

Additional file 1 Supporting Information

Physical Sufficiency and Predictive Identifiability in Amino-Acid Transfer-Energy Representations

Aleksei Novgorodtsev

Independent Researcher, Madrid, Spain

ORCID: 0009-0009-2407-7049 | Correspondence: aleksei.novgorodtsev@gmail.com

This Supporting Information accompanies the main article as Additional file 1. The machine-readable reproducibility archive is Additional file 2. It is self-contained for the reported study. It supplies the finite data ledgers, exact proofs, observer-transfer contracts, predictive-state certificates, nested/null results, physical-predictive bridge, mechanistic gate matrix, adversarial boundaries, prospective contracts, uncertainty status, and claim-boundary ledger used by the manuscript.

S1 Notation and evidence-status vocabulary

- Ω_{mol} : ambient admissible molecular-state space.
- A20: declared canonical 20-residue study slice.
- E6: six-endpoint declared slice.
- D10: ten-rule response-blind predictive descriptor grammar.
- F : free-energy table.
- P_m : orthogonal centering projector in an m -state axis.
- Γ : phase-by-variant double-difference interaction contrast.
- $q_0 = (n_{CS}, n_{ON}, b_{Born})$.
- $q_1 = (q_0, \beta_{br})$.
- $q_2 = (q_1, s)$ with sulfur presence s .
- H1: endpoint-specific low-dimensional predictive compression.
- H2a: shared predictive class with endpoint-specific coefficients.
- H2b: identification of a unique common structural geometry.

Evidence statuses are **THEOREM**, **CERTIFIED**, **EMPIRICAL**, **CONDITIONAL**, **MODEL**, **NONIDENTIFIABLE**, **NO-GO**, and **OPEN**. These are mutually exclusive evidence labels in the current study.

S1.1 Provenance of the retrospective protocol

The primitive family, descriptor grammar, capacity bound and endpoint campaign set are frozen for audit, not asserted to have been independently preregistered before first response exposure. Response-free evaluation of a descriptor is distinct from response-free discovery of its rule. Fiber refinement is response-informed. H1 and H2 nulls repeat computational selection conditional on the final pipeline; they do not adjust for the preceding human design search. The execution plan for the new checks was saved locally before simulation on 16 September 2026, after the existing study and review. It is an execution record, not prospective preregistration. Historical terms such as “declared” in ledgers refer to this reproducibility contract unless a future untouched experiment is explicitly specified.

S2 Endpoint provenance and exact E6 values

S2.1 Endpoint registry

Table S1. Endpoint sources and observation contracts.

Endpoint	Source / quantity	Unit	Observer/state contract	Role / uncertainty
WW Octanol	Wimley-White; whole-residue water-to-octanol transfer free energy	kcal/mol	WW state convention; neutral His; D/E/K/R charged	Anchor; per-residue errors
WW Interface	Wimley-White; whole-residue water-to-membrane-interface transfer free energy	kcal/mol	Published WW interface convention	Dependent secondary; per-residue errors
Hessa	Hessa et al.; translocon-mediated apparent insertion contribution	kcal/mol	Residue-wise charge-state map not sufficiently source-identified	Retrospective; residue-wise uncertainty unknown
Moon-Fleming	Moon & Fleming; OmpLA-context bilayer transfer/folding contrast	kcal/mol	pH 3.8 operational map: D/E neutral; H/K/R positive	Retrospective; row-level uncertainty audit open
Eisenberg	Eisenberg et al.; normalized consensus hydrophobicity scale	unitless	Derived consensus observable	Context only
Fauchere-Pliska	Fauchere & Pliska; Ac-X-amide partition-derived fragment parameter	unitless	Water/octanol contrast; carrier/state ledger differs from WW	Context only; source-row audit pending

S2.2 Exact endpoint matrix

Table S2. Exact E6 endpoint values.

