



**HAL**  
open science

## Comparative analysis of mutational robustness of the intrinsically disordered viral protein VPg and of its interactor eIF4E.

Jocelyne Walter, Justine Charon, Yihua Hu, Joy Lachat, Thomas Leger, Guillaume Lafforgue, Amandine Barra, Thierry Michon

### ► To cite this version:

Jocelyne Walter, Justine Charon, Yihua Hu, Joy Lachat, Thomas Leger, et al.. Comparative analysis of mutational robustness of the intrinsically disordered viral protein VPg and of its interactor eIF4E.. PLoS ONE, 2019, 14 (2), pp.e0211725. 10.1371/journal.pone.0211725 . hal-02628650

**HAL Id: hal-02628650**

**<https://hal.inrae.fr/hal-02628650>**

Submitted on 26 May 2020

**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.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

# Comparative analysis of mutational robustness of the intrinsically disordered viral protein VPg and of its interactor eIF4E

Jocelyne Walter<sup>1\*</sup>, Justine Charon<sup>1,2</sup>, Yihua Hu<sup>1</sup>, Joy Lachat<sup>1</sup>, Thomas Leger<sup>1</sup>, Guillaume Lafforgue<sup>1</sup>, Amandine Barra<sup>1</sup>, Thierry Michon<sup>1\*</sup>

**1** UMR Biologie du Fruit et Pathologie, INRA, Université de Bordeaux, CS, Villenave d'Ornon, France, **2** School of Life & Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

\* [jocelyne.walter@inra.fr](mailto:jocelyne.walter@inra.fr) (JW); [thierry.michon@inra.fr](mailto:thierry.michon@inra.fr) (TM)



## Abstract

Conformational intrinsic disorder is a feature present in many virus proteins. Intrinsically disordered regions (IDRs) have weaker structural requirement than ordered regions and mutations in IDRs could have a lower impact on the virus fitness. This could favor its exploration of adaptive solutions. The potyviral protein VPg contains IDRs with determinants for adaptation to its host plant. To experimentally assess whether IDRs are more resistant to mutations than ordered regions, the biologically relevant interaction between mutant libraries of both VPg and the eukaryotic translation initiation factor 4E (eIF4E) and their respective wild type partner was examined using yeast two hybrid assay. Our data shows that VPg is significantly more robust to mutations than eIF4E and as such belongs to a particular class of intrinsically disordered proteins. This result is discussed from the standpoint of IDRs involvement in the virus adaptive processes.

## OPEN ACCESS

**Citation:** Walter J, Charon J, Hu Y, Lachat J, Leger T, Lafforgue G, et al. (2019) Comparative analysis of mutational robustness of the intrinsically disordered viral protein VPg and of its interactor eIF4E. PLoS ONE 14(2): e0211725. <https://doi.org/10.1371/journal.pone.0211725>

**Editor:** Ron Geller, University of Valencia, SPAIN

**Received:** November 21, 2018

**Accepted:** January 20, 2019

**Published:** February 14, 2019

**Copyright:** © 2019 Walter et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** This work was supported by: YH, fellowship, Hangzhou Normal University, Hangzhou, China; JL, TL, internship special grants, University of Bordeaux, France; JC, PhD grant, Ministère de l'enseignement supérieur, de la recherche et de l'innovation, France.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Conformational intrinsic disorder is now recognized as a functional feature of a large number of proteins. The observation that protein or protein regions lacking stable and unique 3D conformations [1] carry important biological functions has revolutionized the old and well-established protein structure-function dogma. Since the birth in the early 2000s of the concept of protein intrinsic disorder, numerous studies have highlighted its ubiquitous and multifunctional nature, for review see [2,3]. The RNA virus proteome is particularly enriched in intrinsically disordered regions, IDRs [4,5]. Because of their structural plasticity, IDRs are multifunctional. Notably, many IDRs are involved in multi-partnership, thereby developing between the virus and its host, a complex interactome [6]. A common trait of RNA viruses resides in their extraordinary ability to adapt to fast-changing environments, by creating a high degree of genetic diversity in viral populations. This genetic diversity both result from the unparalleled mutation rates observed in viruses, and their ability to evolve by maintaining mutations in populations, *i.e.* mutational robustness [7,8]. Mutational robustness is defined as the ability of an organism to maintain its phenotype despite the disturbances caused by mutation accumulation. In case of viruses, this mutational robustness is the result of combined

multi-scale processes including functional complementation between individual variants in the population (quasi-species) but also mutation buffering at the molecular level. Indeed, non-conservative mutations within viral proteins often lead to an increased instability, partly buffered by an enhanced chaperone assistance [9]. Another and probably more frequent factor accounting for mutational robustness could reside in intrinsic structural properties of viral proteins [10]. We previously proposed that IDRs could contribute to virus adaptive potential [11], and recently, we reported the first experimental evidence for the involvement of a viral protein intrinsically disordered region in the adaptation of a plant RNA virus to its host [12]. An earlier study was aimed at comparing the structural features of 123 single-domain small-proteins from hypothermophilic bacteria, archaea, mesophilic eukaryota and prokaryota, and RNA or DNA viruses, whose crystal structures were available [13]. It was concluded from this analysis that viral proteins and more particularly RNA virus proteins, display (i) higher stability upon simulations of mutation accumulation (lower  $\Delta\Delta G$  per residues compared to the other groups of proteins) and (ii) lower inter-residues contact densities. This latter feature is a typical signature of intrinsic disorder. It has thus been proposed that the large intrinsic disorder content in viral protein could constitute a viral strategy to efficiently buffer mutation effects [13,14], strongly contrasting with non-additive/epistatic stability loss profile expected from ordered proteins as previously reported for a bacterial  $\beta$ -lactamase [15].

The study presented here aimed at experimentally assess if intrinsic disorder within viral proteins is more prone to mutational robustness than ordered proteins. To address this issue, we used the biologically relevant interaction between the viral genome-linked protein (VPg), an intrinsically disordered protein from the potyvirus *Potato virus Y* (PVY), and one of its ordered partner, the eukaryotic translation initiation factor 4E (eIF4E) from pepper *Capsicum annuum* [16]. The VPgs from several potyviruses, namely Lettuce mosaic virus (LMV) [17], Potato virus Y (PVY) [18] and Potato virus A (PVA) [19] have been experimentally characterized as intrinsically disordered. With the exception of its 45 amino acid N-terminal domain, which does not participate in the interaction with VPg [20] [21], eIF4E is an ordered globular protein [22–24]. We compared the mutational robustness of these two proteins subject to random mutagenesis, by monitoring the loss of their respective ability to interact with each other, a molecular feature reflecting their biological function in the infectious process. The effects of artificial mutation accumulation in each of eIF4E and VPg proteins encoding genes was assessed by monitoring VPg-eIF4E interaction using yeast two hybrid (Y2H) system. More specifically, independent screenings of either VPg or eIF4E mutant libraries for the interaction with their respective wild type partner were performed.

