Arc Science
Technical Report

RNAv Co-Folding Platform

Abstract. All-atom biomolecular structure predictors construct quadratic residue-pair tensors scaling as $\mathcal{O}(N_{\text{total}}^2)$ in token count and cubically in triangle operations **[published]** [2, 33]. Closed icosahedral and helical viral assemblies pair dozens to hundreds of symmetric protein subunits with one asymmetric nucleic-acid genome that frequently spans thousands of nucleotides **[published]** [13, 45]. The **RNAv Platform** **[architecture]** replaces full-shell quadratic pair tensors with a symmetry-quotient representation $(\Pi_0, E_0, \Omega_0, F_{g,a})$ over point-group (PG), helical (HX), and Caspar–Klug (CK) manifolds while preserving chirality via proper rotations $\det R_i = +1$ in SE(3) and a parity-odd chiral moment $\chi_{\det}$. Polymer reach $(\ell_c, \ell_p, \ell_K, \hat{\ell})$, soft-core Debye–Hückel electrostatics (ℓ_B, κ_D, I_s), Manning counterion condensation (f_{eff}), cancellation-free Citardauq/asinh Donnan ion partitioning $(c_\Sigma, c_\pm)$, log-domain Sinkhorn proposals (η_{OT}), and an exact dynamic programming oracle (`subset_dp`) supervise two networks inside a cut-posterior M0–M7 ladder. They are an autoregressive placer (`LearnedPlacerNet`; 38,209 parameters) and a co-folding backbone (`RNAvFoldingModel`: 971,323 active zero-shot default parameters; 1,530,941 total with +559,618 optional frozen-encoder adapters). On a local NVIDIA GeForce RTX 4060 Laptop GPU (8,187.5 MB VRAM, PyTorch 2.11.0+cu128, recorded 14.XI.2026), a 50-step BF16 profile run on synthetic input at $L = 128$ takes 300.18 ms/step and 97.34 MB peak VRAM, and total loss falls by 4.20% ($-4.2871 \to -4.4672$) **[measured]**; the run is a profile check, not a training result. On the same GPU, 50-step `LearnedPlacerNet` distillation on $(M = 4, K = 12)$ raises 512-sample importance-sampling effective sample size from 136.73 (26.7%) to 487.71 (95.3%) against exact $\ln Z = -13.1053$ **[measured]**. Beyond implemented mathematics and benchmarks, this report evaluates structural novelty, projects quantitative bounds at the end of the 58-day execution plan, sets PDB archive ceilings (96 searchable global-I Protein/NA assemblies), and defines pre-registered falsifiers. It also states, for four domains, where RNAv helps engineering practice: rigid synthetic nucleocapsids, engineered RNA VLPs, asymmetric cryo-EM reconstructions, and bulk compartment observables, each with its validity bounds.

Contents

1 Platform Architecture and Motivation **2**

2 Empirical Archive and Source Appraisal **3**

3 Symbol Dictionary and Cut Posterior **6**

4 Symmetry Quotient and Chiral Frames **7**

5 Donnan Electrostatics and Compartment Neutrality **9**

6 Packaging-Signal Placement and Pocket Transport **11**

7 Verification Suite and Hardware Benchmarks **13**

8 Architectural Novelty and Prior Foundations **16**

9 Archive Ceilings and Falsification Criteria **16**

10 Engineering Domains and Application Scope **17**

10.1 Boundary Equations and Cut Guarantees 20

11 Execution Schedule and Resource Allocation **21**

12 Whole Platform Architecture **22**

12.1 System Topology and Schemas 23

12.2 Symmetry and Physics Kernels 23

12.3 Neural Trunk and Channels 25

12.4 Platform Algorithmic Specifications 26

12.5 Parameter and Codebase Audit 29

13 Evidence Boundaries and Reproducibility Ledger **32**

14 Conclusions **32**

1. Platform Architecture and Motivation

Full-virion modeling hits two coupled structural walls. First, dense residue-pair representations become intractable once a multi-hundred-subunit shell surrounds a kilobase genome. Bacteriophage MS2 occupies a 180-site $T = 3$ lattice with 178 coat-protein chains ($178 \times 129 = 22{,}962$ residues) plus one 393-residue maturation protein substituting one coat dimer around a 3,569-nucleotide single-stranded RNA genome. This pushes native token counts above 2.6×10^4 before solvent or metal ions enter [published] [13, 45]. Single-GPU attention implementations overflow well below that threshold (baseline EvoAttention peaks at 24.61 GB at $L = 192$ and overflows a single GPU at 590 residues), and distributed sharding frameworks such as Fold-CP require 64 NVIDIA B300 GPUs to handle roughly 30,000 residues at $\mathcal{O}(N^2/P)$ memory per device [published] [33, 36]. Second, naive symmetry enforcement erases the biological asymmetry of the genome inside the lumen, whereas pure $E(3)$-equivariant coordinate networks collapse mirror enantiomers between right-handed $T = 7d$ $(2, 1)$ and left-handed $T = 7l$ $(1, 2)$ Caspar–Klug shells and risk inverting L-amino acid or D-ribose stereocenters [architecture, published] [47, 56].

RNAv resolves both constraints through a dual-track modular pipeline (`rnav`): protein coat sequence features compatible with ESM-2 [published] [35] pass through a symmetry-quotient projector centered on one asymmetric unit (ASU), while genomic RNA transcripts enter through non-overlapping byte-pair encoding (BPE) paired with RNA-FM and Nucleotide Transformer adapters [published] [11, 14]. Physics biases derived from analytical polymer reach and Donnan-screened electrostatics modulate cross-attention logits in FP32 while neural trunk linear

projections execute in mixed BF16 precision. Discrete genomic packaging signals (`M` ordered stem-loops anchored across `K` symmetric capsid pockets) are resolved by a unary log-domain Sinkhorn transport proposal refined through an exact log-domain `subset_dp` oracle and distilled into `LearnedPlacerNet` with quantitative importance-sampling diagnostics [measured]. Figure 6 and §12 trace the complete deep-learning forward stages, stage-by-stage tensor outputs, trainable versus frozen parameter splits, and codebase inconsistency audit.

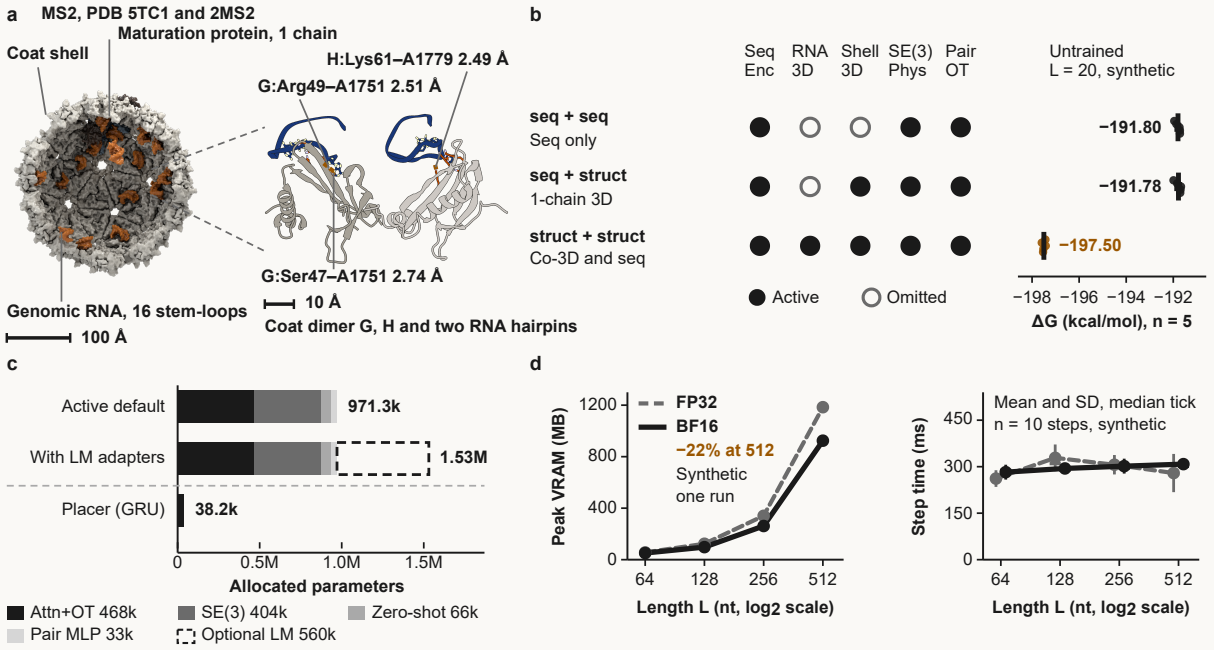

Figure 1: The RNAv co-folding system on the MS2 virion. Panel a shows the MS2 virion cut open (PDB `5TC1` and `2MS2`): grey coat shell, dark maturation protein (1 chain) and orange RNA, with 16 RNA stem-loops traced in the density; scale bar 100 Å. The inset shows the coat dimer G, H with two RNA hairpins, scale bar 10 Å; orange dashed lines join the atoms of three measured contacts: G:Arg49 to A1751, 2.51 Å; G:Ser47 to A1751, 2.74 Å; H:Lys61 to A1779, 2.49 Å. Panel b lists the three prompt modes against the five modules that each mode uses (filled: active, open: omitted): sequence encoder, RNA 3D input, shell 3D input, SE(3) trunk with the physics biases, and the pair and optimal-transport (OT) heads. It gives the free energy of the model before training at length 20 on synthetic input (kcal/mol, 5 seeds; tick: mean, bar: 95% interval stored with the data): −191.80, −191.78 and −197.50. Panel c splits the parameters of `RNAvFoldingModel`: 971.3k in the active default (attention with OT, SE(3), zero-shot encoders without adapters, pair MLP) and 1.53M with the optional language-model adapters (dashed), plus 38.2k in the placer `LearnedPlacerNet`, a GRU. Panel d gives peak VRAM (one run per point) and step time (mean and SD of $n = 10$ steps, tick: median, axis from zero) of FP32 and BF16 on the RTX 4060 for $L \in \{64, 128, 256, 512\}$; BF16 uses 22% less memory at 512 (synthetic input). Orange marks the value that a panel is about. [architecture, measured]

2. Empirical Archive and Source Appraisal

Designing or evaluating any co-folding predictor for whole nucleoprotein particles requires checking what crystallographic and cryo-EM training archives contain. An RCSB Protein Data Bank audit frozen on 26.IX.2026 (recorded in the project technical audit and reconciled against live curation records) counts 260,383 deposited structures total, 17,030 Protein/Nucleic-Acid complexes,

2,277 viral Protein/NA entries, and only 96 searchable global icosahedral (`I`) Protein/NA assemblies [published] [7, 45]. Among those 96 entries, the largest unique deposited polynucleotide model reaches only 199 unique nucleotides (PDB `8ER8`). In the flagship MS2 asymmetric cryo-EM reconstruction (PDB `5TC1` paired with EMDB `EMD-8397`), roughly 80% of the genomic RNA backbone across 16 stem-loops was traced into medium-resolution density, yet only 242 of the 3,569 native nucleotides (6.8%) appear as atomic coordinates in the deposited PDB file [published] [13, 20]. Across curated external databases, VIPERdb catalogs 1,799 icosahedral capsids [published] [45], RMDB stores 1,024 chemical-probing entries (520,709,190 nucleotide reactivity points), EMDB exceeds 46,900 maps, and EMPIAR holds 3,117 raw tilt/particle archives.

To ground every architectural decision, Table 1 scores fourteen primary literature sources from 0 to 3 across publication recency (**Date**), direct relevance to capsid–genome co-folding (**Scope Fit**), and experimental or theoretical rigor (**Validity**). Every score reflects direct inspection of published abstracts, benchmarks, or code-verified metrics in this repository.

Table 1: Source appraisal matrix (0–3 rubric on Date, Scope Fit, and Evidence Validity). Scores: 0 absent/legacy baseline, 1 partial/narrow, 2 strong supporting, 3 primary/direct benchmark. [published]

Source	ID	Date	Scope	Valid.	Key Measured Finding and RNAv Architectural Constraint
Abramson et al.	[2]	3	2	3	AlphaFold 3 unifies protein–NA co-folding ($\leq$ 5,120 tokens stated limit); quadratic/cubic scaling blocks 10^5-token virions; scores 14/60 (23.3%) on two-state conformers.
Baek et al.	[4]	2	2	3	RoseTTAFoldNA predicts small protein–DNA/RNA dimers cleanly yet documents clash and memory limits above $\sim$ 1,000 residues/nucleotides.
Kagaya et al.	[28]	3	2	3	CASP16 multimers and RNA evaluation: symmetry inference fails for most large assemblies; nucleic-acid predictions remain strongly correlated with template availability.
Xu et al. (FoldBench)	[59]	3	2	3	Evaluates 70 protein–RNA interfaces, 330 protein–DNA interfaces, 15 RNA monomers, and 0 viral whole particles; heteromeric DockQ exceeds homomeric DockQ.
Rojas Labra et al. (FoldaVirus)	[46]	3	3	3	Predicts $T \leq 9$ empty icosahedral coat shells via AF template superposition and Amber minimization; trims disordered RNA-binding arms and omits genomic RNA entirely.

Source	ID	Date	Scope	Valid.	Key Measured Finding and RNAv Architectural Constraint
Lin et al. (Fold-CP)	[36]	3	2	2	Distributed pair sharding cuts memory to $\mathcal{O}(N^2/P)$ and reaches $\sim$ 30,000 residues on 64 B300 GPUs; exact sharding complements the quotient trunks.
La et al. (MegaFold)	[33]	3	2	2	EvoFlash-3D proves attention activations scale super-quadratically (3.75 $\rightarrow$ 24.61 GB from $L = 96$ to 192, 6.56$\times$) and yields 4.24$\times$ memory savings via fused kernels.
Ouyang-Zhang et al. (Pairmixer)	[40]	3	2	2	Removing triangle attention keeps triangle multiplication intact, achieving 4$\times$ speedup at 2,048 tokens (1,000 s $\rightarrow$ 250 s) with RNA lDDT 0.59 vs 0.58.
Asor et al. (`mi3`)	[3]	3	3	3	Single-particle mass photometry measures `mi3` dodecahedron trimer self-assembly (15% at 58 nM, 83% at 350 nM), calibrating dimensionless $\kappa = -2.15$ ($k_{\mathrm{B}}T$ units) in RNAv.
Dai et al. (MS2)	[13]	2	3	3	Asymmetric cryo-EM structure (`5TC1`) resolves maturation subunit breaking icosahedral symmetry and 16 genomic stem-loops (242/3,569 atomic nt deposited).
Farafonov et al.	[21]	3	3	3	265 ns unrestrained all-atom MS2 virion simulation fixes genome-mediated pore flux cuts (H_2O -20%, Cl^- -50%) and radial two-shell RNA condensation.
Tarasova et al. (PCV2)	[52, 53]	2	3	3	Half-chloride PCV2 simulations crack 5-fold vertices at RMSD 0.30 nm $<$ 0.35 nm; constant-pH MD (1 μs replicas) isolates the Asp172 pH switch at pH 5.0.
Butterfield; Terasaka	[8, 54]	2	3	3	De novo designed nanocages and evolved nucleocapsids selectively encapsulate bespoke mRNAs ($<$ 4 kb) via interior positive charges and hairpins.
Segel; Banskota	[5, 48]	2	3	3	Retroviral/endogenous VLPs (PEG10 SEND, engineered eVLPs) validate programmable nucleic-acid and CRISPR RNP delivery inside engineered symmetric shells.

3. Symbol Dictionary and Cut Posterior

Table 2 aligns every mathematical identifier one-to-one across the Python package (`rnav`), the Rust geometry workspace (`10_App`), the thesis manuscript, and this technical report [architecture]. Caspar–Klug hexagonal coordinates reserve (h, k) exclusively, assigning $(u_{\mathrm{OT}}, v_{\mathrm{OT}})$ to dual Sinkhorn scaling potentials.

