Biophysical Journal, Volume *120*

Supplemental Information

Frustration and Direct-Coupling Analyses to Predict Formation and Function of Adeno-Associated Virus

Nicole N. Thadani, Qin Zhou, Kiara Reyes Gamas, Susan Butler, Carlos Bueno, Nicholas P. Schafer, Faruck Morcos, Peter G. Wolynes, and Junghae Suh

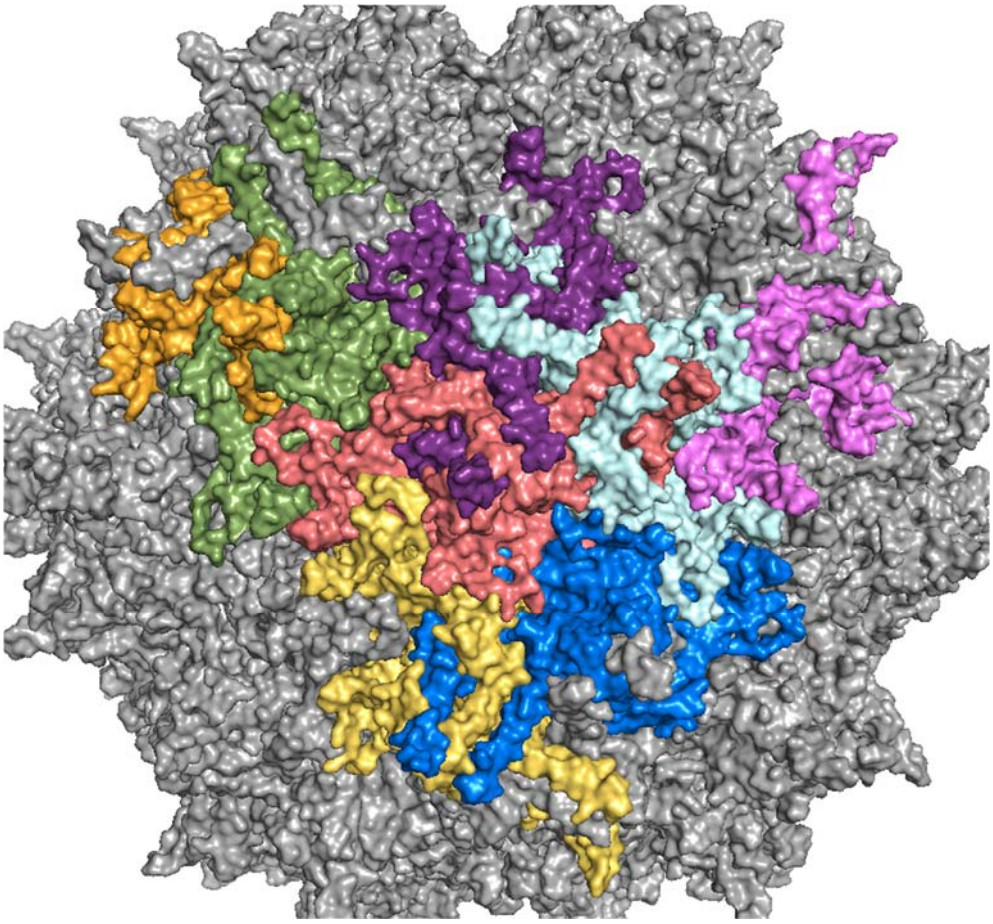


Figure S1: AAV2 monomer and multimer substructures (related to Fig. 1). The AAV2 monomer (coral) is defined as one VP3 subunit. The AAV2 multimer is defined as the central monomer (coral) and the seven surrounding monomers containing a residue with any atom within a radius of 4.5Å around any of the central monomer residues. The rest of the AAV capsid (grey) was not included in this analysis.