AA	WW octanol	WW interface	Hessa	Moon Fleming	Eisenberg	Fauchere Pliska
A	0.5	0.17	0.11	0.0	0.62	0.31
C	-0.02	-0.24	-0.13	0.49	0.29	1.54
D	3.64	1.23	3.49	2.95	-0.9	-0.77
E	3.63	2.02	2.68	1.64	-0.74	-0.64
F	-1.71	-1.13	-0.32	-2.2	1.19	1.79
G	1.15	0.01	0.74	1.72	0.48	0.0
H	0.11	0.17	2.06	4.76	-0.4	0.13
I	-1.12	-0.31	-0.6	-1.56	1.38	1.8
K	2.8	0.99	2.71	5.39	-1.5	-0.99
L	-1.25	-0.56	-0.55	-1.81	1.06	1.7
M	-0.67	-0.23	-0.1	-0.76	0.64	1.23
N	0.85	0.42	2.05	3.47	-0.78	-0.6
P	0.14	0.45	2.23	-1.52	0.12	0.72
Q	0.77	0.58	2.36	3.01	-0.85	-0.22
R	1.81	0.81	2.58	3.71	-2.53	-1.01
S	0.46	0.13	0.84	1.83	-0.18	-0.04
T	0.25	0.14	0.52	1.78	-0.05	0.26
V	-0.46	0.07	-0.31	-0.78	1.08	1.22
W	-2.09	-1.85	0.3	-0.38	0.81	2.25
Y	-0.71	-0.94	0.68	-1.09	0.26	0.96

The WW octanol, WW interface, Hessa, and Moon-Fleming values are expressed in kcal/mol in the current source ledgers. Eisenberg and Fauchere-Pliska are dimensionless/context-only scales. No cross-unit coefficient equality is inferred.

S3 Reference q2 physical-state ledger

The finite physical state used in the WW fiber-closure analysis is reproduced below. $b_{Born} = z_{formal}^2 / \sqrt{A_{polar}}$ under the declared state convention, where A_{polar} is the fragment-based topological polar surface area (TPSA, Å²) of the frozen pH-7 side-chain analogue, using the Ertl fragment convention specified in manuscript reference 32. Thus b_{Born} has units Å⁻¹. The machine-readable column Born_TPSA stores this derived b_{Born} value; it does not store raw A_{polar} .

Table S3. Physical state coordinates on A20.

AA	n_{CS}	n_{ON}	b_{Born}	β_{br}	s
A	1.0	0.0	0.0	0	0
C	2.0	0.0	0.0	0	1
D	2.0	2.0	0.157858	0	0
E	3.0	2.0	0.157858	0	0
F	7.0	0.0	0.0	0	0
G	0.0	0.0	0.0	0	0
H	4.0	2.0	0.0	0	0
I	4.0	0.0	0.0	1	0
K	4.0	1.0	0.190209	0	0
L	4.0	0.0	0.0	0	0
M	4.0	0.0	0.0	0	1
N	2.0	2.0	0.0	0	0
P	3.0	0.0	0.0	0	0
Q	3.0	2.0	0.0	0	0
R	4.0	3.0	0.125353	0	0
S	1.0	1.0	0.0	0	0
T	2.0	1.0	0.0	1	0
V	3.0	0.0	0.0	1	0
W	9.0	1.0	0.0	0	0
Y	7.0	1.0	0.0	0	0

Display values of b_{Born} are rounded to six decimal places in this table for readability. The machine-readable ledger PHYSICAL_STATE_LEDGER_Q2.csv retains the full precision used by the exact structural verifier.

S3.1 Fiber ladder

- $q_0 = (n_{CS}, n_{ON}, b_{Born})$; nontrivial fibers $\{I, L, M\}$ and $\{P, V\}$.
- $q_1 = q_0 + \beta_{br}$; only $\{L, M\}$ remains.
- The L / M WWO fiber defect is 0.58 kcal/mol.
- $q_2 = q_1 + s$; no nontrivial A20 fibers.

The refinement rule is response-blind after a failed-pair identity is established. Endpoint magnitudes, historical fitted coefficients, Q^2 , $D \approx 3$, and later mechanistic constants are not selector inputs.

S4 Exact interaction-quotient proof

Let $F \in \mathbb{R}^{m \times n}$ and

$$A = \{a \mathbf{1}_n^T + \mathbf{1}_m b^T\}. \quad (\text{S1})$$

The generating map has a one-dimensional kernel $(c \mathbf{1}_m, -c \mathbf{1}_n)$, so $\dim A = m + n - 1$. Define centering projectors

$$P_m = I_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^T, P_n = I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T. \quad (\text{S2})$$

Then $\Pi_{int}(F) = P_m F P_n$ is an orthogonal projector with image $\mathbf{1}_m^\perp \otimes \mathbf{1}_n^\perp$ and dimension $(m-1)(n-1)$. Its kernel equals A . Hence $\mathbb{R}^{m \times n} / A \cong \text{im}(\Pi_{int})$.