Table 2: Master symbol dictionary unifying Python, Rust, and mathematical notation ($k_{\mathrm{B}}T$ energy unit). [architecture]

Symbol	Space / Units	Definition and Implementation Module
X	$\mathbb{R}^{N\times 3}$	Cartesian atomic/bead coordinates across coat subunits and nucleic acid
R_i, (Q_i, t_i)	SO(3), SE(3)	Proper rotation ($\det R_i = +1$) and rigid frame; improper $\mathrm{E}(3) \setminus \mathrm{SE}(3)$ reflections excluded
Π_0, E_0, Ω_0, $F_{g,a}$	Quotient Ops	Orbit projection, equivariant expansion, marginal pseudo-precision, quotient features
PG, HX, CK	Symmetry Class	Point-group ($\mathrm{I}, \mathrm{O}, \mathrm{T}, \mathrm{D}_n, \mathrm{C}_n$), helical screw, and Caspar–Klug classes
T_{CK}, n_0, (h, k)	$\mathbb{Z}_{\geq 1}$, $\mathbb{Z}_{\geq 1} \times \mathbb{Z}_{\geq 0}$	Triangulation $T_{\mathrm{CK}} = h^2 + hk + k^2$, ASU chain count n_0 ($N = \lvert G\rvert n_0$), Eisenstein steps (h, k)
M, K, N_m	$\mathbb{Z}_{\geq 1}$	Ordered packaging-signal anchors (M), pocket sites (K), motif count (N_m)
Δz, $\Delta\phi$, N_{turn}, P_{hx}	Helical Params	Rise (Å), twist (deg), subunits/turn $2\pi/\lvert\Delta\phi\rvert$, pitch $N_{\mathrm{turn}}\Delta z$
L, ℓ_c, ℓ_p, ℓ_K, $\hat{\ell}$	Length, Dimless	Token count, contour, persistence, Kuhn length $2\ell_p$, $\hat{\ell} = \ell_c/\ell_p$
ℓ_B, κ_D, I_s	Length, M	Bjerrum length, inverse Debye screening parameter (λ_D^{-1}), ionic strength
$c_\pm$, c_s, c_Σ, ρ_f, f_{eff}	M, Dimless	Counterion/salt concentrations, $c_\Sigma = \sqrt{\rho_f^2 + 4c_s^2}$, Donnan charge, Manning uncondensed fraction
θ_i, γ_i, χ_i	State Vars	Residue protonation fraction, orientational coupling, confinement coordinate
C_{ij}, η_{OT}, u_{OT}, v_{OT}	$k_{\mathrm{B}}T$, $\mathbb{R}_{>0}$, $\mathbb{R}$	Sinkhorn cost, entropic temperature, primal/dual scaling factors
U_{tv}, $W_{t-1,v',v}$, δ_{clip}	$k_{\mathrm{B}}T$, Length	Unary pocket energy, linker log-weight, soft-wall lumen radius margin
$\alpha_t(v, S)$, $\beta_t(v, S)$	$\mathbb{R}$ (log-domain)	Exact forward/backward `subset_dp` messages for anchor t and occupied set S

Symbol	Space / Units	Definition and Implementation Module
t_{asm}, ESS	s, Count	Characteristic assembly timescale and SNIS Kish effective sample size

RNAv enforces a **modular cut-posterior architecture** across rungs M0–M7 **[architecture, published]** [37, 44] (Table 3), so that poorly constrained downstream free energies do not feed back into upstream geometry or thermodynamics, as in a cut posterior. Cutting gradients at designated statistical boundaries prevents non-identifiable compensation between protein interface affinities, oligomerization cooperativity κ, genome WLC flexibility, and explicit solvent force fields.

Table 3: RNAv M0–M7 hierarchical cut-posterior ladder. Cuts isolate well-identified structural and thermodynamic observables from noisy downstream modules. **[architecture, measured]**

Rung	Submodel	Cut Conditioning Input	Observable, Exact Check, or Gate
M0	Symmetry Classifier	Sequence HMM + family template	Point/helical/T_{CK} orbit generator, explicit abstain on novel folds
M1	Shell Quotient Trunk	Coat sequence + ASU geometry	Equivariant SE(3) frames ($\det Q = +1$), VIPERdb deviation $\leq 7.4 \times 10^{-6}$
M2	Oligomer Lattice Gas	Free-trimer mass balance	Grand-canonical cooperativity $\kappa = -2.15$ ($k_{\mathrm{B}}T$ units; `mi3`: $\chi^2_{\mathrm{fit}} = 0.052$)
M3	Electrostatic / pH Gate	Coat titration + Donnan ρ_f, c_s	PCV2 Asp172 pH 5.0 peak (0.785), $c_\pm$ ionic compartment neutrality
M4	Discrete PS Placer	WLC linkers + Sinkhorn unary prior	Exact `subset_dp` ($\ln Z$ to 10^{-14}) and `LearnedPlacerNet` (ESS/$N \geq 0.90$)
M5	All-Atom Stability Gate	Full assembled Virion + Cl^-/Na^+	4-criterion compartment neutrality & R_g/symmetry screen before GROMACS
M6	CG Residual Trajectory	Cut M1–M4 + probing restraints	Asymmetric interior density peaks (80–120 Å) and map-model FSC
M7	Maximum-Entropy Ensemble	Fixed forward model + weights	Bayesian/BME reweighting bounded by Kish ESS diagnostics

4. Symmetry Quotient and Chiral Frames

Icosahedral capsids obey the finite proper rotation group $G = \mathrm{I}$ ($|\mathrm{I}| = 60$ rotations matching VIPERdb conventions to 7.4×10^{-6} in `rnav`) **[measured]** [45], distinct from the chiral octahedral group O ($|\mathrm{O}| = 24$) and tetrahedral group T ($|\mathrm{T}| = 12$). To resolve symbol overlap between point groups and Caspar–Klug geometry (**M0**), $T_{\mathrm{CK}} = h^2 + hk + k^2$ exclusively indexes Caspar–Klug triangulation numbers over hexagonal sheet integers $(h, k) \in \mathbb{Z}_{\geq 1} \times \mathbb{Z}_{\geq 0}$, while n_0 denotes conformers (chains) per asymmetric unit (ASU) so total shell subunits equal $N = |G|\, n_0$. Standard Caspar–Klug shells satisfy $n_0 = T_{\mathrm{CK}}$ and distinguish non-superimposable enantiomorphs by ordered indices (h, k) vs. (k, h)—such as $(2, 1)$ ($T = 7d$) vs. $(1, 2)$ ($T = 7l$)—whereas generalized non-Caspar–Klug

tilings such as PRD1 ($P = 25$, $n_0 = 15$ chains per ASU, $N = 60 \times 15 = 900$ chains, 259.5–340.7 Å) carry $n_0 = 15$, an integer outside $\{h^2 + hk + k^2\}$ **[published]** [56]. Helical filaments such as tobacco mosaic virus (TMV, PDB `6SAG`: $\Delta z = 1.405$ Å, $\Delta\phi = 22.036°$) generate an infinite cyclic screw manifold ($|G| = \infty$). Its screw parameters are:

$$T_{\mathrm{CK}} = h^2 + hk + k^2, \qquad N_{\mathrm{turn}} = \frac{2\pi}{|\Delta\phi|} = 16.337, \qquad P_{\mathrm{hx}} = N_{\mathrm{turn}}\,\Delta z = 22.95 \text{ Å}, \tag{1}$$

The turn number is closely approximated by the rational 3-turn repeat $49/3 = 16.333$ ($49 \times 22.036° = 1079.76°$) **[published]** [57].

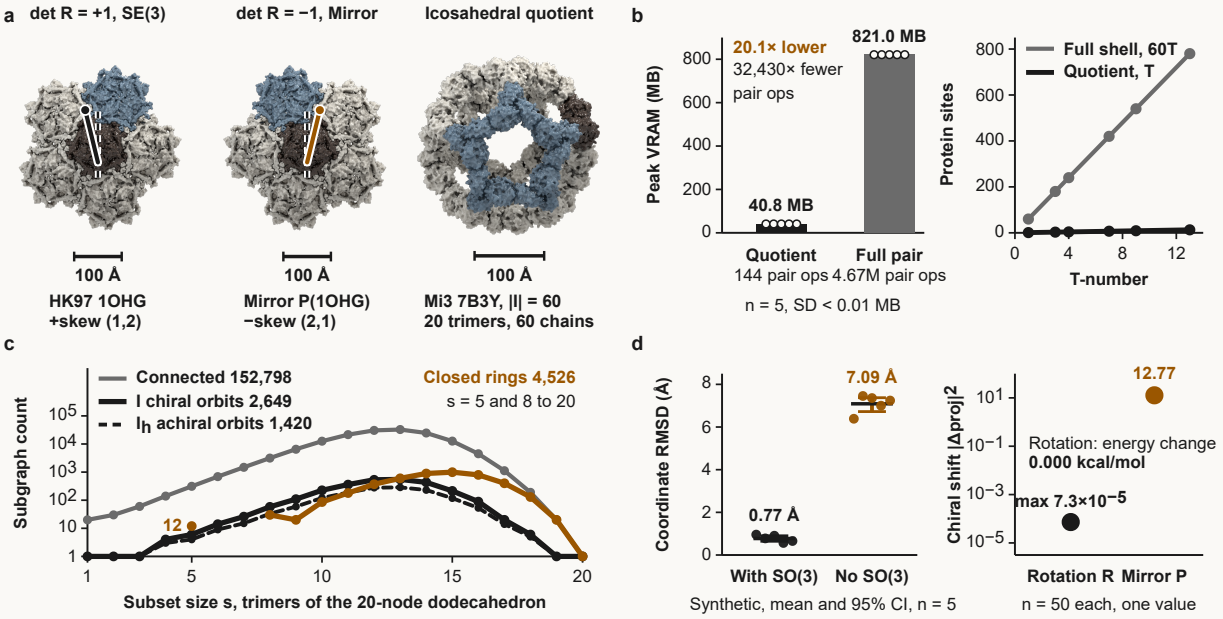

Figure 2: Chiral symmetry quotient, memory saving, `Mi3` subassembly counts and SO(3) projection. Panel a shows the natural `HK97` shell (PDB `1OHG` [26], determinant +1, lattice skew (1,2)), its mirror image (determinant −1, skew (2,1)) and the icosahedral `Mi3` shell (PDB `7B3Y`, 20 trimers, 60 chains, group order 60), each with a 100 Å scale bar; the line over each render is the capsid axis turned by the skew, the white dashes are the lattice axis. Panel b compares peak VRAM of the model with the quotient graph (144 pair operations) and without it (4.67M pair operations) over 5 seeds: 40.8 MB against 821.0 MB, 20.1 times lower, with 32,430 times fewer pair operations; SD below 0.01 MB. The right plot counts protein sites against the Caspar–Klug T-number, a different quantity: $60T$ for the full shell, T for the quotient. Panel c counts the subassemblies of the `Mi3` cage by subset size: 152,798 connected subgraphs, 2,649 chiral orbits under the rotation group I, 1,420 achiral orbits under the group Ih with reflections, and 4,526 closed rings, of which 12 have size 5. Panel d gives the coordinate RMSD with and without the SO(3) projection (0.77 Å and 7.09 Å, mean and 95% interval over 5 seeds, synthetic) and, on a log axis, the chiral shift of the projection: at most 7.3×10^{-5} for 50 rotations and 12.77 for 50 mirror reflections (one value per group); the energy change under rotation is 0.000 kcal/mol. **[architecture, measured, published]**

In `rnav/symmetry/quotient.py` (**M1**), the capsid contact graph collapses onto one ASU of n_0 representative nodes linked by intra-ASU edges (u, v) and symmetry-mate edges (u, v, g) for $g \in G$. Let Π_0 project X to ASU coordinates X_0, $E_0(X_0; G)$ expand X_0 equivariantly over $(Q_g, t_g) \in \mathrm{SE}(3)$, Ω_0 maintain marginal pseudo-precision across quotient neighbors, and $F_{g,a}$ encode orbit-indexed geometric features. Group-invariant tensor contractions apply the idempotent Reynolds projector

($\mathcal{R}^2 = \mathcal{R}$). The expansion and the projection read:

$$X = E_0(X_0; G) = \bigoplus_{g \in G} \big(X_0 Q_g^\top + \mathbf{1}\, t_g^\top\big), \qquad (\mathcal{R}f)(x) = \frac{1}{|G|} \sum_{g \in G} g\, f(g^{-1}x). \tag{2}$$

Under a free G-action on general-position ASU nodes ($N = |G|\, n_0$), the capsid hosts $N^2/|G| = |G|\, n_0^2$ ordered-pair orbits, making lossless pair compression equal to $|G|$ (60× for icosahedral I: collapsing 32,400 MS2 subunit pairs into 540 ordered-pair orbits and 810,000 PRD1 pairs into 13,500 orbits). Restricting pair attention strictly to the $n_0 \times n_0$ intra-ASU block (9 entries for MS2; 225 for PRD1) achieves $|G|^2$ footprint shrinkage (3,600× for $G = \mathrm{I}$; 576× for O; 144× for T) while omitting inter-ASU mate contacts, so rigid multi-ASU interface operators explicitly preserve (u, v, g) mate indices **[architecture]**. Because reflections destroy biological chirality, every frame $T_i = (R_i, t_i) \in \mathrm{SE}(3)$ enforces $\det R_i = +1$. Updates evolve translations in the local frame and exponentiate angular increments $\omega_i^{(\ell)}$ through Rodrigues' formula (switching to a small-angle Taylor series below $\|\omega\| < 10^{-3}$). The updates also apply a 4-step Newton–Schulz polar projection with signed SVD fallback and expose a parity-odd multiscale scalar invariant $\chi_{\det} = \operatorname{asinh}(\det[m_1, m_2, m_3]/10^3)$. The update rules are:

$$t_i^{(\ell+1)} = t_i^{(\ell)} + R_i^{(\ell)}\, \Delta x_i^{(\ell)}, \qquad R_i^{(\ell+1)} = R_i^{(\ell)}\, \mathrm{Exp}\big(\omega_i^{(\ell)}\big) \in \mathrm{SO}(3), \qquad \chi_{\det}(-X) = -\chi_{\det}(X). \tag{3}$$

For exact combinatorial assembly benchmarking on the 20-trimer dodecahedral nanocage `mi3`, `assembly_graph.py` implements Burnside orbit enumeration checked against an independent brute-force program written by GPT-6.1 Sol **[measured]**. Both implementations match identically on all 152,798 connected partial trimer subsets, 2,649 chiral rotation orbits under I, 1,420 achiral orbits when mirrors are identified, 4,526 closed cage subgraphs, and the discrete topological ring-closure sequence (12 five-membered pentagon rings, plus 3 dimeric pentagons up to 30, then +2 and +1 final closures) **[measured]**.

5. Donnan Electrostatics and Compartment Neutrality

Genome encapsidation reflects a thermodynamic balance across interior basic coat-protein tails, condensed divalent/monovalent counterions, semi-flexible RNA conformation, and bulk salt screening **[published]** [6, 15, 43, 49, 55]. Experimentally, cowpea chlorotic mottle virus (CCMV) coat dimers package RNAs from 140 to 12,000 nt at an Author-Corrected critical protein-to-RNA mass ratio of 4.3 **[published]** [9, 10]. BMV RNA1 (3,200 nt) reconstitution across three arginine-rich-motif (ARM) charge mutants shows nearly constant total neutralized tail charge (ARM charge × CP_2:RNA ratio ≈ 2,800–3,200) **[published]** [22]. In contrast, global genome-to-capsid charge slopes vary from 0.3 to 1.69 depending on viral subsets and salt models **[published]** [6, 55]. Mean-field Debye–Hückel electrostatics supplies a quantitative weak-coupling bound once Manning counterion condensation and soft-core clamping prevent short-range singularities in $k_{\mathrm{B}}T$ units. Manning condensation ($\xi = \ell_B |z_c z_p|/b$) yields the general valence law $f_{\mathrm{eff}} = \min(1, \xi^{-1}) \approx 0.49$ for ssRNA spacing $b = 3.5$ Å at $|z_c| = 1$, and $0.24, 0.12, 0.079$ at $z_c = 1, 2, 3$ for $b = 1.7$ Å, omitting redundant extra factors of z_c. Soft-core clamping ($\kappa_D r_{\mathrm{soft}} \ll 1$, where $\lambda_D = \kappa_D^{-1} = 7.85$ Å at 150 mM) keeps the potential finite at short range. The Bjerrum length, the screening parameter, and the potential

are:

$$\ell_B = \frac{e^2}{4\pi\varepsilon_0\varepsilon_r k_{\mathrm{B}}T} \approx 7.15\ \text{Å}, \qquad \kappa_D = \sqrt{8\pi\ell_B\,(10^3 I_s)\,N_A}, \qquad \frac{U_{\mathrm{DH}}(r)}{k_{\mathrm{B}}T} = \ell_B z_1 z_2 \frac{e^{-\kappa_D r}}{\sqrt{r^2 + r_{\mathrm{soft}}^2}}. \quad (4)$$

Inside a semi-permeable shell whose net interior mobile-cation excess is $\rho_f = c_+ - c_- = -\rho_{\mathrm{fixed}}$ against bulk 1:1 salt c_s, small-ion chemical potentials obey $c_+c_- = c_s^2$ and yield dimensionless Donnan potential $y = e(\psi_{\mathrm{bulk}} - \psi_{\mathrm{in}})/k_{\mathrm{B}}T$. Standard quadratic formulas lose mantissa digits via catastrophic subtraction when $|\rho_f|/c_s \ll 1$ and produce 0/0 `NaN` backward gradients in autodiff `torch.where` branches. `RNAv` instead computes Citardauq's formula in NumPy and a smooth hyperbolic-sine parameterization in PyTorch. The resulting quantities are:

$$c_\Sigma = \sqrt{\rho_f^2 + 4c_s^2}, \qquad \Delta c_{\mathrm{osm}} = \frac{\rho_f^2}{c_\Sigma + 2c_s}, \qquad c_\pm = c_s \exp\!\left(\pm\,\mathrm{asinh}\!\left(\frac{\rho_f}{2c_s}\right)\right). \quad (5)$$