## Materials and methods

In order not to lose the high quality (size and diversity) of the libraries during their transfer from bacteria to yeasts, a cloning strategy was optimized. We used high efficiency Gateway recombination system, bacteria strains with maximum transformation efficiency and yeast mating that typically insures a better rate of cotransformed yeast. Corresponding mating efficiencies in yeast are provided in [S2 Table](#). Typically, Libraries transformation in *E.Coli* produced an average of  $3 \times 10^4$  individual clones. The higher cloning-mating efficiency in yeast ( $> 10^6$ ) insured that the original library complexity was conserved during the transfer between bacteria and yeast (see [supporting information](#) for details).

## Gene sequences used in the study

Viral genome linked protein (VPg) was from a variant of PVY isolate SON41 (GenBank accession AJ 439544–2)[25]. The Eukaryotic translation initiation factor 4E used in this study is

encoded by *Capsicum annuum* Yolo Wonder, a bell pepper inbred line susceptible to PVY (GenBank accession AAN74644-1).

### Production of random mutant libraries

Genes encoding Viral protein genome-linked protein (VPg) from Potato virus Y (PVY) SON41g and eIF4E factor from *Capsicum annuum* cv. Yolo Wonder were cloned into pDONR201 entry vector Gateway system (ThermoFisher). GeneMorph II Random Mutagenesis Kit (Agilent) was used to perform random error-prone PCR (epPCR) mutagenesis. Several DNA target amounts were tested to produce various rates of mutations (S3 Table). In case of "Highly-mutated" libraries, an iterative strategy was used, involving two successive rounds of PCR-mutagenesis. EpPCRs were performed following manufacturer instructions and using attL1 and attL2 primers (5' TCGCGTTAACGCTAGCATGGATCTC and 5' GTAACATCAGA GATTTTGAGACAC respectively) that hybridize upstream and downstream of attL Gateway recombination sites. PCR products were loaded on agarose gel 1%. Amplicons of the expected size were extracted from the gel using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). AttL-flanked mutated eIF4E and VPg genes were inserted into Yeast Two Hybrid pDEST-GADT7 (activation domain fusion vector—TAIR resource accession #1010229695) and pDEST-GBKT7 (DNA-binding domain fusion vector—TAIR resource accession #1010229688) respectively. The eIF4E auto-activation prevented us from using the "eIF4E-binding domain" construct. LR recombination reactions were performed at 25°C overnight using LR Gateway LR Clonase II Enzyme mix following manufacturer recommendations. Proteinase K treatment were applied by adding 1μl of proteinase K and incubation at 37°C for 1 minute. LR products were used to transform *Escherichia coli* DH10B MAX-efficiency competent cells (NEB) by electroporation. Transformation products were spread and transformed cells selected on LB plates supplemented with the suitable antibiotics (Kanamycin or Ampicillin). To characterize the libraries, 32 colonies per library were sequenced using the attB1 primer (Gateway). Each resulting sequences were aligned with the WT sequence of the corresponding gene to identify mutations. Total number of mutations, number of non-synonymous, synonymous, STOP and INDEL (insertion-deletion) mutations and their respective position in the protein sequence are displayed in S1 Fig. Standard Kruskal-Wallis statistical tests were performed to estimate the homogeneity of the number of non-synonymous mutations per clone between eIF4E and VPg of a given library (low, medium, high). Statistical homogeneity of INDEL and STOP mutations proportion between eIF4E and VPg libraries were checked using Z-score calculation. Theoretical frequencies of STOP mutations and non-synonymous ones were assessed from PEDEL-AA analysis web-tool [26]. For each mutant library, all isolated colonies were retrieved from LB plates and suspended into liquid LB media for 6 hours. Plasmid extractions were performed using Nucleospin Plasmid kit (Macherey-Nagel). DNA mutant plasmid libraries were suspended in water and stored at -20°C.

### Yeast-Two-Hybrid analysis

The Matchmaker GAL4 Two-Hybrid system 3 (Clontech) was used according to the manufacturer specifications. pDEST GADT7—containing either the wt eIF4E sequence or the different mutant libraries and pDEST GBKT7- containing either the wt VPg sequence or the different mutant libraries were transformed into AH109 and Y187 yeast strains respectively, which contain two independent reporter genes HIS3 and ADE2, conferring histidine and adenine auxotrophy, respectively and driven by hybrid GAL4 promoters. Two-hybrid library screening using yeast mating were performed following the recommendations of the Matchmaker™ pretransformed libraries user manual (Clontech PT 3183–1) and the Make your own “Mate

and Plate™” library system user manual (Clontech PT 4085–1). Each of the yeast mutant libraries “low”, “medium”, “high” were screened using yeast mating with the wild type partner: 5 ml of the wt interacting partner strain ( $>10^8$  cells/ml) was combined with 1 ml aliquot of the mutant library strain ( $>10^7$  cells/ml). Diploid yeasts were selected on a non-interaction selective medium (without leucine and tryptophan, -LW), the number of colonies represented the whole population of variants. Yeast colonies able to grow on stringent selecting medium for interaction (without leucine, tryptophan, histidine and adenine, -LWHA) representing the remaining functional variants were counted on 3 to 5 independent plates. For each experiment, mating efficiencies i.e., % diploid) calculated as the percentages of cfu/ml of diploids (number of cfu on -LW plates) related to cfu/ml of the lower viability partner, was  $>7\%$ . The number of clones screened, calculated as the number of cfu/ml of diploids x resuspension volume, was estimated for each mating and was  $>10^6$ . Yeast colonies were counted both manually and with Open CFU, an open-source software, <http://opencfu.sourceforge.net>, [27]

## Results

### VPg and eIF4E randomized mutant libraries characterization

In this work, we postulated that the mutational robustness of VPg and eIF4E could be compared by monitoring the loss of VPg-eIF4E interaction as a function of a progressive accumulation of random mutations into each of the two proteins.

