Limits of variation, specific infectivity, and genome packaging of massively recoded poliovirus genomes

Yutong Song[a,1,2], Oleksandr Gorbatsevych[a,1], Ying Liu[a,b], JoAnn Mugavero[a], Sam H. Shen[a,c], Charles B. Ward[d,e], Emmanuel Asare[a], Ping Jiang[a], Aniko V. Paul[a], Steffen Mueller[a,f], and Eckard Wimmer[a,2]

[a]Department of Molecular Genetics and Microbiology, Stony Brook University, Stony Brook, NY, 11794; [b]Pathology and Laboratory Medicine, Staten Island University Hospital, Staten Island, NY 10305; [c]Department of Chemistry, University of Iowa, Iowa City, IA 52242; [d]Google, Inc., Mountain View, CA 94043; [e]Department of Computer Science, Stony Brook University, Stony Brook, NY, 11794; and [f]Codagenix Inc., Stony Brook, NY 11794

Contributed by Eckard Wimmer, August 17, 2017 (sent for review May 17, 2017; reviewed by Alexander Gorbalenya and Richard J. Kuhn)

Computer design and chemical synthesis generated viable variants of poliovirus type 1 (PV1), whose ORF (6,189 nucleotides) carried up to 1,297 "Max" mutations (excess of overrepresented synonymous codon pairs) or up to 2,104 "SD" mutations (randomly scrambled synonymous codons). "Min" variants (excess of underrepresented synonymous codon pairs) are nonviable except for $P2^{Min}$, a variant temperature-sensitive at 33 and 39.5 °C. Compared with WT PV1, $P2^{Min}$ displayed a vastly reduced specific infectivity (si) (WT, 1 PFU/118 particles vs. $P2^{Min}$, 1 PFU/35,000 particles), a phenotype that will be discussed broadly. Si of haploid PV presents cellular infectivity of a single genotype. We performed a comprehensive analysis of sequence and structures of the PV genome to determine if evolutionary conserved cis-acting packaging signal(s) were preserved after recoding. We showed that conserved synonymous sites and/or local secondary structures that might play a role in determining packaging specificity do not survive codon pair recoding. This makes it unlikely that numerous "cryptic, sequence-degenerate, dispersed RNA packaging signals mapping along the entire viral genome" [Patel N, et al. (2017) Nat Microbiol 2:17098] play the critical role in poliovirus packaging specificity. Considering all available evidence, we propose a two-step assembly strategy for +ssRNA viruses: step I, acquisition of packaging specificity, either (*a*) by specific recognition between capsid protein(s) and replication proteins (poliovirus), or (*b*) by the high affinity interaction of a single RNA packaging signal (PS) with capsid protein(s) (most +ssRNA viruses so far studied); step II, cocondensation of genome/capsid precursors in which an array of hairpin structures plays a role in virion formation.

poliovirus | genome recoding | packaging signal | specific infectivity

The capsid precursor P1 (881 amino acids) of type 1 poliovirus (PV), mapping to the N terminus of the polyprotein (PP) (Fig. 1*A*), can be encoded in 10^{442} ways (1) due to the degenerate genetic code. The tiniest fraction of these possible sequences defines PV, the cause of poliomyelitis. PV occurs in three serotypes, of which the most neurovirulent type 1 Mahoney [PV1(M)], the main viral species analyzed in this study, was isolated in 1941 from pooled feces of three healthy children (2).

Genome sequence (3, 4) and gene organization (3) of PV1(M) revealed highly complex structures in its 5'-terminal nontranslated region (5'-NTR), followed by a single ORF encoding the PP, followed by a complex 3'-heteropolymeric region and poly(A) tail (Fig. 1*A*) (5, 6). The PP (7) is an active molecule that cleaves itself into ∼29 polypeptides by two viral proteinases ($2A^{pro}$ and $3C^{pro}/3CD^{pro}$) and an enzyme-independent maturation cleavage (Fig. 1*A*) (5, 6, 8).

Capsid domain P1 controls the identity of PV by determining virion structure (9), serotype identity (10), and interaction with the cellular receptor CD155 (10). Since PV replicates as quasispecies at an error rate of $\sim 10^{-4}$ (11), the following questions arise: How conserved is its synonymous sequence given the astronomical number of alternative possibilities? What encoding could have coevolved that would be optimal to specify 881 capsid residues?

If PV, a member of the genus *Enterovirus* of *Picornaviridae*, is an evolutionary descendant of C-cluster Coxsackie viruses (C-CAVs) (12), the evolution of PV nucleotide sequences was constrained as it adhered to the basic architecture of C-CAVs, its evolutionary parents (13). A second well-known restriction of sequence variability in ORFs is "codon bias" (14), the unequal use of synonymous codons. Encoding the PV1 P1 domain with an excess of "rarely used" (human) synonymous codons results in a nonviable viral construct despite sharing an identical amino acid sequence with wild-type (WT) PV (15).

A third barrier to sequence variation is codon pair bias (CPB). This describes the frequency of synonymous codon pairs that does not fit the use of individual codons (16–18) (for details, see *SI Materials and Methods*). Overrepresented codon pairs in an ORF are favorable (good), and underrepresented codon pairs

Significance

We constructed viable poliovirus (PV) variants carrying up to 2,104 synonymous mutations in the ORF. We studied proliferation phenotypes, particularly the specific infectivity (*si*) that defines the probability of a single virion to initiate a single cell infection. A recent hypothesis proposes that dozens of loosely conserved hairpins, formed within the viral genome, determine assembly specificity of +ssRNA viruses. Our analysis of recoded PVs does not support this hypothesis. We propose that the progression of +ssRNA virus assembly follows two steps. (*i*) Specificity: recognition between either (*a*) replication protein and capsid precursor or (*b*) single RNA packaging signal and cognate capsid precursor. (*ii*) Cocondensation, occurring between single genome/single capsid precursor complex and multiple cognate capsids aided by multiple genome-specific hairpins.

Author contributions: Y.S., A.V.P., and E.W. designed research; Y.S., Y.L., J.M., and S.H.S. performed research; O.G. and C.B.W. contributed new reagents/analytic tools; Y.S., O.G., E.A., P.J., A.V.P., S.M., and E.W. analyzed data; and Y.S., O.G., A.V.P., and E.W. wrote the paper.

Reviewers: A.G., Leiden University Medical Center; and R.J.K., Purdue Institute of Inflammation, Immunology, and Infectious Disease.

Conflict of interest statement: S.M. and E.W. are co-founders of Codagenix, Inc., a small biotech company focusing on the design and development of vaccines. They are co-holders of a patent describing the effects of changes of synonymous codon pairs in viral genomes. S.M is an employee of Codagenix. The work described in this manuscript presents no conflict of interest with the focus of Codagenix. A.V.P. has retired with no connection to Codagenix. Y.L., S.C., and C.B.W. have left the laboratory >2 years ago; their employments are unrelated to Codagenix. P.J. is currently unemployed. Y.S., A.G., J.M., E.A. are laboratory members with no connection to Codagenix.

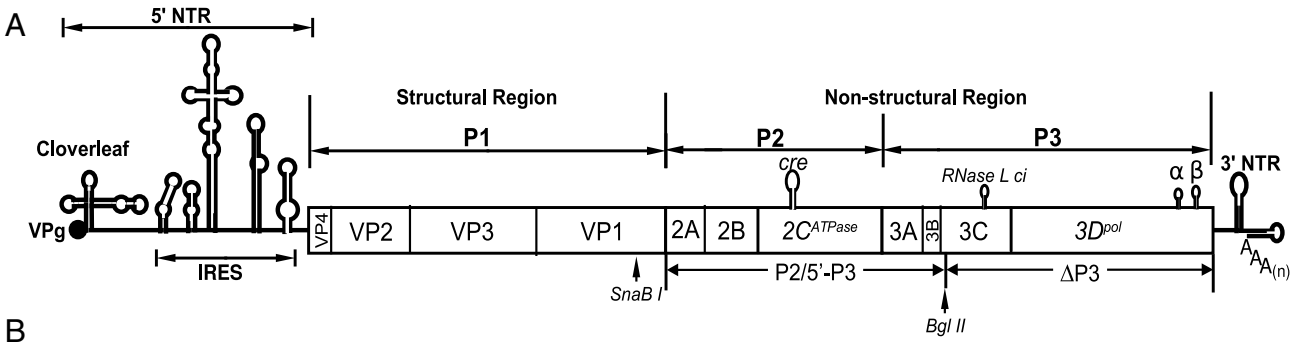
Freely available online through the PNAS open access option.

[1]Y.S. and O.G. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: Eckard.Wimmer@stonybrook.edu or yutong.song@stonybrook.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1714385114/-/DCSupplemental.