For ionizable amino-acid side chains inside the lumen, charge regulation couples y self-consistently to internal acidity $\mathrm{pH_{in}} = \mathrm{pH_{bulk}} - y/\ln 10$ via bracketed root-finding on $F(y) = \sinh y - \rho_f(y)/(2c_s) = 0$. Constant-$\rho_f$ Citardauq agrees only when regulation sensitivity $\mathcal{R} = B/(2c_s\cosh y^*) \ll 1$ (with $B = -d\rho_f/dy$). At $n_A = 0.4$ M, $c_s = 0.1$ M, $\mathrm{pH_{bulk}} = \mathrm{p}K_a$ uncoupled Citardauq overstates Δc_{osm} by 1.93× (0.0828 M vs. self-consistent 0.0429 M). Spatially, Jensen's inequality guarantees $\langle c_+ + c_-\rangle \geq \sqrt{4c_s^2 + \bar{\rho}_f^2}$, so homogeneous compartment Donnan balance is exact only inside the window $\ell_{\mathrm{inh}} \ll \lambda_D \ll R_{\mathrm{in}}$ and must switch to nonlinear shell Poisson–Boltzmann (boundary thickness $\min(\lambda_D, \mu_{\mathrm{GC}})$ with $\mu_{\mathrm{GC}} = 1/(2\pi|z|\ell_B\sigma_s)$) when charges cluster on the capsid wall. For single-stranded RNA linkers bridging consecutive packaging anchors over persistence length ℓ_p ($\ell_K = 2\ell_p$, $\hat{\ell} = \ell_c/\ell_p$), `rnav/physics/wlc.py` evaluates the Kratky–Porod end-to-end variance $\langle R^2\rangle$ **[published]** [32]. It uses an 8th-order Taylor switch whenever $\hat{\ell} < 0.05$ in FP64 (relative error $< 10^{-14}$) or $\hat{\ell} < 0.5$ in FP32 (1.1×10^{-10} at the crossover), which eliminates catastrophic `expm1` subtraction. The variance and the linker log-weight are:

$$\langle R^2\rangle = 2\ell_p\ell_c\left[1 - \frac{\ell_p}{\ell_c}\left(1 - e^{-\ell_c/\ell_p}\right)\right], \qquad W(r;\ell_c) = -\frac{3r^2}{2\langle R^2\rangle} - \frac{3}{2}\ln\langle R^2\rangle \quad (r \leq \ell_c). \quad (6)$$

On thermodynamics and pH gating, `rnav/phase_boundary.py` weights each connected n-trimer subgraph carrying E contacts by $(3^n/60)\exp(gE - \kappa(n-1))$ relative to $c^\circ = 1$ M ($G_0[1] = 0$). Holding the dimensionless net dimer contact affinity $a = g - \kappa = -\ln(4.5K_D(2)/c^\circ) = 11.40$ (in $k_{\mathrm{B}}T$ units) fixed at $K_D(2) = 2.5$ μM from `mi3` single-particle mass photometry **[published]** [3] makes $g(\kappa) = a + \kappa$. On the closed 20-trimer dodecahedron ($E = 30$, Betti cycle rank $\beta_1 = 30-20+1 = 11$), the closed-shell exponent becomes $30g - 19\kappa = 30a + 11\kappa$. The closed-shell critical concentration follows:

$$c^*(\kappa) = c^\circ \exp\!\left[-\frac{30a + 11\kappa + \ln(3^{20}/60)}{19}\right]. \quad (7)$$

Thus cooperativity enters $c^*(\kappa)$ with cycle-rank weight 11/19: fitting `calibration_check.py` recovers dimensionless continuous optimum $\kappa = -2.155$ ($c^* = 20.85$ nM trimers vs. non-cooperative $c^*(\kappa = 0) = 5.99$ nM, ratio 3.48×; rounded $\kappa = -2.15$ gives $g = 9.25$, $c^* = 20.80$ nM, and $e^{11\times 2.15/19} = 3.47\times$). The fit also reproduces measured `mi3` VLP mass fractions under full-species `equilibrium(m/3)` including dimers (14.8% at 58 nM [19.3 nM trimers], 83.1% at 350 nM [116.7 nM trimers], $\chi^2_{\mathrm{fit}} = 0.052$, whereas strict two-state $\kappa = 0$ overpredicts 72.2% and 94.9%) **[measured]**. At the strict two-state 50% midpoint, free trimer sits at $c_1 = 20^{-1/19}c^* = 0.854\,c^*$ and total trimer at

$c_{\mathrm{trim},50} = 2c_1 = 1.708\,c^*$. Converting to monomer concentration ($m = 3c_{\mathrm{trim}}$) places free monomer at $3c_1 = 53.4$ nM and total monomer midpoint at $m_{50} = 6c_1 = 106.9$ nM for $c^* = 20.85$ nM (106.6 nM at rounded $c^* = 20.80$ nM; 108.5 nM in the full lattice gas with dimers). This resolves why 58 nM total monomer yields 14.8% assembly rather than 50%. For pH-responsive capsids, evaluating Wyman's linkage at $\mathrm{pH_{in}}$ across shifted Asp172 constant-pH protonation ($\mathrm{p}K_a^C = 5.8$, $\mathrm{p}K_a^I = 4.2$) against interface carboxylates ($\mathrm{p}K_a = 4.0$) yields a sharp optimum of 0.785 at pH = 5.0 (window $[4.492, 5.552]$), matching negative-stain EM observations for porcine circovirus 2 (PCV2) **[measured, published]** [29, 53, 58].

Whole-capsid MD benchmarks reveal why naive structural stability thresholds fail. Published all-atom simulations show that MS2 (`5TC1`, $+192\,e$ net capsid charge, interior integrated charge peak $+644\,e$ at $r = 12.1$ nm) requires 644 $\mathrm{Cl^-}$ inside and 452 $\mathrm{Na^+}$ outside prior to isotropic 0.9% NaCl addition. Native packaged RNA cuts transmembrane water flux by 20% (5,040 → 4,140 molecules/ns) and chloride flux by half (6.4 → 3.1 ions/ns) **[published]** [19, 21]. Conversely, in paired 10 ns AMBER03 runs of PCV2, depleting internal chloride by half (300 $\mathrm{Cl^-}$ tail-free, 200 $\mathrm{Cl^-}$ with tails vs neutral 606/407) cracks five-fold vertices while the tailed low-chloride trajectory still stays near RMSD ≈ 0.30 nm < 0.35 nm **[published]** [52]. Within the homogeneous 1:1 Donnan validity window $\ell_{\mathrm{inh}} \ll \lambda_D \ll R_{\mathrm{in}}$ (using accessible volume $V(1-\phi)$, integer box neutrality, and charge regulation; nonlinear Poisson–Boltzmann outside), `RNAv` enforces a four-criterion **M5 Stability Gate** prior to any external GROMACS or OpenMM submission **[architecture, measured]** [1, 18]. The criteria are:

1. Pair backbone RMSD directly with radius-of-gyration drift (R_g), internal $\mathrm{Cl^-}$ localization, and icosahedral crack checks (screening monotone drift as instability even under constant variance).
2. Audit and log separate per-compartment neutralizing ion counts (c_+, c_- inside vs outside the shell) before trajectory initiation.
3. Bound 10–50 ns runs strictly as a gross-failure screen alongside longer multi-hundred-nanosecond equilibration when discriminating force fields **[published]** [24].
4. Compare trajectories exclusively under matched genome occupancy, pore diameter, and force-field parameters.

6. Packaging-Signal Placement and Pocket Transport

Mapping M sequential packaging-signal hairpins injectively onto K symmetric interior binding pockets combines unary motif/electrostatic energies U_{tv} ($k_{\mathrm{B}}T$) with pairwise WLC linker penalties $W_{t-1,v',v} \le 0$. A clean counterexample proves why doubly stochastic unary optimal transport alone is insufficient: on a 4-pocket square whose WLC contour limit permits perimeter transitions while forbidding diagonals, uniform unary energies force log-domain Sinkhorn to return uniform 0.25 marginal weights everywhere, whereas only 8 of $4! = 24$ injective permutations remain topologically feasible **[measured]**. RNAv therefore restricts Sinkhorn **[published]** [12] to a unary proposal generator parameterized stably in log-domain potentials $f_i = \ln u_{\mathrm{OT},i}$ and $g_j = \ln v_{\mathrm{OT},j}$. The potential updates are:

$$f_i \leftarrow \ln a_i - \mathrm{logsumexp}_j\left(-\frac{C_{ij}}{\eta_{\mathrm{OT}}} + g_j\right), \qquad g_j \leftarrow \ln b_j - \mathrm{logsumexp}_i\left(-\frac{C_{ij}}{\eta_{\mathrm{OT}}} + f_i\right). \tag{8}$$

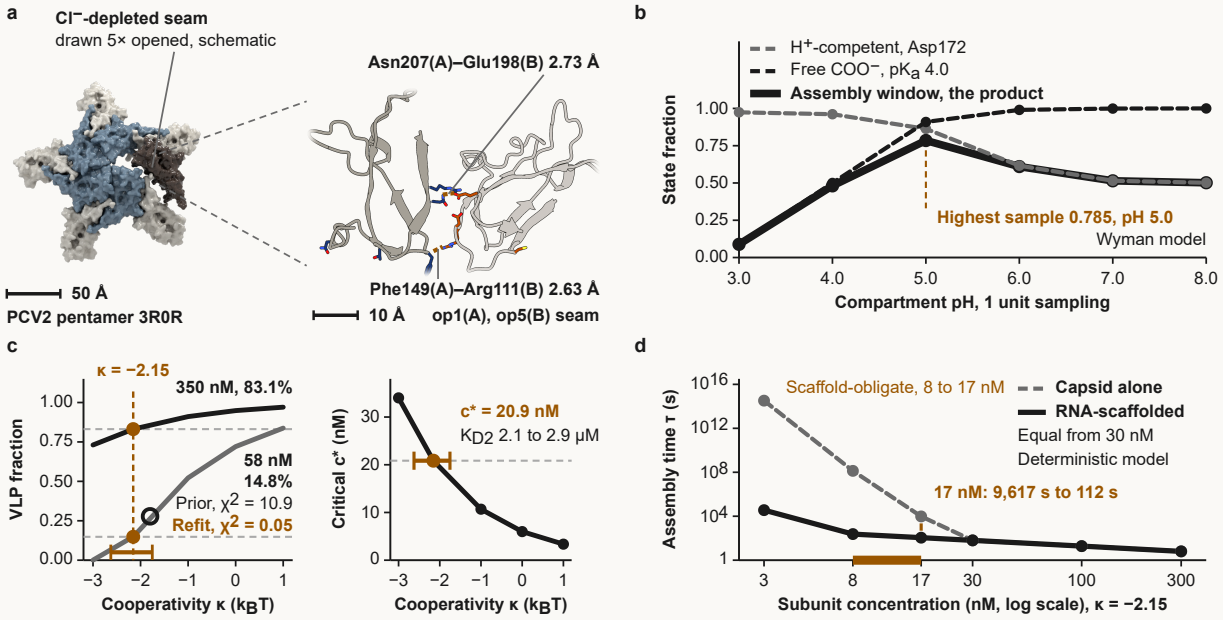

Figure 3: `PCV2` seam, pH curves, `mi3` cooperativity and nucleation. Panel a shows the `PCV2` pentamer (PDB `3R0R`, scale bar 50 Å) with the chloride-depleted seam, drawn 5 times opened as a schematic; the 3.5 Å crack is quoted from the cracking studies behind the report, and no measured overlay is drawn. The inset (scale bar 10 Å) gives two seam contacts between chains A and B, joined by orange dashed lines: Asn207 to Glu198, 2.73 Å, and Phe149 to Arg111, 2.63 Å. Panel b plots the proton-competent state of `Asp172`, the free carboxylate (pK_a 4.0) and their product, the assembly window, against pH at a sampling of 1 pH unit; the highest sample is 0.785 at pH 5.0, and above pH 6 the competent state and the product coincide. Panel c plots the fraction of assembled `Mi3` particles against the cooperativity at 350 nM (83.1%) and 58 nM (14.8%). The prior $\kappa = -1.8$ gives $\chi^2 = 10.9$ (open circle) and the refit $\kappa = -2.15$ gives $\chi^2 = 0.05$. The right plot gives the critical concentration $c^* = 20.9$ nM across $K_{D2} \in [2.1, 2.9]$ μM [3, 58]. Panel d plots the deterministic assembly time against subunit concentration from 3 to 300 nM; at 17 nM the RNA scaffold shortens it from 9,617 s to 112 s, the scaffold-obligate range is 8 to 17 nM, and the two curves are equal from 30 nM. **[measured, simulated, published]**

Exact chain marginals and partition functions run inside `subset_dp` over subset states $S \subseteq \{1,\dots,K\}$ with $|S| = t$ and terminal pocket $v \in S$ starting from $\alpha_1(v,\{v\}) = -U_{1v}$. The recursion and the partition function are:

$$\alpha_t(v,S) = -U_{tv} + \ln \sum_{v' \in S\setminus\{v\}} \exp\big(\alpha_{t-1}(v', S\setminus\{v\}) + W_{t-1,v',v}\big), \qquad \ln Z = \ln \sum_{|S|=M,\, v\in S} \exp\big(\alpha_M(v,S)\big), \tag{9}$$

The recursion matches brute-force chain enumeration to 10^{-14} in floating-point precision **[measured]**. Requiring $M \le K$ and pinning the first anchor by point-group symmetry saves a factor of $|G|$. Under the unified cumulative state count $N_{\text{states}}(M,K) = \sum_{t=1}^{M} t\binom{K}{t}$ guarded by memory ceiling 5×10^6, `subset_dp` passes $(4,12)$ (2,784 states; 22.4 ms), $(6,20)$ (333,280 states; 4.65×10^6 final-layer ops $MK\binom{K}{M}$), $(6,24)$ (1,069,248 states; 16.0 s), and $(4,60)$ (2,056,800 states; 1.17×10^8 ops). It abstains safely in 0.166 ms at $(15,60)$ (1.13×10^{15} states) and whenever final-layer ops exceed budget ($(7,20)$ at 1.09×10^7 ops). Position-only Markov DP on unstructured linkers holds only when seven physical inequalities hold simultaneously:

(i) Geometric reach $d_{\text{chord}} \le \ell_c$.

(ii) Tangent decorrelation $e^{-\ell_c/\ell_p} \le \delta$ ($\ell_c \ge 3\ell_p$ at $\delta = 0.05$) or free-hinge anchors.

(iii) Two-sided rigid-rod window $|d - \ell_c(1 - x/6)| \lesssim \ell_c x/(3\sqrt{10})$ for $x = \ell_c/\ell_p < 0.05$.

(iv) Gaussian regime $x \gtrsim 10$ and $d/\ell_c \lesssim 1/2$.

(v) Cavity clearance $\sqrt{\ell_p \ell_c/3} \ll R_{\text{in}}$ and $\ell_p \ll R_{\text{in}}$.

(vi) Odijk–Skolnick–Fixman screening $\lambda_D \ll 2\sqrt{\ell_B \ell_{p,0}}$ ($I_s \gg 32$ mM for $\ell_{p,0} = 1$ nm).

(vii) Tight anchor affinity $\varepsilon_b - F_{\text{bridge}} \gg k_{\text{B}}T$.

When linkers contain folded hairpins or stiff junctions, through-graph reach replaces ℓ_c, non-local pairing breaks first-order Markovity, and `M4` abstains.

For intractable (K,M), `LearnedPlacerNet` distills prefix conditionals into an autoregressive chain sampler estimated via self-normalized importance sampling (SNIS) and Kish effective sample size $\text{ESS} = (\sum_b w_b)^2 / \sum_b w_b^2$ **[measured]** [30]. Because iterated conditioning yields $(1 + \chi^2_{\min})^M \le 1 + \chi^2(p\|q) \le (1 + \chi^2_{\max})^M$, keeping $\text{ESS}/N \ge 0.5$ at $M = 24$ demands per-step error $\chi^2_{\max} \le 0.029$ (≤ 0.10 for $\text{ESS}/N \ge 0.1$), while empirical Kish ESS alone can miss unvisited modes ($\chi^2 = \infty$). Large-capsid SNIS therefore gates on defensive full-support mixture proposals, SMC resampling across chain steps, and forward MCMC coverage checks. On $(M = 4, K = 12)$ where $\ln Z_{\text{exact}} = -13.1053$ (Fig. 4b–d), 50 RTX 4060 training steps halved negative log-likelihood (1.8495 → 0.9094), reduced $|\Delta \ln Z|$ error (0.0174 → 0.0097), lifted ESS from 136.73 (26.7%) to 487.71 (95.3%) over 512 particles, and kept feasible walk validity at 100% **[measured]**. Across the corresponding 5-seed $(M = 4, K = 12)$ distillation ensemble (`rnav_fig_data.json`), mean NLL falls from 2.14 to 0.92, 512-particle ESS climbs from 131.8 to 383.8 (with 1/5 slower seed reaching 154), and mean $|\Delta \ln Z|$ shrinks from 0.075 to 0.023 **[measured]**.