To introduce molecular diversity into parent sequences of the viral protein VPg and its cellular partner eIF4E, three libraries of randomized VPg and eIF4E mutants were generated using error-prone polymerase chain reaction (epPCR). The mutational rate was modulated by finely tune the amount of matrix. For high mutational rates, two consecutive epPCR amplifications of the starting material were performed (see experimental section). Pools of 32 individual clones were randomly chosen in each library and sequenced. Non synonymous mutations leading to stop codons or frame shifts were not considered. Taking into account nucleotide mutations leading to amino acid changes (non-synonymous mutations, NS), the libraries were classified as “low”, sequences containing an average of 1.5 NS mutations per kilo base (kb), “medium” (3 NS mutations) or “high” (7 NS mutations) (Fig 1).

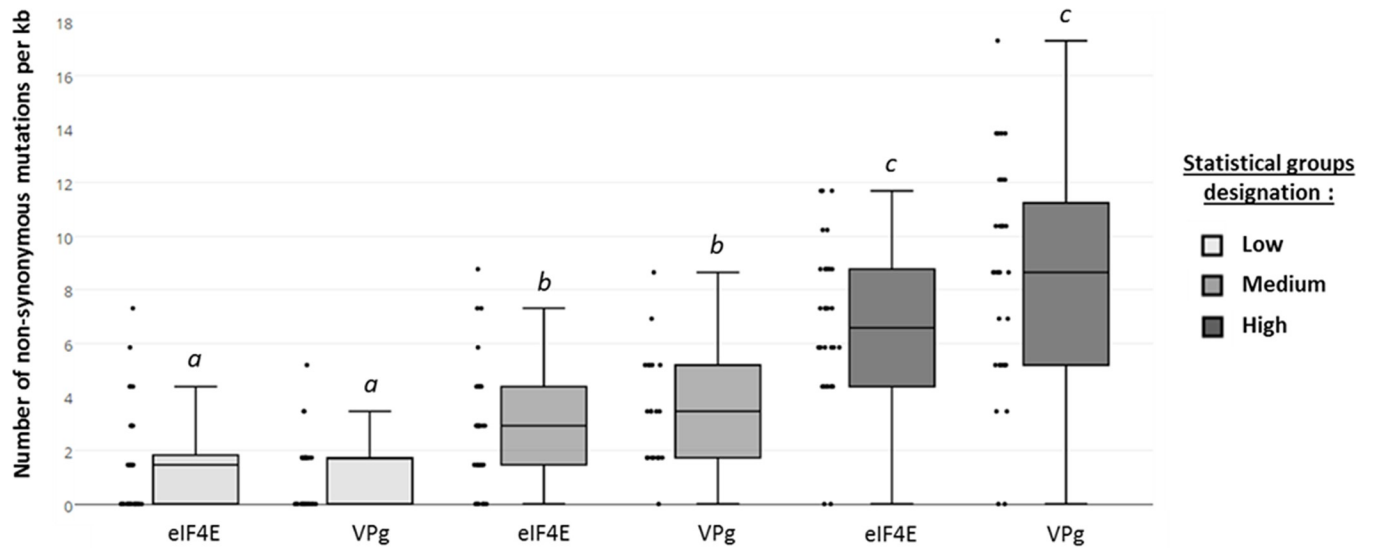
For both eIF4E and VPg coding sequences, the “low”, “medium” and “high” statistical groups were significantly different from each other (p-value between 0.0001 and 0.01, Kruskal-wallis test).

To accurately compare the mutational robustness of the two proteins, the mutation rate of eIF4E and VPg variant libraries within each of the three statistical groups has to be not significantly different (p-value  $>0.01$ , Table 1).

The mutagenesis coverage was estimated by locating on the coding sequence all the mutations observed in each of the sequenced samples. For each of the six libraries within a given statistical class (VPg and eIF4E “low”, “medium” and “high” libraries), mutations were evenly distributed with a homogenous coverage along the gene sequence (Fig 2).

### Analysis of library representativeness

Good library representativeness must satisfy a high diversity (i.e. covering a significant number of possible mutations). This corresponds to a library containing as few redundant sequences (including copies of the non-mutated parental gene) and as many full-length sequences (lacking premature termination codons) as possible. Bacteria were transformed with the six libraries and, from positive clones counting it was deduced that the libraries contained an average of  $3.5 \cdot 10^4$  clones. Sequencing results and their derived nucleotide substitution matrixes (S2 Fig) were used to feed the PEDEL-AA web-tool [26]. Starting from the overall (DNA) mutation



**Fig 1. Distribution of the number of non-synonymous mutations in sequenced variants sampled from each VPg and eIF4E mutant library.** Black dots: number of NS mutation per kb retrieved from each of the 32 variants sequenced. *a*, *b* and *c* are 3 significantly different statistical groups (p-value cut-off at 0.01, Kruskal-Wallis test).

<https://doi.org/10.1371/journal.pone.0211725.g001>

rate, the total size of the library and the nucleotide mutation matrix (which was obtained here by sequencing randomly chosen library members) PEDEL-AA estimates the protein sequence diversity in error prone PCR libraries. Notably, the most useful feature of PEDEL-AA is its ability to quickly estimate the total number of unique proteins in an epPCR library. Using this algorithm, the protein diversity obtained from random mutagenesis of eIF4E and VPg at identical mutation rates was derived (S1 Table). According to this analysis, eIF4E and VPg-based random mutagenesis were expected to provide very similar STOP-mutation frequencies (differences representing less than 10%) in accordance with sequencing results on the 32 randomly chosen variants. NS vs. Synonymous (S) mutation ratios obtained from eIF4E and VPg theoretical libraries revealed minor differences. The VPg “medium” and “high” libraries were expected to provide slightly more non-synonymous mutations (excluding stop codon) per kb than eIF4E. According to this analysis, if a different mutational robustness is observed for the two proteins, it should more likely refer to different non-synonymous mutation tolerance rather than to a differential STOP-codon or INDEL accumulation.

**Table 1. Comparison of the 6 libraries used in this study.**

| Mutants libraries statistical comparison |          | High  |            | Medium |            | Low   |            |
|--|----------|-------|------------|--------|------------|-------|------------|
|  |          | 4E YW | VPg S41g   | 4E YW  | VPg S41g   | 4E YW | VPg S41g   |
| High                                     | 4E YW    | x     | <i>n.s</i> | ***    | ***        | ***   | ***        |
|  | VPg S41g |       | x          | ***    | ***        | ***   | ***        |
| Medium                                   | 4E YW    |       |            | x      | <i>n.s</i> | *     | *          |
|  | VPg S41g |       |            |        | x          | **    | **         |
| Low                                      | 4E YW    |       |            |        |            | x     | <i>n.s</i> |
|  | VPg S41g |       |            |        |            |       | x          |

Side by side comparisons of average non-synonymous mutation number per variant in eIF4E and VPg mutant libraries. “\*\*\*”: p-value < 0.0001

“\*\*”: p-value < 0.001

“\*”: p-value < 0.01; “n.s”, not significant. For p-value > 0.01, Kruskal-Wallis test, libraries were considered as non-significantly different.