A



B

Constructs	Structure of construct P1(743-3385)	P2(3386-5110)	P3(5111-7369)	nts changed/ no. of nts of recoded segment	CPS designed/Wt	CPE @ Pass. #	Titer (lgPFU/ml)	Plaque size
1. PV1	Wt	Wt	Wt	0		0	> 9	large
2. P1^{Max}	Max	Wt	Wt	574/2640	0.246/-0.034	0	> 9	large
3. P1^{Min}	Min	Wt	Wt	632/2640	-0.474/-0.034	None		
4. P1^{SD}	SD	Wt	Wt	929/2628	-0.095/-0.034	0	> 9	large
5. P2^{Max}	Wt	Max^{Cre}	Wt	541/2154	0.223/-0.028	0	> 9	large
6. P2^{Min}	Wt	Min^{Cre}	Wt	471/2154	-0.444/-0.028	3	6.7	tiny
7. P2^{SD}	Wt	SD^{Cre}	Wt	677/2154	-0.075/-0.028	0	> 9	Medium large
8. P3^{Max}	Wt	Wt	Max Wt	182/1395	0.289/0.025	0	> 9	large
9. P3^{Min}	Wt	Wt	Min Wt	330/1395	-0.492/0.025	None		
10. P3^{SD}	Wt	Wt	SD Wt	497/1395	-0.127/0.025	0	8.6	tiny/small

C

Constructs	Structure of construct P1(743-3385)	P2(3386-5110)	P3(5111-7369)	nts changed/ no. of nts of recoded segment	CPS designed/Wt	CPE @ Pass. #	Titer (lgPFU/ml)	Plaque size
1. Wt PV	Wt	Wt	Wt	0		0	> 9	large
2. P(1+2)^{Max}	Max	Max^{Cre}	Wt	1115/4794	0.236/-0.031	0	> 9	large
3. P(2+3)^{Max}	Wt	Max^{Cre}	Max Wt	723/3549	0.249/-0.007	0	> 9	large
4. P(1+3)^{Max}	Max	Wt	Max Wt	756/4032	0.261/-0.014	0	> 9	large
5. P(1+2+3)^{Max}	Max	Max^{Cre}	Max Wt	1297/6189	0.248/-0.019	0	> 9	large + small
6. P(1+2)^{SD}	SD	SD^{Cre}	Wt	1606/4782	-0.086/-0.031	1	7.8	small
7. P(2+3)^{SD}	Wt	SD^{Cre}	SD Wt	1174/3549	-0.095/-0.007	2	6.6	tiny
8. P(1+3)^{SD}	SD	Wt	SD Wt	1426/4020	-0.106/-0.014	1	7.5	small
9. P(1+2+3)^{SD}	SD	SD^{Cre}	SD Wt	2103/6177	-0.095/-0.019	3	6.2	tiny

Fig. 1. The PV genome and details of recoding by Max, Min, and SD designs. (*A*) The PV genome showing the 5′-terminal VPg-linked cloverleaf and IRES, the PP (open box) (divided into the P1 structural and P2, -3 nonstructural domains, and the polyadenylylated 3′NTR (1). The loci of three replication-essential hairpin structures, Cre (81), and α and β (23, 24), as well as the conserved but not-essential RNase L ci element (25, 26), are shown above the PP. The SnaB I and Bgl II restriction sites, used for subcloning the P2 and P3 domains, are indicated. (*B*) Growth phenotypes of PV1(M) and single-domain Max (green), Min (red), and SD (yellow) variants on HeLa monolayer R19 cells (*Materials and Methods*). For connotations of the constructs, see Table S1. Passage 0 means that full CPE emerged after transfection. The sequences of the essential Cre, α and β structures were restored after recoding in P2 and P3 variants (indicated as Max^{cre}, Min^{cre}, or Max^{Wt}, etc.). The total number of nucleotide changes in each recoded segment and the CPS of the recoded fragment relative to the WT in a segment are indicated in column 3 (please note the 12 unchanged nucleotides at the beginning of P1 SD are shown as a small white box). CPS, CPS of the recoded segments (compared with the same WT segment). Growth phenotypes of recoded variants were determined by the time of full CPE, together with virus titer and plaque size after transfection (*Materials and Methods*). (*C*) Same as *B* except more than one domain was recoded.

are unfavorable (bad) in the biological function of the ORF. Although the differences in function between overrepresented and underrepresented (good or bad) pairs is small, rewriting a genome sequence with excessive bad codon pairs kills the virus (16). In this study, we have rewritten PV sequences in the overrepresented ("Max," codon pair optimization) and underrepresented ("Min") [codon pair deoptimization (CPD)] language for testing biological consequences. We used also a third language, the "scrambling design" ("SD"), which involves shuffling synonymous codons at random (15).

Every codon pair has a codon-pair score (CPS), which is the natural logarithm of the ratio of the observed frequency of the codon pair to the expected frequency of the pair [i.e., CPS = ln (Observed/Expected)]. "Expected" is based on the actual occurrences of the codon, and so takes into account codon use. Thus, preferred pairs have positive scores, and dispreferred pairs have negative scores. These values are then used to interpret the average CPS of a viral sequence (16).

All WT PVs have a low specific infectivity (*si*) in HeLa R19 cells, the most favorable tissue culture cells for proliferation. For PV1(M) the ratio of a plaque-forming unit (PFU) to virions is ∼1/120. Our variants described here express vastly different *si* values, which is a viral signature only when determined in related cell lines under comparable conditions. *si* values can dramatically change when a virus is assayed in different cell lines or when cells are treated with drugs that interfere, for example, with the innate immune response. Below, we will discuss the *si* phenotype that is undervalued in virology.

The recoding experiments described here were not only designed to test the PV genome's tolerance to large-scale sequence changes, but also in reference to a recently proposed new model of RNA virus assembly (refs. 19–21 and references therein). This new model implies that the most important determinants for assembly specificity of +ssRNA viruses reside in dozens of genome-specific hairpins that have been described as "cryptic, sequence-degenerate, dispersed RNA packaging signals mapping throughout the genome" (22). We show that this model does not serve to explain PV morphogenesis. Here, we propose a two-step mechanism for assembly of +ssRNA viruses: step I, establishing specificity of assembly, and step II, virion formation by cocondensation of the genome with its cognate capsid proteins (CPs; Fig. 4). We suggest that the multiple loosely conserved hairpin structures are playing a role in step II, possibly contributing to cocondensation of the genome with CPs and virion maturation.

Results

Computer-Designed Recoding of the Entire ORF of the PV1(M) Genome. Using the Max, Min, and SD strategies (*SI Materials and Methods*), we have recoded the PV1(M) PP, by, first, recoding individual PP domains (Fig. 1*B*), followed by recoding two or all three domains (Fig. 1*C*). The identity of individual viral variants is summarized in Table S1.

Growth Properties of Recoded PV Variants. The general outcome of changing the sequences of different PP domains is apparent in Fig. 1 *B* and *C*. Targeting single domains, we confirmed recoding results of the published P1 domain (15, 16). Despite numerous silent mutations, neither $P1^{Max}$ (581/2,640) nor $P1^{SD}$ (930/2,628) individually expressed replication phenotypes different from WT PV under identical conditions. This result confirms that the P1 domain does not harbor essential replication signals (15, 16). Unexpectedly, variant $P1^{Min}$ was nonviable (Fig. 1*B*, line 3; ref. 16). Sequence analyses and subcloning segments of $P1^{Min}$ into WT PV1 produced viable variants (16), which proved that the $P1^{Min}$ plasmid did not harbor fatal flaws. In contrast to $P1^{Min}$, $P2^{Min}$ (Fig. 1*B*, line 6, and Table S1, line 6) was viable (note that a convenient restriction site, Bgl II, inside the P3 domain, was used for cloning). However, transfection with $P2^{Min}$ transcripts required three blind passages to accumulate sufficient inoculum for further studies described below. Sequencing, however, did not indicate adaptive mutations during the blind passages. $P2^{Max}$ revealed growth properties identical to WT PV (Fig. 1*B*, line 5), whereas $P2^{SD}$ (Fig. 1*B*, line 7) produced slightly smaller plaque sizes. We note that in all P2 variants, the essential Cre element (nucleotides 4,444–4,504) was spared from recoding (Fig. 1*B*, indicated as Max^{cre}, Min^{cre}, and SD^{cre}).

The recoded regions of the P3 domain in $P3^{Max}$, $P3^{Min}$, and $P3^{SD}$ variants are relatively small (so is the number of silent mutations) due to the Bgl II restriction site used in cloning and the reconstruction to WT of the important α and β sequences (nucleotides 6,995–7,369) (Table S1, lines 8–10; refs. 23 and 24). $P3^{Max}$ replicated just as WT PV1, whereas $P3^{SD}$ (Table S1, line 10) yielded lower titers and very small plaques (Fig. 1*B*, line 10). $P3^{Min}$ was nonviable, despite carrying fewer silent mutations (330) compared with $P3^{SD}$ (497).

Barton and coworkers have described a complex RNA structure (RNase L ci RNA) mapping to P3 of all C-cluster enteroviruses (PV, C-Coxsackie A viruses) (Fig. 1*A*) that can function as an inhibitor for RNase L (25, 26). Site-directed mutagenesis inactivated the inhibitory activity for RNase L, but it did not disable virus replication (25, 26). This result is supported by our recoding experiments.