7. Verification Suite and Hardware Benchmarks

Implementation spans three precision tiers inside `rnav/model/train.py` and the Rust desktop workspace `10_App`. Tier 1 locks Lie-algebra frame projections, Donnan ion balance, WLC series, and

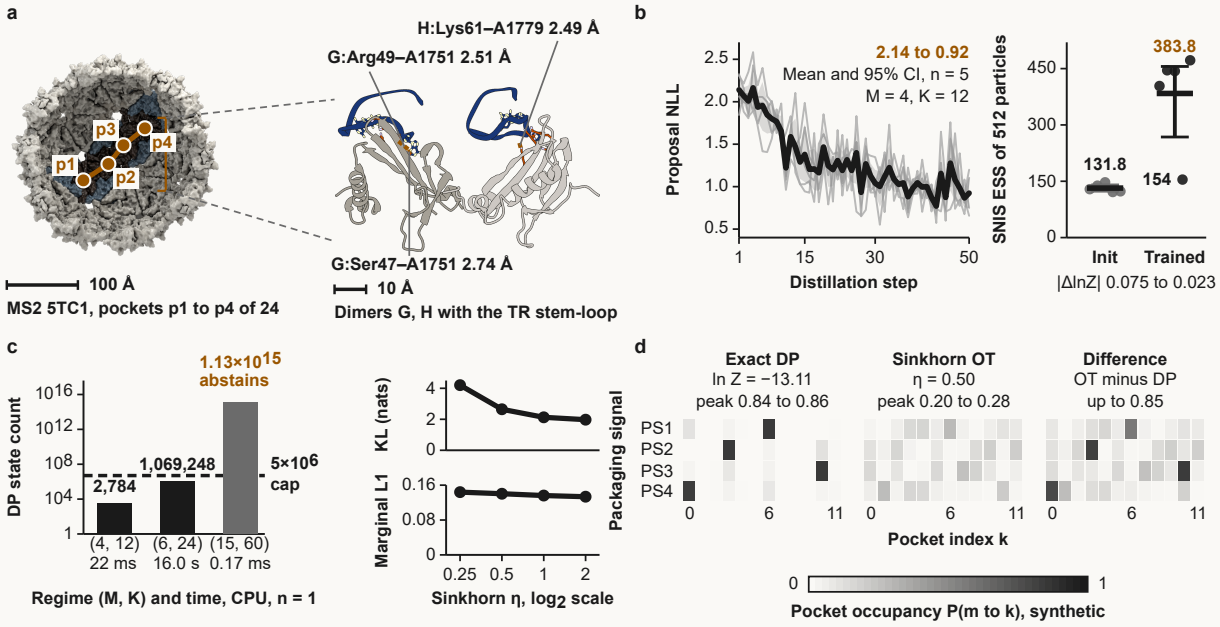

Figure 4: Pocket routing in `MS2`, distillation of the placer, `subset_dp` horizon and Sinkhorn marginals. Panel a shows the pocket walk `p1` to `p4` among the 24 interior pockets of the `MS2` shell (PDB `5TC1`, scale bar 100 Å) and the coat dimer G, H with the TR stem-loop (scale bar 10 Å); orange dashed lines join the atoms of the contacts G:Arg49 to A1751 (2.51 Å), G:Ser47 to A1751 (2.74 Å) and H:Lys61 to A1779 (2.49 Å). Panel b shows the distillation of `LearnedPlacerNet` for M = 4 stem-loops and K = 12 pockets, mean and 95% interval over 5 seeds: the proposal negative log-likelihood (NLL) falls from 2.14 to 0.92, and the effective sample size (ESS) of 512 self-normalised importance-sampling (SNIS) particles rises from 131.8 (init) to 383.8 (trained), with one seed at 154; $|\Delta \ln Z|$ falls from 0.075 to 0.023. Panel c gives the state count of the exact CPU solver of `subset_dp` for (M, K) of (4, 12), 2,784 states in 22 ms, and (6, 24), 1,069,248 states in 16.0 s (one run each); at (15, 60) it abstains after 0.17 ms, above the cap of five million states. The right plots sweep the Sinkhorn temperature eta from 0.25 to 2 on a log scale, with KL(DP ‖ OT) and the marginal L_1 error, each on an axis from zero. Panel d compares the 4 by 12 marginals of the exact DP ($\ln Z = -13.11$) and of Sinkhorn OT (eta 0.50), with their difference, for the packaging signals PS1 to PS4 over the pockets k, on synthetic input: the exact DP peaks at 0.84 to 0.86, Sinkhorn OT at 0.20 to 0.28, Sinkhorn names the same peak pocket as the exact DP in 1 of 4 rows, and the difference reaches 0.85. **[architecture, measured]**

`subset_dp` in FP32. Tier 2 executes the quotient co-folding trunk (`RNAvFoldingModel`: 971,323 active zero-shot default parameters; 1,530,941 total with +559,618 optional frozen-encoder adapters; $d_{\text{model}} = 128$, 3 layers, 4 heads) in BF16 autocast. Tier 3 targets Int8/ONNX export for local GUI execution **[architecture, measured]**.

Every core module passed sequential multi-model audits before hardware profiling **[measured]**. The audits cover TDD unit suites (171 Python unit tests, 60 end-to-end scenarios) and independent answer-key verification in Rust (138 workspace tests agreeing with `oracle_sol.rs` to 10^{-6}). They also include a GPT-6.1 Sol read-only audit (12 findings including 3 high-severity conformer, forbidden-pocket $\ln Z$, and lattice-index bugs repaired with regression tests) and GPT-6 Astra conjugacy-class verification. Table 5.2 xHigh's numerical audit repaired 8 FP32/autodiff edge cases ($\det Q = +1$, Ω_0, WLC $\hat{\ell}$, Citardauq c_Σ, η_{OT}, log-DP indexing, and δ_{clip} soft boundaries).

Tables 4 and 5 report live measurements collected on 14.XI.2026 (`rtx4060_benchmarks.json`) using a local NVIDIA GeForce RTX 4060 Laptop GPU (8,187.5 MB VRAM) under PyTorch 2.11.0+cu128 and CUDA 12.8. Across a 5-seed $L = 128$ BF16 ensemble recorded alongside the figure benchmarks (`rnav_fig_data.json`), synthetic optimization advances from -4.300 to -4.399 at 238.0 ± 9.6 ms/step and 97.2 MB peak VRAM. Section 12 and Figure 6 report the exact per-submodule parameter breakdown and stage-by-stage outputs for these runs.

Table 4: Measured FP32 vs BF16 forward–backward latency (ms) and peak VRAM (MB) for `RNAvFoldingModel` (971,323 active default / 1,530,941 total parameters, RTX 4060, 14.XI.2026). **[measured]**

Token Length L	FP32 (ms)	BF16 (ms)	FP32 VRAM	BF16 VRAM	Saved
64	261.68	282.35	56.08	51.20	8.7%
128	328.50	294.27	123.58	97.34	21.2%
256	306.30	301.78	340.59	261.41	23.2%
512	279.47	308.23	1,184.30	924.16	22.0%

Table 5: Measured 50-step GPU optimization runs and CPU exact `subset_dp` scaling (14.XI.2026). **[measured]**

Benchmark Workload	Measured Trajectory or Exact Output	Runtime	Outcome
`folding_50_steps` ($L = 128$)	Loss $-4.2871 \to -4.4672$ (-4.20%); max $\|\nabla\| = 7.5900$; all finite	300.18 ms/step	97.34 MB VRAM
`placer_50_steps` ($M = 4, K = 12$)	$\ln Z_{\text{exact}} = -13.1053$; NLL $1.8495 \to 0.9094$; ESS $136.73 \to 487.71$; $\|\Delta \ln Z\|$ $0.0174 \to 0.0097$	512-IS	1.000 valid
`subset_dp` ($M = 4, K = 12$)	Exact 2,784 subset states; $\ln Z = -3.7475$	22.36 ms	Exact CPU
`subset_dp` ($M = 6, K = 24$)	Exact 1,069,248 subset states; $\ln Z = -7.1289$	15,975.96 ms	Exact CPU
`subset_dp` ($M = 15, K = 60$)	Analytic state count 1,129,395,641,787,120; memory guard triggered	0.166 ms	Abstained

8. Architectural Novelty and Prior Foundations

Because structural biology already possesses strong general-purpose co-folders and specialized capsid builders, assessing RNAv requires separating pre-existing components from mechanisms that do not appear in prior systems. Table 6 isolates the engineering stack. Language-model embeddings (`ESM-2`, `RNA-FM`, `Nucleotide Transformer`), Sinkhorn transport [published] [12], Debye–Hückel and Donnan equations [published] [15, 55], and constant-pH PCV2/MS2 MD observables [published] [21, 53] are reused foundations. Four integrated mechanisms constitute methodological novelty within the codebase and mathematical design:

1. **Symmetry-Quotient Protein Trunk Coupled to an Asymmetric RNA Track.** Prior capsid predictors such as `FoldaVirus` [published] [46] reconstruct empty $T \leq 9$ coat-only shells by superimposing monomers onto family templates and deleting genome-binding disordered tails to prevent steric clashes, whereas distributed attention frameworks (`Fold-CP` [published] [36]) retain global $\mathcal{O}(N^2/P)$ pair grids across dozens of GPUs. No prior published co-folding architecture collapses the protein coat onto a group-quotient $(\Pi_0, E_0, \Omega_0, F_{g,a})$ representation ($60\times$ –$3600\times$ pair reduction) while concurrently propagating an asymmetric polymer-biased nucleic-acid residual [architecture].
2. **Chiral** $\det(Q) = +1$ SE(3) **Quotient Frames and Parity-Odd Scalar** $\chi_{\det}$**.** Generic $E(3)$-equivariant graph networks [published] [47] commute with improper spatial inversions, conflating chiral $(2,1)$ and $(1,2)$ Caspar–Klug cages and risking mirror-flipped secondary structures. Coupling Rodrigues updates, 4-step Newton–Schulz projections, and signed SVD safeguards with $\chi_{\det}$ guarantees stereochemical and shell enantiomer preservation [architecture, measured].
3. **WLC-Constrained `subset_dp` Oracle + SNIS-Calibrated Neural Placer with Analytic Abstention.** Standard amortized graph samplers and Sinkhorn layers fail when chain connectivity forbids diagonal transitions (§6). Distilling an exact 10^{-14}-verified `subset_dp` forward–backward oracle into an autoregressive neural sampler supplies calibrated uncertainty. The sampler reports finite-sample Kish ESS/N and triggers an 0.166 ms analytic abstention above intractable binomial state counts [measured].
4. **M0–M7 Modular Cut-Posterior Architecture Paired with Per-Compartment Donnan Pre-Flight MD Gating.** Enforcing statistical gradient cuts across symmetry geometry, lattice-gas cooperativity, ion partitioning, discrete placement, and external MD avoids parameter sloppiness across incompatible data densities while turning empirical PCV2/MS2 chloride failure modes into deterministic software pre-checks [architecture, measured].

9. Archive Ceilings and Falsification Criteria

What accuracy and reliability can RNAv realistically attain at Day 58 once the scheduled Tier A–C training completes? Separating projected performance into observables anchored by large structural archives versus regimes blocked by sparse genomic supervision keeps each target within its evidence. Table 7 pairs concrete Day-58 targets against the exact physical or empirical ceilings identified in §2.

Table 6: Methodological novelty audit separating RNAv contributions from assembled components. [architecture, published]

Architectural Layer	Prior Art Capability	RNAv Novelty and Structural Differentiation
Pair Representation	AF3 / Fold-CP keep dense $\mathcal{O}(N^2)$ full grids; FoldaVirus predicts empty $T \leq 9$ coat shells	Couples $(\Pi_0, E_0, \Omega_0, F_{g,a})$ ASU quotient protein trunk with asymmetric nucleic-acid residual inside 1 GPU
Spatial Equivariance	$E(n)$ EGNN permits reflections; AF3 relies on random-rotation data augmentation	Hard SE(3) constraint ($\det R_i = +1$), Newton–Schulz polar projection, and parity-odd invariant $\chi_{\det}$
Genome Placement	Coarse-grained MD (HBV pgRNA 3,200 nt); Sinkhorn ignores chain topology	Exact WLC-coupled `subset_dp` oracle distilling SNIS-calibrated `LearnedPlacerNet` with analytic abstention
Statistical Calibration	Single end-to-end diffusion loss with uncalibrated pLDDT ranking	M0–M7 cut-posterior isolation preserving ESS/N diagnostics and κ parameter identifiability
MD Stability Screening	Post-hoc trajectory inspection and global box neutralization	Pre-flight M5 per-compartment Donnan Cl^-/Na^+ verification preventing 5-fold cracking at < 3.5 Å RMSD

Two fundamental limits govern scientific interpretation at Day 58:

- **The 96-Entry Coordinate Supervision Ceiling.** No neural network trained predominantly on 96 icosahedral Protein/NA structures—where only a single entry (`8ER8`) exceeds 100 unique modeled nucleotides and `5TC1` leaves 93.2% of genomic nucleotides off the coordinate record [published] [13]—can infer unique atomic coordinates for an arbitrary unprobed 3–10 knt RNA sequence. On unprobed kilobase RNAs, thermodynamic branching equilibria produce broad structural ensembles whose base-pairing accuracy saturates near $F_1 \approx 0.35$–0.50 without RMDB/SHAPE/DMS chemical restraints ($F_1 \approx 0.45$–0.70 restrained).
- **Metric Sensitivity and Confounding.** CASP16 evaluations demonstrate that global template metrics often miss local protein–RNA contact failures [published] [28], while FoldBench notes lower homomeric than heteromeric interface accuracy. Consequently, Day-58 acceptance binds to **minimum per-interface i-lDDT**, **leave-one-system-out beat over a symmetric-mean baseline on `5TC1`'s** 242 **deposited nucleotides**, **cross-system validation across STMV SHAPE/RING and FHV vPAR-CL contacts**, and a **sign-inversion ablation control** reversing electrostatic/linker biases while preserving parameter budgets.

10. Engineering Domains and Application Scope

Judging where building and running RNAv helps scientific or biopharma teams requires matching each module—quotient shell geometry (**M1**), lattice-gas assembly (**M2**), Donnan/pH linkage (**M3**), discrete WLC placement (**M4**), and compartment MD pre-flight (**M5**)—against mathematical

Table 7: Projected Day-58 outcomes, empirical/archive bounds, and pre-registered falsification thresholds. [target, published]

Evaluation Target	Projected Day-58 Outcome	Empirical or Physical Ceiling	Pre-Registered Falsifier Condition
Held-out coat ASU & capsomer interfaces	i-lDDT ≈ 0.72–0.82; 100% chiral $(2,1)$ vs $(1,2)$ fidelity	1,825 icosahedral assemblies / 1,799 VIPERdb structures supply rich coat priors	Reject quotient trunk if i-lDDT trails dense baseline by > 0.03 on matched $T = 1, 3$ timers
Discrete PS chain calibration ($M \leq 8, K \leq 30$)	SNIS ESS$/N \geq 0.85$–0.92; $\lvert\Delta \ln Z\rvert \leq 0.02$; 100% valid chains	Exact DP covers $M \leq 6, K \leq 24$ cleanly; autoregressive variance rises past $M \geq 12$	Reject `LearnedPlacerNet` if ESS$/N < 0.25$ at $M = 12, K = 60$ vs uniform baseline
Radial ssRNA shell & cryo-EM density correlation (≤ 3.5 knt)	Interior density peak within 80–120 Å; reproduces 5-fold/q6 enrichment	HBV 3,200 nt CG benchmark + MS2 `5TC1` (242/3,569 nt atomic, 16 SLs)	Reject genome layer if 80–120 Å shell & 5TC1 residual fail to beat symmetric mean
Unprobed kb secondary structure / atomic 3D RNA	2D $F_1 \approx 0.35$–0.50 (unprobed); 0.45–0.70 with SHAPE/DMS restraints	Only 96 global-I Protein/NA entries (`8ER8` 199 nt max); RNA thermal degeneracy	Reject any atomic whole-virion 3D coordinate claim lacking cryo-EM or probing restraints
M5 pre-flight setup & Rust Int8 edge runtime	100% Donnan compartment neutrality; 2.2–3.0× Int8 inference speedup	Exact algebraic balance in Citardauq/asinh FP32; 2.5× typical 8-bit op gain	Reject Int8 export if $\det Q = +1$ fails or SNIS ESS$/N$ drops > 0.05 relative to BF16

validity bounds. Table 8 partitions applications into twelve sub-regimes (**R1a**–**R8**) calibrated against formal mathematical gates.

Table 8: Domain-by-domain engineering verdict partitioned by explicit mathematical validity regimes R1a–R8. [architecture, published]

Regime / Application Domain	Verdict	Governing Validity Bounds and Actionable Value
R1a: Rigid G-symmetric cages, exact DP ($\sum_{t\le M} t\binom{K}{t} \le 5\times 10^6$, $M \le K$; `mi3`, MS2, Butterfield cages)	**STRONG YES**	Symmetry G and pocket count K are fixed; pinning pocket 1 cuts states by $\lvert G\rvert$; satisfies all 7 WLC Markov bounds ($I_s \gg 32$ mM for OSF, $e^{-\ell_c/\ell_p} \le 0.05$, $\varepsilon_b - F_{\text{bridge}} \gg k_{\text{B}}T$) [published] [8, 54].
R1b: Rigid G-symmetric cages, SNIS ($M > 10$, e.g. $M = 24, K = 60$; PEG10/eVLP payloads)	**CONDITIONAL YES**	Conditional on per-step error $\chi^2_{\max} \le 0.029$ (ESS/$N \ge 0.50$ at $M = 24$), defensive full-support proposals, SMC resampling, and forward MCMC coverage checks against missed modes [published] [5, 48].
R1c: Rigid cages with folded linkers or stiff junctions	**ABSTAIN M4 / PASS M0–M3**	Through-graph reach $< \ell_c$ and non-local base pairing break first-order Markov factorization; M4 abstains while shell symmetry, assembly, and Donnan modules M0–M3 hold.
R2: Deformable shells / fluctuating N (elastic sheets, immature VLPs)	**PARTIAL**	M1 covers the $3n_0$-dim G-invariant subspace, missing $1 - 1/\lvert G\rvert$ thermal modes and $l = 2\ldots 5$ elastic modes; M2 needs bending rigidity $\kappa_b/k_{\text{B}}T \gg N_0^2/4\pi$ (32 at $N_0 = 20$; 286 at 60) and $-2\kappa_b < \bar{\kappa} < 0$; only Donnan ($d\ln\rho_f/d\ln N = -3/2$) and geodesic patch spans ($L_{\text{patch}} \le R_{\text{curv}}\sqrt{24\varepsilon}$) transfer.
R3a: Cryo-EM C1 interior asymmetric priors	**CONDITIONAL YES**	Enforces G0–G6 independent-half FSC gates ($\text{FSC}_{\text{raw}} \ge z_\alpha/\sqrt{n_k}$, $\text{FSC}_{\text{free}} \le \sqrt{\text{FSC}_{\text{raw}}} + z_\alpha/\sqrt{n_k}$, $\lambda \le s(k)$, decoy-prior check) plus orientation discriminability $d'^2 \gtrsim 2\ln\lvert G\rvert \approx 8.2$ to prevent Einstein-from-noise priors.
R3b: M5 all-atom MD compartment ion pre-flight	**CONDITIONAL YES**	Exact 1:1 compartment neutrality holds within $\ell_{\text{inh}} \ll \lambda_D \ll R_{\text{in}}$ (7.9 Å at 0.15 M, 96 Å at 1 mM), using $V(1-\phi)$, integer box neutrality, and pH feedback; wall-clustered charge switches to nonlinear Poisson–Boltzmann [measured, published] [21, 52].
R4a: Monovalent & weak-charge interior ($\Xi \le 10$, fixed-point pH solved)	**YES**	Coupling $\Xi = 2\pi z^3 \ell_B^2 \sigma_s = 3.21 z^3 \sigma_s[\text{nm}^{-2}] = 2\Gamma^2$; mean-field controlled for $\Xi \lesssim 1$ ($\sigma_s \le 0.31$ e/nm^2 at $z = 1$), crossover 1–10 ($\sigma_s \le 3.1$); eliminates 1.93× uncoupled Citardauq Δc_{osm} bias at $n_A = 0.4$ M, $c_s = 0.1$ M.
R4b: Strong coupling ($\Xi > 10$; condensed Mg^{2+}, spermine, dense walls)	**ABSTAIN**	At 1 e/nm^2, $z = 2$ gives $\Xi = 25.7$ and $z = 3$ gives $\Xi = 86.7$, causing like-charge correlations outside Poisson–Boltzmann; dilute ion pairing tracks Bjerrum criterion $z_+ z_- \ell_B/a_{\text{ion}} > 2$.

