HAL
open science

Contribution du désordre intrinsèque des protéines aux fonctions impliquées dans le cycle viral et l'évolution adaptative des virus à ARN : étude appliquée au genre modèle Potyvirus

Justine Charon

To cite this version:

Justine Charon. Contribution du désordre intrinsèque des protéines aux fonctions impliquées dans le cycle viral et l'évolution adaptative des virus à ARN : étude appliquée au genre modèle Potyvirus. Biologie végétale. Université de Bordeaux, 2015. Français. NNT : 2015BORD0373 . tel-01456446

HAL Id: tel-01456446

https://theses.hal.science/tel-01456446

Submitted on 5 Feb 2017

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.

THÈSE

PRÉSENTÉE POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé

SPÉCIALITÉ : Biologie Végétale

Par

Justine CHARON

Née le 03 Novembre 1989 à Paris

Contribution du désordre intrinsèque des protéines aux fonctions impliquées dans le cycle viral et l'évolution adaptative des virus à ARN

Etude appliquée au genre modèle *Potyvirus*

Sous la direction de M. Thierry MICHON

Soutenue le 17 Décembre 2015

Membres du jury :

M. BLANCHARD Alain,	Professeur de l'Université de Bordeaux	Président
M. BLANC Stéphane,	Directeur de Recherche INRA	Rapporteur
M. TERRADO, Laurent	Chargé de Recherche CNRS	Rapporteur
M. FORTERRE, Patrick	Professeur de l'Université Paris-Sud	Examinateur
M. MOURY, Benoît	Directeur de Recherche INRA	Membre invité

"Le savant n'est pas l'homme qui fournit les vraies réponses,
c'est celui qui pose les vraies questions"

Claude Lévi-strauss

Remerciements

Et voilà on y est... Après trois années riches en émotions et en rebondissements, le moment que j'ai tant attendu, celui de pouvoir tous vous remercier, est enfin arrivé !

Je souhaite tout d'abord remercier Thierry Candresse pour m'avoir accueilli au sein du l'équipe de virologie végétale de l'UMR 1332 Biologie du Fruit et Pathologie. Thierry, je te suis extrêmement reconnaissante pour ta disponibilité, l'intérêt que tu as a manifesté pour ce travail et bien sûr un grand merci pour avoir validé l'ensemble de mes demandes de formations et autres aventures hors frontières.

J'adresse également mes plus sincères remerciements aux membres du jury qui m'ont fait l'honneur d'évaluer ce travail, Messieurs Stéphane Blanc, Alain Blanchard, Patrick Forterre, Benoît Moury et Laurent Terradot.

Un grand merci également aux membres de mes comités de thèse : Eugénie, Sonia, Benoît et Michel. Merci pour votre intérêt, vos conseils, vos idées, votre enthousiasme et votre humilité qui ont été absolument cruciaux pour le bon déroulement de cette thèse ainsi que pour ma détermination à poursuivre ma voie dans la recherche.

Et merci encore Benoît, pour toute l'aide apportée, pour tes corrections sur ce manuscrit et ta réactivité tellement appréciable en cas de coups durs. Je ne préfère pas imaginer ce qu'il serait advenu de la partie *in planta* si tu n'avais pas volé à notre secours... Merci pour tout !

Un immense merci à toi, cher directeur de thèse ! Merci TiMi pour tout ce que tu m'as apporté au cours de ces trois ans. Ton ouverture d'esprit, ton optimisme à toute épreuve, ton amour inconditionnel pour la science sous toutes ses formes, ta confiance absolue et ton honnêteté m'ont énormément appris. Merci pour toutes ces discussions, pour nos divagations diverses et variées, et pour m'avoir inculqué que "qui peut le plus peut le moins" ou encore que "seul, on est rien". Merci pour la confiance que tu as tout de suite su placer en moi. Merci de m'avoir autant conseillé et réconforté pendant les périodes de doutes et de remises en question qui, j'imagine, sont le propre de tous thésard. Plus qu'un encadrant, tu as été un mentor et je ne réaliserais sûrement jamais assez la chance que j'ai eu de pouvoir faire une thèse avec une personne telle que toi. Bref Merci Doc' !

Et Merci également à mes trois lady de la dreamteam de biochimie !!
Amandaïne, notre wonderwoman de purif, ton soutien sans failles (en témoigne ta conversion à Nova) a permis d'éviter environ un millier de pétages de plombs aigüs et je me demande sincèrement comment j'aurais pu encaisser tous mes déboires biochimiques sans toi ! Ton implication, ta complicité et ta joie de vivre ont rendu ces échecs mineurs et je ne te remercierais jamais assez pour ça. Merci aussi pour ton aide et ton acharnement si précieux lors de la purification de ces protéines capricieuses. Promis, la VPg du PVY ne sera bientôt plus qu'un très lointain souvenir pour toi ! Tu es une personne extraordinaire et j'ai eu une chance incroyable de pouvoir travailler avec toi pendant ces trois ans. Mais surtout souviens toi, le DOP ça se mange pas !

Merci à toi Joce, pour ton aide et ton soutien. Merci pour ta gentillesse, ton brin de folie et nos nombreux éclats de rires ! Merci pour ta confiance, et merci d'avoir toujours été là quand j'avais besoin d'être aiguillée.
Merci Geneviève, pour t'être toujours autant intéressée à ce que je faisais. Toutes nos discussions (scientifiques ou non d'ailleurs) m'ont énormément apporté, et je t'en suis vraiment très reconnaissante !
Et bien sûr un grand merci également à toi Laure (alias ma bobèche préférée) ! Nos journées passées au labo avec Amandine me manque tellement depuis que je suis cloué au bureau pour rédiger ! Merci pour ta bonne humeur communicative (je ne sais même pas à quoi tu ressembles quand tu ne souries pas) qui ne peut que nous faire voir la vie du bon côté (même le lundi matin, en dernière année = exploit !). Merci pour le coup de pouce des dernières manips (dieu sait que découper des feuilles c'est pas passionnant) et merci pour tous nos délires protrudiens !

Sans oublier un grand merci à Yihua et Joy, stagiaires exceptionnelles que nous avons eu la chance d'accueillir. Merci pour votre enthousiasme et votre rigueur qui ont contribué à faire avancer la cause du désordre !

Merci à toi, Groöt, d'être quelqu'un d'aussi exceptionnel ! Tu as tellement contribué à ce que venir au labo chaque matin à 8h ne soit jamais une corvée ! Je ne te trouve toujours pas de défauts, si ce n'est celui de ne pas partager tes sandwichs avec nous... et je te souhaite du fond du cœur toute la reconnaissance que tu mérites, aussi bien en tant qu'espoir majeur du tennis français qu'en tant que technicien prodige.
Merci à toi Pat, pour ta gentillesse, tes conseils, tes blagues et pour avoir fini par accepter ma manie de détraquer toutes les machines auxquelles je touche. J'espère que tu as bien profité de ces 4 mois de répit parce que je reviens maniper très vite ! Et bien sûr merci de pas m'avoir tenu rigueur de toutes les fanny que je t'ai infligé au bab.
Merci à Val pour tes conseils et tes encouragements. Merci à Mélo pour tes sourires, ton soutien et ta gentillesse permanente. Merci à Marie-Christine, notamment pour m'avoir guidé dans mes débuts au labo, et avoir facilité ma découverte de la localisation des kits et autres réactifs. Merci à Pascal pour nos échanges, débats et rires au coin café. Merci à Michel pour tes conseils précieux en cette fin de thèse, et bien sûr pour le recul que tu nous apprends à avoir à nous autres étudiants avec des questions dont toi seul a le secret. Merci à Thierry M pour ta gentillesse, ton aide et tes conseils. Merci à Coco pour ton implication, ta patience, nos délires et pour avoir aussi bien pris soin de nos petites plantes. Merci à Chantal, notre maman de labo à nous, étudiants ! Démarrer sur un nouveau sujet, dans un nouveau labo, devient un jeu d'enfants avec une Chantal Faure pour veiller sur vous ! Merci de tout cœur pour ta bienveillance et tes encouragements qui m'ont reboosté plus d'une fois. Merci à toi "grand" Vincent pour être venu à notre rescousse pour les manips en serre. Même si le succès n'a pas vraiment été au rendez-vous, grâce à elles et bien sûr grâce à toi, les piments et le PVY ont désormais moins de secrets pour moi ! Je te suis très reconnaissante pour tout ça et te remercie également pour tous ces bons moments passés ensemble.
Merci à toi Laurence pour toutes nos discussions, pour tes conseils avisés et ta bienveillance à mon égard.
Merci Armelle, déjà pour ton rire, contagieux même à distance, pour ton aide (on va se la faire un jour cette manip NGS !) et pour toutes ces choses que j'ai apprises grâce à toi. Egalement un grand merci Sylvie et Véro pour tous vos conseils, votre gentillesse, et pour avoir été aussi disponibles à chaque fois que j'ai passé la porte de votre bureau.

Merci Aurélie pour ta bonne humeur, ton entrain et ton aide lors de ma découverte du monde de l'ELISA. Merci Jean-Phi pour ta gentillesse et tes sourires journaliers. Merci Stéphane, compagnon de corvée ménage, pour les discussions, pour tes blagues et pour l'intérêt que tu as eu à mon égard. Merci Sandy, pour le coup de pouce en bioinfo et pour ta sympathie. Je te souhaite pleins de bonheur dans ta nouvelle vie lilloise !

Merci Laure et Alain pour m'avoir accompagné dans ma première expérience d'enseignement.

Un grand merci à toi Claudine pour ta bienveillance et ton aide précieuse dans toutes ces démarches administratives.
Merci à Flo, Isa, Laurence, Muriel et Marie-Lou pour votre bonne humeur et votre aide précieuse quand il a fallu organiser le congrès doctorants.

Un immense merci à mes collocs de boulot (promus au rang de collocs "de vie" lors de ces six dernier mois !) anciens et actuels : Maud, Sylvain, Julia, Julie, Vincent, Bernie, Xav, Cécile ainsi que les tout derniers arrivés, Yuxin et Shuo.
Vincent M, tu es un voisin idéal, merci de n'avoir pas craqué lorsque la vague de publis et autres protocoles ont sournoisement commencé à envahir ton espace vital (t'inquiètes, je range tout demain !). Merci Bernie, pour tes bavardages nocturnes aussoiens, pour tes encouragements et pour toutes nos discussions. Merci Kékile et Xav pour votre entrain, votre bonne humeur et ces bons moments de rire qui font tellement de bien. Bref merci à vous tous pour cette cohabitation plus qu'agréable et votre soutien psychologique permanent. Comme le veut la tradition, c'est désormais mon tour de passer le relais en vous disant (et en y croyant bien sûr) que vous êtes les meilleurs, que vous allez y arriver, que vous allez faire chacun une thèse extraordinaire (oui oui toi aussi Xavier !). Je vous souhaite de bien savourer chaque instant passé dans ce labo ! Et merci au pourvoyeur de nain de jardin anonyme (un sacré mystère cette histoire !).

Un énorme merci à mes quatre compères, collègues devenus amis en moins de temps qu'il n'en faut pour le dire, vous avez été des soutiens tellement précieux. Il y a des rencontres pour lesquelles on se dit que la vie réserve de bien belles surprises !
Merci à ma ritale préférée pour tous ces délires, pour tes mots rassurants, et aussi pour m'avoir élue digne héritière de ton ordi, pour tes tortellinis... et j'en passe !
Merci à toi ma blondasse, pour tous ces moments de joie, de peine, de détente et de fous rires partagés avec toi, merci pour ton amitié et pour tes protocoles tellement beaux que j'en pleurerais, pour ton écoute et pour tout ce que tu m'as apporté.
Merci à toi mon petit marseillais. Merci d'avoir fait une entorse à la règle en devenant ami avec une parisienne, merci pour tes blagues, pour nos débats enflammés sur des sujets toujours extrêmement sérieux, pour m'avoir emmené lors des démotivations de pédalage, et pour t'être autant inquiété de savoir comment j'allais tout au long de la thèse.
Merci à toi mon petit cantalou, pour toutes ces pauses revigorantes d'oxygénation des poumons, pour m'avoir fait découvrir le monde obscur de Linux, pour avoir eu pitié de moi lors d'égarements snowboardiens... Merci d'être également un digne ambassadeur du concept de "le désordre c'est la vie" et surtout merci pour ton amitié et pour être la personne que tu es, à savoir un gars trop génial !

Merci à toi ma Valie, mon roc, mon mentor, mon ange gardien, mon remède pour lutter contre ce satané "bad karma". Si je suis arrivée à bout de cette aventure sans trop d'encombres, je te le dois en grande partie. Je n'ose même pas imaginer ce que sera la vie au labo sans toi mais bon, c'est comme ça, il faut savoir partager les perles comme toi, il reste encore tant de gens qui n'ont pas encore eu la chance de croiser ton chemin. Notre passion commune pour les poules et tous ces fous rires ont été d'une aide inestimable pour traverser mes nombreuses péripéties et je t'en serais reconnaissante à vie et tu es sans nul doute la plus belle personne que j'ai jamais rencontré !
Un grand merci à mes compagnons de galère de thèse non-virologistes (*nobody is perfect*) : Lapinou (alias Kiwiman), Julien, Anto, Benoît, Chachou, Léa, Ccori, Tracey. Courage pour les très futurs docteurs: on est bientôt arrivé au bout du tunnel ! Pour les plus petits (ou les "redoublants"...) qui doivent encore rester un an dans le tunnel, je n'aurais qu'un seul conseil à vous adresser : toujours préparer vos com' en anglais ! Merci à vous tous pour ces moments de décompression intense, de fous rires et de solidarité. J'adresse bien sûr une mention spéciale à notre Rémi national. Merci au destin de m'avoir fait un jour croiser ta route. Un an après, on ne s'est toujours pas remis de ton départ. J'espère pour toi et pour nous tous que tu deviendras un jour ce grand chercheur en évènementiel que nous aimons tant ! Je vous souhaite en tout cas tout le bonheur du monde pour cette nouvelle aventure à 4 !

Merci à mes formidables encadrants de la première heure, David et Fatma, pour m'avoir si bien guidé lors de mes premiers pas et initié à la rigueur et aux joies de la recherche en labo.

Un merci tout particulier à M. CHARRON, pour vos cours et votre pédagogie qui m'ont fait découvrir, aimer et finalement épouser la biologie !

Un grand merci à tous mes amis de Bordeaux, de Paris et d'ailleurs pour votre soutien et pour avoir eu la délicatesse de me demander quasi à chaque fois si j'avais de nouveaux résultats, tout en assumant la réponse incompréhensible d'une demi-heure qui suit généralement ce genre de question. Merci pour tous vos encouragements et surtout merci d'être des gens aussi géniaux !

Je remercie également de tout mon cœur ma famille, mes parents adorés, mes frères, sœurs, demi-frères, demi-sœurs, mes extraordinaires mamies et papys parce que votre amour, votre soutien, votre intérêt et votre fierté sans failles m'ont toujours permis d'avancer de façon la plus sereine possible. Merci d'avoir fait de mes escapades parisiennes et basques des moments de réconfort et de bonheur pur. Cette expérience m'a permis de me rendre compte plus que jamais de la chance que j'ai de vous avoir. Je vous aime très fort.

Enfin merci à toi mon Dou, merci pour ton amour, ta tendresse et ta force tranquille. Essuyer les déboires d'une thésarde n'a pas dû être de tout repos et je te suis tellement reconnaissante de l'avoir si peu montré ! Merci de m'avoir chouchouté tout ce temps, merci de partager ma vie et de la rendre aussi belle. Je t'aime de tout mon cœur et je te dédie cette thèse.

Production scientifique

- **Publication**

Charon J., Theil S., Nicaise V., Michon T. (2016) Protein intrinsic disorder within the Potyvirus genus: from proteome-wide analysis to functional annotation. *Molecular BioSystems*, **12** (634).

- **Communications**

Charon J., Walter J., Longhi S., Moury B., Michon T. (2013) Intrinsic Disorder of viral proteins, mutational robustness and host adaptation using the potyviral VPg as a reporter. *14èmes Rencontres de Virologie Végétale.* Aussois, France, 13-17 janvier 2013. [Poster]

Charon J., Walter J., Longhi S., Moury B., Michon T. (2013) Le désordre intrinsèque des protéines virales : permissivité mutationnelle et adaptation à l'hôte. 5èmes *Journées des Doctorants du Département SPE.* Montpellier, France, 20-22 juin 2013. [Oral]

Charon, J., Walter J., Barra A., Hébrard E., Moury B., Michon T. (2014) Intrinsic disorder in viral proteins : A strategy in viral adaptation ? *14èmes Journée Scientifique de l'Ecole Doctorale Sciences de la Vie et de la Santé.* Arcachon, France, 9 avril 2014. [Poster]

Charon J., Walter J., Barra A., Hébrard E., Moury B., Michon T. (2014) An attempt to experimentally assess the contribution of intrinsic disorder to the virus adaptation. 6èmes *Journées des Doctorants du Département SPE.* Bordeaux, France, 03-05 juin 2014. [Oral]

Charon J., Theil S., Walter J., Longhi S., Hébrard E., Moury B., Michon T. (2014) An attempt to experimentally assess the contribution of intrinsic disorder to the virus adaptation. *Intrinsically disordered proteins - Gordon Research Seminar.* Boston (MA) USA, 05-06 juillet 2014. [Poster]

Charon J., Theil S., Walter J., Barra A., Longhi S., Moury B., Michon T. (2014) An attempt to experimentally assess the contribution of intrinsic disorder to the virus adaptation. *Intrinsically disordered proteins - Gordon Research Conference*. Boston (MA) USA, 06-11 juillet 2014. [Poster]

Justine Charon (2014) Le désordre intrinsèque des protéines : clé de l'adaptation virale ? – *Conférence sur invitation du Réseau Hôte-Pathogène de la SFR-Tansbiomed* Bordeaux, France, Novembre 2014. [Oral]

Charon J., Theil S., Walter J., Barra A., Hébrard E., Moury B., Michon T. (2015) Characterization of intrinsic disorder involvement in potyvirus adaptation using PVY/Pepper pathosystem as model. *15èmes Rencontres de Virologie Végétale*. Aussois, France, 18-22 Janvier 2015. [Oral]

Charon J., Moury B., Walter J., Barra A., Hébrard E., Michon T. (2015) Characterization of protein intrinsic disorder involvement in RNA virus adaptation using a plant virus model. *15ème Journée scientifique de l'Ecole Doctorale Sciences de la Vie et de la Santé.* Arcachon, France, 16 Avril 2015. [Oral]

Charon J., Moury B., Walter J., Barra A., Hébrard E., Michon T. (2015) Characterization of protein intrinsic disorder involvement in RNA virus adaptation using a plant virus model. *7èmes Journées des Doctorants du Département SPE.* Rennes, France, 01-03 juillet 2015. [Oral]

Formations et activités transversales

Activité transversale complémentaire

- **Activité de Monitorat à l'Université de Bordeaux**

UFR Sciences de la Vie - Spécialité d'enseignement : Microbiologie
Durée : 10h de Travaux Dirigés (Master) + 54h de Travaux Pratiques (Master et Licence)
Période : Septembre 2013 – Juin 2014

Formations suivies

- **Formation "De l'enseignement par l'enseignement"**

Séances théoriques - Formateurs : Mme Véronique LAPOSTOLLE et M. André QUINTON
Durée : 12h
Séances Pratiques - Formateur: M. David LECOURIEUX
Durée : 24h

- **Formation "Manager et conduire un projet"**

Formateur : Afpi Sud-Ouest
Durée : 3jrs
Période : 28-29/03/2013 et 11/04/2013

- **Formation initiale au Sauvetage et secourisme du travail**

Formateur : M. Jean-Pierre PETIT
Durée : 25h
Période : du 23/09/2013 au 25/09/2013

- **Formation MISTER 3 : "Eléments pour une stratégie de publication pour communiquer les résultats de ses travaux de recherche"**

Formatrice : M. Marie-Hélène BRIDET
Durée : 8h
Période : 06/12/2013

- **Formation de sensibilisation aux risques professionnels**

Formateurs : M. Gilles DELATTRE, Mme Zohra HACHEM et M. Jean-Pierre PETIT
Durée : 7h
Période : 22/03/2013

Autres activités conduites au cours de la thèse :

- Encadrement d'un stagiaire de niveau Master 1 (Durée 2 mois)
- Co-encadrement d'une stagiaire de niveau Master 2 (Durée 6 mois)
- Organisation de la 6ème édition du congrès des doctorants du département Santé des Plantes et Environnement
- Participation active à la mise en place et l'organisation d'ateliers dédiés au personnel non-permanent, au sein de l'unité Biologie du Fruit et Pathologie.

Listes des virus

Acronyme	Espèce	Genre	Famille
ACMHV-2	*Avian carcinoma virus Mill Hill 2*	*Alpharetrovirus*	*Retroviridae*
AgMV	*Agropyron mosaic virus*	*Rymovirus*	*Potyviridae*
BYMV	*Bean Yellow mosaic virus*	*Potyvirus*	*Potyviridae*
DENV	*Dengue virus*	*Flavivirus*	*Flaviviridae*
HAV	*Hepatite A virus*	*Hepatovirus*	*Picornaviridae*
HCoV-NL63	*Human corona virus NL63*	*Alphacorona virus*	*Coronaviridae*
HCV	*Hepatite C virus*	*Hepacivirus*	*Flaviviridae*
HDV	*Hepatite D virus*	*Deltavirus*	*Unassigned*
HIV	*Human Immundeficiency virus*	*Lentivirus*	*Retroviridae*
HoMV	*Hordeum mosaic virus*	*Rymovirus*	*Potyviridae*
HPV	*Human papilloma virus*	*Alphapapillomavirus*	*Papillomaviridae*
JEV	*Japanese encephalitis virus*	*Flavivirus*	*Flaviviridae*
LMV	*Lettuce mosaic virus*	*Potyvirus*	*Potyviridae*
MeV	*Measle virus*	*Morbillivirus*	*Paraxymoviridae*
n.d	*MS2 RNA bacteriophage*	*Levivirus*	*Leviviridae*
n.d	*Pandoravirus salinus*	Not yet classified	Not yet classified
NoV	*Nodamura virus*	*Alphanodavirus*	*Nodaviridae*
PCV	*Porcine circovirus*	*Circovirus*	*Circoviridae*
PPV	*Plum pox virus*	*Potyvirus*	*Potyviridae*
PRSV	*Papaya ringspot virus*	*Potyvirus*	*Potyviridae*
PSbMV	*Pea seed-borne mosaic virus*	*Potyvirus*	*Potyviridae*
PVA	*Potato virus A*	*Potyvirus*	*Potyviridae*
PVY	*Potato virus Y*	*Potyvirus*	*Potyviridae*
SARS-CoV	*Severe acute respiratory syndrome-related Coronavirus*	*Betacoronavirus*	*Coronaviridae*
SCMV	*Sugarcane mosaic virus*	*Potyvirus*	*Potyviridae*
SMV	*Soybean mosaic virus*	*Potyvirus*	*Potyviridae*
TEV	*Tobacco etch virus*	*Potyvirus*	*Potyviridae*
TBSV	*Tomato bushy virus*	*Tombusvirus*	*Tombusviridae*
TuMV	*Turnip mosaic virus*	*Potyvirus*	*Potyviridae*
TVMV	*Tobacco vein mottling virus*	*Potyvirus*	*Potyviridae*
TYMV	*Turnip yellow mosaic virus*	*Tymovirus*	*Tymoviridae*
WMV	*Watermelon mosaic virus*	*Potyvirus*	*Potyviridae*
WNV	*West nile virus*	*Flavivirus*	*Flaviviridae*
YFV	*Yellow fever virus*	*Flavivirus*	*Flaviviridae*
ZYMV	*Zucchini yellow mosaic virus*	*Potyvirus*	*Potyviridae*

Liste des abréviations

°C	degré Celcius
µ	micro (10^{-6})
A	Adénine
aa	Acide aminé
AAP	Amino acid polymorphism
ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
APS	Aperoxide persulfate
ARN	Acide ribonucléique
ARNm	ARN messager
BET	Bromure d'éthidium
BLOSUM	Block substitution matrix
c	centi (10^{-2})
C	Cytosine
CA	Capsicum annuum
$CaCl_2$	Chlorure de calcium
CD	Circular dichroism
CI	Cylindrical inclusion protein
CP	Coat protein
C-ter	C-terminal
cv	cultivar
Da	Dalton
DIECA	Diethyl Dithiocaremate de Sodium
DisProt	Database of protein disorder
dNTP	Désoxy nucléotides tri-phosphate
DO	Densité optique
dpi	jours post-inoculation
DUB	Deubiquitinating enzyme
EDTA	Ethylenediamine tetraacetique acid
eF1A	eukaryotic elongation factor 1A
eIF4E	eukaryotic translation initiation factor 4E
ELM	Eukaryotic linear motif
ER	Endoplasmic reticulum
EtOH	Ethanol
FHA	Forkhead-associated domains
Fw	Forward
g	gramme
G	Guanine
HCl	Chlorure d'hydrogène
HC-Pro	Helper component protease
HSP	Heat shock protein

ICTV	International committee on taxonomy of viruses
IDP	Protéine intrinsèquement désordonnée
IDR	Région intrinsèquement désordonnée
IMAC	Immobilized ion affinity chromatography
k	kilo (10^3)
kb	kilo base
l	litre
LB	Luria bertani
LiAc	Acétate de lithium
LOB	Loading buffer
M	Molaire
m	Milli (10^{-3})
$MgCl_2$	Chlorure de magnésium
min	minute
MOI	Multiplicity of infection
mol	mole
MoRF	Molecular recognition feature
NaCl	Chlorure de sodium
NaF	Fluoride de sodium
NaOH	Hydroxyde de sodium
NCBI	National center of biotechnological information
Ne	Taille effective de population
NGS	Next generation sequencing
NIa	Protéine d'inclusion nucléaire a
NIb	Protéine d'inclusion nucléaire b
Ni-NTA	Nickel-Nitrilotriacetic acid
NLS	Nuclear localization signal
NES	Nuclear export signal
nm	nano (10^{-9})
nm	nanomètre
NoLS	Nucleolar localization signal
NS	Non Synonyme
NT	Nicotiana tabacum
nt	nucléotides
N-ter	N-terminal
OD	Région ordonnée
ORF	Open reading frame
PABP	Poly(A)-binding protein
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCaP1	Plasma membrane-associated cation binding protein 1
PCR	Polymerase chain reaction
PDB	Protein databank
pH	potentiel hydrogène

PIPO	Pretty interesting protein
PMSF	Phenylmethylsulfonyl fluoride
PONDR	Predictor of naturel disordered regions
PTM	Post-translational modification
PVP	Polyvinyl pyrrolidone
RdRP	RNA dependant RNA polymerase
Rev	Reverse
RH8	RNA helicase 8
RMN	Résonance magnétique nucléaire
rpm	rotation par minute
S	Synonyme
SCE1	SUMO-conjugating enzyme 1
SDS	Sodium dodécyl sulfate
sec	secondes
SOC	Super Optimal broth with Catabolite repression
ssRNA	ARN simple brin
T	Thymine
Ta	Température d'hybridation
USP7	Ubiquitin Specific Protease 7
UTR	Untranslated region
UV	Ultra-violet
VPg	Viral genome-linked protein
vs	versus
w/v	Ratio masse sur volume
WT	Wild type
YPD	Yeast extract Peptone Dextrose

A Adrien,

Résumé

Les protéines sont des acteurs majeurs dans les processus moléculaires et cellulaires d'un organisme. La remise en question des modalités associées aux fonctions de ces macromolécules a récemment été apportée par le concept de désordre intrinsèque. Celui-ci définit l'absence (transitoire ou permanente) de structure tridimensionnelle de certaines protéines ou régions protéiques comme étant directement liée à leurs fonctions. Chez les virus à ARN, les propriétés des protéines ou régions désordonnées semblent associées aux capacités de ces micro-organismes à détourner la machinerie cellulaire de l'hôte en interagissant avec de multiples partenaires, et à s'adapter aux nombreuses contraintes auxquelles ils doivent faire face en tant que parasites obligatoires. Ce travail porte sur les potyvirus, figurant parmi les pathogènes de plantes les plus dommageables étudiés à ce jour. **L'objectif de cette thèse a été d'explorer les fonctions associées au désordre intrinsèque dans le cycle infectieux des potyvirus ainsi que dans le processus d'adaptation.**

Notre approche a ainsi démontré que : i) le désordre est ubiquitaire chez le genre *Potyvirus* ; ii) les régions de désordre conservées chez plusieurs protéines de potyvirus semblent être associées à leur(s) fonction(s) pendant l'infection ; iii) les régions désordonnées sont généralement associées à moins de contraintes évolutives, suggérant ainsi leur implication dans les processus adaptatifs des potyvirus ; iv) les régions prédites comme désordonnées semblent privilégier l'apparition de mutations et donc la capacité d'un virus à accumuler de la diversité génétique au cours de l'évolution sur son hôte naturel ; v) ce travail a permis de corréler le taux en désordre de la protéine viral genome-linked (VPg) du *Potato virus Y* à sa capacité à s'adapter à la résistance récessive *pvr2*^{3} du piment.

Mots-clés : *Désordre intrinsèque des protéines, Potyvirus, virus à ARN, adaptation, Potato virus Y*

Abstract

Proteins are essential actors involved in a majority of molecular and cellular processes. The features associated with the functions of these macromolecules have been recently questioned with the emergence of the intrinsic disorder concept. It defines the transitory or permanent lack of 3D structure in some proteins or regions as directly related to their functions. Among RNA viruses, the properties of disordered proteins may be linked to the ability of these microorganisms to hijack the host machinery by interacting with multiple partners, as well as to adapt to the multiple constraints they must face as obligatory parasites. This work focuses on the *Potyvirus* genus, which includes some of the most damaging plant pathogens studied to date. **The goal of this thesis was to explore the functions associated with intrinsic disorder in the infectious cycle of this viral genus as well as in its process of adaptation.**

Our studies have shown that i) intrinsic disorder is ubiquitous in potyviruses; ii) intrinsically disordered regions (IDR) of some of potyviral proteins are likely to be associated with important functions for the viral cycle ; iii) IDR are generally less evolutionary constrained, suggesting an adaptive potential of these regions ; iv) predicted IDR seem to favor the appearance of mutations and therefore virus ability to accumulate genetic diversity during its evolution in natural host ; v) an experimental disorder modulation within the Viral genome-linked (VPg) protein has been demonstrated as positively correlated with the adaptive ability of the *Potato virus Y* to overcome the *pvr2*^{3} recessive resistance in pepper.

Keywords : *Protein intrinsic disorder, Potyvirus, RNA viruses, adaptation, Potato virus Y*

Sommaire

INTRODUCTION 1

1. Le désordre intrinsèque : concept et implications dans l'étude des protéines 3

A) L'extension du dogme structure-fonction 3

B) L'étude du désordre : qu'en est-il quinze ans après l'émergence du concept ? 5

2. Les fonctions associées au désordre chez les virus à ARN 17

A) Les virus présentent une variabilité record de désordre dans leur protéomes..... 17

B) Les virus à ARN : des agents aux propriétés très particulières 19

C) La flexibilité structurale au service de la multi-interaction chez les virus 21

D) Le désordre intrinsèque, un moteur de l'adaptation virale ? 23

3. Objectifs de la thèse..... 37

4. Présentation du modèle expérimental 39

A) Les potyvirus : un genre majeur chez les virus de plantes 39

B) Les potyvirus : un modèle expérimental de choix pour l'étude des mécanismes adaptatifs viraux..... 41

CHAPITRE I. Analyse *in silico* du désordre intrinsèque et des fonctions qui lui sont associées au sein du genre *Potyvirus*..... 43

Contexte 47

Article *Charon et al (2016), Molecular BioSystems, 12-634*..... 49

Supplementary data 67

CHAPITRE II. Etude de l'implication du désordre dans le maintien de la variabilité génétique d'une population virale *via* l'analyse de données de séquençage haut débit..... 75

1. Contexte de l'étude 77

2. Résultats 79

A) Les protéines des potyvirus n'ont pas toutes la même propension à accumuler des mutations à l'échelle de l'évolution intra-hôte 79

B) Le désordre est-il associé à des changements en acides aminés plus déstabilisants/contraignants pour la structure de la protéine ?..... 85

3. Discussions, Conclusions et Perspectives..... 87

CHAPITRE III. Rôle du désordre de la région centrale de la protéine VPg dans l'adaptation du PVY à la résistance *pvr2*^{3} médiée par eIF4E chez l'hôte *Capsicum annuum* 123

1. Contexte de l'étude 125

A) L'adaptation des virus de plantes aux résistances récessives variétales. 125

B) Cas du contournement, via la protéine VPg, des résistances médiées par eIF4E chez les Potyvirus 127

2. Résultats 133

A) Construction de mutants de PVY présentant des modulations dans le contenu en désordre de la zone centrale de leur VPg 133

B) Validation expérimentale du changement de désordre introduit dans les 3 mutants de VPg 135

C) Test de fitness des mutants SON41p ID1, ID2 et OD2 sur piment sensible Yolo Wonder, porteur de l'allèle pvr2^{+} 137

D) Estimation de la capacité de contournement de résistance des différents mutants de « désordre » de la VPg 139

E) Détermination des évènements mutationnels intervenus au cours du contournement 139

F) Détermination des évènements moléculaires associés au contournement: analyse de l'interaction VPg/eIF4E 141

3. Discussion et Perspectives 145

A) Cas du mutant ID1 : le désordre au service de la multi-interaction ? 145

B) Cas des mutants ID2 et OD2 : le désordre au service de l'adaptation ? 147

C) Considérations d'ordre plus générales sur l'interprétation des données obtenues 129

CONCLUSIONS ET PERESPECTIVES DE LA THESE 157

L'intégration du désordre dans les processus d'annotation fonctionnelle des protéines 135

Le désordre comme déterminant moléculaire de l'adaptation virale : vers un moyen d'appréhender les capacités adaptatives des virus à ARN ? 163

Conclusion générale 175

Le désordre intrinsèque : un concept à l'épreuve des faits 177

Ordre et désordre occupent-ils des espaces fonctionnels de différente nature ? 177

MATERIEL ET METHODES 159

1. Matériel biologique 183

A) Matériel viral 183

B) Matériel végétal 183

2. Procédures expérimentales 183

A) Stratégies et approches expérimentales 183

a. Obtention des vecteurs porteurs des mutants VPg ID1, ID2 et OD2 183

b. Obtention des vecteurs codant pour les formes $pvr2^{+}$ et $pvr2^{3}$ de eIF4E 185

c. Production de protéines recombinantes pour les mesures physico-chimiques *in vitro*..... 187

d. Purification des protéines recombinantes..... 187

e. Dichroïsme circulaire et protéines désordonnées..... 189

f. Détermination des constantes thermodynamiques de dissociation du complexe VPg/eIF4E par spectroscopie de fluorescence 189

g. Suivi d'infection des piments Yolo Wonder et HD285 par les PVY SON41p WT et porteurs des VPg mutantes 193

h. Séquençage des génomes viraux 199

B) Modes opératoires 199

a. Techniques de biologie moléculaire 199

b. Techniques de biochimie..... 197

c. Test de détection virale par DAS-ELISA 205

d. Méthodes d'inoculation virale 229

REFERENCES BIBLIOGRAPHIQUES

ANNEXES

Introduction

1. Le désordre intrinsèque : concept et implications dans l'étude des protéines

A) L'extension du dogme structure-fonction

En 1894, Emil Fisher établissait un dogme devenu ensuite central dans l'étude des protéines : les enzymes interagissent avec leur substrat via une complémentarité de structures à leur surface, selon le modèle clé-serrure. De nombreuses évidences expérimentales ont progressivement appuyé cette notion de structure nécessaire à la fonction (découverte des protéines chaperonnes et de la mise en évidence de l'absence de fonctionnalité des protéines dans leur état dénaturé). Bien plus tard, le chimiste et biochimiste américain Christian Anfinsen complétait ce dogme, en établissant que la structure tridimensionnelle native d'une protéine, dans un environnement donné, est unique, stable et déterminée par sa séquence en acides aminés (Anfinsen 1973).

Cette vision de la relation « séquence → structure → fonction » des protéines a été, jusqu'à très récemment, considérée comme la vision commune et unique. Si bien que les cas recensant des protéines non repliées mais fonctionnelles (par exemple la caséine qui contient une région déstructurée) étaient considérés comme rares (l'exception qui confirme la règle), ou perçus comme des artefacts d'expérimentations (Receveur-Bréchot & Karlin 2005).

Ce n'est qu'à la fin des années 1990 que des chercheurs ont commencé à considérer ces exceptions et « artefacts ».
Par exemple, l'observation faite que le facteur p21, régulateur du cycle cellulaire chez les eucaryotes, présente une absence de structure dans son état libre, malgré sa capacité à assurer de multiples interactions, a stimulé l'intérêt de certains scientifiques (Kriwacki et al. 1996). C'est ainsi que les pionniers du concept de **désordre fonctionnel** se sont penchés de façon très approfondie sur l'ensemble des « artefacts » de cristallisation disponibles dans la littérature. En compilant et étudiant de plus près les protéines dont la structure n'est pas disponible dans la *Protein DataBank* (PDB), certaines caractéristiques communes à ces protéines dépourvues de structure ont pu être déterminées. Elles présentent une faible complexité de séquence, et un biais important dans leurs compositions en acides aminés par rapport aux protéines structurées. Elles possèdent **une forte proportion de résidus à propriétés hydrophiles et chargés**, et à l'inverse semblent **pauvres en résidus hydrophobes**, promoteurs de structuration (Dunker et al. 1998; Romero et al. 2001). La pertinence fonctionnelle et l'abondance qui semblent sous-jacentes à l'absence de structure ont donc souligné la nécessité d'élargir le dogme central établi par Anfinsen en 1973. Le concept de **désordre intrinsèque des protéines** était né (Wright & Dyson 1999; Uversky et al. 2000; Dunker et al. 2001).

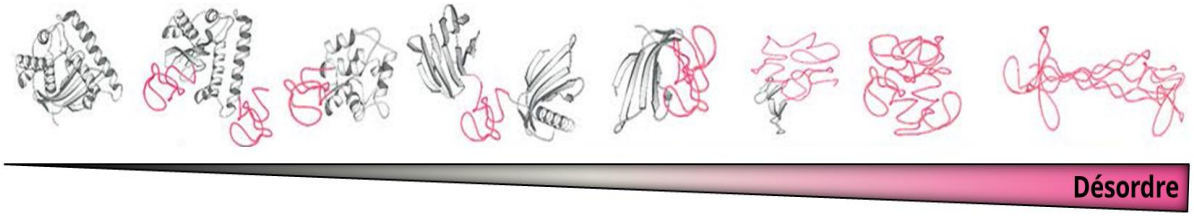


***Figure 1*|Le continuum de désordre intrinsèque à l'intérieur des protéines.** Présentation des intermédiaires possibles existants entre les protéines totalement dépourvues de désordre (à gauche) et totalement désordonnées (à droite). Le désordre peut être présent à différents taux dans une protéine (de gauche à droite) : totalement absent chez une protéine ordonnée (structurée), protéines dont seules les extrémités N- et C-terminales sont désordonnées, protéines avec un petit segment désordonné servant de lien flexible entre deux domaines bien structurés, protéine possédant une grande boucle désordonnée, protéine contenant un domaine désordonné, protéine désordonnée contenant une structure résiduelle, protéine totalement désordonnée collapsée, protéine totalement désordonnée étendue. Les régions désordonnées sont illustrées en rose sur la figure. Adapté de (Habchi et al. 2014; Uversky et al. 2005).

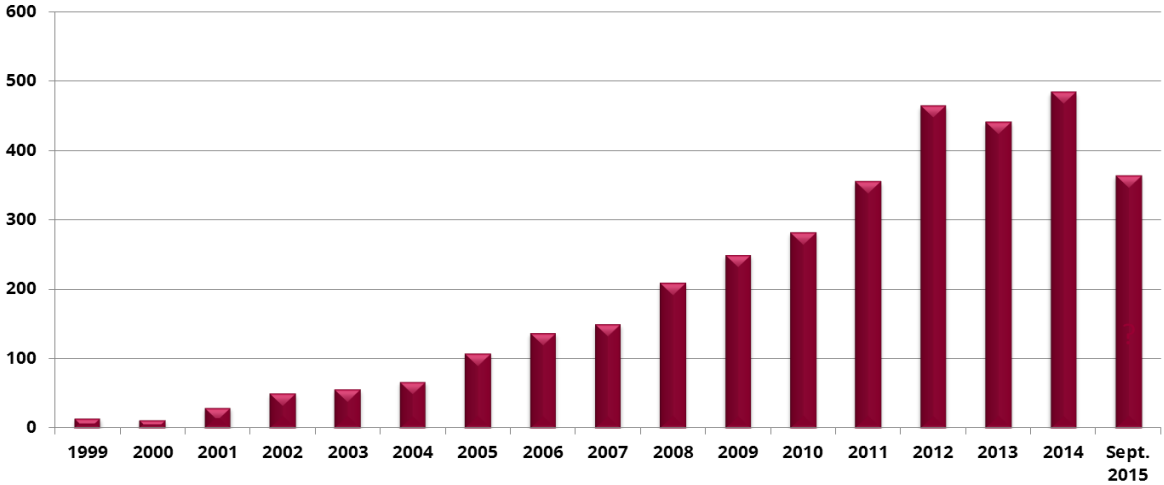


***Figure 2*|Evolution du nombre de publications portant sur le désordre intrinsèque des protéines depuis la naissance du concept en 1999 jusqu'à aujourd'hui.** Le nombre de publications est déterminé via PubMed (http://www.ncbi.nlm.nih.gov/pubmed) en utilisant les paramètres de recherche suivants : "(intrinsically OR natively OR naturally) AND (disordered OR unfolded OR unstructured) AND (protein OR region OR domain)".

B) L'étude du désordre : qu'en est-il quinze ans après l'émergence du concept ?

Le désordre intrinsèque des protéines se définit **par l'absence de structure tridimensionnelle (3D) en conditions physiologiques standards et en l'absence de partenaires associés.** Le désordre intrinsèque peut être plus ou moins local et se définit soit à l'échelle d'une protéine entière, on parlera alors de protéine intrinsèquement désordonnée (*Intrinsically disordered protein*, IDP), ou bien seulement à l'échelle d'une région protéique, on parlera alors de région intrinsèquement désordonnée (*Intrinsically disordered region*, IDR) (Figure 1).

Depuis l'émergence du concept, le désordre intrinsèque des protéines fait l'objet d'un nombre grandissant d'études (Figure 2). Les raisons de cet intérêt accru de la communauté scientifique s'expliquent notamment par les fonctions de plus en plus nombreuses qui lui sont associées, et la mise à disposition d'outils de plus en plus performants permettant de le caractériser.

a. La caractérisation expérimentale des protéines désordonnées

Comme évoqué précédemment, les protéines ou régions désordonnées présentent des propriétés physico-chimiques et dynamiques caractéristiques, pouvant être exploitées lors de leur caractérisation expérimentale.

De façon indissociable, l'émergence et l'acceptation du concept de désordre dans la communauté scientifique se sont accompagnées de progrès importants dans le domaine expérimental, notamment en **Résonance Magnétique Nucléaire** (RMN) qui permet aujourd'hui une analyse dynamique des populations de conformères des IDP (impossible par la technique de cristallisation aux rayons X). Le développement de la RMN *in vivo* (Bodart et al. 2008; Oldfield & Dunker 2014) a constitué une avancée importante en ce qui concerne l'acceptation par les plus sceptiques du concept de désordre intrinsèque, en permettant l'observation d'un état désordonné des protéines dans la cellule, s'affranchissant de biais liés aux expérimentations *in vitro.* Même si elle constitue un outil extrêmement puissant, la RMN reste néanmoins une technique lourde à mettre en place et ne constitue donc pas l'approche la plus adaptée dans le cas d'études préliminaires du désordre d'une protéine. Les propriétés biophysiques sous-jacentes à l'état désordonné d'une protéine peuvent être appréhendées par plusieurs techniques de détection *in vitro,* dont la mise en œuvre est plus accessible que la RMN (Uversky & Dunker 2012).

Parmi ces techniques, on peut notamment citer la **protéolyse ménagée**. Les zones désordonnées sont caractérisées par une faible densité de liaison entre les résidus, et possèdent de façon générale les attributs physico-chimiques associés aux résidus localisés en surface des protéines. Lors d'une digestion par une protéase, les résidus positionnés en surface sont en théorie plus accessibles que ceux enfouis à l'intérieur de la protéine. Dans ce contexte, le suivi de la digestion protéolytique peut permettre d'identifier les régions préférentiellement digérées et donc accessibles en surface.

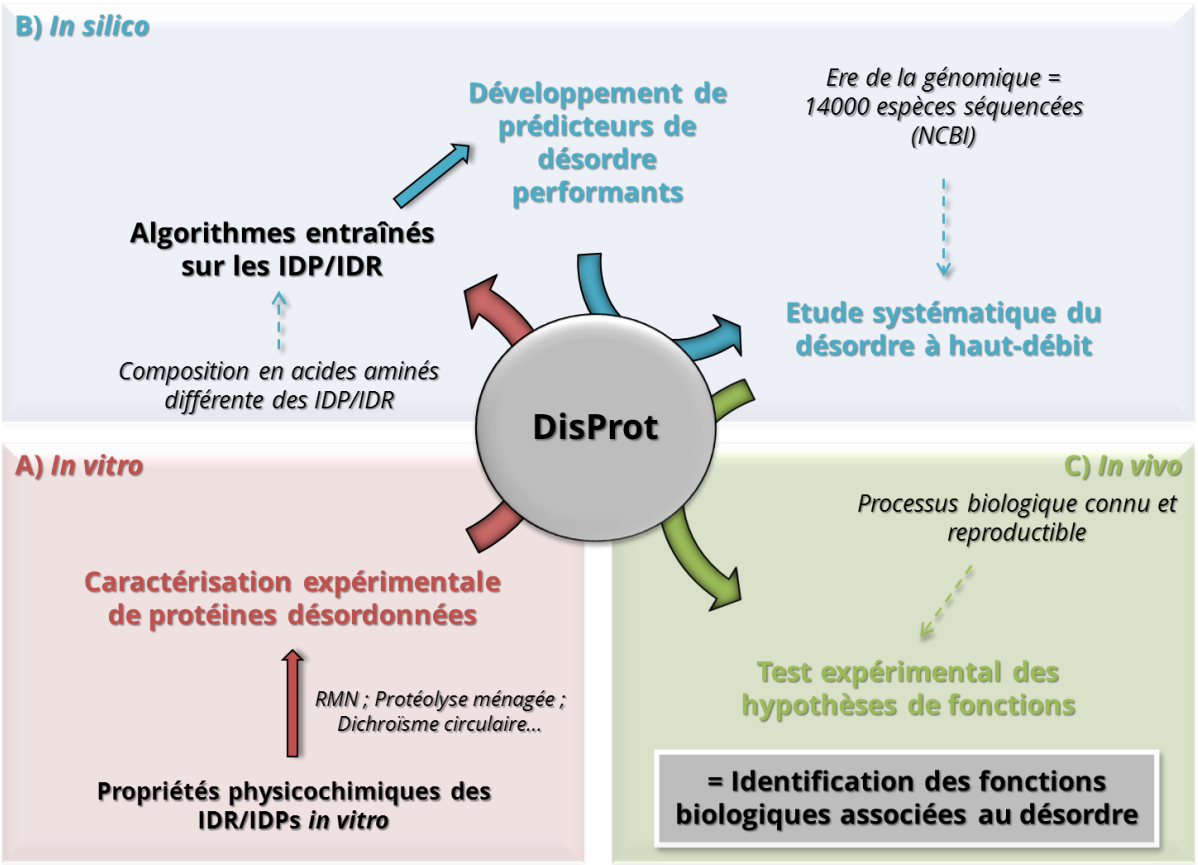


***Figure 3*|Illustration de la transdisciplinarité associée à l'étude du désordre.** La banque de données DisProt compile les 690 IDP et 1500 IDR expérimentalement démontrées comme désordonnées ainsi que leurs fonctions. **A)** Elle est alimentée par les données expérimentales obtenues *in vitro*. **B)** Cette compilation permet d'alimenter et d'entraîner les algorithmes pour développer les prédicteurs de désordre *in silico.* Ces mêmes prédicteurs peuvent permettre l'estimation à haut débit du désordre dans les protéomes des quelques 14000 espèces dont la séquence génomique est désormais disponible sur le NCBI (http://www.ncbi.nlm.nih.gov). Les données systématiques obtenues à partir de ces prédictions haut débit ainsi que la caractérisation fonctionnelle des protéines déterminées comme désordonnées permettent de dresser des hypothèses quant à la fonction du désordre dans les grands processus biologiques. **C)** L'objectif ultime consiste à replacer le désordre dans un contexte biologique afin de valider ou non les hypothèses avancées par l'étude systématique.

Ainsi les protéines intrinsèquement désordonnées présentent un profil de digestion constitué de nombreuses espèces moléculaires alors qu'une protéine très structurée demeurera essentiellement intacte, donnant naissance à très peu de peptides protéolytiques (Johnson et al. 2012).

Le dichroïsme circulaire constitue également une technique relativement facile à utiliser et présente l'avantage d'être beaucoup plus précise et fine que la protéolyse ménagée dans l'analyse du désordre d'une protéine. Largement utilisée en biochimie structurale, cette technique permet de quantifier la contribution des différents types de structures secondaires (hélice α, feuillets et brins β, boucles et segments désordonnés). L'état désordonné possédant un signal spécifique en dichroïsme circulaire, il est possible de déterminer la proportion de résidus désordonnés par cette technique.

Enfin l'analyse des propriétés hydrodynamiques d'une protéine révèle des informations sur sa forme et sa densité moléculaire, des paramètres discriminants d'ordre-désordre. A cet effet, une **chromatographie d'exclusion-diffusion** bien étalonnée s'avère souvent suffisante pour conclure quant aux caractéristiques structurales générales d'une protéine (Uversky 2002).

De par leur facilité de mise en œuvre, ces techniques de détermination expérimentale du désordre peuvent ainsi constituer des étapes préliminaires, et être conduites à haut débit, dans la démarche de caractérisation expérimentale des IDP (Johnson et al. 2012; Uversky & Dunker 2012) (Fig 3A).

b. L'analyse systématique des grandes fonctions biologiques associées au désordre intrinsèque

La banque de données DisProt, qui recense les protéines expérimentalement déterminées comme désordonnées, compte aujourd'hui plus de 690 protéines et 1500 régions, tout organismes confondus (Sickmeier et al. 2007) (http://www.disprot.org). Cette compilation des IDP et IDR caractérisées expérimentalement constitue une source importante et centrale pour l'étude systématique des fonctions biologiques associées au désordre (Figure 3).

Parmi les IDP et IDR étudiées, 39 grandes fonctions moléculaires, cellulaires et biologiques ont été associées à l'absence de structure dans ces protéines (Figure 4) et peuvent être réparties en 6 grandes classes fonctionnelles : **les connecteurs flexibles, les sites exposés, les chaperonnes, les effecteurs, les capteurs et les assembleurs** (Uversky & Dunker 2012) (Figure 5). Une même protéine peut présenter plusieurs régions désordonnées porteuses chacune de fonctions distinctes (Uversky 2013). L'éventail fonctionnel du désordre résulte d'abord des propriétés de ces régions. En effet, leurs longueurs, leurs positions dans la protéine ainsi que leurs capacités à adopter ou non une conformation en présence de partenaires (notion de fonctions associées à l'état flexible désordonné ou à l'inter-conversion désordre-ordre) sont autant de propriétés qui peuvent être reliées à ces grandes fonctions.

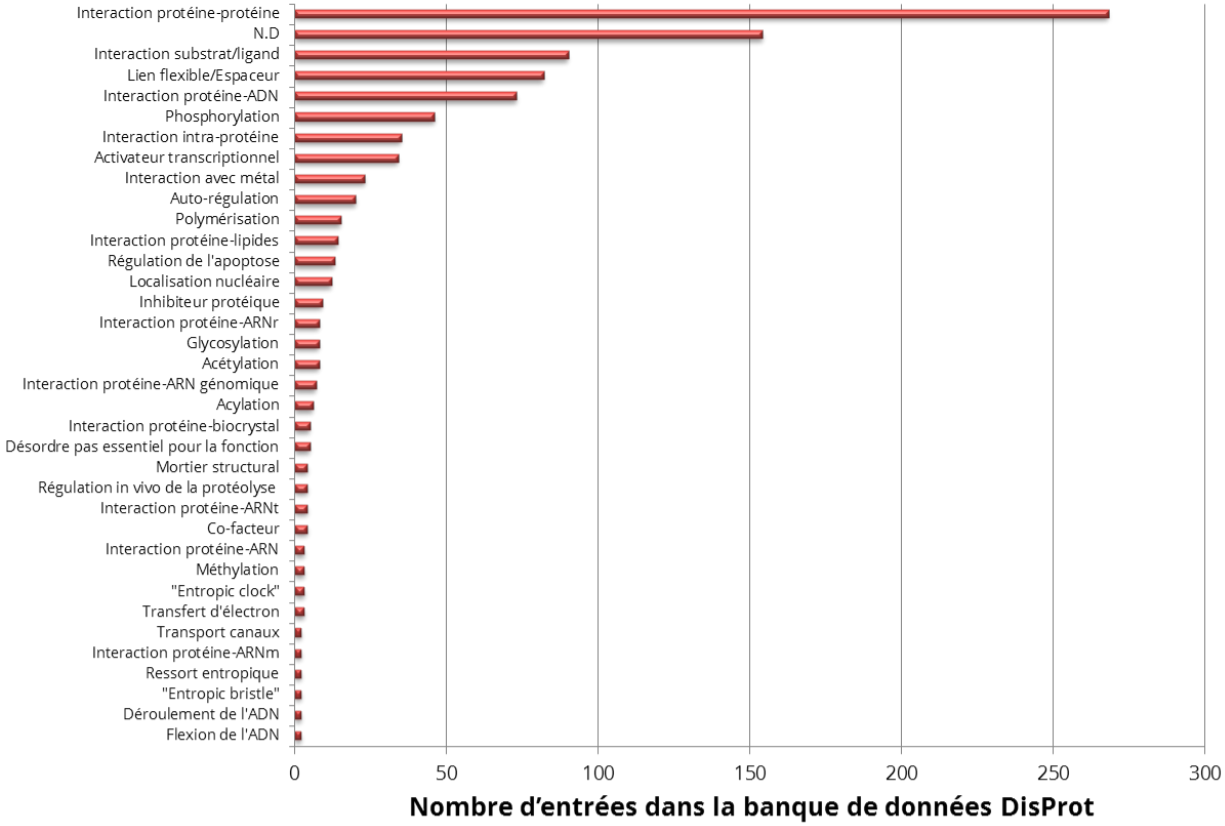


***Figure 4*|Fonctions expérimentalement associées au désordre chez les IDP et IDR désordonnées répertoriées dans la base de données DisProt**. (Source : http://disprot.org) ; *ND* : non déterminé.

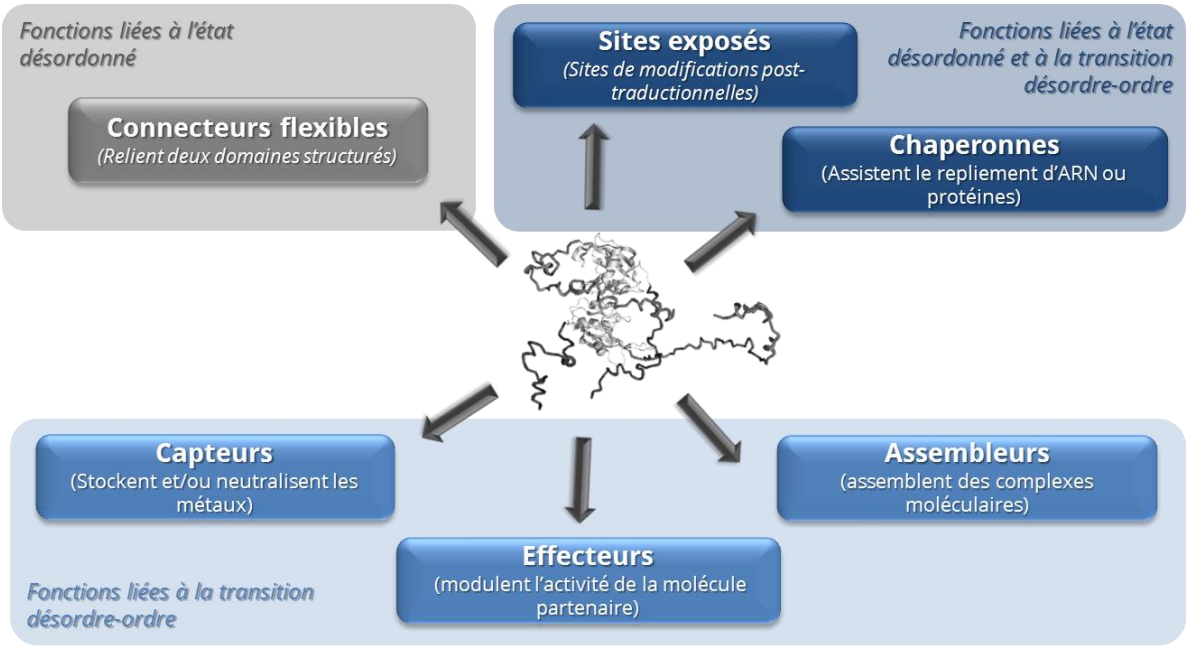


***Figure 5*|Grandes classes fonctionnelles associées aux protéines ou régions désordonnées.**

Les connecteurs flexibles

La fonction de **chaîne entropique** ou encore **connecteurs flexibles** constitue la fonction la plus « passive » du désordre intrinsèque dans les protéines et permet, dans le cas d'une protéine partiellement ordonnée, de lier plusieurs domaines structurés entre eux. L'étude de la protéine eucaryote RPA70 impliquée dans la réplication, la recombinaison et la réparation de l'ADN constitue un exemple de cette fonction. En effet, RPA70 contient une région décrite expérimentalement comme désordonnée, dont la flexibilité est conservée malgré une divergence importante de sa séquence en acides aminés chez les eucaryotes. Le désordre de cette région permet le rapprochement des domaines ordonnés situés de part et d'autre sur la chaîne polypeptidique (Daughdrill et al. 2007). Cet exemple illustre bien le rôle que peut avoir la flexibilité structurale dans la régulation des interactions intra-protéiques entre plusieurs domaines structurés.

Les sites exposés

La composition physico-chimique des IDR, riches en résidus polaires et chargés, favorise l'exposition des chaînes polypeptidiques à la surface des protéines, ainsi que les éventuels motifs et déterminants de reconnaissance moléculaire qu'elles contiennent. Cette exposition ainsi que la flexibilité conformationnelle associée à ces régions peuvent permettre une interaction transitoire rapide avec des enzymes de régulation post-traductionnelle et des protéases ou encore un accès et une reconnaissance facilité des protéines effectrices. Ces modifications ou clivages enzymatiques de surface sont souvent associés à de nombreux processus de signalisation et de régulation. La forte proportion de sites de modifications post-traductionnelles et notamment de phosphorylation dans les régions désordonnées suggère donc de façon générale une implication du désordre dans les processus de régulations cellulaires (Iakoucheva et al. 2004; Collins et al. 2008; Diella et al. 2008; Galea et al. 2008; Fuxreiter et al. 2007).
Cette implication peut notamment être illustrée par l'exemple de p27, régulateur du cycle cellulaire via son interaction avec les complexes CdK/cyclines. La flexibilité d'une de ses régions induit l'exposition et la phosphorylation d'un résidu tyrosine qui déclenche une cascade de régulations conduisant à la dégradation de ce facteur (Galea et al. 2008).

Les chaperonnes

Les chaperonnes sont des protéines qui assistent d'autres protéines et des molécules d'ARN pour atteindre la conformation qui les rend fonctionnelles. Dans le cas des protéines, elles peuvent prévenir leur agrégation et/ou leur dégradation. De nombreuses protéines chaperonnes possèdent des régions désordonnées (plus de 50% de leurs résidus sont localisés dans des IDR) (Tompa & Csermely 2004). L'abondance de désordre dans les chaperonnes peut être d'une part reliée à leur capacité à lier de nombreuses cibles différentes, et d'autre part, à leur fonction de « solubilisateurs ». En effet, le caractère très hydrophile de leurs régions désordonnées permet à certaines chaperonnes de se déployer autour de leurs cibles afin de créer un microenvironnement propice à leur repliement en une conformation fonctionnelle stable (Tompa & Csermely 2004; Ivanyi-Nagy et al. 2005).

Cette fonction associée au désordre a été montré notamment chez la chaperonne eucaryote Hsp90, pour laquelle les deux larges régions désordonnées en C-terminal sont directement impliqués dans l'activité de solubilisation de Hsp90 et permettent de prévenir l'agrégation de deux de ses protéines cibles (Wayne & Bolon 2010; Pursell et al. 2012).

Les effecteurs

Certaines régions désordonnées modulent l'activité des protéines auxquelles elles sont associées (Galea et al. 2008) et assurent un mode de régulation spécifique à chacune d'elles. C'est notamment le cas du facteur p27 qui, en interagissant avec plusieurs complexes Cdk/cyclines, les régule de façon antagoniste, en activent certains (Cdk4 couplé aux cyclines de type D) et en inhibent d'autres (Cdk2 couplé aux cyclines de type A et E) (Galea et al. 2008).

Les capteurs

Le désordre peut également être associé à la capacité d'interaction avec de petits ligands dans le but de les stocker et/ou de les neutraliser, au cours de divers processus de régulation. Dans ce contexte, le désordre de la caséine, phosphoprotéine secrétée dans le lait et fixatrice de calcium, semble directement être associé à la capacité de cette protéine à stocker d'importantes quantités de calcium et de phosphate sous leur forme soluble, en formant des complexes stables et en prévenant ainsi toute calcification du milieu (Holt 2013).

Les assembleurs

Parmi les fonctions du désordre directement liées à l'interaction avec plusieurs partenaires, les régions désordonnées dites assembleuses présentent de larges surfaces d'interaction pouvant servir d'échafaudage ("*scaffold*") en interagissant de façon simultanée avec plusieurs facteurs, et permettent ainsi la formation et l'assemblage de complexes moléculaires (Gunasekaran et al. 2003; Cortese et al. 2008). La protéine CASK-interactive protein 1 constitue un des nombreux exemples du rôle des longues IDR (plus de 30 résidus consécutifs) dans les protéines scaffold, impliquées dans l'assemblage de complexes moléculaires. Cette protéine, retrouvée dans les neurones chez les mammifères, contient une longue région en C-terminale de 800 résidus démontrée comme désordonnée et associée à l'interaction avec 11 partenaires différents, tous impliqués dans des fonctions de transduction du signal (Balázs et al. 2009).

Pour résumer, l'analyse fonctionnelle des régions désordonnées permet de mettre en évidence que l'absence de structure des IDP/IDR est impliquée dans diverses fonctions de régulation. Parmi ces fonctions, **la reconnaissance moléculaire** de plusieurs partenaires distincts constitue un mécanisme central. Elle est souvent associée à la transition d'un état désordonné à un état ordonné chez les IDP/IDR correspondantes (Figure 5).

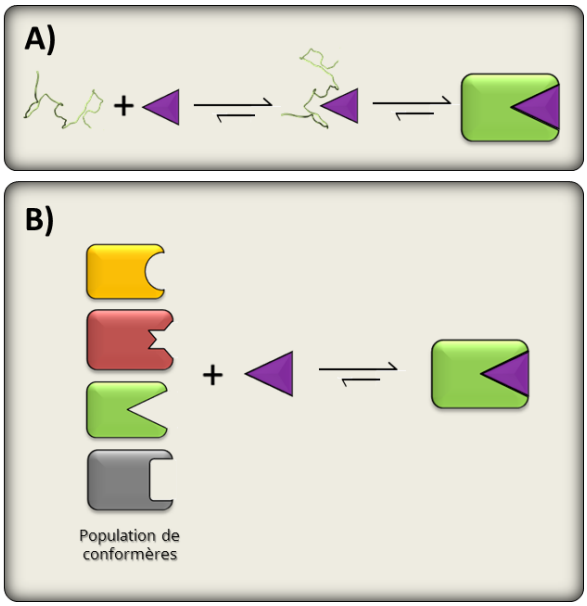


***Figure 6*|Modèles minimums proposés pour décrire les mécanismes d'interactions des IDP/IDR avec un partenaire ordonné. A)** Modèle du repliement couplé à l'interaction. **B)** Modèle de la sélection conformationnelle. *En violet : partenaire ordonné.*

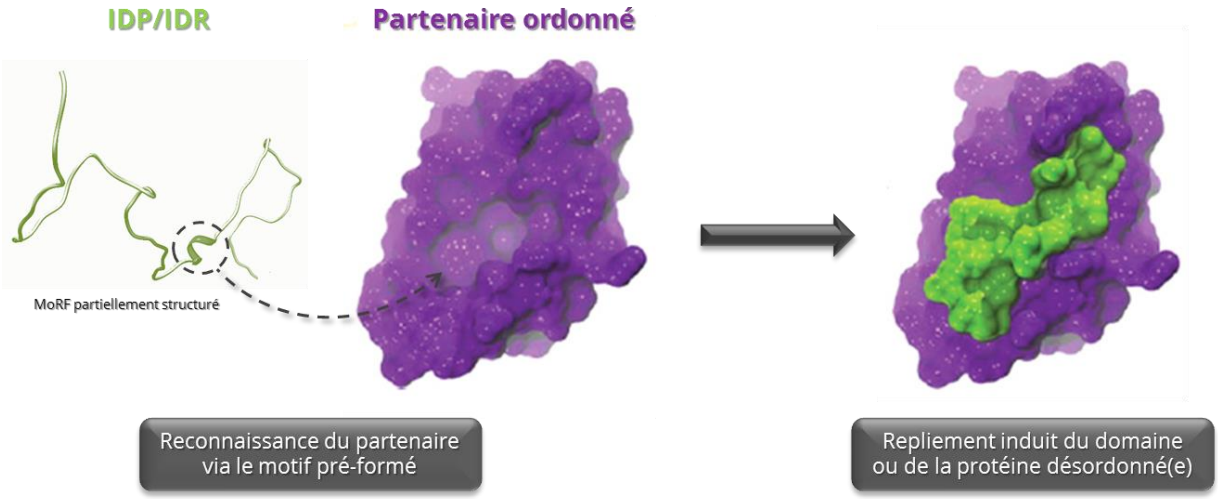


***Figure 7*|Modèle du consensus synergique du mécanisme de repliement et d'interaction des IDP/IDR.** La présence de courts motifs déjà partiellement structurés (MoRFs) (entourés en pointillés à gauche) initie la reconnaissance moléculaire et le recrutement par l'IDP/IDR d'un partenaire donné (violet) Une fois le partenaire recruté, le domaine ou la protéine entière adopte à son tour une conformation (vert) et finalise la formation du complexe.

Mécanistique de la dynamique d'interaction pluri-partenaires mettant en jeux des protéines ou régions désordonnées

Deux modèles minimum ont été proposés pour décrire le repliement des protéines désordonnées. Dans le premier modèle, le repliement et l'interaction sont décrits comme couplés. La présence dans l'environnement d'un partenaire donné induit la fixation rapide à faible affinité de l'IDR/IDP à sa cible, et le repliement de l'IDR/IDP en une conformation complémentaire à la surface du partenaire (*i.e* **repliement induit**) (Figure 6A). Dans le second modèle, l'IDR existe sous forme d'un ensemble de conformères en équilibre les uns avec les autres et parmi lesquels un seul est complémentaire de la cible. Lors de l'interaction, l'équilibre des différentes formes dans la population est déplacé vers le conformère complémentaire. Il s'agit du modèle de **sélection conformationnelle** (Figure 6B).

L'accumulation récente des analyses de la dynamique de repliement et d'interaction des IDP/IDR semblent converger vers l'idée que les deux modèles présentés précédemment ne sont pas exclusifs l'un de l'autre (Espinoza-Fonseca 2009). Ainsi, selon un troisième modèle, le **modèle du consensus synergique**, la première étape d'association d'une IDP/IDR avec son partenaire se fait via une reconnaissance au niveau d'un motif dit pré-formé de l'IDP/IDR (sélection conformationnelle). Cette reconnaissance permet le rapprochement du partenaire, qui va ensuite induire le repliement du reste de la région ou de la protéine (repliement induit) (Espinoza-Fonseca 2009) (Figure 7).
L'interaction avec plusieurs partenaires fait souvent intervenir à la surface des IDP/IDR de courts motifs de reconnaissance moléculaires (MoRF), spécifiques de chacun des partenaires. Ces motifs peuvent être pré-structurés (un tour d'hélice par exemple) et ainsi permettre une première reconnaissance par le partenaire. L'interaction avec ce dernier induit leur repliement en une structure secondaire stable (Fuxreiter et al. 2004).

Comment la flexibilité conformationnelle peut-elle servir la multi-interaction ?

Le désordre des IDP/IDR porteuses d'une capacité de multi-interaction constitue un environnement flexible, permettant à ces déterminants structuraux locaux d'être reconnus en surface et d'adopter différentes structures selon le partenaire.
La flexibilité structurale qui définit ces régions permet de découpler les déterminants structuraux (notion de motifs semi-autonomes) les uns des autres, et de permettre la mise en place de plusieurs structures différentes selon les partenaires.

L'interaction d'une protéine désordonnée avec plusieurs partenaires différents peut se faire par la localisation de motifs de reconnaissance moléculaires distincts, espacés sur la protéine. Dans ce cas, l'interaction avec différents partenaires peut être simultanée. Elle peut aussi avoir lieu dans une même région, qui contient les différents déterminants d'interactions. Cela implique alors une seule interaction à la fois.

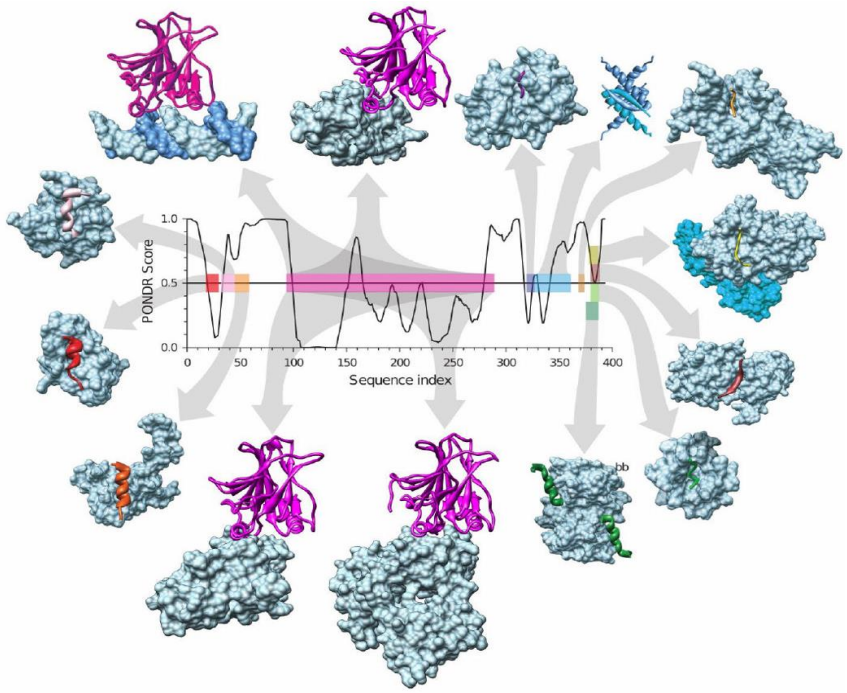


***Figure 8*|Exemple de la multi-interaction médiée par les IDR : Cas du facteur p53.** Scores de désordre prédits par PONDR-VLXT à partir de la séquence en acides aminés de p53 (score des résidus désordonnés > 0.5) et présentation de 14 des 84 conformations adoptées par ces régions (parties colorées) lors de leurs interactions avec les différents partenaires (en gris). La région située à l'extrémité C-terminale de la protéine adopte 4 conformations différentes (représentée en vert clair, vert foncé, jaune et marron) lui permettant d'interagir avec plusieurs partenaires distincts. D'après (Uversky & Dunker 20[10]).

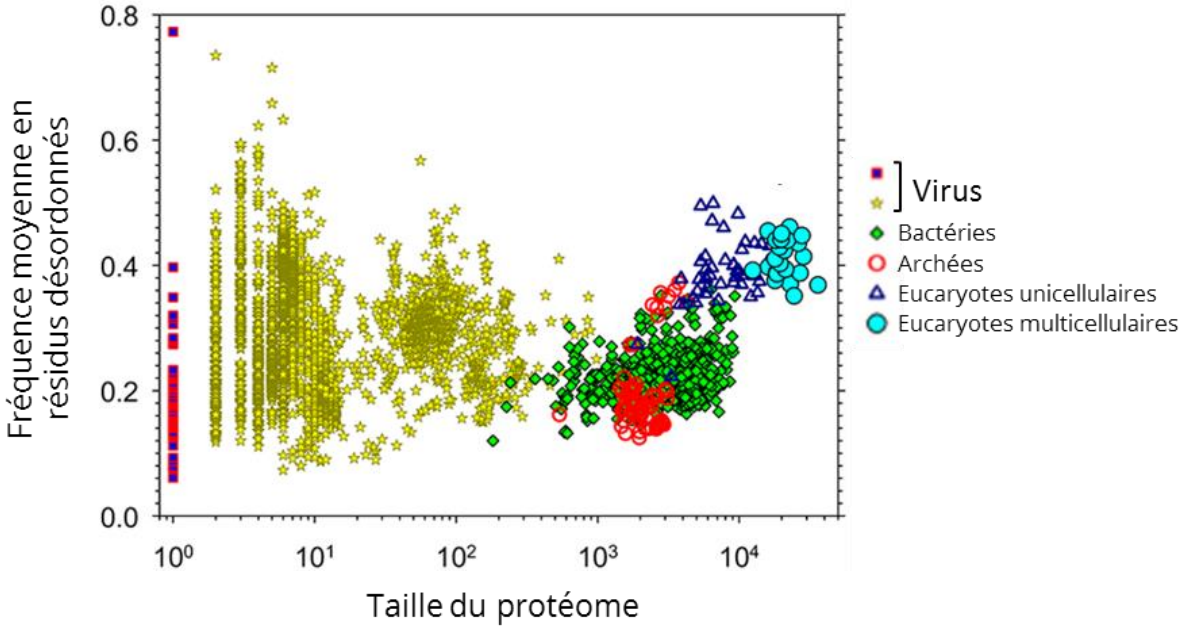


***Figure 9*|Abondance du désordre dans les trois règnes du vivant, Archées, Bactéries et Eucaryotes[1] ainsi que chez les virus.** Les fréquences en résidus désordonnés sont obtenues après prédiction chez 3484 protéomes via le prédicteur PONDR-VSL2B. Adapté de (Xue et al. 2012).

[1] La classification du vivant en trois règnes utilisée ici fait référence à celle de (Woese et al. 1990).

L'exemple le plus représentatif de cette incroyable multi-spécificité d'interaction concerne la protéine suppresseur de tumeur p53. Pour ce facteur, 60 des 84 interactions identifiées sont médiées par des régions désordonnées. Les deux modes d'interactions sont illustrés : certaines régions de p53 sont associées à l'interaction avec un partenaire donné tandis qu'une région à son extrémité C-terminale peut adopter au moins 4 conformations différentes, en fonction du partenaire avec lequel elle interagit (notion de séquence caméléon) (Figure 8).