We then studied variants with two or three recoded domains (Fig. 1*C*). All Max and SD variants [P(1+2), P(2+3), P(1+3), and P(1+2+3)] were constructed with WT Cre and α and β elements, whereas the RNase L ci RNA structures were destroyed. All Max combinations produced WT growth phenotypes (Fig. 2, lines 2–5), an observation suggesting that the overrepresented synonymous codon pairs in these variants are acceptable for efficient replication, despite 1,297 synonymous mutations in $P(1+2+3)^{Max}$ (Fig. 1*C*, line 5). Importantly, the kinetics of virus replication, as analyzed by a one-step growth curve, did not reveal a detectable difference between WT PV1 and $P(1+2+3)^{Max}$ (Fig. S1). Extended SD recoding, however, yielded replication phenotypes as in $P(1+2)^{SD}$, $P(2+3)^{SD}$, $P(1+3)^{SD}$, and $P(1+2+3)^{SD}$ (Fig. 1*C*, lines 6–9). This includes variant $P(2+3)^{SD}$ that has fewer silent mutations (1,174 nt) than the $P(1+2+3)^{Max}$ variant (1,297 nt).

Analysis of Translation and Replication of $P2^{Min}$. The mechanisms by which Max, Min, or SD recoding influences the biological properties of PV are poorly understood. Here, we focused on $P2^{Min}$, the only viable Min construct. Translating $P2^{Min}$ transcripts in a HeLa cell-free extract (27) in the absence of CTP and UTP (to minimize RNA synthesis) did not disclose significant differences in synthesis or processing compared with WT PV transcripts (Fig. 2*A*). Analysis of growth of $P2^{Min}$ in HeLa cells, however, revealed a significant reduction of $P2^{Min}$ virus production when assayed by PFUs (Fig. S1). As we will show below, $P2^{Min}$ virus has a very low si value, which indicates the need of a large number of virions per cell for viral spread and plaque forming on a layer of HeLa cells. Infection of HeLa cells with $P2^{Min}$ virus [multiplicity of infection (MOI) of 2], however, revealed reduced production of $2C^{ATPase}$ and its precursors (to ∼30% compared with similar WT PV infection), as shown by immune precipitation of cellular extracts (Fig. 2*B*). Guanidine hydrochloride (GnHCl; 2 mM) completely inhibits PV RNA replication., Therefore, translation of incoming RNA was reduced (Fig. 2*B*).

We then tested the synthesis of $P2^{Min}$-specific RNA in HeLa cells (MOI of 2) by quantitative RT-PCR (qRT-PCR) and Northern analysis at 4 h postinfection (p.i.) (*SI Materials and Methods*). As shown in Fig. 2*C* (−GnHCl), RNA levels were reduced approximately twofold in $P2^{Min}$-infected HeLa cells

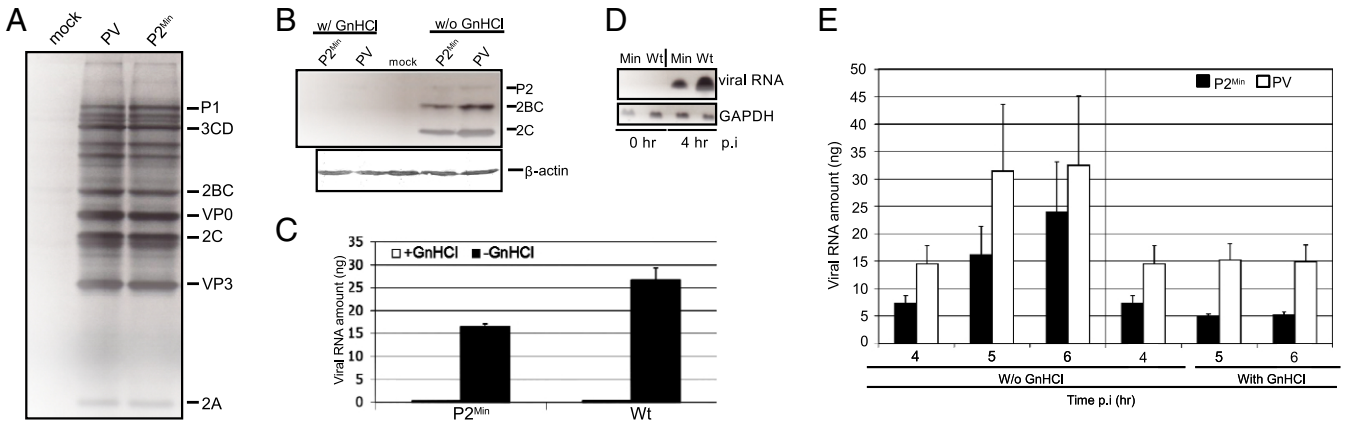


Fig. 2. Protein expression, and RNA levels in cells infected in vitro and in vivo with WT and recoded PVes. (*A*) In vitro translation of WT and $P2^{Min}$ recoded PV RNAs. Transcript RNAs of WT and $P2^{Min}$ constructs were translated in vitro in HeLa cell free extracts at 34 °C for 12 h in the presence of ^{35}S-methionine. Samples were analyzed by SDS/polyacrylamide gel electrophoresis. (*B*) Western analysis of $2C^{ATPase}$ related proteins produced in cells infected with WT and $P2^{Min}$ viruses (*Materials and Methods*). Two sets of HeLa cells were infected with WT or recoded viruses at a MOI of 2, and the incubation continued either in the absence or presence of 2 mM GnHCl. At 4 h p.i., lysates were prepared, and the lysates were probed with antibody to PV $2C^{ATPase}$ and its precursors. (*C*) The synthesis of PV RNA was measured by qRT-PCR 4 h p.i. of HeLa cells that were infected with WT or $P2^{Min}$ viruses (MOI of 2) either in the absence or presence of GnHCl. Total RNA was isolated from cell lysates and subjected to qRT-PCR analyses. (*D*) PV RNA in infected HeLa cells was measured by Northern analyses. The cells were infected at an MOI of 2 with WT and $P2^{Min}$ viruses, at 0 and 4 h p.i. RNAs in lysates were quantitated by Northern analysis. Antisense RNA probes were used for the detection of PV RNA (3′-NTR) and for mRNA of glyceraldehyde-3-phosphate dehydrogenase (GAPDH). (*E*) Stability of newly made WT or recoded RNAs in infected cells. Two sets of HeLa cells were infected with WT and $P2^{Min}$ viruses at a MOI of 2 and incubated for 6 h at 37 °C. At 4 h p.i., 2 mM GnHCl was added to one batch. Lysates of infected cells were made at 4, 5, and 6 h p.i., and the total RNA was isolated. The amount of PV-specific RNA was measured by qRT-PCR.

compared with the WT. If 2 mM GnHCl was added at the time of infection, no RNA synthesis could be observed (Fig. 2*C*, +GnHCl). Northern analyses using a probe complementary to the 3′-NTR of the PV genome were similar to the results obtained with qRT-PCR (Fig. 2*D*). To test the stability of $P2^{Min}$ RNA, RNA replication was aborted after 4 h p.i. by the addition of GnHCl. The fate of viral RNA was then monitored for the next 1 and 2 h. We observed a slight degradation of $P2^{Min}$ RNA compared with WT RNA (Fig. 2*E*). These results are difficult to interpret because PV translation in vivo is linked in cis to genome replication (28): A reduction of translation would interfere with genome synthesis. Defective genome synthesis, in turn, would interfere with translation. Whatever the mechanism, compared with WT PV, there is a small defect in the production of $P2^{Min}$ RNA.

Effect of Temperature on Plaque Assays of Recoded Viruses. An analysis of two of three recoded viruses revealed temperature-sensitive phenotypes (ts), as measured by plaque formation (Table S2). Specifically, $P(1+2+3)^{SD}$ and $P2^{Min}$ did not produce measurable yields of virus at 33 and 39.5 °C, whereas the $P(1+2+3)^{Max}$ variant yielded WT PV titers at 33, 37, and 39.5 °C.

SI PV and of Recoded PV Variants. The probability of a single polio virion to initiate an infection of a favored host cell (HeLa R19) under tissue culture conditions is low. Owing to numerous hostile conditions presented by the cell, most attempts by the virus to invade, establish proliferation, and possibly kill the cell are aborted. This phenomenon of restricted virus/cell interaction is referred to as specific infectivity (si), a phenotype reproducible under identical conditions of experimentation. *si* defines a known minimum number *n* of virions that, when plated on a huge excess of a host cell layer, will allow one virion of "*n*" to successfully initiate an infection. Commonly, the successful infection will be visible as a plaque. The term PFU, however, hides the number *n*: Depending on the virus or the cell, a PFU value may correspond to 100 virions of virus X in cell Y, whereas a PFU of the same virus X tested on a different cell Z may correspond to 1,000 virions.

It is known that plating an average of ∼120 WT PV1(M) virions onto a monolayer of 10^6 HeLa R19 cells produces one plaque (15). The *si* value, therefore, is 1/120. Note that in this experiment, the multiplicity of virion per cell is 0.00012 (120 virions per 10^6 monolayer cells on the experimental plate), which excludes the possibility that more than one virion is needed to produce a plaque. Therefore, the *si* of WT PV1(M) denotes that 1 of 120 virions produced the plaque, whereas 119 virions failed. The selection of the virion or of the single cell that leads to a successful infection is a stochastic process (*Discussion*). Importantly, *si* is the signature of the virus only when determined in related cell lines under comparable conditions.