Regime / Application Domain	Verdict	Governing Validity Bounds and Actionable Value
R4c: Multivalent weak surface charge ($z \geq 2$, $\Xi < 1$; $\sigma_s < 0.039$ e/nm^2)	**NL-PB ONLY**	Use nonlinear Poisson–Boltzmann and two-species Manning competition; omit 1:1 fits whenever Mg^{2+}, phosphate, or citrate chelators are present.
R5: Driven ATPase genome insertion (AAV Rep, phage portal)	**PARTIAL**	Provides a necessary second-law fill limit ($\Delta F_{\mathrm{eq}}(L) \leq \langle W \rangle \leq F_{\mathrm{stall}} L$) and thin-shell hoop wall stress check ($\Delta\Pi\, R_{\mathrm{in}}/(2h) \leq \sigma_{\mathrm{rupture}}$, recovering the 5–10× $R_{\mathrm{in}}/2h$ geometric factor around motor portals such as ϕ29 `1FOU` **[published]** [50]); broken detailed balance ($\Delta\mu \neq 0$, Péclet $\mathrm{Pe} = v\tau_{\mathrm{relax}}/L$) forbids equilibrium kinetic predictions.
R6a / R6b: Interface modulators (HBV CAMs) via directed cut	**CONDITIONAL**	Propagates sensitivity $\delta \ln c^* = (30/19)\Delta\Delta g - (11/19)\Delta\kappa$ (with $\Delta\Delta g = -\delta a = \Delta\Delta G_{\mathrm{dim}}/(k_{\mathrm{B}}T)$ at fixed κ) across 30 contacts (modeled on HBV core-modulator structures such as `5E0I` [31]): holding 2.3× (1σ) precision demands $\sigma_{\Delta G} \leq 0.31$ kcal/mol (**R6a**); 1–1.5 kcal/mol docking scores permit sign/rank only (**R6b**: 54.5× factor at 1.5 kcal/mol; nucleation rates abstain unless $E^{\ddagger}\sigma_{\Delta G}/(k_{\mathrm{B}}T) \leq \ln f_{\mathrm{tol}}$) **[published]** [37, 39, 44].
R7a / R7b: Long unprobed RNA chains ($N_c \geq 3{,}000$ nt)	**NO (3D) / YES (BULK)**	Resolving overlap with R1 on $N_c \in [3{,}000, 4{,}500]$: **R7a** forbids unique 3D coordinates by Bayes risk ($\geq \mathrm{tr}\,\mathrm{Cov}(X\|S)$), Fano ($I \gtrsim c_{\mathrm{eff}} N_c$), G-equivariance, trapping ($\tau_{\mathrm{relax}} \gg \tau_{\mathrm{asm}}$), and 12.8% → 1.3% pair support ($[2L_{\max}N_c - L_{\max}(L_{\max}+1)]/[N_c(N_c-1)]$ at $L_{\max} = 199$ nt for $N_c = 3{,}000 \to 30{,}000$); **R7b** allows self-averaging observables (R_g, ion/radial profiles, fluctuations $\sim N_c^{-1/2}$) via R1/M3.
R8: Exterior envelopes, glycans, receptors, or crowders	**ABSTAIN OUTSIDE RES.**	Exterior enters solely via bulk intensive variables ($c_s, \mathrm{pH}_{\mathrm{bulk}}, c_1$); interior verdicts hold iff coupling satisfies $\beta\|H_c\|_\infty \ll 1$ and shell thickness obeys $h \gg \lambda_D$ (breaking below $I_s \lesssim 10$ mM for $h \approx 3$ nm).

10.1. Boundary Equations and Cut Guarantees

Three formal identities summarize why RNAv restricts autonomous claims to the **R1**, **R3**, **R4a**, and **R7b** corridors. The identities are:

$$P_{\mathrm{err}} \geq 1 - \frac{I(X;Y) + \ln 2}{c_{\mathrm{eff}} N_c}, \qquad \Delta\Pi\, \frac{R_{\mathrm{in}}}{2h} \leq \sigma_{\mathrm{rupture}}, \qquad |\ln p(X) - \ln p_0(X)| \leq 2\beta\|H_c\|_\infty. \quad (10)$$

Respectively, Fano's inequality establishes that identifying a single microstate among $W = e^{c_{\mathrm{eff}} N_c}$ folded RNA basins requires extensive probing information $I(X;Y) \gtrsim c_{\mathrm{eff}} N_c$ (at least $c_{\mathrm{eff}} N_c / \ln N_c$ crosslinks; **R7a**). Laplace thin-shell mechanics supplies a $R_{\mathrm{in}}/(2h)$ amplification factor on equilibrium osmotic pressure checks during active viral packaging (**R5**). Bounded perturbation theory confines exterior coupling error below $2\beta\|H_c\|_\infty$ (**R8**). On CAM drug triage (**R6**), Bayesian cut

conditioning isolates downstream thermodynamic estimation $p_{\mathrm{cut}}(\psi \mid \mathcal{D}_{\mathrm{asm}}, \mathcal{D}_{\mathrm{bind}})$ from an uncalibrated upstream binding scorer [published] [37, 44]. Alternatively it reverses the cut $p(\psi \mid \mathcal{D}^{0}_{\mathrm{asm}})\, p(\phi \mid \psi, \mathcal{D}^{\mathrm{mod}}_{\mathrm{asm}}, \mathcal{D}_{\mathrm{bind}})$ when M2 data is trusted to prevent silent score error propagation, while treating cut boundaries as uncertainty screens rather than docking validation.

Bottom-Line Engineering Assessment. RNAv operates as a **precision design and cryo-EM/MD triage engine for rigid synthetic nucleocapsids, engineered RNA VLPs, G0–G6 asymmetric reconstructions, and bulk compartment observables**, where symmetry, cargo charge, and packaging-signal spacing are experimental control knobs. When $\Xi > 10$, $\kappa_b / k_{\mathrm{B}}T \lesssim N_0^2/(4\pi)$, $\beta \|H_c\|_\infty \sim 1$, or kilobase unprobed trajectories seek unique 3D structures, the platform programmatically abstains from emitting unreliable atomic coordinates.

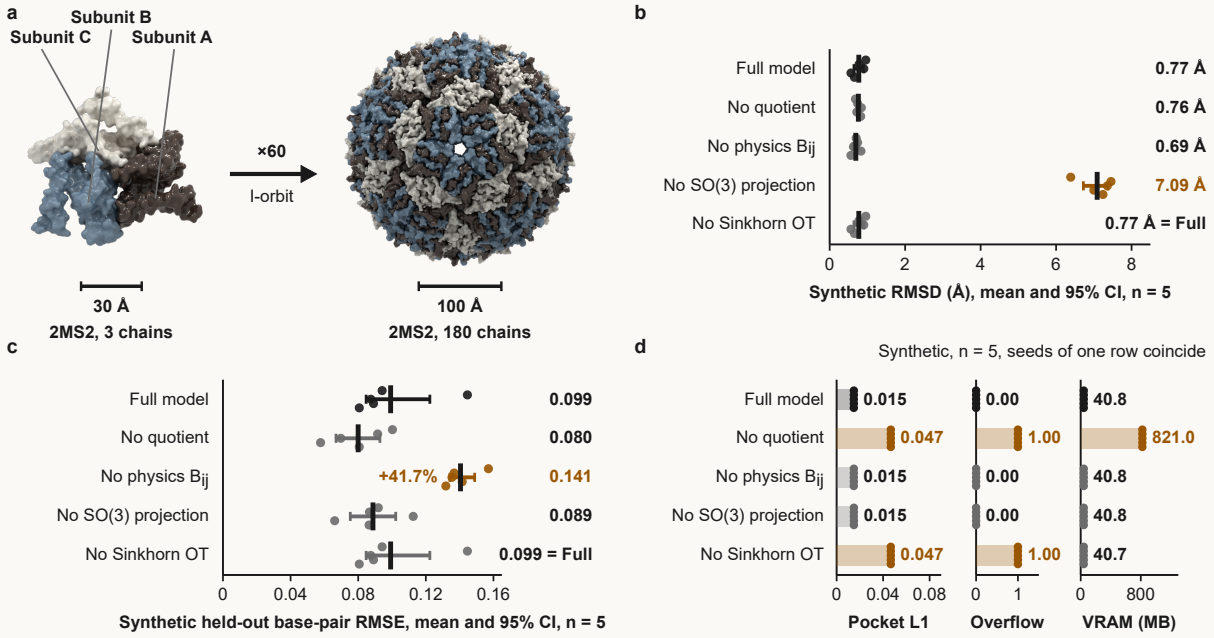

Figure 5: `MS2` quotient lift and 5-seed ablations. Panel a shows the `MS2` coat trimer (PDB `2MS2`, subunits A to C, 3 chains, scale bar 30 Å) lifted by the ×60 icosahedral orbit into the 180-chain capsid (scale bar 100 Å). Panels b to d compare the full model with four variants, each without one component (quotient, physics bias B_{ij}, SO(3) projection, Sinkhorn OT), on 5 synthetic seeds; orange marks the variants that depart from the full model. Panel b: the coordinate RMSD is 0.77 Å for the full model and 7.09 Å without the SO(3) projection. Panel c: the held-out base-pair RMSE is 0.099 for the full model and 0.141 without the physics bias, a rise of 41.7%. Panels b and c show the mean and the 95% interval stored with the data. The variant without Sinkhorn OT has the same five values as the full model in both panels (marked = Full): Sinkhorn OT belongs to the pocket routing, and panel d is where it changes the result. Panel d: without the quotient or without Sinkhorn OT the pocket L1 error rises from 0.015 to 0.047 and the crowding overflow from 0.00 to 1.00; without the quotient, peak VRAM rises from 40.8 MB to 821.0 MB. Panel d shows the mean as a bar and the seeds as dots, which coincide within a row. **[architecture, measured, simulated]**

11. Execution Schedule and Resource Allocation

Following verified foundational completion (**G0**, 14.XI.2026), Table 9 schedules the remaining co-folding campaign across four phases totaling 58 calendar days inside a 60.8-day two-month

horizon [target]. Every phase closes on machine-verifiable numerical gates synchronized with `TRAINING_AND_CODE_PLAN.md`.

Table 9: Four-phase 58-day execution plan aligned with verification gates G1–G6. [target]

Phase / Gate	Duration	Core Deliverables	Quantitative Acceptance Gate
Phase 1 (`G1`/`G2`)	10 days (D1–D10)	SHA-256 accession manifest; frozen ESM-2, RNA-FM, NT feature caches; `run.sh` container check on RTX 4060 and Slurm [published] [61]	All hashes pinned; container resume verified prior to cluster GPU billing
Phase 2 (`G3`/`G4`)	16 days (D11–D26)	Tier A pilot co-folding ($L \leq 512$, BF16); `LearnedPlacerNet` SNIS calibration up to ($M = 6, K = 24$)	ESS$/N \geq 0.90$ and $\lvert\Delta \ln Z\rvert \leq 0.02$ vs `subset_dp`; automatic checkpoint rollback if $\max \lVert\nabla\rVert > 10$
Phase 3 (`G5`)	18 days (D27–D44)	Tier B quotient co-folding ($T_{\mathrm{CK}} \in \{1, 3, 4, 7\}$, HX, PRD1) across `seq+seq`, `seq+struct`, `struct+struct`	$\det Q = +1$ invariant across all steps; chiral $(2, 1)$ vs $(1, 2)$ discrimination verified
Phase 4 (`G6`)	14 days (D45–D58)	Tier C Int8/ONNX distillation; held-out MS2, `mi3`, PCV2, TMV tests + pre-registered falsifiers; M5 compartment ion prep	Per-compartment ion neutrality passed before external GROMACS/OpenMM MD

Table 10 compares two provisioning pathways. The **Hybrid** route is institutional cluster time with rented spot GPUs plus object storage. The on-demand **Commercial Cloud** pathway is converted at 85.55 RUB/USD on base USD lower bounds with nearest-integer +20% contingency headroom on every row.

Table 10: Itemized Tier A–C training budgets in RUB and USD with nearest-integer 20% headroom (85.55 RUB/USD). [target]

Tier	Hybrid (Base + Stor. + 20%)	Hybrid (RUB)	Comm. (RUB)	Commercial USD
Tier A pilot	26,758 + 5,000 + 6,352	38,110	1,272,984	\$14,880–\$25,610
Tier B co-folding	53,515 + 10,000 + 12,703	76,218	2,200,312	\$25,720–\$44,300
Tier C campaign & distill	1,798,699 + 50,000 + 369,740	2,218,439	5,446,114	\$63,660–\$109,680

12. Whole Platform Architecture

RNAv pairs deterministic symmetry and physical kernels in FP64 with a chirality-strict neural trunk (`RNAvFoldingModel`) in mixed BF16/FP32 precision inside the `rnav` Python package and the Rust `10_App` workspace [architecture, measured]. Default zero-shot execution uses 971,323 active weights ($d_{\mathrm{model}} = 128$, 3 layers, 4 heads), while enabling optional adapters for frozen protein and RNA foundation models brings the model to 1,530,941 weights (+559,618). Companion modules outside

the neural forward graph run exact quotient geometry (`QuotientGraph`), Grand-Canonical lattice gas (`LatticeGas`), exact log-domain routing (`subset_dp`), and a 38,209-parameter autoregressive placer (`LearnedPlacerNet`). Figure 6 displays the whole-RNAv pipeline with its cut boundaries, the neural rails with a zoomed SE(3) update block and its 5-seed projection ablation, the three prompt modes with byte-pair encoding, the `subset_dp` and GRU placer, and the Rust and WASM virion studio.

12.1. System Topology and Schemas

Data enters `rnav` through typed dataclasses inside `rnav/schemas.py` and `rnav/model/prompt.py` [architecture]. `ProtomerSpec` computes sequence charge $q(\mathrm{pH})$ from EMBOSS pK_a tables, `VirionSpec` accounts for mature virion stoichiometry ($60T_{\mathrm{CK}}$ less displaced coat subunits), `CapsidSymmetryGrid` stores proper SO(3) site orientations, and `PackagingSignalAnchor` tracks anionic motif charges $-(L-1)$. Interactive co-folding accepts `PromptItem` inputs across Mode (a) dual sequences (`SequencePrompt` validated against `ACGU` or 20 canonical amino acids), Mode (b) one sequence paired with one `StructurePrompt` holding $(N, 3)$ coordinates and optional $(N, 3, 3)$ frames, and Mode (c) paired structures [architecture]. Every forward call returns one typed `FoldingOutput` container holding contact probabilities (L, L), transport assignments (M, K), chiral coordinates $(L, 3)$, physical energy score, and metadata.

Hierarchical Bayesian cuts $\pi_{\mathrm{cut}}(\theta_1, \theta_2) = p(\theta_1 \mid D_1)p(\theta_2 \mid \theta_1, D_2)$ partition upstream FP64 geometry (`M0--M1`) and equilibrium chemistry (`M2--M3`) from neural and discrete inference (`M4,M6`), all-atom stability screening (`M5`), and studio export (`M7`) [architecture, published] [37, 44]. Beside the Python library, the Rust workspace (`10_App/crates/{shared,server,web}`, verified across 138 tests on 15.XI.2026) serves a local Axum API at `127.0.0.1:8787` backed by a 64-entry SHA-256 geometry cache, driving a Leptos 0.8 and Mol* 5.13.1 WebAssembly viewer (`Trunk`). Rust `geometry.rs` builds `10T+2` Caspar–Klug shells matching Python `capsomer_centers` to 10^{-6} along with helical, packaged-spiral, and Fibonacci membrane scaffolds, exporting `mmCIF`, `PDB`, `.mvsj`, allowlisted SAMSON `T0/T1` profiles, and resolution-tagged PNG images. Table 11 maps every Python and Rust subsystem.