Conséquence de la flexibilité structurale des régions désordonnées sur l'évolution des protéines et l'adaptabilité des organismes

Chez de nombreux organismes, les régions désordonnées ont pu être associées à une évolution globalement plus rapide de la séquence protéique (Brown et al. 2002; Brown & Johnson 2011; Chen et al. 2006; Bellay et al. 2011; Nilsson et al. 2011; Kim et al. 2008). Cette capacité évolutive tient vraisemblablement du fait que la faible densité de liaisons inter-résidus dans les zones désordonnées conduit à une plus faible contrainte topologique que celle observée au niveau des régions plus structurées. L'environnement désordonné pourrait être associé à une tolérance plus importante des changements en acides aminés qui surviennent continuellement, au cours du cycle de vie d'un organisme. Cette permissivité aux mutations induite par les IDR conférerait alors aux organismes une capacité particulière à s'adapter aux perturbations environnementales (Pancsa & Tompa 2012). *Les conséquences d'une telle propriété sur l'évolution des protéines seront discutées dans la section 2.D).*

c. Prédiction du désordre *in silico* et estimation globale de l'abondance du désordre dans le vivant

Les régions désordonnées présentent un biais dans leur composition en acides aminés par rapport aux régions structurées. Il est donc possible de développer des algorithmes capables de prédire *in silico* la proportion en désordre d'une protéine en se basant sur sa séquence primaire. Les prédicteurs *in silico* tel que PONDR® (Predictor Of Naturally Disordered Regions) (Romero et al. 2001), utilisé dans cette thèse (voir chapitre 1), sont des programmes dont l'élaboration nécessite une étape dite d'apprentissage. Chez ces prédicteurs basés sur des réseaux neuronaux, l'identification expérimentale de nombreuses régions désordonnées peut donc permettre de « nourrir » et « d'entraîner » ces algorithmes afin d'augmenter la performance des prédictions de désordre *in silico* (Figure 3A et 3B). L'amélioration des prédicteurs ainsi que le nombre grandissant de séquences protéiques désormais disponibles dans les banques de données permettent d'appréhender à grande échelle l'abondance du désordre (Chen et al. 2008; Xue et al. 2012; Peng et al. 2014). Une étude récente conduite sur plus de 3500 protéomes, a ainsi estimé les taux moyens en désordre pour chaque grande classe du vivant (Figure 9).

De façon générale, le **désordre apparaît comme ubiquitaire** dans le monde vivant. Néanmoins, les taux de désordre sont hétérogènes et corrélés à la taille des protéomes.

Ainsi les protéomes eucaryotes présentent une abondance en désordre nettement plus élevée (33-50% des résidus sont prédits comme appartenant à une région désordonnée) que celle déterminée chez les Procaryotes (15-30%) et les Archées (12- 24%). Les études d'annotation fonctionnelle du désordre *in silico* ont permis d'observer que de façon générale le désordre est associé aux processus de régulation, de signalisation et de contrôle (Ward et al. 2004; Xie et al. 2008; Xie et al. 2007; Vucetic et al. 2007). En effet, les facteurs dits "hubs", correspondant aux protéines multipartenaires situées aux carrefours des réseaux de régulation, sont en grande majorité désordonnés (Haynes et al. 2006; Kim et al. 2008; Bellay et al. 2011). L'hypothèse majeure permettant d'expliquer cet enrichissement du désordre chez les eucaryotes consiste donc à associer le désordre au développement chez ces organismes de systèmes de régulation et de signalisation complexes, et impliquant notamment une abondance croissante de ces hubs. Pour ce qui est des eucaryotes unicellulaires, la variabilité de leurs habitats respectifs et les capacités d'adaptation aux changements environnementaux associées peuvent expliquer les taux hétérogènes de désordre observés (Mohan et al. 2008; Pancsa & Tompa 2012).

Ce lien entre complexité et désordre ne semble toutefois pas applicable au monde des virus, pour lequel la taille très réduite des génomes (par rapport aux génomes eucaryotes) est associée à un taux de désordre moyen important pouvant atteindre 80%. **Ceci pose la question de la fonctionnalité associée au désordre chez les virus**. Les avancées et hypothèses concernant ce taux de désordre seront discutées au regard des particularités propres aux virus.

2. Les fonctions associées au désordre chez les virus à ARN

A) Les virus présentent une variabilité record de désordre dans leur protéomes

Par comparaison avec la distribution en désordre observée dans les trois règnes du vivant (Figure 9), les virus présentent un taux extrêmement variable de désordre dans leur protéome, allant de 7,3% de résidus désordonnés pour le *Human corona virus* NL63 (HCoV-NL63) à plus de 75% pour le *Avian carcinoma virus (ACMHV-2)* (Xue et al. 2012).
Les taux record observés chez certains virus interpellent sur la fonctionnalité du désordre dans le cycle viral.

Comment les protéines virales utilisent-elles la flexibilité structurale associée au désordre intrinsèque ?

Comme nous allons le voir dans la section suivante, les virus à ARN possèdent des caractéristiques spécifiques qui peuvent s'illustrer par les propriétés très particulières de leurs protéines.
Pour comprendre comment les fonctions usuellement associées au désordre intrinsèque peuvent être particulièrement bénéfiques aux virus, il convient d'abord de présenter les propriétés de ces micro-organismes.

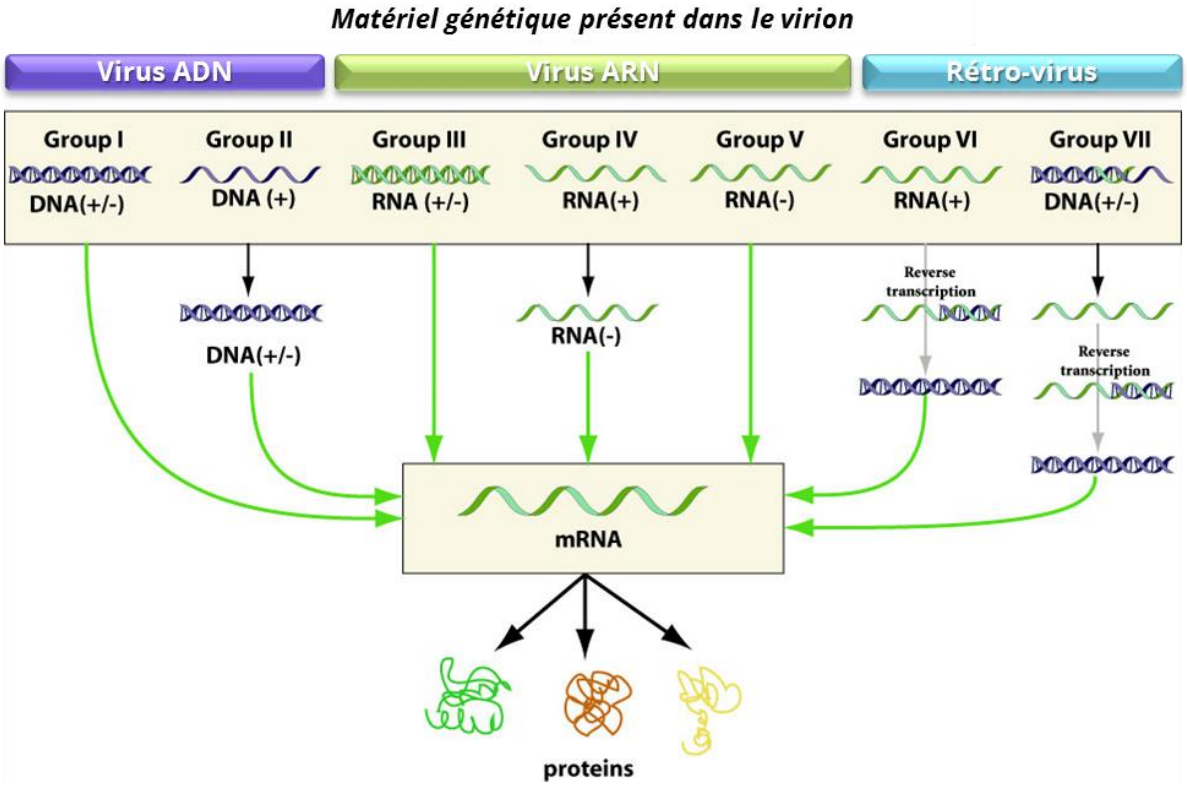


***Figure 10*|Représentation des différents groupes de virus selon la Classification de Baltimore** (Baltimore 1971)**.** Les étapes intermédiaires de réplication et de rétro-transcription conduisant à la synthèse d'un ARN messager, et à la traduction des protéines virales sont représentées. *DNA(+/-) : génome à ADN double brin ; DNA(+) : génome à ADN simple brin ; RNA(+/-) : génome à ARN double brin ; RNA(+) : génome à ARN simple brin de polarité positive ; RNA(-) : génome à ARN simple brin de polarité négative. mRNA : ARN messager viral codant pour les protéines virales.* Adapté de ViralZone (http://viralzone.expasy.org/ ; Swiss Institute of Bioinformatics).

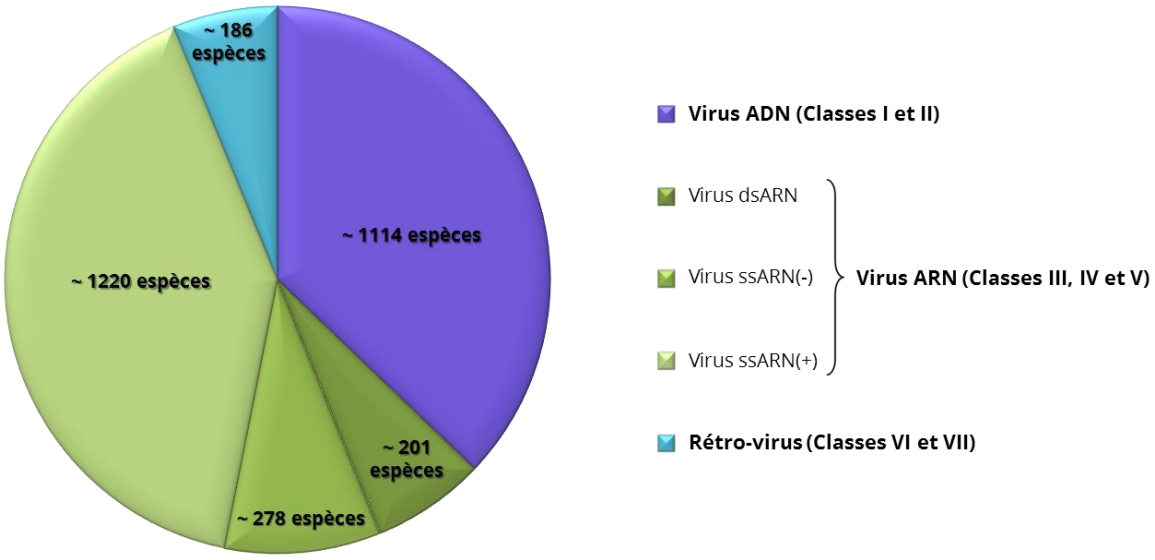


***Figure 11*|Abondance des classes de virus selon le rapport ICTV 2014.**

B) Les virus à ARN : des agents aux propriétés très particulières

Définition et place des virus par rapport au monde du vivant

Les virus peuvent être définis comme des micro-organismes acellulaires, parasites obligatoires, dont le génome composé d'acides nucléiques code des protéines.

Ils constituent l'entité biologique la plus abondante sur Terre [le nombre de particules virales est plus de 10 fois supérieur au nombre total de cellules (Suttle 2007)]. Tout en considérant que la grande majorité d'entre eux n'a certainement pas encore été identifiée, le dernier rapport de *l'International Committee on Taxonomy of Viruses* (ICTV) recense plus de 3000 espèces regroupées en 104 familles virales, capables d'infecter toutes les espèces vivantes peuplant notre planète (Adams et al. 2012; Adams et al. 2015).

L'incroyable diversité chez les virus

De par notamment leurs origines très anciennes et présumées non communes (Forterre 2006), une diversité extrême existe au sein des virus et se caractérise à tous les niveaux (structure et taille du virion, nature et structure du génome, longueur de génome, cycle, hôte, méthode de propagation…). En ce qui concerne la nature de leurs génomes, les virus peuvent en effet être composés soit d'ADN soit d'ARN, sous la forme de double ou simple brin, de polarité positive ou négative. Une classification basée sur ce critère, initialement proposée par David Baltimore (Baltimore 1971), est représentée Figure 10.

Les virus ARN : des cibles majeures de la lutte antivirale

Parmi ces classes de virus, les virus à ARN [englobant les virus ne possédant pas d'étape ADN dans leur cycle, soit les virus des classes III, IV et V (Figure 10)], sont les plus abondants, en représentant plus de la moitié des espèces virales répertoriées en 2014 (Figure 11). Sur un plan anthropomorphe, certains virus à ARN constituent une menace importante pour la santé humaine, l'agronomie et la biodiversité. Ils figurent parmi les agents pathogènes majoritaires des maladies infectieuses émergentes les plus dommageables de notre époque (Anderson et al. 2004; Woolhouse et al. 2013; Howard & Fletcher 2012). Chez les humains, les virus à ARN constituent les agents causaux de maladies telles que la grippe, la rougeole, les fièvres hémorragiques ou encore la dengue. Chez les plantes cultivées également, les dégâts économiques et écologiques associés aux infections par des virus à ARN sont extrêmement lourds (Nicaise 2014). La compréhension de la biologie et des propriétés associées à ces virus apparait donc comme d'intérêt majeur dans la lutte et la prévention des maladies émergentes.

Le cycle des virus à ARN

Chez les virus à ARN, le mode d'entrée dans la cellule de l'hôte est virus-dépendant et peut être directement associé à un processus d'attachement membranaire, un mode de vectorisation ou encore à une blessure. Une fois entrés dans la cellule, les virus à ARN se multiplient en de nombreuses copies en répliquant leur génome, via leur RNA-dependent RNA polymerase (RdRp) – aussi appelée réplicase virale.

Cette étape de réplication est compartimentée dans la cellule au niveau d'usines virales cytoplasmiques. La traduction, qui a lieu avant la réplication dans le cas des virus de la classe IV, est quasi-concomitante avec l'étape de réplication, et conduit à la synthèse des facteurs viraux dans la cellule. La synthèse des protéines virales permet l'accomplissement des étapes suivantes du cycle viral telles que l'encapsidation, le mouvement des particules vers les cellules voisines, la colonisation de l'organisme hôte (systémie) dans le cas d'hôtes multicellulaires, ainsi que la propagation à de nouveaux organismes. L'ensemble de ces étapes fait intervenir le recrutement et le détournement de la machinerie cellulaire de l'hôte.

Compacité de génome et taux élevé de mutations : des caractéristiques intrinsèques aux virus à ARN

A la différence des virus à ADN, dont la taille des génomes s'étend de 1700bp pour le génome du *Porcine circovirus 1* à 2,5Mb pour le génome du « virus géant » *Pandoravirus salinus* (Finsterbusch & Mankertz 2009) (Philippe et al. 2013), les virus à ARN présentent une variation de taille de génome beaucoup plus réduite, comprise entre 3500nt (bactériophage MS2) et 30000nt (coronavirus). Ces génomes de taille réduite (*i.e* compacité de génome) ainsi que le faible nombre de protéines codées constituent des caractéristiques communes à l'ensemble des virus à ARN.

Le taux élevé de mutations introduites par les réplicases virales lors de l'étape de réplication des génomes viraux constitue une autre caractéristique commune à l'ensemble des virus à ARN. Ce taux d'erreurs record, compris entre 10^{-3} et 10^{-5} erreurs/nucléotides/cycle de réplication (Duffy et al. 2008) a été relié à l'extraordinaire capacité d'adaptation de ces virus, puisqu'il leur permet, à des vitesses souvent très rapides, de contrer les mécanismes de défenses de l'hôte et de coloniser de nouvelles espèces (voir section 2.D).

Le désordre intrinsèque des protéines peut-il constituer un moyen pour les virus ARN d'assurer leur cycle infectieux tout en s'adaptant aux perturbations environnementales, dans le contexte de la compacité de génome et du taux élevé de mutations qui les caractérise ?

Les arguments en faveur de cette implication du désordre sont présentés dans les deux sections suivantes (2.C et 2.D). Les répercussions en termes de biologie et d'évolution seront également présentées.

C) La flexibilité structurale au service de la multi-interaction chez les virus

Les virus doivent recruter de nombreux facteurs de l'hôte et dans certains cas, de l'organisme vecteur, pour assurer chacune des étapes de leur cycle viral, de l'entrée dans la cellule à l'exportation des particules virales, et ce tout en esquivant les nombreux mécanismes de défense mis en place par l'hôte. Les protéines virales doivent ainsi interagir entre elles, avec les membranes cellulaires, les facteurs de l'hôte, ainsi

que les acides nucléiques. Malgré cela, les génomes des virus à ARN ne codent généralement pas plus d'une douzaine de protéines.

Une question se pose alors : **comment assurer autant de fonctions/interactions avec si peu de protéines ?**

Pour assurer l'ensemble des fonctions nécessaires au cycle, les protéines virales apparaissent de façon générale comme multifonctionnelles. Dans ce contexte de multifonctionnalité, le désordre intrinsèque peut constituer un avantage non négligeable. Cette flexibilité structurale a été déjà largement décrite comme associée à la multi-interaction chez les organismes eucaryotes, à l'instar du cas extrême de la protéine p53, décrite précédemment (Oldfield et al. 2008).

La capacité des protéines désordonnées à adopter différentes conformations leur permet d'assurer des fonctions variées dans différents contextes physico-chimiques (*e.g* compartiments cellulaires, ligands modulateurs, modifications post-traductionnelles, concentrations locales des différents partenaires…).

Cette propriété de multi-interaction (*i.e* "binding promiscuity") peut donc constituer un avantage clé chez les virus pour assurer les différentes étapes de leur cycle, et le cas échéant, dans différents compartiments cellulaires, tout en disposant d'un nombre très réduit de protéines.

Chez certains virus animaux à ARN, comme le virus de l'Hépatite D (HDV), le virus de l'hépatite C (HCV) ou encore le virus de la Rougeole (MeV), le désordre intrinsèque a d'ores et déjà pu être associé à cette capacité multiple d'interaction (Alves et al. 2010; Macdonald & Harris 2004; Bourhis et al. 2006). Il intervient ainsi dans divers processus au cours du cycle de ces virus, tels que la réplication, l'assemblage de la particule virale ou encore le contournement des voies de signalisation de l'hôte.

D) Le désordre intrinsèque, un moteur de l'adaptation virale ?

La seconde hypothèse en termes de fonctionnalité du désordre intrinsèque concerne son rôle potentiel dans le processus d'adaptation chez les virus à ARN. Pour comprendre l'implication de cette propriété structurale sur le processus d'adaptation, il conviendra d'abord de présenter les processus de l'évolution et de l'adaptation virale, ainsi que les principaux déterminants associés aux différentes étapes de ces processus.

a. L'extraordinaire adaptabilité des virus à ARN passe par la diversité génétique de leurs populations

D'un point de vue très général, la capacité d'un organisme à évoluer passe par sa capacité à accumuler de la diversité génétique tout en continuant à assurer son cycle de vie (notion de compensation diversité *versus* fitness[2]). Les virus à ARN sont à ce jour les entités biologiques décrites comme ayant la plus grande capacité évolutive. Ils constituent un modèle idéal dans la description des conséquences adaptatives liées à la diversité d'une population.

[2] Fitness: Mesure quantitative définissant la capacité d'un individu à accomplir son cycle et à se transmettre aux générations suivantes.

En effet, les virus à ARN existent sous la forme de populations d'individus très divers, appelées quasi-espèces, tous reliés par des évènements de mutations, et capables de coopérer fonctionnellement pour contribuer aux caractéristiques de la population (Lauring & Andino 2010).
Cette théorie des quasi-espèces, développée par Manfred Eigen et Peter Schuster à la fin des années 1970 (Eigen & Schuster 1977), établit l'organisation des quasi-espèces comme un ensemble de variants minoritaires dominés par une séquence majoritaire appelée « master sequence » et présentant la fitness la plus importante.

En termes d'évolution et d'adaptation, la diversité de séquences générée dans les populations de virus à ARN sous-tend une plus vaste exploration des régions de l'espace de séquences, et donc potentiellement des solutions adaptatives, dans un environnement donné. En effet, cette exploration des séquences a d'ores et déjà pu être reliée à l'extraordinaire potentiel adaptatif de ces entités biologiques (Goldhill et al. 2014), de par son implication dans divers processus d'adaptation tels que l'augmentation de pathogénicité (Pfeiffer & Kirkegaard 2005; Vignuzzi et al. 2006), l'échappement à la réponse immune (van Nimwegen 2006), la résistance aux drogues (Domingo et al. 2012), aux vaccins (Davenport et al. 2008) ainsi que lors d'évènements adaptatifs conduisant aux sauts d'espèces et à l'émergence de nouvelles maladies virales (Pepin et al. 2010).

Dans ce contexte, la détermination des acteurs de cette diversité et des mécanismes selon lesquels ils interviennent dans les processus d'adaptation virale sont nécessaires à notre compréhension fondamentale de l'évolution virale et, de façon plus finalisée, à l'optimisation des méthodes de lutte antivirale.

b. Le taux d'erreurs des réplicases virales : un générateur de diversité génétique

La mutation est un processus aléatoire, défini par l'altération de l'information génétique qui, chez l'ensemble du vivant constitue la source majeure de diversité nécessaire à l'évolution. Chez les virus, ces mutations proviennent essentiellement des erreurs de la polymérase, survenant au cours de la réplication, pouvant entraîner des changements de nucléotides (substitutions), et plus rarement des insertions ou des délétions. Contrairement aux organismes eucaryotes, dont les taux de mutations sont de l'ordre de $1x10^{-9}$ erreur/nucléotide/cycle de réplication (Drake et al. 1998), les réplicases virales des virus à ARN présentent un taux d'erreur record, dont l'estimation est comprise entre $1,5x10^{-3}$ à $5x10^{-5}$ erreur/nucléotide/cycle de réplication (Duffy et al. 2008; Sanjuán et al. 2010; Drake et al. 1998). Cette propension accrue à l'accumulation de mutation s'explique par l'absence de mécanismes de réparation associés à la réplication des génomes ARN. Compte-tenu de la taille souvent limitée des génomes des virus à ARN, chaque cycle de réplication engendre donc en moyenne au moins une mutation par génome. Certains facteurs comme la vitesse du cycle de réplication ainsi que la taille du génome contribuent à ce **taux de mutations** et peuvent expliquer les disparités observées au sein des différentes classes de virus (Duffy et al. 2008).

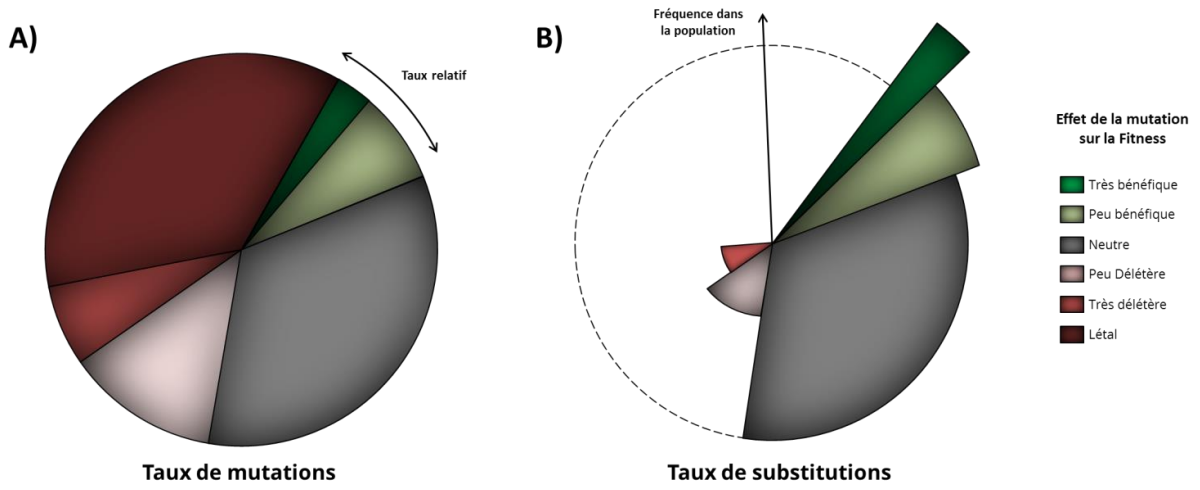


***Figure 12*|Taux de mutations *versus* taux de substitutions chez un virus ARN**. **A)** Taux relatif des mutations spontanées (issues des erreurs des réplicases virales au cours des étapes de réplication) en fonction de leur effets sur la fitness du virus, déterminées pour le *Tobacco etch virus* (TEV) (Sanjuán 2010). **B)** Taux de substitutions des mutations : fréquence de chaque classes de mutation dans la population (après l'effet purgatoire des forces de sélection) selon leur effet sur la fitness. Valeurs théoriques issues de (Barrick & Lenski 2013).

c. La robustesse mutationnelle : gardienne de la diversité génétique au cours de l'évolution des populations de virus à ARN

Le taux de mutations spontanées constitue la source majeure de la diversité observée chez les populations virales, et est directement relié aux **taux de substitution** chez les virus (Sanjuán 2012). Le taux de substitution définit la fréquence à laquelle les mutations s'accumulent et sont maintenues dans la descendance et fait directement référence au taux évolutif d'une population. A la différence du taux de mutations, qui est le résultat de processus aléatoires (erreur de la réplicase), le taux de substitutions est le résultat direct des forces évolutives s'exerçant sur la population dans un environnement donné, dont la plus importante est la sélection.

Selon le principe de la sélection naturelle, énoncé pour la première fois par Charles Darwin en 1859, sous l'effet de cette sélection, les nombreuses mutations vont donc être « triées » selon l'impact qu'elles induisent sur la fitness du virus. Ainsi les mutations dites neutres ou bénéfiques seront maintenues dans la population. De la même façon, les mutations à effets plutôt sur la fitness de l'individu, majoritaires parmi les mutations spontanées chez certains virus ((Sanjuán 2010), Figure 12A), auront plutôt tendance à être purgées. Les mutations létales seront quant à elles complètement éliminées de la population (Figure 12B).

La force de sélection est la résultante de l'action de plusieurs sources de pressions évolutives s'exerçant dans un environnement donné, et qui proviennent essentiellement de l'hôte, du virus lui-même et de l'agent vecteur le cas échéant.

En reliant l'effet d'une mutation sur la fitness de l'organisme à sa fréquence dans la population, les forces de sélection qui s'exercent jouent un rôle majeur dans la composition et l'évolution des populations virales.

Cependant, l'intensité de ces forces de sélection peut varier selon les environnements et plusieurs paysages de variants peuvent ainsi être observés. En effet, dans le contexte d'une forte pression de sélection, seules les rares mutations entraînant un gain important de fitness vont échapper à la purge sélective et être maintenue ce qui aura pour conséquence une diversité génétique très faible de la population.

A l'inverse, dans un contexte environnemental où la pression de sélection est moins importante et pour laquelle la population de départ est déjà bien adaptée, l'émergence de nombreuses mutations à effets neutres voire faiblement déstabilisants pour la fitness virale pourront être maintenues, à fréquence limitée (Barrick & Lenski 2013).

Ainsi, hormis le taux de mutations, plusieurs autres facteurs contribuent à moduler le taux évolutif (taux de substitutions) des génomes viraux, en influençant la façon dont la sélection va façonner la population de variants.

La robustesse phénotypique aux mutations, aussi appelée **robustesse mutationnelle**, est définie par la capacité d'un organisme à maintenir son phénotype malgré les perturbations que constitue l'accumulation de mutations. En dictant l'expression phénotypique de la variation génétique, la robustesse mutationnelle constitue **un des déterminants majeurs de la capacité évolutive d'un organisme** (Montville et al. 2005). La détermination de l'effet de mutations individuelles sur la fitness globale des organismes (MFE pour *Mutational Fitness Effect*) a notamment permis d'observer que chez des organismes tels que *Escherichia coli* ou *Caenorhabditis elegans*, plus de 90% des mutations étaient non létales (Davies et al. 1999; Baba et al. 2006). Avec seulement 20% à 40% des mutations spontanées déterminées comme non-délétères, les virus semblent au contraire moins robustes aux mutations (Figure 12) (Sanjuán 2010).

La capacité à tamponner les effets potentiellement délétères des mutations peut passer par des mécanismes situés à différentes échelles moléculaires. La duplication des gènes induisant la redondance fonctionnelle, l'utilisation de voies métaboliques alternatives, le niveau de ploïdie, l'action de protéines chaperonnes (Fares 2015) ainsi que la structure des protéines (Bloom et al. 2006; Tokuriki & Tawfik 2009) sont autant de facteurs qui peuvent entrer en ligne de compte pour expliquer la capacité d'un organisme à tolérer les mutations.
Compte-tenu de la taille extrêmement réduite des virus à ARN, les stratégies de robustesse basées sur la redondance fonctionnelle des facteurs viraux apparaissent comme très limitées et pourraient expliquer la fréquence importante de mutations délétères ou létales observée chez ces virus. De plus, la multifonctionnalité de leurs protéines ainsi que la présence de cadres de lectures chevauchants renforcent l'idée que les virus ARN apparaissent comme très sensibles aux mutations.
Pourtant le maintien d'un nombre non négligeable de mutations et la diversité tout de même importante retrouvée au sein des populations virales ne reflète pas cette hyper-sensibilité aux mutations.
L'utilisation des protéines chaperonnes de l'hôte, déjà observée chez de nombreux virus (Xiao et al. 2010; Pechmann & Frydman 2014) ainsi que les **caractéristiques structurales des protéines virales**, pourraient constituer les déterminants majeurs de la tolérance aux mutations chez ces organismes, et donc du maintien d'une diversité génétique malgré l'action de la sélection naturelle.

La structure protéique comme déterminant de la robustesse mutationnelle

La capacité évolutive d'une protéine, peut être quantifiée par l'accumulation de mutations non-synonymes (NS), impliquant un changement en acides aminés. Elle est le résultat de plusieurs paramètres intrinsèques à la protéine. Le taux d'expression de la protéine, le nombre d'interactions auxquelles elle participe ainsi que ses propriétés structurales sont impliqués dans la capacité de celle-ci à évoluer et à accumuler des mutations (Toll-Riera et al. 2012; Bloom et al. 2006; Ferrada & Wagner 2008; Kim et al. 2006; Tokuriki & Tawfik 2009; Drummond et al. 2006). La capacité à accumuler des mutations non-synonymes constitue donc une propriété intrinsèque des protéines, et

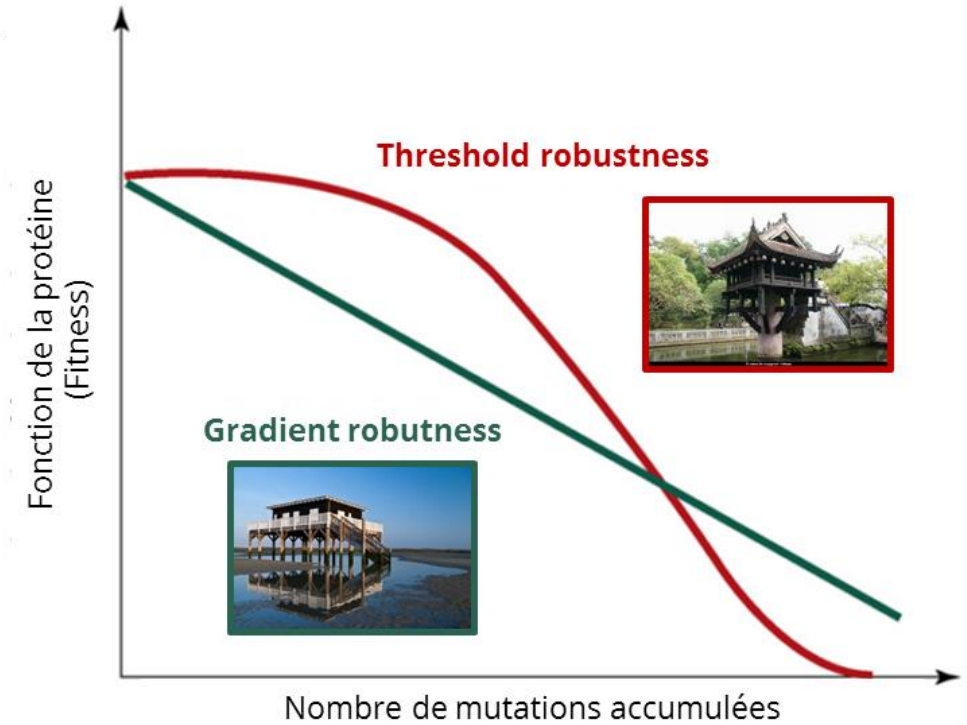


Figure 13*|Modèles de permissivité mutationnelle associés aux protéines selon leur niveau de structuration.** En vert, le comportement de robustesse mutationnelle associé aux protéines désordonnées virales, en rouge celui associé aux protéines ultra-structurées. *Adapté de* (Tokuriki et al. 2009). La robustesse par l'ultra-structuration (compacité structurale) pourrait correspondre à une pagode au pilier unique (encadré rouge), et la robustesse par diminution des contraintes structurales serait assimilée à la maison sur pilotis (encadré vert). Dans cet exemple, la robustesse est alors définie comme la capacité des édifices à tenir debout (fitness de la protéine) malgré les coups de haches (mutations). ***Illustration du "Threshold robustness": l'accumulation des premiers coups de haches affectant uniquement la surface du pied, n'engendre aucune instabilité dans le cas de la pagode au pilier unique. En revanche peu de coups suffisent, une fois que le cœur du pied est à découvert, pour faire effondrer la structure toute entière. ***Illustration du "Gradient robustness":*** contrairement au pilier unique, les pilotis sont eux bien espacés. Dès les premiers coups de haches, les pilotis cèdent mais leur nombre fait que la structure, bien que de plus en plus bancale, est tout de même maintenue (effet neutre ou peu délétère de la mutation), et que dans ce cas, peu de déterminants sont indispensables à la structure globale de l'édifice.

peut se révéler très différente d'une protéine à une autre, dans un même organisme (Drummond et al. 2006). En ce qui concerne le déterminisme structural, des paramètres comme l'accessibilité au solvant et la robustesse de la structure 3D sont directement corrélés à la robustesse mutationnelle des protéines ordonnées. Chez ces mêmes protéines, la notion de « désignabilité » (en anglais, *designability*) fait référence au nombre de séquences différentes permettant à une même protéine d'adopter la même structure et peut donc être vue comme la capacité d'une protéine à évoluer de façon neutre, sans impact pour sa structure et donc pour sa fonction (Bloom et al. 2006). Cette désignabilité est associée aux structures protéiques compactes présentant des densités de contact inter-résidus importantes. Ces caractéristiques structurales ont ainsi pu être corrélées aux taux adaptatifs de protéines de structure 3D connue (Rorick & Wagner 2011). Dans le cas des protéines ordonnées, dont la fonction est strictement liée au repliement en une structure unique, la désignabilité fournie par la compacité structurale constitue un critère significatif de la robustesse mutationnelle d'une protéine et donc de sa capacité évolutive (« évolvabilité »).

d. Le désordre intrinsèque comme moyen alternatif pour accumuler des mutations ?

Que deviennent les concepts de stabilité et d'évolvabilité lorsqu'on considère des protéines intrinsèquement désordonnées, ne présentant pas de structure 3D fixe et unique en condition physiologique ?

Il est concevable que des contraintes évolutives moindres s'exercent sur ces régions dépourvues de structures fixes, bien qu'elles présentent de faibles densités de contact entre résidus. Selon la relation entre stabilité et désignabilité présentée précédemment (cf. paragraphe 2D.c), une capacité d'adaptabilité médiocre serait attendue pour les protéines désordonnées. La désignabilité de ces protéines est pourtant associée à d'importantes capacités évolutives et adaptatives (Brown et al. 2002; Schlessinger et al. 2011). Chez certaines protéines virales en particulier, la faible densité de contact semble être directement reliée à un maintien de la stabilité protéique lors de l'accumulation de mutations (Tokuriki et al. 2009).

Cette contradiction apparente illustre bien que deux mécanismes antagonistes de permissivité mutationnelle coexistent au niveau protéique et s'appliquent respectivement aux protéines ultra-ordonnées (densité de contact et de désignabilité importantes) et désordonnées (densité de contact faible), présentant des propriétés biophysiques opposées.

Dans le cas des protéines très structurées, le nombre accru d'interactions entre les résidus rend la structure « hyperstable ». L'accumulation de mutations est donc tamponnée par les nombreuses interactions restantes, stabilisant la structure d'ensemble. Cependant, il existe un seuil critique de mutations au-delà duquel la structure s'effondre totalement (Bershtein et al. 2006). On parlera alors de « Threshold Robustness » ou robustesse limite, symbolisée par la courbe verte sur la figure 13.

A l'opposé, les protéines désordonnées présentant, par définition, moins de contraintes topologiques, vont être de façon générale moins affectées par l'accumulation de mutations. Leur stabilité est donc supposée décroître de façon graduelle (*i.e* « gradient robustness »), représentée par la courbe verte, Figure 13).

e. Conséquences de la robustesse mutationnelle associée au désordre intrinsèque dans le contexte de l'adaptation virale

Pour comprendre comment la flexibilité structurale et la robustesse mutationnelle associées aux régions désordonnées peuvent favoriser l'adaptation virale, deux scénarios sont présentés et se distinguent l'un de l'autre par la capacité ou non d'un virus à se répliquer et donc à évoluer dans un hôte dit « résistant ».

Scénario 1 : Cas de l'adaptation « indirecte » à des résistances non-hôte

Dans la situation d'une adaptation conduisant à un saut d'espèces, la forme non-adaptée de départ est incapable de se répliquer dans une espèce non-hôte. L'adaptation, qu'on désignera alors comme **« indirecte »**, passe donc par l'émergence de formes dites pré-adaptées lors de l'évolution dans un hôte sensible. Ce n'est que lors de la transmission de ces formes pré-adaptées dans le nouvel hôte que le processus d'adaptation pourra avoir lieu. L'importante pression de sélection présente dans le nouvel hôte va avoir pour effet de réduire drastiquement la diversité en ne laissant émerger que la forme adaptée (Figure 14A). On voit donc dans ce cas que le potentiel adaptatif est directement relié à la capacité d'accumuler de la diversité génétique en conditions favorables (dans l'hôte sensible), lorsque la pression de sélection est moins forte et permet d'accumuler des variants de fitness variable. Ces variants peuvent être considérés comme des intermédiaires, points de départ pour les adaptations évolutives qui surviendront ensuite dans l'espèce non-hôte (Wagner 2008) et constituent donc un « réservoir adaptatif ».

Ainsi la robustesse mutationnelle qui semble être généralement associée au désordre intrinsèque peut, dans ce contexte d'adaptation « indirecte », contribuer au maintien des différents variants dans la population, et permettre notamment ce qu'on appelle la **Variation Productive** (en anglais *« Productive variation »*). Ce terme fait référence à la <u>mise en place de variations génétiques ne compromettant pas (ou marginalement) la fitness de l'organisme dans un environnement donné mais pouvant constituer une solution adaptative dans un nouvel environnement.</u> En effet, un changement en acide aminé peut être neutre dans un contexte environnemental donné, et se révéler très bénéfique dans un nouvel environnement (Bloom et al. 2007).

Selon cette hypothèse, lors de l'évolution d'une population virale en situation favorable (hôte sensible), les mutations s'accumuleraient préférentiellement au niveau des zones désordonnées, afin de limiter leurs effets délétères, d'éviter la purge de la sélection et de maintenir de la diversité dans la population virale. **Tester cette hypothèse dans le contexte de l'évolution sur hôte sensible fait l'objet du second chapitre de ce travail.**

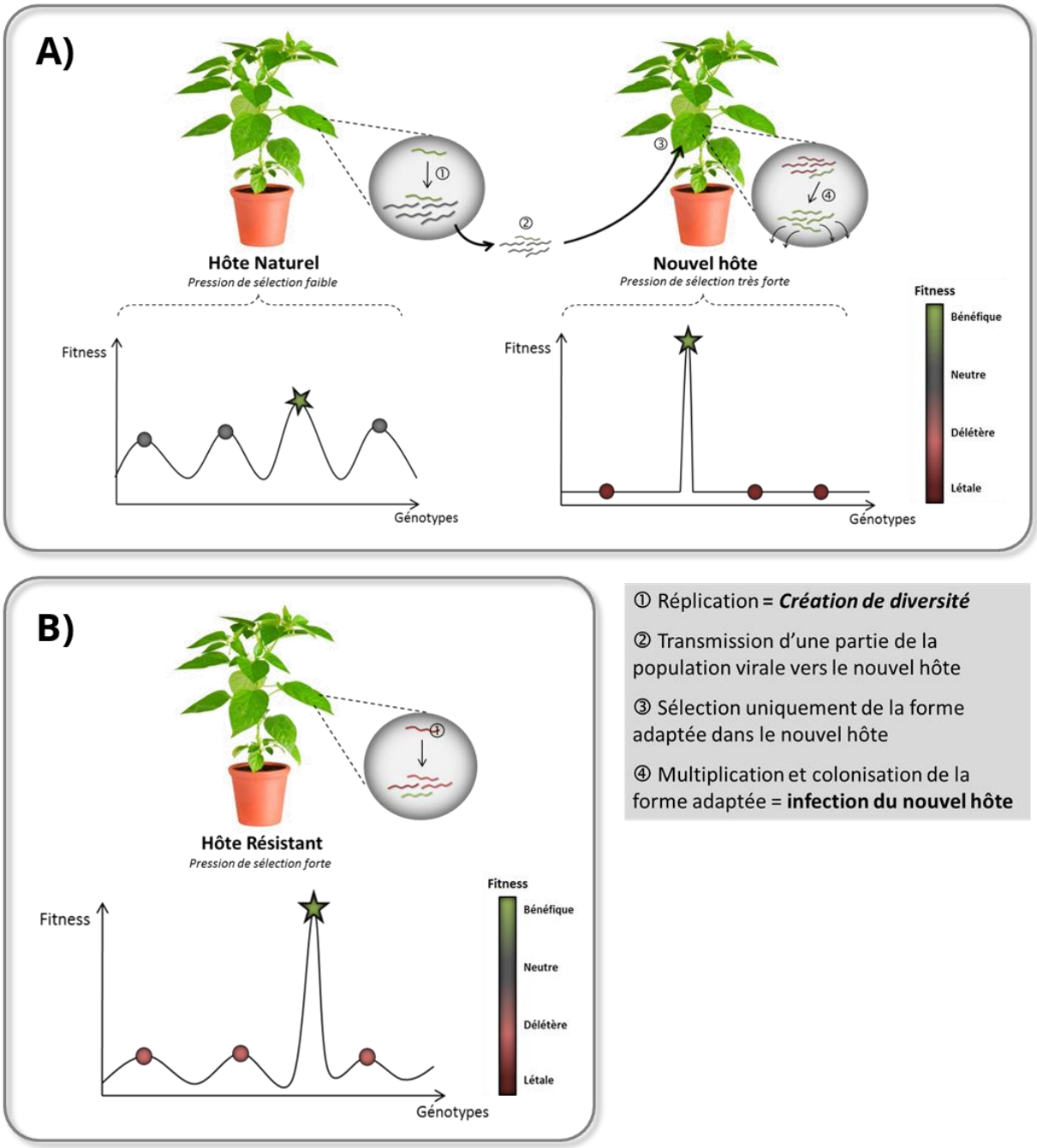


***Figure 14*|Représentation simplifiée des étapes de l'adaptation « directe » et « indirecte » d'un virus à ARN(+) à un hôte résistant. A)** Adaptation indirecte à un hôte dans lequel le virus initial est incapable de se multiplier et d'évoluer. Chaque variant est représenté selon sa fitness. **B)** Adaptation directe à un hôte porteur d'une résistance pour laquelle il y a capacité du virus à se multiplier et à évoluer de façon basale. Les graphiques décrivent la population existante au cours du processus d'adaptation.
La fitness de chaque variant (génotype) est représentée selon un gradient de couleur rouge-vert.

Scénario 2 : Cas de l'adaptation « directe » à des résistances de l'hôte

Suivant le second mécanisme général d'adaptation, des formes virales non-adaptées, dont la fitness est faible, sont tout de même capables de se répliquer et donc d'évoluer *a minima* dans l'environnement pourtant non-favorable de l'hôte résistant. Ce mode d'adaptation, qu'on désignera comme **« directe »** fait référence à l'ensemble des situations pour lesquelles les stratégies antivirales (résistance naturelle de l'hôte, défenses de l'hôte, drogues antivirales…) ne sont pas efficaces à 100% en ce qui concerne l'élimination de la population virale. Celle-ci peut donc se maintenir de façon basale et limitée dans l'hôte. L'adaptation passe donc dans ce cas, non pas par l'accumulation de diversité mais par l'innovation et l'émergence d'une forme adaptée parmi les autres variants délétères. De par son gain de virulence et donc sa fitness importante, le variant sera très rapidement et efficacement sélectionnée et deviendra majoritaire dans la population (Figure 14B).

Dans ce contexte d'adaptation, le désordre peut donc constituer un environnement favorable à l'émergence de ces innovations adaptatives, via les faibles contraintes topologiques qui caractérisent ces régions non structurées et la potentielle robustesse mutationnelle qui leur est liée. La présence de désordre aurait pour conséquence d'augmenter le nombre d'évènements adaptatifs permettant de restaurer une infection en contexte de résistance, et donc finalement d'augmenter le potentiel adaptatif du virus en question.

Tester expérimentalement cette hypothèse fera l'objet du troisième chapitre de ma thèse.

En conclusion, pour ces deux processus adaptatifs, les régions désordonnées contenues dans les protéines des virus à ARN pourraient être la condition nécessaire à l'établissement d'une diversité génétique maintenue dans la population et plus particulièrement de l'émergence de mutations adaptatives, ayant pourtant tendance à être déstabilisantes pour la structure protéique (Tokuriki et al. 2008; Soskine & Tawfik 2010).

Selon notre hypothèse de travail, le désordre intrinsèque pourrait être vu comme un déterminant direct de la diversité et de l'adaptabilité extraordinaire des virus à ARN. Alors que la stratégie de robustesse associée à l'hyper-structuration a largement été étudiée et directement reliée à l'adaptabilité, la robustesse mutationnelle propre au désordre intrinsèque manque encore de preuves expérimentales, les données de Tokuriki *et al.* ayant été obtenues à partir d'expériences *in silico*.

3. Objectifs de la thèse

Malgré l'accumulation de données sur les 15 dernières années plaidant pour une nécessité fonctionnelle flagrante du désordre dans la biologie virale, certaines études à haut débit ont récemment rapporté qu'il existe une variabilité considérable du niveau de désordre dans les virus (Pushker et al. 2013; Xue et al. 2012). Pourquoi certains virus présentent-ils autant de désordre dans leurs protéines alors que d'autres semblent au contraire privilégier l'ordre ? Cette démarche haut-débit présente une limite de par son incapacité à pouvoir interpréter de telles disparités. Ceci sous-entend la nécessité de ramener ces études fonctionnelles à une échelle plus réduite. Nous avons donc pris le parti d'appréhender le désordre et les fonctions qui lui sont associées (interaction avec l'hôte et évolution) à l'échelle plus restreinte d'un genre viral, dont la biologie est bien documentée.

Cette thèse a eu pour objectif de **mieux comprendre l'implication fonctionnelle du désordre dans les protéines virales**, et ce de deux façons :

i) Tout d'abord, à travers une étude *in silico* systématique et descriptive du désordre au sein du protéome des potyvirus, un genre viral pour lequel on ne disposait encore que de très peu de données sur le désordre intrinsèque des protéines et des fonctionnalités qui lui sont associées. L'objectif était de fournir une annotation fonctionnelle du désordre à l'échelle d'un genre majeur de virus de plante. Il s'agit d'une contribution originale dans la mesure où une telle étude intra- et inter-espèce n'avait jamais été entreprise chez des phytovirus en comparaison des nombreuses études déjà menées chez leurs homologues animaux. Ce travail fait l'objet du **Chapitre 1**.

ii) Puis en testant l'hypothèse de la fonction du désordre dans l'adaptation des virus à ARN chez ce même genre viral et ceci de deux manières complémentaires, via une analyse qualitative des données expérimentales disponibles dans la littérature (**Chapitre 2**) et via une approche expérimentale analytique *in vivo* (**Chapitre 3**).

4. Présentation du modèle expérimental

A) Les potyvirus : un genre majeur chez les virus de plantes

Les potyvirus constituent avec les bégomovirus les deux plus grands genres viraux décrits à ce jour chez les virus de plantes. Ce sont des virus à ARN simple brin de polarité positive (ARNss(+)), qui appartiennent à la famille *Potyviridae*. Cette famille, composée de 8 genres, regroupe, avec 158 espèces, plus de 80% des espèces de phytovirus répertoriées (http://www.ictvonline.org/virusTaxonomy.asp) (Adams et al. 2012). Les potyvirus sont d'une importance majeure tant en terme économique que scientifique, comme l'illustre la place que tiennent deux de ses membres, les virus *Potato virus Y* (PVY) et *Plum pox virus* (PPV) dans le top 10 des virus de plantes les plus importants (Scholthof et al. 2011). Ils constituent un problème majeur pour l'agriculture et l'horticulture, en causant des pertes de rendements importantes à travers le monde (Ward & Shukla 1991). Leur transmission par une large gamme d'insectes vecteurs rend leur expansion et leur dispersion très compliquées à contenir et à gérer (Nicaise 2014).

De nombreuses données concernant les différentes étapes du cycle infectieux des potyvirus, ainsi que les protéines intervenant dans chacune d'elles, sont à ce jour disponibles et ont fait très récemment l'objet de plusieurs revues (Ivanov et al. 2014; Mäkinen & Hafrén 2014; Revers & García 2015). Les principales étapes du cycle des potyvirus sont représentées Figure 15.
A l'inverse des virus animaux, le mode d'entrée des virus de plantes dans leur hôte se fait nécessairement via une blessure mécanique induite soit par la piqûre de l'insecte vecteur, soit par inoculation mécanique. Ce mode d'entrée qu'on pourrait qualifier de « brutal » semble être l'unique voie de passage du milieu acellulaire vers la cellule, aucun récepteur membranaire ou autre processus d'internalisation endocytaire n'ayant été découvert jusqu'à présent (Shaw 1999). L'étape suivante est la décapsidation de la particule virale, suivie de façon probablement simultanée par la traduction de l'ARN viral en une polyprotéine. Cette polyprotéine est ensuite clivée sous l'action de trois protéases virales en 10 protéines matures. Un onzième facteur est issu d'un décalage de lecture lors de la traduction de la polyprotéine. Une fois synthétisés, la majorité de ces facteurs viraux s'assemblent au niveau de vésicules cytoplasmiques pour former des complexes de réplication qui vont permettre la synthèse de de l'ARN viral en de multiples copies. Parmi ces génomes viraux néo-synthétisés, certains vont nouveau servir de matrice pour la traduction. D'autres génomes vont quant à eux pouvoir être transportés jusqu'aux cellules voisines via l'adressage des vésicules aux plasmodesmes (jonctions cellulaires des végétaux). Certains ARN viraux seront dégradés, ou encore encapsidés afin de former de nouvelles particules qui pourront être transmises lors d'une nouvelle blessure ou bien de l'acquisition par l'insecte vecteur.

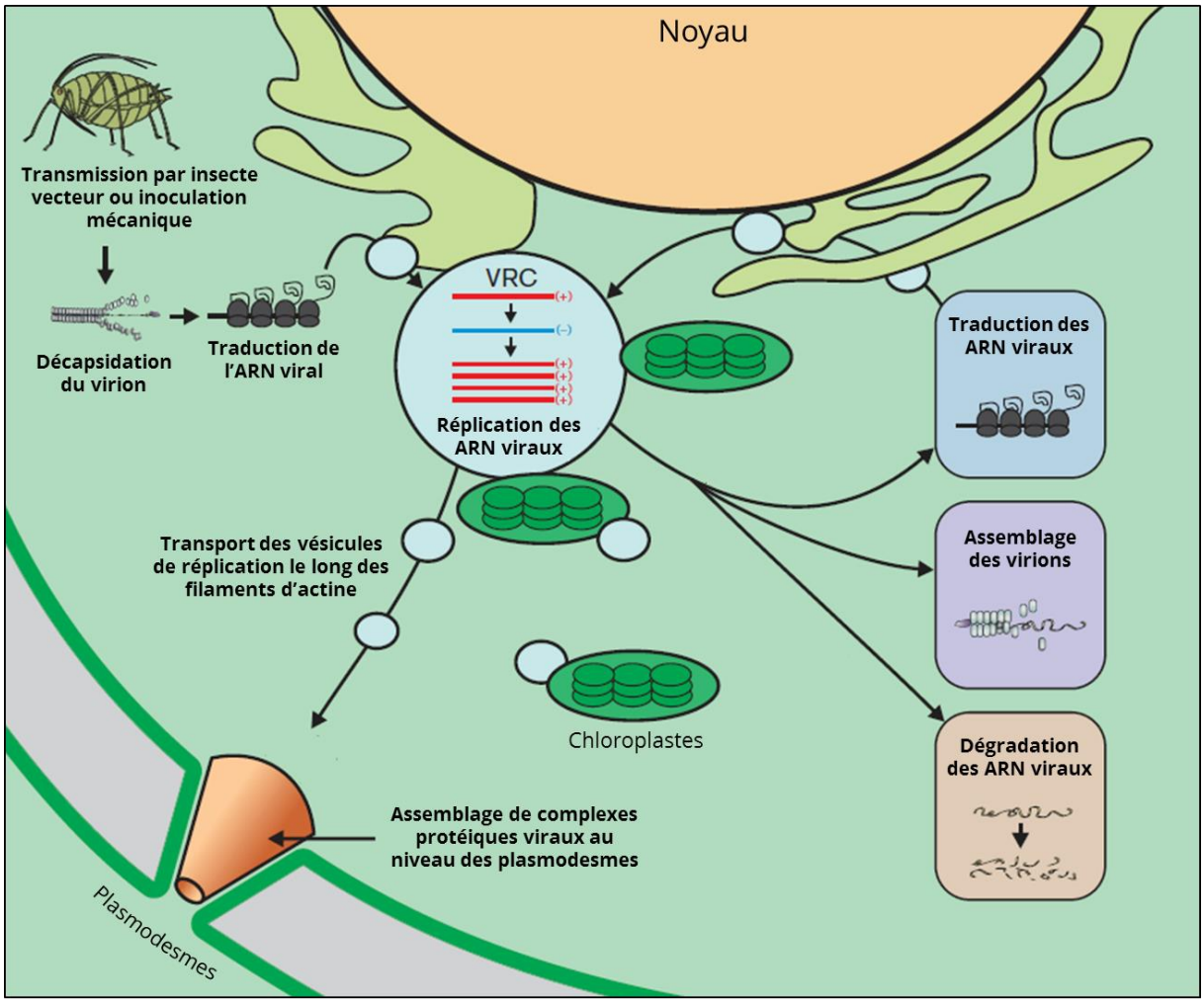


***Figure 15*|Etapes principales du cycle d'infection des Potyvirus.** Les particules virales (virions) entrent dans la cellule végétale suite à une piqûre d'insecte vecteur ou par blessure mécanique. La décapsidation est suivie de la traduction de l'ARN viral en une polyprotéine ensuite clivée en 10 protéines matures. L'ARN viral est ensuite répliqué au niveau de vésicules de réplication grâce à un complexe de réplication. Certains des ARN viraux synthétisés sont ensuite transportés le long des filaments d'actine jusqu'aux plasmodesmes pour permettre leur transport aux cellules voisines. Les ARN viraux néo-synthétisés peuvent également être traduit ou empaquetés pour former de nouvelles particules, tandis que d'autres seront dégradés. *VRC : viral replication complex. Adapté de* (Ivanov et al. 2014).

B) Les potyvirus : un modèle expérimental de choix pour l'étude des mécanismes adaptatifs viraux

L'utilisation des virus de plantes en évolution expérimentale

En évolution expérimentale, l'utilisation de virus de plantes – non-pathogènes pour l'homme - fournit l'avantage de ne pas être soumise au même degré de contraintes concernant les règles d'éthique et de biosécurité, extrêmement drastiques et limitantes, en vigueur dans le cas des virus animaux. Contrairement aux études d'évolution expérimentale conduites sur des modèles de virus animaux infectant des cellules en cultures, les expérimentations de virologie végétale peuvent être réalisées directement sur organismes entiers (hôtes naturels), en serre voire en plein champs.

L'utilisation des potyvirus comme modèle de virus à ARN (+)

L'étude des potyvirus peut permettre de répondre à des questions de portée assez générale sur les mécanismes liés à la compaction de génome et aux processus d'évolution et d'adaptation, chez les virus à ARN. Les virus à ARN (+) regroupent en effet de nombreuses espèces (Figure 11) dont l'étude constitue un intérêt majeur pour la santé humaine. Citons parmi eux le poliovirus et le virus de l'hépatite A (HAV) (*Picornaviridae*), le SARS-CoV agent causal du Syndrome Respiratoire aigu sévère (*Coronaviridae*), le virus de la fièvre jaune et le virus de l'hépatite C (HCV ; *Flaviviridae*).

Pour notre exploration de la contribution du désordre intrinsèque dans les processus évolutifs, la richesse bibliographique nous permet donc **i)** d'utiliser les résultats d'études d'évolution expérimentales conduites sur trois espèces différentes de potyvirus, le *Zucchini yellow mosaic virus* (ZYMV), le *Tobacco etch virus* (TEV) et le PVY, pour évaluer l'existence d'une corrélation entre désordre et apparition de mutations non-synonymes, au sein des populations virales **ii)** d'utiliser la connaissance du mécanisme de contournement (*i.e* adaptation) par le PVY de la résistance récessive médiée par le gène codant pour le facteur d'initiation de la traduction eucaryote 4E (eIF4E) chez le piment. Les déterminants de ce contournement sont localisés dans la région centrale désordonnée de la Viral genome-linked protein (VPg) (Moury et al. 2004; Ayme et al. 2006; Ayme et al. 2007). Ce modèle bien décrit nous a donc permis de tester expérimentalement si le désordre contenu dans la VPg peut être relié à ses capacités adaptatives.

Une présentation approfondie précisant le contexte et les objectifs correspondant à chacune des études composant ce travail est disponible au début de chaque chapitre.

Chapitre 1

Analyse *in silico* du désordre intrinsèque et des fonctions qui lui sont associées au sein du genre Potyvirus

Comme présenté en introduction de ce travail, le désordre intrinsèque a d'ores et déjà été associé à de nombreuses fonctions de reconnaissance moléculaire et présentent des caractéristiques évolutives qui le distingue des régions structurées.

Cette analyse *in silico* du désordre intrinsèque à l'échelle du genre *Potyvirus* présente donc deux objectifs :

i) Etendre l'examen des fonctions potentielles associées au désordre intrinsèque des protéines viral aux virus de plante, pour lesquels les rôles biologiques associés à cette caractéristique structurale sont très peu documentés.

ii) Contribuer à la compréhension de la biologie d'un genre majeur de phytovirus, en établissant des pistes pour nourrir la démarche expérimentale d'annotation fonctionnelle du désordre dans ces protéines.

Ce travail a donc été réalisé en estimant *in silico* le désordre intrinsèque chez 10 espèces de Potyvirus représentatives de la diversité de ce genre viral. Il a ainsi d'abord pu être observé qu'un taux moyen de désordre de l'ordre de 20% était retrouvé chez l'ensemble des protéomes des différentes espèces de potyvirus étudiées.
Les régions désordonnées conservées au sein des différentes espèces de potyvirus ont été examinées et leurs caractéristiques mises en perspectives par rapport aux fonctions habituellement associées au désordre intrinsèque. Cette démarche a permis d'émettre des hypothèses plausibles quant au rôle du désordre dans la biologie des potyvirus.

Les rôles du désordre dans des processus tels que la régulation du clivage protéolytique, l'adaptation ainsi que la reconnaissance moléculaire conduisant aux recrutements des facteurs de l'hôte ont fait l'objet d'analyses approfondies.

Ce travail a fait l'objet d'une publication dans la revue Molecular BioSystems.

Molecular BioSystems

PAPER

Cite this: *Mol. BioSyst.*, 2016, 12, 634

Protein intrinsic disorder within the *Potyvirus* genus: from proteome-wide analysis to functional annotation†

Justine Charon,*[ab] Sébastien Theil,[ab] Valérie Nicaise[ab] and Thierry Michon*[ab]

Received 9th October 2015,
Accepted 11th December 2015

DOI: 10.1039/c5mb00677e

www.rsc.org/molecularbiosystems

Within proteins, intrinsically disordered regions (IDRs) are devoid of stable secondary and tertiary structures under physiological conditions and rather exist as dynamic ensembles of inter-converting conformers. Although ubiquitous in all domains of life, the intrinsic disorder content is highly variable in viral genomes. Over the years, functional annotations of disordered regions at the scale of the whole proteome have been conducted for several animal viruses. But to date, similar studies applied to plant viruses are still missing. Based on disorder prediction tools combined with annotation programs and evolutionary studies, we analyzed the intrinsic disorder content in *Potyvirus*, using a 10-species dataset representative of this genus diversity. In this paper, we revealed that: (i) the *Potyvirus* proteome displays high disorder content, (ii) disorder is conserved during *Potyvirus* evolution, suggesting a functional advantage of IDRs, (iii) IDRs evolve faster than ordered regions, and (iv) IDRs may be associated with major biological functions required for the *Potyvirus* cycle. Notably, the proteins P1, Coat protein (CP) and Viral genome-linked protein (VPg) display a high content of conserved disorder, enriched in specific motifs mimicking eukaryotic functional modules and suggesting strategies of host machinery hijacking. In these three proteins, IDRs are particularly conserved despite their high amino acid polymorphism, indicating a link to adaptive processes. Through this comprehensive study, we further investigate the biological relevance of intrinsic disorder in *Potyvirus* biology and we propose a functional annotation of potyviral proteome IDRs.

Introduction

About 25 years ago, the emergence of the intrinsic disorder concept, which defines proteins or protein regions as devoid of stable and fixed secondary and/or tertiary structures under physiological conditions and in the absence of binding partners, started to challenge the classical view of protein structure/function relationship.[1–8] Contrary to the stable and unique structures found in ordered or globular proteins, intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) share the peculiarity to exist as a dynamic ensemble of conformers in the cell.[2–4,9,10] The key idea behind this discovery, which at the start puzzled many structuralists, was that protein intrinsic disorder had a functional relevance. In this sense, a growing number of studies report the role of disorder in multi-interactions and post-translational regulation processes,[11–14] signal transduction and control pathways, as well as the involvement in evolutive processes.[15–21]

This strong interest of the scientific community encouraged the development of robust *in silico* predictors to estimate the disorder content from amino acid (aa) sequences. This contributed over the last few years to a high-throughput proteome analysis that established the ubiquitous nature of intrinsic disorder in all domains of life.[22,23] Useful databases are now available, such as DisProt that lists today more than 700 IDPs and 1500 IDRs for which disorder prediction matches with experimental data.[24]

With between 7.3% and 75% residues being disordered, the proteome of viruses on the whole presents the highest variability of disorder in the living world.[22,23,25] Viruses are proven masters in genome compactness, as most viral proteomes usually encode for a few proteins that must ensure a diversity of functions through multiple interactions. Recently, intrinsic disorder has been experimentally related to these multi-interactions and binding promiscuity in several studies.[26–29]

Another aspect of protein disorder in the virus lifecycle concerns the extraordinary abilities of these microbes to quickly adapt to various environments. Low topological constraints

[a] *UMR Biologie du Fruit et Pathologie, INRA, Villenave d'Ornon cedex, France. E-mail: justine.charon@bordeaux.inra.fr, thierry.michon@bordeaux.inra.fr*
[b] *UMR Biologie du Fruit et Pathologie, Université de Bordeaux, Villenave d'Ornon cedex, France*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5mb00677e

This journal is © The Royal Society of Chemistry 2016

exerted on disordered regions of some viral proteins have already been experimentally associated with larger mutational tolerance.[30] This could explain in part how the RNA virus can cope with their high mutation rates (10^{-3} to 10^{-5} error per nt per replication cycle).[31] This mutational robustness is even more surprising given the commonly-used frameshift strategy in viral genomes and the multifunctionality of the proteins they encode. Consequent to this observation, it has been postulated that mutations could accumulate in IDRs without strong deleterious effects on protein stability and function, while facilitating viral adaptation[32–36] and the emergence of new functions.[37,38]

In this context, assessing the disorder content and distribution in viral species/families can help to better understand some key features of virus biology, such as adaptive processes leading to host resistance breaking, defense escape and new disease emergence. Overall, while bearing in mind the necessary cautiousness inherent to predictive approaches, addressing ''disordome'' and its functional implications in virus biology can result in finding common traits of disorder functionality. These findings could in turn feed prediction/annotation programs to better apprehend newly discovered viral families.

Functional annotations of disordered regions at the scale of the whole proteome have already been conducted for some well-studied animal viruses, like *Hepatitis C virus* (HCV),[39] *Human immunodeficiency virus-1* (HIV),[40] *Human papilloma virus* (HPV)[41] and more recently *Dengue virus*,[42] with other members of *Flavivirus* genus (*e.g. Yellow fever virus*, *Japanese encephalitis virus* and *West Nile virus*).[43]

To date, similar kinds of analyses applied to plant viruses are still missing. Nevertheless, strong experimental evidence for the presence of functional disorder in a reduced number of *Potyvirus* proteins has been recently reported.[29,44–48]

The genus *Potyvirus* represents one of the largest and most economically destructive genus of plant viruses. Because of their wide host range and spread/dispersion, *Potyviruses* are very hard to contain and manage, and cause dramatic losses in cultural crops worldwide.[49,50] Among *Potyviruses*, both *Potato virus Y* (PVY) and *Plum pox virus* (PPV) rank in the ''Top 10'' plant viruses of major social and economic impact.[51]

Potyvirus virions are flexuous filamentous particles that include a 10 kb positive-sense single-stranded RNA molecule ((+)ssRNA).[52–54] The viral protein genome-linked (VPg) is covalently attached at its 5′ end and a poly(A)-tail terminates at its 3′ end.[55–57] This genome contains untranslated or non-coding regions (UTR) at each of its ends that surround a single-ORF encoding a polyprotein ~3000 residues long. After translation, the polyprotein is proteolytically processed by three *Potyvirus*-encoded proteases P1, Helper component proteinase (HC-Pro) and Nuclear inclusion proteinase (NIa-Pro) into ten mature proteins.[58–62] In addition to the ten proteins ensuing from the polyprotein maturation, an eleventh protein, P3N-PIPO, is translated from a +2 ribosomal frameshift in the P3 sequence.[63] The replication step occurs within viral factories localized into intracellular specific membranes, where the nuclear inclusion b (NIb), the potyviral replicase, performs the synthesis of new genomic RNAs.[64] After the translation and replication processes, viral genomes can be addressed to different fates, comprising encapsidation into new particles, degradation, or cell-to-cell, long distance and vector-mediated movement. Advances in the understanding of processes underlying the regulation of those pathways are reviewed in ref. 65.

In this context, we propose here to examine the occurrence of intrinsic disorder in the entire *Potyvirus* genus and to assess its involvement in viral functions. Owing to the importance of protein intrinsic disorder in viral functions, performing this analysis on the whole plant virus genus is worth doing. In this work, we assessed for the first time the disordome of *Potyvirus*, through *in silico* characterization of intrinsic disorder at the whole genus level. In addition, an attempt was made to identify biologically-relevant IDRs and establish a functional annotation of these regions as part of the study of *Potyvirus* biology.

Materials and methods

Potyviral sequence dataset

Potyviral full length genomic sequences of the 10 *Potyvirus* species considered were retrieved from the National Center for Biotechnology Information (NCBI) resource (http://www.ncbi.nlm.nih.gov/). The 5′ and 3′ non-coding regions (UTRs) were discarded and the remaining genomic sequences were translated into polyproteins using MEGA6.0 software.[66] Sequences containing ambiguous characters were curated. Due to d*N*/d*S* calculation requirements, datasets of more than 40 sequences per species were curated to remove the redundant or almost identical sequences for preserving the database diversity. A total of 288 non-redundant polyproteins was thus obtained (listed in Table S1, ESI†).

To extract individual protein sequences, polyprotein nucleotide sequences were aligned and each open reading frame (ORF) was retrieved from its respective GenBank annotations. ORFs were then translated into protein sequences. A different process was conducted to build the P3N-PIPO dataset, given its intra-species length polymorphism.[67] For each species, the sequence ends were manually adjusted. P3 and P3N-PIPO sequences were not included in the d*N*/d*S* study, because ribosomal frameshifts introduce too much bias for this type of evolutionary sequence analysis.

Determination of disorder in *Potyvirus* proteins

Disorder prediction. Disorder was predicted both in polyproteins and in each individual proteins using PONDR-VLXT® predictor,[68] available on the Disprot database resource (www.disprot.org).[24] This neuronal learning-based predictor employs algorithms using a set of features based on biological knowledge. It was trained on disorder data derived from either X-ray crystallography or nuclear magnetic resonance. It is particularly suitable for proteins that share ordered and disordered regions at the same time, which is typical for viral proteins. This predictor is also very accurate in the detection of sites involved

in molecular recognition, signaling and regulation. There are generally good correlations between PONDR-VLXT® predictions and experimentally probed disorder. It is the case for the potyviral VPg, for which predictions are in very good agreement with the experimental characterization of the disorder properties of VPgs originating from several potyviral species.[29,45,46] PONDR-VLXT® predictions are also in agreement with the experimental disorder characterization of the *Potato virus A* (PVA) CP.[48] The software uses a sliding window (set to 7 amino acid length in this study) to associate a disorder score with each residue. Residues with a score below 0.5 were considered as ordered and residues with a score higher than 0.5 are considered as disordered.

Mean disorder content and conservation calculation. For each viral sequence within a given species, the disorder/order state was determined at each residue position. After intra-species sequence alignment, this disorder/order state was compiled. An average score at each amino acid position along viral sequences representative of all sequences was obtained and expressed as a percentage of disorder state conservation at each position. Heatmaps of these conservation scores were built for each individual protein within the 10 species datasets using Plotly software (https://plot.ly/).

Segmenting the protein sequences into ordered and disordered regions. A comparison of the evolutionary rates associated with ordered and disordered states was initiated by separating ordered and disordered regions. Within each of the 10 species datasets, PONDR-VLXT® disorder scores obtained for all sequences at each amino acid position were used to calculate average scores resulting in a consensus disorder prediction profile along the polyprotein. In the 10 consensus profiles, ordered and disordered regions were defined as a minimum of 5 consecutive amino acids displaying disorder scores >0.5 and <0.5 respectively. The resulting segmentation patterns were used to cut all the polyprotein sequences within each species datasets.

d*N*/d*S* ratio calculations

To evaluate the evolutive constraints exerted on each IDR and ordered regions, we determined d*N*/d*S* ratios, defined as non-synonymous nucleotide substitutions (resulting in amino acid changes) *vs.* synonymous nucleotide substitutions (no amino acid changes). To this end, only domains encoded by nucleotide sequences of at least 60 nucleotides long were retained to maintain the analysis robustness.[69] The PARtitioning approach for Robust Inference of Selection (PARRIS) method[70] was used (default parameters). It is part of the Hyphy package[71] which is available on the website www.datamonkey.org.[72,73] This method is particularly suitable for viral sequence analysis, which often results from recombination events[74] that could lead to misinterpretation of phylogenetic results.[75] Finally, the ω value (corresponding to mean d*N*/d*S* ratio) is inferred from all sequences. Mean d*N*/d*S* values are classified into ordered and disordered domains. A Kruskal–Wallis non-parametric statistical test was applied to mean d*N*/d*S* values obtained from IDRs and ordered regions to detect the significant difference in d*N*/d*S*-values between those 2 classes.

Phylogenetic analysis

To infer the phylogenetic tree of *Potyvirus*, a polyprotein amino acid sequence per species was randomly chosen. They were aligned using the high speed multiple sequence alignment program MAFFT (Multiple Alignment using Fast Fourier Transform)[76] with default parameters. Two outgroup sequences (*Agropyron mosaic virus* and *Hordeum mosaic virus*) belonging to the *Rymovirus* genus were also used for multiple alignment, according to ref. 77. The phylogenetic tree was inferred from polyprotein alignment using the Maximum Likelihood method, using MEGA6.0 software. The resulting phylogenetic tree was customized using FigTree software (http://tree.bio.ed.ac.uk/software/figtree/).

Functional annotations of *Potyvirus* proteins

Amino acid diversity score calculation. Amino acid diversity of protein alignment was determined using Shannon entropy measure, available on the Protein Variability Server (http://imed.med.ucm.es/PVS/).[78] It consists of assigning a diversity score H to each alignment position. Typically, a position is considered as variable when the parameter H is higher than 2 and conserved when H is below 2. Shannon's scores were displayed for each species dataset as heatmaps. A blue (low H-score) to yellow (high scores) color gradient was used to visualize the diversity along protein sequences, using the Plotly program.

Determination of MoRFs. The molecular recognition features (MoRFs) are predicted from amino acid sequences, using the MoRFPred tool[79] (http://biomine-ws.ece.ualberta.ca/MoRFpred/), an algorithm fed with a large dataset of annotated MoRF regions. MoRF predictions were done on each protein sequences and aligned. MoRF signals corresponding to less than 5 consecutive residues were considered as not biologically relevant and were discarded. The intra-species MoRF conservation was determined, and represented as a color gradient (green boxes in protein heatmaps, Fig. S2, ESI†).

Determination of ELMs. To assess the abundance of motif usage in the world of viruses, an *in silico* prediction of the occurrence of 173 known functional ELMs was attempted in a dataset of 2208 non-redundant viruses belonging to various groups and including the 10 *Potyvirus* species used in our study.[37] Given the difficulty to predict motifs with biological relevance, only motifs occurring in IDRs were considered. In order to identify potential new functions/interactions associated with intrinsic disorder in potyviral proteins, we extracted ELM predictions obtained for *Potyviruses* from the whole viral ELM dataset provided by Hagai and co-workers. High probabilities of occurence are inherent to the low complexity of many ELMs. However, given their high biological relevance in the viral context, some of these motifs could not be discarded. Motif conservation strongly correlates with functionality.[80] Consequently, this criterion was used and we applied a 80% conservation cut-off to discriminate the predicted biologically relevant ELMs from false-positive ones. Potentially relevant ELMs are listed in ESI,† Table S2.

Manual annotation of functional domains and proteolytic cleavage sites. Functional annotations of polyproteins and

 This journal is © The Royal Society of Chemistry 2016

proteins were retrieved for each species both from the literature and the UniProtKB website (http://www.uniprot.org/uniprot/).

Results and discussion

Assessment of disorder abundance in the *Potyvirus* genus

To assess whether protein intrinsic disorder represents a strategy employed by *Potyviruses* that would be selected during their evolution, we considered 10 *Potyvirus* species representing the diversity of the whole genus,[77] and for which more than fifteen full-length genome sequences were available on the NCBI GenBank resource (see Materials and methods section). The chosen species and their respective sequence accessions are given in the ESI† (Table S1). For each species dataset, intrinsic disorder was first predicted at the polyprotein level (Fig. 1) using PONDR-VLXT® as a robust predictor (see Materials and methods section). The calculated "average number of disordered residues per polyprotein" is close to 20%, whatever the inter-species phylogenetic distance be. Importantly, starburst radiation occurring during *Potyvirus* evolution[81] prevented us for going deeper into the study of phylogenetic dynamics of intrinsic disorder in the genus. However, the disorder frequency in *Potyviruses* seems to be not directly related to the evolutive history of this group, as no disorder enrichment or depletion seems to be correlated to any evolutive lineages (Fig. 1A). The disorder content was homogenous inside each species, as shown by low mean root square deviations. This was also observed at the inter-species level (Fig. 1B). Such a homogeneity is in contrast to the high variability of disorder content observed in some other viral families and despite the long-scale evolutive divergence of the *Potyvirus* genus.[25,81] Interestingly, a higher propensity of intrinsic disorder is observed for *Lettuce mosaic virus* (LMV) species and results from the particular contribution of the viral protein P3 (discussed in the "P3 and P3N-PIPO contain constrained disordered regions" section). Overall, these results indicate that *Potyvirus* belongs to the group of viruses displaying high disorder in their proteome, similarly to other (+)ssRNA viruses.[22,23,25]

Given the high diversity of this viral genus, the conservation of intrinsic disorder during *Potyvirus* evolution strongly suggests an evolutive hallmark selected during *Potyvirus* evolution and thus a functional advantage of this structural feature.