The *si* values that we observed with WT PV1(M), $P(1+2+3)^{Max}$, and $P2^{Min}$ are presented in Table 1. For comparison, the observed *si* value of WT PV (1/118) was set at 1.0. In this context, the *si* of $P(1+2+3)^{Max}$ was almost identical to that of WT virus, which was surprising considering the large sequence differences in the PP coding region. More surprising was the minute *si* value of $P2^{Min}$ (1/35, 285), as it took 35,285 virions plated onto 10^6 R19 HeLa monolayer cells to produce a single plaque. Even in this experiment, the multiplicity of variant/cell is low (0.035285). Hence, the probability that more than one virion interacted with one cell is low. It has been speculated that small *si* values may reflect defective virions or clumping of the inoculum. However, neither hypothesis is likely to apply for PV (*Discussion*).

The reason for low to minute *si* values of WT PV1 and recoded variants (15, 16) is not yet understood, but might be manifold (ref. 15; *Discussion*). As mentioned, *si* values can change drastically when the conditions of the experiments are modified. To prove this point, we have determined *si* values of WT PV1, $P(1+2+3)^{Max}$, and $P2^{Min}$ on A549 cells that are known to express a robust innate immune response (29). We speculated that using A549 cells may reduce the chance of any of the lonely virions to produce a plaque. This was indeed the case: for PV1(M), the *n* value changed from 1/118 (R19 HeLa cells) to 1/442 (A549 cells), for $P(1+2+3)^{Max}$ from 1/141 (R19 HeLa) to 1/438 (A549), and for $P2^{Min}$ from ∼$1/10^5$ (R19 HeLa) to ∼$1/10^6$ (A549) (Table 1).

Table 1. SI of WT PV and recoded PVes variants in HeLa R19 and A549 cells in the absence and presence of TBK1/IKK inhibitor, BX795

		PV WT		$P(1+2+3)^{Max}$		$P2^{Min}$	
	Viral particles	2.83×10^{12}		2.63×10^{13}		4.9×10^{12}	
	TBK inhibitor (100 nM)	–	+	–	+	–	+
R19	Titer, PFU/mL	2.4×10^{10}	4×10^{10}	1.86×10^{11}	2.8×10^{11}	1.4×10^{8}	2.5×10^{8}
	si, particles/PFU	118 ± 9	71 ± 3	141 ± 6	94 ± 7	$3.5 \pm 0.2 \times 10^{4}$	$1.98 \pm 0.3 \times 10^{4}$
	Relative *si*	1	1.66	0.84	1.26	0.0033	0.006
A549	Titer, PFU/mL	6.4×10^{9}	8×10^{9}	6×10^{10}	6.8×10^{10}	1.2×10^{6}	2.6×10^{7}
	si, particles/PFU	442 ± 18	354 ± 32	438 ± 10	386 ± 39	$4.1 \pm 0.2 \times 10^{6}$	$1.9 \pm 0.2 \times 10^{5}$
	Relative *si*	1	1.25	1.01	1.15	0.0001	0.0023

All virus stocks were purified on CsCl gradients, and the number of particles or genome equivalents (1 genome = 1 virion) was calculated by the formula 9.4×10^{12} particles/mL = 1 OD_{260}. The infectious titer (PFU/mL) of each virus was determined by standard plaque assay (*Materials and Methods*). *si* is defined as the number of particles that are necessary to yield a single PFU after incubation with a vast excess of host cells (cell monolayer comprising 10^6 cells). Here, the *si* values were normalized to the *si* value of WT PV1.

If the *si* is dependent on the A549 host cell and not on the virus brand, an inhibitor that interferes with innate immunity should partially revert the *si* values to those of R19 HeLa cells. We have confirmed this with BX795, a potent inhibitor of *IKK*-related kinases and IFN-β production (30, 31). The altered *si* values with BX795-treated A549 cells did not reach the values with HeLa cells, an observation indicating other effects in the stochastic events not reversed by the drug. Note that BX795 had a significant effect also on HeLa R19 cells, reducing the *n* value for WT PV from 1/118 to 1/71.

We note that the *si* value gives a measure of virus particles produced by a plaque. If *si* is ∼1/35,000 ($P2^{Min}$ on HeLa cells), then the titer at the time of harvest is $(6.7 \times 10^6) \times 35,000$ [where (6.7×10^6) presents the titer of the variant; Fig. 1, line 6), that is, 2.34×10^{11} virions. The titer of WT virus (log10) in the same experiment was $(>10^9) \times 120$; that is, $>1.2 \times 10^{11}$ virions (Fig. 1, line 1, and Table 1). This calculation suggests that once $P2^{Min}$ has successfully established productive infection of a cell, its intracellular replication on HeLa cells is comparable to that of WT PV (*Discussion*). Nevertheless, because of the incredibly large amount of virus necessary to succeed in an infection, the plaques of these virus variants are small, if not minute, and any attempt to the spread of $P2^{Min}$ on the cell layer is greatly reduced.

Genetic Variation of Synthetic PVs. Recoding the ORF of the PV1(M) genome by the Max, Min, or SD strategies leads to profound differences in nucleotide sequences, each specific for the recoding strategy. As shown in Fig. 3 (lines B and C), we have analyzed synonymous site variability using the synplot2 program, which is capable of predicting conserved functional regions shared between similar virus RNAs (32, 33). Synplot2 was previously used to identify conserved sequence and structural elements among different alphavirus serotypes that contain a classical packaging signal ("PS"; below) (34). As shown in Fig. 3*B*, synplot2 predicts the location of the cis-acting elements Cre, RNase L ci, and α and β in genomes of 200 members of the C-Enterovirus genus (32). However, these structural elements do not play a role in PV encapsidation (12, 35), and no other regions have significantly reduced ($P < 0.001$) synonymous site variation. We guess that synonymous changes have likely been saturated since the PV ancestral strain diverged from Coxsackie viruses (12). It is, however, difficult to estimate a natural "limit" of variation due to the naturally high rate of synonymous variation outside of the conserved Cre, RNase L ci RNA, and α and β regions.

Fig. 3 is incomplete, as it lacks presentation of mutations in the capsid domain P1. We omitted P1 because the sequence in this region does not play a role in PV morphogenesis (Assembly of the Poliovirion). The P1 domain, however, has been added in Fig. S2. As expected, nonsynonymous mutations in the P1 domain outnumber those in the P2 and P3 domains, but at cursory inspection, the density of synonymous mutations is similar to that of P2 and P3. To improve the interpretation of the data in Fig. 3*F*, we added Table S3, which depicts the amount of synonymous changes from PV1(M) as a percentage of total nucleotides in the coding domain P1, P2, and P3 (36). Whereas the number of natural synonymous changes in the PVs and CAV20 is large, the changes acquired in $P(1+2+3)^{Max}$ or $PV(1+2+3)^{SD}$ outnumbers natural variation. It should be noted, however, that at a given CPS calculated for our Max or Min constructs, other sequences for Max and Min can exist (with variable identities of mutants) in part because there is no global Max or Min.

Assembly of the Poliovirion. Viral assembly must be highly specific to avoid packaging the abundant heterologous cellular nucleic acids in the infected cell. However, even in the absence of cognate genomes, most viral CPs carry the information to form 3D structures that resemble cognate virions (37), now called "virus like particles" [VLPs; (38)]. Indeed, VLPs expressed by many viruses in host cells fill their internal space with any RNA available (viral or cellular), whereby the abundance of RNAs in the cell can play a critical role in the selection VLP-associated RNAs (39–43).

In PV-infected cells, icosahedral "procapsids" (also known as "empty capsids") (35, 44) are formed that, remarkably, are void of viral or cellular RNAs. These procapsids are structurally closely related to PV virions (8), but the precise role of procapsids in PV morphogenesis is still unknown (*Discussion*). Unusual for +ssRNA viruses, the synthesized PV genome, while *in statu nascendi*, is delivered directly and specifically to cognate capsid precursors by protein/protein interaction. The mechanism used is the recognition of CP precursors by a key protein ($2C^{ATPase}$) of the membrane-bound replication complex. This mechanism is highly specific, as the PV $2C^{ATPase}$ discriminates against even closely related CP precursors (45–48). We suggest that C-cluster enteroviruses and possibly all viruses of the genus *Enterovirus* of *Picornaviridae* use protein/protein recognition for achieving specificity of assembly (Fig. 4 *I*, *A*).

We propose that the framework leading to specificity (Fig. 4 *I*, *A*) triggers morphogenesis of PV as depicted in Fig. 4*II*: perhaps by condensation of capsid precursors ("pentamers") around the genome (49), followed by virion maturation (8). Previous studies have elucidated PV CP assembly in great detail (35), but the role of the PV genome in these steps remains uncertain (*Discussion*). We have borrowed the term cocondensation from other publications (50, 51), as it indicates the progression of delicate steps between CP and the cognate genome (Fig. 4*II*) (*Discussion*).