<https://doi.org/10.1371/journal.pone.0211725.t001>



### Comparative analysis of VPg and eIF4E mutational robustness

Comparing the mutational robustness of two proteins of unrelated functions does not allow to put the results obtained in a true biological perspective. By contrast the choice of comparing VPg and eIF4E mutational robustness through the evaluation of mutation accumulation effects on their interaction is especially relevant. Indeed, the VPg-eIF4E interaction is required for potyvirus infection. More specifically, with respect to evolutive aspects, mutations in eIF4E are related to resistance mechanisms to the virus, which are in turn overcome by virus adaptation through mutations within the intrinsically disordered VPg central region [28]. Among the numerous VPg natural variants and eIF4E alleles provided by the PVY-pepper pathosystem, Pvr<sup>2+</sup>-eIF4E allele and VPg from PVY SON41g isolate were chosen, as Y2H test reveals their strong interaction, thereby reducing the occurrence of false positives (Charron et al. 2008). In order to compare the mutational robustness of both proteins, Y2H experiments were performed by parallel screening of eIF4E mutated libraries against wild type VPg and reversely, mutated VPg libraries against wild type eIF4E. For each of the three groups of libraries (low, medium and high), the VPg and eIF4E mutant populations were compared across all mutagenesis conditions (t-test p-value <0.01, Fig 3) and also within each mutagenesis condition (t-test for low, medium and high libraries, p-values <0.0002, <0.001 and <0.5 respectively). Although no significant difference was observed between the two proteins at high mutation rate, taken together the data show that the remaining functional populations of VPg variants was significantly higher than the corresponding population of eIF4E variants (Fig 3). Such discrepancy could be explained by differences in codon volatility, a property linked to the nucleotide sequence encoding each of the two proteins. Volatility of a codon is defined as the probability that a random point mutation in this codon is non synonymous [29]. If the nucleotide sequence encoding VPg displays a significantly lower codon volatility than the sequence encoding eIF4E, a given mutation rate could generate more synonymous codons in the VPg



**Fig 2. Distribution along eIF4E and VPg sequences of all mutations found in the 32 randomly chosen variants sequenced in each of the six libraries. A) eIF4E; B) VPg.** NS: non-synonymous; S: synonymous; INDEL: insertion/deletion; STOP: mutation involving the appearance of STOP-codons. For comparative purpose, number of mutations is displayed rather than percentages of mutated sites in the coding sequence, as the two proteins are of comparable length.

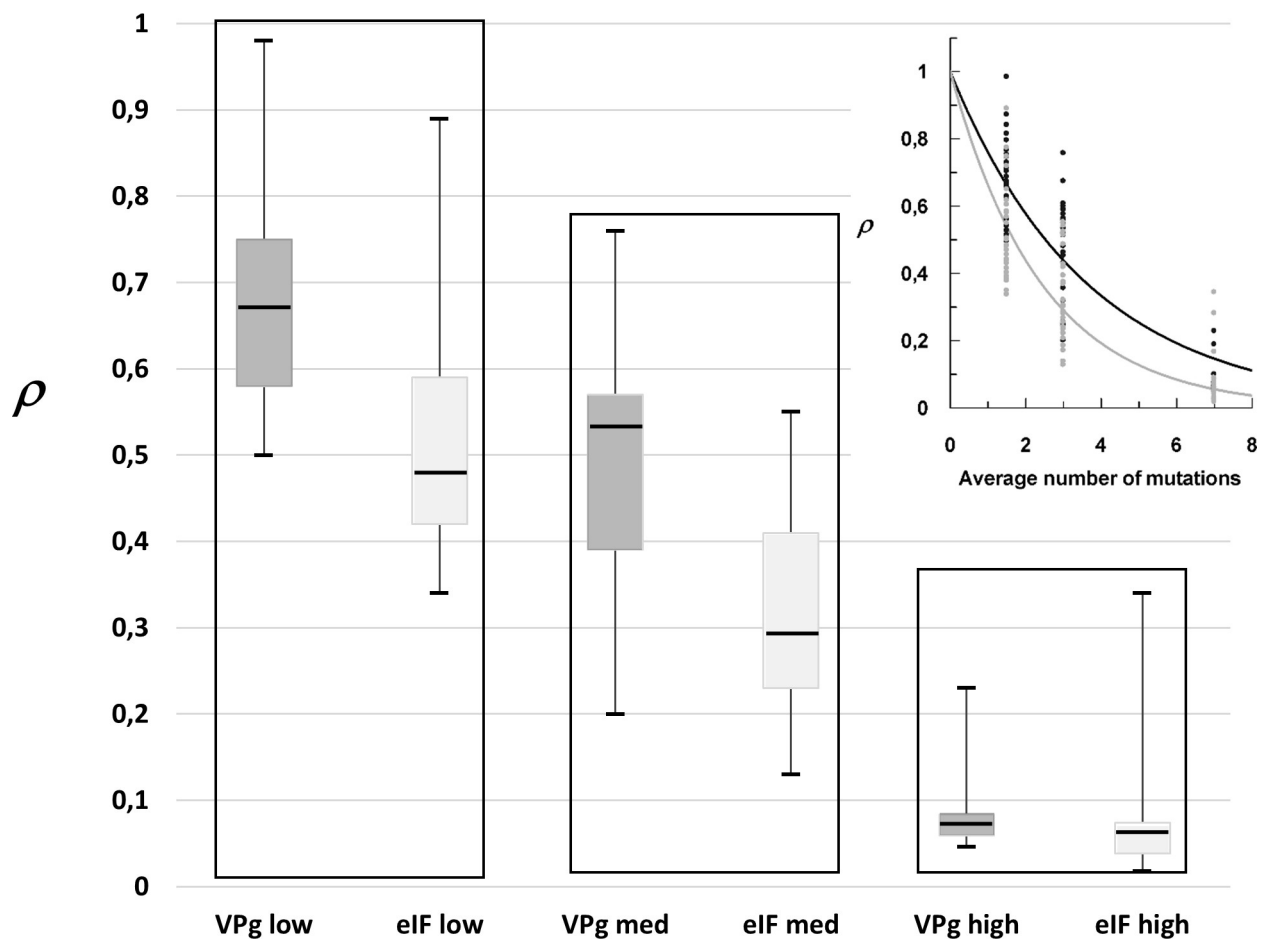
<https://doi.org/10.1371/journal.pone.0211725.g002>

sequence, ensuing into more wild type proteins within the population, and in turn would not imply a higher mutational robustness at the level of the VPg amino acid sequence. There is no bias of this kind, as according to our calculation, VPg and eIF4E nucleotide sequences display comparable probabilities of synonymous codon apparition (0.3) upon random point mutations. An apparently higher mutation robustness of VPg compared to eIF4E can be observed from 1.5 to 7 non-synonymous mutations per kb.