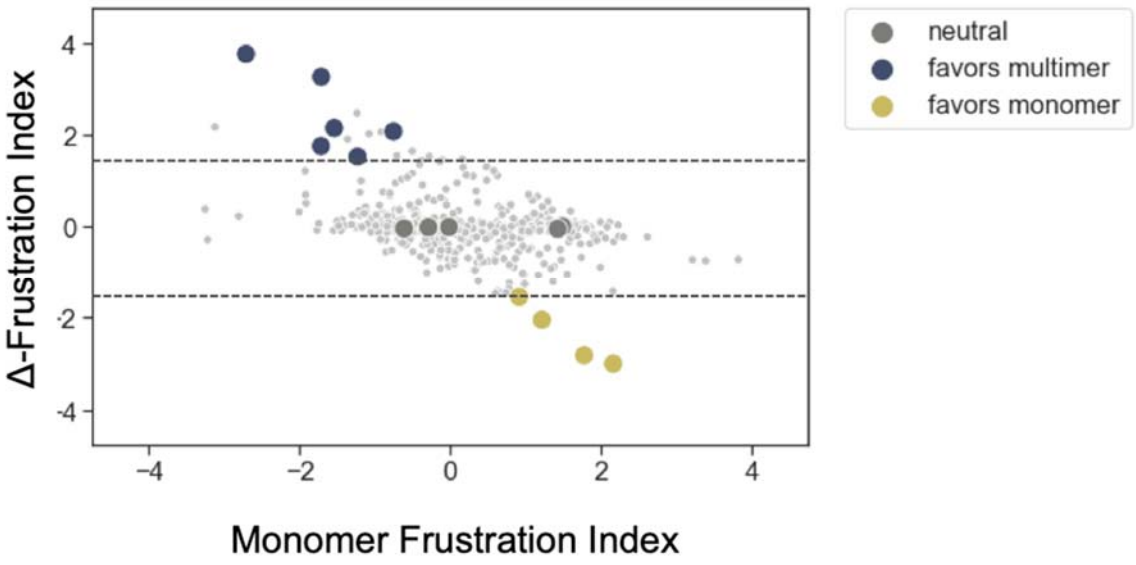


Figure S2: AAV2 Δ-frustration index plotted against the monomer frustration index (related to Fig. 2). The AAV2 Δ-frustration index is computed as the difference between the single-residue mutational frustration index of the multimer and the single-residue mutational frustration index of the monomer. Residues with a positive Δ-frustration index are predicted to energetically favor the assembly state, while residues with a negative Δ-frustration index are predicted to favor the monomer. Residues selected for mutation are large colored circles. Residues not selected for mutation are small grey dots. Dashed lines indicate the +/- 1.5 Δ-frustration index thresholds used to select residues for mutation.

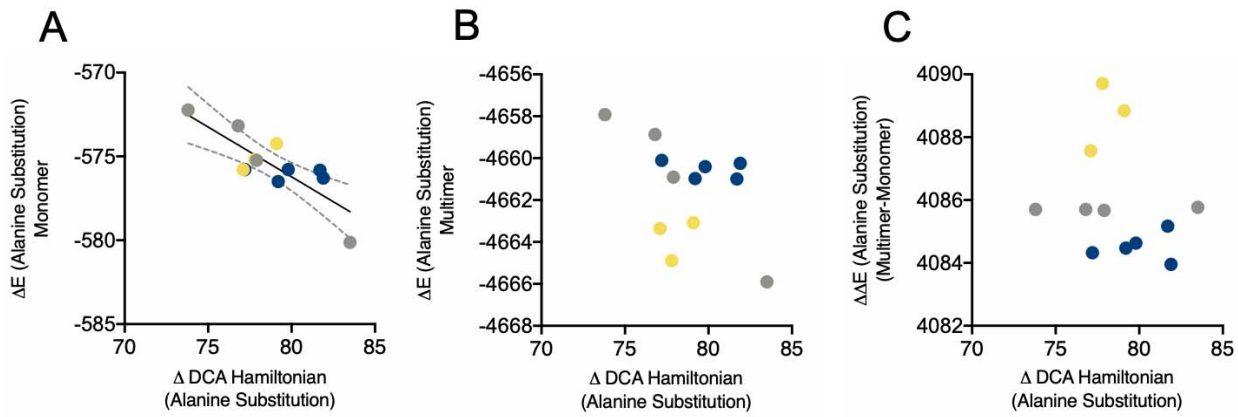


Figure S3: AAV2 $\Delta H_{DCA(Ala)}$ plotted against ΔE-monomer, ΔE-multimer, and ΔΔE of alanine substitutions (related to Fig. 3). AAV2 $\Delta H_{DCA(Ala)}$ is plotted against ΔE-monomer **(A)**, ΔE-multimer **(B)**, and ΔΔE **(C).** N=12. R471A, F712A, and S721A were excluded from these analyses due to a high gap frequency at their alignment position. Regression lines with 95% confidence interval are plotted for statistically significant correlations.