A different mechanism leading to assembly specificity in +ssRNA virus morphogenesis has been recently proposed. It is based on the observation that viral genomes have the propensity to form dozens of genome-specific hairpins [Dykeman, 2013,

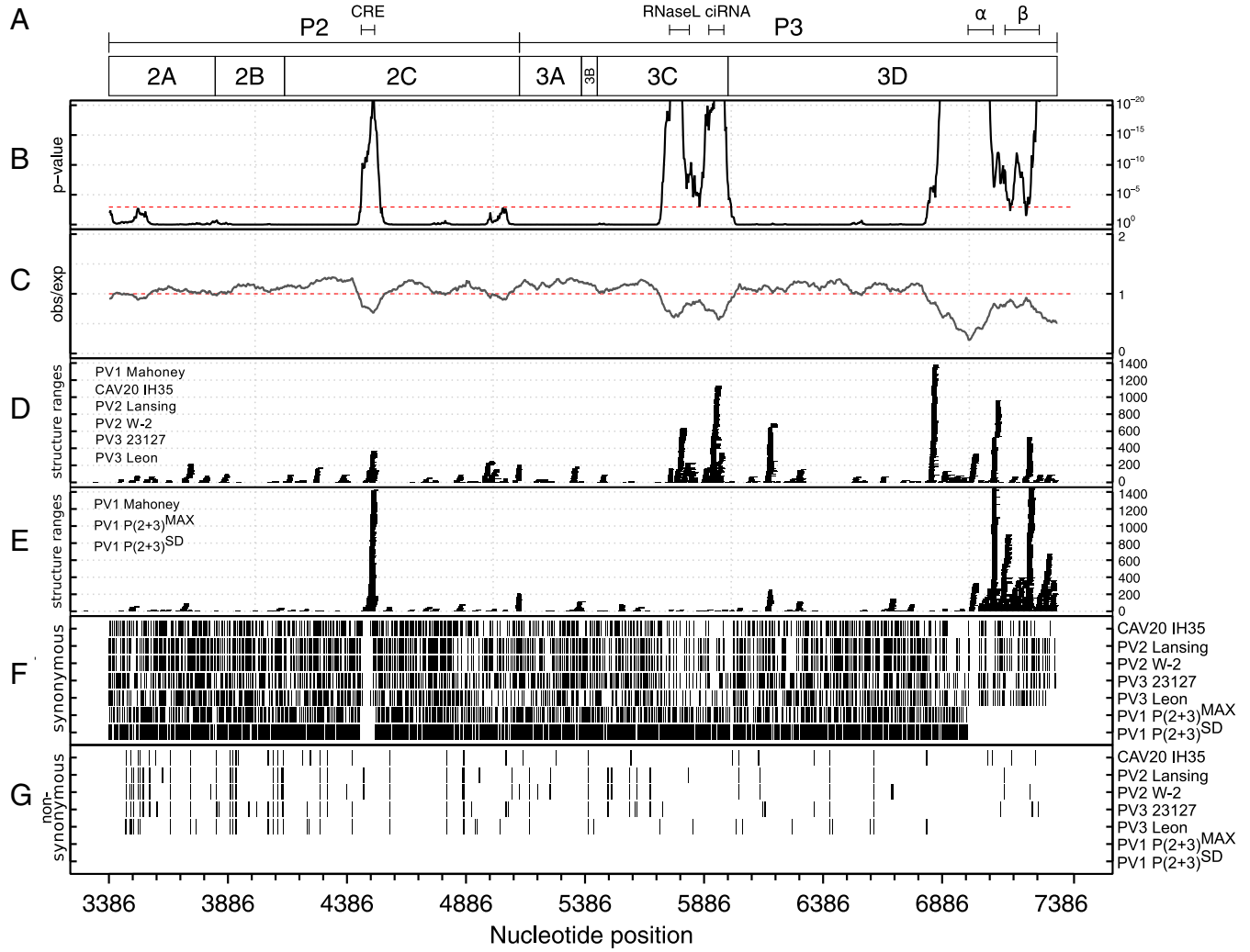


Fig. 3. An analysis of conserved RNA sequences and local secondary structures for viruses capable of assembling with a PV Mahoney capsid to produce infectious virus. (*A*) Cis-acting functional elements conserved in the RNA of enterovirus C cluster viruses are marked on the PVM sequence. The sequence regions mapping to these elements show a statistically significant reduction in variation at synonymous sites with respect to a neutral model of evolution among 200 enterovirus C viruses which share >75% amino acid identity with PVM (*B* and *C*). The dotted line in *B* signifies that the points above it are statistically significant ($P < 0.005$), and the dotted line in *C* represents the level at which variation at synonymous sites is observed as often as it is expected in a neutral model. *D* and *E* depict local secondary RNA structures conserved between PVM and natural viruses: Coxsackie A20, PV type 2, and PV type 3 (*D*), and synthetic viruses: Max and SD which were recoded over the P2 and P3 regions (*E*). A mathematically complete ensemble of structures was generated in 30-nt windows for each sequence with the crumple program (*SI Materials and Methods*). The resulting structures were filtered and compared between the viruses labeled at equivalent nucleotide positions. Each conserved structure is represented by a rectangle that spans the range of the structure's position. (*F* and *G*) Synonymous (*F*) and nonsynonymous (*G*) in the same alignment of virus RNAs as used in *D* and *E* are compared by pairwise alignment to PV1(M); single nucleotide mutations are depicted by dark bars.

no. 1019; no. 1103 (19); Rolfsson, 2016, no. 964 (21)], described as "*cryptic, sequence-degenerate, dispersed RNA packaging signals*" (22). These hairpins, which are loosely conserved and carry common "assembly motifs," have been shown in in vitro experiments to promote RNA-dependent assembly of viral capsids, including elaborate structural rearrangements of capsid subunits (refs. 19 and 21 and references therein). We were curious as to whether the PV genome does also display an array of loosely conserved assembly structures along its genome with essential functions in PV morphogenesis. We hypothesized that the ability of forming these structures would be disturbed by extensive recoding of the PV genome.

Our analysis of the PV1 genome was confined to domains 2 and 3 (Fig. 3). The 5'- and 3'-terminal RNA segments do not determine assembly specificity, as was shown (35, 52). Domain P1 has also been omitted because it does not play any role in morphogenesis based on properties of PV "defective interfering particles" (DI particles) (53, 54). PV DI particles naturally lack the P1 coding region. However, in competition experiments, DI particles outcompete WT virus in genome replication simply because their genome is shorter. Moreover, DI particles "interfere" with WT PV proliferation by stealing (*in trans*) the WT PV CPs. This observation not only supports the mechanism of protein/protein interaction, it also indicates that the large genome segment encoding the capsid precursor is not important for assembly.

Our analyses took into consideration that the choice of synonymous codons for some amino acid residues, and the degeneracy of nucleotide base pairing allows for the formation of conformationally identical RNA structures that do not share significant sequence similarity. We wanted to determine whether any RNA structures could be present that are encoded by

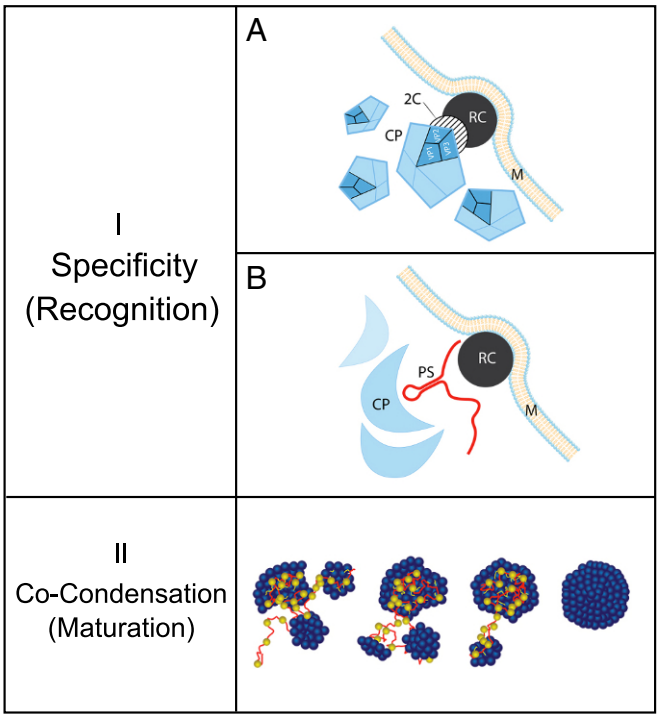


Fig. 4. Hypothesis describing the two-step assembly of +ssRNA viruses. (*I*) Two pathways leading to specificity/recognition. (*I, A*) Binding of a member of the viral replication complex (RC) to the cognate capsid precursor (protein/protein interaction). This interaction has been tailored to the known recognition process in PV assembly (47) where the protein of the RC is the $2C^{ATPase}$ (hatched) recognizing a capsid pentamer [$(VP0, VP3, VP1)_5$]. In this assembly complex, the $2C^{ATPase}$ acts possibly as an oligomer (82). (*I, B*) High-affinity binding of a precursor CP to a classical PS of the viral genome. PS is a cognate RNA structural unit unique to an RNA genome (RNA/protein interaction). M, membrane; PS, classical PS; RC, replication complex. (*II*) Cocondensation of capsid precursors with the cognate RNA genome or with cellular RNAs in the absence of viral genomes (text); orange dots, possible "cryptic, sequence-degenerate, dispersed RNA packaging signals", mapping along the entire viral genome (22) aiding genome/CP condensation. *II* is a modification of a figure published by Perlmutter and Hagan (83) (used with permission).