### Discussion

#### The intrinsically disordered VPg displays a higher mutational robustness than eIF4E, its more structured partner

Several factors could explain the apparently higher mutation robustness observed in VPg compared to eIF4E. Firstly, this difference could be a general property of the source organism (i. e.



**Fig 3.** Comparative mutational robustness of VPg (dark grey) and eIF4E (light gray). Box plots display of the dispersion of remaining functional variants of VPg (dark grey) and eIF4 (light gray) within the low, med and high libraries. Yeasts transformed with “low”, “medium” and “high” libraries of variants were mated with yeasts transformed with the wild type partner encoding gene. The resulting diploid yeasts containing the genes of both protein partners were selected on a non-interaction selective medium (-LW). The corresponding number of colonies,  $\tau$ , represented the whole population of variants. Yeast colonies able to grow on stringent interaction selecting medium (-LWHA) were counted on 3 to 5 independent plates (Three independent panning). The corresponding number,  $\sigma$ , was divided by  $\tau$ .  $\rho = \sigma / \tau$  refers to the rate of remaining functional variants within the mutants libraries relative to the starting whole population of variants.  $\rho$  is indicative of the mutational robustness of the protein considered. (See [Material and Methods](#) for more details). (t-test p-value <0.01). Inset: the rate of remaining functional variants versus mutation accumulation (black circles and gray circles, VPg and eIF4E library survival rates respectively). The mono exponential model of fitness decline as a function of mutation accumulation was fitted to the experimental data  $\rho = \text{EXP}(-k \cdot n)$  with n, the averaged mutation number per protein. Best fit:  $k = 0.274 \pm 0.011$  (VPg black line) and  $k = 0.412 \pm 0.016$  (eIF4E gray line).

<https://doi.org/10.1371/journal.pone.0211725.g003>



virus vs. eukaryotic) rather than specific properties of the proteins involved. This seems unlikely, as in eukaryotic proteomes, IDRs have a higher amino acid polymorphism than ordered regions [30], a feature also observed in virus proteomes. This is well illustrated in potyviruses, where IDRs tend to evolve faster than more structured regions, [11]. Another explanation could be that each amino-acid site of eIF4E and VPg are equally tolerant to mutations, but that a higher proportion of sites in eIF4E are involved in the interaction. This could mean that the distribution over the protein structure, of amino acids involved in the interaction is wider for eIF4E than for VPg. However, the interaction domains, namely the central region of the VPg [31–33] and the C-terminal region of eIF4E [21] are of the same order (20–30 residues length). Hence, it is likely that the VPg structural elements which are involved in the interaction, better tolerate the accumulation of non-synonymous mutations than their eIF4E counterparts. By contrast, the loss of eIF4E interaction properties illustrates the stronger structural constraints exerted on this factor.

It was previously deduced from hydrodynamic parameters determination that VPg behaves as a pre-molten globule [34]. This type of protein possesses loosely packed cores involving secondary structures interspaced with disordered regions [35]. Viral proteins display a high occurrence of such features. A comparative thermodynamic analysis between compact proteins from thermophilic organisms and viral proteins was previously reported. Although the set of structures available was limited, the study strongly suggested that intrinsically disordered viral proteins and compact proteins could display two different types of functional loss, with respect to mutation accumulation [13]. If deleterious mutations interact so that their combined impact on function is greater than expected from an addition of their individual effect, the function decrease would accelerate with the accumulation of mutations, giving rise to negative epistasis. Compact stable globular proteins would comply to this model, namely “threshold robustness”, whereas, many viral proteins would obey the “gradient robustness” model, with a lower average loss per mutation [15]. This hypothesis paved the way to the present work. The impact of mutations depends both on the effect of each mutation on the formation of VPg-eIF4E complex and on the interaction between accumulated mutations. We could not tune the random mutagenesis parameters finely enough such as being able to generate additional mutant libraries in between our three non-overlapping libraries (Fig 1). Consequently, the shape of the curve of  $\rho$  versus the average mutation number could not be more precisely defined in order to explore a possible negative epistasis. Hence, we simply hypothesized that deleterious mutations were not interacting (no epistasis). In such a case, the VPg-eIF4E interaction, expressed as the survival rate, was expected to decline mono-exponentially. It was observed that the functional loss of the two proteins decreased with significantly different rates, which is indicative of a higher mutational robustness of VPg than eIF4E, its more structured partner (Fig 3).

Previous extensive computational analyses of disordered proteins and their evolution [36,37] converge to the agreement that disordered protein regions evolve rapidly. We recently analyzed both intra and interspecies amino acid polymorphism within the potyvirus proteome. Regarding the VPg, there are three intrinsically disordered regions spanning the first fifth of the N-terminal part, the last fifth of the C-terminal part and the central part, roughly between amino-acid 85 and 120. The latter displays a larger amino acid polymorphism than the other parts of the protein [11]. This region, which is involved in the interaction with eIF4E, is under positive selection with respect to amino acids determining the host resistance breakdown [38]. As a matter of fact, large phylogenetic analysis shows a significantly higher degree of positive Darwinian selection in intrinsically disordered regions compared to more structured regions. Such a positive selection is significantly enriched in classes of intrinsically disordered proteins whose functions can be coupled to adaptive processes [39].