To assess the intrinsic disorder distribution inside *Potyvirus* proteomes, polyproteins were segmented into their individual protein sequences and PONDR-VLXT® predictor was used to predict the disorder content in the eleven mature potyviral proteins. A box plot representing the disorder variability for nine of the eleven proteins at the inter-species scale is plotted in Fig. 2A, while the variation at the intra-species scale is figured in Fig. 2B. Except few (P1 from PPV and P3/P3N-PIPO from LMV), all potyviral species share the same distribution of intrinsic disorder for each protein (Fig. 2A and B). Based on a previously described method,[82] the proteins can be classified into three groups according to their mean disorder propensity (Fig. 2A). NIa-Pro and NIb replicase are classified as "structured" with an overall disorder content of less than 10%. HC-Pro, P3, P3N-PIPO and CI are classified as "moderately disordered" with an intermediate disorder content comprised between 11 and 30%. Finally, VPg, P1 and CP have an overall disorder content globally higher than 30% and are classified as "highly disordered" (Fig. 2A). VPgs from LMV, PVA and PVY were previously experimentally proven as disordered proteins, as well as the CP from PVA.[29,44–48] These data validate the strength of our prediction for these two proteins. On the basis of the disorder analysis in our dataset, it is more than likely that the disorder features of these viruses can be extended to the other members of the *Potyvirus* genus. This legitimates the hypothesis of a functional assignment of disorder in *Potyviruses*.

Disorder involvement in molecular functions, multi-interactions and adaptation

No direct correlation was identified so far between the disorder content and functions in potyviral proteins: for instance, the 5 proteins carrying enzymatic functions (P1, HC-Pro, CI, NIa-Pro, NIb) have variable disorder content (from P1 highly disordered to NIb highly ordered, Fig. 2A).

It was hypothesized that intrinsic disorder could confer to a single domain the necessary plasticity to fulfill several different functions.[83–85] Nevertheless, using data on the *Potyvirus* protein interactome reviewed in ref. 86, we observed that the multi-interaction ability of potyviral proteins is not correlated with their mean disorder content ($R < 0.01$; P-value > 0.8) (data not shown).

Intrinsic disorder has been related to adaptive abilities of RNA viruses to escape host immunity responses to breakdown host resistances or enlarge their host specificity.[16,30,37] Interestingly, potyviral VPg, P1 and CP, proteins classified as highly disordered (Fig. 2A), have already been associated with adaptive events (reviewed in ref. 87). It prompted us to perform a deeper analysis of this adaptation–disorder relationship. However, the examination of published data on some plant virus virulence factors, already defined as directly related to adaptive processes, did not give any correlation between these factors and their mean disorder content. Additionally, no clear relationship was observed between evolutive constraints and the disorder content in each protein considered as a whole, as reported with their respective calculated dN/dS ratio (or ω) at the intra-species level (Fig. S1, ESI†).

To conclude on this part, no correlation can be made so far between the whole-protein disorder content and the general molecular and biological functions. This led us to perform a deeper analysis of intrinsic disorder along the protein sequences with the aim of identifying specific conserved disordered regions and discuss their possible involvement in the known and sequence mapped interactions.

Functional annotation of IDRs in *Potyviruses*

***Potyvirus* protein IDRs evolve faster than ordered regions.** We reasoned that because of their multiple roles during plant infection, *Potyvirus* proteins could be viewed, in first instance, as tandems of functional modules. Consequently, we conducted

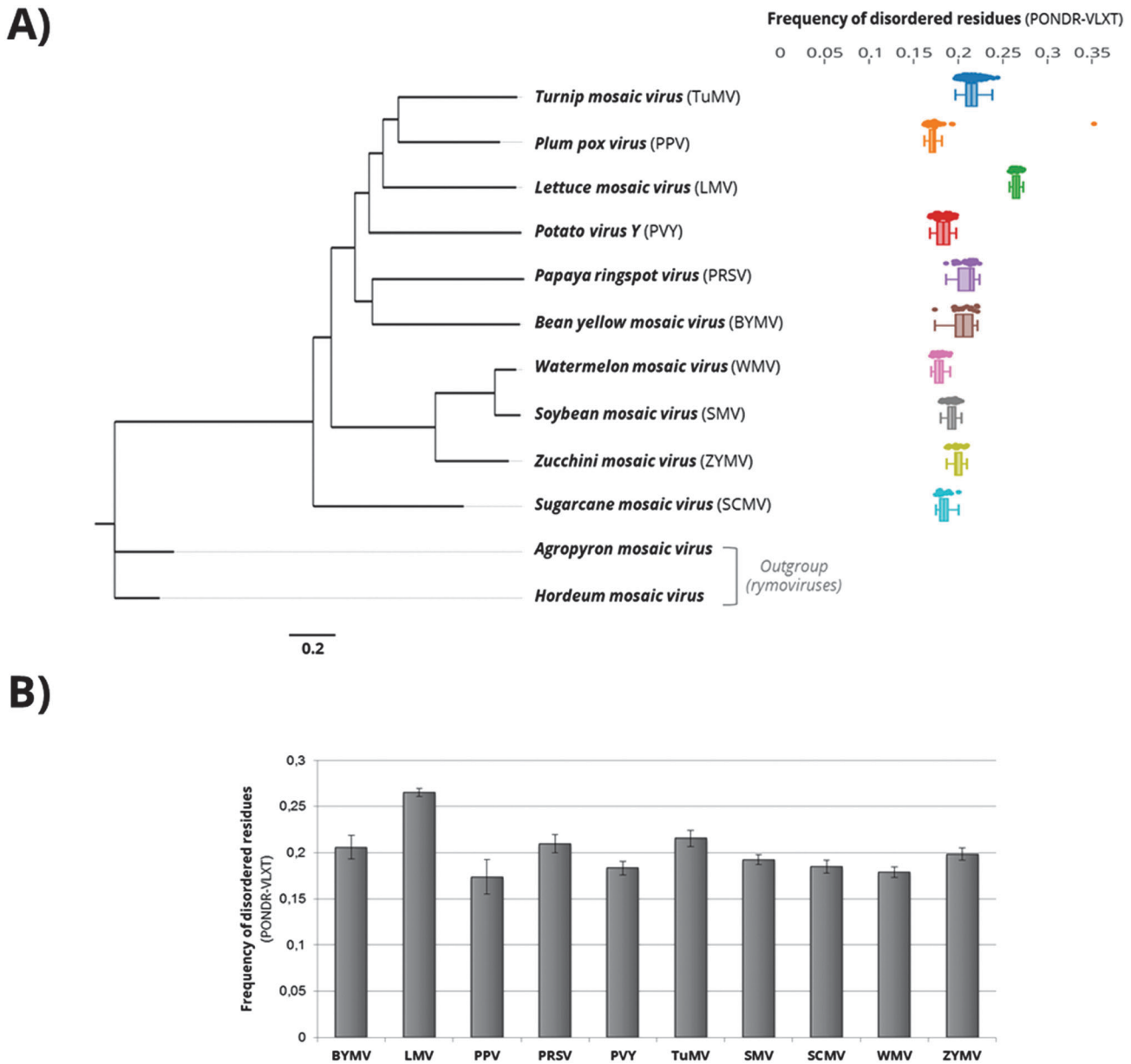


Fig. 1 Intrinsic disorder conservation in the proteome of 10 *Potyvirus* species. (A) Phylogenetic tree of potyviral polyproteins used in this study. Box plots represent variation of residue frequencies predicted as disordered (PONDR-VLXT) among each *Potyvirus* polyprotein dataset. Dots represent disorder frequency calculated for each polyprotein of the dataset used to construct the plot. Tree scale bar correspond to the number of substitutions per site. (B) Bar chart representation of mean disordered residue frequency for each *Potyvirus* proteome (BYMV, *Bean yellow mosaic virus*; LMV, *Lettuce mosaic virus*; PPV, *Plum pox virus*; PRSV, *Papaya ringspot virus*; PVY, *Potato Virus Y*; TuMV, *Turnip mosaic virus*; SMV, *Soybean mosaic virus*; SCMV, *Sugarcane mosaic virus*; WMV, *Watermelon mosaic virus*; and ZYMV, *Zucchini yellow mosaic virus*).

a comparison of the evolutive constraint (d*N*/d*S*) between ordered and disordered regions along the polyprotein of each virus species. Polyproteins were segmented into ordered and disordered regions according to PONDR-VLXT® predictions. The d*N*/d*S* ratio was calculated for all the sequences of at least 20 residues long (see details in the Materials and methods section). Globally, in *Potyviruses*, IDRs display significantly higher d*N*/d*S* values than the ordered ones (Fig. 3), which indicates the true tendency of intrinsically disordered domains to evolve faster than more structured regions during *Potyvirus* evolution. Such weaker evolutive constraints on IDRs strongly suggest the abilities for a faster and easier exploration of adaptive solutions and the emergence of new functions in viral proteins.[30,37] Nevertheless, some potyviral IDRs display a low d*N*/d*S* (Fig. 3), illustrating the concepts of "constrained" and "flexible" disorder discussed below in the section "*Amino acid polymorphism*".

Disorder and proteolytic processing of the potyviral polyprotein. The *Potyvirus* polyprotein is proteolytically processed into ten mature proteins by the three *Potyvirus*-encoded proteases P1, HC-Pro and NIa-Pro[59,61,62] (Fig. 4A). The P3N-PIPO

 This journal is © The Royal Society of Chemistry 2016

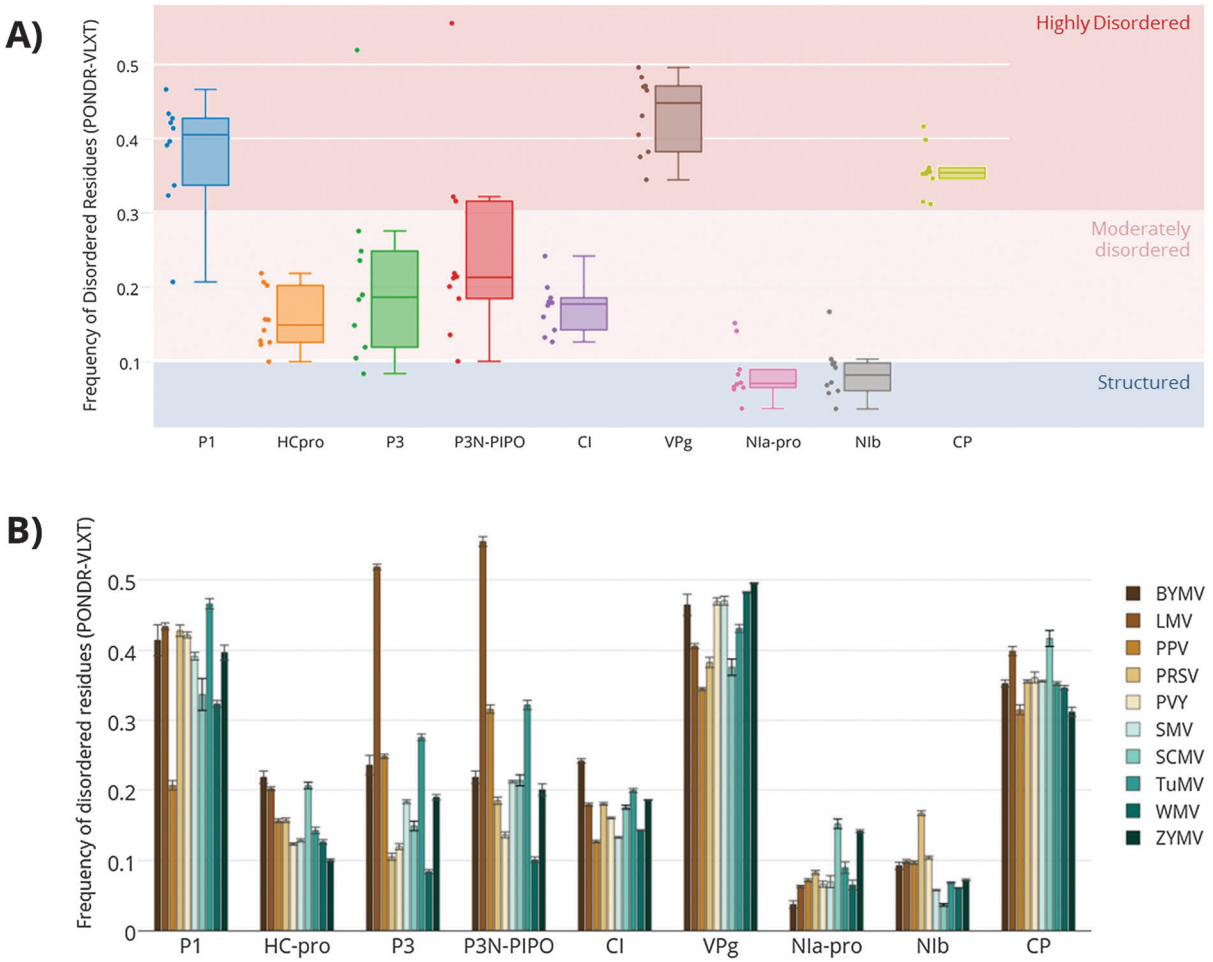


Fig. 2 Conservation of the intrinsic disorder amount in *Potyvirus* proteins. (A) Box plot representation of mean variation in disordered residue frequency (predicted with PONDR-VLXT) for each protein per *Potyvirus* species. Proteins are classified depending on their total disorder content: 0–10% are considered as highly ordered (blue), 11–30% as moderately disordered (white) and 31–100% as highly disordered (red). (B) Bar chart representation of mean disordered residue frequency for each proteins of each species. (BYMV, *Bean yellow mosaic virus*; LMV, *Lettuce mosaic virus*; PPV, *Plum pox virus*; PRSV, *Papaya ringspot virus*; PVY, *Potato Virus Y*; TuMV, *Turnip mosaic virus*; SMV, *Soybean mosaic virus*; SCMV, *Sugarcane mosaic virus*; WMV, *Watermelon mosaic virus*; and ZYMV, *Zucchini mosaic virus*).

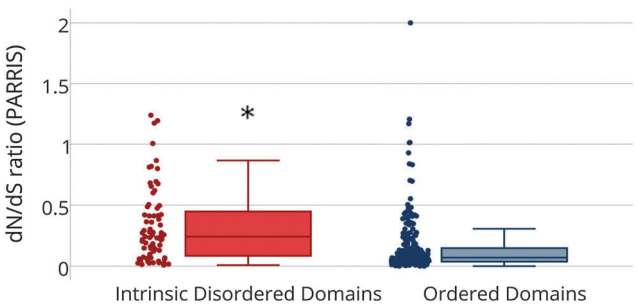


Fig. 3 Evolutive constraints (d*N*/d*S* ratio) exerted on intrinsic disordered domains and ordered domains of *Potyvirus* proteins. Box blots are constructed, respectively, with mean d*N*/d*S* data obtained both for IDRs (red box plot and dots) and ordered domains or OD (blue ones). Dots represent each d*N*/d*S* value obtained. Star represents the statistical significance of difference between mean OD and IDR d*N*/d*S* (p-value < 0.001).

protein results from a translation frameshift and cannot be considered as a proteolytic product.[63] P1 and HC-Pro are self-cleaved at their C-terminal (C-ter) end and NIa-Pro cleaves at 7 different sites in the polyprotein central and C-ter regions[88] (Fig. 4A). The *Potyvirus* genome strategy to encode for a single ORF has been related to a way of tuning the protein function through a sequential proteolytic process during the viral cycle.[58] Indeed, the functional relevant intermediary forms have been identified as key players in viral infection and adaptation. The 6K2-VPg-NIa-Pro intermediate is anchored to the ER membrane during *Potyvirus* replication.[89–91] The 6K2 proteolytic separation from CI and VPg modulates the viral replication rate and movement.[58] Deleterious effects on the viral fitness of NIb relocation at different positions in the polyprotein also highlight the functional relevance of protein placement within the polyprotein and the importance of transient intermediary forms in the viral cycle.[92] The cleavage efficiencies by NIa-Pro at each of its cleavage sites was investigated. Most proteolytic sites were fast processed (6K1/CI, 6K2/VPg, NIa-Pro/NIb and NIb/CP) but three of them, namely P3/6K1, CI/6K2 and VPg/NIa-Pro, were processed at a lower rate.[58] These data strongly suggest that differences in cleavage efficiencies at various NIa-Pro sites are

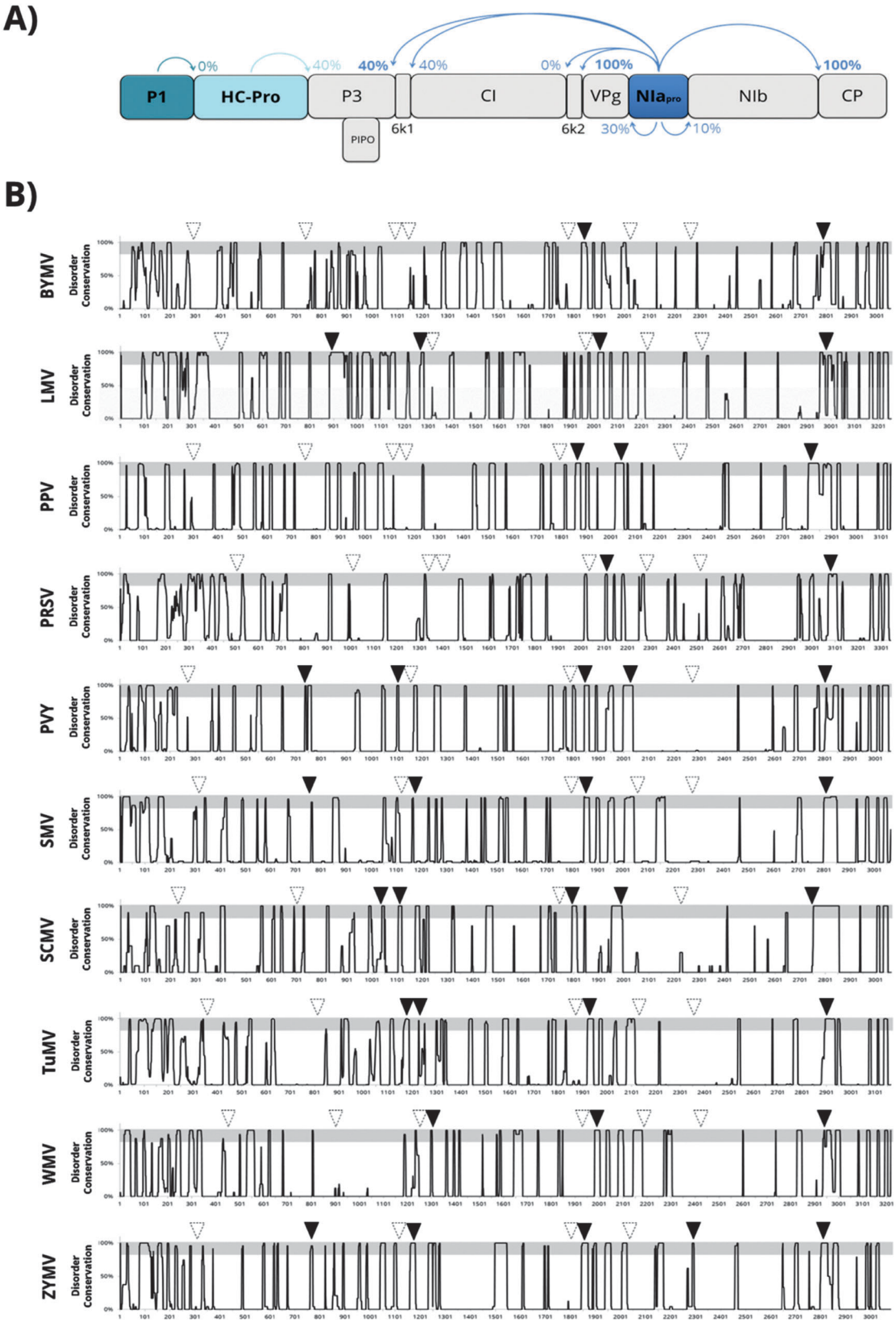


Fig. 4 Disorder status of *Potyvirus* polyprotein cleavage site regions. (A) *Potyvirus* polyprotein representation and cleavage sites. Values represent the percentage of disorder conservation at the level of each cleavage site for the ten *Potyvirus* species. Cyan, light and dark blue arrows indicate sites processed by P1, HC-Pro and NIa-pro respectively. (B) Polyprotein displayed a disorder landscape. Disorder conservation scores obtained along the polyprotein for each species dataset (see Table S1, ESI† for sequence accessions). The grey shaded horizontal bar indicates the 80% cut-off used to define positions where disorder is conserved at the intra-species level. Triangles indicate cleavage sites that are in predicted disordered (dark ones) and ordered regions (white and dotted ones).

 This journal is © The Royal Society of Chemistry 2016

of strong functional relevance. A time-dependent fine-tuning of the polyprotein maturation is likely to be important for the virus biology. For instance, the post-translational phosphorylation of NIa-Pro was demonstrated to inhibit the *trans*-proteolytic cleavage of VPg-Pro.[93]

The site accessibility to the enzyme strongly influences the proteolytic cleavage efficiency. It is well established that intrinsic disordered regions undergo faster proteolytic digestion than the more structured ones.[94–97] In viral polyproteins, the propensity to disorder in the vicinity of cleavage sites has already been predicted for *Hepatitis C virus*[39] and *Dengue virus*. Remarkably, our comparative analysis of the disorder propensity at the inter-species scale shows that the NIb-CP cleavage site is located in a disordered segment in all species (Fig. 4B). The conservation score of the intrinsic disorder state along the polyprotein was also determined for the ten *Potyvirus* species dataset (Fig. 4B). Disorder was considered as conserved intra-species when it was observed in at least 80% of the sequences (Fig. 4B). P3/6K1, CI/6K2 and VPg/NIa-Pro proteolytic sites, which were demonstrated as slowly processed,[58] share a low score of disorder conservation between 0% and 40% (Fig. 4A). By contrast, disorder is conserved in some of the other NIa-Pro cleavage sites with up to 100% conservation in 6K2/VPg and NIb/CP sites (Fig. 4B). The high release rate observed for the CP[58] is directly correlated with high conservation of disorder at this cleavage site and supports well the hypothesis of protease-sensitivity of unstructured/flexible regions. The genome encapsidation requires a large pool of monomeric CP units, and it is likely that a fast release of CPs facilitates the assembly. In addition, it was recently proposed that the pool of free CPs could finely tune the steps of genome translation and encapsidation.[98] This seems to illustrate the evolutive constraint exerted on the disorder status because of its role as the proteolytic "facilitator". CI-6K2 is a stable intermediate form *in vivo*, and has been associated with replicative functions in the viral cycle. It is noteworthy that the cleavage site is part of a conserved ordered region in 100% of the *Potyviruses* considered here (Fig. 4A and B). The prevalence of a structural order in this proteolytic region could slow down the proteolysis by NIa-Pro, resulting in a longer availability of this intermediate required for the viral cycle. Conversely, despite the relatively low disorder-state conservation of NIa-Pro/NIb sites among *Potyviruses*, the proteolytic release of the corresponding intermediary forms is fast.[58] In this case, intrinsic disorder does not seem to be related to any proteolytic facilitation. To conclude, only partial correlations can be established between order/disorder states of the cleavage sites within the polyprotein and their processing kinetics by NIa. It is likely that the order–disorder balance in the vicinity of the polyprotein maturation sites provides a kinetic modulator participating in the necessary chronologic apparition of various functional intermediates during the potyviral cycle. In this respect, the self-release of P1 from HC-Pro requires a special discussion (see the "P1: polymorphism, but conserved IDRs" section).

Protein-by-protein analysis of IDR functions

Although it is difficult to associate the predicted conserved IDRs with known biological processes and functions, the task deserves to be done, as structural disorder proves to be a main strategy for many viral functions.[27,35,39] To this end, a functional annotation of the potyviral IDRs found within each protein was performed through an analysis of: (i) their amino acid polymorphism, (ii) their content in recognition-binding motifs and, (iii) their content in eukaryotic functional linear motifs. The relationship between these three parameters and IDRs is presented as follows.

Amino acid polymorphism analysis. In many organisms, IDRs are more tolerant to aa changes than structured regions, a statement that seems to be also the case for *Potyviruses* (Fig. 3). Bellay and colleagues proposed that it allows sorting conserved disordered regions into two classes according to their evolutive rates: "flexible disorder" for disordered regions which display relatively high amino acid polymorphism (AAP) and "constrained" disorder which refers to disordered regions that are not variable in their aa sequences.[19] Flexible disorder is, in the first instance, found in flexible spacers between functional domains.[99] In addition, flexible disorder is particularly observed in proteins with one or two regions interacting sequentially with several partners through conformational switches, termed "date Hubs".[19,100,101] Constrained disorder, by contrast, is found in IDRs, whose function is linked to more topological requirements. This is the case of multi-interface hubs (named "party hubs"),[100,102,103] that can be involved in simultaneous binding with various partners. Therefore, an analysis of sequence conservation in *Potyvirus* IDRs is expected to give insights on underlying interaction modes.

MoRF analysis. Molecular recognition features (MoRFs) are defined as short motifs (5 to 25 residues long) embedded in IDRs that undergo disorder-to-order transition upon binding to a partner, according to the induced-folding process.[104,105] MoRFs can be classified into three classes, depending of the secondary structure resulting from the folding upon interaction: α-MoRFs fold into α-helices, β-MoRFs into β-sheets and coil-MoRFs into random coils. Functional MoRFs are frequent in viral IDRs.[106] Sequence characteristics (charge, hydrophobicity, conservation and secondary structure properties) allow the prediction of potential MoRFs and help their functional annotation.[79,107,108] This analysis was performed on our potyviral protein datasets using the MoRFPred predictor,[79] to detect highly-conserved MoRF motifs potentially involved in interactions. Their relevance and the possible involvement in *Potyvirus* biology is discussed on a "case-by-case" basis in view of already documented interactions and the evolutive conservation they display at the intra-species level.

ELM analysis. Eukaryotic Linear Motifs (ELMs) consist of 3 to 10 residues long motifs, which are ubiquitous in eukaryotic organisms. According to the ELM database resource[109,110] (http://elm.eu.org/), they are involved in many interaction-based processes and divided into six functional classes: ligand-binding sites, post-translational modifications, targeting sites, proteasome degradation targeting sites, docking sites and proteolytic cleavage sites. ELMs are often located within IDRs,[111] which, with their flexibility, facilitate the ELM-based interaction.[13,37,112,113] Being crucial in the emergence of new functions during

eukaryote evolution, ELMs constitute probably a "Achilles heel" since they have been mimicked by viruses during their evolution.[113,114] The use of such eukaryotic short motifs to hijack the host cellular machinery may explain, at least partially, how viruses establish so many interactions despite their high genome compactness. To date, 46 ELM functional classes among the 219 of the ELM database have been identified in the proteome of viruses from various families, suggesting that ELM mimicry is a common strategy in the world of viruses.[113] To assess the abundance of ELM usage in the world of viruses, an *in silico* prediction of the occurrence of 173 known types of functional ELMs was attempted in 2208 not-redundant viruses of diverse groups, including the 10 *Potyvirus* species used in the present study.[37] Therefore, we selected the potyviral ELMs located in IDRs from the data produced by Hagai *et al.*, in order to identify potential new functions/interactions associated with intrinsic disorder. Given the low complexity and high degenerative nature of these motifs, false positives are expected. In this respect, the localization of conserved ELMs in IDRs constitutes a biologically relevant criterion to discriminate true positive motifs. All motifs predicted in potyviral polyproteins share a high probability of occurrence by chance. However, given their high biological relevance, some of these motifs could not be discarded. Thus, we applied a "conservation cut-off" to discriminate biologically relevant ELMs from false-positive ones. Because motif conservation strongly correlates with functionality,[80] this criterion was used to identify the strong relevant ones. Consequently, all predicted motifs found in at least 80% of the polyprotein IDRs from the 10 species considered are listed in Table S2 (ESI†). Among these motifs, only those which were found relevant regarding plant-virus biology were considered (bold green boxes in Table S2, ESI†) and discussed.

P1: a highly disordered and variable protein

All potyviral P1 display long and conserved IDRs. The P1 protein, localized to the N-terminus (N-ter) of the potyviral polyprotein, is a serine endopeptidase. It is self-released by *cis*-cleavage at its C-ter.[61] A *trans*-activity was also reported in *Tobacco etch virus* (TEV).[115] At the inter-species level, disorder was observed in all P1 (Fig. 2). However, P1 has evolved through recombination and duplication events[116] and displays the highest variability both in the length (30–63 kDa) and in the aa sequence of all potyviral proteins. This prevents the alignment of P1 sequences at the inter-species level. Therefore, such a comparison of the distribution of disorder along the protein was not possible. At the intra-species level, long and conserved IDRs were identified (Fig. 5), suggesting that the conservation of P1 intrinsic disorder in the course of *Potyvirus* evolution is associated with biological functions.

P1 IDRs could contain RNA binding sites. RNA and DNA binding have often been reported as associated with the basic positively charged IDRs,[117] which act as binding facilitators.[118] The predicted central long-IDR of *Turnip mosaic virus* (TuMV) contains a broad specific RNA binding domain, spanning residues 150 to 168, which was experimentally confirmed.[119] It is noteworthy that, although highly diverse in sequence at the inter-species level, the first 180 aa of P1 conserve a high averaged isoelectric point value of 9.2, supporting the presence of a nucleic acid binding function. Hence, IDRs in P1 could favor RNA-binding and participate in viral replication and translation processes. TEV P1 physically interacts with the host 80S cytoplasmic ribosomes in infected cells and stimulates the translation of viral proteins *in vitro*.[120] The RNA-binding ability of P1 IDRs may be related to the RNA-chaperone activity, for which intrinsic disorder is an hallmark.[121] Moreover, such a RNA chaperone activity was already reported in other viral proteins.[122]

N-ter IDR is expected to modulate P1 proteolytic activity. The P1 N-ter domain modulates its proteolytic self-release from the polyprotein, thereby preventing early host defense responses, that would be detrimental to virus systemic infection.[123] The auto-inhibitory effect of intrinsically disordered domains was already reported.[124] Thus, IDRs within the N-terminal part of P1 could fold back to interact with the distant cleavage site at the C-ter end of the protein, slowing down its release. Post-translational modifications, such as phosphorylation, and/or its interaction with another viral/host protein,[125] could act as an activator switch.

MoRF in P1. Some well-conserved MoRFs were found embedded in P1 disordered regions. One of them, located in a C-ter IDR, was shared by 6 of the 10 *Potyvirus* species considered in this study. However, as there is no functional annotation available for this part of the protein, no putative function could be associated with these motifs.

P1 disorder and adaptation. P1 is the most variable potyviral protein in terms of sequence length and AAP. Its involvement in the adaptive process and host range specificity was previously reported[116,126,127] and is strongly suggested by its high mean d*N*/d*S* values (Fig. S1, ESI†). Interestingly, most of the P1 IDRs are well-conserved at the intra-species level (white to red bars, Fig. 5) despite the relatively high diversity of their aa sequence (blue to yellow gradation, Fig. 5). This strongly suggests that: (i) there is an evolutive constraint for the conservation of the disorder state at these positions that is likely of functional relevance and (ii) these regions are prone to aa switches occurring during virus evolution. This kind of evolutive behavior in IDRs has been referred to as "flexible disorder",[19] which relates to regions binding different partners sequentially. The high variability of sequence in P1 (considered as the least conserved protein among *Potyviruses*), associated with its high degree of disorder, strongly supports the hypothesis of mutational robustness and lower-evolutionary constraint effect related to disorder, a feature already observed in various proteins.[128,129] This could be related to adaptive abilities to various environments,[16] a disorder-based feature already discussed for RNA-viruses.[30] In this context, P1 has already been related to host range specificity within the *Potyviridae* family,[116,126,127] reinforcing the idea that disorder could act as an enhancer of adaptation.

P1 disordered regions contain ELMs

Phosphorylation motifs. Post-translational motifs such as phosphorylation sites are abundant in P1. In disordered regions, surface accessible serine and threonine residues are

 This journal is © The Royal Society of Chemistry 2016

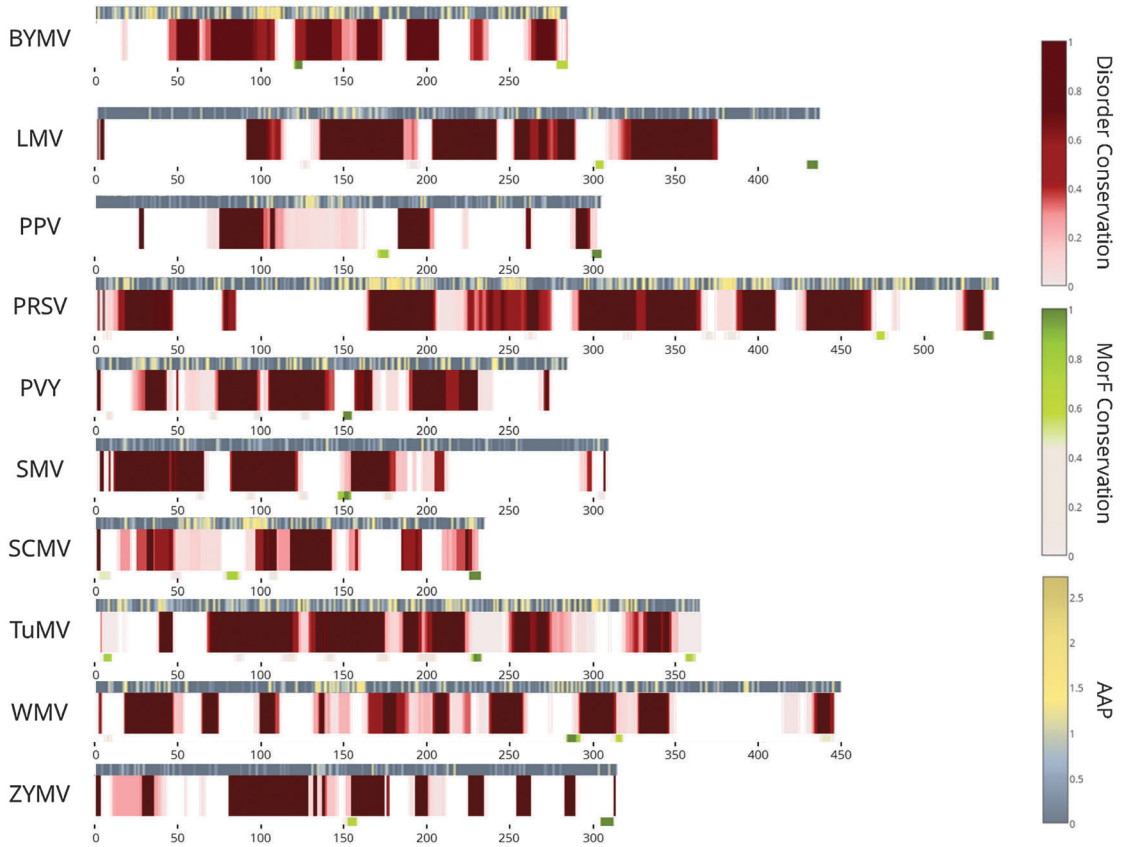


Fig. 5 Conserved IDRs along the P1 protein of ten *Potyvirus* species. The white-to-red gradation bar represents the degree of disorder conservation, from 0% (white) to 100% (dark red). The white-to-green bar represents molecular recognition features (MoRFs) conservation signal, from 0% (white) to 100% (dark green). The blue-to-yellow bar represents amino acid polymorphism (AAP). By definition, 0 to 1 represents highly conserved position, 1 to 2 is considered as moderately conserved and higher than 2 is considered as variable.

more prone to phosphorylation than in ordered regions.[11–14] In many cases, these phosphorylations have been reported to modulate functions by preventing or conversely potentiating the folding of intrinsically disordered regions.[11] According to DISPHOS (Disorder-Enhanced Phosphorylation Sites Predictor), at least four serines and two threonines are highly susceptible to phosphorylation, all located in the P1 N-ter IDR.

14-3-3, FHA and WW binding motifs. These well-conserved eukaryotic motifs are phosphosensors consisting of 6–8 residues with a conserved phosphoserine or threonine binding site. 14-3-3 proteins (also called GRFs for *General Regulatory Factors*) are involved in many regulatory processes in plants (*e.g.* metabolism, hormone signaling, cell-division, stress responses) through their interaction with more than 300 potential targets.[130,131] Several 14-3-3 binding motifs are present within P1 IDRs. The hijacking of those kinds of motifs by *Potyviruses* could have many implications in the viral cycle. It was observed that plants deficient in GRF6 display enhanced resistance to PPV infection.[132] Furthermore, these authors reported that GRF6 degradation by the proteasome is stimulated upon viral infection and contributes to plant defense mechanisms. With 14-3-3, forkhead-associated (FHA) domains are phosphobinding domains involved in numerous signaling processes, such as metabolism and plant development.[131] FHA-binding motifs have already been characterized as functional motifs in plants.[133] They are conserved in *Potyvirus* P1 as a disorder-embedded ELM, and co-localize with the predicted phosphorylation motifs (Table S2, ESI†). The WW domains are known to recruit regulatory protein complexes in various signaling networks. They bind to WW binding motifs, which are short proline or phosphoserine–phosphothreonine-containing motifs. WW-domain proteins have recently been reported as inhibitors of replication of a (+)ssRNA virus, the *Tomato bushy stunt virus* (TBSV, *Tombusviridae*).[134] WW binding motifs were predicted in P1 of all species studied here (Table S2, ESI†). This could provide another example of disorder-mediated ELM mimicry by *Potyviruses* to recruit host factors. Altogether, these data suggest that FHA, 14-3-3 and WW binding motifs, directly associated with post-translational phosphorylation sites in P1, are involved in the viral cycle regulation and/or host defense counteracting activity.

USP7 binding motif. The Ubiquitin Specific Protease USP7 is a member of the large DUB family (deubiquitinating enzymes) in eukaryotes. It catalyzes the ubiquitin removal from proteins, preventing their degradation by the proteasome. Several examples have been reported in plant viruses.[135] Viral DUB activities are encoded by the (+)ssRNA virus *Turnip yellow mosaic virus* (TYMV, *Tymoviridae*) and are involved in controlling the levels of RNA-dependent RNA polymerase (RdRp) and viral infectivity.[136]

It is likely that controlling the host ubiquitination pathway is a prerequisite for the virus cycle. There is no report of ubiquitin protease activity encoded by the *Potyvirus* genome but the presence of conserved USP7 binding motifs in P1 argues in favor of a possible recruitment of deubiquitinating activity.

NLS motif. Nuclear localization signal (NLS) motifs are predicted in P1 IDRs. The propensity of disordered regions to display NLS has already been reported.[137] An active nucleolar localization signal (NoLS) was found in the P1 of TEV within a region encompassing residues 50 to 115, which is predicted as disordered in most of the *Potyvirus* species considered in our study. This NoLS seems to be functional, as during the early stage of infection, P1 is found in the nucleolus, where the pre-ribosomal particle processing takes place. Moreover, the protein has been demonstrated to traffic between the cytoplasm and the nucleolus during infection and interact physically with the cytoplasmic 80S ribosomal subunit, arguing in favor of its involvement in viral translation.[138] Additional nuclear localization signals are predicted for the majority of potyviral species. Interestingly, among these conserved ELMs (Table S2, ESI†), some are not strictly localized in the same IDR. This could suggest a *de novo* apparition of these motifs during *Potyvirus* evolution. Such an evolutionary convergence illustrates both the strong functional benefit underlying these motifs and the evolvability of the corresponding IDRs.[13,37,113]

In vitro experimental approaches like disorder characterization through limited proteolysis and secondary structure analysis through circular dichroism should be now undertaken to further investigate the disorder features of this protein. The biological relevance of the detected ELMs could also be experimentally assessed *in vivo* (for instance, by site-directed mutagenesis).

HC-Pro, a weakly-disordered multifunctional protein

Next to P1 on the polyprotein, HC-Pro is a cysteine endopeptidase of approximately 460 aa long, which self-cleaves at its C-terminus. HC-Pro is a good example of what viral evolution can design in terms of genome economy. This protein is involved in many processes, such as genome replication, aphid transmission, virus-induced gene silencing, viral polyprotein maturation and viral migration within the plant.[139] According to the electron microscopy study of HC-Pro two-dimensional crystals, the N-ter and C-ter domains of the protein are separated by a flexible constriction.[140] Three structural domains can be distinguished: the N- and C-ter regions (approximately 100 aa long each) and the central domain (approximately 250 aa long). The C-ter domain is responsible for the proteolytic activity of the protein. The N-terminal domain is required for aphid transmission of the virus, but most of the HC-Pro functions are located in the central region of the protein. As a whole, only relatively short disordered intra-species conserved regions are predicted in HC-Pro (Fig. S2A, ESI†). These regions are spread all along the protein sequence and are not conserved between species with the exception of the region 161–182 which is predicted as disordered in all the 10 species (Fig. S2A, ESI†). Although resistant to trypsinolysis, HC-Pro from LMV displays two highly exposed sites after Thr170 and Gly176,[140] a feature consistent with this predicted (161–182) IDR (Fig. S2A, ESI†). This region contains a highly conserved FRNK box allowing the binding of siRNA and miRNA duplexes[141] and has been described as involved in the binding of the small RNA methyltransferase HEN1 by HC-Pro from *Zucchini yellow mosaic virus* (ZYMV).[142] Taken together, these features are likely to be related to the gene silencing suppression activity of HC-Pro in the infected cell. This is another example of the potential role of intrinsic disorder in RNA binding processes. This ability of HC-Pro to interact with elements of the silencing pathway well-illustrates a defense and counter-defense interplay between *Potyviruses* and their hosts. Finally, in most of the species considered, a very short segment (at most 10 residues long) is predicted as disordered within the protease domain (Fig. S2A, ESI†). In the case of TuMV, the 3D structure of this protease domain, which spans residues 301–458, was recently solved.[143] It is mostly structured with the exception of a loop (residues 419–426) that was roughly modeled due to weak electron density. However, for TuMV, this loop is not predicted as disordered. This confirms that many short flexible loops on the protein surface should be differentiated from truly disordered segments.

Disorder in P3 and P3N-PIPO: a mediator of protein scaffolding at the membrane surface

In spite of its proven involvement in pathogenicity and symptomatology,[87,144] P3 remains one of the less characterized potyviral proteins. It has been associated with cell-to-cell movement and the formation of replication complexes, through its interaction with the ER and replication vesicles.[145,146] P3 of many *Potyvirus* species contain a N-ter conserved hydrophobic region potentially involved in membrane interactions.[146] This region (between the residues 40 and 80) is preceded by an IDR of variable length, which is conserved intra-species. This IDR-containing N-ter part is important for addressing the protein to the Golgi apparatus.[145] In membranous proteins, an intrinsically disordered tail is commonly observed, protruding into the cytoplasm that is involved in addressing various factors at the membrane surface.[147] This P3 IDR could participate in anchoring the replication vesicles on the ER, or alternatively, on the virus addressing to plasmodesmata through interactions with the cytoskeleton. Indeed, P3 traffics along actin filaments and colocalizes with replication vesicles.[145,146] It is noteworthy that the region including this IDR and the adjacent hydrophobic domain is a common feature of P3 and P3N-PIPO, as the frameshift generating P3N-PIPO is positioned downstream. It could be that both proteins share some functional specificity.

On a broader level, the disorder analysis in P3 showed that there is no clear reproducible IDR pattern at the inter-species level. Nonetheless, some IDRs are highly-conserved intra-species and associated with low AAP (Fig. S2B, ESI†). This enables qualifying P3 disorder as globally constrained. Hence, the protein displays features of "party hubs", namely scaffold proteins which interact simultaneously with several partners and/or anchor them to membranes. It is worth mentioning that

 This journal is © The Royal Society of Chemistry 2016

in LMV more than 30% of the P3 sequence is predicted as disordered with a succession of large and well-conserved IDRs interspaced by shorter ones. A long N-ter IDR (more than 80 residues) is associated with a highly conserved MoRF, a potential interacting domain with other factors. Interestingly, for LMV, the N-ter hydrophobic region was reported around residue 60.[146] According to our data, it is embedded in this long IDR (Fig. S2B, ESI†). This very local hydrophobic signal is typical for MoRF signatures[79] although undetected by MoRFPred.

A 30 aa-long conserved IDR is observed around residue 100 in *Soybean mosaic virus* (SMV) and TuMV (Fig. S2B, ESI†) that matches with a RubisCO-interacting domain.[148] Finally, a central and relatively conserved IDR is identified between residues 200 and 250 of *Bean Yellow mosaic virus* (BYMV), *Sugarcane mosaic virus* (SCMV), LMV, PPV, PVY and TuMV. This region cannot be functionally annotated because of the absence of detectable ELMs.

P3N-PIPO was recently discovered and results from a +2 ribosomal frameshift in P3.[63] Its length is highly variable among *Potyvirus* species and has recently been associated with host-driven specificities.[67] In the infected cells, P3N-PIPO localizes in plasmodesmata, suggesting its involvement in the viral cell-to-cell movement.[149,150] The PIPO domain interacts with PCaP1, a cation-binding protein anchored to membranes through myristoylation.[151] A putative disordered-embedded MoRF is found in the C-ter region. In spite of its heterogeneity in length, and as opposed to P3, the IDR distribution in the PIPO part is conserved throughout in all the 10 species. In addition, similarly to P3, IDRs in the PIPO part can be classified as constrained disordered regions. Like P3, P3N-PIPO could act as a party-hub. As a matter of fact, P3N-PIPO interacts with CI.[149] The IDRs within P3N-PIPO could potentiate the simultaneous interactions of the protein with CI and PCaP1. Although overlapping reading frames are frequent and strategic for virus genome economy, they are expected to be heavily impacted by high mutational rates of viruses. However, many IDRs better tolerate mutations than structured regions. Interestingly, intrinsic disorder was found to be abundantly generated in viral overlapping reading frames.[152] Disorder in both the PIPO and the P3 C-ter part, (orange dashed lines in Fig. S2B and C, ESI†) observed for most of the studied potyviral species, could prevent destabilizing effects of mutations on these overprinting regions.

No ELM was found associated with disorder neither in P3 nor in the PIPO part

Given their small size (about 50 aa each), and owing to the very short disordered segments predicted for 6K1 and 6K2 and obtained in their N- and C-terminal parts (data not shown), these two proteins were discarded from our analysis.

CI, a protein with two ordered domains interspaced by a conserved IDR

The cylindrical inclusion protein (CI) is a viral helicase. It interacts with various viral and host partners and is involved in movement and replication.[153] Similar to HC-Pro, P3 and P3N-PIPO, CI is classified as moderately disordered. This is consistent with CI main functions devoted to catalysis, through its ATPase and RNA-helicase activities, which require well-ordered catalytic domains. The N-terminal region of the CI carries the helicase activity[154] and the C-terminal domain possesses binding sites for (at least) two different proteins, the viral VPg, and the plant eIF4E, an initiation translation factor.[155] These two domains are interspaced by a flexible region (Fig. S2D, ESI†). This less-structured region is the main contributor to the overall disorder scores obtained (Fig. 2). This IDR is conserved in both intra- and inter-species and is roughly located between residues 330 and 370. In the IDR, it is included as an arginine-rich motif constituting a RNA-binding domain, which corresponds to the most carboxy-terminal conserved domain of RNA helicases of the superfamily SF2.[156] In PPV, this RNA-binding domain was mapped in between aa 350 and 402.[154] Interestingly, such a motif is also the hallmark of eIF4A, another SF2 helicase belonging to the translation initiation complex eIF4F highly conserved in all eukaryotes. PONDR-VLXT® prediction obtained for eIF4A displays the same general profile as that of CI, with an ordered background interrupted by short and spaced disordered signals all along the sequence.[157] Hence, it is noteworthy that the CI shares with eIF4A a similar intrinsically disordered RNA binding domain, sustaining the hypothesis of a common function associated with disorder.

VPg contains flexible disorder

In *Potyviruses*, VPg constitutes the N-ter part of the NIa protein. After cleavage, the VPg consists of a 22 to 25 kDa protein. VPgs are not restricted to phytoviruses, also being involved in the genome replication and protein translation of animal (+)ssRNA viruses. VPg is covalently linked to the 5′-end of viral RNA through a conserved tyrosine residue.[158,159] This feature is likely involved in many functional aspects, such as aphid transmission and uncoating process regulation, movement of viral RNA through plasmodesmata, replication initiation through urydilyl VPg priming or the initial steps of translation.[90,91,160–171] VPg can be considered as a hub protein interacting with both various host and viral factors, as well as with the viral RNA.[172] The best documented of these interactions concerns the eukaryotic translation initiation factors eIF4E and eIF(iso)4E.[91,173–177] This is associated with various tasks during potyviral infection such as translation, replication and cell-to-cell movement.[168,178,179] The interaction with nucleolar Fibrillarin has also been reported,[180] suggesting a function in RNA-silencing or host gene expression regulation. The interactions of VPg with the host RNA helicase RH8, a poly(A)-binding protein (PABP) and the eukaryotic elongation factor 1A (eF1A) clearly demonstrate the involvement of VPg in viral RNA replication and translation.[91,176,181,182] Unfortunately, these interactions have not yet been mapped on the VPg sequence. Intrinsic disorder in VPgs was clearly demonstrated by experimental results obtained in several plant viruses comprising *Potyviruses* and sobemoviruses.[29,44–47]

IDRs were predicted in all the potyviral VPgs of our dataset. This intrinsic disorder is highly conserved in intra- and inter-species. Conserved IDRs were found at the N-ter and C-ter of

potyviral VPgs. The N-ter IDR interacts with PVIP,[169] this interaction being linked to the movement of virus. Downstream, a short highly conserved IDR, centered around residue 50, contains functional NTP-binding and RNA-binding sites.[47] The central part of the VPg displays a conserved IDR about 20 to 30 residues long. This region is involved in the interaction of VPg with eIF4E and HC-Pro.[183] Interestingly, although no MoRF was predicted in this region, the IDR spanning residues 89–105 fold into a helix upon binding with eIF4E.[44] At the intra-species level, potyviral VPgs do not share a global high d*N*/d*S* value (Fig. S1, ESI†). But in most of the species, the central and C-ter IDRs of VPgs are associated with high AAP (Fig. S2E, ESI†). Therefore, the disorder contained in potyviral VPgs is typically of flexible nature. In this respect the potyviral VPg is likely to constitute a typical example of "date hub", with its central IDR interacting sequentially with several partners.

Interestingly, the central domain of VPg contains the molecular determinants responsible for overcoming many eIF4E-mediated resistances.[177,184–187] As a matter of fact, VPgs share episodic and localized events of positive selection mostly localized in this central IDR. This leads us to hypothesize about the involvement of the disorder status in such an adaptive process. The central part of PPV VPg is not predicted as disordered and represents a puzzling exception.

Phosphorylation motifs. As in P1, well-conserved phosphorylation motifs are predicted in VPg disordered regions (Table S2, ESI†). Both *in vitro* and *in vivo* experiments have already highlighted that VPg is a highly phosphorylated protein.[188,189] However, no functions have yet been associated with these modifications.

Nuclear localization motifs. The N-ter IDR of TEV and PVA VPgs display bi-partite NLS signals.[180,190] Their biological relevance was validated by VPg nucleus localization experiments in several *Potyviruses*.[176,180,190] Functions associated with this nuclear localization still remain unclear but could be related to a hypothetic involvement in the host silencing suppression process.[180] Such VPg NLS motifs are retrieved in 8 of the 10 species studied here, reinforcing the idea that VPg nuclear localization is shared among all *Potyviruses* and is very probably associated with important functions in the viral cycle.

NIa-Pro displays an inconstant functional C-ter IDR

The 240 residues long C-ter domain of NIa, called NIa-Pro, is a protease, which self-releases from NIa. Its functions in the polyprotein processing have been discussed above (see the "Disorder and polyprotein processing" section). Additionally, NIa-Pro possesses DNase activity possibly involved in the regulation of host gene expression.[191]

A central IDR between residues 95 and 125 is predicted in half of the species examined. The corresponding region in the TEV and *Tobacco vein mottling virus* (TVMV) proteases is part of a well-resolved surface loop in the 3D structures.[192–194] It cannot be excluded that this region is stabilized in the crystal packing. A short IDR (about 20 residues long) is only predicted at the C-ter of SCMV in our dataset (Fig. S2F, ESI†). This inconstant IDR was experimentally confirmed in the structures of the TEV and TVMV NIa-Pro. Interestingly, this IDR increases the proteolytic activity in the case of TEV, while reduces it in the case of TVMV, suggesting that this IDR behaves like a modulator of the enzyme.[193] This illustrates how IDRs could tune the viral functions according to the host specificity.

The NIb replicase protein is ordered

NIb is the potyviral replicase, ensuring virus genome amplification. NIb is addressed to membrane-associated viral factories and participates in the formation of replication complexes, through its interaction with VPg, NIa-Pro and several host factors EF1A, Hsc70-3 and PABP.[182,195,196] In spite of the ordered nature of this protein, a well-conserved IDR spanning residues 400 to 450 was predicted in nine of the ten species. It was not possible to functionally annotate this region. An α-MoRF is predicted for most of the species spanning residues 320–340. However, the low disorder content in the region does not support this prediction.

As many other viral RdRP, the potyviral NIb displays the folding of well-ordered and conserved domains into the typical "right hand" structure.[197] However, the low disorder content observed for potyviral RdRP does not strictly constitute a hallmark in the world of RNA viruses. For instance, HCV RdRP displays a mean disorder content of 19.1%.[39]

Structural coat protein (CP), flexible disorder and adaptation

Potyvirus CPs are filamentous particles with helical symmetry made-up from the supramolecular assembly of about 2000 CP subunits.[54] According to structural studies, many viral CP folding consist of a central globular part with N- and C-ter extended arms, found as flexible and disordered at least in the CP monomeric form.[198] These extended arms have been related to several functions, such as nucleic acid interactions, regulation and control of virion assembly and stabilization. In addition to its role in viral genome protection (encapsidation), potyviral CPs also participate in various non-structural functions such as viral RNA translation, replication and movement.[199,200]

At the inter-species scale, potyviral CPs are heterogeneous in size (from 270 to 350 residues). But all species share the same organization, consisting of: (i) a highly conserved ordered central core flanked by a long N-ter IDR that presents a high intra-species sequence polymorphism and (ii) a more conserved C-ter domain displaying altered short ordered and disordered segments (Fig. S2H, ESI†).[201,202] When assembled, the central globular part of the CP, which is inaccessible to proteolysis, forms a structural core interacting with the genomic RNA. The CP N-ter part is accessible from the particle surface as shown by its sensitivity to proteolysis.[203] In PVA, tritium bombardment gave evidence that the residues 1–15 and 27–50 are exposed to the surface with at least the first 8 residues being disordered.[204] Depending on the virus species, 18 to 20 C-ter aa are also exposed to the surface.[203] Importantly, intrinsic disorder was experimentally detected in the CP within the assembled particle.[48] The N-ter region is likely to be the main contributor to this observed intrinsic disorder. The N-terminal part is not

 This journal is © The Royal Society of Chemistry 2016

structurally essential for the capsid as its deletion has little effect on the virion morphology.[205] This suggests that the structurally flexible N-ter region could participate in non-structural functions in the viral cycle. The N-ter region displays a DAG motif that is involved in the aphid-transmission process[206] by mediating the CP interaction with HC-Pro.[207] This interaction site co-localizes with an inter-species highly conserved MoRF (Fig. S2H, ESI†), supporting plausible induced folding events associated with HC-Pro binding. This N-ter IDR could behave as an extended arm that explores a large area (according to the "fly-casting" model) and folds as it approaches the actual binding site.[208,209] Contrarily to the N-ter part, which is always exposed to the particle surface, the C-ter region surface exposure varies between potyviral species.[203,210] However, in all species, disorder is predicted in this part (Fig. S2H, ESI†). A conserved MoRF signal is associated with these IDRs but no interaction is yet reported in this region.

CPs, amino acid variability and *Potyvirus* adaptation. In the N-ter region of potyviral CPs, the high content in conserved disorder is associated with high amino-acid polymorphism (Fig. S2H, ESI†). This illustrates the relationship between low structural constraints and high variability in disordered regions. Given its N-ter extreme aa diversity (revealed by its high mean d*N*/d*S* value in Fig. S1, ESI†), the CP is expected to be a determinant of *Potyvirus* adaptation. Indeed, host-specific determinant motifs in the CP N-ter region have been reported in *Watermelon mosaic virus* (WMV).[211]

ELMs in CPs. The PVA CP is subjected to phosphorylation. This post-translational modification is directly involved in the RNA-binding regulation, as it reduces the CP affinity for nucleic acids.[212,213] In PPV, CP phosphorylation and O-GlcNAcylation modifications have also been reported in the N-ter of the protein.[214,215] The CP phosphorylation state seems to have an enhancing impact on viral infection.[213] These phosphorylation sites were detected in most of the species of our dataset (Table S2, ESI†). A fine-tuning of the CP pool is required within cells.[199] As a matter of fact, high CP accumulation induces RNA encapsidation, preventing its replication by NIb. To enable replication, this excess of CPs is likely to be addressed to ubiquitination-associated degradation, through an interaction with the chaperones CPIP and HSP70.[98,216] The ELM analysis of CP IDRs revealed a well-conserved USP7-binding motif (Table S2, ESI†). As discussed in the case of P1, this motif could participate in CP recognition by a HSP70 dependent ubiquitin-ligase.

Conclusion

Using a proteome dataset representative of the entire *Potyvirus* genus, we were able to analyze the proteome intrinsic disorder both at inter- and intra-species scales. Our work revealed that *Potyvirus* proteomes display high disorder content. Its maintenance during *Potyvirus* evolution strongly suggests that functional benefits are associated with this structural feature.

A deeper analysis of disorder conservation indicates that, as previously reported for animal viruses, many *Potyvirus* cycle steps and evolutive processes potentially benefit from intrinsic disorder. This feature could favor potyviral adaptation, as IDRs globally evolve faster than ordered regions, suggesting that they are more tolerant to mutation than structured domains. Our results also suggest that intrinsic disorder regulates the polyprotein proteolytic cleavage. Based on their disorder content and their supposed related functions, *Potyvirus* proteins can be mainly classified into three groups. The first group includes P1, VPg and CP, three highly disordered proteins. They all contain IDRs displaying high aa polymorphism, a distinctive feature of flexible disorder. The functional annotation of these proteins allows classifying them as date hubs, which interact sequentially with several partners. Moreover, their numerous IDRs display several conserved ELMs, mostly related to post-translational modifications. This illustrates the potential involvement of intrinsic disorder in host motif mimicry by the virus. The second group, with medium disorder content, includes HC-Pro, P3, P3N-PIPO and CI. Most IDRs of these proteins are rather conserved and belong to the class of constrained disorder. This disorder is especially present in party hubs, which are proteins involved in several simultaneous interactions, as illustrated by the P3 scaffolding role in the intracellular virus factories. The third group, represented by two enzymes, NIa-Pro and NIb, contains a low amount of disorder.

The constant progress in our understanding of *Potyvirus* biology underlines its molecular complexity, and many key features of the virus cycle remain unknown. We believe that this proteome-wide analysis of intrinsic disorder provides an alternative way to functionally annotate *Potyvirus* proteins, helping in defining new experimental paths for exploring the biology of this major virus genus.

Acknowledgements

We thank Dr Jocelyne Walter, Dr Benoît Moury, Dr Eugénie Hébrard and Dr Thierry Candresse for their helpful comments that led to improvement of this work. We also thank SharCo consortium for providing us PPV genome sequences used in this work and Dr Lukasz Kurgan for his essential help concerning MoRFPred predictions.

References

1 P. E. Wright and H. J. Dyson, *J. Mol. Biol.*, 1999, **293**, 321–331.
2 V. N. Uversky, J. R. Gillespie and A. L. Fink, *Proteins: Struct., Funct., Genet.*, 2000, **427**, 415–427.
3 A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner and Z. Obradovic, *J. Mol. Graph. Modell.*, 2001, **19**, 26–59.
4 P. Tompa, *Trends Biochem. Sci.*, 2002, **27**, 527–533.
5 K. W. Plaxco and M. Grob, *Nature*, 1997, **386**, 657–658.

6 K. A. Dill and H. S. Chan, *Nat. Struct. Biol.*, 1997, **4**, 10–19.
7 G. W. Daughdrill, M. S. Chadsey, J. E. Karlinsey, T. H. Kelly and F. W. Dahlquist, *Nature*, 1997, **4**, 285–291.
8 R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed and P. E. Wright, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 11504–11509.
9 K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, Z. Obradovic, C. Science, C. Kissinger, J. E. Villafranca, A. Pharmaceutical and S. Diego, *Pac. Symp. Biocomput.*, 1998, 473–484.
10 V. N. Uversky, *Biopolymers*, 2013, **99**, 870–887.
11 L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic and A. K. Dunker, *Nucleic Acids Res.*, 2004, **32**, 1037–1049.
12 M. O. Collins, L. Yu, I. Campuzano, S. G. N. Grant and J. S. Choudhary, *Mol. Cell. Proteomics*, 2008, **7**, 1331–1348.
13 F. Diella, N. Haslam, C. Chica, A. Budd, S. Michael, N. P. Brown, G. Travé and T. J. Gibson, *Front. Biosci.*, 2008, **13**, 6580–6603.
14 C. A. Galea, Y. Wang, S. G. Sivakolundu and R. W. Kriwacki, *Biochemistry*, 2008, **47**, 7598–7609.
15 C. J. Brown, S. Takayama, A. M. Campen, P. Vise, T. W. Marshall, C. J. Oldfield, C. J. Williams and A. K. Dunker, *J. Mol. Evol.*, 2002, **55**, 104–110.
16 R. Pancsa and P. Tompa, *PLoS One*, 2012, **7**, e34687.
17 J. W. Chen, P. Romero, V. N. Uversky and A. Keith, *J. Proteome Res.*, 2006, **5**, 888–898.
18 C. J. Brown, A. K. Johnson and G. W. Daughdrill, *Mol. Biol. Evol.*, 2010, **27**, 609–621.
19 J. Bellay, S. Han, M. Michaut, T. Kim, M. Costanzo, B. J. Andrews, C. Boone, G. D. Bader, C. L. Myers and P. M. Kim, *Genome Biol.*, 2011, **12**, R14.
20 C. J. Brown, A. K. Johnson and G. W. Daughdrill, *Mol. Biol. Evol.*, 2011, **29**, 443.
21 J. Nilsson, M. Grahn and A. P. H. Wright, *Genome Biol.*, 2011, **12**, R65.
22 Z. Peng, J. Yan, X. Fan, M. J. Mizianty, B. Xue, K. Wang, G. Hu, V. N. Uversky and L. Kurgan, *Cell. Mol. Life Sci.*, 2014, **71**, 1477–1504.
23 B. Xue, A. K. Dunker and V. N. Uversky, *J. Biomol. Struct. Dyn.*, 2012, **30**, 137–149.
24 M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic and A. K. Dunker, *Nucleic Acids Res.*, 2007, **35**, D786–D793.
25 R. Pushker, C. Mooney, N. E. Davey, J.-M. Jacqué and D. C. Shields, *PLoS One*, 2013, **8**, e60724.
26 V. N. Uversky and S. Longhi, *Flexible Viruses: Structural disorder in Viral Proteins*, John Wiley and Sons, Hoboken, 2012.
27 B. Xue, D. Blocquel, J. Habchi, A. V. Uversky, L. Kurgan, V. N. Uversky and S. Longhi, *Chem. Rev.*, 2014, **114**, 6880–6911.
28 C. Alves and C. Cunha, *Future Virol.*, 2012, **7**, 1183–1191.
29 E. Hébrard, Y. Bessin, T. Michon, S. Longhi, V. N. Uversky, F. Delalande, A. Van Dorsselaer, P. Romero, J. Walter, N. Declerck and D. Fargette, *Virol. J.*, 2009, **6**, 23.
30 N. Tokuriki, C. J. Oldfield, V. N. Uversky, I. N. Berezovsky and D. S. Tawfik, *Trends Biochem. Sci.*, 2009, **34**, 53–59.
31 S. Duffy, L. A. Shackelton and E. C. Holmes, *Nat. Rev. Genet.*, 2008, **9**, 267–276.
32 G. K.-M. Goh, A. K. Dunker and V. N. Uversky, *BMC Genomics*, 2008, **9**, S4.
33 B. Xue and V. N. Uversky, *J. Mol. Biol.*, 2013, **426**, 1322–1350.
34 G. K.-M. Goh, A. K. Dunker and V. N. Uversky, *J. Pathog.*, 2012, **2012**, 738590.
35 G. K.-M. Goh, A. K. Dunker and V. N. Uversky, *Virol. J.*, 2009, **6**, 69.
36 G. K.-M. Goh, A. K. Dunker and V. N. Uversky, *PLoS Curr.*, 2013.
37 T. Hagai, A. Azia, M. M. Babu and R. Andino, *Cell Rep.*, 2014, **7**, 1729–1739.
38 L. Gitlin, T. Hagai, A. LaBarbera, M. Solovey and R. Andino, *PLoS Pathog.*, 2014, **10**, e1004529.
39 X. Fan, B. Xue, P. T. Dolan, D. J. Lacount, L. Kurgan and V. N. Uversky, *Mol. BioSyst.*, 2014, **10**, 1345–1363.
40 B. Xue, M. J. Mizianty, L. Kurgan and V. N. Uversky, *Cell. Mol. Life Sci.*, 2012, **69**, 1211–1259.
41 V. N. Uversky, A. Roman, C. J. Oldfield and A. K. Dunker, *J. Proteome Res.*, 2006, **5**, 1829–1842.
42 F. Meng, R. A. Badierah, H. A. Almehdar, E. M. Redwan, L. Kurgan and V. N. Uversky, *FEBS J.*, 2015, **282**, 3368–3394.
43 J. F. Ortiz, M. L. MacDonald, P. Masterson, V. N. Uversky and J. Siltberg-Liberles, *Genome Biol. Evol.*, 2013, **5**, 504–513.
44 J. Chroboczek, E. Hébrard, K. Mäkinen, T. Michon and K. Rantalainen, *Flexible Viruses*, 2012.
45 R. Grzela, E. Szolajska, C. Ebel, D. Madern, A. Favier, I. Wojtal, W. Zagorski and J. Chroboczek, *J. Biol. Chem.*, 2008, **283**, 213–221.
46 K. I. Rantalainen, V. N. Uversky, P. Permi, N. Kalkkinen, A. keith Dunker and K. Makinen, *Virology*, 2008, **377**, 280–288.
47 K. I. Rantalainen, K. Eskelin, P. Tompa and K. Mäkinen, *J. Virol.*, 2011, **85**, 2449–2457.
48 A. L. Ksenofontov, V. Paalme, A. M. Arutyunyan, P. I. Semenyuk, N. V. Fedorova, R. Rumvolt, L. A. Baratova, L. Järvekülg and E. N. Dobrov, *PLoS One*, 2013, **8**, e67830.
49 V. Nicaise, *Front. Plant Sci.*, 2014, **5**, 1–18.
50 C. Ward and D. Shukla, *Intervirology*, 1991, **32**, 269–296.
51 K.-B. G. Scholthof, S. Adkins, H. Czosnek, P. Palukaitis, E. Jacquot, T. Hohn, B. Hohn, K. Saunders, T. Candresse, P. Ahlquist, C. Hemenway and G. D. Foster, *Mol. Plant Pathol.*, 2011, **12**, 938–954.
52 D. Shukla and C. Ward, *Adv. Virus Res.*, 1989, **36**, 273–314.
53 M. J. Adams, F. Zerbini, R. French, F. Rabenstein, D. Stenger and J. P. T. Valkonen, *ICTV*, 2012, 1069–1089.
54 A. Kendall, M. McDonald, W. Bian, T. Bowles, S. C. Baumgarten, J. Shi, P. L. Stewart, E. Bullitt, D. Gore, T. C. Irving, W. M. Havens, S. A. Ghabrial, J. S. Wall and G. Stubbs, *J. Virol.*, 2008, **82**, 9546–9554.
55 S. Lain, J. L. Riechmann, E. Mcndez and J. A. Garcia, *Virus Res.*, 1988, **10**, 325–342.
56 M. F. E. Siaw, M. Shahabuddin, S. Ballard, J. G. Shaw and R. E. Rhoads, *Virology*, 1985, **142**, 134–143.

 This journal is © The Royal Society of Chemistry 2016

57 J. L. Riechmann, S. Lain and J. A. Garcia, *J. Gen. Virol.*, 1989, **70**, 2785–2789.
58 A. Merits, M. Rajama, P. Runeberg-roos and T. Kekarainen, *J. Gen. Virol.*, 2002, 1211–1221.
59 J. C. Carrington, D. D. Freed and C.-S. Oh, *EMBO J.*, 1990, **9**, 1347–1353.
60 J. C. Carrington, S. M. Cary, T. D. Parks and W. G. Dougherty, *EMBO J.*, 1989, **8**, 365–370.
61 J. Verchot, E. V. Koonin and J. C. Carrington, *Virology*, 1991, **185**, 527–535.
62 J. A. Garcia, J. L. Riechmann and S. Lain, *J. Virol.*, 1989, **63**, 2457–2460.
63 B. Y.-W. Chung, W. A. Miller, J. F. Atkins and A. E. Firth, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 5897–5902.
64 Y. Hong and A. G. Hunt, *Virology*, 1996, **226**, 146–151.
65 K. Mäkinen and A. Hafrén, *Front. Plant Sci.*, 2014, **5**, 1–12.
66 K. Tamura, G. Stecher, D. Peterson, A. Filipski and S. Kumar, *Mol. Biol. Evol.*, 2013, **30**, 2725–2729.
67 J. Hillung, S. F. Elena and J. M. Cuevas, *BMC Evol. Biol.*, 2013, **13**, 249.
68 P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown and A. K. Dunker, *Proteins: Struct., Funct., Genet.*, 2001, **48**, 38–48.
69 F.-C. Chen, C.-L. Pan and H.-Y. Lin, *Mol. Biol. Evol.*, 2012, **29**, 187–193.
70 K. Scheffler, D. P. Martin and C. Seoighe, *Bioinformatics*, 2006, **22**, 2493–2499.
71 S. L. K. Pond, S. D. W. Frost and S. V. Muse, *Bioinformatics*, 2005, **21**, 676–679.
72 W. Delport, A. F. Y. Poon, S. D. W. Frost and S. L. Kosakovsky Pond, *Bioinformatics*, 2010, **26**, 2455–2457.
73 S. L. K. Pond and S. D. W. Frost, *Bioinformatics*, 2005, **21**, 2531–2533.
74 S. L. K. Pond, B. Murrell and A. F. Y. Poon, *Methods Mol. Biol.*, 2012, **856**, 239–272.
75 M. Anisimova, R. Nielsen and Z. Yang, *Genetics*, 2003, **1236**, 1229–1236.
76 K. Katoh, K. Kuma, H. Toh and T. Miyata, *Nucleic Acids Res.*, 2005, **33**, 511–518.
77 A. Gibbs and K. Ohshima, *Annu. Rev. Phytopathol.*, 2010, **48**, 205–223.
78 M. Garcia-Boronat, C. M. Diez-Rivero, E. L. Reinherz and P. A. Reche, *Nucleic Acids Res.*, 2008, **36**, W35–W41.
79 F. M. Disfani, W.-L. Hsu, M. J. Mizianty, C. J. Oldfield, B. Xue, A. K. Dunker, V. N. Uversky and L. Kurgan, *Bioinformatics*, 2012, **28**, 75–83.
80 N. E. Davey, K. Van Roey, R. J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel and T. J. Gibson, *Mol. BioSyst.*, 2012, **8**, 268–281.
81 A. J. Gibbs, K. Ohshima, M. J. Phillips and M. J. Gibbs, *PLoS One*, 2008, **3**, e2523.
82 K. Rajagopalan, S. M. Mooney, N. Parekh, R. H. Getzenberg and P. Kulkarni, *J. Cell. Biochem.*, 2011, **112**, 3256–3267.
83 C. Haynes, C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal and L. M. Iakoucheva, *PLoS Comput. Biol.*, 2006, **2**, e100.
84 P. M. Kim, A. Sboner, Y. Xia and M. Gerstein, *Mol. Syst. Biol.*, 2008, **4**, 179.
85 L. C. James and D. S. Tawfik, *Trends Biochem. Sci.*, 2003, **28**, 361–368.
86 S. F. Elena and G. Rodrigo, *Curr. Opin. Virol.*, 2012, **2**, 719–724.
87 F. Revers and J. García, *Adv. Virus Res.*, 2015, **92**, 101–199.
88 M. J. Adams, J. F. Antoniw and F. Beaudoin, *Mol. Plant Pathol.*, 2005, **6**, 471–487.
89 M. A. Restrepo-hartwig and J. C. Carrington, *J. Virol.*, 1994, **68**, 2388–2397.
90 M. C. Schaad, P. E. Jensen and J. C. Carrington, *EMBO J.*, 1997, **16**, 4049–4059.
91 S. Leonard, *J. Gen. Virol.*, 2004, **85**, 1055–1063.
92 E. Majer, Z. Salvador, M. P. Zwart, A. Willemsen, S. F. Elena and J.-A. Daròs, *J. Virol.*, 2014, **88**, 4586–4590.
93 C. Mathur and H. S. Savithri, *Biochem. Biophys. Res. Commun.*, 2012, **427**, 113–118.
94 A. Fontana, P. P. De Laureto, B. Spolaore, E. Frare, P. Picotti and M. Zambonin, *Acta Biochim. Pol.*, 2004, **51**, 299–321.
95 D. E. Johnson, B. Xue, M. D. Sickmeier, J. Meng, M. S. Cortese, C. J. Oldfield, T. Le Gall, A. K. Dunker and V. N. Uversky, *J. Struct. Biol.*, 2012, **180**, 201–215.
96 P. Paulverino De Laureto, V. de Filippis, E. Scaramella, M. Zambonin and A. Fontana, *Eur. J. Biochem.*, 1995, **230**, 779–787.
97 P. Paulverino De Laureto, L. Tosatto, E. Frare, O. Marin, V. N. Uversky and A. Fontana, *Biochemistry*, 2006, **45**, 11523–11531.
98 J. Besong-Ndika, K. I. Ivanov, A. Hafr''n, T. Michon and K. Mäkinen, *J. Virol.*, 2015, **89**, 4237–4248.
99 G. W. Daughdrill, P. Narayanaswami, S. H. Gilmore, A. Belczyk and C. J. Brown, *J. Mol. Evol.*, 2007, **65**, 277–288.
100 J. J. Han, N. Bertin, T. Hao and D. S. Goldberg, *Nature*, 2004, **430**, 88–93.
101 R. Van Der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Old, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, *Chem. Rev.*, 2014, **114**, 6589–6631.
102 P. M. Kim, L. J. Lu, Y. Xia and B. G. Mark, *Science*, 2006, **603**, 1938–1941.
103 S. Agarwal, C. M. Deane, M. A. Porter and N. S. Jones, *PLoS Comput. Biol.*, 2010, **6**, e1000817.
104 C. J. Oldfield, Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky and A. K. Dunker, *Biochemistry*, 2005, **44**, 12454–12470.
105 A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker and V. N. Uversky, *J. Mol. Biol.*, 2006, **362**, 1043–1059.
106 P. T. Dolan, A. P. Roth, B. Xue, R. Sun, A. K. Dunker, V. N. Uversky and D. J. LaCount, *Protein Sci.*, 2015, **24**, 221–235.
107 B. Meszaros, I. Simon and Z. Dosztanyi, *PLoS Comput. Biol.*, 2009, **5**, e1000376.