12.2. Symmetry and Physics Kernels

Point-group rotations in `rnav/symmetry/groups.py` generate exact SO(3) subgroups $(\mathrm{C}_n, \mathrm{D}_n, \mathrm{T}, \mathrm{O}, \mathrm{I})$ via breadth-first matrix closure in FP64 [architecture]. Structure identification in `symmetry/identification.py` classifies coordinates across three symmetry manifolds: point groups (`PG`) via real-projective equal-radius voting and Steinhardt Q_l bond-orientational invariants, helical assemblies (`HX`) via radial-variance axis optimization, and Caspar–Klug cages (`CK`) via convex-hull Euler degree graphs matched by Kabsch alignment. Given triangulation $T_{\mathrm{CK}} = h^2 + hk + k^2$, `expand_exact` replicates T-chain asymmetric units across the 60 VIPER rotations while preserving every crystallographic conformer and disordered atom, and `expand_lattice` orients seed protomers onto `CapsidSymmetryGrid` frames (`+z` outward, `+x` capsomer center). Inside the ASU (`QuotientGraph`), `ZeroCoMPseudoPrecision` computes the translational-null marginal precision $\Omega_0 = (A + cE_0E_0^\top)^{-1} - c^{-1}E_0E_0^\top$ via FP64 Cholesky Schur complement with ridge 10^{-4}, while `CosetInversionKernel` pairs learned group-element embeddings with equivariant residue coordinates to sample genome-induced asymmetric breakings.

Thermodynamic and polymer kernels avoid singular limits analytically [architecture, measured].

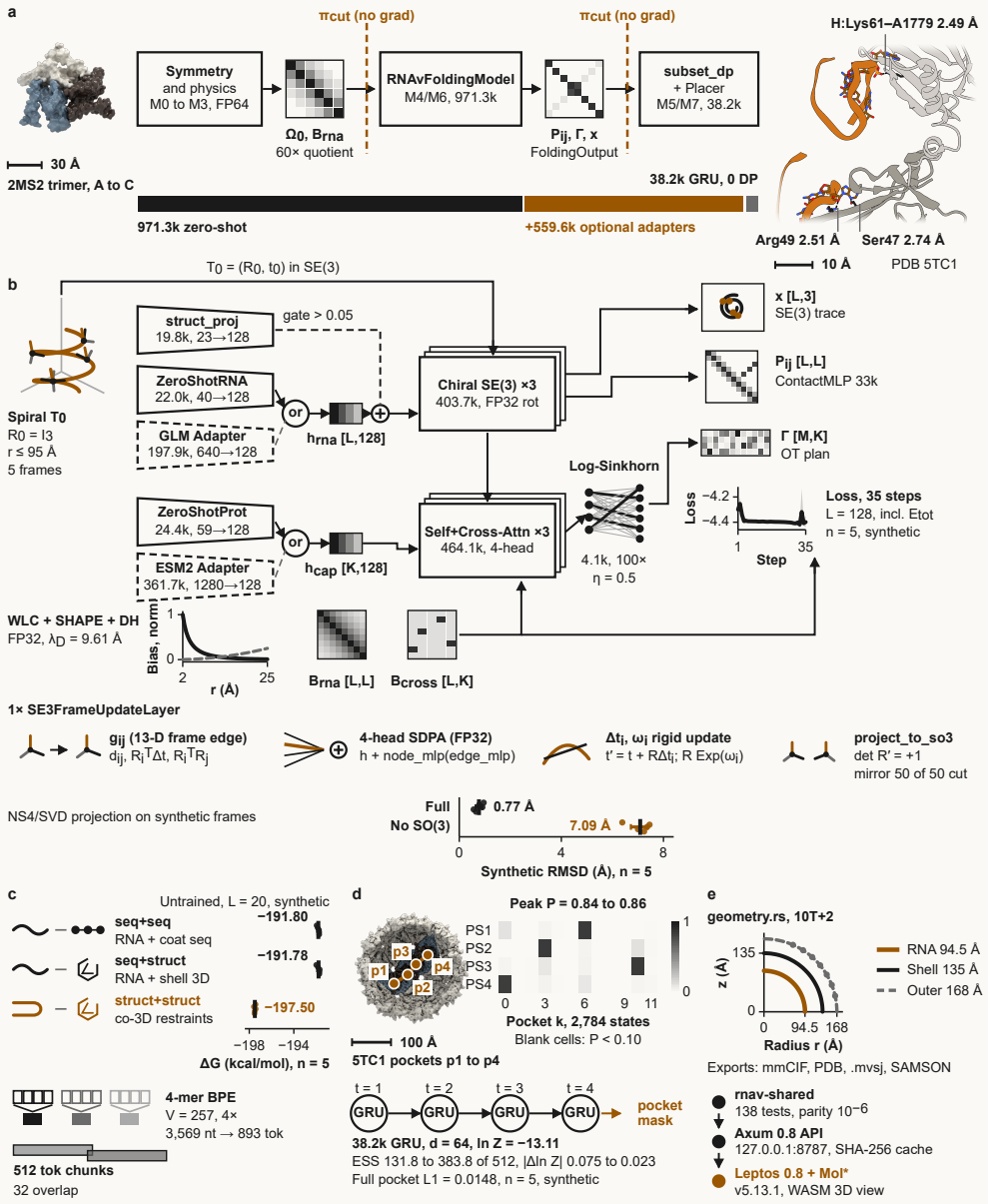

Figure 6: Whole-RNAv pipeline, neural rails, prompt modes, pocket placer and virion studio. Panel a shows the symmetry and physics modules (`M0--M3`, FP64), `RNAvFoldingModel` (`M4`, `M6`; 971.3k) and `subset_dp` with the placer (`M5`, `M7`; 38.2k); no gradient crosses a cut boundary π_{cut}. The bar gives the budget: 971.3k zero-shot, +559.6k optional adapter and 38.2k GRU parameters. The `5TC1` interface (scale bar 10 Å) carries the contacts of Arg49 (2.51 Å), Ser47 (2.74 Å) and Lys61 (2.49 Å). Panel b opens the neural rails: the spiral frames of `T0`, the zero-shot encoders or, dashed, the optional adapters ("or": one of the two feeds the rail), three chiral SE(3) blocks, three attention blocks, log-Sinkhorn transport at temperature eta 0.5 and the outputs; the loss curve shows the mean and the 95% interval of 5 synthetic seeds over 35 steps. Below, one `SE3FrameUpdateLayer` ends in the projection to SO(3), which holds the coordinate RMSD at 0.77 Å against 7.09 Å without it. Panel c shows the three prompt modes with the free energy before training at length 20 (−191.80, −191.78 and −197.50 kcal/mol), the 4-mer byte-pair encoding (3,569 nt, 893 tokens) and chunks of 512 tokens with 32 overlap. Panel d shows the pocket walk `p1` to `p4` (scale bar 100 Å), the exact `subset_dp` marginals of the packaging signals PS1 to PS4 over the pockets k (2,784 states), and the GRU placer, one step t per signal; its effective sample size rises from 131.8 to 383.8 of 512. Panel e shows the Rust and WASM virion studio: shell radii 94.5, 135 and 168 Å, the `rnav-shared` package (138 tests), the Axum API and the Leptos and Mol* viewer, and four export files. **[architecture, measured, simulated]**

Table 11: RNAv software subsystem map across Python (`03_Platform/code/rnav`) and Rust (`10_App`). [architecture, measured]

Subsystem	Live Source Modules	Primary Mathematical & Software Role
Schemas & Prompts	`schemas.py`, `model/prompt.py`	Typed virion specs, EMBOSS titration, Modes (a)–(c) input validation, `FoldingOutput`
Symmetry & Assembly	`icosahedral.py`, `assemble.py`, `symmetry/*`, `assembly_graph.py`	Point-group BFS closure, CK/PG/HX detection, $60T$ lattices, `QuotientGraph`, `CosetInversionKernel`, 2^{20} Burnside orbits
Physics Kernels	`phase_boundary.py`, `physics/{wlc, shape, electrostatics, attention_bias}.py`	Grand-canonical `LatticeGas`, `hypot`/`asinh` Donnan, 8th-order Taylor WLC, blended Deigan SHAPE, pair bias
Neural Trunk	`model/{adapters, folding_model, se3_frames, cross_attention}.py`	Zero-shot / frozen PLM/GLM `if/else` adapters, 23-D `struct_proj`, 3× chiral SE(3) frame trunk, 3× bidir. physics cross-attn, contact/E_{tot} heads
Discrete Routing	`matching.py`, `physics/matching_ot.py`, `model/train.py`	Log-domain Sinkhorn OT, exact FP64 bitmask `subset_dp` ($\leq 5 \times 10^6$ states), `LearnedPlacerNet` GRU, 512-SNIS
Mixed Precision	`model/train.py`	`PrecisionTrainer`: FP32 master weights, BF16 linear autocast, FP32 loss and geometry, norm clip 1.0
Studio Workspace	`10_App/crates/{shared, server, web}`	Deterministic `10T+2` CK/helical/spiral builders, Axum SHA-256 cache API, Leptos + Mol* viewer, `mmCIF`/`.mvsj`/SAMSON export

Screened electrostatics (`phase_boundary.py`, `physics/electrostatics.py`) solve interior/exterior ion partitioning with overflow-free `math.hypot` (PyTorch `asinh`) alongside Manning counterion condensation ($b = 3.5$ Å) and discrete 20-trimer bitmask assembly lattices (`uint32 popcount` for `connected_subsets` and `closed_subsets`). Linker physics (`physics/wlc.py`) switch smoothly to an 8th-order Taylor expansion for $L/\ell_p < 0.05$ in FP64 (< 0.5 in FP32) and evaluate differentiable harmonic lumen walls ($\sqrt{\|x\|^2 + \delta^2}$). Chemical probing (`physics/shape.py`) blends free and capsid-bound reactivities linearly prior to the Deigan transform [published] [25], counting terminal helix pairs once and internal stacked pairs twice while gating out sentinels ≤ -500.0. Pairwise attention bias (`physics/attention_bias.py`) combines canonical base-pair energies (GC $= -3.0$, AU $= -2.0$, GU $= -1.0$ kcal/mol, $|i - j| \geq 4$, $RT = 0.6163$ kcal/mol) with exact WLC linker reach bounds clipped at -10^4.

12.3. Neural Trunk and Channels

Model topology adapts single-sequence and pairwise structural representations from AlphaFold 2 and 3 [published] [2, 27] into a dual-track biomolecular trunk tuned for symmetry-reduced viral shells. Four parallel feature streams initialize the computation inside `RNAvFoldingModel` (`d_model=128`). They are:

1. **RNA Sequence Rail ($\mathbf{h}_{\text{rna}} \in \mathbb{R}^{L \times 128}$):** Via an `if/else` selector in `MultiModalFeatureExtractor`, supplied foundation embeddings pass through `GenomicGLMAdapter` ($640 \to 256 \to 128$, 197,889 params), replacing `ZeroShotRNAExtractor` (22,016 params, 4 base chemistries concatenated with 32 sinusoidal positions, `40->128->LN->GELU->128`). Long transcripts pass

through a non-overlapping $k = 4$ byte-pair tokenizer (`vocab=257`) processed in chunks of 512 with 32-token overlap.

2. **Capsid Pocket Rail ($\mathbf{h}_{\text{cap}} \in \mathbb{R}^{K\times 128}$):** Similarly selected via `if/else`, `ESM2ProteinAdapter` ($1280 \to 256 \to 128$, 361,729 params) replaces `ZeroShotProteinExtractor` (24,448 params, 20 amino-acid biochemical properties plus 32 positional features, `59->128`) when ESM-2 embeddings are supplied.
3. **Analytical Pair Bias Rail ($B_{\text{rna}}, B_{\text{cross}}$):** Pre-computed WLC reach, Deigan SHAPE, and canonical Watson–Crick energies form $B_{\text{rna}} \in \mathbb{R}^{L\times L}$, while FP32 screened Debye–Hückel potentials ($\ell_B = 7.14$ Å, $\lambda_D = 9.61$ Å, $r \geq 2.0$ Å scaled by $1/RT$) populate $B_{\text{cross}} \in \mathbb{R}^{L\times K}$.
4. **Chiral Coordinate and Frame Rail ($T_i = (R_i, t_i) \in \text{SE}(3)$):** Initial spiral beads inside inner radius 95 Å initialize proper rotations ($\det R_i = +1$). When structural conditioning exceeds gate threshold 0.05 (Modes b and c), `InvariantStructureProjection` (19,840 params) projects a 23-D invariant signature—center distance $r_{\text{com}}/100$, 16 radial Gaussian RBF bins (0–200 Å), signed local chirality $\text{asinh}(\det[\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3]/10^3)$ at 20/60 Å decay, neighbor bond ratios $d_{\text{prev}}/20, d_{\text{next}}/20$, bend cosine $\cos\theta$, and backbone dihedral $(\cos\tau, \sin\tau)$—and adds it directly to $\mathbf{h}_{\text{rna}}$.

The main trunk processes these four rails in three sequential stages [architecture, measured]. First, `se3_trunk` ($3 \times$ `SE3FrameUpdateLayer`, 403,710 params) maps 13-D relative local geometry through an edge MLP ($13 \to 32 \to 4$ attention biases), updates $\mathbf{h}_{\text{rna}}$ with 4-head attention and residual MLPs. It predicts $\Delta t_i, \omega_i$ ($128 \to 64 \to 3$), applies $t_i' = t_i + R_i \Delta t_i$, and projects $R_i \,\text{Exp}(\omega_i)$ onto SO(3) via 4-step Newton–Schulz or signed SVD fallback ($\det = +1$). Second, `physics_trunk` (468,284 params) executes one biased RNA self-attention layer (66,308), three bidirectional `CrossModalPhysicsAttention` layers (397,848 total across both directions) modulated by per-head learned scales $\alpha_h B$, and 100-iteration log-domain Sinkhorn transport [published] [12] ($\eta_{\text{OT}} = 0.5$, $c_{\text{unassigned}} = 10.0$, 4,128 projection weights) over pooled $M = \max(1, L//15)$ RNA stem-loop anchors and K capsid pockets. Third, `contact_pair_mlp` (33,025 params) evaluates symmetrized outer-sum contact logits offset by $-\Delta G_{\text{pair}}/RT$, paired with FP32 analytical stacking, WLC step-stretch, and radial wall penalties (E_{tot}). Outside the trunk, `LearnedPlacerNet` (38,209 params) distills exact `subset_dp` conditional marginals using a 64-D pocket encoder, autoregressive `GRUCell(66,64)`, prior bias $s(-U_{tv} + W)$, $\epsilon = 10^{-3}$ uniform mixing, and 512-particle self-normalized importance sampling. Table 12 lists exact channel counts and precision tiers, and Table 13 traces tensor shapes for $L = 128, M = 8, K = 60$.

12.4. Platform Algorithmic Specifications

Algorithms 1–7 specify the core mathematical and neural routines executed across symmetry reduction, physical pre-computation, deep co-folding, exact dynamic programming, and neural importance sampling in `rnav` [architecture].

Table 12: RNAv hyperparameter, channel, and three-tier mixed-precision specification (`rnav/model/train.py` and trunk modules). [architecture, measured]

Component / Quantity	Value / Shape	Precision Tier	Mathematical Constraint
Model width, layers, heads	$d_{\text{model}} = 128$, 3, 4 ($d_h = 32$)	BF16 linear autocast / FP32 master	LayerNorm Post-LN after attention and FFN ($d_{\text{ff}} = 256$)
Genomic tokenizer	$k = 4$ BPE, vocab 257, chunk 512	Integer token IDs	32-token window overlap prevents boundary truncation
`InvariantStructure-Projection`	$23 \to 128$, 16 RBF (0–200 Å)	FP32 geometric statistics	χ_{moment} distinguishes D/L mirror enantiomers
`SE3FrameUpdateLayer`	13-D edge, $128 \to 64 \to 3$ heads	FP32 rotations ($\det = +1$)	NS-4 when $\min \det > 0.5 \wedge \text{err} < 0.25$; SVD otherwise
`DifferentiableSinkhorn-Layer`	100 iters, $\eta_{\text{OT}} = 0.5$, slack 10.0	FP32 log-domain potentials	Pool $M = \max(1, L//15)$ anchors; $0.1\, d_{\text{3D}}$ penalty
Physical loss	$\mathcal{L}_{\text{cmap}} + 10^{-3}(E_{\text{tot}} + H_{\text{row}})$	FP32 loss, AdamW ($10^{-3}, \lambda_{\text{wd}} = 10^{-4}$)	No `GradScaler` in BF16; global grad norm clip 1.0
3-tier precision policy	Tier 1: FP64 / Tier 2: FP32 / Tier 3: BF16	`PrecisionConfig` (`train.py`)	Note Section 12.5 regarding plan deployment Int8 labels

Table 13: End-to-end software stages, module ownership, precision, and exact data result at each step ($L = 128, M = 8, K = 60$). [architecture, measured]

Stage / Module	Input Stream	Software Operation	Exact Result Produced
Stage 0: Prompt intake	`PromptItem` (`SequencePrompt` / `StructurePrompt`), mode (a)–(c)	Validates nucleotide/coat strings, optional $(N, 3)$ coords/frames, builds 60-subunit shell	Aligned $L = L_r$ RNA tokens, $K = 60$ pocket frames, spiral init $\mathbf{x}_0, T_0$ ($\det R_0 = +1$)
Stage 1: Encode + Frames	Token IDs + cached PLM/GLM or `zero_shot`	`MultiModalFeatureExtractor` + `InvariantStructureProjection` (16-bin RBF, χ_{moment})	Latent states $\mathbf{h}_0$ $(128, 128)$, capsid $(60, 128)$, rigid frames $T_0 = (R_0, t_0)$, $\det R_0 = +1$
Stage 2: Chiral SE(3) $\times 3$	$\mathbf{h}_0$, $\mathbf{x}_0$, and rigid frames T_0	3-layer `SE3FrameTrunk`; `rodrigues_exp_map`, NS-4, signed SVD	Updated features $\mathbf{h}$ $(128, 128)$, chiral coords $\mathbf{x}$ $(128, 3)$, frames $(128, 3, 3)$
Stage 3: Physics + Sinkhorn	RNA $\mathbf{h}$, capsid $(60, 128)$, coords	`PhysicsDistillationTrunk` (self-attn WLC/SHAPE B_{ij} + 2 dir. $\times$ 3 Debye–Hückel cross-attn) + Sinkhorn	Conditioned $(\mathbf{h}, \mathbf{c})$, intra-bias B_{ij} $(128, 128)$, pocket transport plan C_M $(8, 60)$
Stage 4: Pair + Energy heads	Conditioned $\mathbf{h}$, $\mathbf{x}$, C_M, coords	Outer-sum symmetrized `contact_pair_mlp` + FP32 $e_{\text{stack}} + e_{\text{wlc}} + e_{\text{wall}} - 150$	`FoldingOutput`: $P \in [0, 0.99]$ $(128, 128)$, C_M $(8, 60)$, $\mathbf{x}$ $(128, 3)$, $E_{\text{tot}} = -6783.71$, metadata
Companion cut (decoupled)	Capsid ASU, pocket graph, unary U_{tv}, W	`QuotientGraph` (Ω_0, E_0); `phase_boundary` (Donnan); log `subset_dp`; `LearnedPlacerNet` GRU	Exact $\ln Z$ & marginals (10^{-14}), G-orbit shell, 512-particle SNIS ESS/N

Algorithm 1. Icosahedral Symmetry Quotient Projection and Lattice Expansion.