degenerate sequences that could have escaped detection with Synplot2. We used the Crumple program (55) and generated all possible RNA structures in 30-nt sliding windows of sequences belonging to the naturally occurring PV1, -2, -3, and CAV20, all of which can be encapsidated into PV1, 2, 3 capsids to produce infectious viruses. We then compared the profile of conserved RNA structures to our synthetic viruses, Max and SD (Figs. 3*D* and 4*E*). The structures, if they were conserved within the recoded segments of MAX and SD, were distinctly reduced compared with those cooccurring in Fig. 3*D*, despite the fact that only synonymous mutations were introduced (Figs. 3*F* and 4*G*). This result demonstrates that few, if any, small conserved RNA structures exist in the P2 and P3 regions of the analyzed viruses that could accommodate the hypothesis of "multiple cryptic, sequence-degenerate, dispersed RNA packaging signals" (22). Manual inspection of structures conserved after synonymous recoding revealed that the majority of prevailing structures occurred over regions encoding amino acids that had reduced codon choice, such that selection of any possible synonymous codon would conserve the presence or absence of base-pairing.

Variants $P(2+3)^{Max}$ (723/3,549 nt changes) and $P(1+2+3)^{Max}$ (1,297/6,189 nt changes) proliferate with WT PV1 characteristics [for $P(1+2+3)^{Max}$, see the one-step growth curve presented in Fig. S1], even though approximately every fifth nucleotide was changed. Absence of any known replication phenotype distinct from WT PV in the life cycle of these variants is astounding. Variants $P(1+2+3)^{SD}$ (2,103/6,177) and $P(2+3)^{SD}$ (1,174/3,549), in which approximately every third nucleotide was changed, were debilitated but they still encapsidated their genomes. The combined data of (*i*) the excessive density of mutations in recoded domains of P2 and P3 variants, yet allowing replication and packaging; (*ii*) the known molecular steps of the PV replication and packaging cycle; and (*iii*) the superior assembly efficiency of DI particles (lacking the P1 domain) over WT PV, lead us to conclude that "multiple cryptic, sequence-degenerate, dispersed RNA packaging signals" (22) do not play a role in PV packaging specificity.

Discussion

Curiosity motivated us to test the extent by which the nucleotide sequence of >80% of the PV genome can be modified while preserving the ability of the variants to proliferate. Avoiding sequences involved in the regulation of macromolecular processes, we followed two strategies: (*i*) changing synonymous codons such that an excess of codon pairs is generated either with average positive (Max) or average negative (Min) CPSs (16); and (*ii*) scrambling (moving) randomly synonymous codons (SD), leaving the average CPS close to that of WT PV (15). In either method, codon use and the protein sequences were retained.

Whereas Max and SD recodings did not destroy viability, CPD (Min constructs), [e.g., an accumulation of underrepresented (unfavorable) codon pairs] inevitably interfered with viral replication and is most often associated with nonviability (Fig. 1 *i* and *C*). By testing CPD of various mammalian viruses (16, 56–60) in tissue culture or in experimental animals, we have concluded that there is no single specific effect of CPD that can be related with certainty to observed deficiencies in viral proliferation of Min constructs.

Possible mechanisms supported in part by this work include: (*i*) an increase in the content of specific dinucleotides in CPD sequences, specifically CpG and UpA, which are two additional dinucleotides that occur in NNCpGNN or NNUpANN positions of new codon pairs and are, therefore, increased without changing codon use. The increase in CpG content (Table S4) may lead to an enhanced innate immune response in the infected cell (56, 61); (*ii*) a robust ts phenotype (59, 60); (*iii*) a reduction (degradation?) of the yield of viral mRNA (57); (*iv*) a reduction of viral protein synthesis in the infected cells (57); (*v*) protein misfolding of a recoded viral protein (62); and (*vi*) a vastly reduced si of virions (16).

Specific Infectivity. The importance of the phenotype *si* is not widely appreciated. In principle, all RNA viruses known to us have low *si* values (e.g., low probabilities that a single virion succeeds in infecting a suitable host cells under tissue culture conditions). Multiple reasons could underlie low *si*: the physical state or genotype of virions or the physiological state of single host cells (in the huge excess of target cells). Generally, abortion of infection can occur at any stage of the infectious cycle. The *si* value of PV1(M) (1/120, HeLa R19) is in a range common to other viruses [e.g., for influenza virus (PR8) (63) or VSV (~1/10–50 for these negative-stranded RNA viruses) (64)]. The surprisingly low *si* values of codon-pair deoptimized PV variants, observed also by Coleman et al. (16), are not the result of a defect in virions because: (*i*) sequence analyses did not reveal mutated genomes; (*ii*) all variants have been purified by CsCl gradient centrifugation where they presented with WT PV density; (*iii*) the *si* phenotype of variants was variable with different host cells and, importantly, similar to changes with WT PV (see also *Note*).

The experiments with A549 cells were revealing because an inhibitor (BX795) of the innate immune response (30, 31)

reversed to some extent the low *si* values in these cells of the three viruses tested. Whereas the innate immune response is likely to play a role in reducing the replication of recoded viruses (57, 61), attempts to identify any specific step related to innate immunity have failed (61).

In determining *si* values, the MOI (virion/cell) never exceeded the value of one. We suggest that the success of one virion over all other virions incubated on a huge excess of host cells is a stochastic process. This conclusion was also reached recently by two other groups (65, 66). Since PV is haploid, studies of *si* present the infection of a single cell with a single genotype, notwithstanding the fact that PV is a quasispecies that will form genetic variants later during its replicative cycle.

Assembly. The key in virus encapsidation is specificity. It requires a mechanism by which viral components (genome and CPs) recognize each other and assemble to near total exclusion of cellular or other viral components. We have discussed that the assembly specificity of C-cluster enteroviruses (PV, several Coxsackie A viruses), and perhaps of all enteroviruses, is achieved through viral protein/protein interaction (Fig. 4 *I*, *A*; ref. 47). A mechanism of protein/protein interaction to promote assembly specificity is involved also in mechanisms of other viral systems (67–69).

A common, alternative mechanism in +ssRNA virus assembly is the specific, high-affinity binding of a genomic RNA element to the cognate CP (Fig. 4 *I*, *B*). Available evidence suggests that this highly specific selection between genome and CP will trigger assembly (Fig. 4 *I*, *B*). For two decades, this RNA element has been referred to as PS (70). The classical PS (*i*) is a single RNA hairpin (70, 71) or a closely mapping cluster of structures (34), which bind strongly to a specific site in a CP; (*ii*) maps to a single genome locus that is distinct for related viruses (34); (*iii*) presents with a sequence that is conserved for a given virus species. Conservation of the classical PS allows the direct discovery of the PS by synplot2 computer analyses (32, 33); (*iv*) triggers assembly in vitro or in vivo that can be prevented by site-directed mutagenesis of even single nucleotides (71); (*v*) can be transplanted into a different RNA bestowing to this heterologous RNA the ability to be encapsidated into the PS-cognate VLP (42); or (*vi*) can be transplanted into a different genome, thereby generating a chimeric virus (70).

The known mechanisms triggering specificity of assembly (protein/protein or RNA/protein recognition) are portrayed in Fig. 4*I*. We propose that the "cryptic, sequence-degenerate, dispersed RNA packaging signals" (22) provide a function in step II of morphogenesis. The cocondensation of genome with cognate CP, which have been characterized in genomes of RNA phage MS2 (19, 21), of plant virus satellite tobacco necrosis virus (72), and of two human viruses (22, 73, 74), is to facilitate "cocondensation" (step II). Cocondensation includes highly complex rearrangements of precursor CPs, ultimately leading to mature virions (19, 22, 50, 51, 75). To avoid confusion, we strongly suggest that, in distinction of the highly conserved entities called PSs, the recently discovered "cryptic, sequence-degenerate, dispersed RNA packaging signals" be abbreviated as dPS.

PV assembly depends on ongoing genome replication (76). Newly synthesized genomes carry a 5'-terminal VPg that is cleaved off (77, 78) by a cellular enzyme once the RNA is released from the replication complex. The resulting infectious viral genomes lack VPg and are never encapsidated, an observation suggesting that putative RNA structures within the PV genome per se do not trigger assembly. The question then arises: How are mature PV virions formed at the replication complex past the recognition process (Fig. 4 *I*, *A*), even though the PV genome lacks dPSs?

Currently, we cannot answer this question. However, we are entertaining the thrilling hypothesis that the empty PV procapsids might serve as direct precursors to virions by first docking to the replication complex via protein/protein interaction followed by being filled with emerging genomic RNA, the "stuffing" catalyzed by the viral ATPase $2C^{ATPase}$. As has been noted before (79), this strategy would be analogous to the morphogenesis of numerous DNA viruses.