108 V. Vacic, C. J. Oldfield, A. Mohan, P. Radivojac, M. S. Cortese, V. N. Uversky and A. K. Dunker, *J. Proteome Res.*, 2007, **6**, 2351–2366.
109 H. Dinkel, K. Van Roey, S. Michael, N. E. Davey, R. J. Weatheritt, D. Born, T. Speck, D. Krüger, G. Grebnev, M. Kuban, M. Strumillo, B. Uyar, A. Budd, B. Altenberg, M. Seiler, L. B. Chemes, J. Glavina, I. E. Sánchez, F. Diella and T. J. Gibson, *Nucleic Acids Res.*, 2014, **42**, D259–D266.
110 H. Dinkel, S. Michael, R. J. Weatheritt, N. E. Davey, K. Van Roey, B. Altenberg, G. Toedt, B. Uyar, M. Seiler, A. Budd, L. Jödicke, M. A. Dammert, C. Schroeter, M. Hammer, T. Schmidt, P. Jehl, C. McGuigan, M. Dymecka, C. Chica, K. Luck, A. Via, A. Chatr-Aryamontri, N. Haslam, G. Grebnev, R. J. Edwards, M. O. Steinmetz, H. Meiselbach, F. Diella and T. J. Gibson, *Nucleic Acids Res.*, 2012, **40**, D242–D251.
111 M. Fuxreiter, P. Tompa and I. Simon, *Bioinformatics*, 2007, **23**, 950–956.
112 R. Pancsa and M. Fuxreiter, *IUBMB Life*, 2012, **64**, 513–520.
113 N. E. Davey, G. Travé and T. J. Gibson, *Trends Biochem. Sci.*, 2011, **36**, 159–169.
114 N. C. Elde and H. S. Malik, *Nat. Rev. Microbiol.*, 2009, **7**, 787–797.
115 J. Verchot and J. C. Carrington, *J. Virol.*, 1995, **69**, 3668–3674.
116 A. Valli, J. J. López-Moya and J. A. García, *J. Gen. Virol.*, 2007, **88**, 1016–1028.
117 J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton and D. T. Jones, *J. Mol. Biol.*, 2004, **337**, 635–645.
118 H. J. Dyson, *Mol. BioSyst.*, 2012, **8**, 97–104.
119 Y. Soumounou and J.-F. Laliberté, *J. Gen. Virol.*, 1994, **75**, 2567–2573.
120 F. Martinez and J. A. Daros, *J. Virol.*, 2014, **88**, 10725–10737.
121 P. Tompa and P. Csermely, *FASEB J.*, 2004, **18**, 1169–1175.
122 R. Ivanyi-Nagy, J. P. Lavergne, C. Gabus, D. Ficheux and J. L. Darlix, *Nucleic Acids Res.*, 2008, **36**, 712–725.
123 F. Pasin, C. Simón-Mateo and J. A. García, *PLoS Pathog.*, 2014, **10**, e1003985.
124 T. Trudeau, R. Nassar, A. Cumberworth, E. T. C. Wong, G. Woollard and J. Gsponer, *Structure*, 2013, **21**, 332–341.
125 A. Merits, D. Guo, L. Järvekülg and M. Saarma, *Virology*, 1999, **263**, 15–22.
126 V. I. Maliogka, B. Salvador, A. Carbonell, P. Sáenz, D. S. León, J. C. Oliveros, M. O. Delgadillo, J. A. García and C. Simón-Mateo, *Mol. Plant Pathol.*, 2012, **13**, 877–886.
127 B. Salvador, P. Saénz, E. Yangüez, J. B. Quiot, L. Quiot, M. O. Delgadillo, J. A. García and C. Simón-Mateo, *Mol. Plant Pathol.*, 2008, **9**, 147–155.
128 C. Brown and A. Johnson, *Curr. Opin. Struct. Biol.*, 2011, **21**, 441–446.
129 A. Schlessinger, C. Schaefer, E. Vicedo, M. Schmidberger, M. Punta and B. Rost, *Curr. Opin. Struct. Biol.*, 2011, **21**, 412–418.
130 R. Lozano-Durán and S. Robatzek, *Mol. Plant-Microbe Interact.*, 2015, **28**, 511–518.
131 D. Chevalier, E. R. Morris and J. C. Walker, *Annu. Rev. Plant Biol.*, 2009, **60**, 67–91.
132 J. L. Carrasco, M. J. Castelló, K. Naumann, I. Lassowskat, M. Navarrete-Gómez, D. Scheel and P. Vera, *PLoS One*, 2014, **9**, e90734.
133 Z. Ding, G. I. Lee, X. Liang, F. Gallazzi, A. Arunima and S. R. Van Doren, *Biochemistry*, 2005, **44**, 10119–10134.
134 J. Qin, D. Barajas and P. D. Nagy, *Virology*, 2012, **426**, 106–119.
135 C. Alcaide-Loridan and I. Jupin, *Plant Physiol.*, 2012, **160**, 72–82.
136 M. Chenon, L. Camborde, S. Cheminant and I. Jupin, *EMBO J.*, 2012, **31**, 741–753.
137 S. E. Dixon, M. M. Bhatti, V. N. Uversky, K. Dunker and W. J. Sullivan Jr, *Mol. Biochem. Parasitol.*, 2011, **175**, 192–195.
138 F. Martínez and J.-A. Daròs, *J. Virol.*, 2014, **88**, 10725–10737.
139 J. Syller, *Physiol. Mol. Plant Pathol.*, 2006, **67**, 119–130.
140 C. Plisson, M. Drucker, S. Blanc, S. German-Retana, O. Le Gall, D. Thomas and P. Bron, *J. Biol. Chem.*, 2003, **278**, 23753–23761.
141 Y. M. Shiboleth, E. Haronsky, D. Leibman, T. Arazi, M. Wassenegger, S. A. Whitham, V. Gaba and A. Gal-On, *J. Virol.*, 2007, **81**, 13135–13148.
142 R. M. Jamous, K. Boonrod, M. W. Fuellgrabe, M. S. Ali-Shtayeh, G. Krczal and M. Wassenegger, *J. Gen. Virol.*, 2011, **92**, 2222–2226.
143 B. Guo, J. Lin and K. Ye, *J. Biol. Chem.*, 2011, **286**, 21937–21943.
144 N. Suehiro, T. Natsuaki, T. Watanabe and S. Okuda, *J. Gen. Virol.*, 2004, **85**, 2087–2098.
145 X. Cui, T. Wei, R. V. Chowda-Reddy, G. Sun and A. Wang, *Virology*, 2010, **397**, 56–63.
146 S. Eiamtanasate, M. Juricek and Y.-K. Yap, *Virus Genes*, 2007, **35**, 611–617.
147 A. B. Sigalov and V. N. Uversky, *Self Nonself.*, 2011, **2**, 55–72.
148 L. Lin, Z. Luo, F. Yan, Y. Lu, H. Zheng and J. Chen, *Virus Genes*, 2011, **43**, 90–92.
149 T. Wei, C. Zhang, J. Hong, R. Xiong, K. D. Kasschau, X. Zhou, J. C. Carrington and A. Wang, *PLoS Pathog.*, 2010, **6**, e1000962.
150 R.-H. Wen and M. R. Hajimorad, *Virology*, 2010, **400**, 1–7.
151 P. Vijayapalani, M. Maeshima, N. Nagasaki-Takekuchi and W. A. Miller, *PLoS Pathog.*, 2012, **8**, e1002639.
152 C. Rancurel, M. Khosravi, A. K. Dunker, P. R. Romero and D. Karlin, *J. Virol.*, 2009, **83**, 10719–10736.
153 M. Sorel, J. A. Garcia and S. German-Retana, *Mol. Plant-Microbe Interact.*, 2014, **27**, 215–226.
154 A. Fernández, S. Laín and J. García, *Nucleic Acids Res.*, 1995, **23**, 1327–1332.
155 G. Tavert-Roudet, A. Abdul-Razzak, B. Doublet, J. Walter, T. Delaunay, S. German-Retana, T. Michon, O. Le Gall and T. Candresse, *J. Gen. Virol.*, 2012, **93**, 184–193.
156 M. E. Fairman-Williams, U.-P. Guenther and E. Jankowsky, *Curr. Opin. Struct. Biol.*, 2010, **20**, 313–324.

 This journal is © The Royal Society of Chemistry 2016

157 V. N. Uversky, *Instrum. Anal. Intrinsically Disord. Proteins*, 2014, **2**, 1–42.
158 I. Oruetxebarria, D. Guo, A. Merits and K. Ma, *Virus Res.*, 2001, **73**, 103–112.
159 R. Anindya, S. Chittori and H. S. Savithri, *Virology*, 2005, **336**, 154–162.
160 K. Eskelin, A. Hafrén, K. I. Rantalainen and K. Mäkinen, *J. Virol.*, 2011, **85**, 9210–9221.
161 M. A. Khan, H. Miyoshi, S. Ray, T. Natsuaki, N. Suehiro and D. J. Goss, *J. Biol. Chem.*, 2006, **281**, 28002–28010.
162 C. Robaglia and C. Caranta, *Trends Plant Sci.*, 2006, **11**, 40–45.
163 M. A. Khan, H. Miyoshi, D. R. Gallie and D. J. Goss, *J. Biol. Chem.*, 2008, **283**, 1340–1349.
164 H. Miyoshi, H. Okade, S. Muto, N. Suehiro, H. Nakashima, K. Tomoo and T. Natsuaki, *Biochimie*, 2008, **90**, 1427–1434.
165 T. Michon, Y. Estevez, J. Walter, S. German-Retana and O. Le Gall, *FEBS J.*, 2006, **273**, 1312–1322.
166 R. Gabrenaite-Verkhovskaya, I. A. Andreev, N. O. Kalinina, L. Torrance, M. E. Taliansky and K. Mäkinen, *J. Gen. Virol.*, 2008, **89**, 829–838.
167 M.-L. Rajamäki and J. P. T. Valkonen, *Mol. Plant-Microbe Interact.*, 2003, **16**, 25–34.
168 Z. Gao, E. Johansen, S. Eyers, C. L. Thomas, T. H. Noel Ellis and A. J. Maule, *Plant J.*, 2004, **40**, 376–385.
169 P. Dunoyer, C. Thomas, S. Harrison, F. Revers and A. Maule, *J. virol.*, 2004, **78**, 2301–2309.
170 J. Herold and R. Andino, *Mol. Cell*, 2001, **7**, 581–591.
171 P. Puustinen and K. Mäkinen, *J. Biol. Chem.*, 2004, **279**, 38103–38110.
172 J. Jiang and J.-F. Laliberté, *Curr. Opin. Virol.*, 2011, **1**, 347–354.
173 S. Wittmann, H. Chatel, M. G. Fortin and J. F. Laliberté, *Virology*, 1997, **234**, 84–92.
174 S. Léonard, D. Plante, S. Wittmann, N. Daigneault, M. G. Fortin and J. F. Laliberté, *J. Virol.*, 2000, **74**, 7730–7737.
175 M. C. Schaad, R. J. Anderberg and J. C. Carrington, *Virology*, 2000, **273**, 300–306.
176 C. Beauchemin, N. Boutet and J.-F. Laliberté, *J. Virol.*, 2007, **81**, 775–782.
177 C. Charron, M. Nicolaï, J.-L. Gallois, C. Robaglia, B. Moury, A. Palloix and C. Caranta, *Plant J.*, 2008, **54**, 56–68.
178 B. Culjkovic, I. Topisirovic and K. L. B. Borden, *Cell Cycle*, 2007, **6**, 65–69.
179 C. L. Bokros, J. D. Hugdahl, H. Kim, V. R. Hanesworth, A. Van Heerden, K. S. Browning and L. C. Morejohn, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 7120–7124.
180 M.-L. Rajamäki and J. P. T. Valkonen, *Plant Cell*, 2009, **21**, 2485–2502.
181 T.-S. Huang, T. Wei, J.-F. Laliberté and A. Wang, *Plant Physiol.*, 2010, **152**, 255–266.
182 K. Thivierge, S. Cotton, P. J. Dufresne, I. Mathieu, C. Beauchemin, C. Ide, M. G. Fortin and J.-F. Laliberté, *Virology*, 2008, **377**, 216–225.
183 G. Roudet-Tavert, T. Michon, J. Walter, T. Delaunay, E. Redondo and O. Le Gall, *J. Gen. Virol.*, 2007, **88**, 1029–1033.
184 B. Moury, C. Charron, B. Janzac, V. Simon, J. L. Gallois, A. Palloix and C. Caranta, *Infect., Genet. Evol.*, 2014, **27**, 472–480.
185 V. Ayme, S. Souche, C. Caranta, M. Jacquemond, J. Chadoeuf, A. Palloix and B. Moury, *Mol. Plant-Microbe Interact.*, 2006, **19**, 557–563.
186 B. Moury, C. Morel, E. Johansen, L. Guilbaud, S. Souche, V. Ayme, C. Caranta, A. Palloix and M. Jacquemond, *Mol. Plant-Microbe Interact.*, 2004, **17**, 322–329.
187 M. Bruun-Rasmussen, I. S. Møller, G. Tulinius, J. K. R. Hansen, O. S. Lund and I. E. Johansen, *Mol. Plant-Microbe Interact.*, 2007, **20**, 1075–1082.
188 P. Puustinen, M. Rajama, K. I. Ivanov and J. P. T. Valkonen, *J. Virol.*, 2002, **76**, 12703–12711.
189 A. Hafrén and K. Mäkinen, *J. Gen. Virol.*, 2008, **89**, 1509–1518.
190 J. C. Carrington, D. D. Freed and A. J. Leinicke, *Plant Cell*, 1991, **3**, 953–962.
191 R. Anindya and H. S. Savithri, *J. Biol. Chem.*, 2004, **279**, 32159–32169.
192 J. Phan, A. Zdanov, A. G. Evdokimov, J. E. Tropea, H. K. Peters, R. B. Kapust, M. Li, A. Wlodawer and D. S. Waugh, *J. Biol. Chem.*, 2002, **277**, 50564–50572.
193 P. Sun, B. P. Austin, J. Tözsér and D. S. Waugh, *Protein Sci.*, 2010, **19**, 2240–2251.
194 C. M. Nunn, M. Jeeves, M. J. Cliff, G. T. Urquhart, R. R. George, L. H. Chao, Y. Tscuchia and S. Djordjevic, *J. Mol. Biol.*, 2005, **350**, 145–155.
195 Y. Wang, C. Liu, D. Yang, H. Yu and Y.-C. Liou, *Mol. Cell*, 2010, **37**, 112–122.
196 P. J. Dufresne, K. Thivierge, S. Cotton, C. Beauchemin, C. Ide, E. Ubalijoro, J.-F. Laliberté and M. G. Fortin, *Virology*, 2008, **374**, 217–227.
197 X. Wang, Z. Ullah and R. Grumet, *Virology*, 2000, **275**, 433–443.
198 L. Liljas, in *Flexible Viruses*, ed. S. Longhi and V. N. Uversky, John Wiley & Sons, Inc, 2012, pp. 35–46.
199 K. I. Ivanov and K. Mäkinen, *Curr. Opin. Virol.*, 2012, **2**, 712–718.
200 P. Ni and K. C. Cheng, *Virology*, 2013, **446**, 123–132.
201 V. Dolja, R. Haldeman-Cahill, A. Montgomery, K. Vandenbosch and J. Carrington, *Virology*, 1995, **206**, 1007–1016.
202 R. Anindya and H. S. Savithri, *Virology*, 2003, **316**, 325–336.
203 D. D. Shukla, P. M. Strike, S. L. Tracy, K. H. Gough and C. W. Ward, *J. Gen. Virol.*, 1988, **69**, 1497–1508.
204 L. A. Baratova, A. V. Efimov, N. Eugenie, N. V. Fedorova, R. Hunt, G. A. Badun, A. L. Ksenofontov, L. Torrance and E. N. Dobrov, *J. Virol.*, 2001, **75**, 9696–9702.
205 A. Voloudakis, C. Malpica, M.-E. Aleman-Verdaguer, D. Stark, C. Fauquet and R. Beachy, *Arch. Virol.*, 2004, **149**, 699–712.
206 P. L. Atreya, C. D. Atreya and T. P. Pirone, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 7887–7891.

207 S. Blanc, J.-J. López-Moya, R. Wang, S. García-Lampasona, D. W. Thornbury and T. P. Pirone, *Virology*, 1997, **231**, 141–147.
208 B. Shoemaker, J. J. Portman and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 8868–8873.
209 Y. Huang and Z. Liu, *J. Mol. Biol.*, 2009, **393**, 1143–1159.
210 R. F. Allison, W. G. Dougherty, T. D. Parks, L. Willis, R. E. Johnston, M. Kelly and F. B. Armstrong, *Virology*, 1985, **147**, 309–316.
211 C. Desbiez, C. Chandeysson and H. Lecoq, *Mol. Plant Pathol.*, 2014, **15**, 217–221.
212 K. I. Ivanov, P. Puustinen, A. Merits, M. Saarma and K. Mäkinen, *J. Biol. Chem.*, 2001, **276**, 13530–13540.
213 K. I. Ivanov, P. Puustinen, R. Gabrenaite, H. Vihinen, L. Rönnstrand, L. Valmu, N. Kalkkinen and K. Mäkinen, *Plant Cell*, 2003, **15**, 2124–2139.
214 Z. W. Subr, M. Kamencayová, S. Nováková, A. Nagyová, J. Nosek and M. Glasa, *Arch. Virol.*, 2010, **155**, 1151–1155.
215 D. Chen, S. Jua, L. Hartweck, J. M. Alamillo, C. Simo and J. J. Pe, *J. Virol.*, 2005, **79**, 9381–9387.
216 A. Hafrén, D. Hofius, G. Rönnholm, U. Sonnewald and K. Mäkinen, *Plant Cell*, 2010, **22**, 523–535.

 This journal is © The Royal Society of Chemistry 2016

Supplementary data

TABLE S1. List of sequence accessions used in this study. TaxID : Taxonomic Identifier in NCBI database (http://www.ncbi.nlm.nih.gov/). GI : GenBank Identifier. Symbols "*" correspond to sequences retrieved from SharCo database (http://www.sharco.eu/).

Species Name	Acronym	TaxID	NCBI Sequences Accessions (GI)
Bean Yellow Mosaic Virus	**BYMV**	**12197**	***Total sequences : 16***
			163930863 ; 218664786 ; 1354081 ; 19881394 ; 1217902 ; 37780263 ; 238035114 ; 109638255 ; 109638257 ; 238035116 ; 238035120 ; 109390061 ; 388240710 ; 238035118 ; 402756072 ; 562745024
Lettuce Mosaic Virus	**LMV**	**12202**	***Total sequences : 23***
			643433345 ; 643433351 ; 643433389 ; 643433371 ; 641483757 ; 1903232 ; 641483753 ; 643433365 ; 643433347 ; 643433349 ; 643433373 ; 643433357 ; 28193444 ; 9864420 ; 643433359 ; 643433353 ; 643433375 ; 643433385 ; 643433381 ; 643433377 ; 643433355 ; 643433367 ; 643433361
Plum Pox Virus	**PPV**	**12211**	***Total sequences : 36***
			90102165 ; 89143150 ; TR98ap* ; GR19ap* ; AL14pe* ; AL11pl* ; TR322ap* ; 5650729 ; AL120my* ; FR67pe* ; 190576805 ; AL70ap* ; AL80pl* ; RO72ap* ; 531731 ; 222408 ; 62766784 ; CZ54pl* ; 294769532 ; SK23pe* ; RO22pl* ; FR66pl* ; 157419767 ; 294769530 ; CZ14my* ; 294769534 ; RO25pl* ; CN2ja* ; CZ7pl* ; FR64pl* ; RS67pl* ; 340025614 ; 40252370 ; 327335415 ; 62766782 ; 327335419
Papaya Ringspot Virus	**PRSV**	**12205**	***Total sequences : 27***
			116109588 ; 121615481 ; 573465033 ; 27542791 ; 329763105 ; 29469899 ; 168997615 ; 87130665 ; 14581404 ; 551702287 ; 157399685 ; 258106 ; 61350 ; 9629244 ; 161168458 ; 195975850 ; 1771470 ; 408362879 ; 87130663 ; 23680941 ; 408362875 ; 85690827 ; 408362871 ; 85690823 ; 408362873 ; 85690825 ; 408362877
Potato Virus Y	**PVY**	**12216**	***Total sequences : 39***
			209395183 ; 482427295 ; 58613914 ; 27371967 ; 336318906 ; 186892492 ; 189187421 ; 56462531 ; 85376485 ; 158347632 ; 90968457 ; 90968455 ; 482661424 ; 482427275 ; 383101016 ; 482661426 ; 482661420 ; 482661432 ; 482661430 ; 257122583 ; 21913301 ; 336318968 ; 340815945 ; 340815947 ; 157679488 ; 116664524 ; 356707357 ; 465487346 ; 465487403 ; 307094821 ; 116664522 ; 307094819 ; 482661434 ; 116664526 ; 242276192 ; 336318898 ; 336318902 ; 336318954 ; 465487447
Soybean Mosaic Virus	**SMV**	**12222**	***Total sequences : 37***
			525569144 ; 525569141 ; 256587878 ; 253297 ; 256587860 ; 256587826 ; 34304610 ; 224593245 ; 32452358 ; 256587876 ; 256587838 ; 256587854 ; 256587856 ; 256587852 ; 256587828 ; 256587874 ; 27877110 ; 256587822 ; 256587842 ; 256587830 ; 256587836 ; 348499965 ; 348499963 ; 256587868 ; 254679812 ; 11995002 ; 226525453 ; 256587840 ; 256587848 ; 364285301 ; 519767289 ; 315419694 ; 519767282 ; 301087113 ; 261872238 ; 301087111 ; 364285303
Sugarcane Mosaic Virus	**SCMV**	**12224**	***Total sequences : 15***
			402216201 ; 27528454 ; 402695592 ; 402695594 ; 402216199 ; 20270985 ; 18621142 ; 18652414 ; 21632589 ; 24415606 ; 77403451 ; 159032828 ; 295647310 ; 340629655 ; 45774601
Turnip Mosaic Virus	**TuMV**	**12230**	***Total sequences : 38***
			460837895 ; 15055163 ; 460837983 ; 460837911 ; 460837903 ; 460837905 ; 33146222 ; 33146220 ; 114217198 ; 460837927 ; 114217152 ; 109706892 ; 460837989 ; 114217188 ; 216409654 ; 33146248 ; 460837947 ; 460837973 ; 109706890 ; 114217178 ; 124507406 ; 33146246 ; 114217234 ; 460837909 ; 114217154 ; 460837995 ; 460837965 ; 114217228 ; 114217214 ; 33146226 ; 114217158 ; 216409656 ; 33146218 ; 460837957 ; 460837907 ; 33146230 ; 114217208 ; 296531308
Watermelon Mosaic Virus	**WMV**	**146500**	***Total sequences : 33***
			327537117 ; 161168460 ; 540342156 ; 194399684 ; 326468188 ; 194399688 ; 194399698 ; 326468200 ; 194399702 ; 326468184 ; 326468182 ; 41058913 ; 51949945 ; 194399692 ; 194399700 ; 194399682 ; 194399704 ; 326468204 ; 268628057 ; 101918494 ; 326468190 ; 194399694 ; 326468186 ; 194399680 ; 194399690 ; 326468194 ; 194399686 ; 326468196 ; 326468198 ; 326468202 ; 88954063 ; 326468192 ; 194399696
Zucchini Mosaic Virus	**ZYMV**	**12232**	***Total sequences : 24***
			13940781 ; 466347 ; 17019510 ; 17059637 ; 73621258 ; 84993539 ; 84993541 ; 532235175 ; 532235177 ; 532235179 ; 532235181 ; 532235183 ; 532235187 ; 532235171 ; 532235173 ; 532235185 ; 118566318 ; 118566320 ; 37778797 ; 388542373 ; 33391222 ; 33391224 ; 161168462 ; 33391220

TABLE S2. Eukaryotic Linear Motifs (ELMs) highly conserved among proteins of 10 potyvirus species. ELM predictions are retrieved from Hagai et al. 2014 (see Material and Methods). Their names refer to ELM database resource. [(1)] Conservation at inter-species level. Correspond to number of species in which corresponding motif is predicted. [(2)] Motifs listed in ELM database as experimentally described and functionally validated in plants and/or viruses. *PTM: post-translational modification; ND: not listed in ELM database.*

ELM Name	Functional class	ELM conservation in potyviral protein[(1)]	Present in taxon	Described in plants[(2)]	Described in viruses[(2)]	Relevance in plant virus context
CLV_C14_Caspase3-7	Cleavage	P1(90%) ; CP(90%)	Metazoa	ND	ND	×
CLV_NDR_NDR_1	Cleavage	P1(100%) ; VPg(60%)	Metazoa	ND	ND	×
CLV_PCSK_PC1ET2_1	Cleavage	P1(100%)	Vertebrates	ND	ND	×××
CLV_PCSK_SKI1_1	Cleavage	P1(100%)	Vertebrates	ND	ND	×××
LIG_14-3-3_3	**Ligand**	**P1(90%) ; CP(90%)**	**Eukaryota**	***Arabidopsis thaliana***	***Mouse polyomavirus***	**✓✓✓**
LIG_Clathr_ClatBox_1	Ligand	P1(90%)	Eukayota	ND	ND	×
DOC_CYCLIN_1	Docking Ligand	P1(90%)	Eukaryota	ND	*Human papillomavirus*	✓
LIG_FHA_1	**Ligand**	**P1(90%)**	**idem**	***Arabidopsis thaliana***	**ND**	**✓✓✓**
LIG_FHA_2	Ligand	P1(100%); VPg(70%) ; CP(100%)	Eukaryota	ND	ND	×
DOC_MAPK_1	Docking Ligand	P1(90%)	Eukaryota	ND	ND	×
LIG_SH2_STAT5	Ligand	P1(80%); VPg(60%)	Metazoa	ND	ND	×
LIG_SH3_3	Ligand	P1(80%)	Metazoa	ND	*Influenza A virus ; Hepatitis E virus Herpesvirus saimiri ;*	✓
LIG_USP7_1	Ligand	P1(100%) ; CP(80%)	Eukaryota	ND	*Human Herpes virus*	✓
DOC_WW_Pin1_4	**Docking Ligand**	**P1(100%)**	**Eukaryota**	***Arabidopsis thaliana***	***Hepatitis B virus***	**✓✓✓**
MOD_CK1_1	PTM	P1 ; VPg(70%) ; CP(80%)	Eukaryota	ND	ND	×
MOD_CK2_1	PTM	P1(100%) ; CP(100%)	Eukaryota	ND	*Human immunodeficiency virus*	✓
MOD_GlcNHglycan	PTM	P1(90%)	Metazoa	ND	ND	×
MOD_GSK3_1	PTM	P1 ; VPg(70%) ; CP(100%)	Eukaryota	ND	ND	×
MOD_PIKK_1	PTM	P1(100%)	Eukaryota	ND	ND	×
MOD_PKA_2	**PTM**	**P1 ; VPg(80%) ; CP(80%)**	**Eukaryota**	***Arabidopsis thaliana***	**ND**	**✓✓✓**
MOD_ProDKin_1	PTM	P1(100%)	Eukaryota	ND	ND	×
TRG_NLS_MonoExtN_4	**Targeting**	**P1(80%); VPg (80%)**	**Eukaryota**	**ND**	***Avian Sarcoma virus ; Murine polyomavirus ; Simian virus 40 ; Cercopithecine herpesvirus ; Human herpesvirus ; Nipah virus***	**✓✓✓**

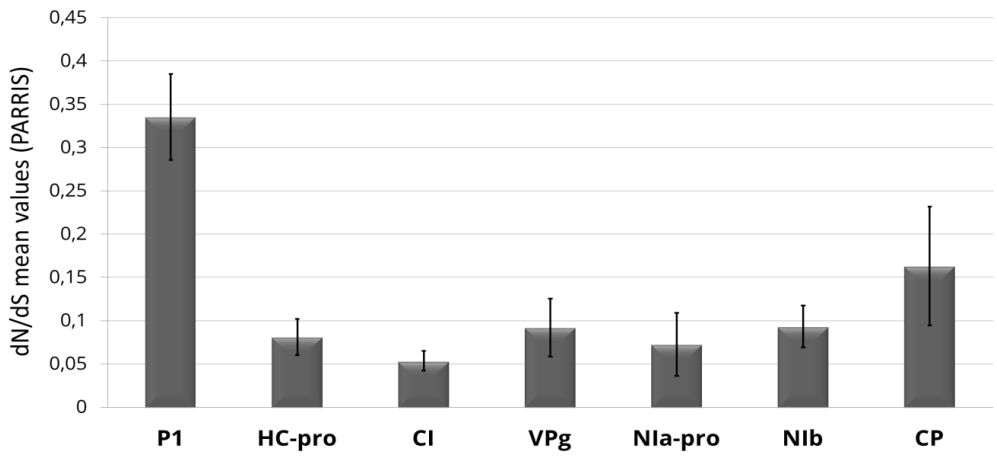


FIG S1|Mean evolutionary constraints (dN/dS ratio or ω value) exerted on Potyviral protein from 10 Potyvirus species. dN/dS ratio were calculated with PARRIS method on each species protein dataset. Bar scaled represent variation of mean dN/dS value between species dataset.

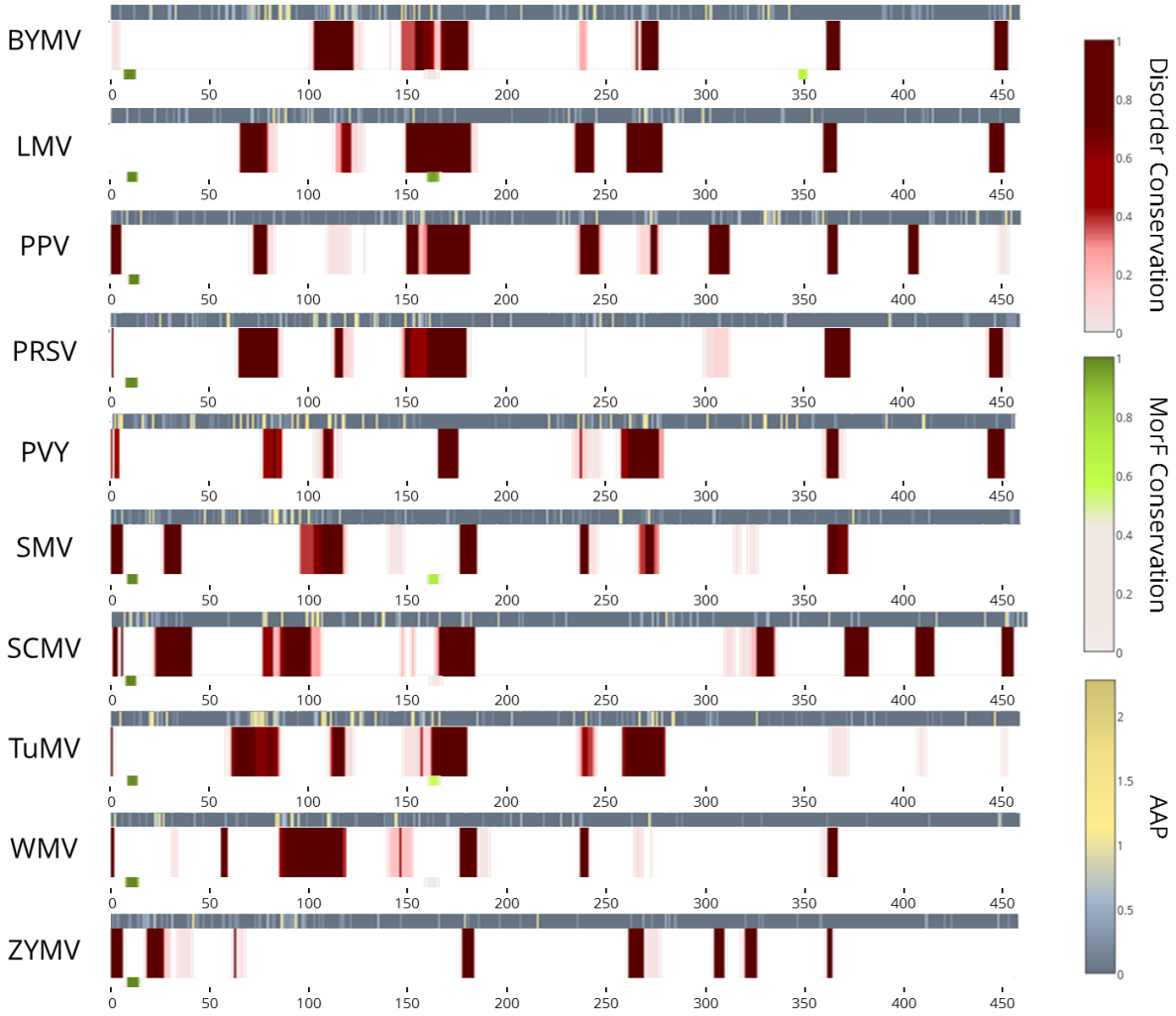


FIG S2.A|Intrinsic disorder conservation of HC-pro protein. For all figures of S2, white-to-red gradation bar represent degree of disorder conservation, from 0% (white) to 100% (dark red). White-to-green bar represent MoRFs conservation signal, from 0% (white) to 100% (dark green). Blue-to-yellow bar represents amino acid polymorphism (AAP). By definition, 0 to 1 represent highly conserved position, 1 to 2 is considered as conserved and higher than 2 is considered as variable.

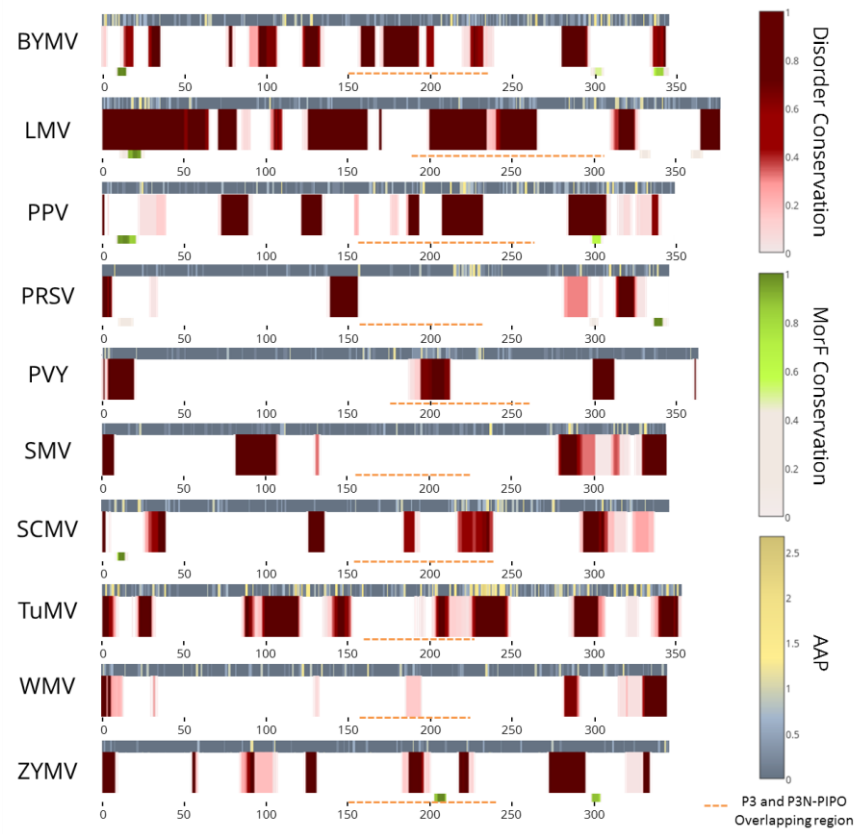


FIG S2.B|Intrinsic disorder conservation of P3 protein.

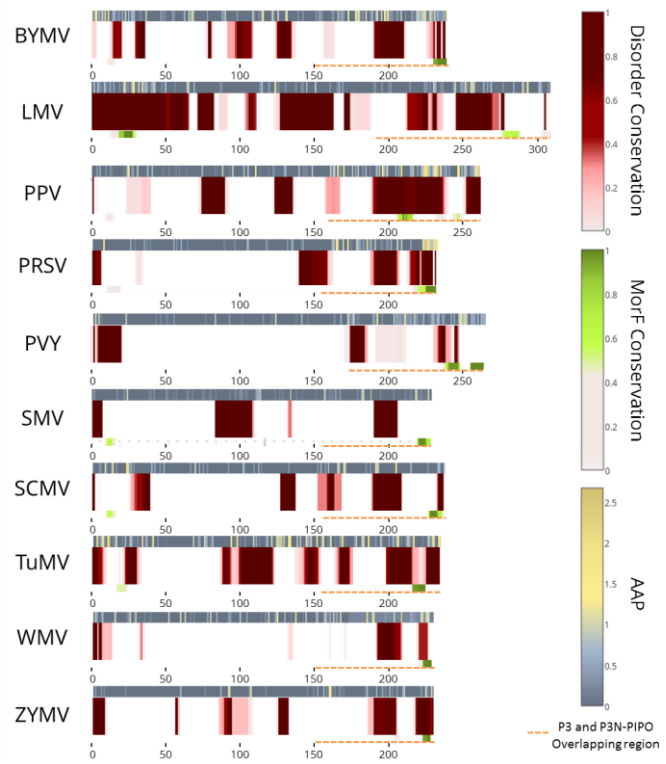


FIG S2. C| Intrinsic disorder conservation of P3N-PIPO protein.

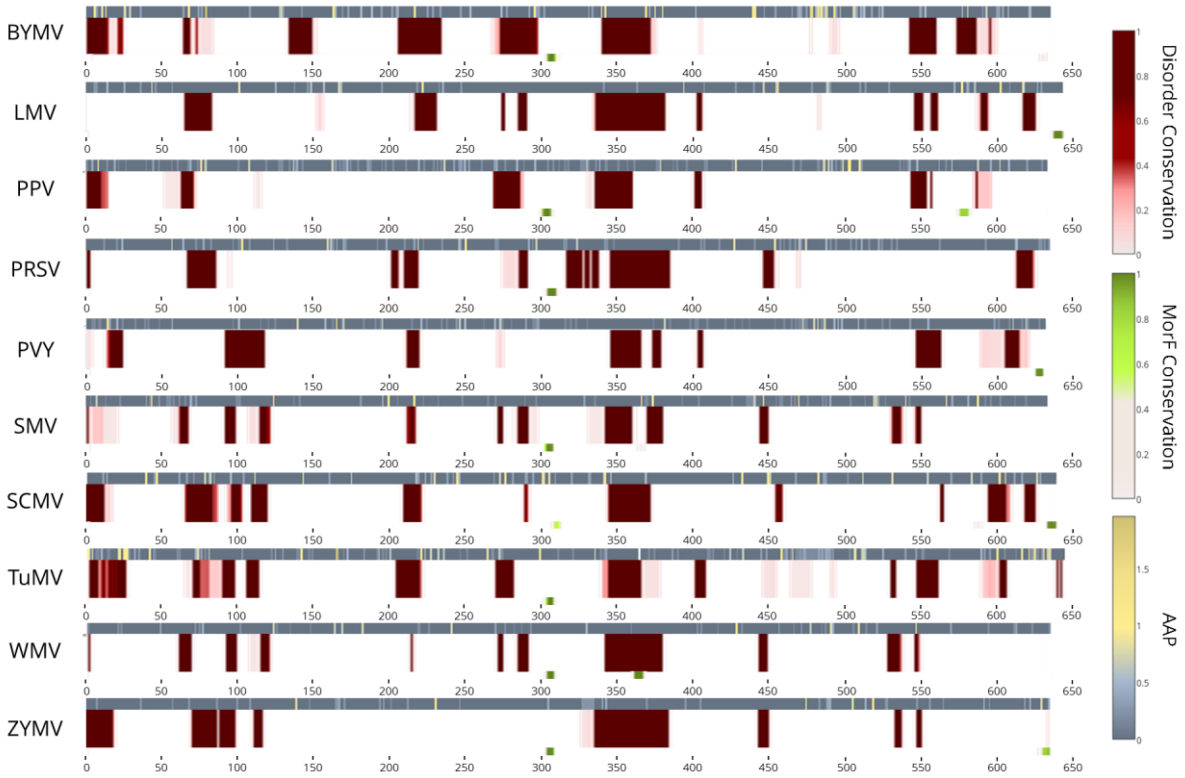


FIG S2.D|Intrinsic disorder conservation of CI protein.

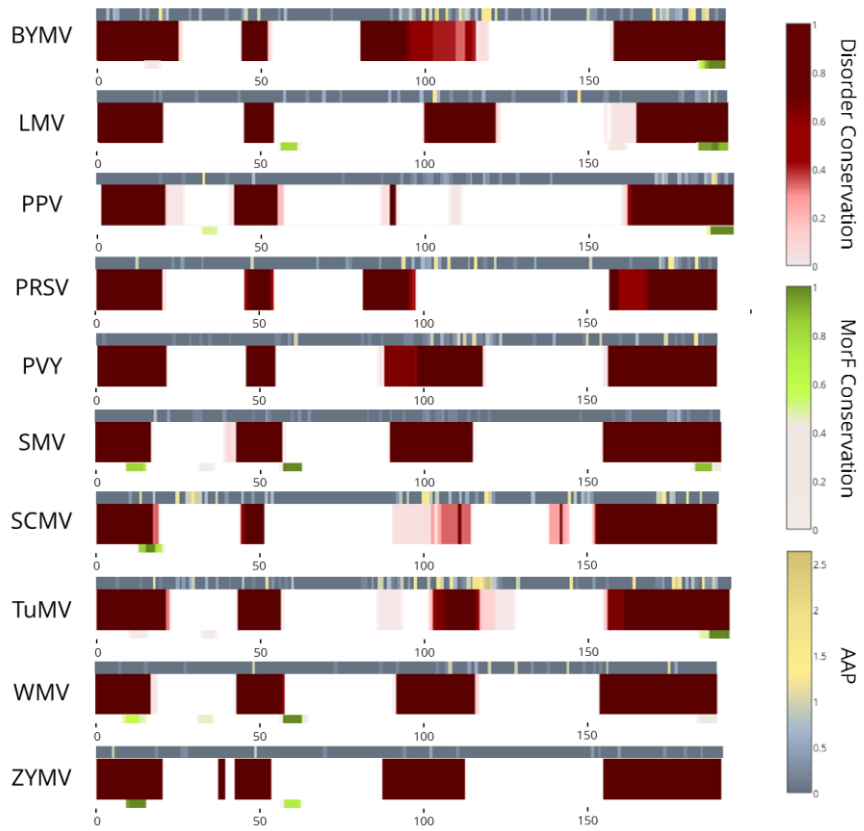


FIG S2.E|Intrinsic disorder conservation of VPg protein.

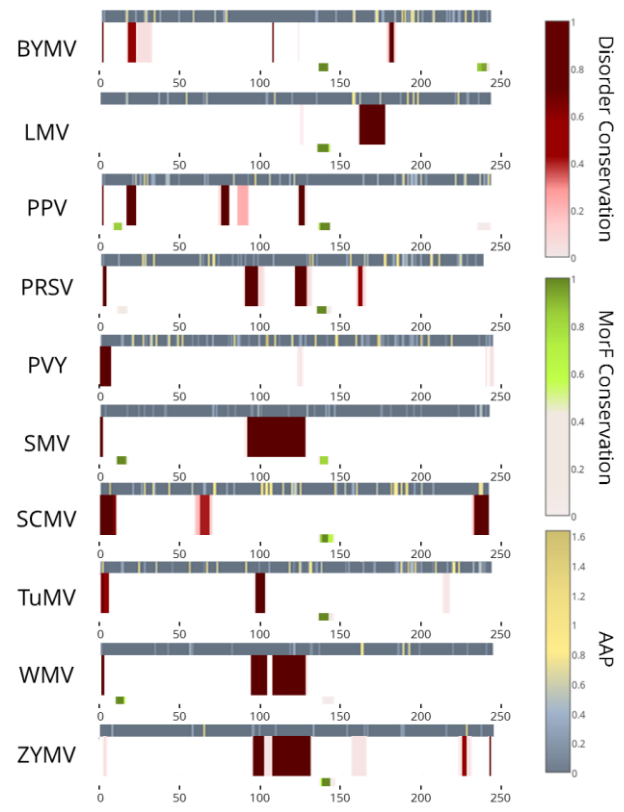


FIG S2.F|Intrinsic disorder conservation of NIa-pro protein.

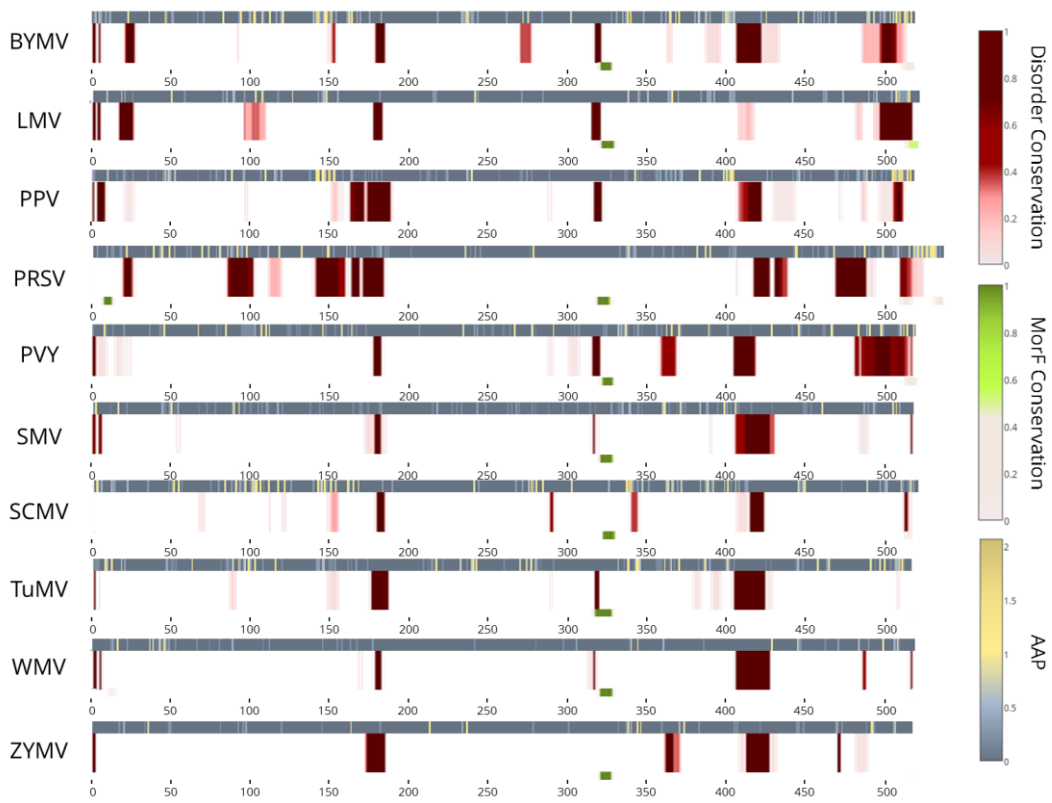


FIG S2.G|Intrinsic disorder conservation of NIb protein.

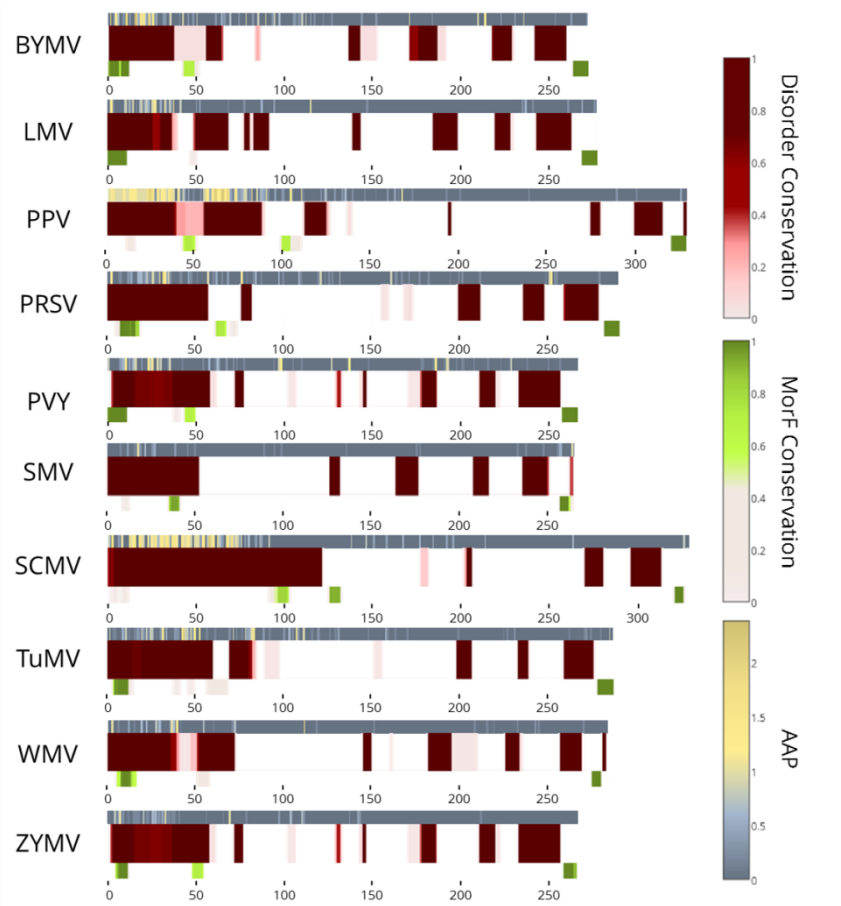


FIG S2.H|Intrinsic disorder conservation of Coat Protein (CP).

Chapitre II

Etude de l'implication du désordre dans le maintien de la variabilité génétique d'une population virale *via* l'analyse de données de séquençage à haut débit.

1. Contexte de l'étude

Comme présenté en introduction, les récentes évidences expérimentales reliant la diversité des populations de virus à ARN à leur capacité à infecter de nouveaux hôtes (adaptation indirecte, Partie Introduction Figure 14A) illustrent l'importance de pouvoir décrire cette diversité et le réseau de variants de ces populations.

Le séquençage haut-débit : un outil de choix dans l'identification et la compréhension des déterminants de l'adaptation virale. Les progrès technologiques associés au séquençage profond ou *Next Generation Sequencing* (NGS) permettent d'accéder aux séquences des génomes complets de l'ensemble des variants viraux – y compris les plus minoritaires - qui composent une population (Schadt et al. 2010).
Ainsi, de nombreuses études portant sur la description exhaustive des variants contenus dans des populations virales ont vu le jour ces dernières années (Orton et al. 2013; Wright et al. 2011; Acevedo et al. 2014; Seifert et al. 2014). Ce passage de la théorie à la description contribue donc activement à augmenter notre compréhension des processus, encore assez méconnus, qui façonnent la structure génétique des populations virales et déterminent leur potentiel adaptatif.
Chez les virus de plante également, de telles approches ont très récemment été utilisées pour décrire les structures génétiques de populations virales évoluant dans un hôte donné, notamment en ce qui concerne notre modèle d'étude, le genre *Potyvirus* (Kutnjak et al. 2015; Dunham et al. 2014; Cuevas et al. 2015).

En se basant sur les données de séquençage obtenues pour chacune de ces 3 études, l'objectif de ce chapitre est d'estimer l'impact du désordre intrinsèque sur l'accumulation de la diversité génétique d'une population virale, en contexte d'infection sur hôte sensible. Autrement dit, cette étude vise à analyser si les zones prédites comme désordonnées sont plus promptes à accueillir des substitutions en acides aminés (mutations non-synonymes[1]) que les zones plus structurées, comme le laissent supposer les faibles contraintes topologiques associées aux IDR.

[1]*A l'échelle des protéines, les mutations synonymes n'ont, par définition, pas d'impact structural et fonctionnel. Selon le principe du ratio dN/dS, ces mutations synonymes ne rendent pas compte de la contrainte exercée sur une protéine mais servent plutôt à quantifier les changements « silencieux » non soumis aux forces évolutives. Cependant, il parait important de noter que l'apparition de mutations synonymes pourraient, en lien avec l'adaptabilité des virus à ARN, avoir un effet sur les structures secondaires du génome à ARN et le potentiel évolutif de par la volatilité des codons* (Lauring et al. 2012; Lauring et al. 2013; Acevedo et al. 2014)*. Cependant, notre étude portant sur le désordre, à l'échelle des protéines, le rôle des mutations synonymes dans l'adaptabilité ne sera pas abordé dans la suite de ce travail.*

	Dunham et al. (2014)	Kutnjak et al. (2015)	Cuevas et al. (2015)
Potyvirus	Zucchini Yellow Mosaic Virus (ZYMV)	Potato Virus Y (PVY)	Tobacco Etch Virus (TEV)
Hôte	Cucurbita *pepo* (hôte naturel)	Solanum *Tuberosum* (hôte naturel)	Nicotiana *Tabacum* (hôte naturel) Capsicum *annuum* (hôte non naturel)
Mode d'inoculation	Mécanique	Mécanique	Mécanique (calibration de l'inoculum)
Modalités d'échantillonnage			
• **Nombre de plantes analysées (âge)**	Une plante (0 à 2 mois).	60 plantes (3 semaines post inoculation)	2 réplicas biologiques par hôte x 15 passages séquentiels
• **Prélèvements dans le temps**	Oui	Non	Oui (16 passages)
• **Prélèvements dans l'espace**	Oui (Prélèvement séquentiel de 23 feuilles matures)	Non (Pool des parties vertes).	Oui (3 feuilles par plantes)
Modalités de séquençage			
• **Technologie de séquençage**	Illumina	Illumina Hiseq 2000	Illumina Hiseq 2000
• **Seuil de fréquence des variants**	1%	1%	1%
• **Utilisation de réplicas techniques**	Non	Oui	Oui (7 PCR et RT indépendantes qui sont ensuite poolées)
Résultats de séquençage			
• **Profondeur moyenne**	27150x	31298x	25840x (NT) 32821x (CA)

Tableau II.1 | Présentation des modes opératoires utilisés dans les 3 études de suivi des populations de potyvirus évoluant sur hôtes sensibles. *CA : Capsicum annuum ; NT : Nicotiana tabacum*.

2. Résultats

La nature des mutations, leur position sur le génome viral ainsi que leurs fréquences ont été analysées par rapport aux prédictions du désordre intrinsèque dans les 3 protéomes considérés.
Les résultats présentés dans cette partie ont donc été obtenus en parallèle, pour chacun des 3 jeux de données. Ils nous permettent de discuter le déterminisme du désordre dans l'accumulation des mutations non-synonymes identifiées. Les biais induits par les différentes méthodes, ainsi que la pertinence générale de ce type d'analyse seront également discutés.

Ces 3 études sont conduites sur des pathosystèmes différents. La conception de chacune des expériences, ainsi que les techniques de séquençage utilisées et les principaux résultats sont fournis dans le tableau II.1.

A) Les protéines des potyvirus n'ont pas toutes la même propension à accumuler des mutations à l'échelle de l'évolution intra-hôte

Les différences de contraintes exercées sur les protéines, à large échelle évolutive, sont-elles observables à l'échelle de l'évolution intra-hôte ?
A de larges échelles évolutives, les données de variabilité naturelle existant au sein d'une même espèce de potyvirus témoignent que toutes les protéines n'accumulent pas les mutations au même taux (Chapitre 1). Certaines sont dites plus permissives que d'autres, et plus propices à la diversité, comme l'illustrent les résultats de dN/dS présentés dans le Chapitre 1 et obtenus à partir de 10 espèces de potyvirus différents. Ainsi, des protéines comme P1 et CP sont soumises à moins de contraintes et évoluent plus rapidement que d'autres, comme la réplicase NIb.
Notre étude *in silico* montre également que le désordre était associé de façon significative à moins de contraintes évolutives, permettant à de telles régions d'évoluer plus rapidement.
Les séquences constituant les jeux de données ayant conduit à cette observation étaient issues de prélèvements en des lieux et temps différents. Les variations observées entre ces séquences résultaient donc de l'action de l'ensemble des contraintes sélectives exercées sur les populations depuis la divergence des isolats échantillonnés.

Dans ce contexte, il parait intéressant de déterminer si de telles propriétés pourraient être observées à une échelle de temps et d'espace beaucoup plus réduite, à savoir au cours de l'infection virale sur hôte sensible. Ce travail vise donc à répondre aux questions suivantes : **i. Est-ce que cette courte échelle évolutive permet d'ores et déjà d'observer des différences en termes de capacités évolutives entre les différentes protéines ? ii. De telles différences peuvent-elles être corrélées aux propriétés désordonnées de certaines régions protéiques ?**

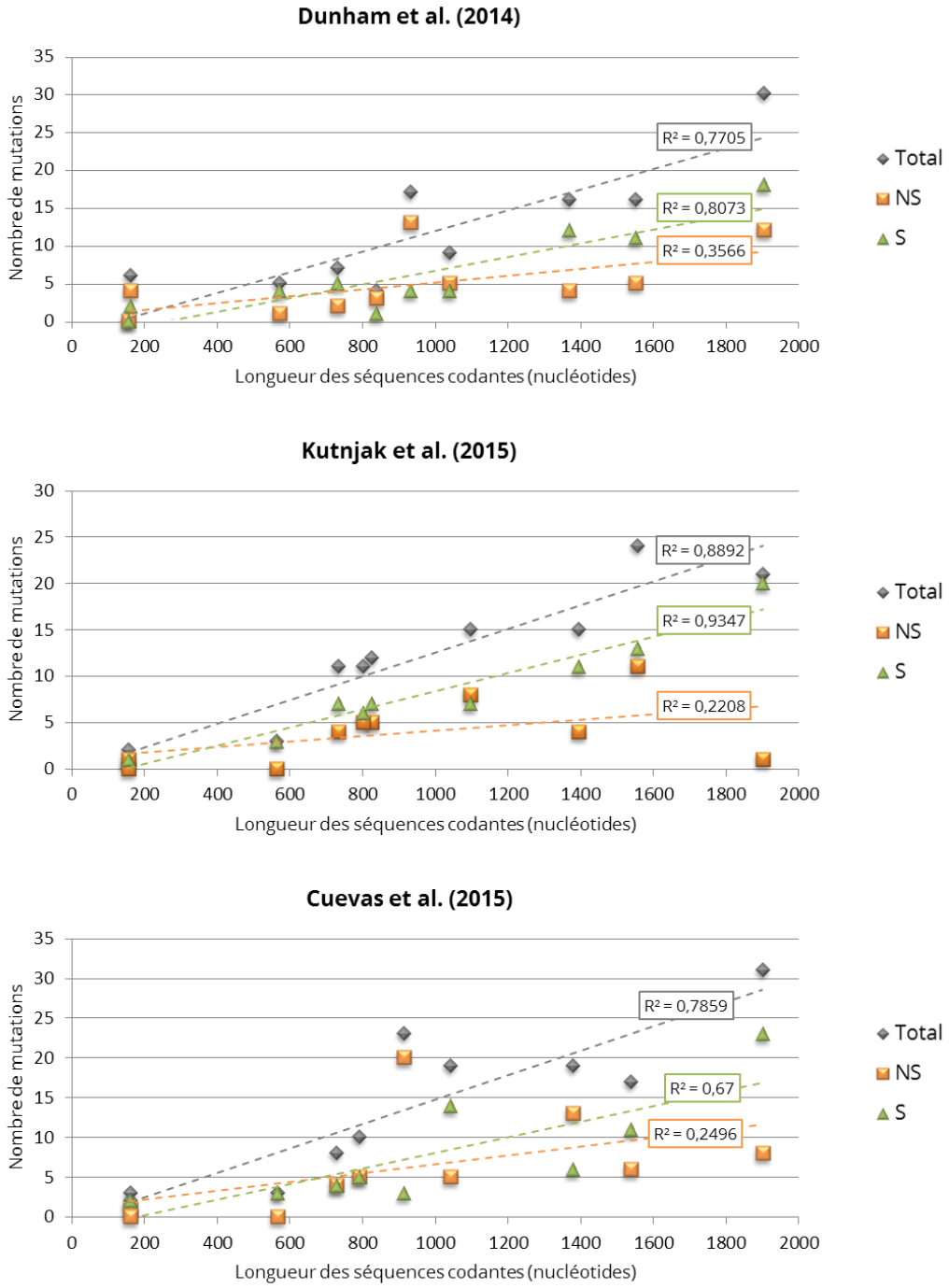


Figure II.1|Accumulation mutationnelle de chaque séquence codant pour une protéine virale. Le nombre de mutations est porté en fonction de la longueur de la séquence codant chaque protéine, et ce pour chaque type de mutations (totales - gris, non-synonymes NS-orange et synonymes S-vert). Le coefficient R^2 est obtenu par régression linéaire et correspond au degré de corrélation entre nombre de mutations et longueur de la séquence nucléotidique.

L'accumulation de mutations non-synonymes à l'échelle d'une infection intra-hôte est-elle protéine-dépendante ?

Afin de déterminer si à l'échelle de l'hôte, les protéines de potyvirus présentent ou non la même propension à accumuler des mutations induisant un changement d'acides aminés (mutations non-synonymes), le nombre de mutations obtenues par séquence codant chacune des protéines a été reporté en fonction de la taille en nucléotides de celles-ci. Les résultats sont présentés sur les graphes Figure II.1 et montrent que, en ce qui concerne l'accumulation de mutations totales, une corrélation significative ($R^2 >= 0.8$; $P < 0.05$) est obtenue entre la taille des séquences nucléotidiques codant pour les différentes protéines et le nombre de mutations qu'elles présentent, et ce pour l'ensemble des 3 jeux de données analysés. En revanche, lorsque l'on discrimine les mutations synonymes des non-synonymes, on observe une corrélation significative pour les mutations synonymes ($R^2>=0.7$; $P < 0.005$) tandis que l'accumulation de mutations non-synonymes ne semble pas corrélée à la longueur de la séquence des différentes protéines ($R^2<=0.4$; $P > 0.05$). Ces résultats montrent que l'accumulation de mutations non-synonymes lors de l'évolution intra-hôte est protéine-dépendant, ce qui n'est en revanche pas le cas pour les mutations synonymes, qui se distribuent de façon homogène le long du génome viral.

Quelles sont les protéines qui ont tendance à accumuler plus de mutations ?

Pour chaque protéine, la fréquence de mutations (définie par le nombre de mutations totales observées dans la séquence codante de la protéine divisée par la longueur de cette séquence) a été déterminée. Ces fréquences sont rapportées sur les graphes de la Figure II.2. La faible taille de certaines protéines (en particulier 6K1 et 6K2) ainsi que le faible nombre de mutations observées sont parfois limitant pour conclure quant à la significativité des différences de fréquences observées. Cependant, on note que **la protéine P1 accumule significativement plus de mutations non-synonymes que les protéines HC-Pro, P3, CI, NIb et CP lors de l'évolution du TEV sur piment et tabac** (étude sur Cuevas *et al.* 2015) ($P < 0.02$; Test de Z). De la même façon, **P1 accumule préférentiellement les substitutions en acides aminés lors de l'évolution du ZYMV dans l'hôte *Cucurbita pepo*** ($P < 0.04$, Test de Z).

Cependant, aucune tendance significative n'a été observée en ce qui concerne l'accumulation de mutations Non-Synonymes (NS) lors de l'analyse des variants formés au cours de l'évolution du PVY dans *S.tuberosum*.

Comme présenté en introduction, la propriété de chaque protéine à accumuler des mutations induisant un changement en acides aminés (*i.e* capacité à évoluer) peut avoir différents déterminants.

Le désordre intrinsèque présente une répartition/abondance hétérogène entre les différentes protéines de potyvirus (cf. Chapitre 1). La protéine P1 figure parmi les protéines fortement désordonnées chez l'ensemble des potyvirus.

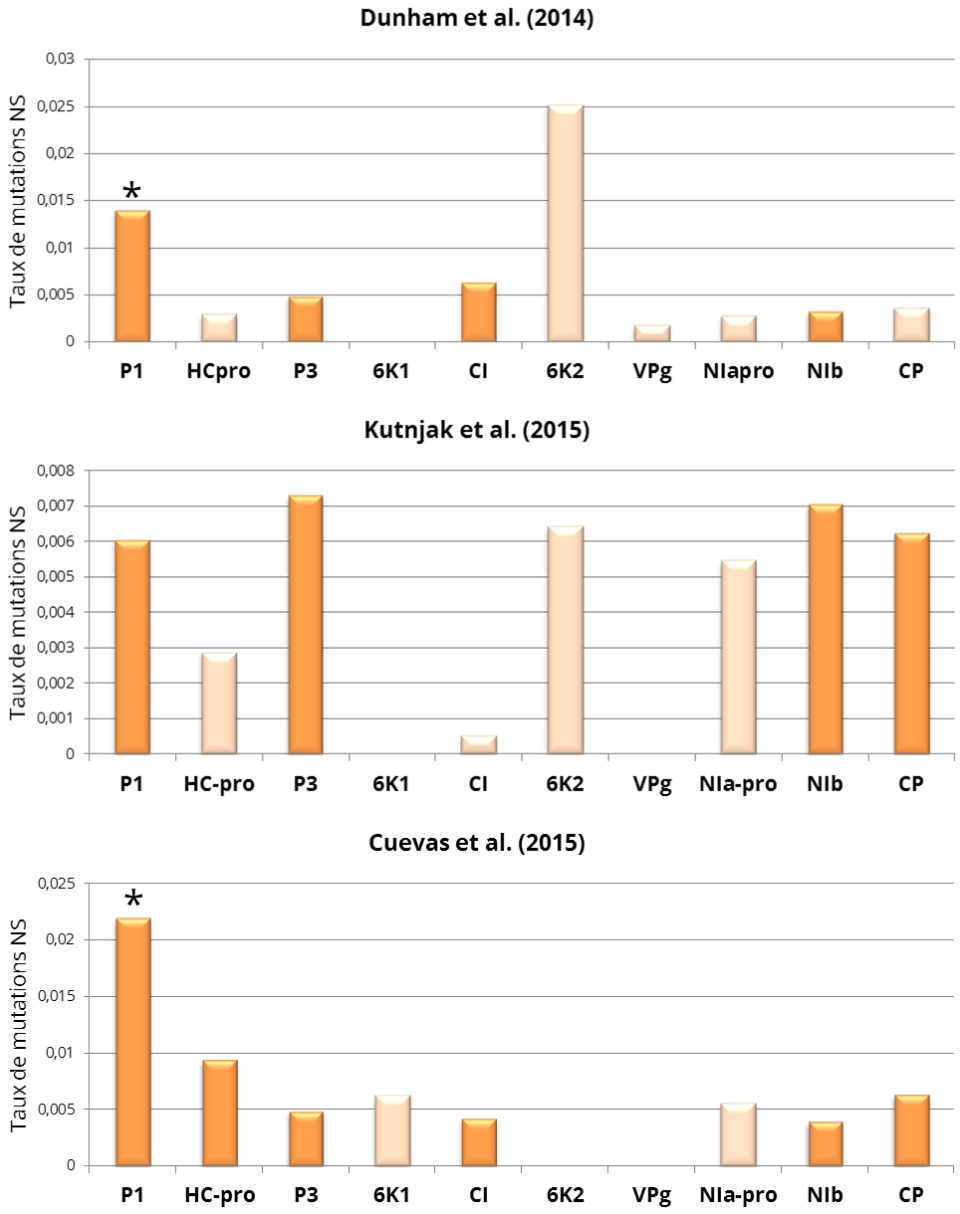


Figure II.2|Fréquence de mutations non-synonymes observées par protéine. Les fréquences de mutations correspondent au nombre de mutations observée dans chaque protéine par rapport à la longueur totale en nucléotides de leur séquence codante. Les barres orange clair correspondent aux protéines dont le nombre de mutations est inférieur à 5 et ne permet donc pas de réaliser de test statistique. Le symbole "*" correspond aux valeurs significativement différentes de toutes les autres (P<0.05) selon le test de Z. *NS : Non-Synonyme.*

Peut-on relier l'abondance en désordre observée dans la protéine P1 à sa capacité à accumuler plus de mutations que les autres protéines ? De façon générale, le désordre intrinsèque peut-il constituer un déterminant de la capacité évolutive des protéines de potyvirus à l'échelle de l'évolution intra-hôte ? Pour tenter de répondre à ces questions, chaque mutation non-synonyme sera analysée selon l'environnement structural dans lequel elle est apparue (région ordonnée ou désordonnée).

Les régions désordonnées accumulent-elles plus de mutations non-synonymes ?

La suite de l'étude consiste donc à analyser si le désordre peut constituer un déterminant de l'accumulation de mutations NS dans les protéines de potyvirus à cette échelle évolutive. Ainsi, pour chacune des trois études, le désordre des protéines codées par les génomes de référence (ayant servi à inoculer les plantes au départ) est prédit via l'utilisation du prédicteur PONDR-VLXT. Au cours des expériences, les échantillons de chaque prélèvement sont séquencés, comparés aux séquences de référence. A chaque mutation identifiée, nous avons attribué un statut « ordonné » (OD) ou « désordonné » (ID) selon qu'elle est localisée dans l'un ou l'autre type de région. Les mutations sont ensuite classées selon **i)** leur localisation ID *versus* OD ; **ii)** leur nature Synonyme *versus* Non-Synonyme. Les résidus désordonnés ne représentent que 20% en moyenne de la totalité de la séquence en acides aminés des polyprotéines de potyvirus (cf. partie 1). Le nombre de mutations est donc rapporté au nombre total de résidus prédits appartenant à chacune des classes Ordre vs Désordre, afin de ne pas biaiser les résultats. Les résultats de la répartition ID vs OD des mutations identifiées sont ainsi reportés sur les graphes de la figure II.3. Pour l'ensemble des 3 études, les mutations non-synonymes tendent à s'accumuler préférentiellement dans les régions désordonnées que dans les régions ordonnées. Cependant, seule l'analyse des jeux de données de Cuevas *et al.* permet de montrer que **les régions désordonnées accumulent significativement plus de mutations non-synonymes que les régions ordonnées** ($P < 0.02$, Test de Z).

Si on tient maintenant compte de l'effet hôte-dépendant, les résultats sont différents entre l'évolution du TEV sur hôte *N. tabacum* (NT) (hôte naturel) et celle observée sur hôte piment *C.annuum* (CA) (nouvel hôte). Le nombre total de mutations observées est nettement plus important pour l'expérience d'évolution conduite sur hôte naturel (43 mutations non-synonymes au total pour NT contre 19 pour CA). En lien direct ou pas avec cette observation, l'accumulation significativement plus importante de mutations non-synonymes dans les zones désordonnées est observée uniquement dans la condition d'infection sur l'hôte naturel NT ($P < 0,04$, Test de Z).

Ainsi, lors de l'évolution du TEV dans son hôte naturel, les régions désordonnées accumulent plus de mutations non-synonymes que les régions ordonnées.

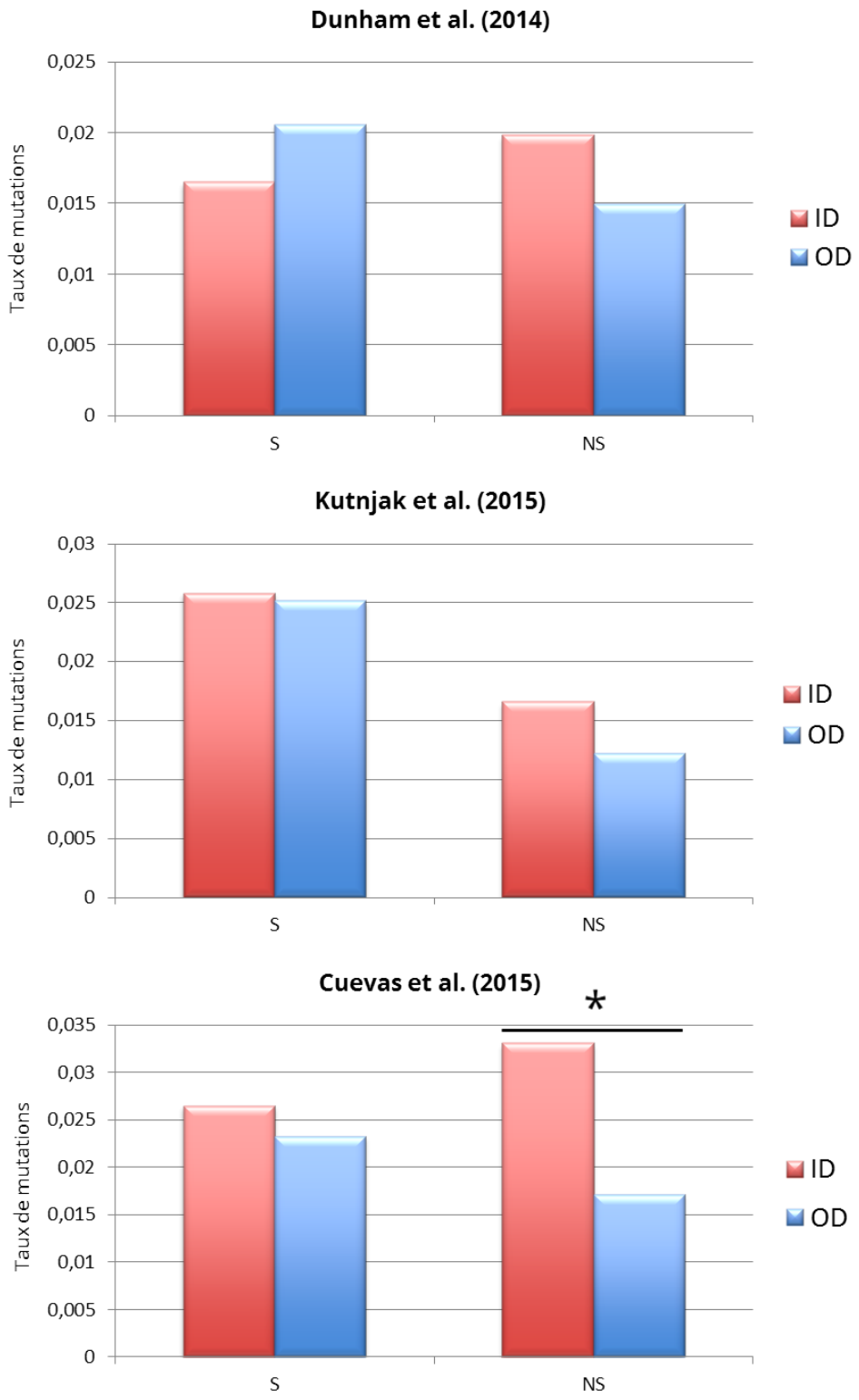


Figure II.3|Fréquence de mutations synonymes (S) et non-synonymes (NS) des segments génomiques codant pour les régions désordonnées (ID) et ordonnées (OD). Les fréquences de mutations sont calculées à partir du nombre de mutations S ou NS totales observées dans chacune des 3 études, rapportés à la longueur totale en acides aminés des segments ordonnés et désordonnés qui constituent la polyprotéine. Le symbole "*" correspond à une *p-value* inférieure à 0.05.

B) Le désordre est-il associé à des changements en acides aminés plus déstabilisants/contraignants pour la structure de la protéine ?

Nous avons ensuite souhaité évaluer la nature plus ou moins déstabilisante de ces mutations non-synonymes co-localisant avec des zones de désordre. En d'autres termes, les régions désordonnées, de par la flexibilité structurale qui les caractérise, sont-elles plus susceptibles de tolérer des substitutions entraînant un changement de contexte physico-chimique que des zones plus structurées ?

En effet, les innovations évolutives et donc adaptatives passent plus probablement par l'apparition de mutations ayant des effets plutôt impactant sur la structure et la stabilité d'une protéine (Soskine & Tawfik 2010; Tokuriki & Tawfik 2009; Tokuriki et al. 2008). De précédentes observations ont permis de relier le désordre aux mutations plutôt déstabilisantes (Brown et al. 2010; Pechmann & Frydman 2014).

Nous proposons donc de tester ce lien entre désordre et mutations déstabilisantes qui apparaissent dans une population virale.

Deux classifications basées sur la nature physico-chimiques des acides aminés ont été utilisées. La première classification utilise un score de substitution basé sur la matrice BLOSUM 62 (cf. Annexes Figure A2.1). Un score a été associé à chacune des mutations non-synonymes répertoriées. Dans cette classification, un score faible est associé à des changements physico-chimiques drastiques et donc supposés plus déstabilisants en termes de structure. La seconde classification, basée sur la pénalité structurale potentiellement imposée par le changement en acides aminés, classe chaque substitution selon sa nature soit conservative (C), c'est-à-dire peu déstabilisante pour la structure, soit non-conservative (NC), plutôt déstabilisantes (Pechmann & Frydman 2014) (cf. Annexes Figure A2.2).

La répartition des scores BLOSUM62 ainsi que la nature C ou NC des substitutions, selon qu'elles sont localisées dans une région ordonnée ou désordonnée, est représentée figure II.4. A un seuil de confiance de 95%, aucune corrélation significative n'est observée entre la nature déstabilisante des substitutions et leur localisation dans les régions ordonnées ou désordonnées, tant pour les scores BLOSUM62 (figure II.4.A) que pour la classification NC *versus* C (figure II.4.B) malgré une tendance observée dans cette dernière.

Le nombre faible de mutations non-synonymes à étudier peut néanmoins constituer une limite majeure à cette analyse.