Input: T-chain ASU structure $X_0 \in \mathbb{R}^{n_0 \times 3}$, contact precision A, ridge $\epsilon = 10^{-4}$, seed frame F.
1: Generate proper rotations $\{R_g\}_{g=1}^{60} \subset \mathrm{SO}(3)$ via BFS closure of VIPER 5-fold and 2-fold generators.
2: Expand E_0: place ASU atom a on orbit copy (g, a) by $X_{g,a} \leftarrow R_g X_{0,a} + t_g$ (preserving conformers).
3: Build quotient adjacency A across intra-ASU and symmetry-mate contacts (c = soft-wall precision).
4: Factor Schur complement $(A + cE_0E_0^\top)$ via FP64 Cholesky; set $\Omega_0 \leftarrow (A + cE_0E_0^\top)^{-1} - c^{-1}E_0E_0^\top$.
5: Sample symmetry breaking via `CosetInversionKernel`: $z_g \leftarrow \mathrm{MLP}([\mathrm{Emb}(g), \mathrm{vec}(\mathbf{h}_0), x_{\mathrm{anchor}}])$.
Return: Chiral $60T$ virion coordinates, quotient precision Ω_0, and orbit-equivariant features $F_{g,a}$.

Algorithm 2. Cancellation-Free Donnan, Taylor WLC, and Blended SHAPE Bias.

Input: Fixed coat charge ρ_f, salt c_s, contour L_{ij}, $\ell_p = 10$ Å, reactivities $(r^{\mathrm{free}}, r^{\mathrm{bound}}, \theta)$, seq s.
1: Compute Donnan density $c_\Sigma \leftarrow \mathrm{hypot}(\rho_f, 2c_s)$ and potential $\eta \leftarrow \mathrm{asinh}(\rho_f/(2c_s))$ in FP64.
2: Compute $x \leftarrow L_{ij}/\ell_p$; evaluate 8th-order Taylor series if $x < 0.05$ (FP64) or 0.5 (FP32), else $2\ell_p(L_{ij} + \ell_p \mathrm{expm1}(-x))$.
3: Blend occupancies linearly: $r_{\mathrm{eff}} \leftarrow (1-\theta)\max(0, r^{\mathrm{free}}) + \theta \max(0, r^{\mathrm{bound}})$ (sentinel $\le -500 \to 0$).
4: Map Deigan energy [25]: $g_i \leftarrow 1.8 \ln(r_{\mathrm{eff},i} + 1) - 0.6$ (2× weight internally stacked, 1× terminal).
5: Sum canonical Watson–Crick energies ($|i-j| \ge 4$) and WLC linker log-weight divided by $RT = 0.6163$ kcal/mol.
Return: Intra-RNA pair bias matrix $B_{\mathrm{rna}} \in \mathbb{R}^{L \times L}$ (clamped at -10^4) and screened ion state.

Algorithm 3. Multi-Modal Forward Pass (`RNAvFoldingModel`).

Input: `PromptItem` pair (Mode a: seq+seq, Mode b: seq+struct, Mode c: struct+struct), optional adapters.
1: Encode RNA ($L \times 40 \to 128$) and capsid ($K \times 59 \to 128$) via `if/else` selection of GLM/ESM2 adapters or zero-shot extractors.
2: Initialize spiral $\mathbf{x}_0 \in \mathbb{R}^{L \times 3}$, $R_0 = I_3$; compute 23-D RBF + χ_{moment} + $(\cos\tau, \sin\tau)$ features via `struct_proj`.
3: If conditioning gate > 0.05, inject $\mathbf{h}_{\mathrm{rna}} \leftarrow \mathbf{h}_{\mathrm{rna}} + \mathbf{h}_{\mathrm{struct}}$ in 128-D hidden space.
4: Run 3 chiral `SE3FrameUpdateLayer` blocks (**Alg. 4**), updating $(\mathbf{h}_{\mathrm{rna}}, R_i, t_i)$ and coords $\mathbf{x}$.
5: Run self-attention + 3× bidirectional cross-attention + log-Sinkhorn (**Alg. 5**) to obtain $\Gamma_{M \times K}$.
6: Compute symmetrized $P_{ij} = \sigma(\frac{1}{2}(Z_{ij} + Z_{ji}) - \Delta G_{ij}/RT)$ and FP32 free energy E_{tot}.
Return: `FoldingOutput`$(P, \Gamma, \mathbf{x}, E_{\mathrm{tot}}, \mathrm{metadata})$.

Algorithm 4. Chirality-Strict SE(3) **Frame Update and Homochiral Projection.**

Input: Node features $\mathbf{h} \in \mathbb{R}^{L \times 128}$, proper frames $T_i = (R_i, t_i)$ in FP32.
1: Rotate pairwise offsets into local frames: $u_{ij} \leftarrow R_i^\top (t_j - t_i)$; form 13-D invariant vector $\mathbf{g}_{ij}$.
2: Evaluate edge biases $b_{ij}^{(h)} \leftarrow$ `edge_mlp`$(\mathbf{g}_{ij})$ $(\mathbb{R}^{13} \to \mathbb{R}^{32} \to \mathbb{R}^4)$ in FP32.
3: Compute 4-head context $\mathbf{c} \leftarrow \mathrm{SDPA}(\mathbf{h}, \mathbf{b})$ and update $\mathbf{h} \leftarrow \mathbf{h} +$ `node_mlp`$(\mathbf{c})$ $(128 \to 256 \to 128)$.
4: Predict local translation Δt_i and axis-angle ω_i $(128 \to 64 \to 3)$; update $t_i' \leftarrow t_i + R_i \Delta t_i$.
5: Map $R_{\mathrm{cand}} \leftarrow R_i \,\mathrm{Exp}(\omega_i)$ (Taylor series if $\|\omega_i\| < 10^{-3}$).
6: Project $R_i' \leftarrow$ `project_to_so3`(R_{cand}) via 4-step Newton–Schulz or signed SVD ($\det R_i' = +1$).
Return: Updated features $\mathbf{h}$ and strictly proper frames (R_i', t_i') ($\det R_i' = +1$ verified by gate).

Algorithm 5. Physics-Biased Cross-Attention and Log-Domain Sinkhorn Transport.

Input: $\mathbf{h}_{\text{rna}}$, $\mathbf{h}_{\text{cap}}$, pairwise coords $(\mathbf{x}, \mathbf{y})$, pair biases $B_{\text{rna}}, B_{\text{cross}}$.
1: Apply Post-LN self-attention on $\mathbf{h}_{\text{rna}}$ with learned per-head scaling $\alpha_h B_{\text{rna}}$.
2: Compute FP32 Debye–Hückel potential matrix $\Phi_{ik} = (\ell_B / r_{ik}) \exp(-r_{ik}/\lambda_D)/RT$ ($\lambda_D = 9.61$ Å).
3: For $\ell = 1 \dots 3$, update $\mathbf{h}_{\text{rna}} \leftarrow \text{CrossAttn}(\mathbf{h}_{\text{rna}}, \mathbf{h}_{\text{cap}}; \Phi)$ and symmetrically $\mathbf{h}_{\text{cap}} \leftarrow \text{CrossAttn}(\mathbf{h}_{\text{cap}}, \mathbf{h}_{\text{rna}}; \Phi^\top)$.
4: Segment-pool $\mathbf{h}_{\text{rna}}$ into $M = \max(1, L//15)$ anchors; project anchors & pockets $128 \to 16$.
5: Form transport cost $C_{mk} = \|\mathbf{a}_m - \mathbf{p}_k\|_2 + 0.1\,\|\bar{\mathbf{x}}_m - \mathbf{y}_k\|_2$ with slack cost 10.0.
6: Iterate 100 log-Sinkhorn updates at $\eta_{\text{OT}} = 0.5$: $u \leftarrow -\eta_{\text{OT}} \text{LSE}_k(-C/\eta_{\text{OT}} + v)$, $v \leftarrow -\eta_{\text{OT}} \text{LSE}_m(-C/\eta_{\text{OT}} + u)$.
Return: Conditioned $(\mathbf{h}_{\text{rna}}, \mathbf{h}_{\text{cap}})$ and doubly-sub-stochastic pocket matrix $\Gamma \in [0,1]^{M \times K}$.

Algorithm 6. Exact Injective Bitmask Subset DP with State-Cap Abstention.

Input: Anchors M, pockets K, unary potentials $U \in \mathbb{R}^{M \times K}$, WLC pair tensor W, cap $N_{\text{cap}} = 5 \times 10^6$.
1: Count exact injective states $N_{\text{states}} = \sum_{t=1}^{M} \binom{K}{t} \cdot t$; if $N_{\text{states}} > N_{\text{cap}}$, raise `SubsetDPAbstentionError`.
2: Initialize FP64 log-forward table $\alpha_0(v, 2^v) \leftarrow -U_{0,v}$ for all pockets $v \in \{0, \dots, K-1\}$.
3: For step $t = 1 \dots M-1$, transition $v \notin S$ into mask $S' = S \cup \{v\}$:
$\alpha_t(v, S') \leftarrow \text{LSE}_{v' \in S}\big[\alpha_{t-1}(v', S) + W_{t-1,v',v}\big] - U_{t,v}$.
4: Compute exact log-partition $\ln Z \leftarrow \text{LSE}_{v,S}[\alpha_{M-1}(v, S)]$ in FP64 (to 10^{-14} tolerance).
5: Run symmetric backward recursion $\beta_t(v, S)$ and accumulate exact marginals $\mu_t(v) = \exp(\alpha_t + \beta_t - \ln Z)$.
Return: Exact partition $\ln Z$, marginal matrix μ, and conditional distributions `prefix_conditional`.

Algorithm 7. Teacher-Forced GRU Placer Distillation and 512-Particle SNIS.

Input: Exact DP teacher conditionals p_{DP}, pocket descriptors $\mathbf{e}_k$, unary/pair (U, W), sample budget $N = 512$.
1: Encode pockets $\tilde{\mathbf{e}}_k \leftarrow \texttt{pocket_enc}(\mathbf{e}_k)$ ($\mathbb{R}^2 \to \mathbb{R}^{64}$) and key vectors $\mathbf{k}_k \leftarrow \texttt{key_proj}(\tilde{\mathbf{e}}_k)$.
2: During distillation, teacher-force anchor t: compute logits $\ell_{t,k} = \frac{1}{8}\mathbf{q}_t^\top \mathbf{k}_k + s(-U_{t,k} + W_{t-1,v_{t-1},k})$ (masked on visited S).
3: Minimize masked KL/NLL against exact `prefix_conditional` marginals using AdamW in FP32 (38,209 params).
4: At inference, sample $N = 512$ injective trajectories from defensive mixture $\tilde{\pi}_\theta = (1-\epsilon)\pi_\theta + \epsilon\,\text{Unif}_{\setminus S}$ ($\epsilon = 10^{-3}$).
5: Evaluate exact log importance weights $\log w^{(i)} = \log p_{\text{phys}}(\mathbf{v}^{(i)}) - \log \tilde{\pi}_\theta(\mathbf{v}^{(i)})$; normalize via log-sum-exp.
Return: $\widehat{\ln Z}$ estimate, SNIS-weighted pocket marginals, and Kish effective sample size ESS/N [30].

12.5. Parameter and Codebase Audit

Table 14 disaggregates trainable weights across every PyTorch module and separates deterministic mathematical oracles from components profiled locally or scheduled for cluster execution. To confirm whether prior codebase and manuscript inconsistencies were resolved, all five Iteration-1 blocker defects in `GATE_STATUS.md` passed the 15.XI.2026 regression suite across 171 Python unit tests, 60 E2E scenarios, numerical physics stress tests, and 138 Rust workspace tests, plus a 60,000-sample `CosetInversionKernel` audit resolving the unseeded `adversarial_m5_harness.py` Pearson variance **[measured]**. Table 15 records every fixed blocker alongside seven residual non-blocking discrepancies observed directly in the repository at Gate G0.

All deterministic software suites above remain reproducible from the repository root using these PowerShell commands (verified 15.XI.2026; note the unseeded M5 variance in Table 15):

```
$env:PYTHONPATH="RNAv-live\RNAv_Platform\03_Platform\code;."
python -m pytest RNAv-live/RNAv_Platform/03_Platform/code/tests -q
python -m pytest tests/e2e -q
python tests/adversarial_m5_harness.py
python tests/stress_physics_numerical.py
cargo test --manifest-path RNAv-live/RNAv_Platform/10_App/Cargo.toml --workspace --quiet
```

Table 14: Measured submodule parameter breakdown (`rnav/model/folding_model.py` and `train.py`, 15.XI.2026). [measured, target]

Component / Module	Parameters	Share / Prec.	Training Scope & Current Empirical State
`feature_extractor`	606,082	39.59% / BF16	`esm_adapter` (361,729), `glm_adapter` (197,889), zero-shot heads (46,464); upstream PLM/GLM frozen; adapters at random initialization
`struct_proj`	19,840	1.30% / FP32	Pairwise RBF + chirality coordinate projection prior to SE(3) trunk (Stage 1 active zero-shot subtotal 66,304; total w/ adapters 625,922); random initialization
`se3_trunk` ($3 \times$ SE(3) update)	403,710	26.37% / FP32 rot	Rotation updates fixed orthogonal in FP32; linear weights at random initialization; profiled 50 BF16 steps on RTX 4060
`physics_trunk` + Sinkhorn	468,284	30.59% / FP32 bias	2 directions $\times$ 3 cross-attn (198,924 ea. dir.), RNA self-attn (66,308), anchor/pocket proj (4,128); Sinkhorn 0 weights; random initialization
`contact_pair_mlp`	33,025	2.16% / FP32	Symmetric pairwise contact head; random initialization (synthetic $-4.29 \rightarrow -4.47$ profile check only)
`RNAvFoldingModel` total	1,530,941	971,323 active	Zero-shot default (971,323) plus optional adapters (+559,618); random initialization at Gate G0; domain training is planned for Tiers A–C (58 days Slurm)
`LearnedPlacerNet` (cut)	38,209	Standalone GRU	`pocket_enc` (4,480), `step_gru` (25,344), `query_proj` (4,160), `key_proj` (4,160); 50-step synthetic $(4, 12)$ smoke run passed
Analytic & Rust oracle stack	0 learnable	Exact / FP32	`QuotientGraph`, WLC, Donnan, Wyman, `subset_dp`, Rust `10_App`; deterministic math with no learned weights

Table 15: Resolution of prior Iteration-1 blockers and open residual codebase/document inconsistencies (verified 15.XI.2026). [measured]

Defect / Item	Status	Code / Document Evidence and Exact Scope
Icosahedral triangulation ($T \geq 16$)	Fixed	`rnav/symmetry/triangulation.py`: `_build_ico_grid` and free 60-element I-orbit pass $T \in \{13, 16, 25, 49\}$ and 3,600/3,600 closure
Citardauq / asinh Donnan stability	Fixed	`rnav/physics/electrostatics.py`: overflow-free `math.hypot`; sweeps reach max rel err $= 1.42 \times 10^{-16}$
WLC Taylor regime	Fixed	`rnav/physics/wlc.py`: raises `ValueError` on $\ell_c \leq 0$; 8th-order Taylor matches FP64 reference to 8.52×10^{-15}
Exact log-DP & forbidden pockets	Fixed	`rnav/matching.py` (`subset_dp`, `SubsetDPAbstentionError`): agrees with brute force to 10^{-14}; 0.166 ms abstention above 5×10^6 states
Chiral SE(3) frame & reflection gate	Fixed	`rnav/model/se3_frames.py` (`SE3FrameTrunk`, `project_to_so3`, `assert_homochiral_frames`): 100/100 equivariance trials (max 2.38×10^{-7}); 50/50 improper reflections rejected
Cross-attention λ_D / temperature	Open (residual)	`rnav/model/cross_attention.py:252--254` hardcodes $\lambda_D = 9.61$ Å (100 mM) and divides ℓ_B/d by $RT = 0.6163$ (310.15 K) without explicit $z_i z_j$, versus 7.85 Å (150 mM) and 298.15 K (0.592485 kcal/mol) in M3/CAM checks
Precision tier numbering drift	Open (docs/code)	`rnav/model/train.py:4--6` numbers Tiers 1–3 as FP64/FP32/BF16, whereas `TRAINING_AND_CODE_PLAN.md:62--66` and §7 number Tiers 1–3 as FP32/BF16/Int8 ONNX
E2E `contracts.py` shadow math	Open (residual)	`tests/e2e/contracts.py:163--251` retains standalone `wlc_mean_square` ($\ell \leq 0 \to 0.0$ vs. `ValueError`) and `test_tier1:443,456` stub shape hints $(100, 100)/(40, 180)$, while stub models and Donnan delegate to `rnav`
Unseeded M5 Curie chi-square variance	Open (test)	`tests/adversarial_m5_harness.py:318--340` runs 4 unseeded $\alpha = 0.05$ tests (18.5% false rejection risk; one run hit $p = 0.0039$ at $N = 3{,}000$), while $N = 60{,}000$ direct `CosetInversionKernel` audit verifies $\chi^2 = 60.28$, $p = 0.4292$
Trunk vs. companion cut split	Open (design)	`RNAvFoldingModel.forward()` uses dense (L, L) attention ($L \leq 512$); `QuotientGraph`, `phase_boundary`, and `subset_dp` run outside `forward()`
M0–M7 & doc count drift	Open (docs)	Table 3 renumbers archive milestones cleanly (M1 quotient/M3 Donnan vs. archive M2/M4); `PROJECT.md` still lists M5 planned / 165 unit tests instead of measured 171 Python + 138 Rust tests
Thesis & bibliography handoff	Open (docs)	`06_Reference/citation_audit/AUDIT_SUMMARY.md` flags 16 closed bibliography PDFs lacking full text and leaves `07_Thesis` unmodified outside this technical report

13. Evidence Boundaries and Reproducibility Ledger

Following Harness[2] recursive verification guidelines [published] [60], Table 16 pins the verification state of every deliverable at Gate G0 so downstream cluster campaigns preserve clean separation between implemented software and future training outcomes.