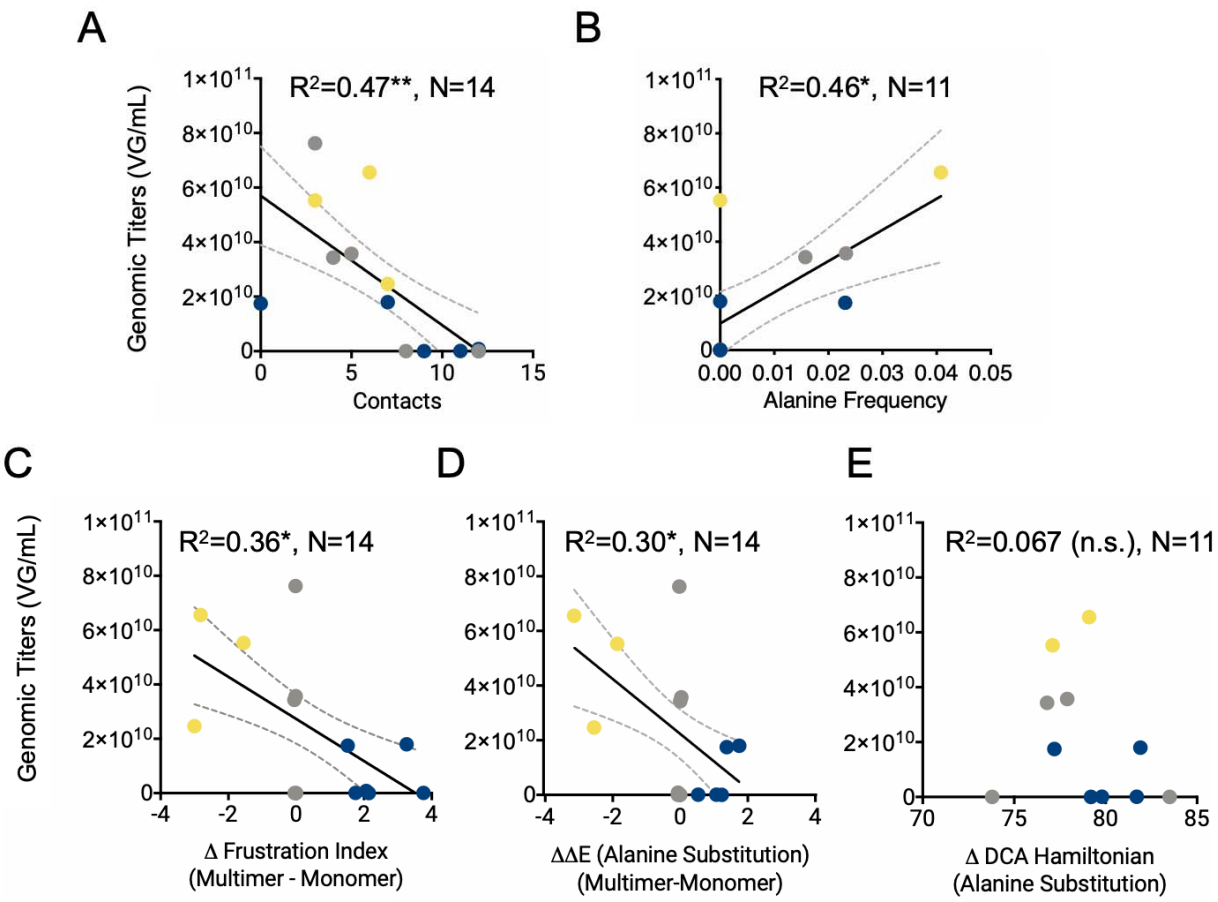
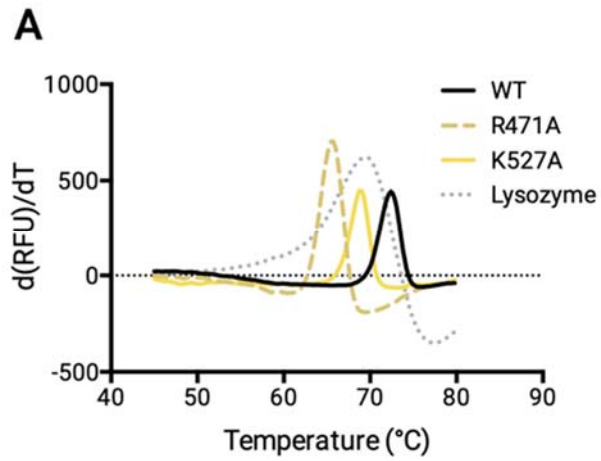


Figure S4: Correlation of AAV2 genomic titers with computational metrics excluding the P657A mutant (related to Fig. 5). Genomic titers are plotted against (**A**) number of residue-residue contacts in the native capsid structure, (**B**) alanine frequency in the parvovirus sequence family, (**C**) Δ-frustration index, (**D**) ΔΔE of alanine substitutions, and (**E**) $\Delta H_{DCA(Ala)}$. N=14 mutants excluding P657A for structure-based analyses. N=11 mutants excluding P657A for sequence-based analyses as R471A, F712A, and S721A were excluded from these analyses due to a high gap frequency at their alignment position. Regression lines with 95% confidence interval are plotted for statistically significant correlations. R^2 values and significance levels are reported (*p<0.05, **p<0.01).