Les substitutions dans les zones désordonnées peuvent-elles être associées à une fréquence plus importante du variant correspondant (*i.e* sa fitness) dans la population ?

L'objectif est de tester la corrélation entre désordre et mutations non-synonymes (*i.e* diversité de séquences à l'échelle protéique), en partant de l'idée que les zones structuralement moins contraintes (désordonnées) tolèrent mieux les mutations.

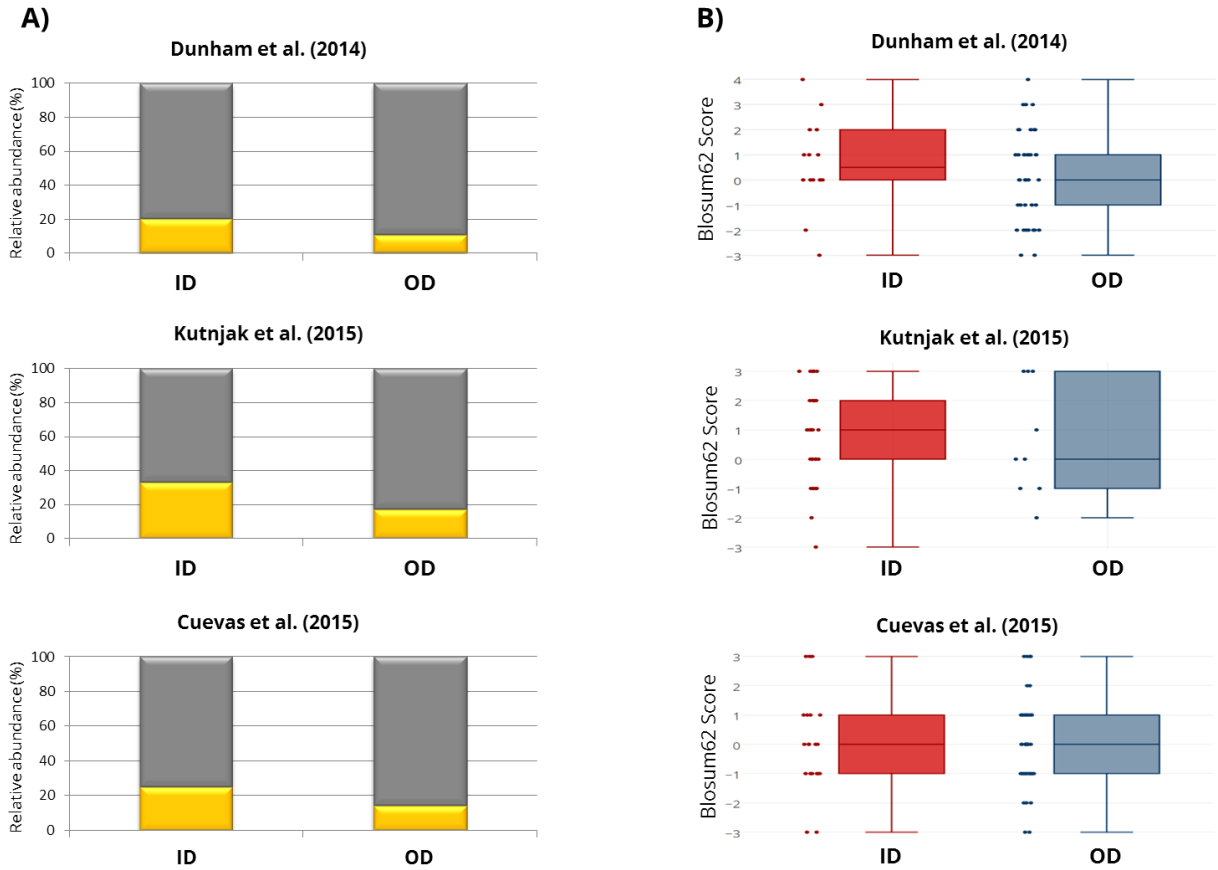


Figure II.4|Changements physico-chimiques induits par les mutations non-synonymes localisées respectivement dans les zones désordonnées (ID) et ordonnées (OD). A) Classification des mutations non-synonymes selon leur effet fortement déstabilisant (non-conservatives, en jaune) ou faiblement déstabilisant (conservatives, en gris) sur la structure de la protéine correspondante, définie dans (Pechmann & Frydman 2014). **B)** Répartition des scores BLOSUM62 associés à chaque substitution selon leur localisation dans des zones désordonnées (ID) ou ordonnées (OD).

Ces mutations dans les régions désordonnées sont donc attendues comme moins "impactantes" pour la fitness du virus que celles associées aux régions ordonnées. Partant du principe qu'une mutation est maintenue dans la population proportionnellement à l'effet qu'elle confère sur la fitness du variant, la fréquence à laquelle celui-ci est rencontrée est un indicateur de la fitness associée à la mutation.
Pour tenter de déterminer si les substitutions observées dans des zones de désordre présentent un impact moindre en terme de fitness par rapport à celles contenues dans les zones plus structurées, nous avons donc classé les différentes mutations selon la fréquence à laquelle elles sont retrouvées dans chacun des échantillons.
A un seuil de confiance de 95%, aucune corrélation robuste n'a pu être établie entre le statut ordonné ou désordonné de la région porteuse de la mutation et la fitness/fréquence de cette mutation dans la population, et ce pour les 3 jeux de données étudiés (*p-values* de 0.36, 0.82 et 0.12 obtenues respectivement pour Dunham (2014), Kutnjak (2015) et Cuevas (2015), selon le test non-paramétrique de Kruskal-Wallis).

Aucune corrélation n'est observée entre le désordre et la fréquence des variants dans un échantillon. La récurrence entre les différentes feuilles et/ou plantes a également été testée en fonction de la position de chaque mutation dans des régions ordonnées ou désordonnées. Là encore, aucune corrélation n'a pu être établie entre l'occurrence des mutations partagées par plusieurs échantillons et leur position dans des régions ordonnées ou désordonnées.

3. Discussions, Conclusions et Perspectives

L'analyse de la répartition des mutations accumulées au cours de l'infection de 3 potyvirus, le ZYMV, le PVY et le TEV sur leurs hôtes naturels et pour le TEV également sur un hôte dit « nouveau », a d'ores et déjà permis de mettre en évidence que les protéines de ces virus présentent une hétérogénéité dans leur capacité à accumuler des changements en acides aminés (mutations non-synonymes). Cependant, en ce qui concerne le rôle du désordre sur cette capacité, les résultats sont à nuancer et diffèrent entre les jeux de données. Ainsi, au cours de l'évolution du TEV sur son hôte naturel *N.tabacum*, les régions désordonnées accumulent plus de substitutions en acides aminés que les régions ordonnées. Néanmoins, cette observation n'a pas été confirmée pour les autres jeux de données. Certaines hypothèses peuvent être apportées afin d'expliquer cette disparité : **i)** Le biais ordre/désordre s'exerce au cours de l'évolution intra-hôte mais le nombre de mutations trop restreint limite l'interprétation statistique de la tendance observée ; **ii)** L'effet du désordre sur la robustesse mutationnelle et la création d'un réseau de variants neutres n'est visible ou ne s'exerce qu'à plus large échelle évolutive ; **iii)** En étant relié au mode évolutif du virus dans un environnement donné, l'effet du désordre sur la robustesse mutationnelle est pathosystème dépendant.

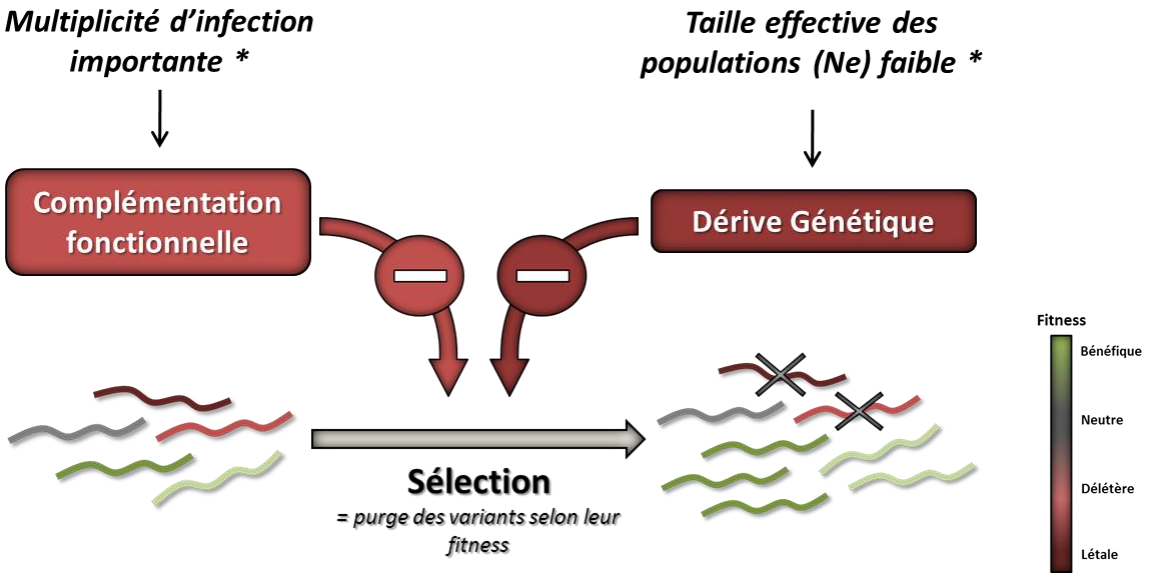
Vérifier ces hypothèses passe donc par la détermination des facteurs pouvant biaiser nos observations, au vu des différentes méthodes et matériels qui ont conduit à l'obtention de chacun de ces jeux de données.
En effet, les différences entre les modes opératoires (récapitulés dans le tableau II.1) notamment en ce qui concerne le pathosystème utilisé, le nombre de plantes ayant servi pour l'analyse, l'utilisation ou non de passages séquentiels au cours de l'évolution expérimentale, la profondeur de séquençage et l'utilisation de réplicats biologiques et techniques sont autant de paramètres qu'il convient d'intégrer lors de l'interprétation de nos résultats.

i) Le nombre trop restreint de mutations limite l'observation de tendances robustes

Oui et non. D'une part, on observe une fréquence de variations non-synonymes dans les zones de désordre systématiquement supérieure à celle retrouvée dans les zones ordonnées (Figure II). On peut supposer qu'un jeu de données plus large aurait pu permettre de confirmer cette tendance et la rendre statistiquement significative.
Dans l'analyse de celles de Cuevas *et al*., l'accumulation significative des mutations NS dans les régions désordonnées est observée à partir de 43 mutations NS, lors de l'accumulation du TEV sur hôte NT. Or, le nombre de mutations NS totales observées à partir des données de Kutnjak et Dunham, pourtant du même ordre de grandeur (respectivement 49 et 39), ne permettent pas de conclure à une tendance significative. Le nombre de mutations ne permet donc pas à lui seul d'expliquer la non-significativité des tendances observées. D'autres facteurs sont donc vraisemblablement impliqués.

ii.a) Effets des paramètres associés au séquençage

Comme résumé dans le tableau II.1, les techniques de séquençage utilisées couvrent la quasi-intégralité du génome avec une profondeur moyenne de séquençage sensiblement identique pour chacune des trois expériences. Nous pouvons cependant noter l'usage différent de réplicats techniques. En effet, la limite majeure du séquençage profond pour identifier des évènements mutationnels rares réside dans la difficulté à discriminer les variations nucléotidiques présentes dans la population de départ, de celles introduites au cours des différentes étapes techniques de RT, PCR et réactions de séquençage. L'utilisation de réplicats techniques vise donc à limiter ce biais en répétant en parallèle les différentes étapes techniques, à partir d'un même échantillon de départ. Les mutations retrouvées dans un seul des réplicats seront donc considérées comme un artefact technique et pourront être écartées. Cependant, au vu des résultats basés sur l'étude conduite par Kutnjak *et al*. qui inclut l'utilisation de réplicats, ce facteur n'explique pas les différences observées quant au rôle du désordre sur l'apparition des variants en intra-hôte.



** Facteurs hôte, temps et organe dépendant*

Figure II.5|Schéma récapitulatif des forces qui s'exercent, ainsi que de leurs déterminants, sur une population virale au cours de l'évolution intra-hôte.

ii.b) Les différences observées sont dûes aux variations dans les modes opératoires, qui ne se situent pas à la même échelle évolutive

Le postulat qui a été fait jusqu'à présent consiste à admettre que l'existence et la fréquence d'un variant est directement corrélé à sa fitness intra-hôte, sous l'action des forces de sélection. Or le nombre limité de plantes (une seule pour l'analyse conduite dans Dunham *et al.*) ou l'absence de répétitions indépendantes (1 pool de 60 plantes dans l'analyse de Kutnjak *et al*.) peuvent entrainer un biais dans ce postulat. En effet, à l'échelle de l'évolution intra-hôte, la dérive génétique et la complémentation fonctionnelle peuvent également contribuer activement au façonnage de la population de variants et ce, de façon antagoniste à la sélection naturelle (Figure II.5). Ces forces et leur mise en évidence chez certains virus de plantes sont présentées dans les deux paragraphes suivants.

La dérive génétique décrit le processus de variations évolutives impliquant la « sélection aléatoire » (*random sampling*) de gènes de la génération parentale pour conduire à la génération suivante. Ce processus stochastique s'établit dans des populations de faible nombre, comme par exemple après le passage d'une étape biologique (voir plus loin) constituant un important goulet d'étranglement, pour lesquelles le nombre de variants ne permet pas que chacun soient représentés de façon égale à la génération suivante. Les variations génétiques vont alors pouvoir se transmettre de façon totalement aléatoire, et ce indépendamment de leur valeur sélective (fitness).

Ainsi, quantifier la force qu'exerce la dérive sur une population donnée permet de relier ou non la fréquence obtenue pour chaque variant à sa fitness. Ceci est indispensable à des études comme la nôtre, portant sur les déterminants de la robustesse mutationnelle, qui sous-entend l'analyse du coût de fitness induit par chaque mutation. La force de la dérive qui agit dans l'évolution d'une population finie peut être appréhendée via l'estimation de la taille de population efficace (Charlesworth 2009). Plus précise et rigoureuse que la simple estimation de la taille totale de la population, la taille efficace de population (*Ne*) correspond au nombre d'individus qui contribuent à l'établissement de la génération suivante. La *Ne* est souvent reportée comme beaucoup plus faible que la taille totale de la population chez les virus à ARN (García-Arenal et al. 1984; Frankham 1995). En effet, une large fraction de la population consiste en des mutants "non fit", incapables de se multiplier. En conséquence, les conclusions faites sur l'efficacité de la sélection peuvent être biaisées lorsqu'elles sont uniquement basées sur la taille totale de la population. En effet, plus la *Ne* est faible, plus la dérive génétique va avoir d'impact sur la structure génétique de la population suivante. Elle va ainsi conduire au maintien et à l'expansion de variants délétères, pouvant même amener à la situation extrême dite de Muller's Ratchet, dans laquelle un mutant "non fit" domine la population virale.

Pour estimer la *Ne* et donc par association, la force de la dérive, une solution peut consister à analyser l'évolution de la structure d'une population (par séquençage) parallèlement sur deux plantes hôtes (réplicats biologiques), inoculées avec le même virus et se développant dans les mêmes conditions. La pression de sélection est donc théoriquement identique dans les deux plantes. La différence de population observée après plusieurs temps de génération révélera directement la force exercée par la dérive génétique dans ces populations. Ainsi, si on retrouve de nombreux variants identiques entre les deux expériences, la sélection constitue alors la force majoritaire dans le façonnage de la population de variants, et ces mêmes variants partagés entre les deux expériences peuvent être considérés comme "fit" (bénéfiques). A l'inverse, une différence importante entre les paysages de variants au sein des populations des deux expériences traduira un effet non négligeable des processus stochastiques tels que la dérive génétique, et il sera alors risqué de tenter de relier la fitness des variants à leurs fréquences respectives dans la population.

Chez les virus de plantes, de nombreuses études ont permis d'estimer la taille des goulets d'étranglement présents notamment au niveau des étapes de transmission horizontal par vecteur (Moury et al. 2007; Ali et al. 2006; Betancourt et al. 2008; Simmons et al. 2012) ainsi qu'au niveau de étapes de colonisation dans la plante hôte (Hall et al. 2001; French & Stenger 2003; Fraile & Sacristán 2008; Sacrista et al. 2003; Dunham et al. 2014; Tromas et al. 2014). Ces goulets réduisent drastiquement et brutalement la taille effective de la population virale, avec un nombre de particules transmises parfois estimé comme inférieur à 20.

A l'opposé, il a été observé dans d'autres études que les fluctuations stochastiques intervenaient finalement assez peu dans les structures génétiques des populations virales (Monsion et al. 2008; Elena et al. 2011). A titre indicatif, la *Ne* au niveau des goulets d'étranglement constitués par le mouvement systémique des particules vers les feuilles apicales serait plutôt de l'ordre d'une centaine de particules (Fabre et al. 2014; Gutiérrez et al. 2012). Dans ce cas, la *Ne* permet une meilleure efficacité du processus déterministe de sélection naturelle, en limitant l'effet antagoniste de la dérive génétique. Ces estimations très variables de la taille des goulets d'étranglement qui se produisent au cours du mouvement des populations virales dans un même hôte semblent refléter l'hétérogénéité d'un tel processus d'un pathosystème à l'autre. Enfin, des biais comme la dose initiale de l'inoculum ainsi que le mode de transmission ont été identifiés comme pouvant affecter ces processus (Zwart et al. 2011; Gutiérrez et al. 2012).

Pour résumer, les résultats contradictoires fournis par la littérature en ce qui concerne la taille effective de population virale déterminée lors de l'infection d'une plante hôte ne nous permettent pas de dresser un postulat net sur la force de la dérive génétique à l'échelle de l'hôte, dans le cas des virus à ARN de plantes. Ainsi, tenter de relier de façon fine la fitness d'un variant à sa fréquence dans la population nécessiterait d'abord d'estimer la force des processus stochastiques via l'estimation du *Ne* global à l'échelle de l'hôte, ce qui n'a pas été réalisé dans les différentes études que nous avons analysées.

Dans ce contexte, le nombre de plantes utilisées constitue un critère pouvant avoir un impact non négligeable sur les résultats et donc l'interprétation d'expériences d'évolution expérimentale chez les plantes. En effet, un nombre trop restreint de plantes limite l'estimation de la force de la dérive, ce qui peut donc gêner à terme l'interprétation des résultats.

La complémentation fonctionnelle est un processus de robustesse mutationnelle à l'échelle de la population observé chez les quasi-espèces. Elle correspond à la capacité de coopération fonctionnelle entre quasi-espèces, lorsqu'elles infectent la même cellule, et peut en quelque sorte être interprétée comme le niveau de ploïdie d'un génome viral à un instant donné. Ce processus, au même titre que la dérive, permet à certains variants "non fit" de se maintenir à des fréquences pouvant être élevées (cas extrême du virus de la Dengue dans lequel le variant majoritaire est délétère (Aaskov et al. 2006; Ke et al. 2013)) et peut ainsi affecter la structure génétique d'une population. Chez les virus de plantes, le maintien de mutants défectueux (porteurs de mutations très délétères voire létales) a également pu être observé (Garcia-Arenal et al. 2001; Malpica et al. 2002). Ce processus est directement lié au nombre de génomes viraux infectant une même cellule, défini par la « Multiplicité d'Infection » (MOI). Une étude a permis d'estimer la MOI au cours de l'infection d'un virus à ARN de plante, le TMV, sur l'hôte *Nicotiana benthamiana* (González-Jara et al. 2009). Il a ainsi pu être observé que la MOI décroit au cours de l'infection, passant de 6 aux étapes précoces à 1 ou 2. Contrairement à la dérive, dont l'effet peut être estimé en comparant différents réplicats biologiques, l'impact de la complémentation ne peut pas être mesuré juste en analysant la diversité génétique des populations et nécessite des mesures de fitness individuelles. Il parait néanmoins important de garder à l'esprit que de tels évènements peuvent survenir lors de l'analyse de l'évolution virale intra-hôte ainsi que les éventuels biais qu'ils peuvent induire sur notre interprétation de ces processus évolutifs.

La dérive et la complémentation peuvent donc, à l'échelle intra-hôte du moins, biaiser de façon importante le postulat reliant la fréquence d'un variant à sa fitness. Chez l'expérience de suivi de l'évolution du ZYMV, l'occurrence d'un variant supposé totalement létal (décalage du cadre de lecture au niveau de la séquence codante de P1 soit en N-ter de la polyprotéine) dans plusieurs échantillons prélevés à des temps successifs, semble bien illustrer l'action de la dérive ou de la complémentation dans le maintien des variants dans la population malgré leur fitness faible ou nulle.

iii) Les contraintes exercées par l'hôte limitent l'analyse en limitant la diversité

De nombreuses différences fondamentales sont également à noter en ce qui concerne le matériel biologique utilisé ainsi que le mode opératoire utilisé dans les 3 études. En effet, il est par exemple envisageable qu'un effet pathosystème puisse expliquer les différences de résultats obtenus, mais toutes choses n'étant pas égales par ailleurs, on ne peut pas discuter quant à l'impact de la variable « virus » sur ces résultats.

En ce qui concerne la variable « hôte », les contraintes exercées par les hôtes ont déjà été démontrées comme spécifiques à chaque système et directement impliquées dans le déterminisme de la purge sélective et donc la composition des populations virales (Moury et al. 2002; Lalic et al. 2011). Les résultats obtenus par Cuevas *et al* illustrent très bien ce propos. En effet, à modes opératoires équivalents, avec pour seule différence l'hôte infecté, la composition des populations virales au cours de l'infection se trouve être totalement différente. En effet, alors que de nombreuses mutations s'accumulent au cours des différents passages dans l'hôte naturel, peu sont recensées lors de l'évolution sur le nouvel hôte. Comme discuté dans leur étude, des pressions de sélection plus fortes sont probablement responsables de cette évolution limitée dans le nouvel hôte (expliquant notamment pourquoi le TEV s'accumule également moins) et conduisent donc à une purge plus stricte des variants par rapport à leur fitness, en limitant donc leur nombre dans la population. L'évolution dans le nouvel hôte semble donc se faire sur un mode « innovant » (Barrick & Lenski 2013), favorisant l'émergence de formes très "fit", aux dépends d'une accumulation de variants neutres ou peu délétères. A l'inverse, l'hôte naturel constitue un environnement plus favorable pour le virus. L'évolution dans ce dernier est donc soumise à une force plus modérée de la sélection (sélection diversificatrice, sur un mode d'optimisation), et considérée donc comme plus permissive à l'accumulation de variants neutres et peu délétères, ce qui mène donc à terme à une diversité globale plus importante dans la population.
Dans le cas très précis de l'étude conduite par Cuevas *et al*, le nombre réduit de variant dans CA pourrait expliquer l'absence de déterminisme structural observé dans la diversité génétique virale pour cet hôte.
Dans le même ordre d'idée, pour l'hôte NT, la tendance significative observée est associée à un plus grand nombre de mutants dans la population considérée. Ceci nous permet de suggérer que le **désordre constituerait un déterminant de la robustesse mutationnelle et ce, dès l'échelle de l'évolution intra-hôte.**

Perspectives à envisager pour des analyses complémentaires

L'observation dans d'autres pathosystèmes du rôle du désordre dans l'accumulation de mutations non-synonymes nécessite de pouvoir estimer les facteurs tels que la MOI et la taille effective de la population, pour appréhender les biais pouvant limiter les effets de la sélection. Dans ce contexte, l'analyse de plusieurs plantes, constituant des réplicats biologiques, apparait comme judicieux.
Une approche différente pour tenter d'établir de façon plus directe un lien entre le désordre et la robustesse mutationnelle des variants lors d'une infection, consiste à insérer par mutagénèse dirigée des mutations à plusieurs endroits dans le génome viral et de mesurer directement l'impact de ces mutations sur la fitness du variant ainsi créé.

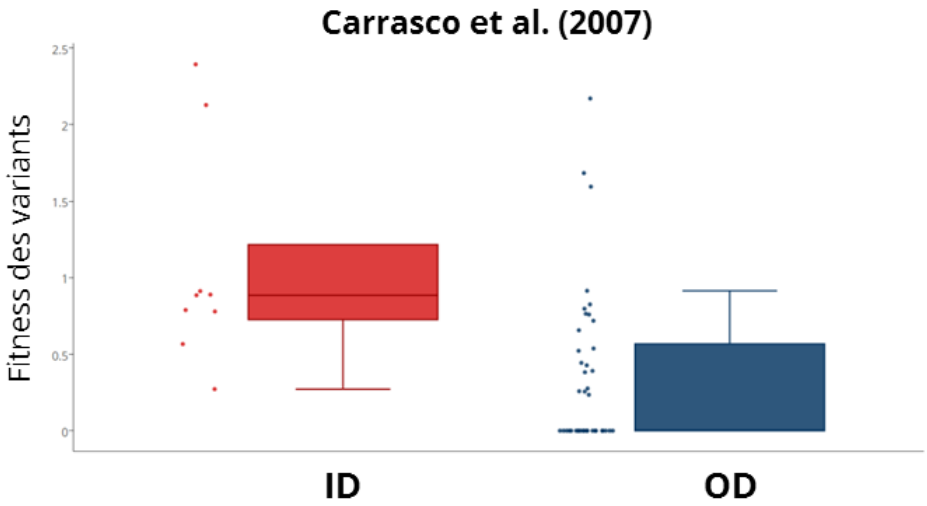


***Figure II.6*|Effets sur la fitness virale de mutations non-synonymes introduites dans les régions désordonnées et ordonnées du *Tobacco etch virus*.** Les points représentent chacune des valeurs de fitness relatives (*W*) déterminées pour chaque mutation. Les valeurs de *W* sont obtenues lors de test de compétition avec la souche sauvage PC1 du TEV, via leurs co-inoculation sur l'hôte *Nicotiana tabacum* (Carrasco et al. 2007).

Lien entre désordre et fitness de mutations obtenues par mutagénèse aléatoire

Une telle approche a été conduite par (Carrasco et al. 2007). Des mutations ont été introduites sur le génome du TEV afin d'analyser l'impact de chacune d'elles sur la fitness du virus, au cours de l'infection sur l'hôte *Nicotiana tabacum*.

Pour déterminer si le désordre est associée à des mutations de fitness globalement plus importante, nous avons classé chacune des mutations et leurs valeurs respectives de fitness relatives selon leur position dans les régions ordonnées ou désordonnées (figure II.6).

Bien qu'une tendance semble se dessiner, le nombre trop réduit de mutations étudiées ne permet pas de conclure quant à un lien significatif entre désordre et robustesse mutationnelle (p-value > 0.05 ; Test de Kruskal-Wallis). Cette démarche constitue néanmoins une base intéressante pour des études futures, qui nécessiteront alors d'augmenter le nombre de mutations à étudier.

Chapitre III

Rôle du désordre de la région centrale de la protéine VPg dans l'adaptation du PVY à la résistance *pvr2*^{3} médiée par eIF4E chez l'hôte *Capsicum annuum*

1. Contexte de l'étude

A) L'adaptation des virus de plantes aux résistances récessives variétales.

Comme présenté précédemment, les virus sont des parasites obligatoires et l'accomplissement de leur cycle infectieux passe donc nécessairement par le recrutement des facteurs cellulaires de l'hôte. Ainsi des formes de résistance dite « passives » résident dans l'incompatibilité de ces facteurs. Cette incompatibilité peut être due à la distance évolutive de l'hôte résistant par rapport à l'hôte naturel (résistance non-hôte) ou résulter de modifications survenues au cours de l'évolution de l'hôte naturel (résistances récessives), dans le contexte de la course à l'armement existant entre l'hôte et son pathogène. Dans ces deux cas de figure, le virus, n'ayant pas ou plus la possibilité de recruter ces composants essentiels à une ou plusieurs étapes de son cycle, ne pourra donc pas s'accumuler dans l'hôte correspondant.

Dans le domaine de la pathologie végétale, l'identification de ces résistances récessives et les déterminants de leur durabilité est d'autant plus importante que l'utilisation de telles résistances au champ constitue une des stratégies les plus efficaces et les plus durables pour lutter contre les phytovirus et les pertes économiques très importantes que représentent les maladies virales dans les domaines de l'agriculture (Nicaise 2014; Wang & Krishnaswamy 2012; Truniger & Aranda 2009).
L'utilisation de variétés porteuses de ces résistances a pour conséquence d'imposer une importante pression de sélection sur les populations virales, pouvant conduire de façon plus ou moins rapide (notion de durabilité de la résistance) à l'émergence d'isolats dits contournants, capables de s'adapter à ces résistances génétiques (cf. Introduction Figure 14B). L'identification des déterminants de cette durabilité et notamment des mécanismes mis en jeux au cours du contournement des résistances récessives constituent donc un enjeu majeur dans la lutte contre les virus de plantes. Chez les virus de plantes (constitués en grande partie de virus à ARN), de nombreux facteurs viraux impliqués dans la **virulence**[1] vis-à-vis des gènes de résistance (appelés facteurs de virulence/avirulence) ont ainsi été identifiés. Dans la grande majorité des cas, la restauration de la virulence du virus dans l'hôte résistant se joue au niveau protéique, et se traduit par l'apparition d'un nombre très limité de substitutions en acides aminés au niveau de ces facteurs de virulence (Moury et al. 2010). Le nombre de mutations nécessaires au contournement par le virus constitue un des déterminants de la durabilité d'une résistance (on comprend bien que plus le nombre de mutations nécessaires au contournement est important, plus la probabilité que le facteur de virulence parvienne à toutes les accumuler est faible) (Harrison 2002; Fabre et al. 2009).

[1] *Virulence = composante qualitative binaire du pouvoir pathogène indiquant la capacité ou non d'un agent pathogène à induire une infection, pouvant être mesuré par l'accumulation de la charge virale et/ou l'apparition de symptômes, dans un génotype d'hôte particulier.*

Cependant il arrive qu'une résistance qualifiée de durable soit pourtant associée à un nombre de mutations limité. Ce sera le cas d'une mutation déstabilisante pour la protéine et donc difficile à maintenir dans la population.
Le nombre de mutations indépendantes capables de conduire à une adaptation (ou encore le nombre de solutions adaptatives différentes) constitue également un déterminant de la durabilité d'une résistance et est inversement corrélé à celle-ci.
Ce déterminisme moléculaire peut directement être relié aux bases moléculaires de l'interaction, perturbée dans le contexte d'une résistance récessive, et donc aux propriétés structurales de chaque facteur impliqué dans cette interaction.

En assurant un environnement topologique plus favorable à l'émergence de mutations que les régions structurées (voir Introduction section 2.D), le désordre intrinsèque retrouvé chez certains facteurs de virulence pourrait faciliter l'émergence et le nombre de solutions adaptatives. En conséquence de quoi **le désordre constituerait un des déterminants de l'adaptation virale aux résistances variétales et influencerait la durabilité de celles-ci**.

B) Cas du contournement, via la protéine VPg, des résistances médiées par eIF4E chez les Potyvirus

Le facteur d'initiation de la traduction eIF4E, un déterminant majeur des résistances récessives aux virus de plantes

Parmi les cas de résistances récessives naturelles recensées chez diverses espèces de plantes ces dix dernières années, une très grande majorité (12 sur les 14 recensées en 2012) concerne celles médiées par le facteur eucaryote d'initiation de la traduction 4E (eIF4E) ou de son isoforme eIF(iso)4E (Wang & Krishnaswamy 2012; Lecampion et al. 2009).

Le recrutement d'eIF4E par la protéine VPg : déterminant majeur dans l'infection par les Potyvirus.

Bien que de récentes évidences montrent qu'elle ne constitue pas nécessairement l'unique déterminant de cette résistance (Tavert-Roudet et al. 2012; Nakahara et al. 2010; Ala-Poikela et al. 2011; Roudet-Tavert et al. 2007; Abdul-Razzak et al. 2009; Estevan et al. 2014), l'interaction entre eIF4E et la protéine virale VPg est directement corrélée à la sensibilité de la plante pour le virus, chez de nombreux pathosystèmes. D'abord décrit chez le TuMV (Wittmann et al. 1997), l'interaction entre la VPg et les facteurs eIF4E a en effet été directement relié à la capacité d'infection virale chez plusieurs virus végétaux, parmi lesquels de nombreux potyvirus (Léonard et al. 2000; Nicaise et al. 2003; Díaz et al. 2004; Stein et al. 2005; Ruffel et al. 2002; Beauchemin et al. 2007; Charron et al. 2008; Moury et al. 2004), ainsi que chez certains virus animaux membres de la famille des *Caliciviridae* (Goodfellow et al. 2005; Chung et al. 2014; Hosmillo et al. 2014).

L'interaction VPg/eIF4E au cœur de nombreuses fonctions dans le cycle viral

Etant donné l'importance d'une telle source de résistance, les conséquences fonctionnelles de cette interaction sur le cycle viral ont été largement étudiées ces 10 dernières années.

Rôle de VPg/eIF4E dans la traduction des ARN viraux :

La VPg pouvant être considéré comme un analogue de la coiffe eucaryote de par sa liaison covalente à l'extrémité 5' des ARN viraux, il a d'abord été envisagé que l'interaction VPg/eIF4E permette le recrutement de la machinerie de traduction eucaryote pour la traduction des génomes viraux. En effet, des études *in vitro* ont révélé l'existence d'une compétition entre la protéine VPg et la coiffe des ARN messagers eucaryotes (Michon et al. 2006; Khan et al. 2008; Miyoshi et al. 2008), qui aurait pour conséquence de réduire le taux de traduction des ARNm en présence de la VPg. Cette hypothèse a été remise en question suite à la découverte des sites internes d'entrée des ribosomes (Internal Ribosome Entry Site, IRES) en 5' des génomes de potyvirus (Khan et al. 2006; Robaglia & Caranta 2006). Les données montrent en effet que la VPg participerait à l'initiation de la traduction en augmentant l'affinité d'eIF4E pour l'IRES situé à l'extrémité 5' terminal de l'ARN génomique du TEV. Aucune évidence à ce jour n'a pu démontrer la présence en 5' de la VPg chez les ARN en cours de traduction. De plus, chez le poliovirus (famille des *Picornaviridae*) la VPg n'est pas attachée au génome viral au cours de la traduction (Nomotot et al. 1977).

Pour autant, cette interaction peut tout de même avoir une implication dans le processus d'initiation de la traduction, le taux de traduction IRES-dépendante ayant été montré comme augmenté en présence de la protéine VPg (Khan et al. 2008).

Pour résumer, la VPg servirait plus vraisemblablement de stabilisateur des ARN viraux, prévenant la dégradation de ces structures instables et favorisant ainsi la traduction via les sites IRES (Eskelin et al. 2011), tout en inhibant la traduction des ARN messagers de l'hôte, via son interaction avec eIF4E (Khan et al. 2008).

Autres fonctions associées à l'interaction VPg/eIF4E :

- ***Rôle dans la réplication***

D'autres fonctions pourraient être associées à cette interaction VPg/eIF4E. Le recrutement d'eIF4E par la VPg lié en 5' au cours des étapes précoces de la réplication virale pourrait ainsi permettre la circularisation du génome viral via l'interaction entre eIF4E et la poly-A Binding protein (PABP) à l'instar de ce qui a déjà été décrits chez les poliovirus (Herold et al. 2001). Compte-tenu du rôle d'amorçage de la VPg pour la réplication (cf. Chapitre I), cette circularisation en permettant le rapprochement entre l'extrémité 3' du génome viral et la VPg pourrait faciliter l'initiation de la synthèse du nouveau brin par la réplicase virale RdRP.

- ***Rôle dans le mouvement***

Le complexe eIF4F, dont fait partie eIF4E, interagit avec les microtubules (Bokros et al. 1995) et intervient dans l'export du noyau de certains ARNm (Culjkovic et al. 2007). L'interaction d'eIF4E avec la VPg pourrait donc servir à l'adressage des particules virales aux plasmodesmes. Ces structures, situées au niveau des jonctions cellulaires végétales, permettent le passage des génomes viraux de cellule à cellule. Cette hypothèse est également suggérée par la corrélation qui existe entre la surexpression du facteur eIF4E et l'accroissement du mouvement du *Pea seed-borne mosaic virus* (PSbMV) (Gao et al. 2004).

Mécanismes moléculaires de l'interaction VPg/eIF4E

Une étude conduite au laboratoire a permis de montrer que la région centrale de la VPg interagit fortement avec eIF4E (Roudet-Tavert et al. 2007). De plus, des tests d'interaction en double hybride entre eIF4E et VPg montrent que des changements d'acides aminés dans cette région centrale sont corrélés à une modulation de la force d'interaction (Walter *et al*., Manuscrit en préparation). En utilisant des peptides synthétiques mimant le domaine central de la VPg, il a également pu être montré que l'association s'établit suivant un modèle en deux étapes : une étape rapide de reconnaissance non spécifique, suivie d'une étape de repliement du peptide en hélice α (Chroboczek et al. 2012).

Chez les potyvirus, le contournement des résistances médiées par eIF4E est lié à l'apparition de substitutions dans la protéine VPg.

Les évènements moléculaires conduisant à l'émergence de formes adaptées (contournants) de la VPg compatibles avec eIF4E, ont été largement étudiés chez les Potyvirus et particulièrement chez le pathosystème PVY/piment. Chez ce pathosystème, les substitutions en acides aminés dans la **région centrale de la VPg comprise entre les résidus 101 et 123** sont associées aux évènements de contournement, et ce pour plusieurs formes alléliques de l'allèle pvr2, codant pour le facteur eIF4E (Ruffel et al. 2002; Moury et al. 2004; Ayme et al. 2006; Ayme et al. 2007). Ainsi dans le cas du contournement par l'isolat SON41p de la résistance médiée par l'allèle $pvr2^3$, 5 événements de mutations indépendants sont expérimentalement démontrés comme impliqués dans le contournement conduisant à l'infection d'environ 40% des individus, chez l'hôte HD285, porteur homozygote de l'allèle de résistance $pvr2^3$ (Figure III.1) (Ayme et al. 2006).

Cas du pathosystème PVY/piment : une course à l'armement à deux composants

En considérant les séquences d'eIF4E codées par les différentes formes alléliques de pvr2, ainsi que les différents déterminants du contournement identifiés sur la VPg, une stricte corrélation a pu être établie entre l'interaction eIF4E-VPg observée en double hybride et la virulence des virus sur les plantes inoculées (Charron et al. 2008).

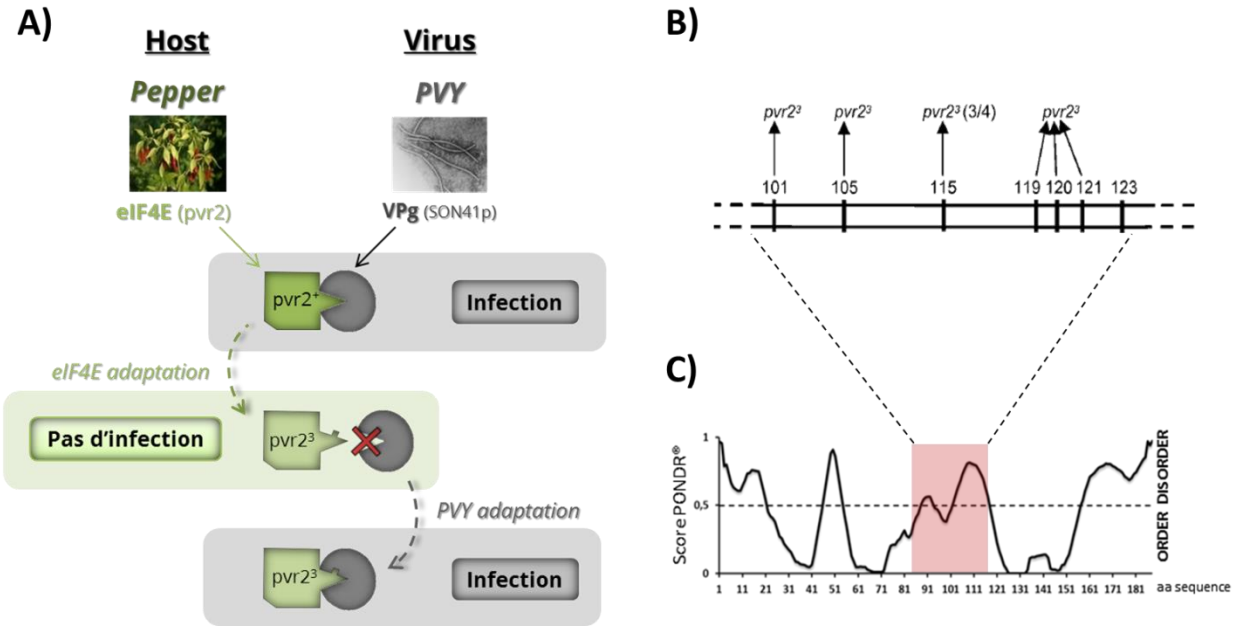


Figure III.1|Contournement par modification de la VPg du PVY-SON41p, de la résistance pvr2^3 du piment HD285. A) Illustration des étapes de la course à l'armement existant entre le piment et le PVY, jusqu'à la mise en place d'un contournement. **B)** Les six mutations dans la région centrale de la VPg de l'isolat SON41p du PVY identifiées comme responsables du contournement de la résistance portée par le gène pvr2^3 (d'après (Ayme et al. 2007). **C)** Prédiction du désordre de la VPg-SON41p. Le carré rose indique la zone centrale dans laquelle sont contenues les mutations de contournement.

Ce pathosystème semble donc suivre le modèle « simple » à deux composants, de la course à l'armement qui existe entre un hôte et son pathogène (Moury et al. 2014), ce qui facilite dans ce cas particulier l'analyse des déterminants associés au processus d'adaptation (Figure III.1).

Les déterminants du contournement sont situés dans la région centrale désordonnée de la VPg

Fait intéressant, le domaine central porteur des déterminants du contournement, a été expérimentalement montré comme intrinsèquement désordonné chez la VPg du PVY (Grzela et al. 2008; Chroboczek et al. 2012) ainsi que chez plusieurs autres VPg de potyvirus et sobemovirus (Rantalainen et al. 2008; Hébrard et al. 2009). Ces données sont en accord avec les régions prédites comme désordonnées dans la VPg de l'isolat SON41p (Figure III.1C).

Pour résumer, l'identification du facteur d'avirulence (VPg), du gène de résistance (eIF4E) ainsi que des mutations contrôlant la spécificité de l'interaction entre virus et hôte, constituent un jeu de données solide faisant du pathosystème PVY/piment un modèle idéal pour déterminer expérimentalement la contribution du désordre dans le processus adaptatif de contournement d'une résistance récessive.

En pratique, l'implication du désordre de la VPg dans l'adaptation sera testée en analysant le contournement de la résistance pvr2^3 du piment par des variants de VPg de l'isolat SON41p de PVY.

2. Résultats

L'objectif de cette partie de mon travail de thèse était donc de déterminer si le désordre décrit dans la zone centrale de la VPg du PVY, et qui semble constituer une caractéristique commune à l'ensemble des VPg des Potyvirus (cf. Chapitre 1) (Hébrard et al. 2009), est contributif de l'adaptabilité du PVY, et plus précisément de sa capacité à contourner la résistance portée par l'allèle pvr2^3, au cours de l'infection de la lignée HD285 de *Capsicum annuum* (Figure III.1).

A) Construction de mutants de PVY présentant des modulations dans le contenu en désordre de la zone centrale de leur VPg

L'approche choisie pour déterminer la contribution du désordre à l'adaptabilité du PVY a été de modifier artificiellement ce désordre dans le facteur de virulence VPg, directement impliqué dans l'adaptation du PVY. Plusieurs mutants de la VPg altérés dans leur contenu en désordre sont ainsi conçus. Afin de limiter la potentialité délétère des mutations introduites, ces mutations ont été choisies en se basant sur la diversité naturelle qui existe au sein du PVY. Un alignement de 44 VPg du groupe C1 du PVY (groupe capable d'infecter le piment) a été réalisé afin d'identifier les variabilités contenues dans la zone centrale.

VPg groupe C polymorphic sites		PVY SON41p																
		93	94	95	96	98	100	101	102	105	108	114	115	117	118	119	121	123
	aa seq	I	**R**	**D**	I	E	**F**	S	**E**	R	V	**E**	T	A	L	D	H	S
93	V	-																
94	**G**		**D**															
95	**E,N**			**D**														
96	V				-													
98	D					-												
100	**L**						**D**											
101	G							**RB**										
102	**Q**								**O**									
105	K									**RB***								
108	I										-							
114	**Q**											**O**						
115	A,M,R,T												**RB**					
117	V													-				
118	V														-			
119	S,E,H															**RB**		
121	N																**RB***	
123	K,E,N,T																	**RB***

Tableau III.1|Variabilité de séquences de la région centrale de 44 VPg du PVY[C]. Les substitutions entrainant respectivement une augmentation de désordre (rouge) et d'ordre (bleu) par rapport à la séquence de référence SON41p sont retenues pour construire les mutants VPg contenant plus ou moins de désordre. *RB : substitutions correspondant à des évènements de contournement de la résistance pvr2[3] par SON41p* (Ayme et al. 2006). *RB* : sites pour lesquels aucun évènement de contournement n'a pu être observé, mais dont la pression de sélection positive (dN/dS>1) suggère un potentiel rôle dans le contournement* (Benoît Moury et al. 2014).

A)

```
SON41p    61 INMYGFDPTEYSFIQFVDPLTGAQIEENVYADIRDIQERFSEVRRKMVEDDEIETQALDS 120
OD2       61 INMYGFDPTEYSFIQFVDPLTGAQIEENVYADIRDIQERFSQVRRKMVEDDEIQTQALDS 120
ID1       61 INMYGFDPTEYSFIQFVDPLTGAQIEENVYADIGEIQERLSEVRRKMVEDDEIETQALDS 120
ID2       61 INMYGFDPTEYSFIQFVDPLTGAQIEENVYADIGNIQERFSEVRRKMVEDDEIETQALDS 120
             ******************************** :****:*:***********:******
```

B)

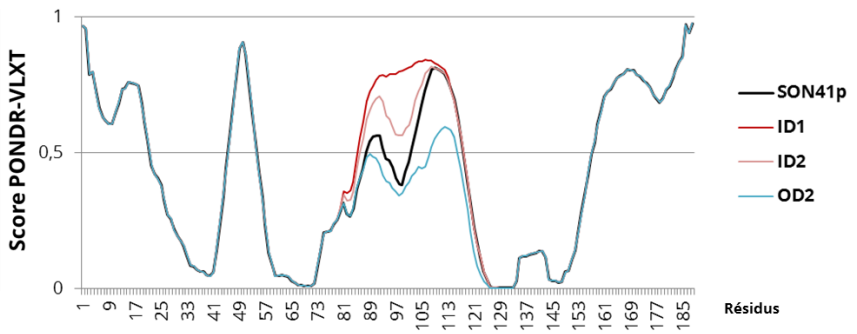


Figure III.2|Présentation des mutants PVY ID1, ID2 et OD2. A) Séquences en acide aminés de la région centrale de la VPg (résidus 61 à 120) correspondant à l'isolat SON41p du PVY ainsi qu'aux 3 mutants. Les sites qui varient entre ces 4 protéines sont indiqués en gras. Les résidus colorés correspondent aux mutations promotrices de désordre (rouges) et d'ordre (bleu). B) Scores de désordre obtenus via PONDR-VLXT pour les 3 mutants ainsi que pour l'isolat naturel SON41p. Un score supérieur à 0,5 correspond à un résidu se situant dans une région prédite comme désordonnée.

L'ensemble des variations en acides aminés observées est reporté dans le tableau III.1. L'impact de chacune de ces substitutions sur le désordre de la VPg-SON41p est ensuite estimé via la prédiction de la séquence chimère correspondante sous PONDR-VLXT. Les positions identifiées comme impliquées dans le contournement sont ignorées afin de ne pas biaiser les résultats de l'expérience de contournement (Tableau III.1).

3 mutants sont ainsi obtenus : les mutants ID1 et ID2 résultent respectivement d'une triple mutation (R94G ; D95E ; F100L) et d'une double mutation (R64G ; D95N) par rapport à la séquence de référence correspondant à l'isolat SON41p. Ces deux mutants présentent un contenu en désordre prédit comme augmenté (environ 12% de résidus désordonné en plus, avec des scores de probabilité PONDR qui diffèrent). A l'inverse, OD2 obtenu à partir d'une double substitution (E102Q ; E114Q) est prédit comme ayant 17% de résidus désordonnés en moins par rapport à la VPg entière de l'isolat SON41p, ainsi qu'une probabilité de désordre moindre. Les séquences de ces mutants et leurs profils de désordre sont présentés Figure III.2.

B) Validation expérimentale du changement de désordre introduit dans les 3 mutants de VPg

Afin de vérifier expérimentalement que les changements introduits pour chacun des mutants impactent bel et bien la proportion en désordre de la région centrale de la VPg, des expériences de dichroïsme circulaire (CD) sont réalisées sur les protéines correspondant aux 3 mutants. Pour cela, les séquences codant pour chacun des mutants ainsi que la forme de référence SON41p sont introduites dans des vecteurs d'expression d'*Escherichia coli* (cf. Section Matériel et Méthodes). Les différentes VPg sont produites puis extraites et purifiées, et des mesures de dichroïsme circulaire sont réalisées pour chacune d'elles. Le CD est une technique biophysique couramment utilisée pour inférer la contribution de chaque structure secondaire (hélice α, feuillet ou brin β et peptide non structuré) à la structure totale de la protéine (Kelly et al. 2005). Le signal de « désordre » se traduit par un minimum à environ 205nm associé à une inflexion du signal vers 220nm. Le rapport entre les valeurs du signal dichroïque à 205 et 220nm permet de comparer entre elles les teneurs en désordre des différentes formes de VPg précédemment obtenues (Woody 2010) (cf. Matériel et Méthodes).
Pour chacune des VPg, 10 spectres indépendants sont enregistrés (figure III.2A), la moyenne des ratios 205nm/220nm est déterminée ainsi que l'écart-type associé à l'écart obtenu entre ces différentes répétitions de lectures (figure III.2B).

Le mutant OD2 apparait ainsi comme étant le moins désordonné, suivi de SON41p, puis ID2 et enfin ID1, qui présente le signal de désordre le plus important. Ces différences, faibles mais significatives (p-values < 0,001) valident les prédictions de désordre réalisées précédemment. Les différentes mutations introduites dans la forme sauvage VPg-SON41p ont bel et bien permis de diminuer le désordre de la VPg dans le cas du mutant OD2, ou de l'augmenter dans le cas des mutants ID2 et ID1.

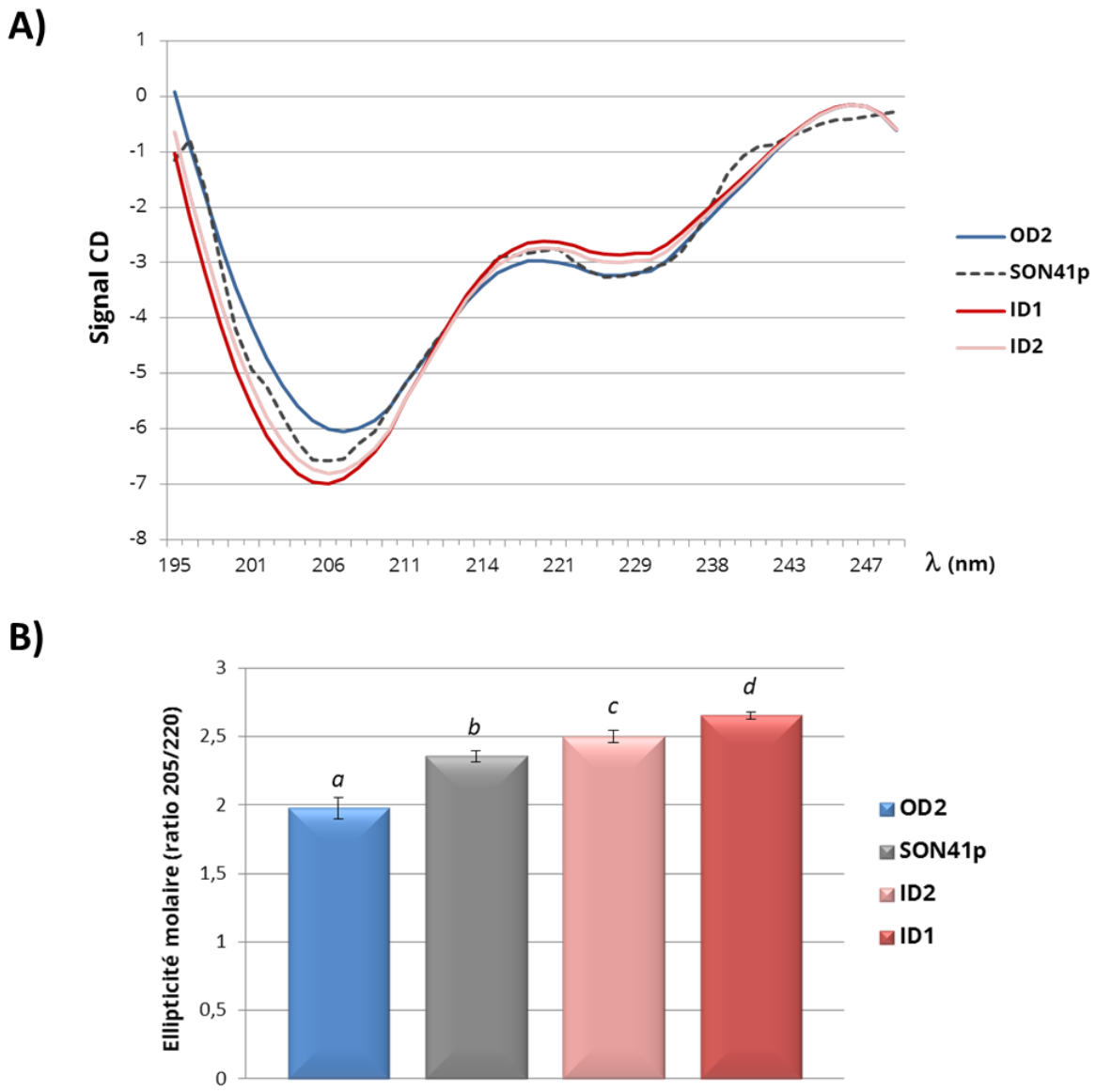


Figure III.3|Détermination de la proportion de désordre dans chaque forme de la VPg par dichroïsme circulaire. A) Moyenne des ratios d'absorbance obtenus après 10 lectures indépendantes pour chacune des protéines testées. B) Détermination de l'ellipticité molaire définie par le ratio des absorbances lues à 205 et 220 nm. *p-values < 0,001 (Test de Kruskal-Wallis)*.

Le signal dichroïque enregistré résulte de la contribution de l'ensemble de la structure de la VPg. Même s'il ne peut être totalement exclu que les modifications dans la zone centrale aient des répercussions à distance, il est vraisemblable que les variations observées traduisent l'état plus ou moins ordonné de cette région centrale.

C) Test de fitness des mutants SON41p ID1, ID2 et OD2 sur piment sensible Yolo Wonder, porteur de l'allèle pvr2^{+}

Afin de déterminer l'impact des mutations introduites pour chacune des formes de VPg sur la fitness globale du PVY, des piments Yolo Wonder, porteur de l'allèle de sensibilité au PVY pvr2^{3} sont inoculés avec les virus correspondant à l'isolat SON41p ainsi que les 3 mutants ID1, ID2 et OD2.
Le nombre de plantes présentant des symptômes de type mosaïques ainsi que l'accumulation virale dans chacune d'elles (mesurée par la technique de DAS-ELISA) ont été déterminé 30 jours après inoculation (figure III.4).

Comme ce qui est observé lors de l'infection de Yolo Wonder par l'isolat SON41p, **100% des plantes inoculées présentent des symptômes dès 10 jours post inoculation pour chacun des 3 mutants ID1, ID2 et OD2.** En revanche l'accumulation virale dans les plantes inoculées avec OD2 est significativement plus faible que celle observées chez les plantes inoculées avec SON41p (WT), (ID1) et (ID2).
Enfin, pour déterminer la stabilité des mutations au cours de l'infection, une extraction d'ARN ainsi qu'un séquençage en Sanger de la région génomique codant pour la VPg ont été réalisés chez 10 des 20 plantes inoculées par condition. **Aucune mutation n'est observée, par rapport à la séquence du virus inoculé, ce qui démontre la stabilité des mutations au cours de l'infection dans l'hôte sensible Yolo Wonder.** En particulier, aucun phénomène de réversion n'est détecté, ce qui permet d'associer les effets observés, tant en termes de symptômes que d'accumulation virale, aux mutations introduites.
On peut donc conclure que, malgré une plus faible accumulation virale détectée pour OD2, l'ensemble des mutants sont capables de s'accumuler de façon assez importante pour conduire à l'infection de 100% des plantes inoculées. Les mutations introduites dans ID1, ID2 et dans une moindre mesure OD2, ne semblent donc pas affecter la fitness du virus lors du processus d'infection sur hôte sensible.
Pour déterminer si les différents degrés de désordre introduit dans la région centrale de la VPg du PVY peuvent avoir un effet sur sa capacité à s'adapter et à contourner la résistance pvr2^{3}, les mutants de PVY codant pour les différentes VPg ont été inoculé à des piments de variété HD285, homozygotes pour l'allèle pvr2^{3} et donc initialement résistants au SON41p.

A)

	PVY SON41p			
	OD2	**WT**	**ID2**	**ID1**
Plantes inoculées	20	20	20	20
Plantes inféctées	20	20	20	20
% infection	**100%**	**100%**	**100%**	**100%**

B)

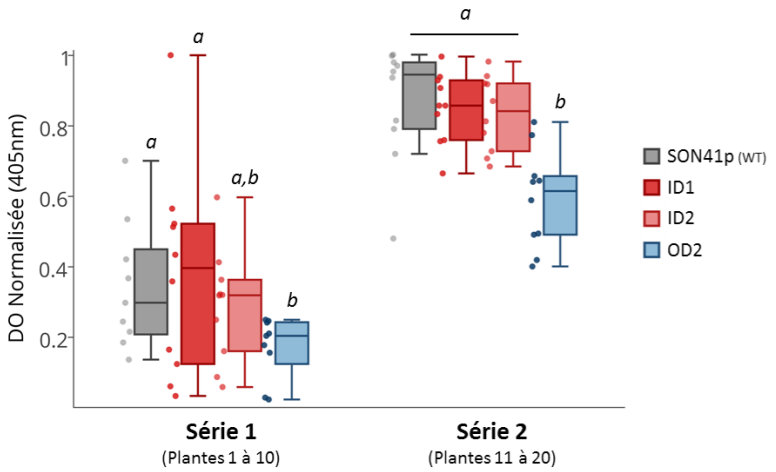


Figure III.4|Suivi d'infection sur piments Yolo Wonder par l'isolat SON41p (WT) et les formes dérivées porteuses des VPg mutantes ID1, ID2 et OD2. A) Nombre de plantes infectées (présentant des symptômes de type mosaïques dès 10 jours post-inoculation. B) Accumulation relative moyenne des différentes formes de PVY à 30 jours post-inoculation. *p-value < 0,001 (Test de Kruskal-Wallis).*

A)

	PVY SON41p			
	OD2	**WT**	**ID2**	**ID1**
Inoculated plants	120	120	120	119
Infected plants	0	47	74	111
% infection	**0%[a]**	**39%[b]**	**62%[c]**	**93%[d]**

B)

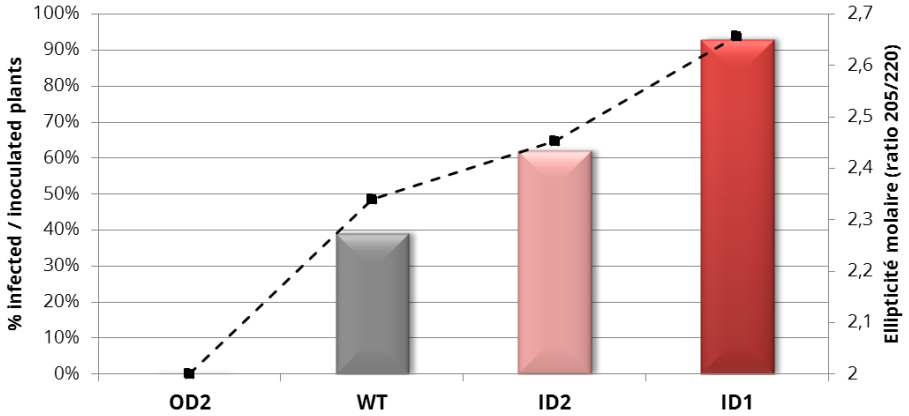


Figure III.5|Taux de contournement de la résistance pvr2[3] associé aux différentes constructions de PVY, selon le désordre dans la VPg. A) Taux de contournement observé chez les piments HD285 par rapport au nombre total de plantes inoculées pour chaque virus. **B)** Comparaison entre le taux de contournement observé et le rapport des ellipticités molaires 205/220nm (ligne en pointillée) correspondant à chaque forme de VPg. *p-values < 0.001 (Test de Z).*

D) Estimation de la capacité de contournement de résistance des différents mutants de « désordre » de la VPg

Les différentes formes de PVY, correspondant à chacune des 4 constructions SON41p-WT, SON41p-ID1, SON41p-ID2 et SON41p-OD2, ont ensuite été inoculées sur des piments HD285 (cf. Matériel et Méthodes). 36 jours post-inoculation, le nombre de plantes présentant des symptômes de type nécrose, caractéristiques du contournement de pvr2[3] par SON41p, ont été dénombrées. Afin d'identifier les faux-négatifs, l'accumulation virale a été vérifiée chez les plantes ne présentant pas de symptômes, par détection DAS-ELISA. Les résultats, présentés figure III.5, tiennent compte à la fois des plantes symptomatiques et de celles présentant un titre viral caractéristique d'une infection.

Les changements de désordre de la VPg sont corrélés aux capacités de l'isolat SON41p du PVY à restaurer une infection sur l'hôte initialement résistant HD285

Les pourcentages de plantes infectées obtenues montrent **que les mutants présentent tous une capacité de contournement différente** (*p-value* < 0,0001 ; test de Kruskal-Wallis en comparaison 2 à 2). Aucune des plantes inoculées avec la construction OD2 n'a présenté de signes d'infection. En revanche, environ 40% des plantes inoculées avec la construction wild-type (WT) ont été infectées, 62% des plantes inoculées avec ID2 et 93% des plantes inoculées avec ID1 (figure III.5A).

Si on rapporte ces résultats à ceux concernant la détermination du taux de désordre dans chacun des mutants VPg (figure III.5B), force est de constater que **la fréquence de contournement est corrélée au degré de désordre de la zone centrale de la VPg.**

L'étape suivante a donc consisté à essayer de comprendre quels processus sous-jacents sont mis en jeu au niveau moléculaire lors du contournement pour ces différents virus, et notamment d'émettre des hypothèses sur l'implication du degré de désordre dans les différences de capacités adaptatives observées.

E) Détermination des évènements mutationnels intervenus au cours du contournement

Les VPg de 20 plantes par virus ont été séquencées afin d'identifier les déterminants moléculaires associés aux différents taux de contournement observés précédemment. L'ensemble des mutations retrouvées ainsi que leurs fréquences respectives sont rassemblées dans la figure III.6.

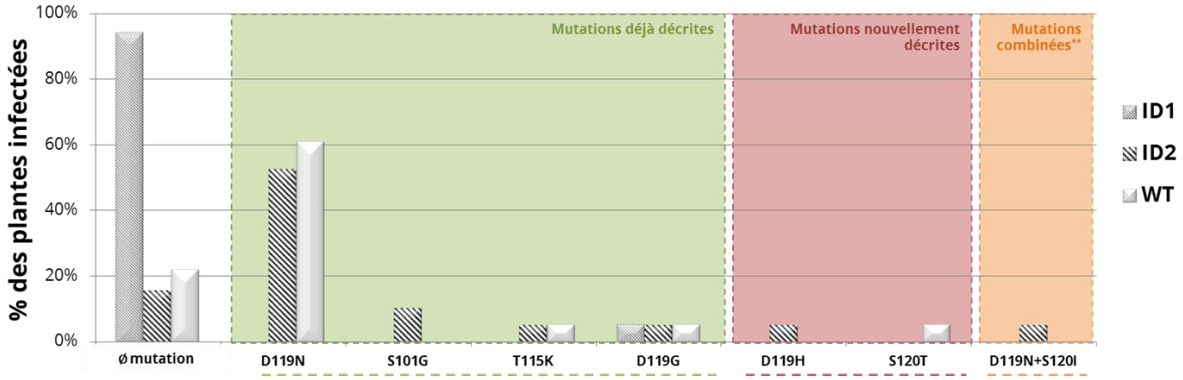
L'infection par WT et ID2 passe par l'apparition de mutations de contournement

La majorité des mutations de contournement de la résistance pvr2[3] identifiées lors du séquençage des contournants ID2 et WT ont d'ores et déjà été décrites dans des études précédentes (Ayme et al. 2006; Montarry et al. 2011) (Figure III.6). Deux « nouvelles » simples mutations sont identifiées comme associées au contournement (D119H chez ID2 et S120T chez WT).

A)

		Mutations uniques						Combinaisons**
	Aucune mutation	D119N	D119G	S101G	T115K	**D119H**	**S120T**	**D119N+S120I**
ID1	(17/18)	-	-	-	(1/18)	-	-	-
ID2	(3/19)	(10/19)	(1/19)	(2/19)	(1/19)	**(1/19)***	-	**(1/19)***
OD2	-	-	-	-	-	-	-	-
WT	(4/18)	(11/18)	(1/18)	-	(1/18)	-	**(1/18)***	-

B)



C)

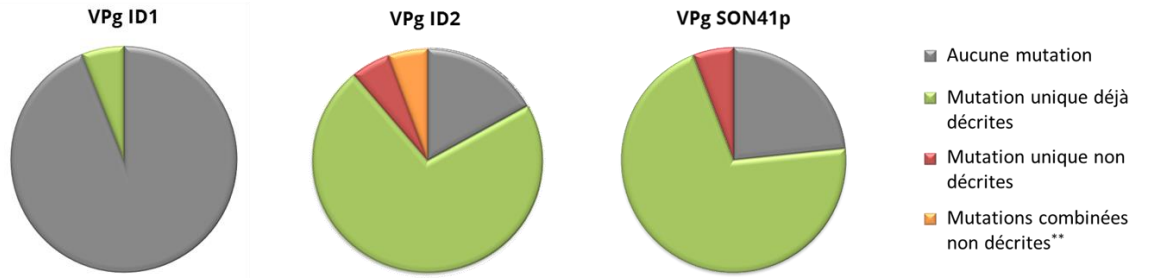


Figure III.6|Evènements mutationnels conduisant au contournement de la résistance pvr2[3] par SON41p (WT) et les mutants VPg. **A)** Effectifs des mutations selon le nombre de plantes étudiées. Le symbole « * » indique que les échantillons ont été séquencés à deux reprises. **B)** Répartition des profils mutationnels des virus présents sur plantes HD285 infectées après inoculation par ID1, ID2 et WT. Les mutations déjà décrites dans (Ayme et al. 2006; Montarry et al. 2011) sont indiquées en vert, les mutations uniques nouvellement décrites sont indiquées en rouge et les mutations combinées nouvellement décrites sont indiquées en orange. **C)** Proportion parmi les génomes PVY séquencés de ceux ne présentant pas de mutations (gris) ou des mutations de contournement déjà décrites (vert), ou nouvelles (orange et rouge). *"**": le séquençage en Sanger de la séquence consensus au sein d'une population ne permet pas de discriminer un évènement de double mutation d'un mélange de deux simples mutants dans la population.*

La combinaison de deux substitutions (D119N et S120I) sont également observées dans une des populations virale de ID2 ayant conduit au contournement. La méthode Sanger de séquençage utilisée ne permet cependant pas de déterminer si ces deux mutations sont en interaction sur les mêmes génomes ou bien si les simples mutants, porteurs soit de l'une soit de l'autre de ces mutations, existent en mélange équimolaire dans la population. Il est également à noter que, pour une faible proportion de plantes inoculées avec les constructions WT et ID2, aucune mutation n'est associé au contournement. Pour conclure chez WT et ID2, la capacité à restaurer une infection chez les plantes initialement résistantes HD285 passe donc en très large majorité par l'apparition des mêmes mutations dans la VPg des populations virales descendantes. Cependant le taux de succès plus important d'ID2 par rapport au WT pourrait être imputable à une plus grande permissivité mutationnelle imputable au désordre.

L'infection par ID1 n'est associée à aucune mutation du PVY

En ce qui concerne les plantes infectées suite à l'inoculation du mutant ID1, à l'exception d'une seule, aucune mutation dans la VPg n'est associée à la mise en place de l'infection. La capacité à induire l'infection dans le cas d'ID1 ne semble donc pas passer par le processus de contournement classique habituellement décrit pour ce pathosystème.

Afin de déterminer si les mutations, ou l'absence de celles-ci, observées dans les VPg constituent les uniques déterminants des différences phénotypiques observées, des séquençages en Sanger sur génome complet ont également été réalisés sur 3 des plantes infectées pour chaque condition (voir Matériel et Méthodes). Les 3 échantillons sont choisis de façon à avoir les deux situations différentes : des échantillons porteurs de mutations dans leur VPg et des échantillons porteurs d'aucune mutation dans la VPg. Ce séquençage en génome complet permet de voir que, en dehors de la séquence codant pour la VPg, aucun autre évènement de mutation non-synonyme n'est détecté, y compris pour les génomes ne présentant pas de mutation de contournement sur leur VPg. Nous pouvons donc conclure qu'aucun évènement de mutations alternatives ou collaboratives (épistasie), faisant intervenir d'autres protéines que la VPg, ne semblent rentrer en jeux dans les évènements de contournement décrits pour le WT, ID1 et ID2. Les changements de désordre introduits n'ont pas modifié le déterminisme du contournement, validé une fois de plus comme étant uniquement associé à la VPg, chez ce pathosystème (Montarry et al. 2011).

F) Détermination des évènements moléculaires associés au contournement: analyse de l'interaction VPg/eIF4E

Afin de tenter d'expliquer les différences observées en termes de capacités adaptatives des PVY porteurs des VPg WT et mutantes pour restaurer une infection dans les piments résistants HD285, une analyse quantitative des forces d'interaction entre les VPg et les formes d'eIF4E, correspondant aux expérimentations *in planta*, a été réalisée via des expériences de spectrofluorescence.

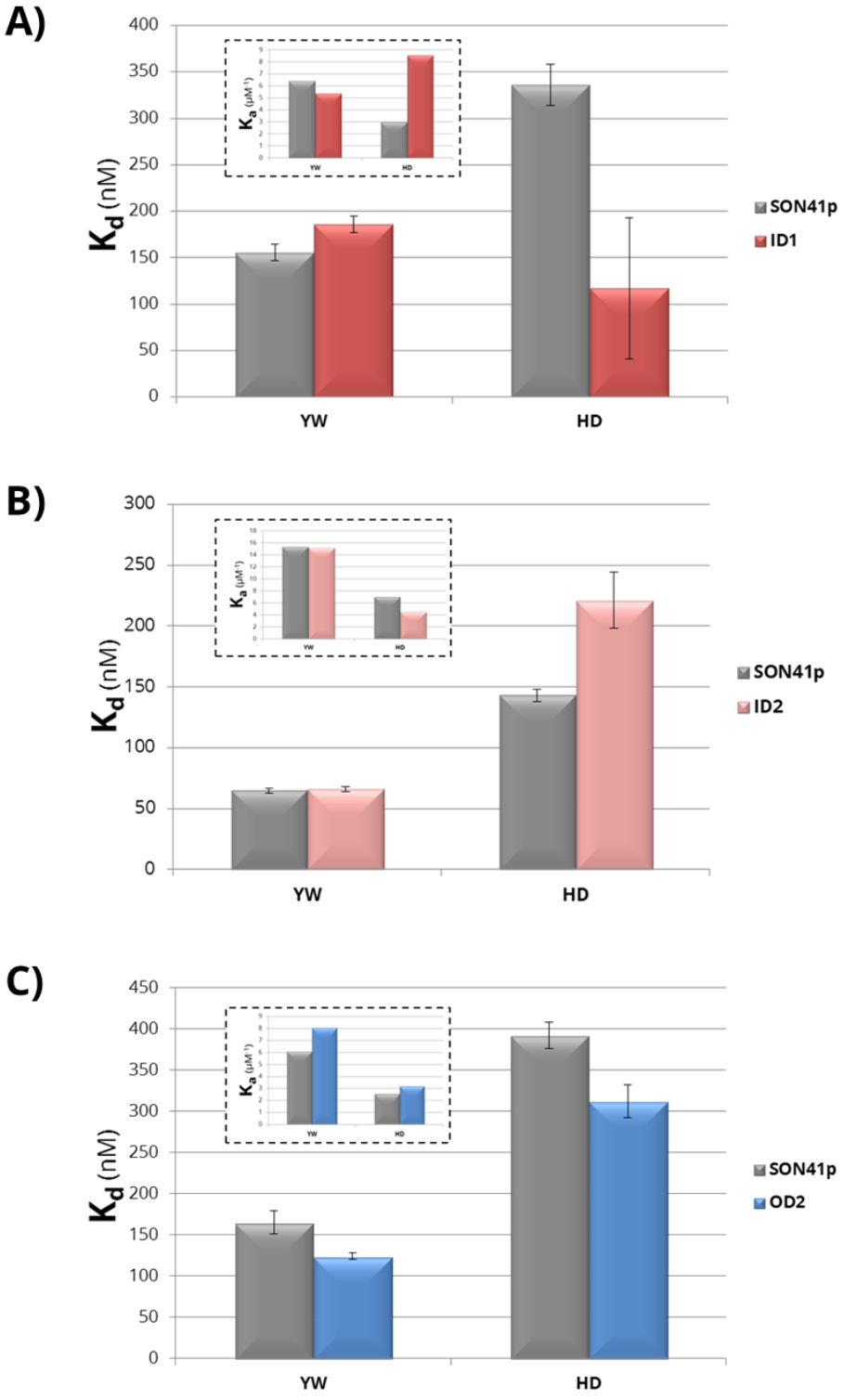


Figure III.7|Constantes d'affinités des interactions eIF4E/VPg déterminées par spectrofluorescence. Les valeurs d'affinités sont représentées par les constantes thermodynamiques de dissociation (K_d) mesurées expérimentalement ainsi que les constantes d'association (K_a=1/K_d) déduites (graphes en inserts). Comparaison des interactions de eIF4E correspondant aux génotypes de piments Yolo Wonder (YW) et HD285 (HD) avec la VPg de l'isolat SON41p avec **A)** VPg-ID1. **B)** VPg-ID2. **C)** VPg OD2.

Les protéines eIF4E contiennent plusieurs résidus tryptophane dont la fluorescence décroit partiellement lors de l'interaction avec la VPg. Cette décroissance est proportionnelle à la quantité de complexe VPg-eIF4E qui se forme. Il est donc possible de titrer l'interaction entre les deux protéines en suivant la décroissance de la fluorescence par ajout de quantités croissantes de VPg à une solution d'eIF4E (cf. Matériel et méthodes). De ces mesures on déduit la constante thermodynamique de dissociation (K_d) du complexe binaire. Cette constante est par définition **inversement corrélée à l'affinité** entre les deux interacteurs.

Comme les VPg WT et mutantes, les deux formes d'eIF4E codées respectivement par les variétés Yolo Wonder et HD285 (c'est-à-dire celles utilisées pour les expérimentations *in planta*), ont été produites sous forme recombinantes dans E.coli et utilisées pour les mesures (cf. Matériel et méthodes). La figure III.7 présente les constantes de dissociation (K_d) et d'association (K_a) obtenues pour chaque interaction testée.