Table 16: Verification provenance across codebase, visual figures, and deliverables (ledger state 15.XI.2026). [measured]

Artifact / Suite	Measured Result	Verification Basis & Scope Boundary
Python `rnav`	171 unit tests; 60 E2E scenarios	Matches VIPERdb matrices to 7.4×10^{-6}, exact `subset_dp` to 10^{-14}, and regenerates `results.json`
Rust Workspace (`10_App`)	138 workspace tests pass	Agrees with independent `oracle_sol.rs` answer key to 10^{-6}
RTX 4060 Hardware Profile	$L \in \{64, 128, 256, 512\}$, 50-step BF16	Recorded 14.XI.2026 in `rtx4060_benchmarks.json`; Tier B/C kilobase scaling is planned on Slurm
M5 Stability Pre-Flight	Compartment neutrality tests pass	Enforces Donnan balance numerically; external GROMACS/OpenMM runs are specified for Gate G6
Molecullage Figures (6)	6/6 strict `qa_figure` ok + blind-read PASS	ChimeraX + Molecullage MCP Chromium + Gemini Pro / Claude blind reads & Opus 5.5 checks (15.XI.2026); Arial $\geq$ 6.55 pt, vector PDFs

14. Conclusions

RNAv combines four verified mathematical building blocks—group-quotient protein attention $(\Pi_0, E_0, \Omega_0, F_{g,a})$, chirality-strict SE(3) frame updates with $\det R_i = +1$ and χ_{det}, cancellation-free polymer and Donnan electrostatics feeding M5 ion-compartment gating, and exact log-domain `subset_dp` distilling SNIS-calibrated `LearnedPlacerNet`—inside one cut-posterior M0–M7 architecture tested locally on consumer RTX 4060 hardware. The platform targets verifiable physical observables, explicit statistical abstention, and synthetic nanocage, VLP, and cryo-EM prior workflows where symmetry, cargo charge, and packaging-signal spacing are experimental control knobs. The platform does not regress whole unprobed viral genomes in 3D; the 96-entry PDB archive limits such regression.

References

[1] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. doi:10.1016/j.softx.2015.06.001.

[2] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016), 493–500. doi:10.1038/s41586-024-07487-w.

[3] Asor, R., Loewenthal, D., Melnyk, D., Tan, T. K., and Kukura, P. (2026). Molecular-level observation of the self-assembly of a virus-like particle. *Nature*, 657(8132), 653–660. doi:10.1038/s41586-026-10948-z.

[4] Baek, M. et al. (2023). Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nature Methods*. doi:10.1038/s41592-023-02086-5.

[5] Banskota, S., Raguram, A., Liu, D. R., et al. (2022). Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell*. doi:10.1016/j.cell.2021.12.021.

[6] Belyi, V. A. and Muthukumar, M. (2006). Electrostatic origin of the genome packing in viruses. *Proceedings of the National Academy of Sciences USA*, 103(46), 17174–17178. doi:10.1073/pnas.0608311103.

[7] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242. doi:10.1093/nar/28.1.235.

[8] Butterfield, G. L., Lajoie, M. J., Baker, D., King, N. P., et al. (2017). Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature*. doi:10.1038/nature25157.

[9] Cadena-Nava, R. D., Comas-Garcia, M., Garmann, R. F., Rao, A. L. N., Knobler, C. M., and Gelbart, W. M. (2012). Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. *Journal of Virology*, 86(6), 3318–3326. doi:10.1128/JVI.06566-11.

[10] Cadena-Nava, R. D., Comas-Garcia, M., Garmann, R. F., Rao, A. L. N., Knobler, C. M., and Gelbart, W. M. (2019). Correction for Cadena-Nava et al., Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes. *Journal of Virology*, 93(4), e01939-18. doi:10.1128/JVI.01939-18.

[11] Chen, J., Hu, Z., Sun, S., Tan, Q., Wang, Y., Yu, Q., et al. (2022). Interpretable RNA Foundation Model from Unannotated Data for Highly Accurate RNA Structure and Function Predictions. *arXiv preprint*. arXiv:2204.00300.

[12] Cuturi, M. (2013). Sinkhorn distances: lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems 26 (NeurIPS 2013)*. arXiv:1306.0895.

[13] Dai, X., Li, Z., Lai, M., et al. (2017). In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature*, 541(7635), 112–116. doi:10.1038/nature20589.

[14] Dalla-Torre, H., Gonzalez, L., Mendoza-Revilla, J., Lopez Carranza, N., Grzywaczewski, A. H., Oteri, F., et al. (2025). Nucleotide Transformer: building and evaluating robust foundation models for human genomics. *Nature Methods*, 22(2), 287–297. doi:10.1038/s41592-024-02523-z.

[15] Debye, P. and Hückel, E. (1923). Zur Theorie der Elektrolyte. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Physikalische Zeitschrift*, 24(9), 185–206.

[16] Dykeman, E. C., Stockley, P. G., and Twarock, R. (2013). Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *Journal of Molecular Biology*, 425(17), 3235–3249. doi:10.1016/j.jmb.2013.06.005.

[17] Dykeman, E. C., Stockley, P. G., and Twarock, R. (2014). Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proceedings of the National Academy of Sciences USA*, 111(14), 5361–5366. doi:10.1073/pnas.1319479111.

[18] Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., et al. (2017). OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7), e1005659. doi:10.1371/journal.pcbi.1005659.

[19] Farafonov, V. S. and Nerukh, D. (2019). MS2 bacteriophage capsid studied using all-atom molecular dynamics. *Interface Focus*, 9(3), 20180081. doi:10.1098/rsfs.2018.0081.

[20] Farafonov, V. S., Stich, M., and Nerukh, D. (2022). Reconstruction and validation of entire virus model with complete genome from mixed resolution cryo-EM density. *Faraday Discussions*, 240, 152–167. doi:10.1039/D2FD00053A.

[21] Farafonov, V. S., Stich, M., and Nerukh, D. A. (2023). Complete virion simulated: all-atom model of an MS2 bacteriophage with native genome. *Journal of Chemical Theory and Computation*, 19(21), 7924–7933. doi:10.1021/acs.jctc.3c00846.

[22] Garmann, R. F., Comas-Garcia, M., Koay, M. S. T., et al. (2014). Role of electrostatics in the assembly pathway of a single-stranded RNA virus. *Journal of Virology*, 88(18), 10472–10479. doi:10.1128/JVI.01044-14.

[23] Garmann, R. F., Goldfain, A. M., and Manoharan, V. N. (2019). Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome. *Proceedings of the National Academy of Sciences USA*, 116(45), 22485–22490. doi:10.1073/pnas.1909223116.

[24] Hadden, J. A., Perilla, J. R., Schlicksup, C. J., Venkatakrishnan, B., Zlotnick, A., and Schulten, K. (2018). All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits. *eLife*, 7, e32478. doi:10.7554/eLife.32478.

[25] Hajdin, C. E. et al. (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proceedings of the National Academy of Sciences USA*. doi:10.1073/pnas.1219988110.

[26] Helgstrand, C., Wikoff, W. R., Duda, R. L., Hendrix, R. W., Johnson, J. E., and Liljas, L. (2003). The Refined Structure of a Protein Catenane: The Hk97 Bacteriophage Capsid at 3.44Å Resolution. *Journal of Molecular Biology*, 334, 885. doi:10.1016/J.JMB.2003.09.035.

[27] Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. doi:10.1038/s41586-021-03819-2.

[28] Kagaya et al. (2025). Structure Modeling Protocols for Protein Multimer and RNA in CASP16 With Enhanced MSAs, Model Ranking, and Deep Learning. *Proteins: Structure, Function, and Bioinformatics*. doi:10.1002/prot.70033.

[29] Khayat, R., Brunn, N., Speir, J. A., et al. (2011). The 2.3-angstrom structure of porcine circovirus 2. *Journal of Virology*, 85(15), 7856–7862. doi:10.1128/JVI.00737-11.

[30] Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.

[31] Klumpp, K., Lam, A. M., Lukacs, C., Vogel, R., Ren, S., Espiritu, C., Baydo, R., Atkins, K., Abendroth, J., Liao, G., Efimov, A., Hartman, G., and Flores, O. A. (2015). High-resolution crystal structure of a hepatitis B virus replication inhibitor bound to the viral core protein. *Proceedings of the National Academy of Sciences USA*, 112, 15196–15201. doi:10.1073/pnas.1513803112.

[32] Kratky, O. and Porod, G. (1949). Röntgenuntersuchung gelöster Fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas*, 68(12), 1106–1122. doi:10.1002/recl.19490681203.

[33] La et al. (2025). MegaFold: Efficient Training of Next-Generation 3D Attention Protein Models on Cross-Platform GPUs. *arXiv preprint*. arXiv:2506.20686.

[34] Lewis, S., Hempel, T., Jiménez-Luna, J., et al. (2025). Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389(6761), eadv9817. doi:10.1126/science.adv9817.

[35] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637), 1123–1130. doi:10.1126/science.ade2574.

[36] Lin et al. (2026). Fold-CP: A Context Parallelism Framework for Biomolecular Modeling. *arXiv preprint*. arXiv:2603.14806.

[37] Liu, F., Bayarri, M. J., and Berger, J. O. (2009). Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1), 119–150. doi:10.1214/09-BA404.

[38] Mattiotti, G., Micheloni, M., Petrolli, L., Rovigatti, L., Tubiana, L., Pasquali, S., and Potestio, R. (2024). Molecular dynamics characterization of the free and encapsidated RNA2 of CCMV with the oxRNA model. *Macromolecular Rapid Communications*, 45(24), e2400639. doi:10.1002/marc.202400639.

[39] Mohajerani, F., Tyukodi, B., Schlicksup, C. J., Hadden-Perilla, J. A., Zlotnick, A., and Hagan, M. F. (2022). Multiscale modeling of hepatitis B virus capsid assembly and its dimorphism. *ACS Nano*, 16(9), 13845–13859. doi:10.1021/acsnano.2c02119.

[40] Ouyang-Zhang et al. (2025). Triangle Multiplication Is All You Need For Biomolecular Structure Representations. *arXiv preprint*. arXiv:2510.18870.

[41] Peralta, B., Gil-Carton, D., Castaño-Díez, D., Bertin, A., Boulogne, C., Oksanen, H. M., Bamford, D. H., and Abrescia, N. G. A. (2013). Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLOS Biology*, 11(9), e1001667. doi:10.1371/journal.pbio.1001667.

[42] Perlmutter, J. D., Perkett, M. R., and Hagan, M. F. (2014). Pathways for virus assembly around nucleic acids. *Journal of Molecular Biology*, 426(18), 3148–3165. doi:10.1016/j.jmb.2014.07.004.

[43] Perlmutter, J. D. and Hagan, M. F. (2015). Mechanisms of virus assembly. *Annual Review of Physical Chemistry*, 66, 217–239. doi:10.1146/annurev-physchem-040214-121637.

[44] Plummer, M. (2015). Cuts in Bayesian graphical models. *Statistics and Computing*, 25(1), 37–43. doi:10.1007/s11222-014-9503-z.

[45] Reddy, V. S., Natarajan, P., Okerberg, B., et al. (2001). Virus Particle Explorer (VIPER), a website for virus capsid structures and their computational analyses. *Journal of Virology*, 75(24), 11943–11947. doi:10.1128/JVI.75.24.11943-11947.2001.

[46] Rojas Labra et al. (2026). FoldaVirus, a knowledge-based icosahedral capsid prediction tool using AlphaFold. *Proceedings of the National Academy of Sciences USA*. doi:10.1073/pnas.2613741123.

[47] Satorras, V. G., Hoogeboom, E., and Welling, M. (2021). E(n) Equivariant Graph Neural Networks. *arXiv preprint*. arXiv:2102.09844.

[48] Segel, M., Lash, B., Zhang, F., et al. (2021). Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudotyped for mRNA delivery. *Science*. doi:10.1126/science.abg6155.

[49] Šiber, A., Božič, A. L., and Podgornik, R. (2012). Energies and pressures in viruses: contribution of nonspecific electrostatic interactions. *Physical Chemistry Chemical Physics*, 14(11), 3746–3765. doi:10.1039/c1cp22756d.

[50] Simpson, A. A., Tao, Y., Leiman, P. G., Badasso, M. O., He, Y., Jardine, P. J., Olson, N. H., Morais, M. C., Grimes, S., Anderson, D. L., Baker, T. S., and Rossmann, M. G. (2000). Structure of the bacteriophage phi29 DNA packaging motor. *Nature*, 408, 745–750. doi:10.1038/35047129.

[51] Sun, Z., El Omari, K., Sun, X., Ilca, S. L., Kotecha, A., Stuart, D. I., Poranen, M. M., and Huiskonen, J. T. (2017). Double-stranded RNA virus outer shell assembly by bona fide domain-swapping. *Nature Communications*, 8, 14814. doi:10.1038/ncomms14814.

[52] Tarasova, E., Farafonov, V., Khayat, R., et al. (2017). All-atom molecular dynamics simulations of entire virus capsid reveal the role of ion distribution in capsid's stability. *Journal of Physical Chemistry Letters*, 8(4), 779–784. doi:10.1021/acs.jpclett.6b02759.

[53] Tarasova, E., Okimoto, N., Feng, S., et al. (2021). Constant pH molecular dynamics of porcine circovirus 2 capsid protein reveals a mechanism for capsid assembly. *Physical Chemistry Chemical Physics*, 23(43), 24617–24626. doi:10.1039/d1cp02874j.

[54] Terasaka, N., Azuma, Y., and Hilvert, D. (2018). Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proceedings of the National Academy of Sciences USA*. doi:10.1073/pnas.1800527115.

[55] Ting, C. L., Wu, J., and Wang, Z.-G. (2011). Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. *Proceedings of the National Academy of Sciences USA*, 108(41), 16986–16991. doi:10.1073/pnas.1109307108.

[56] Twarock, R. and Luque, A. (2019). Structural puzzles in virology solved with an overarching icosahedral design principle. *Nature Communications*, 10, 4414. doi:10.1038/s41467-019-12367-3.

[57] Weis, F., Beckers, M., von der Hocht, I., and Sachse, C. (2019). Elucidation of the viral disassembly switch of tobacco mosaic virus. *EMBO Reports*, 20(11), e48451. doi:10.15252/embr.201948451.

[58] Wyman, J. (1964). Linked functions and reciprocal effects in hemoglobin: a second look. *Advances in Protein Chemistry*, 19, 223–286. doi:10.1016/s0065-3233(08)60190-4.

[59] Xu et al. (2025). Benchmarking all-atom biomolecular structure prediction with FoldBench. *Nature Communications*. doi:10.1038/s41467-025-67127-3.

[60] Xu, S. et al. (2026). Harness^2: Recursive Agent Harnessing for an Open World. *arXiv preprint*. arXiv:2610.14224.

[61] Yoo, A. B., Jette, M. A., and Grondona, M. (2003). SLURM: Simple Linux Utility for Resource Management. In *Job Scheduling Strategies for Parallel Processing*, Springer, 44–60. doi:10.1007/10968987_3.