B

Capsid	Melting Temperature (°C)
WT AAV2	72.4±0.2
R471A	65.6±0.2 ***
K527A	68.9±0.1 ***

Figure S5: Melting temperature of alanine mutant capsids R471A and K527A (related to Fig. 6). The melting temperatures of selected capsids with reductions in thermal stability were determined using differential scanning fluorescence. (**A**) Derivative in fluorescence intensity with heating of capsids. Fluorescence increases upon the binding of Sypro Orange to exposed hydrophobic pockets as capsids denature. Traces are the average of N=3 independent experiments. Lysozyme is included as a control, which denatures at 69.4°C±0.2, comparable to the previously reported melting temperature of 68.6°C (1). (**B**) Melting temperature of capsids were determined as the peak of plots shown in panel A. N=3 independent experiments in duplicate, SEM reported. One-way ANOVA was performed with Dunnett's post-hoc multiple comparison test to compare melting temperatures of mutants to WT. ***p<0.001

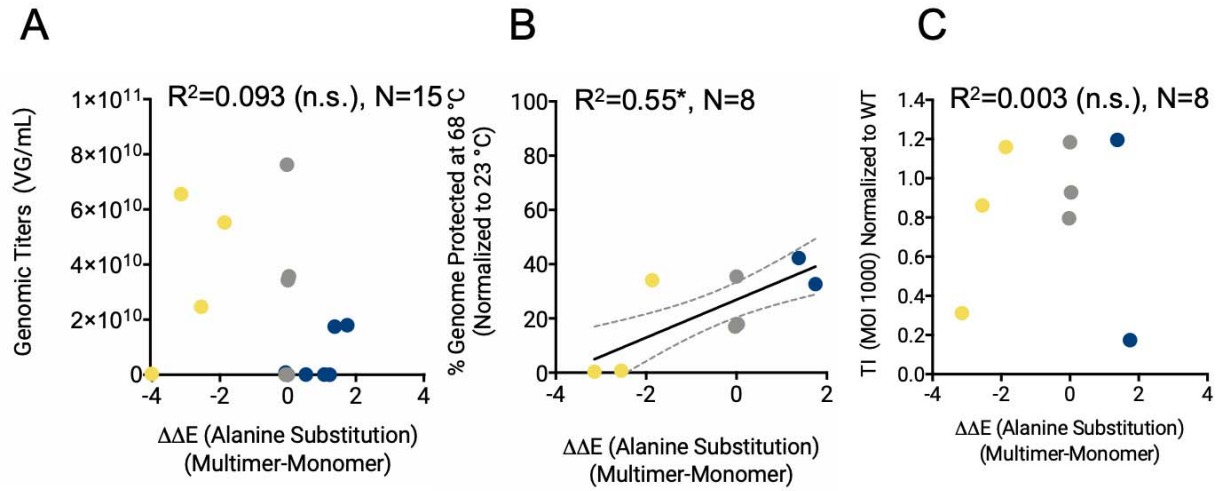


Figure S6 (related to Fig. 5-7): Correlation of ΔΔE of alanine substitutions with experimental measures: ΔΔE of alanine substitutions are plotted against (**A**) genomic titers, (**B**) % genomic protection at 68 °C, and (**C**) transduction index. Regression lines with 95% confidence interval are plotted for statistically significant correlations. R^2 values and significance levels are reported (*p<0.05).