Les formes recombinantes eIF4E obtenues étant peu sables et parfois contaminées, leur capacité d'interaction avec un analogue de la coiffe des ARNm eucaryotes, le m^7-GTP a été systématiquement testée afin de s'assurer de leur fonctionnalité lors de la titration des variants de VPg. Pour l'ensemble des expériences d'interactions, les eIF4E ont présenté des affinités pour la coiffe, du même ordre de grandeur que celle définie pour l'eIF4E de laitue (Michon et al. 2006), soit de l'ordre de 300nM. Nous avons cependant observé des disparités assez importantes dans les valeurs obtenues. De cette manière, il s'est avéré difficile de comparer des blocs de mesure distincts. C'est pourquoi en plus de la titration de eIF4E par le mutant VPg, chaque bloc d'expérience comprenait systématiquement une titration comparative de l'interaction eIF4E-VPg SON41p-WT ainsi que le contrôle de la fonctionnalité d'eIF4E par le test d'interaction eIF4E-m7GTP. Les caractéristiques d'interactions de chaque mutant avec les formes d'eIF4E peuvent donc être discutées au regard de celles obtenues pour la VPg WT au sein du bloc correspondant.

La forme VPg-WT est plus affine pour eIF4E Yolo Wonder que pour HD285

Il n'est pas surprenant d'observer que la VPg de SON41p interagit plus fortement avec le facteur eIF4E issu de la variété sensible Yolo Wonder qu'avec son homologue de la variété résistante HD285 (Barres grises sur la Figure III.7). Ces résultats confirment la corrélation entre l'interaction VPg-eIF4E et la capacité d'interaction préalablement observée pour ce pathosystème (Charron et al. 2008; Ayme et al. 2006; Ayme et al. 2007). La mesure d'une interaction significative entre eIF4E-HD et VPg-SON41p semble de plus expliquer que SON41p parvienne à évoluer dans l'hôte HD285 avec émergence des mutants contournants de la résistance, durant les 36 jours qui suivent l'inoculation.

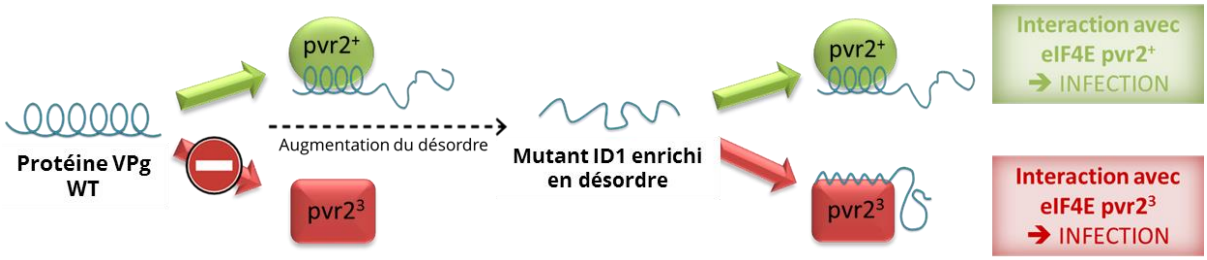


***Figure III.8*|Modèle du rôle du désordre dans le cas du mutant ID1.** L'augmentation du désordre chez ce mutant permet l'élargissement de son spectre d'interaction et donc de son spectre de virulence face aux génotypes pvr2.

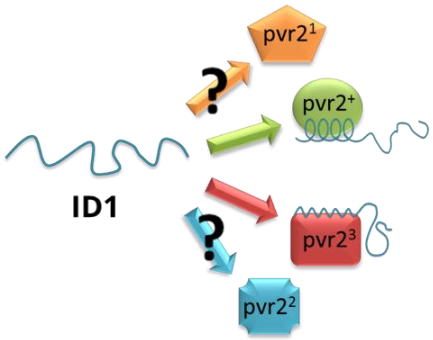


***Figure III.9*|Hypothèse de l'élargissement du spectre d'interaction du mutant ID1.** Conclure sur un élargissement de spécificité associé au mutant ID1 ou contraire à un transfert de spécificité nécessite de tester la capacité d'infection de ce mutant chez les autres variétés de piments porteuses des allèles de résistance pvr2^1 et pvr2^2 et initialement sensibles à l'isolat SON41p.

La forme VPg-ID1 interagit de façon comparable avec les deux formes d'eIF4E

La force d'interaction de VPg-ID1 pour les deux facteurs eIF4E est du même ordre que celle observée pour l'interaction de la VPg-SON41p WT avec la forme eIF4E du piment sensible Yolo Wonder.

Les formes VPg-ID2 et VPg-OD2 présentent le même profil d'affinité pour les formes eIF4E sensibles et résistantes que la VPg-SON41p

Les formes mutantes ID2 et OD2 de la VPg présentent des caractéristiques d'interactions avec les différents eIF4E semblables à celles observées le cas de VPg-SON41p WT, à savoir une interaction forte pour le facteur eIF4E Yolo Wonder et deux à quatre fois plus faible pour le facteur eIF4E HD285.

Remarque : les expériences d'interactions *in vitro* ont été réalisées à partir de protéines exprimées dans *E.coli*. Certains biais sont donc à considérer et parmi eux l'absence de modifications post-traductionnelles chez les organismes procaryotes. Des expériences complémentaires d'interactions *in planta* seront donc à envisager par la suite.

3. Discussion et Perspectives

A) Cas du mutant ID1 : le désordre au service de la multi-interaction ?

Notre étude a permis de montrer que les 3 mutations introduites pour parvenir au mutant ID1 de la VPg ont permis d'augmenter le désordre de celle-ci, au niveau de sa région centrale. Ces trois mutations et l'augmentation du désordre qu'elles induisent chez ID1 sont associées à une capacité accrue (93% des plantes inoculées) à infecter le piment HD285 porteur de l'allèle de résistance pvr2[3]. L'absence de mutations détectée chez les virus extraits de ces plantes infectées montre que cette augmentation importante de la virulence du PVY n'est pas associée à un évènement de contournement qui aurait eu lieu post-inoculation. L'accroissement significatif de l'affinité d'interaction détectée entre la VPg ID1 et le facteur eIF4E HD285 par rapport à la VPg WT est un contributeur très probable de l'augmentation de virulence observée pour le mutant ID1. Dans ce cas précis, l'augmentation du désordre associée à ID1 via les mutations introduites sont associées à une augmentation du spectre d'interaction de ce facteur (Figure III.8). Ces résultats semblent donc illustrer un des avantages fonctionnels du désordre de la zone centrale de la VPg qui, par sa plasticité conformationnelle, permettrait un élargissement de spécificité d'interaction via la mise en place d'une compatibilité et une interaction moléculaire de haute affinité avec la forme HD285 d'eIF4E, tout en conservant sa capacité d'interaction avec la forme eIF4E du piment sensible Yolo Wonder. Jusqu'à quel degré pourrait s'étendre cette augmentation du spectre d'interaction ? S'agit-il d'un élargissement de spécificité ou bien d'un transfert de spécificité ? Cette question nécessite de développer des expériences complémentaires faisant intervenir d'autres formes d'eIF4E issues d'autres allèles de résistances pvr2 (Figure III.9), initialement capables d'interagir avec la VPg sauvage de l'isolat SON41p.

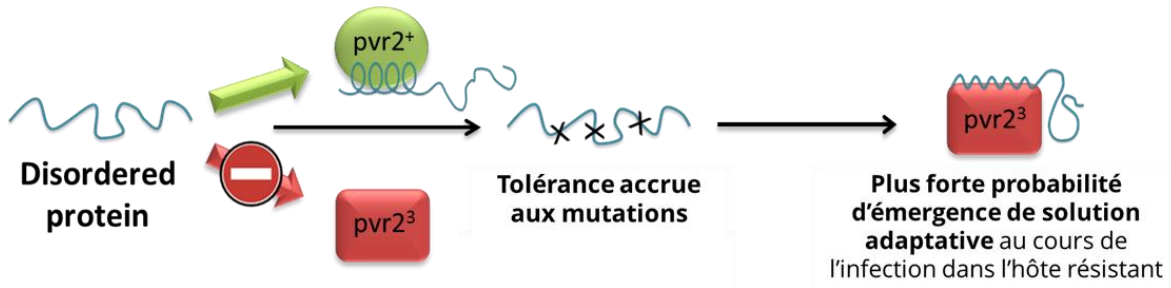


***Figure III.10*|Modèle du rôle du désordre dans le cas du mutant ID2.** L'augmentation de désordre dans la zone centrale favorise, via sa robustesse mutationnelle accrue, la probabilité d'émergence des mutations de contournement pour restaurer l'interaction avec le facteur eIF4E codé par pvr2^3 et donc l'infection sur hôte HD285.

B) Cas des mutants ID2 et OD2 : le désordre au service de l'adaptation ?

Le mutant VPg-ID2, également caractérisé par une augmentation du désordre dans la zone centrale, est également associé à une plus grande capacité d'infection du PVY correspondant sur piment résistants HD285, porteurs de l'allèle de résistance pvr2^3. Pourtant les mécanismes moléculaires qui sont potentiellement liés à ces observations sont différents de ceux observés pour ID1. En effet, dans le cas d'ID2, la capacité à infecter les piments HD285 est quasi-exclusivement associée à des évènements mutationnels donc à la situation du contournement de résistance observé dans le cas de l'isolat naturel SON41p du PVY. Les affinités de la VPg-ID2 pour les formes eIF4E sont également semblables voire inférieures à celles déterminées pour la forme WT de la VPg et ne permettent donc pas d'expliquer la différence significative de la proportion de plantes infectées par ID2.

Compte-tenu de ces résultats et des rôles généralement attribués au désordre (cf. Introduction), une hypothèse serait que **la plus grande proportion de désordre associée à la zone centrale d'ID2 augmenterait la robustesse mutationnelle de cette région.** Ceci se pourrait se traduire ainsi par **i) une tolérance accrue des mutations de contournement ; ii) l'émergence de nouvelles solutions adaptatives**.

Dans ces deux cas de figure, l'augmentation du désordre conduirait à un potentiel adaptatif plus important (Figure III.10).

En effet l'adaptation de l'isolat SON41p à la résistance pvr2^3 peut être vue comme une véritable course contre la montre. Malgré sa capacité à se répliquer dans l'hôte résistant, le virus ne peut pas rester dans sa forme non adaptée indéfiniment. Pour qu'il y ait contournement, les formes adaptatives doivent donc émerger dans un laps de temps donné, au-delà duquel les mécanismes de défense de la plante hôte ainsi que la dérive génétique conduisent à l'élimination totale du virus. La probabilité d'apparition des mutations de contournement ainsi que leur impact sur la fitness du virus sont donc des paramètres déterminants dans la mise en place du contournement dans un temps donné. Plus le nombre de solutions est élevé, plus la probabilité d'atteindre aléatoirement ces solutions est importante. La réduction de l'impact de ces mutations sur la structure, et indirectement sur la fitness du virus, peut également conduire à l'augmentation de la capacité adaptative, en postulant que leur apparition même tardive permettrait au virus de s'accumuler rapidement, et de restaurer une infection.

A l'heure actuelle, les résultats concernant le mutant ID2 ne permettent pas de comprendre précisément comment l'augmentation du désordre chez ce mutant peut être reliée à ses capacités adaptatives accrues. Des expériences complémentaires sont donc à envisager pour permettre de vérifier les hypothèses présentées plus haut.

La première hypothèse visait à corréler le désordre à l'augmentation de la fitness des variants contournants. Pour tester cela, il pourrait être envisagé de comparer par DAS-ELISA les capacités d'accumulation sur hôte HD285 d'un contournant donné, selon qu'il est introduit dans la VPg WT ou dans la VPg ID2.

Conclure quant au rôle du désordre dans l'augmentation du nombre de solutions adaptatives chez le mutant ID2 passe par l'identification d'évènements mutationnels retrouvés uniquement chez les contournants issus de l'évolution de SON41p-ID2. Pour maximiser nos chances d'isoler ces hypothétiques évènements mutationnels propres à ID2, il sera nécessaire de séquencer un plus grand nombre de contournants de SON41p WT et ID2 prélevés sur les piments HD285 infectés. De plus, l'analyse de l'effet de ces mutations potentiellement « désordre-dépendantes » sur la fitness du virus lorsque celles-ci sont introduites dans un environnement moins désordonné, comme celui de la VPg-WT, devrait permettre de renforcer cette hypothèse de robustesse mutationnelle.

Les résultats obtenus dans le cas du mutant OD2 convergent également vers l'hypothèse du désordre promoteur de l'adaptation de la VPg à la résistance pvr2[3].
En effet, la diminution du désordre dans la zone centrale de la VPg-OD2 est associée à une suppression totale de la capacité de ce mutant à restaurer une infection sur le piment résistant HD285.
Les mutations introduites dans ce mutants ne semblent pas affecter la capacité d'infection du PVY correspondant sur plantes sensibles et la VPg-OD2 présente les mêmes caractéristiques d'interactions que la VPg-WT pour les formes eIF4E YW et HD. Les différences de capacités adaptatives observées semblent donc plus probablement découler d'une incapacité à faire émerger des formes adaptées que d'un effet direct de ces mutations pour le cycle viral. L'environnement plus structuré de la région centrale du mutant OD2 diminuerait la robustesse mutationnelle et empêcherait ainsi le maintien des mutations adaptatives décrites chez les formes WT et ID2.
Le désordre de la zone centrale de la VPg apparait comme ubiquitaire chez les potyvirus. S'il s'avère que l'une de ces fonctions est de d'explorer un champ mutationnel plus large, ou encore de tamponner les effets délétères des mutations de contournement en vue de l'adaptation à l'hôte, il s'agirait là d'un exemple démonstratif de la fonction adaptative du désordre chez les virus.
Toutefois, il est important de noter la diminution significative d'accumulation virale détectée pour le mutant OD2 sur piments sensibles, qui pourrait indiquer un effet délétère faible de ces mutations sur le cycle du virus. Ceci pourrait aussi expliquer l'absence de contournement observé chez ce mutant suite à l'inoculation sur plantes résistantes.

C) Considérations d'ordre plus générales sur l'interprétation des données obtenues

Désordre et différences adaptatives : causalité ou simple corrélation ?

Le désordre est séquence-dépendant. Toute tentative de modification de ce dernier passe nécessairement par des changements en acides aminés. Compte-tenu de notre démarche de changement artificiel du désordre dans la VPg, nous ne pouvons exclure que certaines de nos observations soient imputable à la nature même d'une ou plusieurs des substitutions introduites, plutôt qu'au désordre en lui-même. De plus, la VPg est une protéine multifonctionnelle impliquée dans de nombreuses étapes du cycle des Potyvirus. Les expériences conduites sur plantes sensibles servent néanmoins à s'affranchir *a minima* de ce biais, en partant du principe que si les substitutions introduites dans les différents mutants affectaient de façon importante la fitness viral en interférant dans une autre fonction de la VPg que l'interaction avec eIF4E, ceci se traduirait d'abord au cours de l'infection en plante sensible. De telles différences ne sont pas observées, du moins entre ID1, ID2 et WT (OD2 présente une accumulation moins importante dans l'hôte sensible Yolo Wonder).

Le désordre de la zone centrale de la VPg, pas indispensable pour accomplir le cycle viral ?

L'observation d'une capacité d'infection dans l'hôte sensible maintenue pour le mutant OD2 malgré une diminution notable du désordre permet de voir que dans ce contexte donné, le désordre ne semble pas associé à des fonctions cruciales du cycle viral. Cependant OD2 s'accumule plus faiblement dans les piments sensibles par rapport aux trois autres virus (SON41p, ID1 et ID2) dont les VPg sont plus désordonnées. La différence d'accumulation pourrait donc être reliée à la différence de désordre observée. L'interaction de la VPg-OD2 avec eIF4E ne semble cependant pas être significativement affectée. L'ensemble de ces résultats suggère donc un rôle du désordre dans une ou plusieurs des étapes du cycle viral, qui serait indépendant de l'interaction entre VPg et eIF4E. La VPg interagit avec de nombreux facteurs viraux (Jiang & Laliberté 2011), le désordre de la zone centrale pourrait donc constituer un facteur clé dans la mise en place ou la stabilisation de certaines de ces interactions. L'étude de l'impact des changements en désordre dans les trois mutants ID1, ID2 et OD2 sur les interactions de la VPg avec des partenaires déjà identifiés, et autres que eIF4E, pourrait donc constituer une piste d'étude intéressante.

Pourquoi les mutants ID1 et ID2, porteurs d'un avantage fonctionnel certain dans notre contexte expérimental, ne sont-ils pas retrouvés dans la nature ?

Les mutants ID1 et ID2 sont tous deux porteurs d'un avantage fonctionnel et donc à priori sélectif par rapport au mutant WT, en étant capables de mieux infecter la plante résistante HD285, tout en conservant une fitness importante dans l'hôte sensible.

En effet, il est surprenant de constater que les triples (ID1) et doubles mutations (ID2) entraînant respectivement soit un gain de virulence dans l'hôte résistant soit un gain adaptatif, ne sont pas associées à une perte de fitness dans l'hôte sensible (en se basant à la fois sur l'accumulation virale et les symptômes induits sur Yolo Wonder).

Pourtant, la plupart des mutations de virulence naturelles ayant été caractérisées dans la nature sont très fréquemment désavantageuses dans l'hôte sensible, ce qui explique que des isolats non contournants puisse encore se maintenir dans la nature (García-Arenal & Fraile 2013). En effet, des contraintes importantes s'exercent sur les génomes viraux, les mutations pouvant entrainer plusieurs conséquences fonctionnelles sur par exemple les frameshifts génomiques ou encore la multifonctionnalité des protéines virales.

Plusieurs explications peuvent alors être envisagées pour expliquer pourquoi nos mutants, compte-tenu de leur avantage sélectif théorique, ne sont pas retrouvés en abondance dans la nature :

i) *Ces mutants existent dans la nature mais n'ont pas encore été identifiés.*

ii) *Ces mutants sont moins fit que le WT lorsqu'ils sont mis en compétition dans la même plante.* Tester cette hypothèse reviendrait à réaliser un test de compétition *in planta*, en co-inoculant les deux clones infectieux et en mesurant à différents temps, l'accumulation relative de chacun d'eux.

iii) *Ces mutants ne sont pas présents dans la nature car ils sont moins fit que la forme PVY^{SON41p}-WT.* Une des raisons à cela pourrait être que les mutations introduites sont délétères pour le virus en impactant des processus tels que la transmission vectorielle par puceron, ignorée lors notre analyse en inoculation mécanique. Un tel effet délétère n'aurait ainsi pas pu être observé au cours de notre analyse en raison des biais que présente notre approche expérimentale par rapport au processus naturel d'infection et la pression de sélection qui lui est associée. Cependant, l'hypothèse d'un effet délétère des mutations associées à la VPg est peu probable, cette protéine n'ayant pas été identifiée comme un déterminant de la transmission vectorielle par puceron, chez ce pathosystème (Ayme et al. 2007).

iv) *Les mutants ne sont pas survenus naturellement au cours de l'évolution du PVY car le chemin adaptatif vers ces mutations est trop improbable ou difficile à franchir.* Les triples et doubles mutations respectivement introduites dans les mutants ID1 et ID2 présentent, de par leur nombre, une très faible probabilité d'apparition simultanée. Selon toute vraisemblance, les mutations n'auraient pu apparaitre que successivement au cours de l'évolution du virus. Ceci sous-tendrait que les formes intermédiaires aient une fitness suffisante pour pouvoir être maintenu au cours de l'évolution. Un moyen de tester cela expérimentalement consisterait à construire les simples et doubles mutants (dans le cas d'ID1) intermédiaires et de tester leur fitness *in planta*.

La mise en perspectives de ces résultats dans le contexte applicatif plus large de prédiction des durabilités des résistances variétales est présentée dans la dernière section de ce travail (cf. Partie « Conclusions et perspectives »).

Conclusions et Perspectives de la thèse

Ce travail s'inscrit dans une démarche de compréhension fondamentale des fonctions biologiques associées au désordre intrinsèque des protéines chez les virus à ARN.
En particulier, chez certains virus animaux, le désordre a notamment pu être associé à la multifonctionnalité des protéines virales et leur capacité à recruter de nombreux facteurs de l'hôte. Il est également considéré comme un déterminant potentiel de l'incroyable capacité évolutive et adaptative de ces micro-organismes.

L'objectif en commençant cette thèse a donc été de contribuer à l'étude des processus et fonctions associées au désordre intrinsèque, en focalisant nos études sur un genre majeur de phytovirus, les *Potyvirus*.

Le genre Potyvirus : un outil autant qu'une cible

Focaliser notre étude aux virus de plantes a permis d'une part de déterminer si les fonctions précédemment mises en évidences chez les virus animaux, peuvent être étendues aux virus de plantes. Si tel est le cas, l'étude du désordre dans les protéomes des virus de plantes peut s'inscrire dans une démarche systématique d'annotation fonctionnelle pour améliorer nos connaissances sur la biologie de ce genre majeur de phytovirus.

D'autre part, les phytovirus constituent un outil de choix pour mener des expériences d'évolution expérimentales nous permettant de tester l'hypothèse du rôle du désordre sur l'adaptation virale.

"Montres-moi ton désordre, je te dirais qui tu es" : l'intégration du désordre dans les processus d'annotation fonctionnelle des protéines

Les méthodes d'annotation fonctionnelle des protéines

Les progrès croissants en matière de séquençage ont permis d'engendrer un nombre considérable de séquences génomiques et protéomiques. Le développement d'outils bioinformatiques permettant une annotation fonctionnelle préliminaire de ces séquences est donc devenu primordiale et constitue désormais une démarche incontournable pour l'ensemble de la communauté scientifique. Dans le cas des protéines, cette annotation fonctionnelle est basée sur la comparaison de séquences. Elle nécessite principalement l'identification de domaines[1], d'abord démontrés comme fonctionnels chez les espèces modèles et conservés au cours de l'évolution. Ces domaines sont, par définition, structurés et évolutivement contraints. Mais cette annotation fonctionnelle par identification de domaines n'est pas toujours fructueuse. On estime en effet que 40% des séquences en acides aminés présentes dans la banque de données UniprotKB/Swiss-Prot ne sont assignées à aucune fonction (Cozzetto & Jones 2013). Dans ce contexte, et de par sa complémentarité, l'identification du désordre

[1] Le terme de domaine protéique fait référence aux unités structurales et fonctionnelles des protéines et définit une région protéique capable de se replier et de fonctionner de façon autonome.

dans les protéines peut permettre d'enrichir ces annotations. Ceci illustre l'importance d'introduire les outils associés à la prédiction et l'annotation fonctionnelle des régions désordonnées aux démarches systématiques d'annotation des protéines.
La mise à disposition de nombreux outils *in silico* de prédiction permet aujourd'hui de prédire le désordre intrinsèque et ses caractéristiques de façon assez robuste (Ghalwash et al. 2012; Dolan et al. 2015). Parmi ces caractéristiques, le potentiel de nombreuses IDP/IDR à interagir avec d'autres partenaires ainsi que les résidus impliqués dans cette reconnaissance moléculaire, peuvent par exemple être prédits. De plus, le nombre croissant de données associées aux fonctions biologiques que portent les régions désordonnées permet de consolider des pistes de recherche pour des investigations expérimentale futures.

L'estimation in silico du désordre chez les potyvirus pour inspirer l'exploration expérimentale de leur biologie

C'est dans cette optique que la première partie de ce travail de thèse a été réalisée. L'estimation du désordre dans les protéomes de plusieurs espèces virales appartenant au genre *Potyvirus* a d'abord permis de constater que les potyvirus possèdent 20% de résidus désordonnés dans leurs protéomes, et que les protéines peuvent être classées selon leur taux de désordre en 3 groupes : fortement désordonnées (VPg, CP et P1), modérément désordonnées (HC-Pro, P3, P3N-PIPO, CI) et au contraire très ordonnées (NIa-pro et NIb). Cette homogénéité inter-espèces de la répartition protéomique du désordre est associée à une importante variabilité de séquence, ce qui suggère fortement que le maintien du désordre résulte d'une contrainte évolutive et donc, implicitement, de fonctionnalités associées.
Une cartographie plus fine du désordre au sein de chaque protéine a permis d'identifier des régions désordonnées conservées, attendues donc comme pertinentes biologiquement parlant, au sein des différentes protéines.

Les attributs des fonctions associées à la reconnaissance moléculaire, considérées comme largement majoritaires dans les IDR, ont été appréhendés *in silico* au sein des régions conservées identifiées précédemment. Dans un premier temps, la prédiction de motifs de repliement (MoRF) et de motifs linéaires (ELM) suggère des potentialités d'interactions impliquant les IDR conservées des protéines de potyvirus. Dans un second temps, la mise en perspective des résultats obtenus par rapport aux fonctions moléculaires et cellulaires de chacune des protéines a permis de formuler des hypothèses de fonctions biologiques portées par le désordre chez certaines protéines des potyvirus.
Ainsi le désordre prédit dans la protéine P1 semble par exemple être associé à la présence de motifs de reconnaissance moléculaire capables d'interagir avec les domaines protéiques eucaryotes, tels que les domaines 14-3-3, FHA ou encore WW. La pertinence biologique que constitue la présence de tels motifs sur le protéome des potyvirus mérite d'être expérimentalement testée.

Cependant, l'hétérogénéité rapportée dans certains cas nécessite une validation expérimentale des prédictions de désordre *in silico*. Par exemple, l'importante proportion de désordre prédit dans la région N-terminal des protéines P3 et P3N-PIPO spécifique au LMV peut constituer, si elle est démontrée expérimentalement, une piste exploratoire intéressante. De même, l'absence de désordre prédit dans la région centrale de la VPg du PPV constitue une exception assez intrigante qui nécessiterait une vérification expérimentale.
L'étude approfondie de ces exceptions et l'hétérogénéité du contenu en désordre constituent un autre point d'entrée dans l'étude des fonctions associées au désordre. En effet les études conduites à haut débit sur l'ensemble de la diversité des virus a permis de constater que ceux-ci présentent une variation record dans leurs protéomes (Xue et al. 2012). Alors que la majorité des études sur le rôle du désordre néglige cet aspect et a tendance à ne se focaliser que sur les virus présentant des taux de désordre de plus de 50%, les virus présentant une utilisation moindre de ce dernier sont tout aussi intéressants à étudier, et enrichissent nécessairement notre compréhension du désordre. En effet, si les grandes fonctions virales (réplication, traduction, assemblage…) doivent être nécessairement satisfaites, et si l'économie de génome ainsi que les interactions multiples sont des caractéristiques relativement bien conservées dans le monde des virus à ARN, l'évolution peut avoir opté pour des bases moléculaires différentes n'impliquant pas systématiquement le désordre intrinsèque.

Dans ce contexte, un travail d'identification à haut débit des déterminants pouvant être à la base de cette hétérogénéité a été récemment conduit et s'est basé sur l'analyse de 2,278 génomes viraux très divers (Pushker et al. 2013). Compte-tenu de l'extrême diversité qui existe au sein des virus, cette étude n'a pas permis d'établir de corrélations globales entre le contenu en désordre et les caractéristiques générales (taille de génome, virulence, cycles d'infection, taux de mutations) des différents 2,278 virus étudiés. Les déterminants de la variation en désordre apparaissent à 80% comme famille-dépendants et nécessitent donc, pour être étudiés, de se placer à une échelle évolutive plus réduite (Pushker et al. 2013).

Notre étude s'est donc focalisée à l'échelle d'un genre, les *Potyvirus*, dont la biologie est déjà bien documentée. Néanmoins, au sein de ce genre, l'homogénéité du désordre global obtenu pour l'ensemble des espèces étudiées nous limite donc pour discuter des raisons sous-jacentes à l'hétérogénéité du désordre chez les virus. Cette analyse pourrait donc être étendue à l'échelle intermédiaire de la famille des *Potyviridae,* qui compte, parmi ses huit genres viraux, de nombreuses espèces déjà bien décrites.

Quoi qu'il en soit, la prise en compte du désordre dans l'annotation fonctionnelle des protéomes viraux nécessite au préalable d'accumuler des données expérimentales au sein de chaque famille virale.

Dans ce contexte, l'objectif de la seconde partie de mon travail a été d'évaluer expérimentalement une des hypothèses fonctionnelles associée au désordre chez les potyvirus : son rôle dans l'adaptation.

Le désordre comme déterminant moléculaire de l'adaptation virale : vers un moyen d'appréhender les capacités adaptatives des virus à ARN ?

De façon générale, les régions désordonnées, de par la faible contrainte topologique qu'elles présentent, sont associées à une plus grande capacité à accumuler des changements en acides aminés. Dans le contexte de l'évolution d'un organisme, cette acquisition de diversité de séquence se traduit par une plus grande capacité à évoluer et potentiellement à s'adapter aux éventuelles perturbations environnementales.

Les régions désordonnées des protéines de potyvirus sont sujettes à moins de contraintes évolutives.

Notre analyse bioinformatique a permis d'étendre cette observation au genre *Potyvirus*, chez lequel les IDR sont globalement moins contraintes que les régions ordonnées. C'est le cas des protéines P1, VPg et CP qui sont directement impliquées dans des processus adaptatifs chez plusieurs espèces de potyvirus et pour lesquelles un contenu important en désordre est prédit. Ces résultats suggèrent donc un rôle moteur de certaines régions désordonnées dans les processus adaptatifs des potyvirus.

Les régions désordonnées sont le siège de l'accumulation de diversité génétique chez des populations virales évoluant, au cours de passages successifs, sur hôte naturel.

Le premier processus adaptatif concernait l'adaptation aux résistances dites "non-hôte" pour lesquelles l'incompatibilité entre le virus et l'hôte est totale et aucun isolat viral n'est capable d'infecter les différents génotypes de l'hôte. Ce mode d'adaptation nécessite au préalable l'émergence et le maintien de variants dans la population virale évoluant dans un hôte naturel (hôte réservoir). Parmi ces variants, certains peuvent constituer des formes adaptées capables de se multiplier et d'infecter de nouveaux hôtes. Leur transmission à des hôtes, leur multiplication et leur dispersion pourra donc entraîner l'émergence d'une nouvelle maladie. Dans ce mode d'adaptation, l'accumulation de diversité dans l'hôte réservoir est liée à la capacité adaptative d'un virus.

Dans ce contexte, nous avons voulu voir si le désordre constituait un déterminant de l'accumulation des mutations au sein de populations virales, au cours de l'infection sur un hôte naturel. Les résultats de données de séquençage haut-débit effectués sur les populations de 3 espèces de potyvirus au cours de 3 études indépendantes ont été analysés. Une accumulation significativement plus importante de mutations non-synonymes dans les régions désordonnées n'a été observée que pour un seul jeu de données. Les paramètres expérimentaux et notamment l'échelle évolutive à laquelle ces processus sont observés semblent interférer de façon non négligeable sur la structure

génétique des populations séquencées. Les forces antagonistes à la sélection, comme la dérive génétique et la complémentation fonctionnelle, pourraient en effet s'exprimer différentiellement selon les modes opératoires utilisés. L'analyse de l'impact des mutations apparues sur la fitness pourrait ainsi être compromise. En conséquence, l'absence significative de corrélation entre désordre et accumulation de mutations non-synonymes dans les expériences d'évolution sur hôtes ZYMV et PVY ne permettent pas de conclure quant à l'implication ou non du désordre dans la robustesse mutationnelle.
Néanmoins, on peut s'attendre à ce que la purge de variants délétères soit facilitée lors d'une évolution induite par plusieurs passages successifs *in planta*. Cette modalité expérimentale conduit vraisemblablement à une pression de sélection plus efficace. C'est le cas de l'expérience d'évolution du TEV dont l'analyse montre une différence significative d'accumulation dans les régions désordonnées, ce qui constitue un argument certain en faveur de la robustesse mutationnelle des régions désordonnées.
Des expériences complémentaires optimisées pour permettre l'identification de la valeur sélective associée aux différentes mutations apparaissant dans les populations virales seront donc à envisager pour pouvoir conclure de façon robuste quant à un déterminisme ou non du désordre intrinsèque dans l'établissement de la diversité génétique des populations virales.

De façon plus fondamentale, la démonstration de la propriété d'une protéine désordonnée à être plus robuste à l'accumulation de mutations qu'une protéine ordonnée est actuellement en cours dans notre équipe.

Ce travail basé sur l'évolution dans un hôte sensible, s'est focalisé sur l'étude de la diversité dans sa globalité. Parmi les mutations apparues, certaines mutations "innovantes" constituent les clés d'une adaptation potentielle à un nouveau contexte environnemental. Leur identification est nécessaire pour relier le désordre à la capacité adaptative sous-jacente au franchissement de la barrière d'espèces.

Le désordre de la VPg, condition sine qua none de l'adaptation du PVY à la résistance pvr2^3 médiée par eIF4E ?

La VPg des potyvirus est une protéine expérimentalement démontrée comme désordonnée (Hébrard et al. 2009; Grzela et al. 2008; Rantalainen et al. 2008) dont l'IDR centrale, conservée au sein des différentes espèces de ce genre viral, comporte les sites de contournement associés à l'adaptation aux résistances variétales médiées par le facteur eIF4E chez le PVY (Moury et al. 2004; Ayme et al. 2006; Ayme et al. 2007). Dans le cas du pathosystème PVY/piment, cette adaptation peut directement être reliée à l'interaction de la VPg avec eIF4E. La co-localisation entre désordre et sites de contournement pourrait témoigner de la corrélation entre désordre et processus d'adaptation, voire même à une causalité du désordre. Pour tester cette hypothèse, notre approche originale a consisté à construire différents mutants de la VPg, porteurs d'un niveau de désordre variable (validé expérimentalement) dans cette zone centrale.

Les capacités biologiques et adaptatives des différents mutants ont été ensuite évaluées au cours d'expériences d'évolution *in planta*, conduites parallèlement sur hôte sensible Yolo Wonder et sur hôte résistant HD285.

Les expériences conduites sur hôte sensible ont d'abord permis d'étudier l'impact des mutations introduites dans les mutants VPg sur le cycle du PVY. La VPg est une protéine multipartenaire, impliquée dans de nombreuses étapes du cycle des potyvirus. La modification de sa séquence peut donc perturber d'autres fonctions que celle sur laquelle nous basons nos observations, à savoir son interaction avec eIF4E. Les différentes mutations introduites n'affectent pas de façon significative la capacité de ces virus à infecter l'hôte sensible. Cependant une plus faible accumulation virale est détectée dans le cas du mutant le moins désordonné (OD2). L'affinité de la VPg de OD2 pour le facteur eIF4E n'est cependant pas significativement différente de celle observée pour la forme sauvage. Le lien entre la différence de capacité multiplicative du mutant OD2 et l'appauvrissement en flexibilité conformationnelle de sa VPg, reste donc à élucider. Le mutant OD2 offre ainsi un outil de choix pour l'étude du rôle du désordre de la région centrale de la VPg dans le cycle des potyvirus.

Concernant les expériences d'évolution sur hôtes résistants HD285, celles-ci ont révélé une différence adaptative significative entre chacun des trois mutants, de telle sorte que l'abondance en désordre a pu être directement corrélée aux capacités adaptatives de ces mutants.
L'analyse approfondie des évènements moléculaires associés a permis de relier l'introduction des mutations dans le mutant ID1 à sa capacité à restaurer une interaction forte avec la forme HD285 d'eIF4E. Il est à noter que la mise en place de l'infection n'est associée à l'apparition d'aucune mutation. Ce sont donc bien les mutations introduites au départ pour augmenter le désordre qui sont responsables d'une affinité moléculaire accrue pour eIF4E et d'une plus grande efficacité d'infection.
En ce qui concerne les évènements moléculaires associés aux différences d'adaptation à la résistance *pvr2*[3] du mutant ID2, le contournement est associé à l'apparition de mutations. Cependant aucun évènement mutationnel innovant n'a pu être observé par rapport à ceux détectés lors du contournement par la forme sauvage du PVY. L'interaction des VPg ID2 et sauvage avec le facteur eIF4E de HD285 semble se faire avec une affinité semblable. L'absence de contournants observés chez OD2 est également associée à une interaction avec eIF4E d'affinité semblable à celle du sauvage. A ce stade de l'étude, la conclusion qui peut être envisagée concernant les résultats obtenus pour ID2 et OD2 est que les mutations introduites pour modifier le désordre influencent plus vraisemblablement la capacité du PVY à "trouver" ou à tolérer les solutions adaptatives (*i.e* capacité adaptative), que la capacité de la VPg à interagir avec eIF4E.
Sans qu'aucun lien de causalité strict n'ait pu être encore établi, il n'en demeure pas moins que le degré de désordre est corrélé à ces différences de capacités adaptatives.

Il est important de souligner qu'afin de limiter les biais fonctionnels pouvant être associés à l'introduction de mutations dans la VPg, les sites impliqués dans le contournement des résistances préalablement identifiés n'ont pas été considérés. D'autre part, les changements introduits préexistent tous dans la variabilité naturelle au sein des VPg du PVY groupe C1. Ces sites sont polymorphiques et jusqu'à présent, n'ont jamais été reliés au contournement. Même si cela n'est pas formellement vérifié, ils ne sont pas attendus comme des déterminants cruciaux dans les évènements de reconnaissance moléculaire de la VPg (auxquels cas ils seraient conservés) et notamment de son interaction avec eIF4E (auxquels cas ils constitueraient des déterminants du contournement). Ces sites peuvent vraisemblablement être associés à l'environnement désordonné nécessaire à la fonctionnalité des motifs de reconnaissance moléculaire et déterminant donc de la structuration des complexes que forme la VPg avec ses partenaires. Le fait que la modification de ces sites n'affecte pas la capacité des virus correspondants à infecter les piments sensibles (du moins en ce qui concerne les mutants les plus désordonnés ID1 et ID2) renforce également notre postulat que les mutations introduites chez ces mutants ne constituent pas de déterminants directs de reconnaissance moléculaire.

Pour conclure, nos observations vont clairement dans le sens d'un **lien entre flexibilité de structure et capacité à accumuler des mutations adaptatives**, avec des conséquences moindres sur la stabilité et donc la fonction globale de la protéine.

Le désordre dans l'adaptation du PVY à la résistance pvr2^3 par la VPg : cas isolé ou caractéristique ubiquitaire ?

Comme le souligne le titre de cette section, il s'agit d'une étude préliminaire, conduite sur un pathosystème donné, dans un contexte de résistance donnée. Une perspective directe de ce travail réside donc dans la tentative d'étendre ces observations à d'autres pathosystèmes. Il s'agira de choisir d'autres modèles expérimentaux pour lesquels les déterminants de l'adaptation sont localisés dans des régions prédites comme désordonnées et permettant d'appréhender la généralité d'une telle observation. Chez les potyvirus, la VPg n'est pas la seule protéine possédant des déterminants du contournement dans une zone désordonnée. A titre d'exemple, nous avons pu identifier que certaines des mutations dans la protéine CI, préalablement décrites comme associées au contournement par le LMV des allèles de résistance *mo1^1* et *mo1^2* chez la laitue (Abdul-Razzak et al. 2009; Sorel 2013), sont localisées au niveau de l'IDR conservée de la région C-terminale. Le développement d'une approche similaire à celle conduite au cours de ce travail, cette fois sur le pathosystème LMV/laitue pourrait permettre d'estimer si le désordre est aussi impliqué dans l'adaptation aux résistances variétales portée par d'autres protéines du virus. Cependant afin d'appréhender le rôle éventuellement ubiquitaire du désordre dans l'adaptation, l'analyse de pathosystèmes plus éloignés est incontournable. A cet égard, chez les sobemovirus, des évènements mutationnels adaptatifs, localisés dans des régions prédites comme désordonnées de la

VPg, ont d'ores et déjà été identifiés (Hébrard et al. 2009) et constitue également de bons candidats.

Le désordre comme crible supplémentaire dans la démarche d'anticipation des évènements d'adaptation virale ?

Cette partie de mon travail de thèse s'inscrit dans une démarche globale de compréhension fondamentale des déterminants moléculaires associés à la capacité des virus à ARN à s'adapter à leur environnement. Les données obtenues vont dans le sens d'un rôle du désordre dans l'évolution virale et l'adaptation.

La démonstration aboutie d'une fonction adaptative portée par le désordre protéique chez les virus serait d'une portée très générale et de nature à ouvrir de nouveaux champs conceptuels en termes d'évolution. Pour tester cette hypothèse, les travaux de cette thèse contribuent à un front de science encore timidement illustré par la littérature. A titre d'exemple, une IDR à l'intérieur de la polymérase du *Nodamura virus* (famille *Nodaviridae*) constitue un réservoir de diversité susceptible de jouer un rôle significatif dans l'adaptation du virus à de nouveaux environnements (Gitlin et al. 2014).

D'un point de vue plus applicatif, et sous réserve bien sûr que ces observations puissent être généralisables, il peut être envisagé que la composante "désordre" soit prise en compte dans les processus permettant d'appréhender le risque d'occurrence d'évènements adaptatifs.
L'anticipation des risques d'émergence de maladies virales, ainsi que la durabilité des résistances mises en jeu notamment en agronomie, sont directement intégrés au développement des méthodes de lutte virale.

Le désordre comme moyen complémentaire d'appréhender le risque d'émergence de nouvelles maladies virales ?

L'émergence de nouvelles maladies virales se produit lors du passage d'un virus à un nouvel hôte (franchissement de la barrière d'espèces). Le challenge que représente la prédiction de ces évènements passe notamment par l'identification des déterminants et mécanismes moléculaires sous-jacents à l'émergence de formes dites "pré-adaptées" dans l'hôte dit "réservoir" (Howard & Fletcher 2012). De par l'hétérogénéité qui existe au sein des virus en termes de capacité adaptatives conduisant à de tels évènements, des études systématiques sont envisageables pour dresser l'ensemble des caractéristiques intrinsèques à chaque virus pouvant conduire à leurs spécificités adaptatives. En effet, les virus présentent une importante variabilité de gammes d'hôtes : certains sont capables d'infecter une à deux espèces (virus spécialistes) tandis que d'autres peuvent infecter plus d'une centaine d'espèces différentes, pouvant se répartir dans différents groupes taxonomiques (virus généralistes) (Elena et al. 2008). Le taux de désordre est hétérogène au sein du monde viral (Pushker et al. 2013). Une analyse *a priori* pertinente consisterait à comparer le désordre intrinsèque entre virus généralistes et spécialistes.

Le but à long terme de ces études systématiques est l'élaboration de modèles permettant l'anticipation la plus précise possible des risques d'émergence, compte-tenu de la menace que ces évènements peuvent constituer que ce soit en terme de pandémie humaine ou agronomique. Si un lien existe entre désordre et capacité à franchir la barrière d'espèce, l'optimisation de notre capacité à prédire le désordre intrinsèque contribuerait à anticiper les émergences potentielles, aussi bien chez les virus végétaux que chez leurs homologues animaux.

Le désordre a-t-il une place dans les programmes de prédictions de durabilité des résistances variétales ?

Le challenge majeur dans la lutte antivirale, et plus généralement dans la lutte contre les micro-organismes à fort potentiel évolutif, réside dans la capacité à mettre en place des moyens de lutte durables, ne pouvant pas être rapidement contournés par l'agent pathogène.

Les virus ne présentent pas tous la même capacité à s'adapter aux résistances auxquelles ils sont confrontés. De même que, pour un virus donné, certaines résistances peuvent être très rapidement contournées alors que d'autres sont, au contraire, utilisées depuis longtemps, et n'ont pas fait émerger de formes capables de s'y adapter (résistances qualifiées de durables).

L'identification et l'estimation des facteurs sous-jacents à ces spécificités adaptatives est donc primordiale dans le contexte de mise en place et de gestion de résistances aux maladies virales les plus durables possibles.

La probabilité d'émergence d'une maladie virale ou encore la durabilité d'une résistance sont multifactorielles et résultent de la combinaison de processus situés aussi bien à l'échelle moléculaire (déterminants de la diversité dans un hôte donné) qu'à l'échelle de populations entières d'individus (déterminants de la diversité à l'échelle de l'écosytème). Cette thèse n'a donc pas la prétention de défendre l'idée que le désordre pourrait être l'élément moteur majeur de l'adaptation, et par conséquent que sa prédiction soit la solution ultime à l'éradication des maladies virales dans le monde. Ce travail s'inscrit plutôt dans une démarche plus générale d'identification des caractéristiques moléculaires associées à chaque mécanisme d'adaptation.

Un objectif consisterait à prédire la durabilité d'une résistance. Des corrélations intéressantes ont d'ores et déjà été obtenues dans le cas des résistances variétales. En effet, des paramètres tels que le nombre de mutations nécessaires, et le coût en fitness associé à ces mutations semblent reliés à la durabilité des résistances (Ayme et al. 2007; Harrison 2002; Jenner et al. 2002).

D'après une étude conduite par (Janzac et al. 2009), le degré de contrainte évolutive exercée sur les facteurs de virulence a pu également être relié à la durabilité de la résistance faisant intervenir ces facteurs. Cette étude illustre bien la pertinence que peut représenter l'intégration des facteurs moléculaires dans la prédiction des durabilités, observées pourtant à très larges échelle de temps et d'espace.

L'intégration de la composante "désordre" présente l'avantage certain de pouvoir être prédit, de façon robuste, à partir d'une seule séquence, contrairement à la détermination de la contrainte évolutive, qui nécessite quant à elle de pouvoir comparer un jeu de séquences conséquent.

Conséquence du rôle du désordre dans l'adaptation : les régions désordonnées seraient-elles de mauvaises cibles pour l'élaboration des drogues antivirales ?

Compte-tenu du challenge que représente la mise au point de méthodes durables de lutte antivirale, aussi bien dans le domaine de l'agronomie que celui de la santé humaine, il apparait comme intéressant de déterminer si la prédiction de désordre peut rentrer en compte dans la sélection de cibles moléculaires. L'objectif de la lutte antivirale est de développer des résistances les plus complexes à détourner possibles. Alors qu'on peut envisager le pyramidage des gènes de résistance chez les végétaux, les croisements de génomes humains destinés à générer des individus résistants ne sont pas d'actualité. Dans ce contexte, les vaccins préventifs et les antiviraux thérapeutiques sont donc de rigueur. L'accumulation de plusieurs gènes de résistance ou plusieurs molécules antivirales ciblant différents facteurs viraux constitue en principe une démarche payante, mais le choix des cibles moléculaires reste primordial. On pourrait penser *a priori* que les protéines virales interagissant avec plusieurs partenaires sont des cibles de choix. En effet, il semble assez difficile à concevoir que la protéine virale puisse par une combinaison mutationnelle complexe restaurer l'interaction avec deux facteurs d'hôtes déficients pour le virus. Cependant, il a été souligné à plusieurs reprises que la promiscuité d'interaction siège souvent au niveau d'IDR. A y regarder de plus près, la tolérance mutationnelle dans ces régions pourrait faciliter le contournement. En conséquence, il semblerait plus rationnel de cibler des régions protéiques extrêmement contraintes, pour lesquelles le virus va avoir du mal à modifier sa séquence et donc à s'adapter. Compte-tenu de nos résultats préliminaires, et l'observation générale du taux évolutif élevé dans les IDR, les régions désordonnées ne semblent donc pas être de bonnes cibles.

Conclusion générale

Depuis le début des années 2000, la quantité croissante de fonctions cellulaires associées au désordre intrinsèque, très abondant dans le monde du vivant, n'a cessé de renforcer la validité du concept de désordre fonctionnel.

En se focalisant sur le genre *Potyvirus,* ce travail constitue la première tentative d'annotation fonctionnelle du désordre dans les protéomes de phytovirus. Les résultats de ce travail confirment la pertinence de la prise en compte du désordre intrinsèque dans la démarche d'annotation fonctionnelle des protéines. Dans la seconde partie de ce travail, l'analyse de diversité accumulée au cours de l'évolution a permis d'évaluer l'implication du désordre dans l'accumulation de la diversité virale sur hôte naturel.

Enfin, une approche expérimentale nous a permis de tester de façon originale le rôle du désordre dans un cas précis d'adaptation virale à une résistance variétale. Les résultats obtenus renforcent l'idée que le désordre constitue un déterminant central des capacités adaptatives des virus à ARN.
Ainsi, de façon similaire aux virus animaux, le désordre chez les phytovirus est associé à certaines fonctions. Mais ces fonctions sont-elles propres au désordre ?

Le désordre intrinsèque : un concept à l'épreuve des faits

L'intérêt de la communauté scientifique pour les prédictions du désordre intrinsèque est naturellement lié aux fonctions centrales que celui-ci assure. Il convient désormais d'assumer et d'intégrer la complémentarité ordre/désordre dans notre démarche d'appréhension des mécanismes biologiques impliquant les protéines. Mais la remise en question du dogme structure-fonction et sa nécessaire acceptation par la communauté scientifique s'est, à juste titre ou non, accompagnée d'une démarche de valorisation parfois exagérée de l'importance et des rôles joués par le désordre intrinsèque. Cette tendance pourrait presque faire oublier le fait que le désordre n'est pas la seule voie empruntée par l'évolution, loin de là. Pour preuve, la *Protein Data Bank*, historiquement dépositaire de l'ordre, renferme à ce jour plus de 70000 structures protéiques.

Ordre et désordre occupent-ils des espaces fonctionnels de différente nature ?

Certaines fonctions sont directement associées aux caractéristiques propres aux régions ordonnées. Tel est le cas des systèmes nécessitant une topologie atomique unique (*e.g* catalyse, complexation de métaux, reconnaissance de très haute affinité, transferts électroniques) ou encore ceux nécessitant un environnement hydrophobe (protéine membranaire).
Au contraire certaines caractéristiques, comme par exemple la capacité d'une protéine à interagir avec 84 partenaires différents, ne semblent envisageable qu'au moyen de l'importante flexibilité structurale fournie par le désordre intrinsèque.

Cependant notre étude n'a pas permis d'identifier de fonctions ou de processus exclusivement dépendant du désordre. Que ce soit dans le cas du processus de clivage protéolytique, de celui de l'adaptation ou bien de la capacité de multi-interaction, aucune binarité fonctionnelle ordre/désordre n'a pu être observée. A titre d'exemple, nos données expérimentales sont en faveur de l'hypothèse selon laquelle la région centrale de la VPg tirerait profit du désordre pour faire émerger plus facilement les mutations nécessaires à l'adaptation du PVY. Cependant, un rapide criblage d'évènements mutationnels impliqués dans l'adaptation d'autres phytovirus à d'autres espèces de plantes ne montre pas de corrélation stricte entre localisation des évènements mutationnels adaptatifs et désordre intrinsèque.

A l'instar de la stratégie de robustesse mutationnelle associée aussi bien aux protéines extrêmement compactes qu'aux protéines dépourvues de structures (Tokuriki & Tawfik 2009), l'ordre et le désordre, bien qu'opposés, occupent des espaces fonctionnels communicants. Ainsi, si ordre et désordre sont fonctionnellement liés, un challenge consisterait à identifier des répertoires fonctionnels partagés entre ordre et désordre, et à déterminer quelles sont les conditions dans lesquelles ces combinaisons apparaissent. Ce genre de questionnement pourrait notamment permettre d'interpréter les variations importantes de désordre observées chez les virus et de tenter d'identifier, si elles existent, les conditions pour lesquelles le désordre est indispensable. Il permettrait, pour un organisme donné, de mieux appréhender les coûts évolutifs associés à l'utilisation de l'ordre ou au contraire du désordre. D'après les données de prédiction haut-débit dont nous disposons, l'évolution a favorisé de façon hétérogène l'expansion du désordre (ou la diminution de l'ordre ?).

En conclusion, même si l'utilisation ubiquitaire du désordre intrinsèque est désormais compréhensible au vu des nombreuses fonctions qu'il porte, les raisons de son utilisation hétérogène parmi les virus, et plus largement dans le monde du vivant, restent à élucider et promet des découvertes passionnantes.

Matériel et Méthodes

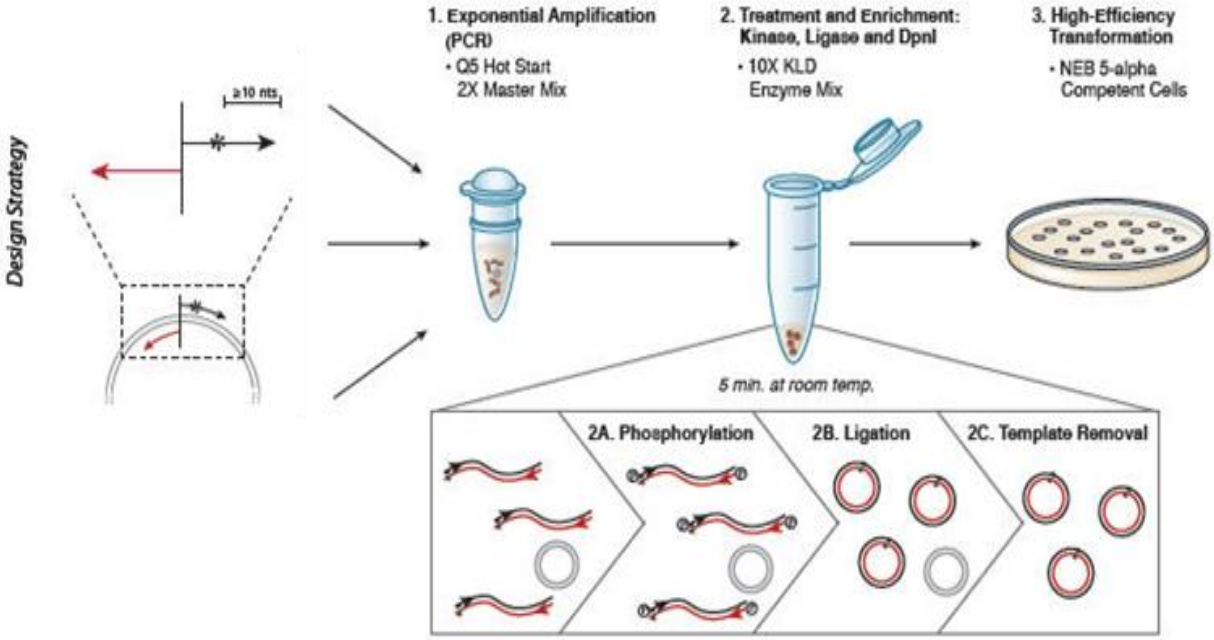


***Figure M1*|Etapes de la stratégie de mutagénèse dirigée via le kit Q5 Directed Mutagenesis (NEB™).**

1. MATERIEL BIOLOGIQUE

A) Matériel viral

L'isolat du *Potato virus Y* (PVY) utilisé dans cette étude est le PVY-SON41p. Il est issu de la diversité naturelle qui existe au sein du PVY et appartient au groupe C1, capable d'infecter le piment. Son numéro d'accession GenBank est AJ439544. **Les séquences des gènes utilisés sont fournies en Annexe 3.**

B) Matériel végétal

L'ensemble des plantes ayant servi à cette étude ont été cultivées en serre, avec une température contrôlée variant de 20 à 26°C. Suite à des problèmes de reproductibilité intervenus lors des essais conduits à Bordeaux, les expérimentations *in planta* présentées dans cette thèse ont été réalisées à l'unité de Pathologie Végétale de l'INRA d'Avignon, en étroite et aimable collaboration avec Benoît Moury et Grégory Girardot.

Deux espèces végétales ont été utilisées: *Nicotiana benthamiana* et *Capsicum annuum*. Les plantes *Nicotiana benthamiana* sont utilisées pour accumuler et multiplier le PVY afin d'obtenir des broyats foliaires suffisamment concentrés en virus pour permettre l'inoculation sur piments.

Les expérimentations d'évolution et d'adaptation sont conduites sur des piments *Capsicum annuum* de variété Yolo Wonder, porteurs de l'allèle de sensibilité au PVY $pvr2^+$ à l'état homozygote, ou de lignée HD285, porteurs de l'allèle de résistance $pvr2^3$ à l'état homozygote. Les séquences de chacun des gènes et protéines correspondant(e)s sont fournies en Annexe 3.

Yolo Wonder est une "bell pepper inbred line" ($pvr2^+/pvr2^+$) sensible au PVY.

HD285 est une lignée haploïde doublée issue de la F1 du croisement des variétés Perennial x Yolo Wonder, et sélectionné pour l'allèle de résistance $pvr2^3$ et les allèles de sensibilité aux autres loci (Caranta et al. 1997).

2. PROCEDURES EXPERIMENTALES

A) Stratégies et approches expérimentales

a. Obtention des vecteurs porteurs des mutants VPg ID1, ID2 et OD2

Conception des amorces mutagènes

Afin d'introduire les substitutions en acides aminés identifiées comme susceptibles de modifier le contenu en désordre de la zone centrale de la VPg SON41p (cf. Chapitre III), une expérience de mutagénèse dirigée à partir de la séquence de la VPg-SON41p sauvage (WT) a été réalisée. Pour cela, le kit Q5® Site-Directed Mutagenesis (NEB™) a été utilisé. Les différentes étapes sont décrites Figure M1.

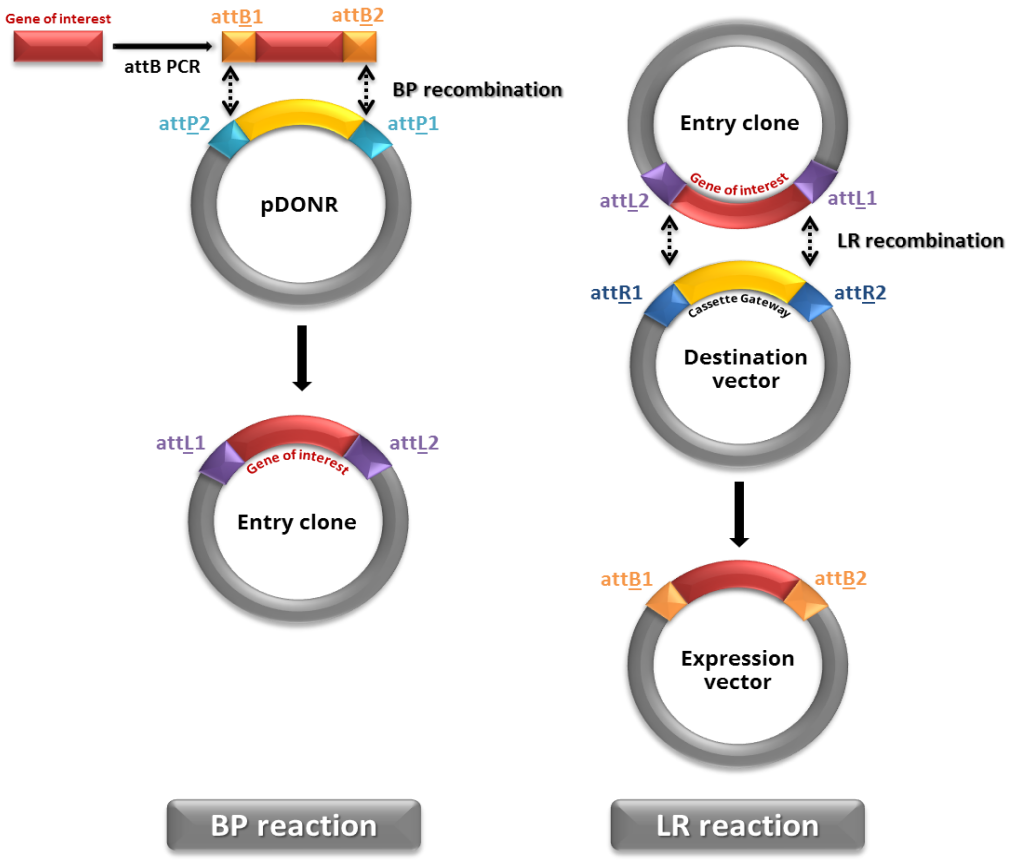


***Figure M2*|Principe de la stratégie de clonage par le système Gateway®.**

La première étape a consisté à concevoir des amorces contenant les mutations à introduire pour chacun des mutants et ce, grâce au logiciel NEBaseChanger™ (www.NEBaseChanger.neb.com). Ainsi, en fournissant la séquence d'origine ainsi que la nature des mutations à introduire et leurs positions respectives, le logiciel propose les couples d'amorces les plus adaptés, compte-tenu de leur longueur et de leur température d'hybridation (Ta) respectives. Les changements nucléotidiques introduits, ainsi que les couples d'amorces utilisées pour l'obtention de chaque mutant, sont fournies en annexe (Tableau A4.1). Le protocole détaillé utilisé pour réaliser la mutagénèse dirigée est présenté à la section « Modes opératoires ».

Technique de clonage dans le système Gateway®

Pour rentrer les séquences codantes des VPg du PVY (SON41p, ID1, ID2 et OD2) et des eIF4E de piment (pvr2^{+} et pvr2^{3}) dans des vecteurs de clonage puis d'expression, la stratégie de clonage Gateway® a été utilisée. Le principe de la stratégie par clonage Gateway® est décrit dans la Figure M2. Elle utilise des sites de recombinaison spécifiques afin d'intégrer la séquence du gène d'intérêt dans des vecteurs de clonage dits « d'entrée » (pENTR ou pDONR) (recombinaison par réaction BP). L'avantage incontestable de cette technologie est qu'elle permet ensuite, par recombinaison (réaction LR), de transférer la séquence du gène d'intérêt dans l'ensemble des vecteurs d'expression dits de « destination » (pDEST) disponibles en système Gateway®. Elle facilite ainsi grandement les étapes de clonage, par comparaison avec la méthode classique utilisant sites de restriction et endonucléases (enzymes de restriction), spécifiques à chaque vecteur. La technique Gateway® offre également la possibilité de contre-sélectionner les vecteurs vides n'ayant pas intégré le gène d'intérêt. En effet, la partie du vecteur Gateway® (pENTR ou pDEST) destiné à être remplacé par l'insert lors de la recombinaison contient le gène *ccdB* qui code pour une protéine toxique pour la bactérie. Ainsi, l'absence de recombinaison aura pour conséquence l'expression de cette protéine toxique et la mort des bactéries qui l'expriment.

Le clonage de nos gènes d'intérêt nécessite d'abord d'intégrer les sites de recombinaison de part et d'autres de leurs séquences respectives. Les amorces utilisées sont présentées en annexes (Tableau A4.2). Le protocole de PCR utilisé est présenté dans la partie "PCR attB" de la section Mode opératoire.

b. Obtention des vecteurs codant pour les formes pvr2^{+} et pvr2^{3} de eIF4E

Nécessité de clonage alternatif dans le vecteur d'expression pETGB1a

Des problèmes au niveau de l'expression des protéines de fusions His^{6x}-eIF4E codées par les vecteurs Gateway® pDEST17 notamment la formation importante en corps d'inclusions de nos protéines d'intérêt nous ont contraint à tester d'autres vecteurs d'expression.

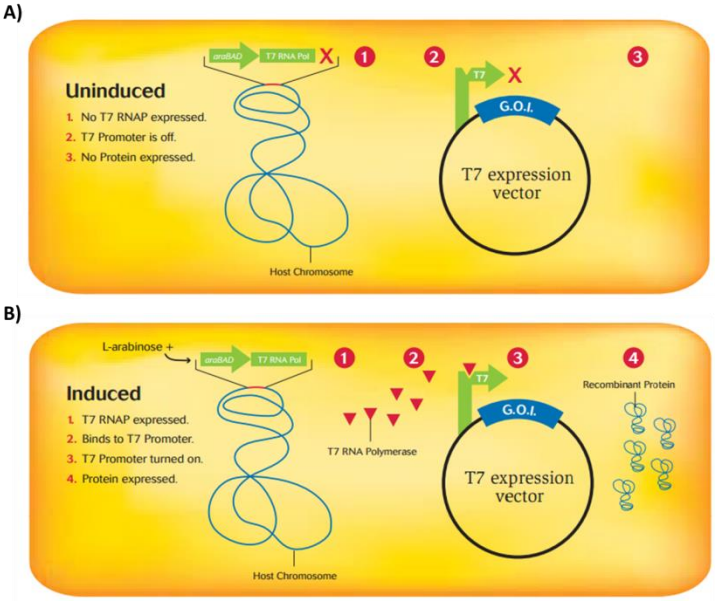


***Figure M3*|Mécanismes de régulation de l'expression des protéines sous contrôle du promoteur T7 chez les souches *E.coli* BL21-AI™. A)** Avant induction, en l'absence d'arabinose, le gène codant pour la polymérase à ARN T7 est réprimé. L'absence de la polymérase T7 empêche l'activation de l'expression des gènes d'intérêt sous contrôle du promoteur T7. **B)** L'ajout de L-arabinose (+) lors de l'induction lève la répression de l'expression de la polymérase T7 par activation du promoteur araB. L'expression de la polymérase T7 va permettre l'activation du promoteur T7 et donc l'expression du gène d'intérêt.

Ces souches permettent une régulation plus fine de l'induction de l'expression par le promoteur T7. Le risque de fuite d'expression des protéines d'intérêt en l'absence d'inducteur pendant la phase de croissance est réduit et assure une meilleure croissance des bactéries avant induction dans le cas de protéines d'intérêt toxiques. *Adapté de « Competent Cells Brochure », www.lifetechnologies.com. G.O.I : gène d'intérêt ; T7RNAP : Polymérase ARN T7.*

La production de quantités suffisantes de protéines eIF4E porteuses de l'étiquette His^{6x} en phase soluble a nécessité l'utilisation d'un vecteur d'expression différent, le pETGB1a, optimisé pour favoriser le repliement de protéines instables dans *E.coli*. Ce vecteur n'étant pas compatible avec la technologie Gateway®, une technologie de clonage classique par restriction-ligation a dû être utilisée (cf. section Modes opératoires).

Les deux sites de clonage NcoI et BamHI présents sur le plasmide pETGB1a ont été utilisés (cf. Annexe 5). Des amorces PCR contenant chacun des sites à leur extrémités 5' ont été élaborées afin de flanquer les sites NcoI et BamHI respectivement en 5' et 3' de la séquence codante des gènes *eIF4E* de piment (cf. Annexes - Figure A4.2). Les conditions de PCR utilisées, ainsi que les étapes de traitement à l'enzyme DpnI et la purification sont identiques au mode opératoire décrit dans la section PCR attB. Le protocole utilisé pour l'insertion des gènes d'intérêt dans ces vecteurs est décrit dans la section "Clonages".

c. *Production de protéines recombinantes pour les mesures physico-chimiques in vitro.*

Les mesures de dichroïsme circulaire sur les différentes VPg et la détermination des paramètres thermodynamiques d'interaction VPg-eIF4E nécessitaient des quantités de protéines purifiées actives de l'ordre du milligramme. Pour cela, ses protéines ont été exprimées et purifiées à partir de systèmes hétérologues dans *E. coli*. Au cours d'essais préliminaires, nous avons observé que l'expression des protéines VPg était associée à un ralentissement significatif de la croissance de la souche standard d'expression d'*E. coli* BL21. La souche E.coli BL21-AI™ Chemically Competent (ThermoFisher™) de génotype [F-ompT hsdSB (rB- mB-) gal dcm araB::T7RNAP-tetA] qui permet l'expression de protéines toxiques pour la cellule a donc été utilisée (cf. Figure M3 pour détails). Cette souche est de plus dépourvue des protéases OmpT et Lon souvent responsables de dégradations protéolitiques des protéines hétérologues lors de leur extraction. Les protocoles d'expression et de purification sont décrits dans la section "Modes opératoires -Purification de protéines recombinantes".

d. *Purification des protéines recombinantes*

Les protéines recombinantes ont été purifiées à partir de la totalité des protéines bactériennes solubles par chromatographie d'affinité sur cations métalliques (IMAC). Le support chromatographique est constitué d'ions Ni^{2+} complexés à de l'acide nitroacétique sur des billes d'agarose (technologie Ni-NTA Protino® Machery Nagel). Les étiquettes histidine fusionnées en N-ter des protéines d'intérêt se fixent aux ions Ni^{2+} sur la résine d'agarose. Après lavage, elles sont éluées par déplacement compétitif avec de l'imidazole, un analogue de l'histidine.

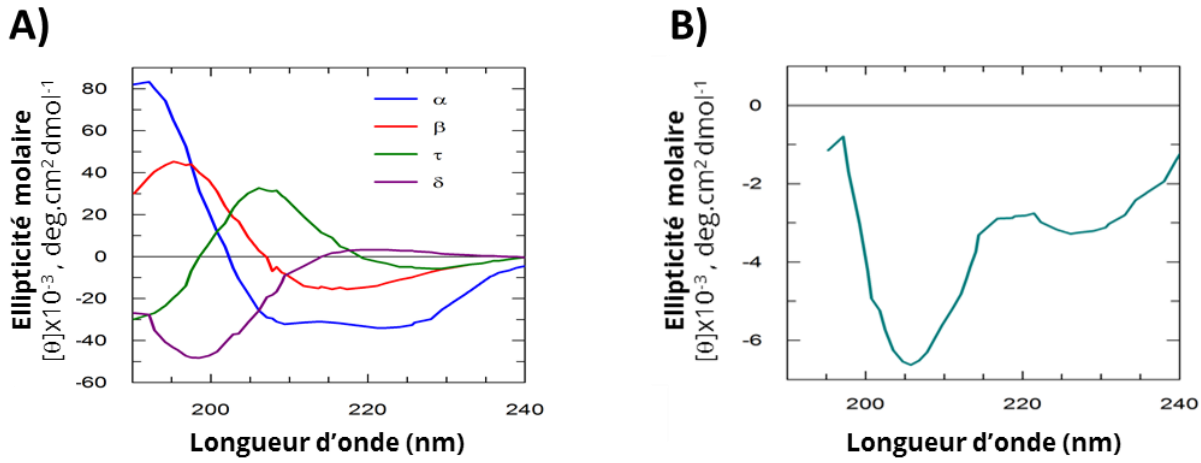


***Figure M4*|Représentation des spectres d'ellipticité molaire obtenus en dichroïsme circulaire correspondant à : A)** chaque structure secondaire (hélice α, feuillet β, coude τ et désordonné δ) ; **B)** la VPg du PVY SON41p.

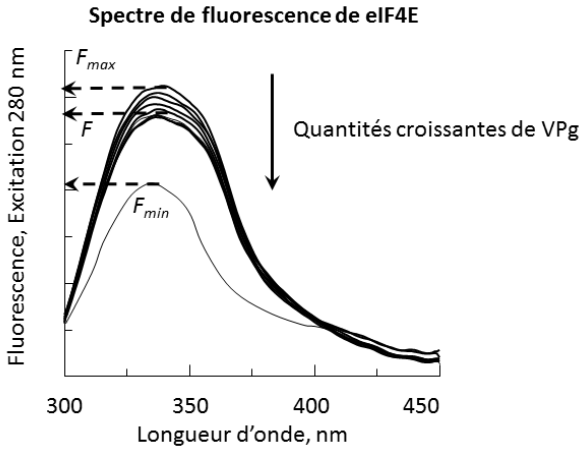


***Figure M5*|Spectres de fluorescence de la protéine eIF4E obtenus pour différentes quantités de VPg.** Après excitation à 280nm, le signal de fluorescence est enregistré à plusieurs longueurs d'ondes. Ce signal décroit à mesure que la concentration en VPg augmente. F_{max} : signal maximum de fluorescence enregistré à 340nm. Correspond à la situation sans VPg. F_{min} : signal minimum de fluorescence enregistré à 340nm.

e. Dichroïsme circulaire et protéines désordonnées

Le dichroïsme circulaire (CD) dans les régions de l'UV lointain constitue une des techniques les plus utilisées pour caractériser la contribution des différentes structures secondaires (les hélices alpha (α), les brins bêta (β), les coudes β (τ) et les polypeptides non ordonnés (δ)) à la structure d'une protéine. Très simplement, un échantillon protéique est excité par une lumière polarisée. En fonction des différentes contributions de sa structure secondaire, la protéine n'absorbera pas les mêmes quantités d'énergie des composantes droite et gauche de cette lumière. Le spectre est le reflet des rapports des intensités droite et gauche absorbées en fonction de la longueur d'onde de la lumière incidente. L'ellipticité molaire θ observée à une longueur d'onde λ est donc une combinaison linéaire, somme de la contribution des différents éléments de structures secondaires :

$$\theta_\lambda = \alpha\theta_\lambda^\alpha + \beta\theta_\lambda^\beta + \tau\theta_\lambda^\tau + \delta\theta_\lambda^\delta \text{ avec } \alpha + \beta + \tau + \delta = 1$$

Les régions ou protéines désordonnées, qui par définition, ne possèdent pas ou peu d'éléments structuraux stables, vont donc présenter en plus large proportion le signal CD correspondant au « polypeptide non ordonné », et qui se traduit par un minimum très négatif aux environs de 205nm et un épaulement aux alentours de 220nm (Woody 2010)**.** Le rapport des signaux de dichroïsme enregistrés respectivement à 205nm et 220nm permet d'estimer le contenu relatif en désordre d'une protéine (cf. Figure M4).

f. Détermination des constantes thermodynamiques de dissociation du complexe VPg/eIF4E par spectroscopie de fluorescence

Les résidus tryptophane ont un maximum d'absorption dans l'ultraviolet, centré autour de 280 nm. Une partie de la lumière absorbée est réémise sous forme radiative (fluorescence). La longueur d'onde de cette lumière émise est très dépendante de l'environnement du tryptophane. La fluorescence est donc un excellent senseur des modifications structurales que peut subir la protéine au contact d'un ligand par exemple.

Les facteurs eIF4E YW et HD contiennent tous deux 9 résidus tryptophanes et émettent donc un signal de fluorescence dont le maximum est centré à 340nm lorsque excités à 280nm. La formation du complexe eIF4E/VPg s'accompagne d'une diminution de la fluorescence de certains des tryptophanes d'eIF4E. Cette diminution est proportionnelle à la quantité de complexe formé. L'enregistrement des variations de fluorescence d'eIF4E associées à l'ajout progressif de petites quantités de VPg permet de titrer la formation du complexe eIF4E/VPg. On a ainsi accès à la constante de dissociation K_d du complexe, un paramètre thermodynamique traduisant l'affinité des deux macromolécules (Michon et al. 2006; Roudet-Tavert et al. 2007). La figure M5 illustre la diminution de fluorescence d'eIF4E lorsqu'elle est mélangée à des quantités croissantes en protéine VPg.

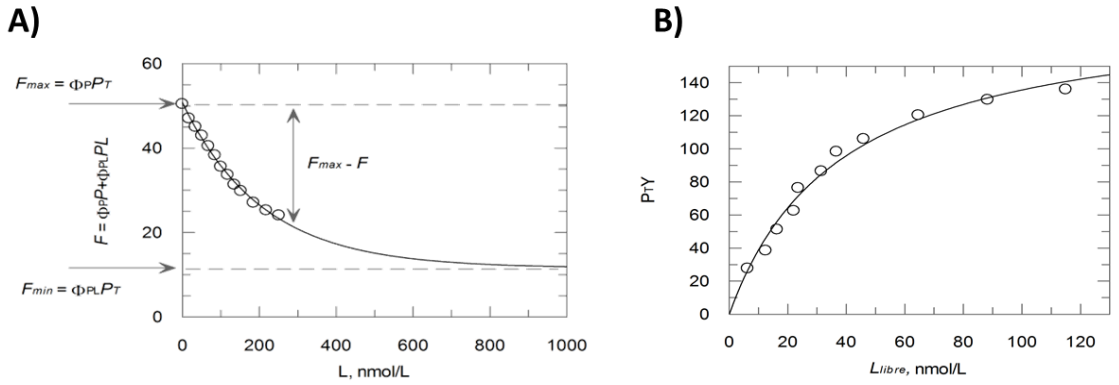


***Figure M6*|Spectre de fluorescence de la protéine P au cours de l'ajout d'un ligand L. A)** avant transformation ; **B)** après transformation.

Traitement du signal

Le modèle minimum d'interaction binaire de la VPg (*L*) avec eIF4E (*P*) peut se formaliser de la manière suivante :

$$P + L \overset{K_D}{\rightleftharpoons} PL \tag{1}$$

$$P_{tot} = P + PL \tag{2}$$

$$PL = P_{tot}\frac{L_{libre}}{K_D + L_{libre}} = P_{tot}\overline{Y}\ ^{1} \tag{3}$$

Dans le cas qui nous intéresse, l'interaction PL est associée à une diminution de la fluorescence de P (eIF4E) (cf. figure M6.A)(Φ : coefficients molaires de fluorescence).

$$F_{max} = \Phi_P P_{tot} \tag{3}$$

$$F = \Phi_P P + \Phi_{PL} PL \tag{4}$$

$$F_{min} = \Phi_{PL} P_{tot} \tag{5}$$

Les paramètres Φ_P et Φ_{PL} sont les coefficients de fluorescence molaires spécifiques à *P* et au complexe *PL*.

$$F_{max} - F = \Phi_P P_{tot} - (\Phi_P P + \Phi_{PL} PL) \tag{6}$$

$$F_{max} - F = (\Phi_P - \Phi_{PL})PL \tag{7}$$

$$F_{max} - F_{min} = (\Phi_P - \Phi_{PL})P_{tot} \tag{8}$$

$$\frac{PL}{P_{tot}} = \frac{(F_{max} - F)}{(F_{max} - F_{min})} = \overline{Y} \tag{9}$$

En combinant les équations (3) et (9) on voit qu'il est en principe possible de déterminer K_d à partir de l'expérience (cf. Figure M6.B).

$$\frac{(F_{max} - F)}{(F_{max} - F_{min})} = \frac{L_{libre}}{K_d + L_{libre}} \tag{10}$$

Néanmoins il subsiste deux inconnus : d'une part, la connaissance de F_{min} est nécessaire, mais en pratique il est difficile d'obtenir sa valeur par l'expérience. Dans notre cas, elle serait obtenue pour des concentrations très élevées en VPg, non compatibles avec son domaine de solubilité. D'autre part, on connait la concentration totale en VPg (L_{tot}) ajoutée mais pas la concentration en VPg (L_{libre}) non complexée à eIF4E.

