acids Res* **30**: 1902-1910.

Abréviations

ACP : Analyse en Composantes Principales
ADN : Acide DesoxyRiboNucléique
AFV9 : Acidianus Filamentous Virus 9
ANME : Archées anaérobies oxydant le méthane
ARN : Acide RiboNucléique
ARNP : ARN polymérase
ARNr : ARN ribosomique
b : base
BMGE : Unité de Biologie Moléculaire du Gène chez les Extrêmophiles
CFU : Colony Forming Units
CH : Clustering hiérarchique
CRISPRs : Clustered Regularly Interspaced Short Palindromic Repeats
DMSO : Dimethyl Sulfoxide
DO : Densité Optique
dUTPase : dUTP pyrophosphatase
ESCRT : Endosomal Sorting Complex Required for Transport
HIV : Human Immunodeficiency Virus
ICTV : International Commitee on Taxonomy of Viruses
LUCA : Last Universal Common Ancestor
ORB : Origin Recognition Box
ORC : Origin recognition complex
ORF : Open Reading Frame (cadre ouvert de lecture)
ORFan : orphan ORF (cadre de lecture ouvert orphelin)
MCM : Mini-chromosome Maintenance
nm : nanomètre
pb : paires de base
PCNA : Proliferating Cell Nuclear Antigen
PEG : polyéthylène glycol
PFU : Plaque Forming Units
Pol : Polymérase
PSU-ARNr : petite sous-unité de l'ARN ribosomique
RHH : Ribon-Helix-Helix
SIRV2 : *Sulfolobus islandicus* rod-shaped virus
SRV : *Stygiolobus* rod-shaped virus
T : température
TBP : TATA binding protein
TF : Transcrition Factor
TMV : Tobacco Mosaic Virus
UV : Ultra-violet
VIP : Viral Induced Protrusion
Vsr : Very Short Patch Repair

ABV : *Acidianus* bottle-shaped virus
AFV1 à AFV9 : *Acidianus* filamentous virus 1 to 9
ARV1 : *Acidianus* rod-shaped virus 1
ATV : *Acidianus* two-tailed virus
BJ1 : archaeal BJ1
HF1 : Halovirus HF1
HF2 : Halovirus HF2
His1 : His1 virus
His2 : His2 virus
PAV1 : *Pyrococcus abyssi* virus 1
PhiH et PhiCh1 (ΦH et Φch1) : *Natrialba* phage PhiH et PhiCh1
PHV1 : *Pyrococcus horikoshii* provirus 1
PsiM2 et PsiM100 (ΨM2 ΨM100) : archaeal phage PsiM2 et archaeal prophage PsiM100
PSV : *Pyrobaculum* spherical virus
SIFV : *Sulfolobus islandicus* filamentous virus
SH1 : *Haloarcula* phage SH1
SIRV1 et SIRV2 : *Sulfolobus islandicus* rod-shaped virus 1 et 2
SNDV : *Sulfolobus neozealandicus* droplet-shaped virus
SRV : *Stygiolobus* rod-shaped virus
SSV-RH : *Sulfolobus* spindle-shaped virus Ragged Hills
SSV-K1 : *Sulfolobus* spindle-shaped virus Kamchatka 1
SSV1 à SSV6 : *Sulfolobus* islandicus spindle-shaped virus 1 à 6
STIV : *Sulfolobus* turreted icosahedral virus
STSV1 : *Sulfolobus tengchongensis* spindle-shaped virus 1
TKV1 à TKV4 : *Thermococcus* kodakarensis prophage 1 à 4
TTSV1 :*Thermoproteus tenax* spherical virus 1
TTV1 à TTV4 : *Thermoproteus tenax* virus 1