It is likely that the pathway of +ssRNA virus assembly shown in Fig. 4 may have to be modified in the future as more details of +ssRNA virus morphogenesis emerge. It may be prudent, therefore, not to follow just one "Yellow Brick Road" (80) because this may lead to the wrong wizard.

Materials and Methods

Recoding and Plasmids Design. These are described in *SI Materials and Methods*.

RNA Structure Alignment. This is described in *SI Materials and Methods*.

Cell Culture. Both HeLa R19 cells and A549 (human lung epithelial carcinoma) cells were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin, and 100 μg/mL streptomycin. All cells were grown at 37 °C in a 5% CO_2 incubator.

Virus Growth and Virus Titration by Plaque Assay. Viruses were amplified by infection of HeLa R19 cell monolayers with ≥2 PFUs per cell. Infected cells were incubated in DMEM (2% FBS) at 37 °C until a complete cytopathic effect (CPE) was detected. After three rounds of freezing and thawing, the lysate was clarified of cell debris by a brief high-speed centrifugation, and the supernatant containing the virus was used for further passaging or reinfection. For temperature sensitivity test, SD-, Max-, and Min-containing constructs were transfected into HeLa R19 cells. If there was no viable virus, the cell lysate was harvested for blind passages up to six times. Viral titers were determined by a standard plaque assay on HeLa R19 cell monolayers using a semisolid overlay of 0.6% tragacanth gum (Sigma-Aldrich) in minimal Eagle's medium. Plaques were visualized after 2 d of incubation (for normal plaque size) or 3- to 4-d incubation (for pinpoint plaque size), by staining cells with crystal violet.

General Procedures. Reverse transcription-PCR (RT-PCR); sequencing of virus variants; in vitro transcription and RNA transfection; labeling of PV proteins in vitro; virus purification and determination of viral particles via OD_{260} absorbance; Northern blot analysis; measurement of viral protein levels by Western blot analysis; and assay for specific infectivity are described in *SI Materials and Methods*.

Note. During review of this paper, Aguilera et al., published a study [mBio, March/April 2017 (84)] concluding that in infections with purified PV (expressing no phenotype) at low MOI (0.000001 PFU/cell) approximately 5% of the cells were infected with >1 virion. Dual infections increased (to 8%) when infections with two different PV mutants were analyzed and assayed phenotypically.

ACKNOWLEDGMENTS. We thank numerous colleagues with whom we discussed aspects of RNA virus assembly, amongst them H. G. Kräusslich, J. M. Hogle, M. F. Hagan, R. J. Kuhn, S. B. Larson, B. D. Lindenbach, E. I. Frolova, I. Frolov, A. L. N. Rao, D. J. Rowlands, P. G. Stockley, P. E. Prevelidge, and S. P. J. Whelan. We thank D. Peabody for introducing us into the genetics and assembly of phage MS2 and M. F. Hagan for the permission to use a modified figure published previously. This work was supported in part by National Institute of Health Grants R01AI07521901 and 5R37AI15122.

1. Wimmer E, Paul AV (2011) Synthetic poliovirus and other designer viruses: What have we learned from them? *Annu Rev Microbiol* 65:583–609.
2. Hahn EEA (1972) Polioviruses. *Strains of Human Viruses*, ed Plotkin SA (S. Karger AG, Basel), pp 155–176.
3. Kitamura N, et al. (1981) Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature* 291:547–553.
4. Racaniello VR, Baltimore D (1981) Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proc Natl Acad Sci USA* 78:4887–4891.
5. Wimmer E, Hellen CU, Cao X (1993) Genetics of poliovirus. *Annu Rev Genet* 27: 353–436.
6. Wimmer E, Paul AV (2010) The making of a picornavirus genome. *The Picornaviruses*, ed Ehrenfeld E (ASM, Washington, DC), pp 33–55.
7. Jacobson MF, Baltimore D (1968) Polypeptide cleavages in the formation of poliovirus proteins. *Proc Natl Acad Sci USA* 61:77–84.
8. Basavappa R, et al. (1994) Role and mechanism of the maturation cleavage of VP0 in poliovirus assembly: Structure of the empty capsid assembly intermediate at 2.9 a resolution. *Protein Sci* 3:1651–1669.