[1] *$\overline{Y}$ correspond à la fonction de saturation. Elle est définie comme le nombre moyen de sites de fixations occupés par le ligand par rapport au nombre total de sites de fixation. Les valeurs de $\overline{Y}$ sont donc comprises entre 0 et 1.*

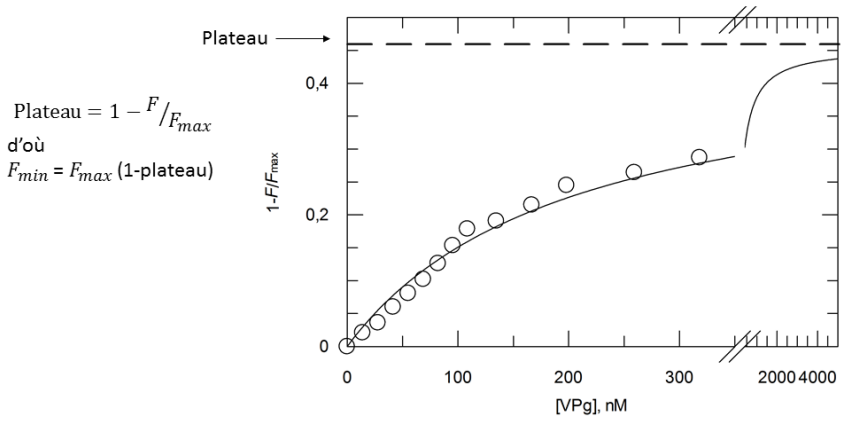


***Figure M7*|Détermination du plateau théorique traduisant la saturation du complexe VPg/eIF4E en solution.** La valeur du plateau correspond à la valeur maximum de 1-F/F_{max}, calculée par régression non linéaire des mesures de fluorescence enregistrées après ajout de 0 à 325nM de VPg dans la solution initiale d'eIF4E. Le F_{min} peut ainsi être déduit.

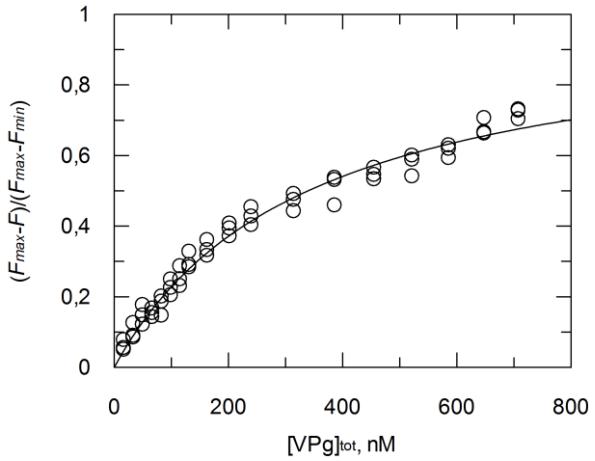


***Figure M8*|Détermination par régression non-linéaire de la valeur de $\overline{Y}$.** $\overline{Y}$ correspond à la valeur maximum (à saturation) que peut atteindre (F_{max}-F)/(F_{max}-F_{min}), et est calculée à partir des données de fluorescence expérimentales.

On procède donc en deux étapes:

ETAPE 1

Tout d'abord, on représente :

$$1 - {}^{F}/_{F_{max}} = f(L_{tot}) \qquad (11)$$

L'équation (12) est ajustée sur les données expérimentales par régression non linéaire (Figure M7). Cela permet d'estimer de manière robuste le paramètre « *Plateau* ».
Notons que la valeur de K_d obtenue n'a pas de sens physique puisque c'est la concentration totale en VPg qui est ici utilisée.

$$1 - {}^{F}/_{F_{max}} = Plateau \times \frac{L_{tot}}{K_D + L_{tot}} \qquad (12)$$

F_{min} peut alors être déterminé à partir de la valeur du plateau (cf. Figure M7).

ETAPE 2

Il est alors possible de déterminer une valeur robuste de K_d en portant :

$$\frac{(F_{max} - F)}{(F_{max} - F_{min})} = f(L_{tot}) \qquad (13)$$

K_d est obtenu en ajustant les paramètres de l'équation (14) sur les données expérimentales. Y_{sat} est la valeur de $(F_{max}-F)/(F_{max}-F_{min})$ à saturation. Plus elle sera proche de 1 plus la détermination de K_d sera robuste (cf. Figure M8).

$$\frac{(F_{max} - F)}{(F_{max} - F_{min})} = \sqrt{(K_d + [VPg]_{tot} + Y_{sat})^2 - 4[VPg]_{tot}Y_{sat}} \qquad (14)$$

g. Suivi d'infection des piments Yolo Wonder et HD285 par les PVY SON41p WT et porteurs des VPg mutantes

Obtention des clones PVY infectieux porteurs des VPg SON41p, ID1, ID2 et OD2

Les différents clones infectieux de PVY sont obtenus à partir de la méthode de recombinaison dans la levure, qui consiste à intégrer un fragment d'ADN dans un vecteur en s'affranchissant de l'étape de ligation. Ainsi la co-transformation dans la levure d'un produit d'amplification obtenu par PCR et d'un plasmide délété d'une partie de la région correspondant à l'amplicon PCR et préalablement linéarisé à cet endroit, conduit à une recombinaison au niveau de la zone de chevauchement dite homologue. Les séquences des VPg mutantes et WT ont été recombinées avec le plasmide SON41p-pCAPM13-Trp2micronΔVPg-CD.

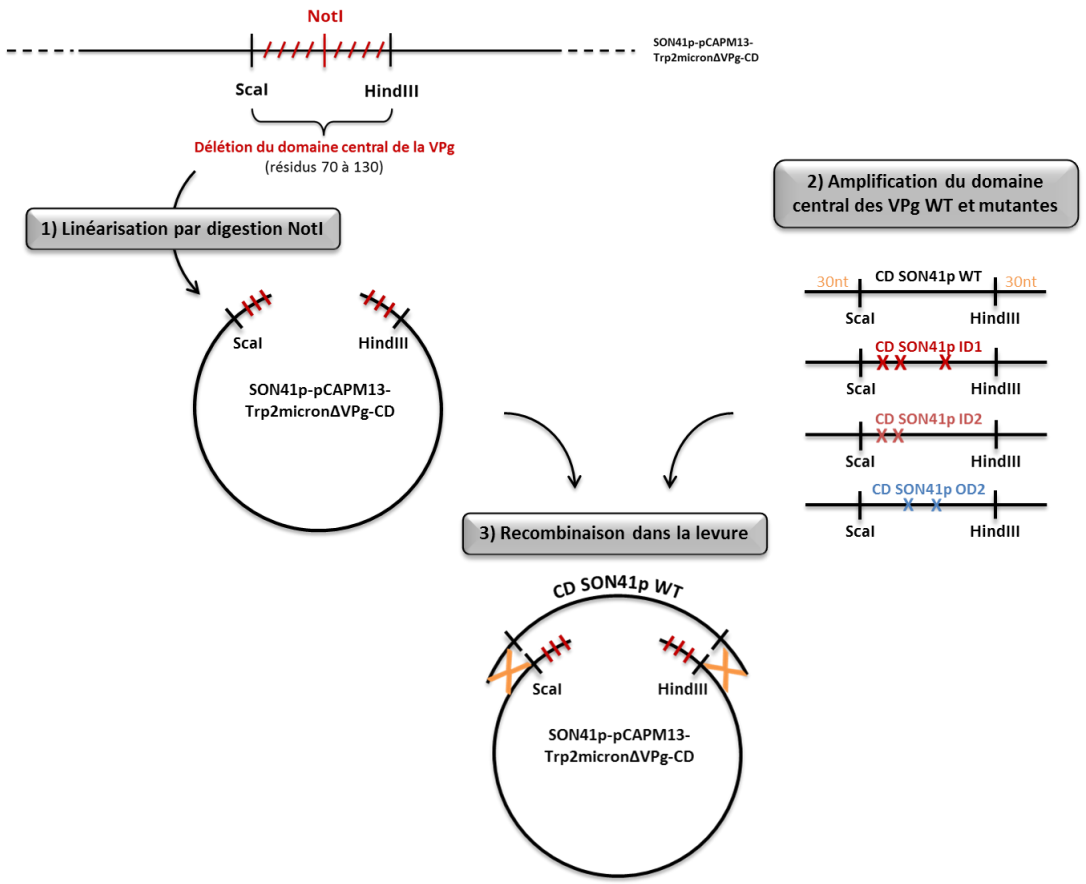


***Figure M9*|Obtention des clones infectieux complets PVY -SON41p WT et mutants ID1, ID2 et OD2 par stratégie de recombinaison dans la levure.**

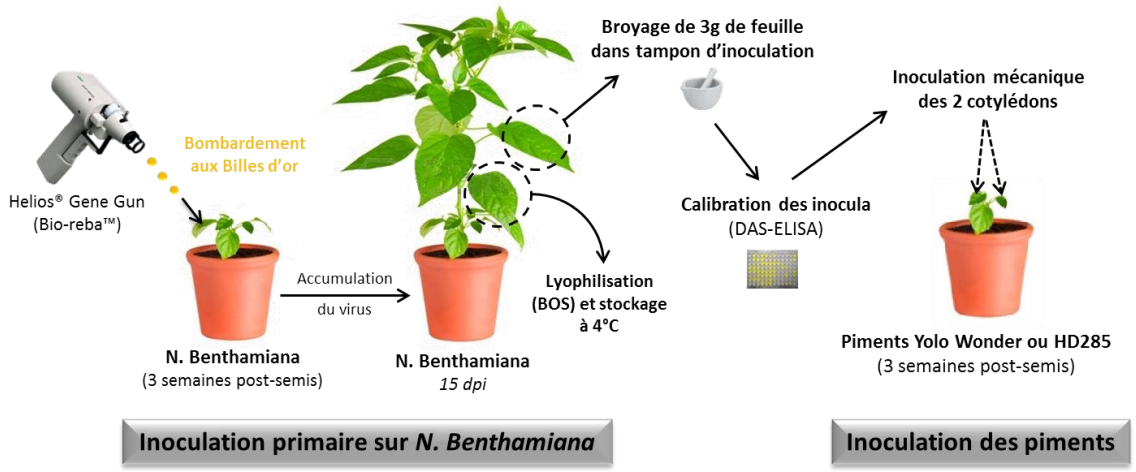


***Figure M10*|Stratégie d'inoculation du PVY dans les piments *Capsicum annuum* Yolo Wonder et HD2785.**

Ce plasmide porte le génome complet du PVY mais est délété au niveau de la région centrale de la VPg (entre les sites ScaI et HindIII), elle-même remplacée par un site de restriction NotI (cf. Figure M9) qui va permettre la linéarisation du vecteur dans la zone où l'on souhaite insérer les fragments VPg. Les régions centrales des VPg SON41p WT, ID1, ID2 et WT sont amplifiées par PCR (en suivant le protocole de la PCR attB), avec des amorces propres au domaine central (cf. Annexes A4.3) de telle façon qu'il y ait environ 30 nucléotides de part et d'autre du domaine central.

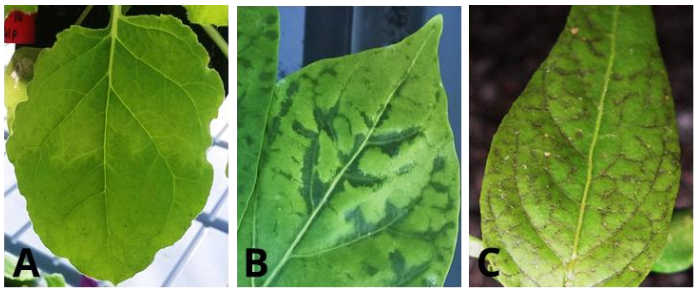
Inoculation des constructions virales dans les piments Yolo Wonder et HD285

L'infection virale d'une plante peut être recréée artificiellement après avoir fait pénétrer un ADN plasmidique viral dans les cellules végétales. Le plasmide contenant le génome du PVY (isolat SON41p) est cloné sous forme d'ADNc, sous contrôle d'un promoteur de transcription 35S et d'un terminateur de transcription NOS, est inoculé par bombardement biolistique. Cette méthode d'inoculation consiste à projeter sous pression d'hélium des billes d'or enrobées d'ADN plasmidique. Cette technologie ainsi que le protocole de préparation du matériel sont fournis par Bio-Rad™. Suite au bombardement, l'ADNc pénètre dans les cellules végétales et est alors transcrit en ARN viral, ce qui donne lieu à la traduction des protéines virales et à l'infection de la plante.

- **Inoculation primaire du PVY dans *Nicotiana benthamiana* par méthode biolistique**

Dans le cas du piment *Capsicum annuum*, des études précédentes ont d'ores et déjà mis en évidence qu'il n'était pas possible de passer directement par une méthode biolistique pour inoculer cette plante (Moury et al. 2004). Nous avons donc été contraint de passer par une étape intermédiaire d'inoculation biolistique sur *Nicotiana benthamiana* afin de pouvoir récolter une charge virale qui sera ensuite transférée, lors d'une seconde inoculation, sur les plants de piments (cf. Figure M10).

Lors de l'inoculation des plants de *Nicotiana benthamiana*, 3 feuilles étalées par plante sont ainsi tirées en biolistique avec chacun des clones infectieux de PVY d'isolat SON41p, porteurs respectivement de la VPg WT, ID1, ID2 et OD2. Une fois les symptômes apparus, les feuilles apicales les plus symptomatiques sont récoltées, lyophilisées (stocks appelés BOS) et stockées à 4°C pour pouvoir ensuite être réutilisées lors des expérimentations sur piments.



A) Nicotiana Benthamiana infectées avec PVY-SON41p

B) Capsicum annuum (Yolo Wonder) infecté avec PVY-SON41p

C) Capsicum annuum (HD285) infecté avec PVY-SON41p

***Figure M11*|Symptômes de l'infection induite par l'isolat SON41p du *Potato virus Y* (PVY) chez** : **A)** N. benthamiana ; **B)** *C. annuum* Yolo Wonder ; **C)** *C. annuum* HD285.

- **Accumulation de la charge virale dans l'hôte *Nicotiana benthamiana* et inoculation de la plante modèle *Capsicum annuum***

La charge virale contenue dans ces BOS n'est cependant pas suffisante pour infecter directement les piments. Une inoculation mécanique intermédiaire sur *N. benthamiana* est nécessaire avant chaque nouvelle expérimentation, afin d'augmenter la charge virale des inocula qui, seulement ensuite, pourront être utilisés pour inoculer les piments. Afin de calibrer les inocula et de s'assurer de l'absence de biais trop important en ce qui concerne la concentration en virus dans chacun des inocula, un test de DAS-ELISA est réalisé (une présentation de cette technique est disponible section suivante). Les volumes des inoculas sont ajustés avec du tampon de broyage si nécessaire (cf. Figure M10). Une fois les inocula calibrés, les inoculations mécaniques sont réalisées au niveau de chaque cotylédon de piments *Capsicum annuum* âgés de trois semaines post-semis. Des contrôles négatifs sont également réalisés en inoculant 3 piments de chaque variété Yolo Wonder et HD285 avec du broyat de feuilles de *Nicotiana benthamiana* saines afin de pouvoir discriminer les symptômes apparus suite à la présence de virus de ceux dus aux éventuels stress de culture pouvant survenir au cours de l'expérimentation.

Suivi d'infection des différents clones infectieux de PVY sur Nicotiana benthamiana et piments Yolo Wonder et HD285

Dans ce travail, l'infection par le PVY des piments Yolo Wonder et HD285 est déterminée respectivement à 30 et 36 jours post-inoculation. L'infection peut se caractériser à la fois par la présence de symptômes et par une accumulation en particules virales significative. En ce qui concerne les infections intermédiaires sur *Nicotiana benthamiana*, le prélèvement des feuilles n'est réalisé qu'une fois les symptômes apparus.

- **Symptomatologie**

Les plants de *Nicotiana benthamiana* infectés par le PVY-SON41p présentent des symptômes de type mosaïque et "leaf curling" dès 7 jours post-inoculation (Figure M11.A), dans les conditions de cultures décrites ci-dessus.

Les piments de variété Yolo Wonder (YW) infectés par l'isolat SON41p présentent des symptômes de type mosaïque détectables de 10 à 14 jours post-inoculation, au niveau des feuilles apicales (non-inoculées), dans les conditions de cultures énoncées ci-dessus (Figure M11.B).

Les piments de variété HD285 infectés suite au contournement de l'isolat SON41p (représentant en moyenne 40% des plantes inoculées au total) présentent des symptômes plus tardifs, apparaissant en moyenne entre 21 et 68 jours post-inoculation et se traduisant d'abord par l'apparition de mosaïques. Apparaissent ensuite des symptômes de type nécrose, tout d'abord au niveau des nervures foliaires (Figure M11.C), qui s'étendent sur l'ensemble de la feuille et de la plante, et conduisent à la mort de celles-ci environ 50 jours post-inoculation.

1. Coating: Sensibilisation des microplaques avec un anticorps spécifique
Diluer au 1/1000 les anticorps de coating (IgG) dans le tampon de coating ; i.e. 20 µl IgG dans 20 ml de tampon. Ajouter 200 µl de la solution obtenue dans chaque puit de la microplaque.

Couvrir les plaques et placer les dans une étuve humide.

Incuber 4 h à 30°C ou une nuit à +4°C.

1a. Lavage : Vider les plaques et laver les au moins 3 fois avec le tampon de lavage. Eliminer tout le liquide en tapant les plaques sur du papier absorbant. Le laveur de microplaque BIOREBA «EASY WASH 2000» permet un lavage optimum.

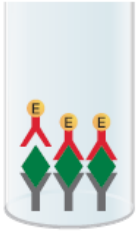
2. Echantillon: Incubation de l'extrait de plante (antigène)
Diluer l'extrait végétal au 1/20 dans le tampon d'extraction (la nature du tampon d'extraction ainsi que le taux de dilution peut varier selon les échantillons; se reporter à la fiche «information produit» du réactif).

Ajouter 200 µl par puit.

Couvrir les plaques et placer les dans une étuve humide.

Incuber une nuit à +4°C.

2a. Laver comme en 1a.



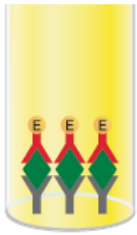
3. Conjugué: Incubation avec les anticorps conjugués à l'enzyme (conjugate)
Diluer au 1/1000 les anticorps conjugués dans le tampon de conjugué.

Ajouter 200 µl de la solution obtenue par puit.

Couvrir les plaques et placer les dans une étuve humide.

Incuber 5 h à 30°C.

3a. Laver comme en 1a.



4. Substrat: Une réaction colorée indique la présence d'un échantillon infecté
Diluer les pastilles de substrat pNPP à 1 mg/ml dans le tampon de substrat.

Ajouter 200 µl de la solution obtenue par puit.

Incuber à température ambiante (18-25°C) à l'obscurité.

Observez la réaction et notez le développement de la coloration entre 30 et 120 min, soit visuellement soit avec un photomètre à 405 nm. Les photomètres à double lecture 405/492 nm permettent de réduire les effets de bruits de fond.

***Figure M12*|Principe du test DAS-ELISA.** *Adapté de "Mode Opératoire du test DAS-ELISA" (disponible sur www.bioreba.com).*

- **Détermination de l'accumulation virale chez les plantes inoculées non symptomatiques**

Pour les plantes HD285 ne présentant pas de symptômes suite à l'inoculation par SON41p WT et les trois mutants, plusieurs feuilles sont prélevées et broyées afin de déterminer leur charge virale par la technique de DAS-ELISA, dont le principe est décrit Figure M12. La première étape du test de DAS-ELISA dite du « coating » consiste à fixer l'anticorps dirigé contre la capside du *Potato virus Y* au fond des puits de plaques ELISA 96 puits. Des lavages successifs sont ensuite réalisés, puis un second anticorps, couplé à l'enzyme phosphatase alcaline, est incubé sur la plaque. La révélation colorimétrique est ensuite réalisée par ajout du substrat de l'enzyme. Cette technique permet donc indirectement la détermination de la quantité de virion initialement présent dans le broyat.

h. *Séquençage des génomes viraux*

Afin de déterminer la stabilité des mutations introduites dans nos 3 mutants ou encore d'identifier les mutations associées au contournement par le PVY, les ARN viraux de certaines plantes sont extraits selon le protocole décrit dans la section "MODES OPERATOIRES - Extraction des ARN viraux". En ce qui concerne l'infection sur hôte sensible Yolo Wonder, les ARN viraux extraits sont séquencés au niveau de la région codant pour la VPg (cf. section MODES OPERATOIRES). Pour ce qui est de l'expérience de contournement de l'isolat SON41p sur l'hôte HD285, pour 3 des 18 échantillons prélevés, un séquençage sur génome complet a été réalisé. La méthode d'amplification des génomes entiers ainsi que les différentes amorces ayant servi pour le séquençage sont présentées dans la section "MODE OPERATOIRES - Séquençage des génomes viraux".

B) Modes opératoires

a. Techniques de biologie moléculaire

Amplification d'acides nucléiques

- **PCR mutagène**

La PCR mutagène est réalisée en suivant les instructions du kit Q5® Site-Directed Mutagenesis (NEB™). 12,5µl de Master Mix 2X sont mélangés avec 1,25µl de chaque primer de concentration 10µM, 1µl de plasmide matrice à 25ng/µl et 9µl d'eau milliQ. Le cycle de PCR est le suivant : une dénaturation initiale est réalisée à 98°c pendant 30s, puis 25 cycles sont effectués, comprenant chacun une étape de dénaturation à 98°C pendant 10s, une étape d'hybridation de 30s dont la température dépend du couple d'amorces utilisé (cf. Annexes A4.1), et une étape d'élongation de 30s à 72°C. Enfin une élongation finale de 72°C pendant 2min termine la réaction d'amplification.

La seconde étape consiste à réaliser une réaction KLD (Kinase, Ligase et DpnI) permettant i) de rabouter les extrémités des brins amplifiés via l'ajout de groupement phosphate ii) de faire la ligation des brins raboutés et, iii) d'éliminer l'ADN matrice non muté via l'utilisation de l'enzyme DpnI, endonucléase qui reconnait et clive spécifiquement l'ADN au niveau des sites de méthylation Gm6A^TC (dans notre cas uniquement situés sur le plasmide ayant servi de matrice). Pour cela, 1µl de produit PCR est mélangé avec 5µl de Tampon KLD 2X, 1µl de mix enzymatique KLD et 3µl d'eau milliQ. La réaction se fait pendant 5min à température ambiante.

- **PCR attB**

Dans le cas où les gènes sont manipulés dans le système Gateway®, la première étape consiste à introduire les sites de recombinaison attB aux abords de la séquence des gènes d'intérêts. Les PCR sont réalisées selon les recommandations du kit de la *Taq Phusion® High-fidelity polymerase* (NEB™, 2U/µl). Les conditions sont les suivantes : 1µl de matrice, concentration 25ng/µl, 1µl dNTP 10mM, 1µl d'amorce attB1 et 1µl d'amorce attB2 à 25µM, 10µl de tampon HF, 1µl de Phusion DNA polymerase, dans un volume réactionnel total de 50µl. Les cycles d'amplification sont ensuite réalisés suivant le protocole « 2-step » de la Phusion, en raison de la longueur importante des amorces et donc de leur température d'hybridation élevée: la dénaturation initiale à 98°C pendant 1min est suivie de 35 cycles contenant chacun une étape de dénaturation à 98°C de 10s, puis une étape d'hybridation et d'élongation à 72°C pendant 1min. Une élongation finale est ensuite réalisée à 72°C pendant 10min. Afin d'éliminer l'ADN plasmidique ayant servi de matrice pour l'amplification, les produits PCR sont traités avec 1µl d'enzyme DpnI (ThermoFisher™, 10U/µl) additionnés de 5µl de tampon Tango et incubés pendant 15min à 37°C suivie de 15min à 65°C pour inactiver l'enzyme.

- **PCR sur colonies bactériennes**

Après clonage en vecteur d'expression pETGB1, 30 clones par construction sont testés par PCR dans les conditions suivantes (volumes donnés pour une réaction de PCR) : 0,25µl de Taq DNA polymerase (NEB®, 5U/µl) sont mélangés à 1,5µl de tampon 10X, 1,5µ de BSA, 0,2µl de dNTP (10µM), 1µl de chaque amorce dans 15µl de volume total. Tous les clones prélevés sont à la fois introduits dans le mix de PCR et repiqués sur boîte de Pétri. Le cycle d'amplification suivant est ensuite réalisé : 3min de dénaturation initiale à 94°C, suivi de 30 cycles comprenant une première étape de dénaturation à 94°C pendant 30s, puis une étape d'hybridation des amorces à 55°C pendant 30s et une étape d'élongation à 72°C pendant 1min. Enfin une élongation finale est réalisée pendant 5min à 72°C.

- **PCR sur colonies de levures**

Des colonies sont prélevées sur boîtes de culture et resuspendues dans 100µl d'eau stérile. 5µl sont ajouté au mix suivant : 0,3µl de Taq DNA polymerase (NEB®, 5U/µL), 5µl du tampon Taq, 0,3µl de dNTP 10mM, 1µl de chaque amorce Fw et Rev (cf. Annexes A4.3), complété à 45µl par de l'eau milliQ. Le cycle d'amplification par PCR est identique à celui utilisé lors des PCR sur colonies bactériennes.

- **Amplification des génomes viraux par RT-PCR**

Deux stratégies de reverse-transcription et PCR découplées ont été utilisées en fonction de la quantité initiale de virus dans les échantillons.

Synthèse d'ADNc avec l'enzyme ReverTAID (ThermoScientific™). Un premier mix réactionnel est réalisé contenant 3µl d'ARN totaux, 1µl d'amorce oligodT(18) à 100µM et ajusté à 12,5µl avec de l'eau milliQ. Ce mélange est incubé 5min à 65°C puis placé dans la glace. Sont ensuite ajoutés au mélange réactionnel : 4µl de tampon 5X RT fourni avec la ReverTAID, 0,5µl d'eau milliQ, 2µl de dNTP à 10mM et 1µl de reverse-transcriptase ReverTAID. Les tubes sont placés dans un thermocycleur à 42°C pendant 1h puis à 70°C pendant 10min.

Synthèse d'ADNc avec l'enzyme SuperScript II (Invitrogen™). Pour les échantillons présentant une trop faible concentration en virus ne permettant pas l'amplification par l'enzyme ReverTAID, les ADNc sont synthétisés selon le protocole suivant : un pré-mix est réalisé en mélangeant 5µl d'ARN, 0,4µl de dNTP (25mM), 0,4µl d'amorce N6 (héxamère aléatoire) à 100µM, 0,2µl d'oligodT(18) à 100µM et 6µl H_2OmQ. Les échantillons sont incubés 5min à 65°C puis immédiatement placés sur la glace. Le mix suivant est ensuite rajouté : 4µl de Buffer SSII, 2µl de DTT, 1µl de la reverse transcriptase SuperScript II, et 1µl d'inhibiteur de RNase Ribolock (ThermoScientific™, 20u). Une étape favorisant l'hybridation des amorces est réalisé via une incubation à 25°C pendant 10min, puis une étape d'élongation et de synthèse du cDNA est réalisée à 42°C pendant 50min. Enfin l'enzyme RT est désactivée via une étape à 70°C pendant 15min. 0,3µl de RNase H (ThermoScientific™, 5U/µl) est ensuite rajouté pour permettre la dégradation des ARN restants, via une incubation de 20min à 37°C.

L'étape d'amplification à partir des ADNc totaux a été réalisée en suivant le protocole de la Taq Phusion® décrit pour les PCR attB. Différents cycles ont été utilisés en fonction des amorces utilisées et de la taille du produit d'amplification. Pour les PCR visant à amplifier uniquement la VPg, le cycle suivant a été utilisé : Dénaturation initiale à 98°C pendant 30s, puis 40 cycles comportant les étapes de dénaturation à 98°C pendant 10s, hybridation à 72°C pendant 30s et élongation à 72°C pendant 10min. Une étape d'élongation finale est réalisée à 72°C pendant 10s.

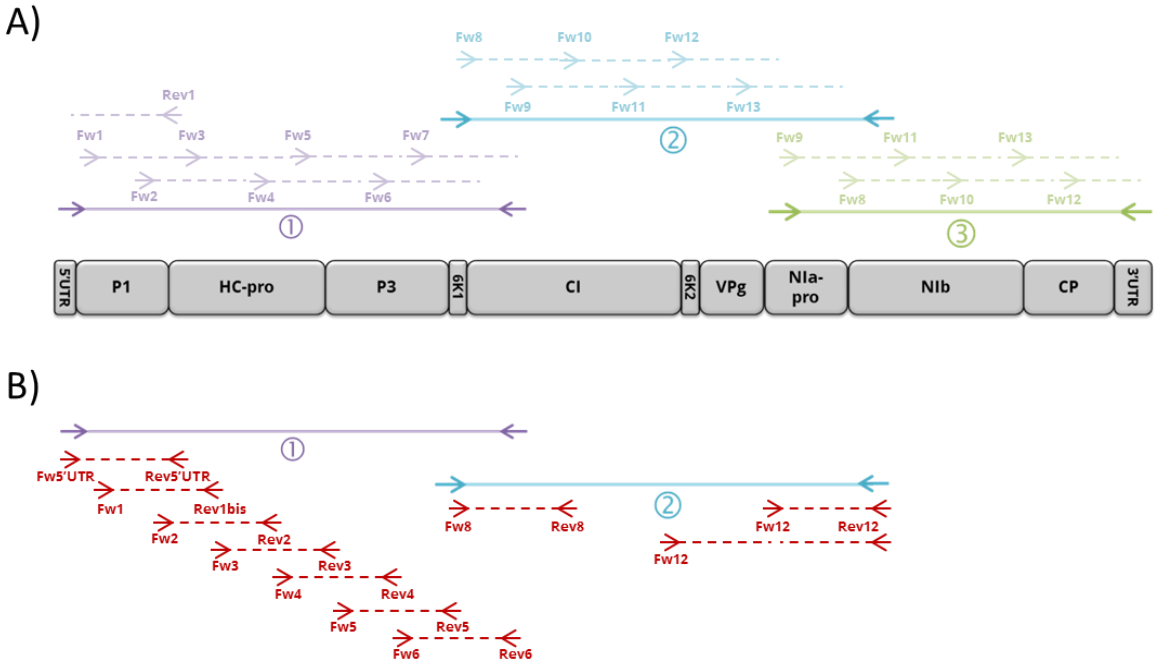


***Figure M13*|Stratégie de séquençage des génomes viraux. A)** Les lignes continues représentent les 3 amplicons de PCR obtenus après amplification des cDNA avec chacun des 3 couples de primers décrits en annexes, tableau A4.4. Les lignes pointillées représentent les produits de séquençage Sanger et les primers correspondants. **B)** Les lignes en pointillés représentent les couples d'amorces internes utilisées pour la seconde amplification en Nested-PCR sur les amplicons notés ①,② and ③.

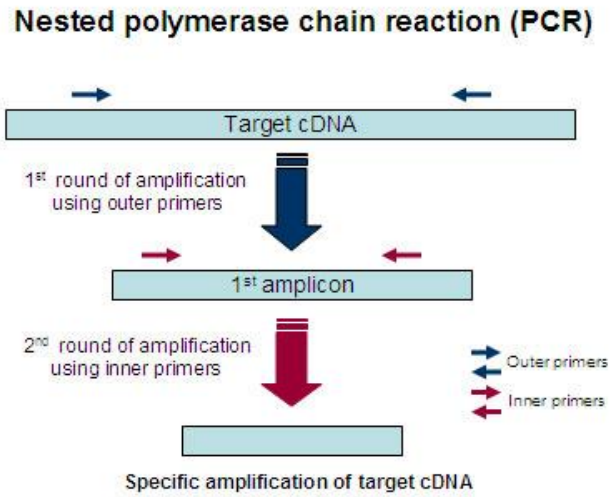


***Figure M14*|Principe de la stratégie de PCR nichée (Nested-PCR).**

Pour l'amplification du génome complet en 3 produits (cf. Figure M13.A et Annexes A4.4), le même mélange réactionnel a été utilisé, mais certaines étapes du cycle d'amplification ont été modifiées étant donné la taille des produits d'amplification et la difficulté à amplifier certains échantillons : la dénaturation initiale a été réalisée à 98°C pendant 2min, ensuite 40 cycles comprenant les étapes suivantes ont été réalisés : 45 sec à 98°C, 45 sec à (62°C pour la PCR PVYfull-part1, 68°C pour PVYfull-part2 et 72°C pour PVYfull-part3[2]) et une élongation de 3min et 30s à 72°C. Enfin, une élongation finale est faite à 72°C pendant 10min. Pour certains échantillons, une stratégie de PCR nichée (cf. Figure M14 et annexes A4.4) a dû être utilisée en raison du rendement trop faible de la PCR sur génome complet probablement dû à une très faible concentration en ADNc viral. Des amorces dites "internes" ont donc été désignées et utilisées pour amplifier les produits de PCR PVYfull-part1 et PVYfull-part2 (Figure M13.B) de certains échantillons (cf. tableau A4.4 fourni en annexe).

Révélation des acides nucléiques

Pour l'ensemble des PCR réalisées, la présence de l'amplicon est déterminée en faisant migrer par système d'électrophorèse (100mV) 1 à 2µl de produit PCR ou 5µl de produit de digestion plasmidique sont déposés sur gel de TAE-agarose 1%, préalablement supplémenté en Bromure d'Ethidium (BET). Les acides nucléiques sont révélés aux UV et la conformité des tailles des produits d'amplification est vérifiée en utilisant du marqueur de taille 1kb (+) (Invitrogen™).

Séquençage des produits d'amplification

L'ensemble des produits d'amplification sont séquencés par la technique de Sanger, qui permet d'accéder à la séquence consensus (majoritaire) présente dans chaque tube.

- **Séquençage des produits de clonage**

Les produits de clonage ou de PCR intermédiaires sont automatiquement envoyés au séquençage chez GATC Biotech pour vérifier l'orientation et l'absence de mutation des inserts introduits/amplifiés au cours de la réaction de clonage/PCR.

- **Séquençage des génomes viraux**

Pour le séquençage sur génome complet, 19 amorces sont utilisées (présentées en annexe Annexes A4.5). De façon à assurer une couverture totale de la séquence codante des génomes viraux, ces amorces sont définies de façon à ce qu'elles s'hybrident à environ 500 nt les unes des autres, et sont réparties sur les 3 produits d'amplification (cf. Figure M13.A). Les amplicons obtenus par Nested-PCR sont séquencés avec les amorces utilisées pour la seconde PCR. Les séquences ainsi obtenues sont traitées via le logiciel CLC-genomics (QIAGEN®) : les séquences (reads) sont triées (« trimmées ») en fonction de la qualité du séquençage, déterminée par l'amplitude du signal du chromatogramme.

---

[2] *les T_a de chaque couple d'amorce ont été calculés avec l'outil Tm calculator disponible sur www.neb.com.*

Les portions des séquences correspondant à un séquençage de faible qualité (cut off à 0.05), et pouvant donc conduire à une mauvaise interprétation des résultats, sont écartées de l'analyse. Pour chacun des échantillons, les reads ainsi trimmés sont ensuite mappés sur le génome de référence du PVY-SON41p, et la séquence consensus issue de l'alignement de tous les reads est extraite. Les changements nucléotidiques identifiés entre la référence et la séquence consensus sont analysés, et dans le cas d'un conflit, un second séquençage est réalisé pour s'assurer de la véracité du changement en question. Pour chaque échantillon, la séquence consensus est ensuite analysée sur le logiciel Serial Cloner 2.6 (http://serialbasics.free.fr/Serial_Cloner.html). La séquence en acides aminés correspondante est comparée avec la séquence de la polyprotéine de référence.

Obtention des vecteurs d'expressions

- **Protocole de clonage Gateway®**

Vecteurs d'entrée. Réaction de recombinaison BP dans pDONR201™

La réaction de recombinaison BP est réalisée via le kit Gateway® BP Clonase® II enzyme mix (ThermoScientific™) selon le protocole suivant : 150ng du vecteur pDONR™201 sont mélangés à 150ng de produit PCR-attB et le volume est ajusté à 9µl avec de l'eau milliQ. 1µl d'enzyme BP recombinase est ensuite ajouté puis les tubes sont laissés toute la nuit à 25°C. Le lendemain, 2µl de protéinase K (fournie avec le kit) sont rajoutés au mélange réactionnel et sont mis à incuber pendant 10min à 37°C pour stopper la réaction.

Vecteurs de destination

Les vecteurs d'expression en système Gateway® sont tous obtenus à partir d'une réaction de recombinaison dite « LR » (cf. Figure M2) à partir des vecteurs donneurs pDONR™201+inserts obtenus par la méthode décrite dans la section précédente.

La réaction LR est réalisée via le kit Gateway® LR Clonase® II enzyme mix (ThermoScientific™) selon le protocole suivant: 150ng de vecteur de destination sont mélangés avec 150ng de vecteur d'entrée, le mélange réactionnel est ajusté à 9µl avec de l'eau milliQ puis 1µl d'enzyme LR recombinase est ajouté.

Dans le cas où les vecteurs d'entrée et de destination possèdent tous deux le gène de résistance à la kanamycine, il existe alors un risque de sélectionner des clones contenant les vecteurs d'entrée. Pour éliminer cette possibilité, on procède à un clonage dit « par PCR ». Le gène contenu dans le vecteur d'entrée est amplifié par PCR grâce à deux amorces contenant les séquences attL1 et attL2 nécessaires au clonage par recombinaison dans le vecteur de destination (cf. annexes A4.2). L'amplifiat PCR obtenu est un fragment contenant le gène d'intérêt flanqué par les sites attL1 et 2.

Le cycle de PCR est le suivant : une première étape de dénaturation à 98°C pendant 30s est réalisée. Elle est suivie de 30 cycles définis comme suit la première étape de dénaturation à 98°C pendant 10s est suivie d'une étape d'hybridation des amorces à 60°C pendant 30s, puis d'une étape d'élongation à 72°C pendant 30s. Enfin, une élongation finale est réalisée à 72°C pendant 5min. Comme les produits PCR-attB, les produit PCR-attL sont traités à la DpnI, purifiés, dosés et analysés sur gel 1%. 150ng de ces produits PCR-attL sont ensuite utilisés pour la réaction LR. La réaction LR se fait à 25°C toute la nuit et est suivie d'un traitement à la protéinase K identique à précédemment.

- **Protocole d'obtention des vecteurs d'expression pETGB1a-eIF4E**

Digestion des inserts et du plasmide pETGB1a avec NcoI et BamHI

Les digestions du vecteur pETGB1a et des produits de PCR obtenus précédemment sont réalisées parallèlement selon les conditions suivantes : pour la digestion plasmidique, 1µg de pETGB1a est mélangé avec 5µl de tampon 10X 3.1 (NEB®) et 1µl de chaque enzyme NcoI (NEB®, 10U/µl) et BamHI (NEB®, 20U/µl), et le volume réactionnel final est ajusté à 50µl avec de l'eau milliQ. Parallèlement, 500ng de produits PCR purifiés sont mélangés avec 3µl de buffer 10X 3.1 (NEB®), 1µl de chaque enzyme NcoI et BamHI et le volume final est ajusté à 30µl avec de l'eau milliQ. Les digestions sont réalisées toute la nuit à 37°C et les produits purifiés le lendemain selon deux protocoles différents : les produits PCR digérés sont purifiés selon le protocole décrit précédemment pour les produits de PCR attB. Contrairement au système Gateway®, le plasmide pETGB1a non-recombinant n'est ici pas contre-sélectionné après transformation. Une digestion partielle peut donc générer de nombreux faux positifs (obtention d'une grande quantité de clones porteurs du plasmide de départ dépourvus de l'insert). Les produits de digestion du plasmide pETG1a sont donc séparés du plasmide non clivé par migration sur gel d'agarose (voir section « purification d'ADN).

Ligation des produits de digestion des inserts eIF4E avec le plasmide pETGB1a

Afin de liguer les produits issus de la double digestion NcoI/BamHI, une réaction de ligation consistant à reformer des liaisons phosphodiesters entre les extrémités chevauchantes et complémentaires des inserts et vecteurs est réalisée. 1µl de T4 DNA ligase (Invitrogen®, 1U/µl) est ainsi mélangé avec 2µl de tampon de ligase 10X et 50ng de vecteur pETGB1a et produits de PCR préalablement digérés correspondant à un ratio molaire 8:1 (insert : vecteur). Le volume du mélange réactionnel est ajusté à 20µl avec de l'eau milliQ. La réaction de ligation se fait à 22°C toute la nuit, puis l'enzyme est inactivée par incubant les tubes 10min à 65°C. Les produits de ligation sont ensuite clonés et amplifiés en bactéries (cf. section « Transformations bactériennes ».

Tableau M1 **| Milieux utilisés pour les cultures bactériennes de *E.coli*.**

Milieu SOC	
Tryptone	2%
Extraits de levure	0,5%
NaCl	10mM
KCl	2,5mM
MgCl2	10mM
MgSO4	10mM
Glucose	20mM
Milieu LB	
Tryptone	1%
Extraits de levure	0,5%
NaCl	10mM

Transformations bactériennes

- **Produits de mutagénèse dirigée**

5µl de mix KLD est ajouté à 50µl de bactéries *E.coli* NEB 5-alpha Competent (High Efficiency) fournies avec le kit de mutagénèse (NEB). Le mélange est laissé à incuber 30min sur la glace, puis un choc thermique de 30s à 42°C est réalisé pour perméabiliser les membranes cellulaires et permettre ainsi la transformation des bactéries par les plasmides. Les tubes sont laissés dans la glace 5min puis 950µl de SOC (cf. Tableau M1) sont ajoutés aux bactéries avant de les incuber une heure à 37°C. 100µl de culture sont ensuite étalés sur boîtes de Pétri contenant du milieu Luria-Bertani (LB) (cf. Tableau M1) supplémenté en Kanamycine (50µg/ml). Les bactéries sont mises en culture toute la nuit à 37°C. Le lendemain, 3 clones par conditions sont ensemencés dans 5ml de LB liquide + Kanamycine et mis à incuber sous agitation toute la nuit.

- **Produits de recombinaison en systèmes Gateway®**

3µl de produits de réaction BP sont ajoutés à 20µl de bactéries *E. coli* MAX efficiency® DH10B™ Competent Cells (ThermoFisher™, 18297-010), et 40µl de glycérol stérile. Le mélange est introduit dans une cuve d'électroporation, et un pulse électrique (paramètres 1.8/kV/cm, capacité de 25µF et résistance à 200Ω) est appliqué. Les bactéries sont immédiatement remises en suspension dans 350µl de milieu SOC et incubées 1h à 37°C. 150µl de culture sont ensuite étalés sur boîtes de Pétri contenant du milieu Luria-Bertani (LB) inclus dans de l'agar solide et l'antibiotique de sélection approprié (cf. tableau Annexe $). Les boîtes sont mises en culture à l'étuve à 37°C toute la nuit.

- **Produits de ligation pETGB1-eIF4E**

4µl de produits de ligation sont utilisés pour transformer 40µl de bactéries *E. coli* électrocompétentes DH10B, supplémentées avec 60µl de glycérol. Les électroporations sont réalisées dans les mêmes conditions que précédemment. 150µl de produits de transformation sont étalés sur boîtes LB + Kanamycine (50µg/ml). Après une culture durant une nuit à 37°C, les clones ayant poussé sont testés en PCR sur colonie pour vérifier la présence de l'insert eIF4E.

- **Vecteurs d'expression pETGB1-eIF4E**

Pour chacune des constructions (cf. Annexe Tableau A5.1), 1µl de plasmide d'expression est mélangé avec 50µl de bactéries BL-21 AI™ puis l'ensemble est incubé 30min dans la glace. Un choc thermique de 30s à 42°C est réalisé puis les cultures sont reprises dans 250µl de SOC et laissées 1h à 37°C sous agitation. 25µl de culture sont récupérés et dilués dans 25µl d'eau stérile et ensuite étalés sur milieu LB + antibiotique pendant une nuit à 37°C.

***Tableau M2*|Milieux de cultures des levures *S. cerevisiae*.**

Milieu YPD	
Difco™ Yeast extract	1% (p/v)
Bacto™ Peptone	2% (p/v)
D-glucose	2% (p/v)
+/- Agar-agar	2% (p/v)
Milieu CAU	
Clontech™ Minimal SD Base	2,67% (p/v)
Clonetch™ -Trp DO supplemt	0,074%
+/- Agar-agar	2% p/v

***Tableau M3*|Tampons utilisés lors de la transformation des levures YP105.**

Tampon LiAc	
Tris HCl (pH8)	10mM
EDTA	1mM
Acétate de Lithium	100mM
Tampon PEG/LiAc	
Tris HCl (pH8)	10mM
EDTA	1mM
Acétate de Lithium	100mM
PEG 3350	50% (p/v)

Transformation et recombinaison dans la levure

- **Linéarisation du vecteur PVYΔVPg-CD**

La linéarisation du plasmide SON41p-pCAPM13-Trp2micronΔVPg-CD est réalisée en digérant 1µg de ce dernier 1µl d'enzyme NotI (NEB®, 10U/µl), 5µl de tampon 3.1, complété par de l'eau milliQ pour atteindre 50µl de volume réactionnel total. Le produit de digestion est purifié puis vérifié par migration sur gel d'agarose 1%.

- **Transformation des levures**

La souche de levure utilisée pour la réaction de recombinaison est la souche Saccharomyces cerevisiae (*S. cerevisiae*) YPH501. Celle-ci présente la caractéristique d'être délétée au niveau du gène *TRP2*, ce qui l'empêche de synthétiser le tryptophane. La sélection de levures transformées avec le plasmide SON41p-pCAPM13-Trp2micronΔVPg-CD (et donc complémenté par le gène *TRP2* porté par le plasmide), se fera donc sur milieu dépourvu en tryptophane.

Des levures YPH501 sont donc ensemencées dans 5ml de milieu YDP liquide (cf. Tableau M2) sous agitation à 30°C toute la nuit. Le lendemain, 20ml de milieu de culture YPD sont ajoutés, et les levures sont remises en cultures sous agitation jusqu'à atteindre une DO_{600nm} d'environ 0,7. Les cultures sont ensuite centrifugées 10min à 4000rpm puis les culots sont re-suspendus dans 10ml d'eau stérile. Une nouvelle centrifugation de 5min à 4000rpm est réalisée puis les culots sont repris dans 500µl d'une solution de LiAc (cf. Tableau M3), centrifugés 5min à 4000rpm puis resuspendus dans 500µl de LiAc et laissés à température ambiante pendant 3 heures.

Pour l'étape de transformation, 100µl des levures compétentes sont ajoutées à 15µl d'un mélange insert/vecteur (ratio 3:1) contenant respectivement 300ng et 100ng. 70µg de sperme de saumon préalablement dénaturé à 50°C est ajouté au mélange, ainsi que 700µl de solution de PEG/LiAc. Les tubes sont ensuite mis à incuber 1 heure à 30°C sous agitation à 250rpm.

Un choc thermique à 42°C pendant 15min est appliqué afin de permettre la perméabilisation transitoire des membranes de levure, et permettre ainsi l'entrée de l'ADN exogène. Les échantillons sont ensuite centrifugés 5min à 5000rpm puis le culot est repris dans 400µl d'eau stérile. 200µl de cette suspension cellulaire sont ensuite étalés sur boîtes de milieu sélectif CAU dépourvu en tryptophane et incubées 48 heures à 30°C.

Tableau M4 **| Tampons utilisés lors de l'extraction plasmidique à partir des levures YP105.**

Tampon de resuspension	
KH2PO4	67mM
Tampon de lyse	
NaOH	0,2N
SDS	1%
Tampon de Neutralisation	
Ko.Ac.	3M/5M

Extraction d'acides nucléiques

- **Purification de plasmides à partir de cultures de levures**

Pour purifier les plasmides des clones positifs, les 4ml de cultures sont centrifugés 15min à 15000xg et le culot est repris dans 100µl de tampon de re-suspension (cf. Tableau M4). 5µl de Zymolyase (10mg/ml) sont ajoutés, puis les tubes sont mis à incuber 1h à 37°C. Par la suite, 200µl de tampon de lyse sont ajoutés et l'ensemble est incubé 5min à température ambiante. L'ajout de 150µl de tampon de neutralisation permet ensuite de stopper la réaction de lyse et une centrifugation de 12min à 15000xg (4°C) permet de récupérer le surnageant contenant les acides nucléiques. Ce surnageant est transféré dans un tube contenant 2 volumes d'éthanol 100% froid et placés ensuite à -20°C pendant 10min pour permettre la précipitation des acides nucléiques. Le culot contenant l'ADN est récupéré grâce à une étape de centrifugation de 15min à 15000xg (4°C) puis re-suspendu et lavé avec 200µl d'éthanol à 70%. Une nouvelle étape de centrifugation est réalisée pendant 5min à 15000xg (4°C) pour éliminer l'éthanol. Les culots sont ensuite séchés au speed-vaccuum pendant 10min jusqu'à élimination totale des traces d'éthanol. Enfin, les ADN plasmidiques sont repris dans 15µl d'eau milliQ et stockés à 4°C en attendant d'être transformés dans *E.coli*.

- **Purification de plasmide à partir de cultures bactériennes**

L'extraction plasmidique est réalisée avec le kit Wizard® Plus SV Minipreps DNA Purification Systems (Promega™). Le protocole est le suivant : 5ml de culture bactérienne sont centrifugés à 4000rpm pendant 10min afin d'éliminer le milieu de culture. Les culots bactériens sont ensuite remis en suspension dans 250 µl de solution de re-suspension cellulaire (50mM Tris-HCl pH 7.5, 10mM EDTA, 100µg/ml RNAse A). Après transfert dans un tube de 1,5ml, 250µl de solution de lyse cellulaire (0.2M NaOH, 1% SDS) sont ensuite ajoutés et les tubes sont mélangés par inversion. L'ajout de 10µl de solution de protéase alcaline permet l'inactivation des endonucléases et autres protéines libérées lors de la lyse cellulaire et qui pourraient altérer le rendement et la qualité de l'extraction plasmidique. Les tubes sont incubés 5min à température ambiante, puis 350µl de solution neutralisante (4.09M d'hydrochlorure de guanidine, 0.759M d'acétate de potassium, 2.12M acide acétique glacial, pH 4.2) sont ajoutés pour stopper l'activité protéasique et prévenir la dégradation de l'ADN. Le lysat bactérien est centrifugé à 14000xg pendant 10min à température ambiante. Le surnageant (phase translucide) est déposé sur des colonnes *Wizard® SV Minicolumns*. Les colonnes sont centrifugées à 14000xg pendant 1min à température ambiante, le volume non retenu est éliminé et les colonnes sont lavées avec 750µl puis 250µl de solution de lavage (60mM Acétate de potassium, 8.3mM Tris-HCl pH 7.5, 0.04mM EDTA pH 8.0, 60% EtOH). Entre chaque lavage, une étape de centrifugation (respectivement 1min et 2min) à 14.000xg, est effectuée.

Les tubes sont ensuite à nouveau centrifugés à vide pour s'assurer de l'élimination totale de l'éthanol sur la membrane, puis l'ADN plasmidique est élué par ajout de 50µl d'eau milliQ et centrifugation à 14.000xg pendant 1min.
Les plasmides ainsi purifiés sont ensuite dosés au spectrophotomètre. Leur taille est analysée par digestion de l'enzyme de restriction BamHI possédant 1 site de clivage sur le plasmide pDONR201, selon le protocole suivant : 1µl de BSA (NEB®, 20mg/ml), 1µl de tampon NEBuffer 3.1, 0,5µl d'enzyme de restriction (NEB®, 20U/µl) et 150ng de plasmide sont mélangés dans un tube 1,5ml et le volume est ajusté à 10µl avec de l'eau mQ. La réaction de digestion est réalisée à 37°C pendant 1h. Les produits de digestion sont révélés après migration sur gel d'agarose. Les séquences des vecteurs sont vérifiées par séquençage Sanger, en utilisant l'amorce universelle pDONR-FP.

- **Purification des produits PCR**

Les produits de PCR traités à la DpnI sont purifiés via le kit NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel™, référence 740609.50) en suivant les instructions fournies : 100µl de tampon d'accrochage NTI sont ajoutés aux 50µl de produits PCR. Les sels chaotropiques contenus dans le NTI permettent l'accrochage et la rétention de l'ADN sur la membrane de silice contenue dans la colonne. Le tout est vortexé puis transféré sur colonne. Une première centrifugation de 30s à 13000xg est réalisée puis 2 lavages avec 700µl d'éthanol sont effectués afin d'enlever les contaminants. Afin d'éliminer toutes trace d'éthanol, les colonnes sont centrifugées à vide pendant 1min à 13000xg. L'élution des produits PCR purifiés est réalisée en conditions faiblement salines via l'ajout de 30µl de tampon NE (5mM Tris/HCl, pH 8.5), après une incubation de 1min à température ambiante et une centrifugation de 1min à 13000xg. La pureté et la taille des produits PCR sont ensuite vérifiées par migration sur gel d'agarose. Les échantillons sont dosés au spectrophotomètre puis envoyés au séquençage.

- **Purification des produits de digestion plasmidique**

Le mix de digestion plasmidique est soumis à une électrophorèse sur gel 0.8% agarose. Après migration, la bande d'intérêt est révélée sur table UV puis excisée du gel et récoltée dans un tube 1.5ml. 200µl de tampon NTI pour 100mg de gel prélevé sont ajoutés. Le mix est dissout à 55°C pendant 10min. Le mélange est ensuite purifié via le kit NucleoSpin® Gel and PCR Clean-up utilisé précédemment (voir section "Purification des produits PCR").

- **Extraction de l'ARN viral**

Les extractions d'ARN totaux ont été réalisées via le kit TRI-reagent à partir de 2 types d'échantillons, soit 200µl de broyat (feuilles broyées dans 3 volumes (p/v) de PBS-Tween PVP), soit de dix disques foliaires broyés à sec (expérience de contournement).
Selon le matériel de départ, 500µl ou 1ml de TRI-reagent sont ajoutés aux broyats foliaires, les échantillons sont vortexés et laissés 10min à température ambiante.

200µl de chloroforme sont ensuite ajoutés, puis le mélange est vortexé pendant 30s avant d'être centrifugé 15min à 14000rpm à 4°C. Après centrifugation, la phase supérieure est transférée dans des nouveaux tubes contenant un volume égal d'isopropanol. Puis les tubes sont mélangés vigoureusement pendant 30s, avant d'être laissés 10min à température ambiante, pour ensuite être centrifugés 15min à 14000rpm (4°C). Les culots contenant les acides nucléiques sont ensuite repris dans 500µl d'éthanol 70%, puis à nouveau centrifugés 5min à 14000rpm (4°C). Les culots ainsi lavés sont débarrassés de l'éthanol et mis à sécher au speed-vacuum pendant 5min, puis re-solubilisés dans 20µl d'eau "RNase free". Les ARN sont enfin dosés au spectrophotomètre afin de déterminer leur concentration et leur pureté.

b. Techniques de biochimie

Expression et purification des formes recombinantes de VPg et eIF4E

- **Analyse rapide de produits d'expression à partir des clones positifs.**

Après transformation par le vecteur d'expression d'intérêt, 3 clones positifs sont utilisés pour ensemencer 3 fois 1ml de LB contenant l'antibiotique de sélection. Après une nuit à 37°C, 50 à 100 µl de culture sont inoculés dans 5ml de milieu frais. Une culture est conduite à 37°C jusqu'à une valeur de densité optique de 0.4 et l'expression hétérologue induite par ajout de l'inducteur (1mM IPTG ou 13mM arabinose en concentrations finales). La culture est ensuite conduite dans des conditions variables de température (20-37°C) et de durée variables (3 à 16h) afin d'optimiser les conditions de production. 1ml de culture est prélevé et les cellules culotées par centrifugations sont reprises dans 40 µl d'eau milliQ. 10 µl de tampon de charge (LOB) sont ajoutés et après incubation à 90°C pendant 3min, un aliquot de 10 µl est analysé par SDS-PAGE (SDS-polyacrylamide gel electrophoresis) en conditions dénaturantes. La présence de la protéine d'intérêt est révélée par western blot en utilisant des anticorps anti-His^{6x}. Les clones produisant de la protéine sont conservés sous forme de stock glycérolé (500µl de culture dans 500µl de glycérol stérile 50%) à -80°C.

- **Production, extraction et purification des protéines recombinantes d'intérêt**

Les stocks sont utilisés à chaque nouvelle expérience d'expression pour relancer les pré-cultures sont ensuite réensemencées dans des erlens autoclavés contenant 1litre de milieux LB+antibiotique, puis cultivés sous agitation à 37°C jusqu'à l'obtention d'une DO_{600nm} d'environ 0,5. Pour l'induction de l'expression des protéines d'intérêt, 0,2% (W/V) de L-arabinose sont ajoutés au milieu. Dans le cas des vecteurs pETGB1a, 1mM d'IPTG est également ajouté. L'induction est ensuite réalisée pendant 3h à 28°C. Les cultures sont ensuite culotées par centrifugation 15min à 20000 g. Le surnageant est éliminé et les culots bactériens sont repris dans 10ml de LB puis transférés dans des tubes Falcon de 50ml. Ils sont centrifugés 8min à 4000rpm et stockés à -20°C jusqu'à l'extraction.

Tableau M5 **| Tampon d'extraction des protéines totales.**

Tampon Extraction (pH8)	
HEPES	20mM
EDTA	250µM
Tween 20	0,25% (p/v)
DTT	1mM
β-mercaptoéthanol	2,5mM
$MgCl_2$	20mM
Cocktail Protéase	1 pastille

Tableau M6 **| Tampons utilisés lors des chromatographies sur colonnes.**

Tampons de base Ni-NTA (pH8)	
Phosphate	20mM
NaF	500mM
Tampons de lavage Ni-NTA (pH8)	
Phosphate	20mM
NaF	500mM
Imidazole	25mM et 50mM
Tampon Elution Ni-NTA (pH8)	
Phosphate	20mM
NaF	500mM
Imidazole	250mM

Extraction des protéines totales

Deux tampons d'extractions différents sont utilisés. Pour les expériences de dichroïsme circulaire, les culots bactériens sont doucement suspendus dans 20ml de tampon d'extraction Phosphate/NaF à pH.8 (cf. Tableau M5.A) à l'aide d'un potter. Pour ce qui est des expériences de spectrofluorescence, le tampon présenté Tableau M5.B est utilisé. Ensuite, pour faciliter la lyse cellulaire, 0,4ml de solution de lysozyme à 50X et 100µl d'inhibiteur de protéase PMSF (0,2M) sont ajoutés. Le lysat est incubé 1h à 4°c sous agitation lente. Le mélange devient visqueux (cette viscosité témoigne du relargage de l'ADN et donc de l'efficacité de la lyse cellulaire). Le lysat est potterisé et soumis à 3 cycles de sonication successifs de 1min. 100µl de PMSF (0,2M) sont à nouveau ajoutés dans la solution, puis les échantillons sont centrifugés 30min à 13000rpm (4°C).

Purification des protéines recombinantes par Immobilized Metal Ion Affinity Chromatography

0,8ml de résine Ni-NTA Protino® est tout d'abord équilibrée avec 5ml de tampon de base (cf. Tableau M6). Les extraits protéiques solubles issus des cultures sont mis à incuber au moins 30min avec la résine sous agitation lente à 4°C, pour permettre aux histidines de se fixer aux billes de nickel. Le mélange est ensuite tassé dans une petite colonne et l'éluant non fixé est éliminé. Un premier lavage est réalisé avec 15ml de tampon de lavage contenant 25mM puis un second à 50mM d'imidazole. Ces lavages successifs à faibles concentrations en imidazole permettent d'éluer les interacteurs de faible affinité au Ni^{2+} qui pourraient contaminer le produit de purification final. Afin de pouvoir décrocher nos protéines tagguées His_6, il sera donc nécessaire de réaliser une élution en concentrations d'imidazole élevée soit en utilisant 8ml de tampon d'élution, contenant de l'imidazole à 250mM. L'imidazole est un réactif qui interagit également avec les billes de nickel de façon très affine et rentre donc en compétition avec les interacteurs accrochés aux billes. Les produits de purification sont récupérés en fractions de 0,8ml.

Migration des protéines totales en électrophorèse sur gel d'acrylamide

Un aliquot de chaque extrait (intermédiaire et final) après purification est soumis à une analyse sur gel d'acrylamide-bisacrylamide en conditions dénaturantes (SDS-PAGE). Pour cela, 10µl de produit d'extraction (T_0 prélevé avant induction, culot et surnageant prélevés après l'étape de centrifugation des lysats) sont mélangés avec du tampon de charge (LOB 5X) (cf. Tableau M7), puis laissés 5min à température ambiante et enfin chauffés 5min à 90°C. Les échantillons sont déposés sur gel d'acrylamide à 12.5% composé d'un gel de concentration et d'un gel de séparation, puis mis à migrer dans un tampon de migration contenant du SDS (cf. Tableau M8). Le SDS permet la dénaturation et le chargement négatif de toutes les protéines. L'électrophorèse va alors permettre de séparer les protéines de l'échantillon selon leur taille.

***Tableau M7*|Réactifs du tampon de charge (5X) utilisés en SDS-PAGE.**

Tampon de charge 5X	
Tris HCl pH 6,8	60 mM
Glycérol	25 % (v/v)
SDS	2 % (v/v)
B-mercaptoéthanol	14,4 mM
Bleu de Bromophénol	0,25 % (p/v)

***Tableau M8*|Tampons utilisés en SDS-PAGE**

Gel de séparation	
Acrylamide/BisAcrylamide	15 %
SDS	0,1 %
Tris HCl pH 8,8	375 mM
APS	0,1 %
TEMED	0,08 %
Gel de concentration	
Acrylamide/BisAcrylamide	5 %
SDS	0,1 %
Tris HCl pH 6,8	125 mM
APS	0,1 %
Tampon de migration	
Tris-base	3 g/l
Glycine	14,4 g/l
SDS	2 %

***Tableau M9*|Tampons utilisés pour le transfert et l'immunodétection par Western blot.**

Tampon de lavage TTBS	
Tris pH 7,4	10 mM
NaCl	150 mM
Tween 20	0,05 %
Tampon d'hybridation	
Tampon de lavage	100 ml
Lait en poudre	5g
Tampon de révélation	
Tris pH 9,5	100 mM
NaCl	250 mM
MgCl2	1 mM

Tampon de transfert	
Tris-base	3 g/l
Glycine	14,4 g/l
Méthanol	200 ml/l

La migration est réalisée 15min à 110V puis lorsque que le front de migration parvient à la limite du gel de concentration, la migration dans le gel de séparation se fait à 130V jusqu'à ce que le premier front de migration atteigne la limite basse du gel. Les gels sont réalisés en double afin de pouvoir à la fois voir les protéines totales par révélation au Bleu de Coomassie et réaliser parallèlement des immunomarquages spécifiques (Western blot) afin de valider la présence des protéines d'intérêt (cf. ci-dessous).

Coloration des protéines totales au Bleu de Coomassie

Après l'électrophorèse, le premier gel d'acrylamide est lavé 3 fois à l'eau puis placé dans la solution de coloration Instant Blue™ (Expedeon). Une fois qu'une coloration satisfaisante est obtenue, la réaction est arrêtée en plaçant le gel dans l'eau.

Révélation des protéines d'intérêt par Western blot

Les protéines totales ayant migré sur le second gel sont transférées sur une membrane de nitrocellulose (Protran® Whatman). Des feuilles de papier Whatman imbibée de tampon de transfert est disposée sur la plaque anode de l'appareil, puis une feuille de nitrocellulose. Le gel est positionné sur la nitrocellulose puis recouvert d'une seconde couche de papier Whatman imbibé (conditions de transfert dit semi-sec). La cathode est ensuite clipsée sur l'ensemble (Tableau M9). Un champ électrique est appliqué à 150mA durant 50min. Les protéines chargées négativement sont transférées sur la membrane de nitrocellulose en migrant vers l'anode (l'appareil Trans-Blot SD semidry transfer cell de chez Bio-Rad™ est utilisé). Après transfert, la membrane est lavée 2 fois pendant 5min dans du TTBS 1X, puis placée dans une solution de TTBS1X + lait 5% durant une nuit à 4°C (Tableau M9).

Le lendemain, la membrane est incubée 1h dans du tampon TTBS1X + lait 3% contenant la solution d'anticorps primaires monoclonaux anti-histidine de souris, diluée au 5000ème (HRP-Sigma®).

Les anticorps primaires en excès sont ensuite éliminés grâce à 3 lavages successifs de 5min au TTBS 1X. La membrane est incubée 30min dans une solution de TTBS 1X contenant la solution diluée au 10000ème d'anticorps secondaires polyclonaux de lapins, dirigés contre les anticorps primaires et couplés à la phosphatase alcaline (Sigma®). La membrane est ensuite lavée 3 fois 5min dans du tampon TTBS 1X, puis un dernier lavage de 5min dans le tampon de révélation est réalisé.

Enfin, la révélation de la membrane se fait par réaction colorimétrique, en ajoutant le substrat de la phosphatase alcaline, le NBT-BCIP (Roche) dilué au 1/100ème dans du tampon de révélation. Lorsque le signal observé est satisfaisant (coloration violette), la réaction est rapidement stoppée en transférant la membrane dans l'eau, afin de ne pas la saturer en signal aspécifique.

Mesures de dichroïsme circulaire

Les mesures de dichroïsme ont été réalisées sur un spectropolarimètre Jasco-815 équipé d'un régulateur de température à effet Pelletier. Les spectres ont été acquis à 20°C, sous atmosphère d'azote, dans l'UV lointain (190-240 nm). Les réglages étaient les suivants : demi largeur de bande 1 nm ; intervalles de mesure 0,2 nm ; acquisition 1s par intervalle ; constante de l'instrument 100 ms. Au-dessous de 230 nm, le signal transmis par l'échantillon diminue considérablement (beaucoup d'absorption) et l'appareil doit augmenter progressivement la tension du photomultiplicateur pour maintenir une intensité de signal constant. Généralement, au-dessus de 500 Volt, le signal n'est plus exploitable. Pour aucun des jeux de données la tension du photomultiplicateur n'a été supérieure à 380 V, à la longueur d'onde la plus faible soit 195nM. Beaucoup de tampons et sels donnent un signal dichroïque important et le bruit de fond généré au-dessous de 230 nm ne permet pas d'obtenir un rapport signal sur bruit suffisant pour exploiter les spectres dans l'UV lointain. C'est pourquoi toutes les préparations des protéines et les mesures ont été réalisées dans du tampon phosphate 50mM et NaF 0.5M (pH 7.8), qui présente un faible signal dichroïque.

Pour chaque échantillon, 10 spectres cumulés ont été réalisés à 0,1 mg/ml de protéines dans des cuves de 100µl et 1mM de trajet optique. Le spectre moyen obtenu a été lissé (polynôme d'ordre 4). Les lignes de bases ont été soustraites avant la conversion du signal en ellipticité molaire moyenne (deg, cm^2 $dmol^{-1}$). Pour la conversion des données en proportions de structures secondaires, les programmes de déconvolution hébergés par le serveur DichroWeb (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) ont été utilisés.

Mesures de fluorescence à l'état stationnaire

Les titrations de l'association des différentes formes de VPg avec les eIF4E Yolo Wonder et HD ont été réalisées à l'aide d'un spectrofluorimètre Xenius Safas (Monaco). Les échantillons étaient excités à 280 nm (fente 5nm) et la fluorescence émise collectée à 340 nm (fente 7 nm). La tension du photomultiplicateur a été ajustée pour un signal de fluorescence de 80% en l'absence de VPg (typiquement de 750 à 880 Volts). Les mesures étaient réalisées à une concentration constante de 210nM en eIF4E dans un volume initial de 500µl contenant de l'HEPES 20mM, KCl 300mM, DTT 1mM à pH 7,9 et 20°C. La titration était réalisée par introduction dans la cuve de prises aliquotées de 2 à 10µl d'une solution concentrée de VPg (entre 3 et 5µM). A chaque ajout, la fluorescence à 340 nm était enregistrée pendant 2min (7 mesures cadencées toutes les 20s). La valeur moyenne était retenue pour traitement. Les VPg utilisées possèdent un résidu tryptophane. Afin de s'affranchir de cette contribution, à chaque point de titration, une cuvette référence ne contenant pas de eIF4E a reçu le même volume de VPg et le signal de ce blanc a été déduit de la fluorescence de la cuve d'essai.

***Tableau M10*|Bloc de titration utilisé lors des mesures de fluorescence réalisées au cours de la formation du complexe formé par eIF4E (HD) et VPg (SON41p).**

cuve	Test	eIF4E (µL)	Tampon (µL)	Commentaire
1	HD+m7GTP	62	438	Contrôle fonctionnalité eIF4E réplica 1
2	baseline S41p	0	500	Contribution autofluorescence de la VPg SON41p
3	HD+S41p (1)	62	438	titration eIF4E HD par VPg SON41p réplica 1
4	HD+S41p (2)	62	438	titration eIF4E HD par VPg SON41p réplica 2
5	HD+S41p (3)	62	438	titration eIF4E HD par VPg SON41p réplica 3
6	HD+m7GTP	62	438	Contrôle fonctionnalité eIF4E réplica 2
7	baseline OD2	0	500	Contribution autofluorescence de la VPg OD2
8	HD+OD2 (1)	62	438	titration eIF4E HD par VPg OD2 réplica 1
9	HD+OD2 (2)	62	438	titration eIF4E HD par VPg OD2 réplica 2
10	HD+OD2 (3)	62	438	titration eIF4E HD par VPg OD2 réplica 3

***Tableau M11*|Tampons utilisés en DAS-ELISA**

Tampon de coating	
Na_2CO_3	15 mM
$NaHCO_3$	30 mM
Tampon PBS-Tween	
NaCl	136,9 mM
KH_2PO_4	1,47 mM
KCl	2,68 mM
NA_2HPO_4	8,1 mM
Tween 20	0,05 % (v/v)
Tampon substrat (pH 9,8)	
Diéthanolamine	9,7 % (v/v)
Tampon conjugué	
Ovalbumine	0,2%

Les signaux de fluorescence ont été aussi ajustés par rapport à la dilution due à l'augmentation de volume au cours de la titration. Dans une expérience typique, la titration a été réalisée à partir de 18 à 20 points de mesure (concentration finale en VPg entre 0 et 700 nM). Les mesures ont été réalisées en simultanée sur trois cuvettes (réplicats techniques). Les trois jeux de données ont été rassemblés et utilisés pour ajuster les valeurs des paramètres Y_{sat} et K_d du modèle de fixation binaire VPg-eIF4E. Chaque « bloc » de titration a ainsi intégré systématiquement deux titrations de l'association eIF4E-m7GTP (analogue de coiffe des ARNm) afin de contrôler la qualité fonctionnelle du lot d'eIF4E utilisé. Dans le même bloc, chaque triplicat de titration d'eIF4E par l'un des mutants de VPg est accompagné d'un triplicat de titration par la VPg du virus de référence (isolat SON41p). A titre d'exemple, un bloc de titration est représenté tableau M10.

c. Test de détection virale par DAS-ELISA

La détection et la quantification relative de particule virale a été réalisée via la méthode de DAS-ELISA ("*Double-antibody sandwich enzyme linked immuno sorbent assay")*. La composition des différents tampons utilisés est récapitulée dans le Tableau M11.

La révélation est réalisée à partir d'échantillons de feuilles broyées par presse dans 3 volumes (p/v) de PBS-Tween PVP et clarifiés par centrifugation. Dans le cas de test semi-quantitatifs ou de calibrage d'inocula, 3 et 6 points de dilutions au 25ème sont réalisés respectivement pour chaque échantillon.

Le coating est réalisé par ajout de l'anticorps polyclonal anti-CP(PVY) (Bioreba, Rèf 110511) est dilué au 1000ème dans le tampon de coating et 100µl sont déposés dans chaque puit de la plaque ELISA. Les plaques sont mises à incuber 3h à 37°C, puis sont lavées à 3 reprises avec du tampon PBS-Tween. 100µl de broyat ou dilutions sont ensuite déposés et mis à incuber toute la nuit à 4°C, ou 3h à 37°C dans le cas de la calibration des inocula.

Par la suite, les plaques sont lavées au PBS-Tween 3 fois avant d'y ajouter 100µl de la solution d'anticorps anti-CP(PVY) conjugués à la phosphatase alcaline (Bioreba), elle-même préalablement diluée au 1000ème dans le tampon du conjugué. Les plaques sont mises à incuber 1h à 37°C, puis à nouveau lavées 3 fois avec du PBS-Tween. Le substrat de révélation (Paranitrophénylphosphate, Sigma®) est préparé extemporanément par dissolution de 2 pastilles dans 12ml de tampon Substrat. La plaque est laissée selon la concentration initiale en virus entre 1h et 4h à température ambiante, le temps pour les substrats d'être transformé en un produit coloré jaune, le paranitrophénol, en présence de phosphatase alcaline. L'absorbance à 405nm des différents puits est mesurée via le spectrophotomètre multiplaque EPOCH™ (Biotek®) et est directement corrélée à la quantité de paranitrophénol produit et donc de phosphatase alcaline fixée sur les puits.

d. Méthodes d'inoculation virale

Inoculation par biolistique

La technologie ainsi que le protocole de préparation du matériel sont fournis par Bio-Rad. 5mg de billes d'or de 1µm de diamètre sont suspendus dans 100µl de spermidine (50mM) puis mis en présence de 1µg d'ADN plasmidique. Cet ADN est précipité sur les billes d'or par ajout, sous agitation, de 100µl de $CaCl_2$ (1M) et incubé 10min à température ambiante. Après 3 rinçages avec 1ml d'éthanol absolu, les billes sont resuspendues dans 600µl d'une solution de PVP-éthanol (50µg/ml). La solution ainsi obtenue est introduite dans un tube en plastique « Tefzel Tubbing » qui est séché sous flux d'azote. Ce tube couvert d'or est découpé en une dizaine de cartouches qui sont insérées dans le barillet d'un pistolet à particule Helios® Gene Gun (Bio-rad). Lors du bombardement, les particules d'or sont propulsées sur le tissu végétal par un flux d'hélium à 180psi.

Inoculation mécanique

Les feuilles sont inoculées par frottement manuel avec un mélange composé de 1g de BOS (feuilles infectées prélevées lors de l'inoculation primaire) broyé dans 4 volumes p/v de tampon Na_2HPO_4 (30mM, pH7), DIECA (2%), additionné de carborundum et de charbon végétal (20mg/ml).

Références bibliographiques

Aaskov, John, Katie Buzacott, Hlaing Myat Thu, Kym Lowry, and Edward C Holmes. 2006. "Long-Term Transmission of Defective RNA Viruses in Humans and Aedes Mosquitoes." *Science (New York, N.Y.)* 311(5758): 236–38.

Abdul-Razzak, Anas, Thomas Guiraud, Martine Peypelut, Jocelyne Walter, Marie Christine Houvenaghel, Thierry Candresse, Olivier Le Gall, and Sylvie German-Retana. 2009. "Involvement of the Cylindrical Inclusion (CI) Protein in the Overcoming of an eIF4E-Mediated Resistance against Lettuce Mosaic Potyvirus." *Molecular Plant Pathology* 10(1): 109–13.