Table des illustrations

Figures

Numéro	Titre	Page
Figure 1	Carl Woese photographié a côté d'un gel d'électrophorèse bi-dimensionnel	17
Figure 2	Les membranes d'archées et de bactéries	19
Figure 3	Exemples de cellules d'archées aux morphologies variées	21
Figure 4	Température et pH requis pour la croissance d'archées et de bactéries	22
Figure 5	Arbre phylogénétique schématique des archées cultivées et non cultivées	24
Figure 6	Lac salé en Utah, aux Etats-Unis, coloré par la présence d'archées halophiles	26
Figure 7	Identification *in situ* de microorganismes	29
Figure 8A	L'Arbre universel d'Haeckel (1866)	32
Figure 8B	Les 5 règnes de Whittaker (1969)	33
Figure 9	Exemple schématiques d'arbres du vivant	34
Figure 10	Virus et boucle microbienne	38
Figure 11	Exemples représentatifs de morphotypes de virus d'archées	41
Figure 12	Similarités de gènes chez les virus de crenarchées hyperthermophiles	44
Figure 13	Les virus à ADN double-brin dans les 3 domaines du vivant	46
Figure 14	L'origine des virus	47
Figure 15	Cellules de crenarchées infectées par des virus	53
Figure 16	Représentation schématique du cycle cellulaire chez *Sulfolobus*	58
Figure 17	Carte des ORFs (cadres ouverts de lecture) dans le génome de SIRV2	60
Figure 18	Décroissance du titre de différentes préparations de virus en fonction du temps	108
Figure 19	ACP de la composition en mots de 4 lettres des séquences d'éléments extra-chromosomiques	122
Figure 20	Analyse par clustering hiérarchique de la composition en mots de 4 lettres des séquences d'éléments extra-chromosomiques	123
Figure 21	Analyse par clustering hiérarchique de la composition en mots de 4 lettres des séquences d'éléments extra-chromosomiques (sous-groupes de Sulfolobales)	124
Figure 22	Scores d'exceptionnalité des mots de 4 lettres pour les séquences concaténées LIPOTHRIX et pour *Sulfolobus acidocaldarius*	128
Figure 23	Scores d'exceptionnalité des mots de 5 lettres chez *Pyrobaculum aerophilum* et *Pyrobaculum arsenaticum*	129
Figure 24	Scores d'exceptionnalité des mots de 6 lettres dans le génome de *Natronomonas pharaonis* et les séquences concaténées HALO_EXTRA	129
Figure 25	Scores d'exceptionnalité des mots de 4 lettres dans les séquences d'*Haloquadratum walsbyi* et les séquences concaténées HALO_EXTRA	130
Figure 26	Scores d'exceptionnalité des mots de taille 4 dans les séquences concaténées CONJUG, et LIPOTHRIX	133

Figure 27 Scores d'exceptionnalité des palindromes de 4 lettres dans le génome de *Methanopyrus kandleri* 133
Figure 28 Exemples d'exceptionnalité globale des mots symétriques 134
Figure 29 Scores d'exceptionnalité des mots symétriques de 7 lettres chez *Ignicoccus hospitalis* et *Aeropyrum pernix* 135
Figure 30 Le motif AAAAAA dans le génome de *Sulfolobus acidocaldarius* 138
Figure 31 Distribution du mot AATCC dans le génome de *Sulfolobus solfataricus* 139
Figure 32 Scores des mots de 6 lettres dans les génomes d'espèces du genre *Pyrococcus* 140
Figure 33 Distance entre deux motifs successifs CCTGGG chez *Pyrococcus horikoshii* 141
Figure 34 Sections de cellules de *Sulfolobus islandicus* infectées par SIRV2, en fin de cycle infectieux 167

Tableaux

Numéro	Titre	Page
Tableau 1	Comparaison des principales caractéristiques des trois domaines du vivants	18
Tableau 2	Les habitats des archées hyperthermophiles, thermophiles et acidophiles	28
Tableau 3	Les réactions productrices d'énergie chez les archées hyperthermophiles	28
Tableau 4	La diversité des virus d'archées : principales caractéristiques des virus d'archées isolés et décrits dans la littérature	42
Tableau 5	Données expérimentales sur l'interaction hôte-virus chez les crenarchées hyperthermophiles	52
Tableau 6	Stabilité des particules virales de SIRV2 dans différentes solutions	107
Tableau 7	Exemples de motifs fonctionnels non-codants chez *Escherichia coli*	112
Tableau 8	Répartition des séquences utilisées pour l'étude	116
Tableau 9	Séquences concaténées, reflétant les ordres auxquels appartiennent les hôtes des virus et plasmides	125
Tableau 10	Séquences concaténées, reflétant les familles auxquels appartiennent les virus et plasmides de Sulfolobales	125
Tableau 11	Rang d'exceptionnalité de palindromes correspondant à des systèmes de restriction-modification (ou de méthylation) déjà connus chez les archées	127
Tableau 12	Palindromes et pseudo-palindromes exceptionnels de 4 à 6 lettres, dans les séquences d'archées et de leurs éléments extra-chromosomiques	131
Tableau 13	Motifs les plus exceptionnels (selon leu rang) communs à plusieurs espèces d'un même genre ou d'un même ordre	137
Tableau 14	Régions d'évitement du mot CCTGGG dans les génomes d'espèces de *Pyrococcus*	140

Tableau A	Rangs d'exceptionnalité des palindromes de 4 lettres dans les différentes séquences	143
Tableau B	Rangs d'exceptionnalité de certains pseudo-palindromes de 5 lettres dans les diverses séquences	144
Tableau C	Rangs d'exceptionnalité de certaines palindromes de 6 lettres dans les diverses séquences	145
Tableau D	Rangs d'exceptionnalité de divers motifs de 4 lettres dans les séquences	146
Tableau E	Rangs d'exceptionnalité de divers motifs de 5 et 6 lettres dans les séquences	147
Tableau F	Exceptionnalité globale des palindromes et pseudo-palindromes de 5 à 8 lettres dans les différentes séquences	148
Tableau G	Exceptionnalité globale des mots symétriques de 5 à 8 lettres dans les différentes séquences	149