IDRs are known to adopt different conformations upon interaction with different binding partners [40]. Interestingly, interactions with multiple partners and high rates of evolution are both associated with disorder [41,42] and conformational flexibility [43]. It was shown that the intrinsically disordered central region of VPg participates in the interaction with the virus helper component HC-Pro, eIF4E being a competitive binder of this latter [33]. It is likely that intrinsic disorder brings the conformational flexibility required for this multipartnership. Indeed, binding promiscuity is one of the functional features of intrinsically disordered regions [44,45]. Although it seems counterintuitive that a region participating to the definition of a precise surface complementation in the context of a protein-protein interaction displays mutational robustness, it was observed that there are several possibilities to bind a given protein. The surface overlap may be defined by alternative combinations of amino acids [46].

### Mutational robustness in VPg relates to the particular class of “flexible disorder”

We previously observed that, globally, in potyviruses, IDRs display significantly higher dN/dS (non-synonymous substitutions over synonymous substitutions) values than ordered ones, that indicates a tendency of intrinsically disordered domains to evolve faster than more structured regions during potyvirus evolution [11]. Such weaker evolutive constraints on IDRs strongly suggest abilities for a faster and easier exploration of adaptive solutions and the emergence of new functions in viral proteins [13,47]. However, disorder is not systematically related to mutational robustness. Bellay and colleagues proposed to sort conserved disordered regions in two classes according to their evolutive rates: “flexible disorder” for disordered regions which display a relatively high amino acid polymorphism, and “constrained” disorder which refers to disordered regions that are not variable in their amino acid sequences [48]. Flexible disorder is, in the first instance, found in flexible spacers between functional domains [49]. It is also observed in proteins with one or two regions interacting sequentially with several partners through conformational switches, termed “date Hubs” [50]. VPg belongs to this family, as its intrinsically disordered central region sequentially interacts with eIF4E and the virus helper component HC-Pro [33]. In addition VPg interacts also with CI, the viral helicase [51]. During the preparation of this manuscript, the involvement of VPg N-ter disordered region in an interaction with NIa-Pro, the protease domain responsible for the viral polyprotein maturation was reported. NIa-Pro protease activity and VPg ATPase activity are modulated through this binary interaction [52]. This finding supports the hypothesis that VPg belongs to the family of conformational switches, a family of IDPs displaying mutational robustness. By contrast, constrained disorder is found in IDRs, whose function is linked to more strict topological requirements. This is the case of multi-interface hubs (named “party hubs”) [50] that can be involved in simultaneous binding with various partners.

### Conclusion

The experimental analysis of intrinsically disordered protein mutational robustness is still scarcely documented. Recently, the demonstration that the C-terminal IDR of Nodamura Virus polymerase can accommodate very diverse amino acid sequences without losing its function suggests that IDRs can be reservoirs for evolutionary innovations towards virus adaptation to environment changes [14]. This hypothesis needs to be assessed using other examples. The strength of our experimental model lies in the fact that it makes it possible to compare the mutational robustness of two naturally interacting proteins in a biological context. The results of the present study supports previous conclusions. Indeed, using the same experimental model, we recently observed *in planta* that a modulation of the VPg disorder

content impacts the success of the virus to overcome the host resistance [12]. The questions raised by the possible link between IDR mutational robustness and adaptive processes is of very general relevance. This type of study requires to be further developed as it will feed the debate regarding the current reassessment of the dogma of the relationship between protein structure and function. To this end, reverse genetics is especially suitable. Using phytovirus as experimental models, the biological properties of mutants can be quickly and accurately tested *in vivo* on large populations of host plants (1000 and more). Notably, a comparative analysis in *planta* of the impact on infection phenotype of mutations introduced either in ordered or disordered regions within the virus genome will allow the assessment of IDRs mutational robustness in a true biological context and ultimately, to better understand how viruses cope with mutations.

## Supporting information

### S1 Fig. Mean substitution rates matrixes.

(PDF)

### S2 Fig. Distribution of VPg and eIF4E codon volatility.

(PDF)

### S1 Table. Library characteristics (PEDEL-AA analysis).

(PDF)

### S2 Table. Mating efficiency in yeast.

(PDF)

### S3 Table. Error prone PCR conditions used to produce “low”, “medium” and “high” libraries.

(PDF)

### S4 Table. Mutations within the six (low, med, high) samples sequenced to characterize VPg and eIF4 mutant libraries.

(PDF)

### S1 Code. Rcode for codon volatility calculation.

(PDF)

## Acknowledgments

We are grateful to Dr Jean Luc Gallois, INRA Avignon, France for kindly providing the yeast two hybrid vectors bearing the wild type eIF4E and VPg genes.

## Author Contributions

**Conceptualization:** Jocelyne Walter, Thierry Michon.

**Data curation:** Jocelyne Walter, Justine Charon, Yihua Hu, Thomas Leger, Thierry Michon.

**Formal analysis:** Jocelyne Walter, Justine Charon, Yihua Hu, Guillaume Lafforgue, Thierry Michon.

**Funding acquisition:** Thierry Michon.

**Investigation:** Jocelyne Walter, Justine Charon, Yihua Hu, Joy Lachat, Thomas Leger, Amandine Barra.

**Methodology:** Jocelyne Walter, Justine Charon, Yihua Hu, Joy Lachat, Thomas Leger, Guillaume Lafforgue, Amandine Barra.