Acevedo, Ashley, Leonid Brodsky, and Raul Andino. 2014. "Mutational and Fitness Landscapes of an RNA Virus Revealed through Population Sequencing." *Nature* 505(7485): 686–90.

Adams, M. J., E. J. Lefkowitz, a. M. Q. King, D. H. Bamford, M. Breitbart, a. J. Davison, S. a. Ghabrial, a. E. Gorbalenya, N. J. Knowles, P. Krell, R. Lavigne, D. Prangishvili, H. Sanfaçon, S. G. Siddell, P. Simmonds, and E. B. Carstens. 2015. "Ratification Vote on Taxonomic Proposals to the International Committee on Taxonomy of Viruses (2015)." *Archives of Virology* 160(7): 1837–50.

Adams, M. J., F.M Zerbini, R French, F Rabenstein, D.C Stenger, and Jari P T Valkonen. 2012. "Virus Taxonomy : Ninth Report of the International Comittee on Taxonomy Viruses. Family Potyviridae." *ICTV*: 1069–89.

Ala-Poikela, Marjo, Elisa Goytia, Tuuli Haikonen, Minna-Liisa Rajamäki, and Jari P T Valkonen. 2011. "Helper Component Proteinase of the Genus Potyvirus Is an Interaction Partner of Translation Initiation Factors eIF(iso)4E and eIF4E and Contains a 4E Binding Motif." *Journal of virology* 85(13): 6784–94.

Ali, Akhtar, Hongye Li, William L Schneider, Diana J Sherman, Stewart Gray, Dawn Smith, and Marilyn J Roossinck. 2006. "Analysis of Genetic Bottlenecks during Horizontal Transmission of Cucumber Mosaic Virus." *Journal of virology* 80(17): 8345–50.

Alves, Carolina, Hong Cheng, Heinrich Roder, and John Taylor. 2010. "Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen." *Virology* 407(2): 333–40.

Anderson, Pamela K., Andrew a. Cunningham, Nikkita G. Patel, Francisco J. Morales, Paul R. Epstein, and Peter Daszak. 2004. "Emerging Infectious Diseases of Plants: Pathogen Pollution, Climate Change and Agrotechnology Drivers." *Trends in Ecology and Evolution* 19(10): 535–44.

Anfinsen, Christian B. 1973. "Principles That Govern the Folding of Protein Chains." *Science* 181(4096): 223–30.

Ayme, Valérie, Julien Petit-Pierre, Sylvie Souche, Alain Palloix, and Benoît Moury. 2007. "Molecular Dissection of the Potato Virus Y VPg Virulence Factor Reveals Complex Adaptations to the pvr2 Resistance Allelic Series in Pepper." *The Journal of general virology* 88(Pt 5): 1594–1601.

Ayme, Valérie, Sylvie Souche, Carole Caranta, Mireille Jacquemond, Joël Chadoeuf, Alain Palloix, and Benoît Moury. 2006. "Different Mutations in the Genome-Linked Protein VPg of Potato Virus Y Confer Virulence on the pvr2(3) Resistance in Pepper." *Molecular plant-microbe interactions : MPMI* 19(5): 557–63.

Baba, Tomoya, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill a Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. 2006. "Construction of Escherichia Coli K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection." *Molecular systems biology* 2: 2006.0008.

Balázs, Annamária, Veronika Csizmok, László Buday, Marianna Rakács, Robert Kiss, Mónika Bokor, Roopesh Udupa, Kálmán Tompa, and Peter Tompa. 2009. "High Levels of Structural Disorder in Scaffold Proteins as Exemplified by a Novel Neuronal Protein, CASK-Interactive protein1." *FEBS Journal* 276(14): 3744–56.

Baltimore, D. 1971. "Expression of Animal Virus Genomes." *Bacteriological reviews* 35(3): 235–41.

Barrick, Jeffrey E, and Richard E Lenski. 2013. "Genome Dynamics during Experimental Evolution." *Nature reviews. Genetics* 14(12): 827–39.

Beauchemin, Chantal, Nathalie Boutet, and Jean-François Laliberté. 2007. "Visualization of the Interaction between the Precursors of VPg, the Viral Protein Linked to the Genome of Turnip Mosaic Virus, and the Translation Eukaryotic Initiation Factor Iso 4E in Planta." *Journal of virology* 81(2): 775–82.

Bellay, Jeremy, Sangjo Han, Magali Michaut, Taehyung Kim, Michael Costanzo, Brenda J Andrews, Charles Boone, Gary D Bader, Chad L Myers, and Philip M Kim. 2011. "Bringing Order to Protein Disorder through Comparative Genomics and Genetic Interactions." *Genome biology* 12(2): R14.

Bershtein, Shimon, Michal Segal, Roy Bekerman, Nobuhiko Tokuriki, and Dan S Tawfik. 2006. "Robustness-Epistasis Link Shapes the Fitness Landscape of a Randomly Drifting Protein." *Nature* 444(7121): 929–32.

Betancourt, Mónica, Alberto Fereres, Aurora Fraile, and Fernando García-Arenal. 2008. "Estimation of the Effective Number of Founders That Initiate an Infection after Aphid Transmission of a Multipartite Plant Virus." *Journal of virology* 82(24): 12416–21.

Bloom, Jesse D, D Allan Drummond, Frances H Arnold, and Claus O Wilke. 2006. "Structural Determinants of the Rate of Protein Evolution in Yeast." *Molecular biology and evolution* 23(9): 1751–61.

Bloom, Jesse D, Alpan Raval, and Claus O Wilke. 2007. "Thermodynamics of Neutral Protein Evolution." *Genetics* 175(1): 255–66.

Bodart, Jean-François, Jean-Michel Wieruszeski, Laziza Amniai, Arnaud Leroy, Isabelle Landrieu, Arlette Rousseau-Lescuyer, Jean-Pierre Vilain, and Guy Lippens. 2008. "NMR Observation of Tau in Xenopus Oocytes." *Journal of Magnetic Resonance* 192(2): 252–57.

Bokros, Carol L, Jeffrey D Hugdahl, Hyong-ha Kim, Virginia R Hanesworth, A N N V A N Heerdent, Karen S Browningt, and Louis C Morejohn. 1995. "Function of the p86 Subunit of Eukaryotic Initiation Factor (Iso) 4F as a Microtubule-Associated Protein in Plant Cells." 92(July): 7120–24.

Bourhis, Jean-Marie, Bruno Canard, and Sonia Longhi. 2006. "Structural Disorder within the Replicative Complex of Measles Virus: Functional Implications." *Virology* 344(1): 94–110.

Brown, Celeste J, Audra K Johnson, and Gary W Daughdrill. 2010. "Comparing Models of Evolution for Ordered and Disordered Proteins." *Molecular biology and evolution* 27(3): 609–21.

Brown, Celeste J, Sachiko Takayama, Andrew M Campen, Pam Vise, Thomas W Marshall, Christopher J Oldfield, Christopher J Williams, and a Keith Dunker. 2002. "Evolutionary Rate Heterogeneity in Proteins with Long Disordered Regions." *Journal of molecular evolution* 55(1): 104–10.

Brown, CJ, and AK Johnson. 2011. "Evolution and Disorder." *Current opinion in structural biology* 21(3): 441–46.

Caranta, Carole, Véronique Lefebvre, Alain Palloix, Plantes Maraîchères, and Saint Maurice. 1997. "Polygenic Resistance of Pepper to Potyviruses Consists of a Combination of Isolate-Specific and Broad-Spectrum Quantitative Trait Loci." 10(7): 872–78.

Carrasco, Purificación, Francisca de la Iglesia, and Santiago F Elena. 2007. "Distribution of Fitness and Virulence Effects Caused by Single-Nucleotide Substitutions in Tobacco Etch Virus." *Journal of virology* 81(23): 12979–84.

Charlesworth, Brian. 2009. "Fundamental Concepts in Genetics: Effective Population Size and Patterns of Molecular Evolution and Variation." *Nature reviews. Genetics* 10(3): 195–205.

Charron, Carine, Maryse Nicolaï, Jean-Luc Gallois, Christophe Robaglia, Benoît Moury, Alain Palloix, and Carole Caranta. 2008. "Natural Variation and Functional Analyses Provide Evidence for Co-Evolution between Plant eIF4E and Potyviral VPg." *The Plant journal : for cell and molecular biology* 54(1): 56–68.

Chen, Jessica Walton, Pedro Romero, Vladimir N Uversky, and A Keith. 2006. "Conservation of Intrinsic Disorder in Protein Domains and Families: II. Functions of Conserved Disorder." *Journal of proteome research* 5(4): 888–98.

Chen, Jessica Walton, Pedro Romero, Vladimir N Uversky, and A Keith. 2008. "Conservation of Intrinsic Disorder in Protein Domains and Families: I. A Database of Conserved Predicted Disordered Regions." *Journal of Proteome research* 5(4): 879–87.

Chroboczek, Jadwiga, Eugénie Hébrard, Kristiina Mäkinen, Thierry Michon, and Kimmo Rantalainen. 2012. "Intrinsic Disorder in Genome-Linked Viral Proteins VPgs of POtyviruses." In *Flexible Viruses*,.

Chung, Liliane, Dalan Bailey, Eoin N Leen, Edward P Emmott, Yasmin Chaudhry, Lisa O Roberts, Stephen Curry, Nicolas Locker, and Ian G Goodfellow. 2014. "Norovirus Translation Requires an Interaction between the C Terminus of the Genome-Linked Viral Protein VPg and Eukaryotic Translation Initiation Factor 4G." *The Journal of biological chemistry* 289(31): 21738–50.

Collins, Mark O, Lu Yu, Iain Campuzano, Seth G N Grant, and Jyoti S Choudhary. 2008. "Phosphoproteomic Analysis of the Mouse Brain Cytosol Reveals a Predominance of Protein Phosphorylation in Regions of Intrinsic Sequence Disorder." *Molecular & cellular proteomics : MCP* 7(7): 1331–48.

Cortese, Marc S., Vladimir N. Uversky, and A.Keith Dunker. 2008. "Intrinsic Disorder in Scaffold Proteins : Getting More from Less." *Progress in Biophysics and Molecular Biology* 98(1): 85–106.

Cozzetto, Domenico, and David T Jones. 2013. "The Contribution of Intrinsic Disorder Prediction to the Elucidation of Protein Function." *Current Opinion in Structural Biology* 23(3): 467–72.

Cuevas, José M, Anouk Willemsen, Julia Hillung, Mark P Zwart, and Santiago F Elena. 2015. "Temporal Dynamics of Intrahost Molecular Evolution for a Plant RNA Virus." *Molecular biology and evolution*.

Culjkovic, Biljana, Ivan Topisirovic, and Katherine L.B. Borden. 2007. "Controlling Gene Expression through RNA Regulons: The Role of the Eukaryotic Translation Initiation Factor eIF4E." *Cell Cycle* 6(1): 65–69.

Daughdrill, Gary W., Pranesh Narayanaswami, Sara H. Gilmore, Agniezka Belczyk, and Celeste J. Brown. 2007. "Dynamic Behavior of an Intrinsically Unstructured Linker Domain Is Conserved in the Face of Negligible Amino Acid Sequence Conservation." *Journal of Molecular Evolution* 65(3): 277–88.

Davenport, Miles P., Liyen Loh, Janka Petravic, and Stephen J. Kent. 2008. "Rates of HIV Immune Escape and Reversion: Implications for Vaccination." *Trends in Microbiology* 16(12): 561–66.

Davies, E K, a D Peters, and P D Keightley. 1999. "High Frequency of Cryptic Deleterious Mutations in Caenorhabditis Elegans." *Science (New York, N.Y.)* 285(5434): 1748–51.

Díaz, Juan A, Cristina Nieto, Enrique Moriones, Verónica Truniger, and Miguel A Aranda. 2004. "Of a Melon Necrotic Spot Virus Strain That Overcomes the Resistance in Melon and Nonhost Plants." 17(6): 668–75.

Diella, Francesca, Niall Haslam, Claudia Chica, Aidan Budd, Sushama Michael, Nigel P. Brown, Gilles Travé, and Toby J. Gibson. 2008. "Understanding Eukaryotic Linear Motifs and Their Role in Cell Signaling and Regulation." *Frontiers in Bioscience* 13(1): 6580–6603.

Dolan, Patrick T, Andrew P Roth, Bin Xue, Ren Sun, a Keith Dunker, Vladimir N Uversky, and Douglas J LaCount. 2015. "Intrinsic Disorder Mediates Hepatitis C Virus Core-Host Cell Protein Interactions." *Protein science : a publication of the Protein Society* 24(2): 221–35.

Domingo, Esteban, Julie Sheldon, and Celia Perales. 2012. "Viral Quasispecies Evolution." *Microbiology and molecular biology reviews : MMBR* 76(2): 159–216.

Drake, John W., Brian Charlesworth, Deborah Charlesworth, and James F. Crow. 1998. "Rates of Spontaneous Mutation." *Genetics* 148(4): 1667–86.

Drummond, D. Allan, Alpan Raval, and Claus O. Wilke. 2006. "A Single Determinant Dominates the Rate of Yeast Protein Evolution." *Molecular Biology and Evolution* 23(2): 327–37.

Duffy, Siobain, Laura A Shackelton, and Edward C Holmes. 2008. "Rates of Evolutionary Change in Viruses: Patterns and Determinants." *Nature reviews. Genetics* 9(4): 267–76.

Dunham, J P, H E Simmons, E C Holmes, and a G Stephenson. 2014. "Analysis of Viral (zucchini Yellow Mosaic Virus) Genetic Diversity during Systemic Movement through a Cucurbita Pepo Vine." *Virus research* 191: 172–79.

Dunker, A.Keith, J.David Lawson, Celeste J Brown, Ryan M Williams, Pedro Romero, Jeong S Oh, Christopher J Oldfield, Andrew M Campen, Catherine M Ratliff, Kerry W Hipps, Juan Ausio, Mark S Nissen, Raymond Reeves, ChulHee Kang, Charles R Kissinger, Robert W Bailey, Michael D Griswold, Wah Chiu, Ethan C Garner, and Zoran Obradovic. 2001. "Intrinsically Disordered Protein." *Journal of Molecular Graphics and Modelling* 19(1): 26–59.

Dunker, K, E Garner, S Guilliot, P Romero, K Albrecht, Z Obradovic, Computer Science, C Kissinger, J E Villafranca, Agouron Pharmaceutical, and San Diego. 1998. "Protein Disorder and the Evolution of Molecular Recognition : Theory, Predictions and Observations." *Pac Symp Biocomput.*: 473–84.

Eigen, Manfred, and Peter Schuster. 1977. "The Hypercyde Emergence of the Hypercycle." *Die Naturwissenschaften* 64: 541–65.

Elena, S F, P Agudelo-Romero, P Carrasco, F M Codoñer, S Martín, C Torres-Barceló, and R Sanjuán. 2008. "Experimental Evolution of Plant RNA Viruses." *Heredity* 100(5): 478–83.

Elena, Santiago F, Stéphanie Bedhomme, Purificación Carrasco, José M Cuevas, Francisca De Iglesia, Guillaume Lafforgue, Jasna Lali, Àngels Pròsper, Nicolas Tromas, and Mark P Zwart. 2011. "The Evolutionary Genetics of Emerging Plant RNA Viruses." 24(3): 287–93.

Eskelin, Katri, Anders Hafrén, Kimmo I Rantalainen, and Kristiina Mäkinen. 2011. "Potyviral VPg Enhances Viral RNA Translation and Inhibits Reporter mRNA Translation in Planta." *Journal of virology* 85(17): 9210–21.

Espinoza-Fonseca, L. Michel. 2009. "Reconciling Binding Mechanisms of Intrinsically Disordered Proteins." *Biochemical and Biophysical Research Communications* 382(3): 479–82.

Estevan, Joan, Aramata Maréna, Caroline Callot, Séverine Lacombe, André Moretti, Carole Caranta, and Jean-Luc Gallois. 2014. "Specific Requirement for Translation Initiation Factor 4E or Its Isoform Drives Plant Host Susceptibility to Tobacco Etch Virus." *BMC plant biology* 14: 67.

Fabre, Frédéric, Claude Bruchou, Alain Palloix, and Benoît Moury. 2009. "Key Determinants of Resistance Durability to Plant Viruses: Insights from a Model Linking within- and between-Host Dynamics." *Virus research* 141(2): 140–49.

Fabre, Frédéric, Benoît Moury, Elisabeth Ida Johansen, Vincent Simon, Mireille Jacquemond, and Rachid Senoussi. 2014. "Narrow Bottlenecks Affect Pea Seedborne Mosaic Virus Populations during Vertical Seed Transmission but Not during Leaf Colonization." *PLoS pathogens* 10(1): e1003833.

Fares, Mario a. 2015. "The Origins of Mutational Robustness." *Trends in Genetics* 31(7): 373–81.

Ferrada, Evandro, and Andreas Wagner. 2008. "Protein Robustness Promotes Evolutionary Innovations on Large Evolutionary Time-Scales." *Proceedings. Biological sciences / The Royal Society* 275(1643): 1595–1602.

Finsterbusch, Tim, and Annette Mankertz. 2009. "Porcine Circoviruses-Small but Powerful." *Virus Research* 143(2): 177–83.

Forterre, Patrick. 2006. "The Origin of Viruses and Their Possible Roles in Major Evolutionary Transitions." *Virus Research* 117(1): 5–16.

Fraile, A, and S Sacristán. 2008. "A Quantitative Analysis of Complementation of Deleterious Mutants in Plant Virus Populations." 6: 195–200.

Frankham, Richard. 1995. "Effective Population Size/adult Population Size Ratios in Wildlife: A Review." *Genetical research* 66: 95–107.

French, Roy, and Drake C Stenger. 2003. "Evolution of Wheat Streak Mosaic Virus: Dynamics of Population Growth within Plants May Explain Limited Variation." *Annual review of phytopathology* 41: 199–214.

Fuxreiter, Monika, István Simon, Peter Friedrich, and Peter Tompa. 2004. "Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins." *Journal of Molecular Biology* 338(5): 1015–26.

Fuxreiter, Monika, Peter Tompa, and István Simon. 2007. "Local Structural Disorder Imparts Plasticity on Linear Motifs." *Bioinformatics (Oxford, England)* 23(8): 950–56.

Galea, Charles A., Yuefeng Wang, Sivashankar G. Sivakolundu, and Richard W. Kriwacki. 2008. "Regulation of Cell Division by Intrinsically Unstructured Proteins; Intrinsic Flexibility, Modularity and Signaling Conduits." *Biochemistry* 47(29): 7598–7609.

Gao, Zhihuan, Elisabeth Johansen, Samantha Eyers, Carole L Thomas, T H Noel Ellis, and Andrew J Maule. 2004. "The Potyvirus Recessive Resistance Gene, sbm1, Identifies a Novel Role for Translation Initiation Factor eIF4E in Cell-to-Cell Trafficking." *The Plant journal : for cell and molecular biology* 40(3): 376–85.

García-Arenal, F, P Palukaitis, and M Zaitlin. 1984. "Strains and Mutants of Tobacco Mosaic Virus Are Both Found in Virus Derived from Single-Lesion-Passaged Inoculum." *Virology* 132(1): 131–37.

García-Arenal, F., and a. Fraile. 2013. "Trade-Offs in Host Range Evolution of Plant Viruses." *Plant Pathology* 62(S1): 2–9.

Garcia-Arenal, Fernando, Aurora Fraile, and José M Malpica. 2001. "VARIABILITY AND GENETIC STRUCTURE OF PLANT VIRUS POPULATIONS." *Annual review of phytopathology* 39: 157:186.

Ghalwash, Mohamed F, a Keith Dunker, and Zoran Obradović. 2012. "Uncertainty Analysis in Protein Disorder Prediction." *Molecular bioSystems* 8(1): 381–91.

Gitlin, Leonid, Tzachi Hagai, Anthony LaBarbera, Mark Solovey, and Raul Andino. 2014. "Rapid Evolution of Virus Sequences in Intrinsically Disordered Protein Regions" ed. Daniel R. Perez. *PLoS Pathogens* 10(12): e1004529.

Goldhill, Daniel, Angela Lee, Elizabeth S C P Williams, and Paul E Turner. 2014. "Evolvability and Robustness in Populations of RNA Virus Φ6." *Frontiers in microbiology* 5(February): 35.

González-Jara, Pablo, Aurora Fraile, Tomás Canto, and Fernando García-Arenal. 2009. "The Multiplicity of Infection of a Plant Virus Varies during Colonization of Its Eukaryotic Host." *Journal of virology* 83(15): 7487–94.

Goodfellow, Ian, Yasmin Chaudhry, Ioanna Gioldasi, Andreas Gerondopoulos, Alessandro Natoni, Louisette Labrie, Jean-François Laliberté, and Lisa Roberts. 2005. "Calicivirus Translation Initiation Requires an Interaction between VPg and eIF 4 E." *EMBO reports* 6(10): 968–72.

Grzela, Renata, Ewa Szolajska, Christine Ebel, Dominique Madern, Adrien Favier, Izabela Wojtal, Wlodzimierz Zagorski, and Jadwiga Chroboczek. 2008. "Virulence Factor of Potato Virus Y, Genome-Attached Terminal Protein VPg, Is a Highly Disordered Protein." *The Journal of biological chemistry* 283(1): 213–21.

Gunasekaran, Kannan, Chung-Jung Tsai, Sandeep Kumar, David Zanuy, and Ruth Nussinov. 2003. "Extended Disordered Proteins: Targeting Function with Less Scaffold." *Trends in Biochemical Sciences* 28(2): 81–85.

Gutiérrez, Serafín, Michel Yvon, Elodie Pirolles, Eliza Garzo, Alberto Fereres, Yannis Michalakis, and Stéphane Blanc. 2012. "Circulating Virus Load Determines the Size of Bottlenecks in Viral Populations Progressing within a Host." *PLoS pathogens* 8(11): e1003009.

Habchi, Johnny, Peter Tompa, Sonia Longhi, and Vladimir N Uversky. 2014. "Introducing Protein Intrinsic Disorder." *Chemical reviews*.

Hall, J S, R French, G L Hein, T J Morris, and D C Stenger. 2001. "Three Distinct Mechanisms Facilitate Genetic Isolation of Sympatric Wheat Streak Mosaic Virus Lineages." *Virology* 282(2): 230–36.

Harrison, Bryan D. 2002. "Virus Variation in Relation to Resistance-Breaking in Plants." *Euphytica* 124(2): 181–92.

Haynes, C, C J Oldfield, F Ji, N Klitgord, M E Cusick, P Radivojac, V N Uversky, M Vidal, and L M Iakoucheva. 2006. "Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes." *PLoS Comput Biol* 2(8): e100.

Hébrard, E, Y Bessin, T Michon, Sonia Longhi, Vladimir N Uversky, F Delalande, A Van Dorsselaer, P Romero, J Walter, N Declerk, and D Fargette. 2009. "Intrinsic Disorder in Viral Proteins Genome-Linked: Experimental and Predictive Analyses." *Virology journal* 6:23.

Hébrard, Eugénie, Yannick Bessin, Thierry Michon, Sonia Longhi, Vladimir N Uversky, François Delalande, Alain Van Dorsselaer, Pedro Romero, Jocelyne Walter, Nathalie Declerck, and Denis Fargette. 2009. "Intrinsic Disorder in Viral Proteins Genome-Linked: Experimental and Predictive Analyses." *Virology journal* 6(1): 23.

Herold, Jens, Raul Andino, and San Francisco. 2001. "Circularization through a Protein – Protein Bridge." 7: 581–91.

Holt, Carl. 2013. "Unfolded Phosphopolypeptides Enable Soft and Hard Tissues to Coexist in the Same Organism with Relative Ease." *Current Opinion in Structural Biology* 23(3): 420–25.

Hosmillo, Myra, Yasmin Chaudhry, Deok-Song Kim, Ian Goodfellow, and Kyoung-Oh Cho. 2014. "Sapovirus Translation Requires an Interaction between VPg and the Cap Binding Protein eIF4E." *Journal of virology* 88(21): 12213–21.

Howard, Colin R, and Nicola F Fletcher. 2012. "Emerging Virus Diseases: Can We Ever Expect the Unexpected?" *Emerging Microbes & Infections* 1(12): e46.

Iakoucheva, Lilia M, Predrag Radivojac, Celeste J Brown, Timothy R O'Connor, Jason G Sikes, Zoran Obradovic, and a Keith Dunker. 2004. "The Importance of Intrinsic Disorder for Protein Phosphorylation." *Nucleic acids research* 32(3): 1037–49.

Ivanov, K I, K Eskelin, a Lõhmus, and K Mäkinen. 2014. "Molecular and Cellular Mechanisms Underlying Potyvirus Infection." *The Journal of general virology* 95(Pt 7): 1415–29.

Ivanyi-Nagy, R, L Davidovic, E W Khandjian, and J-L Darlix. 2005. "Disordered RNA Chaperone Proteins: From Functions to Disease." *Cellular and molecular life sciences : CMLS* 62(13): 1409–17.

Janzac, Berenger, Frederic Fabre, Alain Palloix, and Benoit Moury. 2009. "Constraints on Evolution of Virus Avirulence Factors Predict the Durability of Corresponding Plant Resistances." *Molecular plant pathology* 10(5): 599–610.

Jenner, Carol E, Xiaowu Wang, Fernando Ponz, and John a Walsh. 2002. "A Fitness Cost for Turnip Mosaic Virus to Overcome Host Resistance." *Virus Research* 86(1-2): 1–6.

Jiang, Jun, and Jean-François Laliberté. 2011. "The Genome-Linked Protein VPg of Plant Viruses-a Protein with Many Partners." *Current opinion in virology* 1(5): 347–54.

Johnson, Derrick E, Bin Xue, Megan D Sickmeier, Jingwei Meng, Marc S Cortese, Christopher J Oldfield, Tanguy Le, A.Keith Dunker, and Vladimir N Uversky. 2012. "High-Throughput Characterization of Intrinsic Disorder in Proteins from the Protein Structure Initiative." *Journal of Structural Biology* 180(1): 201–15.

Ke, Ruian, John Aaskov, Edward C. Holmes, and James O. Lloyd-Smith. 2013. "Phylodynamic Analysis of the Emergence and Epidemiological Impact of Transmissible Defective Dengue Viruses." *PLoS Pathogens* 9(2).

Kelly, Sharon M., Thomas J. Jess, and Nicholas C. Price. 2005. "How to Study Proteins by Circular Dichroism." *Biochimica et Biophysica Acta - Proteins and Proteomics* 1751(2): 119–39.

Khan, Mateen a, Hiroshi Miyoshi, Daniel R Gallie, and Dixie J Goss. 2008. "Potyvirus Genome-Linked Protein, VPg, Directly Affects Wheat Germ in Vitro Translation: Interactions with Translation Initiation Factors eIF4F and eIFiso4F." *The Journal of biological chemistry* 283(3): 1340–49.

Khan, Mateen a, Hiroshi Miyoshi, Sibnath Ray, Tomohide Natsuaki, Noriko Suehiro, and Dixie J Goss. 2006. "Interaction of Genome-Linked Protein (VPg) of Turnip Mosaic Virus with Wheat Germ Translation Initiation Factors eIFiso4E and eIFiso4F." *The Journal of biological chemistry* 281(38): 28002–10.

Kim, Philip M, Andrea Sboner, Yu Xia, and Mark Gerstein. 2008. "The Role of Disorder in Interaction Networks: A Structural Analysis." *Molecular systems biology* 4(179): 179.

Kim, Philip M., Long J. Lu, Yu Xia, and B. Gerstein Mark. 2006. "Relating Three-Dimensional Structures to Protein Networks Provides Evolutionary Insights." *Science* 603(February).

Kriwacki, Richard W, Ludger Hengst, Linda Tennant, Steven I Reed, and Peter E Wright. 1996. "Structural Studies of p2 lWa1CiPl / Sdil in the Free and Cdk2-Bound State : Conformational Disorder Mediates Binding Diversity." 93(October): 11504–9.

Kutnjak, Denis, Matevž Rupar, Ion Gutierrez-Aguirre, Tomaž Curk, Jan F Kreuze, and Maja Ravnikar. 2015. "Deep Sequencing of Virus Derived Small Interfering RNAs and RNA from Viral Particles Shows Highly Similar Mutational Landscape of a Plant Virus Population." *Journal of virology* 89(9): 4760–69.

Lalic, Jasna, M Cuevas, and Santiago F Elena. 2011. "Effect of Host Species on the Distribution of Mutational Fitness Effects for an RNA Virus." *PLoS genetics* 7(11).

Lauring, Adam S, Ashley Acevedo, Samantha B Cooper, and Raul Andino. 2012. "Codon Usage Determines the Mutational Robustness, Evolutionary Capacity, and Virulence of an RNA Virus." *Cell host & microbe* 12(5): 623–32.

Lauring, Adam S, and Raul Andino. 2010. "Quasispecies Theory and the Behavior of RNA Viruses." *PLoS pathogens* 6(7): e1001005.

Lauring, Adam S, Judith Frydman, and Raul Andino. 2013. "The Role of Mutational Robustness in RNA Virus Evolution." *Nature reviews. Microbiology* 11(5): 327–36.

Lecampion, Cécile, Carole Caranta, and Christophe Robaglia. 2009. "La Lutte Pour La Traduction : Utilisation Des Facteurs D ' Initiation de La Traduction eIF4E Par Les Phytovirus À ARN." 13(4): 223–32.

Léonard, S, D Plante, S Wittmann, N Daigneault, M G Fortin, and J F Laliberté. 2000. "Complex Formation between Potyvirus VPg and Translation Eukaryotic Initiation Factor 4E Correlates with Virus Infectivity." *Journal of virology* 74(17): 7730–37.

Macdonald, Andrew, and Mark Harris. 2004. "Hepatitis C Virus NS5A: Tales of a Promiscuous Protein." *The Journal of general virology* 85(Pt 9): 2485–2502.

Mäkinen, Kristiina, and Anders Hafrén. 2014. "Intracellular Coordination of Potyviral RNA Functions in Infection." *Frontiers in Plant Science* 5(March): 1–12.

Malpica, José M, Aurora Fraile, Ignacio Moreno, Clara I Obies, John W Drake, and Fernando García-Arenal. 2002. "The Rate and Character of Spontaneous Mutation in an RNA Virus." *Genetics* 162(4): 1505–11.

Michon, Thierry, Yannick Estevez, Jocelyne Walter, Sylvie German-Retana, and Olivier Le Gall. 2006. "The Potyviral Virus Genome-Linked Protein VPg Forms a Ternary Complex with the Eukaryotic Initiation Factors eIF4E and eIF4G and Reduces eIF4E Affinity for a mRNA Cap Analogue." *The FEBS journal* 273(6): 1312–22.

Miyoshi, H, H Okade, S Muto, N Suehiro, H Nakashima, K Tomoo, and T Natsuaki. 2008. "Turnip Mosaic Virus VPg Interacts with Arabidopsis Thaliana eIF(iso)4E and Inhibits in Vitro Translation." *Biochimie* 90(10): 1427–34.

Mohan, Amrita, William J Sullivan, Predrag Radivojac, a Keith Dunker, and Vladimir N Uversky. 2008. "Intrinsic Disorder in Pathogenic and Non-Pathogenic Microbes: Discovering and Analyzing the Unfoldomes of Early-Branching Eukaryotes." *Molecular bioSystems* 4(4): 328–40.

Monsion, Baptiste, Rémy Froissart, Yannis Michalakis, and Stéphane Blanc. 2008. "Large Bottleneck Size in Cauliflower Mosaic Virus Populations during Host Plant Colonization." *PLoS pathogens* 4(10): e1000174.

Montarry, Josselin, Juliette Doumayrou, Vincent Simon, and Benoìt Moury. 2011. "Genetic Background Matters: A Plant-Virus Gene-for-Gene Interaction Is Strongly Influenced by Genetic Contexts." *Molecular plant pathology* 12(9): 911–20.

Montville, Rebecca, Remy Froissart, Susanna K. Remold, Olivier Tenaillon, and Paul E. Turner. 2005. "Evolution of Mutational Robustness in an RNA Virus." *PLoS Biology* 3(11): 1939–45.

Moury, B, C Charron, B Janzac, V Simon, J L Gallois, A Palloix, and C Caranta. 2014. "Evolution of Plant Eukaryotic Initiation Factor 4E (eIF4E) and Potyvirus Genome-Linked Protein (VPg): A Game of Mirrors Impacting Resistance Spectrum and Durability." *Infection, genetics and evolution* 27: 472–80.

Moury, Benoît, Frédéric Fabre, Josselin Montarry, Bérenger Janzac, Valérie Ayme, and Alain Palloix. 2010. "L ' Adaptation Des Virus de Plantes Aux Résistances Variétales." *Virologie* 14(4): 227–39.

Moury, Benoît, Frédéric Fabre, and Rachid Senoussi. 2007. "Estimation of the Number of Virus Particles Transmitted by an Insect Vector." *Proceedings of the National Academy of Sciences of the United States of America* 104(45): 17891–96.

Moury, Benoît, Bérenger Janzac, Youna Ruellan, Vincent Simon, Mekki Ben Khalifa, Hatem Fakhfakh, Frédéric Fabre, and Alain Palloix. 2014. "Interaction Patterns between Potato Virus Y and eIF4E-Mediated Recessive Resistance in the Solanaceae." *Journal of virology* 88(17): 9799–9807.

Moury, Benoît, Caroline Morel, Elisabeth Johansen, Laurent Guilbaud, Sylvie Souche, Valérie Ayme, Carole Caranta, Alain Palloix, and Mireille Jacquemond. 2004. "Mutations in Potato Virus Y Genome-Linked Protein Determine Virulence toward Recessive Resistances in Capsicum Annuum and Lycopersicon Hirsutum." *Molecular plant-microbe interactions : MPMI* 17(3): 322–29.

Moury, Benoît, Caroline Morel, Elisabeth Johansen, and Mireille Jacquemond. 2002. "Evidence for Diversifying Selection in Potato Virus Y and in the Coat Protein of Other Potyviruses." *The Journal of general virology* 83(Pt 10): 2563–73.

Nakahara, Kenji S, Ryoko Shimada, Sun-Hee Choi, Haruko Yamamoto, Jun Shao, and Ichiro Uyeda. 2010. "Involvement of the P1 Cistron in Overcoming eIF4E-Mediated Recessive Resistance against Clover Yellow Vein Virus in Pea." *Molecular plant-microbe interactions : MPMI* 23(11): 1460–69.

Nicaise, ValÃ©rie. 2014. "Crop Immunity against Viruses: Outcomes and Future Challenges." *Frontiers in Plant Science* 5(November): 1–18.

Nicaise, Valérie, Sylvie German-retana, Raquel Sanjua, Marie-pierre Dubrana, Marianne Mazier, Brigitte Maisonneuve, Thierry Candresse, Carole Caranta, and Olivier Legall. 2003. "The Eukaryotic Translation Initiation Factor 4E Controls Lettuce Susceptibility to the Potyvirus Lettuce Mosaic Virus 1." *Plant physiology* 132(July): 1272–82.

Nilsson, Johan, Mats Grahn, and Anthony P H Wright. 2011. "Proteome-Wide Evidence for Enhanced Positive Darwinian Selection within Intrinsically Disordered Regions in Proteins." *Genome biology* 12(7): R65.

Van Nimwegen, Erik. 2006. "Epidemiology. Influenza Escapes Immunity along Neutral Networks." *Science (New York, N.Y.)* 314(5807): 1884–86.

Nomotot, Akio, Naomi Kitamura, Fred Golini, and Eckard Wimmerf. 1977. "The 5 ′ -Terminal Structures of Poliovirion RNA and Poliovirus mRNA Differ Only in the Genome-Linked Protein VPg *." 74(12): 5345–49.

Oldfield, Christopher J, and a Keith Dunker. 2014. "Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions." *Annual review of biochemistry* (February): 1–32.

Oldfield, Christopher J, Jingwei Meng, Jack Y Yang, Mary Qu Yang, Vladimir N Uversky, and a Keith Dunker. 2008. "Flexible Nets: Disorder and Induced Fit in the Associations of p53 and 14-3-3 with Their Partners." *BMC genomics* 9 Suppl 1: S1.

Orton, Richard J, Caroline F Wright, Marco J Morelli, Nicholas Juleff, Gaël Thébaud, Nick J Knowles, Begoña Valdazo-González, David J Paton, Donald P King, and Daniel T Haydon. 2013. "Observing Micro-Evolutionary Processes of Viral Populations at Multiple Scales." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 368(1614): 20120203.

Pancsa, Rita, and Peter Tompa. 2012. "Structural Disorder in Eukaryotes." *PloS one* 7(4): e34687.

Pechmann, Sebastian, and Judith Frydman. 2014. "Interplay between Chaperones and Protein Disorder Promotes the Evolution of Protein Networks." *PLoS computational biology* 10(6): e1003674.

Peng, Zhenling, Jing Yan, Xiao Fan, Marcin J Mizianty, Bin Xue, Kui Wang, Gang Hu, Vladimir N Uversky, and Lukasz Kurgan. 2014. "Exceptionally Abundant Exceptions: Comprehensive Characterization of Intrinsic Disorder in All Domains of Life." *Cellular and molecular life sciences : CMLS*.

Pepin, Kim M, Sandra Lass, Juliet R C Pulliam, Andrew F Read, and James O Lloyd-Smith. 2010. "Identifying Genetic Markers of Adaptation for Surveillance of Viral Host Jumps." *Nature reviews. Microbiology* 8(11): 802–13.

Pfeiffer, Julie K., and Karla Kirkegaard. 2005. "Increased Fidelity Reduces Poliovirus Fitness and Virulence under Selective Pressure in Mice." *PLoS Pathogens* 1(2): 0102–10.

Philippe, Nadège, Matthieu Legendre, Gabriel Doutre, Olivier Poirot, Magali Lescot, Defne Arslan, Virginie Seltzer, Lionel Bertaux, Christophe Bruley, and Jean-michel Claverie. 2013. "Pandoraviruses : Amoeba Viruses." *Science (New York, N.Y.)* 341(July): 281–86.

Pursell, Natalie W., Parul Mishra, and Daniel N a Bolon. 2012. "Solubility-Promoting Function of Hsp90 Contributes to Client Maturation and Robust Cell Growth." *Eukaryotic Cell* 11(8): 1033–41.

Pushker, Ravindra, Catherine Mooney, Norman E Davey, Jean-Marc Jacqué, and Denis C Shields. 2013. "Marked Variability in the Extent of Protein Disorder within and between Viral Families." *PloS one* 8(4): e60724.

Rantalainen, Kimmo I, Vladimir N Uversky, Perttu Permi, Nisse Kalkkinen, A. keith Dunker, and Kristina Makinen. 2008. "Potato Virus A Genome-Linked Protein VPg Is an Intrinsically Disordered Molten Globule-like Protein with a Hydrophobic Core." 377(2): 280–88.

Receveur-Bréchot, Véronique, and David Karlin. 2005. "Le Désordre, Clé de Voûte Des Protéines." *La Recherche* 387: 52–55.

Revers, Frédéric, and Juan-antonio García. 2015. "Molecular Biology of Potyviruses." In *Advances in Virus Research, Vol.92*, , 101–99.

Robaglia, Christophe, and Carole Caranta. 2006. "Translation Initiation Factors: A Weak Link in Plant RNA Virus Infection." *Trends in plant science* 11(1): 40–45.

Romero, Pedro, Zoran Obradovic, Xiaohong Li, Ethan C Garner, Celeste J Brown, and A Keith Dunker. 2001. "Sequence Complexity of Disordered Protein." 48(August 2000): 38–48.

Rorick, Mary M., and Günter P. Wagner. 2011. "Protein Structural Modularity and Robustness Are Associated with Evolvability." *Genome Biology and Evolution* 3(1): 456–75.

Roudet-Tavert, G, T Michon, J Walter, T Delaunay, E Redondo, and O Le Gall. 2007. "Central Domain of a Potyvirus VPg Is Involved in the Interaction with the Host Translation Initiation Factor eIF4E and the Viral Protein HcPro." *The Journal of general virology* 88(Pt 3): 1029–33.

Ruffel, Sandrine, Marie-Hélène Dussault, Alain Palloix, Benoît Moury, Abdelhafid Bendahmane, Christophe Robaglia, and Carole Caranta. 2002. "A Natural Recessive Resistance Gene against Potato Virus Y in Pepper Corresponds to the Eukaryotic Initiation Factor 4E (eIF4E)." *The Plant journal : for cell and molecular biology* 32(6): 1067–75.

Sacrista, Soledad, Jose M. Malpica, Aurora Fraile, and Fernando García-Arenal. 2003. "Estimation of Population Bottlenecks during Systemic Movement of Tobacco Mosaic Virus in Tobacco Plants." *Journal of virology* 77(18): 9906–11.

Sanjuán, Rafael. 2010. "Mutational Fitness Effects in RNA and Single-Stranded DNA Viruses: Common Patterns Revealed by Site-Directed Mutagenesis Studies." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365(1548): 1975–82.

Sanjuán, Rafael. 2012. "From Molecular Genetics to Phylodynamics: Evolutionary Relevance of Mutation Rates across Viruses." *PLoS Pathogens* 8(5): 1–5.

Sanjuán, Rafael, Miguel R Nebot, Nicola Chirico, Louis M Mansky, and Robert Belshaw. 2010. "Viral Mutation Rates." *Journal of virology* 84(19): 9733–48.

Schadt, Eric E., Steve Turner, and Andrew Kasarskis. 2010. "A Window into Third-Generation Sequencing." *Human Molecular Genetics* 19(R2): 227–40.

Schlessinger, Avner, Christian Schaefer, Esmeralda Vicedo, Markus Schmidberger, Marco Punta, and Burkhard Rost. 2011. "Protein Disorder--a Breakthrough Invention of Evolution?" *Current opinion in structural biology* 21(3): 412–18.

Scholthof, Karen-Beth G, Scott Adkins, Henryk Czosnek, Peter Palukaitis, Emmanuel Jacquot, Thomas Hohn, Barbara Hohn, Keith Saunders, Thierry Candresse, Paul Ahlquist, Cynthia Hemenway, and Gary D Foster. 2011. "Top 10 Plant Viruses in Molecular Plant Pathology." *Molecular plant pathology* 12(9): 938–54.

Seifert, D., F. Di Giallonardo, K. J. Metzner, H. F. Gunthard, and N. Beerenwinkel. 2014. "A Framework for Inferring Fitness Landscapes of Patient-Derived Viruses Using Quasispecies Theory." *Genetics* 199(1): 191–203.

Shaw, John G. 1999. "Tobacco Mosaic Virus and the Study of Early Events in Virus Infections." (1886).

Sickmeier, Megan, Justin a Hamilton, Tanguy LeGall, Vladimir Vacic, Marc S Cortese, Agnes Tantos, Beata Szabo, Peter Tompa, Jake Chen, Vladimir N Uversky, Zoran Obradovic, and a Keith Dunker. 2007. "DisProt: The Database of Disordered Proteins." *Nucleic acids research* 35(Database issue): D786–93.

Simmons, H E, J P Dunham, J C Stack, B J a Dickins, I Pagán, E C Holmes, and a G Stephenson. 2012. "Deep Sequencing Reveals Persistence of Intra- and Inter-Host Genetic Diversity in Natural and Greenhouse Populations of Zucchini Yellow Mosaic Virus." *The Journal of general virology* 93(Pt 8): 1831–40.

Sorel, Maud. 2013. "La Protéine D'inclusion Cylindrique (CI) : Un Facteur Viral Clé Dans L'adaptation Du LMV À La Résistance Conférée Par eIF4E Chez La Laitue."

Soskine, Misha, and Dan S Tawfik. 2010. "Mutational Effects and the Evolution of New Protein Functions." *Nature reviews. Genetics* 11(8): 572–82.

Stein, Nils, Dragan Perovic, Jochen Kumlehn, Bettina Pellio, Silke Stracke, Stefan Streng, Frank Ordon, and Andreas Graner. 2005. "The Eukaryotic Translation Initiation Factor 4E Confers Multiallelic Recessive Bymovirus Resistance in Hordeum Vulgare (L.)." *The Plant journal : for cell and molecular biology* 42(6): 912–22.

Suttle, Curtis A. 2007. "Marine Viruses-Major Players in the Global Ecosystem." *Nature reviews. Microbiology* 5: 801–12.

Tavert-Roudet, G, A Abdul-Razzak, B Doublet, J Walter, T Delaunay, S German-Retana, T Michon, O Le Gall, and T Candresse. 2012. "The C Terminus of Lettuce Mosaic Potyvirus Cylindrical Inclusion Helicase Interacts with the Viral VPg and with Lettuce Translation Eukaryotic Initiation Factor 4E." *The Journal of general virology* 93: 184–93.

Tokuriki, Nobuhiko, Christopher J Oldfield, Vladimir N Uversky, Igor N Berezovsky, and Dan S Tawfik. 2009. "Do Viral Proteins Possess Unique Biophysical Features?" *Trends in biochemical sciences* 34(2): 53–59.

Tokuriki, Nobuhiko, Francois Stricher, Luis Serrano, and Dan S. Tawfik. 2008. "How Protein Stability and New Functions Trade off." *PLoS Computational Biology* 4(2): 35–37.

Tokuriki, Nobuhiko, and Dan S Tawfik. 2009a. "Protein Dynamism and Evolvability." *Science (New York, N.Y.)* 324(April): 203–7.

Tokuriki, Nobuhiko, and Dan S Tawfik. 2009b. "Stability Effects of Mutations and Protein Evolvability." *Current opinion in structural biology* 19(5): 596–604.

Toll-Riera, Macarena, David Bostick, M Mar Albà, and Joshua B Plotkin. 2012. "Structure and Age Jointly Influence Rates of Protein Evolution." *PLoS computational biology* 8(5): e1002542.

Tompa, Peter, and Peter Csermely. 2004. "The Role of Structural Disorder in the Function of RNA and Protein Chaperones." *The FASEB journal* 18(11): 1169–75.

Tromas, Nicolas, Mark P Zwart, Maïté Poulain, and Santiago F Elena. 2014. "Estimation of the in Vivo Recombination Rate for a Plant RNA Virus." *The Journal of general virology* 95(Pt 3): 724–32.

Truniger, Verónica, and Miguel A Aranda. 2009. "Recessive Resistance to Plant Viruses." *Advances in Virus Research* 75: 119–59.

Uversky, V N. 2002. "Natively Unfolded Proteins: A Point Where Biology Waits for Physics." *European Journal of Biochemistry* 12(May 2001): 2–12.

Uversky, Vladimir N. 2013. "The Most Important Thing Is the Tail: Multitudinous Functionalities of Intrinsically Disordered Protein Termini." *FEBS letters* 587(13): 1891–1901.

Uversky, Vladimir N, and a Keith Dunker. 2012. "Multiparametric Analysis of Intrinsically Disordered Proteins: Looking at Intrinsic Disorder through Compound Eyes." *Analytical chemistry* 84(5): 2096–2104.

Uversky, Vladimir N, Joel R Gillespie, and Anthony L Fink. 2000. "Why Are ' Natively Unfolded ' Proteins Unstructured Under Physiologic Conditions ?" 427(July): 415–27.

Uversky, Vladimir N., and A. Keith Dunker. 2010. "Understanding Protein Non-Folding." *Biochim Biophys Acta* 1804(6): 1231–64.

Uversky, Vladimir N., Christopher J. Oldfield, and a. Keith Dunker. 2005. "Showing Your ID: Intrinsic Disorder as an ID for Recognition, Regulation and Cell Signaling." *Journal of Molecular Recognition* 18(5): 343–84.

Vignuzzi, Marco, Jeffrey K Stone, Jamie J Arnold, Craig E Cameron, and Raul Andino. 2006. "Quasispecies Diversity Determines Pathogenesis through Cooperative Interactions in a Viral Population." *Nature* 439(7074): 344–48.

Vucetic, Slobodan, Hongbo Xie, Lilia M. Iakoucheva, Christopher J. Oldfield, A. Keith Dunker, Zoran Obradovic, and Vladimir N. Uversky. 2007. "Functional Anthology of Intrinsic Disorder. II. Cellular Components, Domains, Technical Terms, Developmental Processes and Coding Sequence Diversities Correlated with Long Disordered Regions." *Journal of Proteome research* 6(5): 1899–1916.

Wagner, Andreas. 2008. "Neutralism and Selectionism: A Network-Based Reconciliation." *Nature reviews. Genetics* 9(12): 965–74.

Wang, Aiming, and Sowmya Krishnaswamy. 2012. "Eukaryotic Translation Initiation Factor 4E-Mediated Recessive Resistance to Plant Viruses and Its Utility in Crop Improvement." 13: 795–803.

Ward, CW, and DD Shukla. 1991. "Taxonomy of Potyviruses: Current Problems and Some Solutions." *Intervirology* 32: 269–96.

Ward, J J, J S Sodhi, L J McGuffin, B F Buxton, and D T Jones. 2004. "Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life." *Journal of molecular biology* 337(3): 635–45.

Wayne, Natalie, and Daniel N Bolon. 2010. "Charge-Rich Regions Modulate the Anti-Aggregation Activity of Hsp90." *Journal of molecular biology* 401(5): 931–39.

Wittmann, S, H Chatel, M G Fortin, and J F Laliberté. 1997. "Interaction of the Viral Protein Genome Linked of Turnip Mosaic Potyvirus with the Translational Eukaryotic Initiation Factor (iso) 4E of Arabidopsis Thaliana Using the Yeast Two-Hybrid System." *Virology* 234(1): 84–92.

Woese, C R, O Kandler, and M L Wheelis. 1990. "Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya." *Proceedings of the National Academy of Sciences of the United States of America* 87(12): 4576–79.

Woody, Robert W. 2010. "Circular Dichroism of Intrinsically Disordered Proteins." In *Instrumental Analysis of Intrinsically Disorered Proteins : Assessing Structure and Conformation*, eds. Vladimir N. Uversky and Sonia Longhi. , 303–21.

Woolhouse, Mark E J, Kyle Adair, and Liam Brierley. 2013. "RNA Viruses : A Case Study of the Biology of Emerging Infectious Diseases." : 1–11.

Wright, Caroline F, Marco J Morelli, Gaël Thébaud, Nick J Knowles, Pawel Herzyk, David J Paton, Daniel T Haydon, and Donald P King. 2011. "Beyond the Consensus: Dissecting within-Host Viral Population Diversity of Foot-and-Mouth Disease Virus by Using next-Generation Genome Sequencing." *Journal of virology* 85(5): 2266–75.

Wright, P E, and H J Dyson. 1999. "Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm." *Journal of molecular biology* 293(2): 321–31.

Xiao, Allen, Jerry Wong, and Honglin Luo. 2010. "Viral Interaction with Molecular Chaperones: Role in Regulating Viral Infection." *Archives of Virology* 155(7): 1021–31.

Xie, Hongbo, Slobodan Vucetic, Lilia M Iakoucheva, Christopher J Oldfield, A Keith Dunker, Zoran Obradovic, and Vladimir N Uversky. 2007. "Functional Anthology of Intrinsic Disorder . 3 . Ligands , Post-Translational Modifications , and Diseases Associated with Intrinsically Disordered Proteins Research Articles." : 1917–32.

Xie, Hongbo, Slobodan Vucetic, Lilia M. Iakoucheva, Christopher J. Oldfield, A. Keith Dunker, Vladimir N. Uversky, and Zoran Obradovic. 2008. "Functional Anthology of Intrinsic Disorder. I. Biological Processes and Functions of Proteins with Long Disordered Regions." *Journal of Proteome research* 6(5): 1882–98.

Xue, Bin, A Keith Dunker, and Vladimir N Uversky. 2012. "Orderly Order in Protein Intrinsic Disorder Distribution : Disorder in 3500 Proteomes from Viruses and the Three Domains of Life." *Journal of Biomolecular Structure and Dynamics* 30(2): 137–49.

Zwart, Mark P., José-Antonio Daròs, and Santiago F. Elena. 2011. "One Is Enough: In Vivo Effective Population Size Is Dose-Dependent for a Plant RNA Virus" ed. Peter D. Nagy. *PLoS Pathogens* 7(7): e1002122.

Annexes

Annexe 1. Ressources bioinformatiques utilisées au cours de la thèse

Nom du programme	Adresse URL
Dichroweb	http://dichroweb.cryst.bbk.ac.uk/html/
DISPHOS (Disorder-enhanced Phosphorylation Sites predictor)	http://www.dabi.temple.edu/disphos/
DisProt (The database of Protein Disorder)	http://www.disprot.org
ELM database (Eukaryotic linear motif ressource)	http://elm.eu.org/
FigTree	http://tree.bio.ed.ac.uk/software/figtree/
ICTV (International Committee for taxonomy of viruses)	http://www.ictvonline.org
MAFFT (Multiple Alignment using Fast Fourier Transform)	http://mafft.cbrc.jp/alignment/server/
MEGA6 (Molecular Evolutionary Genetics Analysis)	http://www.megasoftware.net/
MoRFPred (Molecular Recognition Feature predictor)	http://biomine-ws.ece.ualberta.ca/MoRFpred/
NCBI (National Center for Biotechnology Information)	http://www.ncbi.nlm.nih.gov
NEBase Changer	www.NEBaseChanger.neb.com
PARRIS (Partitioning Approach for Robust Inferences of Selection in recombining sequences)	http://www.datamonkey.org
Plotly	https://plot.ly/
PONDR-VLXT (Predictor of Naturally Disordered regions)	http://www.pondr.com/
PVS (Protein Variability Server)	http://imed.med.ucm.es/PVS/
Serial Cloner 2.6	http://serialbasics.free.fr/Serial_Cloner.html
Tm calculator	http://www.neb.com
UniProtKB (Universal Protein resource Knowledgebase)	http://www.uniprot.org/uniprot/
ViralZone	http://viralzone.expasy.org/

Annexe 2. Estimation des effets déstabilisants de substitutions en acides aminés sur la structure d'une protéine

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-2	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-1	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

***Figure A2.1*|Représentation de la matrice BLOcks SUbstitution Matrix 62 (BLOSUM62).** Des scores élevés traduisent une similitude de propriétés physicochimiques; A l'inverse, des scores négatifs traduisent un changement de propriété physicochimique.

	Conservative	Non-conservative
A	D, E, G, S, T	P, V
C	G, R, S, W, Y	F
D	A, E, G, H, N, V, Y	
E	A, D, G, K, Q, V	
F	I, L, Y	C, S, V
G	A, C, D, E, R	S, V, W
H	D, L, N, P, Q, R, Y	
I	F, L, M, N, V	K, R, S, T
K	E, M, N, Q, R, T	I
L	F, H, I, M, P, Q, R, V, W	S
M	I, K, L, R, T, V	
N	D, H, I, K, S, T, Y	
P	H, L, Q, R, S	A, T
Q	E, H, K, L, P, R	
R	C, G, H, K, L, M, P, Q, T, W	I, S
S	A, C, N, P, T, W, Y	F, G, I, L, R
T	A, K, M, N, R, S	I, P
V	D, E, I, L, M,	A, F, G
W	C, L, R, S	G
Y	C, D, F, H, N, S	

***Figure A2.2*|Classification des substitutions en acides aminés selon leur nature conservative ou non-conservative.** Les substitutions qui impliquent des changements en résidus de propriétés physico-chimiques semblables sont dites conservatives alors que celles impliquant un changement physico-chimique important (changement de charge, polaire/apolaire...) sont dites non-conservatives. *D'après Pechmann et al. (2014).*

Annexe 3. Séquences nucléotidiques et protéiques

- **VPg SON41p** (GenBank accession = AJ439544)

Séquence nucléotidique :

GGGAAAAATAAATCCAAGAGAATTCAAGCCTTGAAGTTTCGCCATGCTCGCGACAAGAGGGCTGGCTTCGAAATCGACAA
TAATGATGACACAATAGAGGAATTCTTTGGATCTGCATACAGGAAGAAAGGAAAAGGTAAAGGCACCACAGTTGGTATGG
GCAAGTCGAGTAGGAGATTCATCAACATGTATGGATTTGATCCAACAGAGTACTCATTCATCCAATTCGTTGATCCACTC
ACTGGAGCACAAATAGAAGAGAATGTCTATGCTGACATTAGGGATATTCAAGAAAGATTTAGTGAAGTGCGAAGGAAGAT
GGTTGAGGATGACGAAATAGAAACGCAAGCCTTGGATAGTCACACGAGCATACATGCTTACTTCAGGAAAGATTGGTCTG
ACAAAGCTTTGAAGATCGACTTAATGCCACATAACCCACTTAAGGTTTGTGACAAAACAAATGGCATTGCTAAATTCCCT
GAGAGAGAGTTTGAACTGAGGCAGACTGGACCAGCTGTGGAAGTTAATGTGAAGGACATACCAAAACAGGAAGTAGAGCA
TGAA

Séquence protéique :

GKNKSKRIQALKFRHARDKRAGFEIDNNDDTIEEFFGSAYRKKGKGKGTTVGMGKSSRRFINMYGFDPTEYSFIQFVDPLTGAQIEENV
YADIRDIQERFSEVRRKMVEDDEIETQALDSHTSIHAYFRKDWSDKALKIDLMPHNPLKVCDKTNGIAKFPEREFELRQTGPAVEVNVK
DIPKQEVEHE

- **eIF4E-Yolo Wonder** (GenBank accession = AAN74644.1)

Séquence nucléotidique :

ATGGCAACAGCTGAAATGGAGAAAACGACGACGTTTGATGAAGCTGAGAAGGTGAAATTGAATGCTAATGAGGCAGATGA
TGAAGTTGAAGAAGGTGAAATTGTTGAAGAAACTGATGATACGACGTCGTATTTGAGCAAAGAAATAGCAACAAAGCATC
CATTAGAGCATTCATGGACTTTCTGGTTTGATAATCCAGTGGCGAAATCGAAACAAGCTGCTTGGGGTAGCTCGCTTCGC
AACGTCTACACTTTCTCCACTGTTGAAGATTTTTGGGGTGCTTACAATAATATCCACCACCCAAGCAAGTTAGTTGTGGG
AGCAGACTTACATTGTTTCAAGCATAAAATTGAGCCAAAGTGGGAAGATCCTGTATGTGCCAATGGAGGGACATGGAAAA
TGAGTTTTTCAAAGGGTAAATCTGATACCAGCTGGCTATATACGCTGCTTGCAATGATTGGACATCAATTCGATCATGAA
GATGAAATTTGTGGAGCAGTAGTTAGTGTCAGAGGTAAGGGAGAAAAAATATCTTTGTGGACCAAGAATGCTGCAAATGA
AACGGCTCAGGTTAGCATTGGTAAGCAATGGAAGCAGTTTCTGGATTACAGCGACAGTGTTGGCTTCATATTTCACGACG
ATGCAAAGAGGCTCGACAGAAATGCAAAGAATCGTTACACAGTATAA

Séquence protéique :

MATAEMEKTTTFDEAEKVKLNANEADDEVEEGEIVEETDDTTSYLSKEIATKHPLEHSWTFWFDNPVAKSKQAAWGSSLRNVYTFSTVE
DFWGAYNNIHHPSKLVVGADLHCFKHKIEPKWEDPVCANGGTWKMSFSKGKSDTSWLYTLLAMIGHQFDHEDEICGAVVSVRGKGEKIS
LWTKNAANETAQVSIGKQWKQFLDYSDSVGFIFHDDAKRLDRNAKNRYTV

- **eIF4E-HD** (GenBank accession = AAV88614.1)

Séquence nucléotidique :

ATGGCAACAGCTGAAATGGAGAAAACGACGACGTTTGATGAAGCTGAGAAGGTGAAATTGAATGCTAATGAGGCAGATGA
TGAAGTTGAAGAAGGTGAAATTGTTGAAGAAACTGATGATACGACGTCGTATTTGAGCAAAGAAATAGCAACAAAGCATC
CATTAGAGCATTCATGGACTTTCTGGTTTGATAATCCAGAGGCGAAATCGAAACAAGCTGCTTGGGGTAGCTCGCTTCGC
AACGTCTACACTTTCTCCACTGTTGAAGATTTTTGGGGTGCTTACAATAATATCCACCACCCAAGCAAGTTAGTTGTGGG
AGCAGACTTACATTGTTTCAAGCATAAAATTGAGCCAAAGTGGGAAGATCCTGTATGTGCCAATGGAGGGACATGGAAAA
TGAGTTTTTCAAAGGGTAAATCTGATACCAGCTGGCTATATACGCTGCTTGCAATGATTGGACATCAATTCGATCATGAA
GATGAAATTTGTGGAGCAGTAGTTAGTGTCAGAGGTAAGGGAGAAAAAATATCTTTGTGGACCAAGAATGCTGCAAATGA
AACGGCTCAGGTTAGCATTGGTAAGCAATGGAAGCAGTTTCTGGATTACAGCGGCAGTGTTGGCTTCATATTTCACGACG
ATGCAAAGAGGCTCGACAGAAATGCAAAGAATCGTTACACAGTATAA

Séquence protéique :

MATAEMEKTTTFDEAEKVKLNANEADDEVEEGEIVEETDDTTSYLSKEIATKHPLEHSWTFWFDNPEAKSKQAAWGSSLRNVYTFSTVE
DFWGAYNNIHHPSKLVVGADLHCFKHKIEPKWEDPVCANGGTWKMSFSKGKSDTSWLYTLLAMIGHQFDHEDEICGAVVSVRGKGEKIS
LWTKNAANETAQVSIGKQWKQFLDYSGSVGFIFHDDAKRLDRNAKNRYTV

Annexe 4. Séquences des amorces utilisées au cours de la thèse

***Tableau A4.1*|Amorces utilisées pour la construction des mutants VPg ID1, ID2 et OD2 par la stratégie de mutagénèse dirigée.** *Ta : température d'hybridation des amorces.*

Mutant Name	Amino acid substitutions	Corresponding Nucleotide substitutions	Mutator Primers sequences	Ta (°C)
ID1	R94G + D95E + F100L	A280G + T285G + T300A	Fw 5' AGAAAGATTAAGTGAAGTGCGAAGGAAG Rev 5' TGAATCTCCCCAATGTCAGCATAGACATTCTC	60
ID2	R94G + D95N	A280G + T285A	Fw 5' TGCTGACATTGGGAATATTCAAGAAAG Rev 5' TAGACATTCTCTTCTATTTGTG	58
OD2	E102Q + E114Q	G304C + G340C	Fw 5' TTGAGGATGACGAAATACAAACGCAAGCCTTGGATAG Rev 5' CCATCTTCCTTCGCACTTGACTAAATCTTTCTTGAATATCCCTAATG	62

***Tableau A4.2*|Amorces utilisées pour le clonage des gènes VPg et eIF4E dans les vecteurs pDONR201 et pETGB1a**. Les parties en rouge représentent les sites de clonage ajoutés en 5' de la séquence des gènes. Les codons STOP sont représentés en gras et surlignés.

Primer name	Sequence	Product size
PCR attB		
eIF4E-attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCAACAGCTGAAATGGAGAAAA	748
eIF4E-attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTC**TTA**TACTGTGTAACGATTCTTTGCA	
VPgSON41p-attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGGGAAAAATAAATCCAAGAGAATTC	628
VPgSON41p-attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTC**CTA**TTCATGCTCTACTTCCTGTTTTGGT	
PCR NcoI-BamHI		
4E-NcoI	GGGGCCATGGGCATGGCAACAGCTGAAATGGAGAAA	711
4E-BamHI	CGCGGATCCGCGTTATACTGTGTAACG	
PCR ORF eIF4E		
4E-Fw	ATGGCAACAGCTGAAATGGA	687
4E-Rev	TTATACTGTGTAACGATTCTTTG	
PCR attL		
attL1	TCGCGTTAACGCTAGCATGGATCTC	
attL2	GTAACATCAGAGATTTTGAGACAC	

***Tableau A4.3*|Amorces utilisées pour la construction des clones cDNA complet de PVY par recombinaison dans la levure.**

Primer name	Sequence
PCR VPg Central domain	
VPgSON41p-CD(Fw)	GATTCATCAACATGTATGGATTTGATCCAACAG
VPgSON41p-CD(Rev)	GTGGGTTATGTGGCATTAAGTCGATCTTCA
PCR VPg ORF	
VPgSON41p-ORF(Fw)	GGTGGCTAAAGATTTGATCATAGCAGG
VPgSON41p-ORF(Rev)	GGCTGCAGTTTGCTATTATGTAAGCCC

***Tableau A4.4*|Amorces utilisées pour l'amplification des génomes viraux PVY.** *Ta : température d'hybridation.*

	Primers sequence (5'→3')	Product	Size (bp)	Ta used with Taq Phusion (°C)
Amplification of VPg				
PVY(SON41p)-VPgFw	GGTGGCTAAAGATTTGATCATAGCAGG	VPg cistron	803	**72**
PVY(SON41p)-VPgRev	GGCTGCAGTTTGCTATTATGTAAGCCC			
Amplification of PVY genome				
PVYfull-Fw(1)start	AAATTAAAACAACTCAATACAACATAAGAA	PVYfull-part1	4291	**62**
PVYfull-Rev(1)4271	TCAACCTCCCTTCCCACTGGG			
PVYfull-Fw(2)2893	CAACATGGCATTCATACAGAGCAAG	PVYfull-part2	4435	**68**
PVYfull-Rev(2)7395	TTCTTGCCACCATACATAGCTCC			
PVYfull-Fw(3)6300	TTGAGAGATTTCAATCCAATTGCCCAAACAG	PVYfull-part3	3072	**72**
PVYfull-RevCP	TGTCGACATTCACATGTTCTTGACTCC			
Nested PCR on amplifiat				
PVYfull-Fw5'UTR	CTCAAGCAATTTTCTAAGTTTCAGC	Nested-5'UTR	781	**64**
PVYfull-Rev5'UTR	CATTTATCTGTTCGAGCGAGACAT			
SeqPVYfull-Fw1	TGAATATCACTTTCAGATGGC	Nested-1	745	**56**
PVYfull-Rev1bis	GGGGCAAGTCATCTTATAGC			
SeqPVYfull-Fw2	TATTCATAGTGCGTGGATCA	Nested-2	759	**56**
PVYfull-Rev2	CCTCGAAGAAGTTTGAGAAA			
SeqPVYfull-Fw3	AAGGAAAAGAGAATACAGCTCA	Nested-3	763	**59**
PVYfull-Rev3	GGGTAGAAAATTCTCATTTGAG			
SeqPVYfull-Fw4	CAACCAAAAAGCACCTTG	Nested-4	754	**54**
PVYfull-Rev4	GCTGCTGACACTCGTAGA			
SeqPVYfull-Fw5	AGGAGGATGCATAATGTCTG	Nested-5	768	**56**
PVYfull-Rev5	GTGACAAAATTTGGCAAGC			
SeqPVYfull-Fw6	GATTTAACTTGGCGAGAAAA	Nested-6	796	**58**
PVYfull-Rev6	CCAAAACCGATGATGTTC			
SeqPVYfull-Fw8	CTTAGTGTGGCTGGATCTGT	Nested-8	756	**57**
PVYfull-Rev8	GATTCTTTCCCCATAGCTGA			
SeqPVYfull-Fw10	GGTCAATGCATCCTGTTATAC	Nested-10	1748	**58**
PVYfull-Rev12	CCAATGCTTCCAAAAAGT			
SeqPVYfull-Fw12	CACTCACTGGAGCACAAATA	Nested-12	758	**55**
PVYfull-Rev12	CCAATGCTTCCAAAAAGT			

***Tableau A4.5*|Amorces utilisées pour les séquençages complets en Sanger des génomes de PVY.**

	Primers sequence (5'→3')
Primers used for PVY sequencing	
SeqPVYfull-Fw1	TGAATATCACTTTCAGATGGC
SeqPVYfull-Fw2	TATTCATAGTGCGTGGATCA
SeqPVYfull-Fw3	AAGGAAAAGAGAATACAGCTCA
SeqPVYfull-Fw4	CAACCAAAAAGCACCTTG
SeqPVYfull-Fw5	AGGAGGATGCATAATGTCTG
SeqPVYfull-Fw6	GATTTAACTTGGCGAGAAAA
SeqPVYfull-Fw7	GTTAAGTTTGCACAAGCACA
SeqPVYfull-Fw8	CTTAGTGTGGCTGGATCTGT
SeqPVYfull-Fw9	GGCCAAACTCTTAACAGATAAA
SeqPVYfull-Fw10	GGTCAATGCATCCTGTTATAC
SeqPVYfull-Fw11	AAGTGAGAAGCCAATTGAAA
SeqPVYfull-Fw12	CACTCACTGGAGCACAAATA
SeqPVYfull-Fw13	CATGCATGGCACGTTTAG
SeqPVYfull-Fw14	AACAACTAAGCTTGTGCAGG
SeqPVYfull-Fw15	AATTGTTATGCAGAGCTGTTT
SeqPVYfull-Fw16	TGGATAACTCTCTTATGGTTGTT
SeqPVYfull-Fw17	GCACTGAGGAAACTATACATGG
SeqPVYfull-Fw18	AGTTTGGGTTATGATGGATG
SeqPVYfull-Rev1	TGTGGCTCAAACTTCGAT

Annexe 5. Présentation des vecteurs utilisés

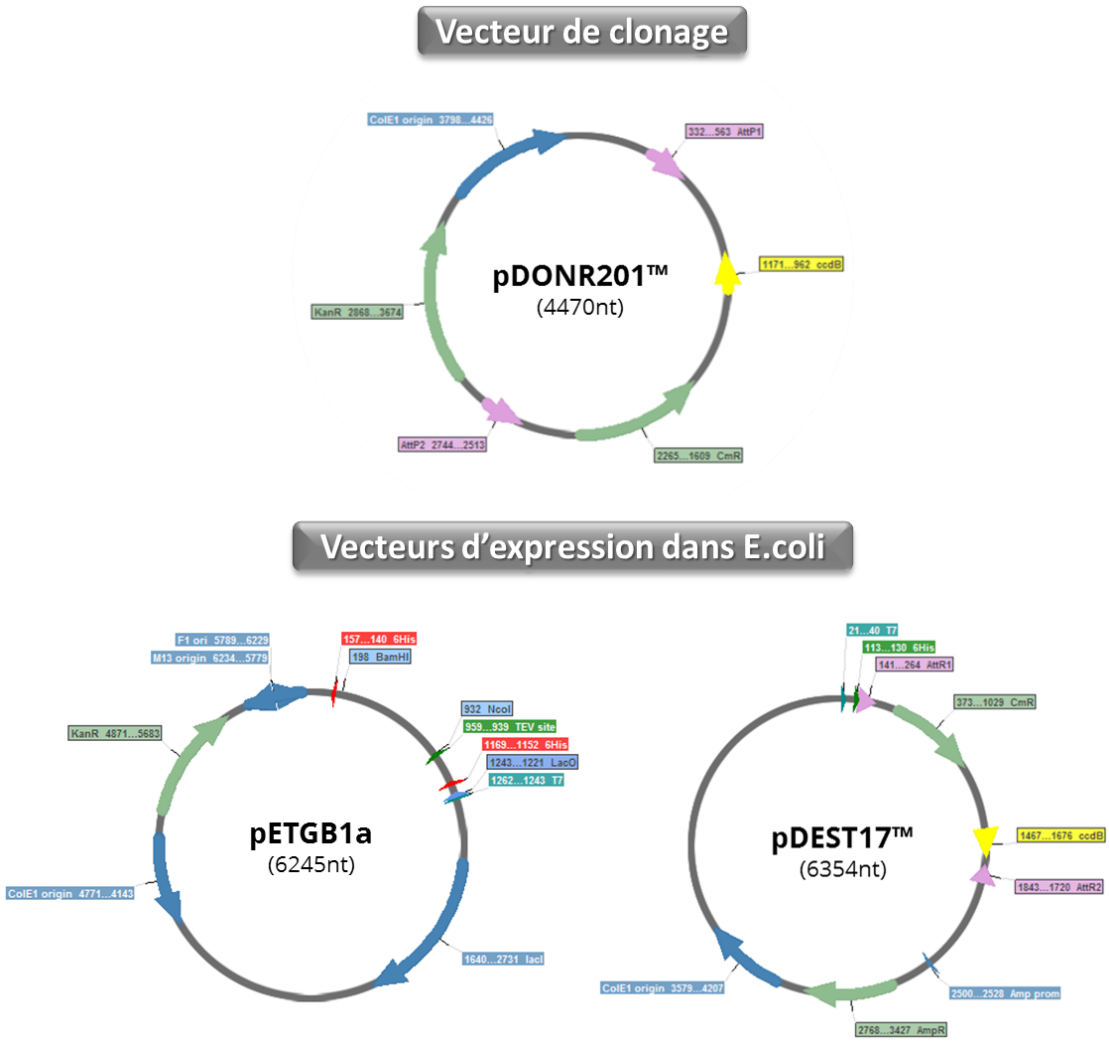


***Figure A5.1*|Carte des plasmides utilisés pour le clonage des VPg (pDONR201™) et l'expression des protéines eIF4E (pETGB1a) et des VPg (pDEST-17) dans *E.coli*.**

Tableau A5.1 **| Caractéristiques des différents vecteurs utilisés au cours de la thèse.**

Nom du clone	Vecteur utilisé	Application	Hôte	Gène de Sélection	Promoteur et conditions d'induction	Protéine exprimée	Disponible/Généré
Vecteurs Gateway							
pENTR VPg-SON41p	pDONR-201™	**Vecteurs d'entrée**	Bactérie (E.coli)	KanamycineR	-	-	Généré
pENTR VPg-SON41g					-	-	Généré
pENTR VPg-SON41p(ID1)					-	-	Généré
pENTR VPg-SON41p(ID2)					-	-	Généré
pENTR VPg-SON41p(OD2)					-	-	Généré
pENTR 4E-pvr2^{+}					-	-	Généré
pENTR 4E-pvr2^{3}					-	-	Généré
pEXP (His_6)VPg-SON41p	pDEST-17™	**Vecteur d'expression (His tag) pour purification sur Ni-NTA**	Bactérie (E.coli)	AmpicillineR	T7 (IPTG)	His_{x6}-VPg(SON41p)	Généré
pEXP (His_6)VPg-SON41p(ID1)						His_{x6}-VPg(ID1)	Généré
pEXP (His_6)VPg-SON41p(ID2)						His_{x6}-VPg(ID2)	Généré
pEXP (His_6)VPg-SON41p(OD2)						His_{x6}-VPg(OD2)	Généré
Vecteurs non Gateway							
pGEMT 4E-pvr2^{+}	pGEM®T	**Vecteur de clonage**	Bactérie (E.coli)	AmpicillineR	-	-	Fournis par JL Gallois
pGEMT 4E-pvr2^{3}	pGEM®T			AmpicillineR	-	-	Fournis par JL Gallois
pETGB1a (His_6+TEV)4E-pvr2^{+}	pETGB1a	**Purification sur Ni-NTA**		KanamycineR	Promoteur T7/lacO	Hisx6-TEV-4E(pvr2+)	Généré
pETGB1a (His_6+TEV)4E- pvr2^{3}	pETGB1a					Hisx6-TEV-4E(pvr2(3))	Généré
Clones ADNc complet PVY							
SON41p-pCAPM13-Trp2µΔVPg-CD	pCAPM13-Trp2µΔCD	**Infection virale *in planta***	Levure (S.cerevisiae) + Bactérie (E.coli) + Plante (C.annuum)	TRP(-) et AmpicillineR	T7 et ADH1	-	Fourni par Benoît Moury
SON41p-pCAPM13-Trp2µ	pCAPM13-Trp2µΔCD					Polyprotéine du PVY SON41p	Généré
SON41p-VPgID1	pCAPM13-Trp2µΔCD					Polyprotéine du PVY SON41p(ID1)	Généré
SON41p-VPgID2	pCAPM13-Trp2µΔCD					Polyprotéine du PVY SON41p(ID2)	Généré
SON41p-VPgOD2	pCAPM13-Trp2µΔCD					Polyprotéine du PVY SON41p(OD2)	Généré

SI ON MET UN PEU DE DÉSORDRE DANS UN ORDRE PARFAIT
C'EST TOUT DE SUITE LE BORDEL
PAR CONTRE, SI ON MET UN PEU D'ORDRE AU MILIEU DU DÉSORDRE
C'EST TOUJOURS LE BORDEL
ÇA VEUT DIRE QUE LE DÉSORDRE EST LE PLUS FORT