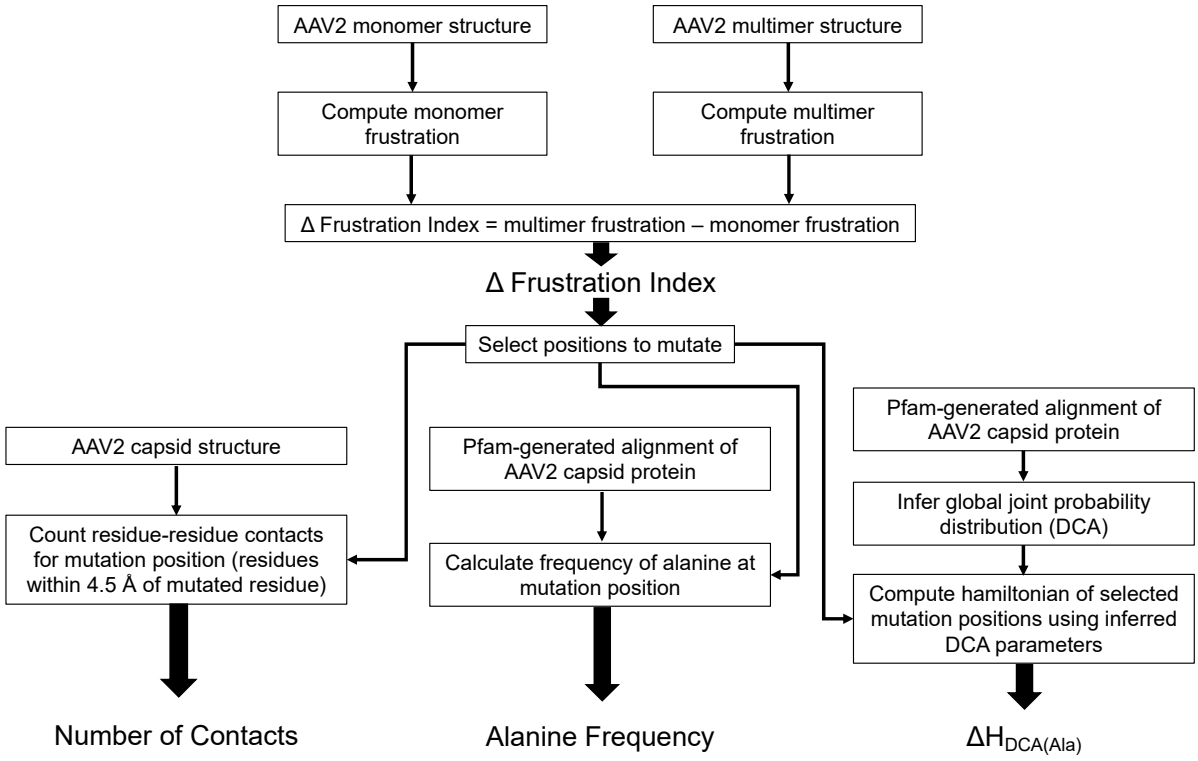


Figure S7: Overview of computational methods and their application in this study.

Supporting Data File 1 (related to methods): Hidden markov model used to find sequences in the parvovirus sequence family (Text file)

Supporting Data File 2 (related to methods): Multiple sequence alignment generated by using HMM to query for sequences in the parvovirus sequence family (Text file)

Supporting Data File 3 (related to methods): Raw qPCR fluorescence curves and plate layouts (Excel files)

Supporting Data File 4 (related to methods): Flow cytometry counts and thresholds (CSV files)

Table S1 (related to Fig. 3): List of top 300 couplings found using DCA and their direct information scores. (Excel file)

Table S2 (related to Fig. 5-7): R-squared values of linear regression between structure-based variables and experimental measures excluding residues that were excluded for sequence-based measures (R471A, F712A, and S721A).

	Titers	**Genomic Protection**	**Transduction**
Structural Contacts (distance < 4.5 Å)	0.28 *(n.s.)*	0.36 *(n.s.)*	0.66*
ΔΔE	0.17 *(n.s.)*	0.49 *(n.s.)*	0.007 *(n.s.)*
ΔF	0.31 *(n.s.)*	0.40 *(n.s.)*	0.017 *(n.s.)*

Table S3 (related to Fig. 5): Mutations introduced in the AAV assembly-activating protein and x-genes in alternate reading frames on the *cap* gene.

***Cap* gene mutation**	**AAP gene mutations**	**X gene mutations**
W228A	I52M, G53R	N/A
R238A	E63Q	N/A
R245A	E70Q	N/A
E347A	S172R	N/A
Y397A	N/A	N/A
Y441A	N/A	N/A
R471A	N/A	N/A
K527A	N/A	N/A
V611A	N/A	C36R
L647A	N/A	F71L, S72P
N656A	N/A	I81L
P657A	N/A	I81M
W694A	N/A	G119R
F712A	N/A	N/A
S721A	N/A	I145M

Supporting References

1. Wright, T.A., J.M. Stewart, R.C. Page, and D. Konkolewicz. 2017. Extraction of Thermodynamic Parameters of Protein Unfolding Using Parallelized Differential Scanning Fluorimetry. *J. Phys. Chem. Lett.* 8:553–558.