**Software:** Guillaume Lafforgue.

**Supervision:** Jocelyne Walter, Thierry Michon.

**Validation:** Jocelyne Walter, Justine Charon, Thierry Michon.

**Visualization:** Jocelyne Walter.

**Writing – original draft:** Jocelyne Walter, Justine Charon.

**Writing – review & editing:** Jocelyne Walter, Thierry Michon.

## References

1. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol.* 1999; 293: 321–331. <https://doi.org/10.1006/jmbi.1999.3110> [pii] ET—1999/11/05 PMID: 10550212
2. Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. *Chem Rev. American Chemical Society;* 2014; 114: 6561–6588. <https://doi.org/10.1021/cr400514h> PMID: 24739139
3. Oldfield CJ, Dunker a K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu Rev Biochem.* 2014; 1–32. <https://doi.org/10.1146/annurev-biochem-041612-095236>
4. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn.* 2012; 30: 137–149. <https://doi.org/10.1080/07391102.2012.675145> PMID: 22702725
5. Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, et al. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci.* 2014; <https://doi.org/10.1007/s00018-014-1661-9> PMID: 24939692
6. Xue B, Blocquel D, Habchi J, Uversky A V, Kurgan L, Uversky VN, et al. Structural Disorder in Viral Proteins. *Chem Rev. American Chemical Society;* 2014; 114: 6880–6911. <https://doi.org/10.1021/cr4005692> PMID: 24823319
7. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet.* 2008; 9: 267–76. <https://doi.org/10.1038/nrg2323> PMID: 18319742
8. Sanjuán R. From molecular genetics to phylodynamics: Evolutionary relevance of mutation rates across viruses. *PLoS Pathog.* 2012; 8: 1–5. <https://doi.org/10.1371/journal.ppat.1002685> PMID: 22570614
9. Pechmann S, Frydman J. Interplay between chaperones and protein disorder promotes the evolution of protein networks. *PLoS Comput Biol.* 2014; 10: e1003674. <https://doi.org/10.1371/journal.pcbi.1003674> PMID: 24968255
10. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol.* 2009/09/22. 2009; 19: 596–604. <https://doi.org/10.1016/j.sbi.2009.08.003> [pii] 10.1016/j.sbi.2009.08.003 PMID: 19765975
11. Charon J, Theil S, Nicaise V, Michon T. Protein intrinsic disorder within the Potyvirus genus: from proteome-wide analysis to functional annotation. *Mol BioSyst. The Royal Society of Chemistry;* 2016; 12: 634–652. <https://doi.org/10.1039/C5MB00677E> PMID: 26699268
12. Charon J, Barra A, Walter J, Millot P, Hébrard E, Moury B, et al. First experimental assessment of protein intrinsic disorder involvement in an RNA virus natural adaptive process. *Mol Biol Evol.* 2018; 35: 38–49. <https://doi.org/10.1093/molbev/msx249> PMID: 29029259
13. Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. Do viral proteins possess unique biophysical features? *Trends Biochem Sci.* 2009; 34: 53–59. <https://doi.org/10.1016/j.tibs.2008.10.009> PMID: 19062293
14. Gitlin L, Hagai T, LaBarbera A, Solovey M, Andino R. Rapid evolution of virus sequences in intrinsically disordered protein regions. *PLoS Pathog. United States;* 2014; 10: e1004529. <https://doi.org/10.1371/journal.ppat.1004529> PMID: 25502394
15. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature.* 2006; 444: 929–932. <https://doi.org/10.1038/nature05385> PMID: 17122770

16. Charron C, Nicolai M, Gallois JL, Robaglia C, Moury B, Palloix A, et al. Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. *Plant J.* 2008; 54: 56–68. <https://doi.org/10.1111/j.1365-313X.2008.03407.x> PMID: 18182024
17. Hebrard E, Bessin Y, Michon T, Longhi S, Uversky VN, Delalande F, et al. Intrinsic disorder in Viral Proteins Genome-Linked: experimental and predictive analyses. *Virology.* 2009/02/18. 2009; 6: 23–36. <https://doi.org/10.1186/1743-422X-6-23> [pii] 10.1186/1743-422X-6-23 PMID: 19220875
18. Grzela R, Szolajska E, Ebel C, Madern D, Favier A, Wojtal I, et al. Virulence Factor of Potato Virus Y, Genome-attached Terminal Protein VPg, Is a Highly Disordered Protein. *J Biol Chem.* 2008; 283: 213–221. <https://doi.org/10.1074/jbc.M705666200> PMID: 17971447
19. Rantalainen KI, Uversky VN, Permi P, Kalkkinen N, Dunker AK, Makinen K. Potato virus A genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core. *Virology.* 2008; 377: 280–288. <https://doi.org/10.1016/j.virol.2008.04.025> PMID: 18533220
20. Miyoshi H, Okade H, Muto S, Suehiro N, Nakashima H, Tomoo K, et al. Turnip mosaic virus VPg interacts with *Arabidopsis thaliana* eIF(iso)4E and inhibits *in vitro* translation. *Biochimie.* 2008/06/28. 2008; 90: 1427–1434. S0300-9084(08)00188-0 [pii] <https://doi.org/10.1016/j.biochi.2008.03.013> PMID: 18582528
21. Okade H, Fujita Y, Miyamoto S, Tomoo K, Muto S, Miyoshi H, et al. Turnip mosaic virus genome-linked protein VPg binds C-terminal region of cap-bound initiation factor 4E orthologue without exhibiting host cellular specificity. *J Biochem.* 2009/01/06. 2009; 145: 299–307. mvn180 [pii] <https://doi.org/10.1093/jb/mvn180> PMID: 19122207
22. Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK. Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell.* 1997; 89: 951–961. PMID: 9200613
23. Monzingo AF, Dhaliwal S, Dutt-Chaudhuri A, Lyon A, Sadow JH, Hoffman DW, et al. The structure of eukaryotic translation initiation factor-4E from wheat reveals a novel disulfide bond. *Plant Physiol.* 2007; 143: 1504–1518. <https://doi.org/10.1104/pp.106.093146> PMID: 17322339
24. Ashby JA, Stevenson CEM, Jarvis GE, Lawson DM, Maule AJ. Structure-based mutational analysis of eIF4E in relation to sbm1 resistance to Pea seed-borne mosaic virus in Pea. *PLoS One.* 2011; 6. <https://doi.org/10.1371/journal.pone.0015873> PMID: 21283665
25. Caranta C, Lefebvre V, Palloix A. Polygenic Resistance of Pepper to Potyviruses Consists of a Combination of Isolate-Specific and Broad-Spectrum Quantitative Trait Loci. *Mol Plant-Microbe Interact. The American Phytopathological Society;* 1997; 10: 872–878. <https://doi.org/10.1094/MPMI.1997.10.7.872>
26. Firth AE, Patrick WM. GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Res.* 2008; 36: 281–285. <https://doi.org/10.1093/nar/gkn226> PMID: 18442989
27. Geissmann Q. OpenCFU, a New Free and Open-Source Software to Count Cell Colonies and Other Circular Objects. *PLoS One.* 2013; 8. <https://doi.org/10.1371/journal.pone.0054072> PMID: 23457446
28. Moury B, Charron C, Janzac B, Simon V, Gallois JL, Palloix A, et al. Evolution of plant eukaryotic initiation factor 4E (eIF4E) and potyvirus genome-linked protein (VPg): A game of mirrors impacting resistance spectrum and durability. *Infect Genet Evol.* 2014; 27: 472–480. <http://dx.doi.org/10.1016/j.meegid.2013.11.024> PMID: 24309680
29. Plotkin JB, Dushoff J, Fraser HB. Detecting selection using a single genome sequence of *M. tuberculosis* and *P. falciparum*. *Nature.* Nature Publishing Group; 2004; 428: 942–945. <https://doi.org/10.1038/nature02458> PMID: 15118727
30. Ahrens J, Dos Santos HG, Siltberg-Liberles J. The Nuanced Interplay of Intrinsic Disorder and Other Structural Properties Driving Protein Evolution. *Mol Biol Evol.* 2016; 33: 2248–2256. <https://doi.org/10.1093/molbev/msw092> PMID: 27189555
31. Leonard S, Plante D, Wittmann S, Daigneault N, Fortin MG, Laliberte JF. Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *J Virol.* 2000; 74: 7730–7737. PMID: 10933678
32. Khan MA, Miyoshi H, Gallie DR, Goss DJ. Potyvirus genome-linked protein, VPg, directly affects wheat germ *in vitro* translation: interactions with translation initiation factors eIF4F and eIFiso4F. *J Biol Chem.* 2008; 283: 1340–1349. <https://doi.org/10.1074/jbc.M703356200> PMID: 18045881
33. Roudet-Tavert G, Michon T, Walter J, Delaunay T, Redondo E, Le Gall O. Central domain of a potyvirus VPg is involved in the interaction with the host translation initiation factor eIF4E and the viral protein HcPro. *J Gen Virol.* 2007; 88: 1029–1033. <https://doi.org/10.1099/vir.0.82501-0> PMID: 17325377
34. Chroboczek J, Hebrard E, Makinen C, Michon T. Intrinsic disorder in genome-linked viral proteins VPgs of potyviruses. In: Uversky VN, Longhi S, editors. *Flexible viruses: structural disorder within viral proteins.* John Wiley and Sons; 2011. p. 277–312 ST–Intrinsic disorder in genome-linked.



35. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. *Chem Rev.* 2014/04/30. 2014; 114: 6589–6631. <https://doi.org/10.1021/cr400525m> PMID: 24773235
36. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. *Curr Opin Struct Biol.* 2011/04/13. 2011; 21: 441–446. S0959-440X(11)00036-4 [pii] <https://doi.org/10.1016/j.sbi.2011.02.005> PMID: 21482101
37. Schlessinger A, Schaefer C, Vicedo E, Schmidberger M, Punta M, Rost B. Protein disorder—a breakthrough invention of evolution? *Curr Opin Struct Biol.* 2011/04/26. 2011; 21: 412–418. S0959-440X(11)00066-2 [pii] <https://doi.org/10.1016/j.sbi.2011.03.014> PMID: 21514145
38. Ayme V, Souche S, Caranta C, Jacquemond M, Chadoeuf J, Palloix A, et al. Different mutations in the genome-linked protein VPg of potato virus Y confer virulence on the pvr2(3) resistance in pepper. *Mol Plant Microbe Interact.* 2006; 19: 557–563. <https://doi.org/10.1094/MPMI-19-0557> PMID: 16673943
39. Nilsson J, Grahn M, Wright AP. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. *Genome Biol.* BioMed Central Ltd; 2011; 12: R65. <https://doi.org/10.1186/gb-2011-12-7-r65> PMID: 21771306
40. Wright PE, Dyson HJ. Linking folding and binding. *Curr Opin Struct Biol.* 2009/01/23. 2009; 19: 31–38. S0959-440X(08)00179-6 [pii] <https://doi.org/10.1016/j.sbi.2008.12.003> PMID: 19157855
41. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, et al. Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol.* 2002; 55: 104–110. <https://doi.org/10.1007/s00239-001-2309-6> ET—2002/08/08 PMID: 12165847
42. James LC, Tawfik DS. Conformational diversity and protein evolution—A 60-year-old hypothesis revisited. *Trends in Biochemical Sciences.* 2003. pp. 361–368. [https://doi.org/10.1016/S0968-0004\(03\)00135-X](https://doi.org/10.1016/S0968-0004(03)00135-X) PMID: 12878003
43. Boehr DD, Wright PE. Biochemistry. How do proteins interact? *Science (80-).* 2008; 320: 1429–1430. <https://doi.org/10.1126/science.1158818> [pii] 10.1126/science.1158818 ET—2008/06/17 PMID: 18556537
44. Schreiber G, Keating AE. Protein binding specificity versus promiscuity. *Curr Opin Struct Biol.* 2011; 21: 50–61. <https://doi.org/10.1016/j.sbi.2010.10.002> PMID: 21071205
45. Cumberworth A, Lamour G, Babu MM, Gsponer J. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem J.* 2013; 454: 361–369. <https://doi.org/10.1042/BJ20130545> PMID: 23988124
46. Martin J. Beauty is in the eye of the beholder: Proteins can recognize binding sites of homologous proteins in more than one way. Nussinov R, editor. *PLoS Comput Biol.* 2010; 6: 1–13. <https://doi.org/10.1371/journal.pcbi.1000821> PMID: 20585553
47. Hagai T, Azia A, Babu MM, Andino R. Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell Rep.* 2014; 7: 1729–39. <https://doi.org/10.1016/j.celrep.2014.04.052> PMID: 24882001
48. Bellay J, Han S, Michaut M, Kim T, Costanzo M, Andrews BJ, et al. Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.* 2011/02/18. 2011; 12: R14. gb-2011-12-2-r14 [pii] <https://doi.org/10.1186/gb-2011-12-2-r14> PMID: 21324131
49. Daughdrill G, Narayanaswami P, Gilmore S, Belczyk A, Brown C. Dynamic Behavior of an Intrinsically Unstructured Linker Domain Is Conserved in the Face of Negligible Amino Acid Sequence Conservation. *J Mol Evol.* 2007; 65: 277–288. <https://doi.org/10.1007/s00239-007-9011-2> PMID: 17721672
50. Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang L V, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature.* 2004; 430: 88–93. <https://doi.org/10.1038/nature02555> PMID: 15190252
51. Taverit-Roudet G, Abdul-Razzak A, Doublet B, Walter J, Delaunay T, German-Retana S, et al. The C terminus of lettuce mosaic potyvirus cylindrical inclusion helicase interacts with the viral VPg and with lettuce translation eukaryotic initiation factor 4E. *J Gen Virol.* 2012; 93: 184–193. <https://doi.org/10.1099/vir.0.035881-0> PMID: 21918009
52. Sabharwal P, Srinivas S, Savithri HS. Mapping the domain of interaction of PVBV VPg with N1a-Pro: Role of N-terminal disordered region of VPg in the modulation of structure and function. *Virology.* 2018; 524: 18–31. <https://doi.org/10.1016/j.virol.2018.08.002> PMID: 30138835