9. Hogle JM, Chow M, Filman DJ (1985) Three-dimensional structure of poliovirus at 2.9 A resolution. *Science* 229:1358–1365.
10. Mueller S, Wimmer E, Cello J (2005) Poliovirus and poliomyelitis: A tale of guts, brains, and an accidental event. *Virus Res* 111:175–193.
11. Domingo E, Holland JJ (1988) High error rates population equilibrium, and evolution of RNA replication systems. *RNA Genetics*, eds Domingo EHJ, Ahlquist P (CRC, Boca Raton, FL), Vol 3, pp 3–36.
12. Jiang P, et al. (2007) Evidence for emergence of diverse polioviruses from C-cluster coxsackie A viruses and implications for global poliovirus eradication. *Proc Natl Acad Sci USA* 104:9457–9462.
13. Gromeier M, Wimmer E, Gorbalenya AE (1999) Genetics, pathogenesis and evolution of picornaviruses. *Origin and Evolution of Viruses*, eds Domingo E, Parrish CR, Holland JJ (Academic, London), pp 287–343.
14. Plotkin JB, Kudla G (2011) Synonymous but not the same: The causes and consequences of codon bias. *Nat Rev Genet* 12:32–42.
15. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E (2006) Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol* 80:9687–9696.
16. Coleman JR, et al. (2008) Virus attenuation by genome-scale changes in codon pair bias. *Science* 320:1784–1787.
17. Gutman GA, Hatfield GW (1989) Nonrandom utilization of codon pairs in Escherichia coli. *Proc Natl Acad Sci USA* 86:3699–3703.
18. Moura GR, et al. (2007) Codon-triplet context unveils unique features of the Candida albicans protein coding genome. *BMC Genomics* 8:444.
19. Dykeman EC, Stockley PG, Twarock R (2013) Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *J Mol Biol* 425:3235–3249.
20. Patel N, et al. (2015) Revealing the density of encoded functions in a viral RNA. *Proc Natl Acad Sci USA* 112:2227–2232.
21. Rolfsson O, et al. (2016) Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. *J Mol Biol* 428:431–448.
22. Patel N, et al. (2017) HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. *Nat Microbiol* 2:17098.
23. Song Y, et al. (2012) Identification of two functionally redundant RNA elements in the coding sequence of poliovirus using computer-generated design. *Proc Natl Acad Sci USA* 109:14301–14307.
24. Burrill CP, et al. (2013) Global RNA structure analysis of poliovirus identifies a conserved RNA structure involved in viral replication and infectivity. *J Virol* 87:11670–11683.
25. Han JQ, et al. (2007) A phylogenetically conserved RNA structure in the poliovirus open reading frame inhibits the antiviral endoribonuclease RNase L. *J Virol* 81:5561–5572.
26. Townsend HL, et al. (2008) A viral RNA competitively inhibits the antiviral endoribonuclease domain of RNase L. *RNA* 14:1026–1036.
27. Molla A, Paul AV, Wimmer E (1991) Cell-free, de novo synthesis of poliovirus. *Science* 254:1647–1651.
28. Novak JE, Kirkegaard K (1994) Coupling between genome translation and replication in an RNA virus. *Genes Dev* 8:1726–1737.
29. Dove BK, et al. (2012) A quantitative proteomic analysis of lung epithelial (A549) cells infected with 2009 pandemic influenza A virus using stable isotope labelling with amino acids in cell culture. *Proteomics* 12:1431–1436.
30. Clark K, Plater L, Peggie M, Cohen P (2009) Use of the pharmacological inhibitor BX795 to study the regulation and physiological roles of TBK1 and IkappaB kinase epsilon: A distinct upstream kinase mediates Ser-172 phosphorylation and activation. *J Biol Chem* 284:14136–14146.
31. Feldman RI, et al. (2005) Novel small molecule inhibitors of 3-phosphoinositide-dependent kinase-1. *J Biol Chem* 280:19867–19874.
32. Firth AE (2014) Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. *Nucleic Acids Res* 42:12425–12439.
33. Firth AE, Wills NM, Gesteland RF, Atkins JF (2011) Stimulation of stop codon readthrough: Frequent presence of an extended 3′ RNA structural element. *Nucleic Acids Res* 39:6679–6691.
34. Kim DY, Firth AE, Atasheva S, Frolova EI, Frolov I (2011) Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. *J Virol* 85:8022–8036.
35. Jiang P, Liu Y, Ma HC, Paul AV, Wimmer E (2014) Picornavirus morphogenesis. *Microbiol Mol Biol Rev* 78:418–437.
36. Jorba J, Campagnoli R, De L, Kew O (2008) Calibration of multiple poliovirus molecular clocks covering an extended evolutionary range. *J Virol* 82:4429–4440.
37. Berger B, Shor PW, Tucker-Kellogg L, King J (1994) Local rule-based theory of virus shell assembly. *Proc Natl Acad Sci USA* 91:7732–7736.
38. Frietze KM, Peabody DS, Chackerian B (2016) Engineering virus-like particles as vaccine platforms. *Curr Opin Virol* 18:44–49.
39. Comas-Garcia M, Davis SR, Rein A (2016) On the selective packaging of genomic RNA by HIV-1. *Viruses* 8:E246.
40. Lu K, Heng X, Summers MF (2011) Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol* 410:609–633.
41. Peabody DS (1997) Role of the coat protein-RNA interaction in the life cycle of bacteriophage MS2. *Mol Gen Genet* 254:358–364.
42. Pickett GG, Peabody DS (1993) Encapsidation of heterologous RNAs by bacteriophage MS2 coat protein. *Nucleic Acids Res* 21:4621–4626.
43. Rulli SJ, Jr, et al. (2007) Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J Virol* 81:6623–6631.
44. Putnak JR, Phillips BA (1981) Picomaviral structure and assembly. *Microbiol Rev* 45:287–315.
45. Asare E, Mugavero J, Jiang P, Wimmer E, Paul AV (2016) A single amino acid substitution in the nonstructural protein 2CATPase of poliovirus causes conditional defects in encapsidation and uncoating. *J Virol* 90:6174–6186.
46. Jia XY, Van Eden M, Busch MG, Ehrenfeld E, Summers DF (1998) trans-encapsidation of a poliovirus replicon by different picornavirus capsid proteins. *J Virol* 72:7972–7977.
47. Liu Y, et al. (2010) Direct interaction between two viral proteins, the nonstructural protein 2C and the capsid protein VP3, is required for enterovirus morphogenesis. *PLoS Pathog* 6:e1001066.
48. Wang C, Jiang P, Sand C, Paul AV, Wimmer E (2012) Alanine scanning of poliovirus 2CATPase reveals new genetic evidence that capsid protein/2CATPase interactions are essential for morphogenesis. *J Virol* 86:9964–9975.
49. Ghendon Y, Yakobson E, Mikhejeva A (1972) Study of some stages of poliovirus morphogenesis in MiO cells. *J Virol* 10:261–266.
50. Schneemann A (2006) The structural and functional role of RNA in icosahedral virus assembly. *Annu Rev Microbiol* 60:51–67.
51. Prasad BV, Schmid MF (2012) Principles of virus structural organization. *Adv Exp Med Biol* 726:17–47.
52. Barclay W, et al. (1998) Encapsidation studies of poliovirus subgenomic replicons. *J Gen Virol* 79:1725–1734.
53. Cole CN, Smoler D, Wimmer E, Baltimore D (1971) Defective interfering particles of poliovirus. I. Isolation and physical properties. *J Virol* 7:478–485.
54. Nomoto A, Jacobson A, Lee YF, Dunn J, Wimmer E (1979) Defective interfering particles of poliovirus: Mapping of the deletion and evidence that the deletions in the genomes of DI(1), (2) and (3) are located in the same region. *J Mol Biol* 128:179–196.
55. Bleckley S, Stone JW, Schroeder SJ (2012) Crumple: A method for complete enumeration of all possible pseudoknot-free RNA secondary structures. *PLoS One* 7:e52414.
56. Shen SH, et al. (2015) Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. *Proc Natl Acad Sci USA* 112:4749–4754.
57. Yang C, Skiena S, Futcher B, Mueller S, Wimmer E (2013) Deliberate reduction of hemagglutinin and neuraminidase expression of influenza virus leads to an ultraprotective live vaccine in mice. *Proc Natl Acad Sci USA* 110:9481–9486.
58. Ni YY, et al. (2014) Computer-aided codon-pairs deoptimization of the major envelope GP5 gene attenuates porcine reproductive and respiratory syndrome virus. *Virology* 450–451:132–139.
59. Le Nouen C, et al. (2014) Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. *Proc Natl Acad Sci USA* 111:13169–13174.
60. Le Nouen C, et al. (2017) Genetic stability of genome-scale deoptimized RNA virus vaccine candidates under selective pressure. *Proc Natl Acad Sci USA* 114:E386–E395.
61. Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P (2014) The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res* 42:4527–4545.
62. Wang B, et al. (2015) Recoding of the vesicular stomatitis virus L gene by computer-aided design provides a live, attenuated vaccine candidate. *MBio* 6:e00237–15.
63. Pauly MD, Lauring AS (2015) Effective lethal mutagenesis of influenza virus by three nucleoside analogs. *J Virol* 89:3584–3597.
64. Arstila P (1976) Quantitative studies on adsorption, elution, and haemagglutination of vesicular stomatitis virus. *Arch Virol* 51:51–58.
65. Schulte MB, Andino R (2014) Single-cell analysis uncovers extensive biological noise in poliovirus replication. *J Virol* 88:6205–6212.
66. Staring J, et al. (2017) PLA2G16 represents a switch between entry and clearance of Picornaviridae. *Nature* 541:412–416.
67. Chaturvedi S, Rao AL (2014) Live cell imaging of interactions between replicase and capsid protein of Brome mosaic virus using bimolecular fluorescence complementation: Implications for replication and genome packaging. *Virology* 464–465: 67–75.
68. Hesketh EL, et al. (2015) Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. *Nat Commun* 6:10113.
69. Venter PA, Schneemann A (2008) Recent insights into the biology and biomedical applications of flock house virus. *Cell Mol Life Sci* 65:2675–2687.
70. Qu F, Morris TJ (1997) Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. *J Virol* 71:1428–1435.
71. Witherell GW, Gott JM, Uhlenbeck OC (1991) Specific interaction between RNA phage coat proteins and RNA. *Prog Nucleic Acid Res Mol Biol* 40:185–220.
72. Larson SB, McPherson A (2001) Satellite tobacco mosaic virus RNA: Structure and implications for assembly. *Curr Opin Struct Biol* 11:59–65.
73. Kalynych S, Palkova L, Plevka P (2015) The structure of human parechovirus 1 reveals an association of the RNA genome with the capsid. *J Virol* 90:1377–1386.
74. Shakeel S, et al. (2017) Genomic RNA folding mediates assembly of human parechovirus. *Nat Commun* 8:5, and erratum (2017) 8:83.
75. Perlmutter JD, Hagan MF (2015) The role of packaging sites in efficient and specific virus assembly. *J Mol Biol* 427:2451–2467.
76. Nugent CI, Johnson KL, Sarnow P, Kirkegaard K (1999) Functional coupling between replication and packaging of poliovirus replicon RNA. *J Virol* 73:427–435.
77. Nomoto A, Kitamura N, Golini F, Wimmer E (1977) The 5′-terminal structures of poliovirion RNA and poliovirus mRNA differ only in the genome-linked protein VPg. *Proc Natl Acad Sci USA* 74:5345–5349.
78. Virgen-Slane R, et al. (2012) An RNA virus hijacks an incognito function of a DNA repair enzyme. *Proc Natl Acad Sci USA* 109:14634–14639.
79. Jacobson MF, Baltimore D (1968) Morphogenesis of poliovirus. I. Association of the viral RNA with coat protein. *J Mol Biol* 33:369–378.

80. Prevelidge PE (2016) Follow the yellow brick road: A paradigm shift in virus assembly. *J Mol Biol* 428:416–418.
81. Goodfellow I, et al. (2000) Identification of a cis-acting replication element within the poliovirus coding region. *J Virol* 74:4590–4600.
82. Adams P, Kandiah E, Effantin G, Steven AC, Ehrenfeld E (2009) Poliovirus 2C protein forms homo-oligomeric structures required for ATPase activity. *J Biol Chem* 284:22012–22021.
83. Perlmutter JD, Hagan MF (2015) Mechanisms of virus assembly. *Annu Rev Phys Chem* 66:217–239.
84. Aguilera ER, Erickson AK, Jesudhasan PR, Robinson CM, Pfeiffer JK (2017) Plaques formed by mutagenized viral populations have elevated coinfection frequencies. *mBio* 8:e02020-16.
85. Cello J, Paul AV, Wimmer E (2002) Chemical synthesis of poliovirus cdna: Generation of infectious virus in the absence of natural template. *Science* 297:1016–1018.
86. van der Werf S, Bradley J, Wimmer E, Studier FW, Dunn JJ (1986) Synthesis of infectious poliovirus RNA by purified T7 RNA polymerase. *Proc Natl Acad Sci USA* 83:2330–2334.
87. Rueckert RR (1985) Picornaviruses and their replication. *Virology*, eds Fields BN, Knipe DM, Chanock RM, Melnick JL, Roizman B, Shope RE (Raven, New York), Vol 1, pp 705–738.
88. Bleckley S, Schroeder SJ (2012) Incorporating global features of RNA motifs in predictions for an ensemble of secondary structures for encapsidated MS2 bacteriophage RNA. *RNA* 18:1309–1318.
89. Lawrence M, et al. (2013) Software for computing and annotating genomic ranges. *PLoS Comput Biol* 9:1–10.
90. Rima BK, McFerran NV (1997) Dinucleotide and stop codon frequencies in single-stranded RNA viruses. *J Gen Virol* 78:2859–2870.
91. Rothberg P, Wimmer E (1981) Mononucleotide and dinucleotide frequencies, and codon usage in poliovirion RNA. *Nucleic Acids Res* 9:6221–6230.