For any rows ϕ, ψ and columns x, y ,

$$D_{\phi\psi;xy}(F) = F_{\phi x} - F_{\phi y} - F_{\psi x} + F_{\psi y} \quad (\text{S3})$$

annihilates A . If all double differences vanish, choose reference row ψ_0 and column y_0 , set $a_\phi = F_{\phi y_0}$ and $b_x = F_{\psi_0 x} - F_{\psi_0 y_0}$, and obtain $F_{\phi x} = a_\phi + b_x$. Thus all double differences vanish if and only if $P_m F P_n = 0$.

For an equilibrium partition function Z , $F = -\beta^{-1} \log Z$ yields the cross-ratio form

$$D_{\phi\psi;xy} = -\beta^{-1} \log \frac{Z_{\phi x} Z_{\psi y}}{Z_{\phi y} Z_{\psi x}}. \quad (\text{S4})$$

This theorem identifies the exact interaction observable, not a microscopic decomposition of Z .

S5 Primitive minimality and rank certificates

Every proper subset of q_2 fails exact WWO point-estimate sufficiency on declared A20. Five one-coordinate deletion witnesses are:

Table S4. Deletion witnesses for primitive minimality.

Deleted coordinate	Collision	Absolute WWO defect (kcal/mol)	Sum of reported errors (kcal/mol)	Conservative interval gap (kcal/mol)
n_{CS}	F/G	2.86	0.22	2.64
n_{ON}	H/L	1.36	0.22	1.14
b_{Born}	E/Q	2.86	0.30	2.56
β_{br}	P/V	0.60	0.22	0.38
sulfur presence	L/M	0.58	0.22	0.36

For all five one-coordinate deletions the reported-error intervals remain non-overlapping under simple summed-error bounds. This is a conservative sensitivity check on the finite witnesses, not a confidence calculation because cross-residue covariance is unavailable.

The rationalized q2 matrix has column rank and affine rank 5. One exact nonzero 5×5 minor using rows C/H/I/L/R is $-391728289503047 / 390625000000000$.

S5.1 Rank nonidentifiability construction

Let $y \in \mathbb{R}^{20}$ be the scalar WWO record. For any $1 \leq r \leq 20$, define $Z_r = [y, e_1, \dots, e_{r-1}]$ and observer $O(z) = z_1$. Then $\text{rank}(Z_r) = r$ on the finite sample while $O(Z_r) = y$ for every r . Therefore the scalar record is compatible with many latent sample ranks and does not identify intrinsic response rank.

S5.2 Direct q2 validation-depth certificate

For $A_{q_2} = [1, q_2]$, exact arithmetic on the supplied decimal ledger gives rank six. All 20 single-row deletions preserve rank. Exactly one of the 190 unordered pairs, Cys/Met, reduces rank to five. The sulfur coordinate has support exactly on that pair. Thus $\text{cospark}(A_{q_2}) = 2$: ordinary LOO passes, outer-LOO/inner-LOO fails. This is a rank certificate for the recorded finite design, not a robustness theorem under uncertainty in the physical coordinates. The all-function reconstruction afforded by finite injectivity must not be confused with exact affine fitting.

The certificate is independent of the predictive capacity bound. Relaxing that bound to admit five coordinates would not remove the failed Cys/Met training domain. The stronger no-factorization claim elsewhere remains restricted to the declared D10 family; it is not a theorem about all possible encodings.

S6 Observer/protocol transfer ledger

Table S5. Observer transfer and coverage.

Endpoint	q2 state	Observer/protocol	Coverage	Boundary
WW Octanol	Transfer	Transfer	5/5	Reference identity
WW Interface	Transfer	Partial transfer	5/5	Same host-guest/state family; target phase changes
Hessa	Not identified	No transfer	4/5 + 1 unknown	Charge-state map missing; Sec61 apparent-insertion observer
Moon-Fleming	No transfer	No transfer	4/5 + 1 incompatible	pH 3.8 charge map and OmpLA folding observer differ
Fauchere-Pliska	Not identified	Partial transfer	4/5 + 1 unknown	Same phase pair/log-ratio layer, different carrier/state ledger
Eisenberg	No transfer	No transfer	label adapter only	Consensus index, no direct physical observer

The transfer map is defined before predictive comparison. PARTIAL_TRANSFER never means full physical identity; NONIDENTIFIABLE does not permit imputation of missing state information.

S7 D10 descriptor ledger and exact semantic quotient

Table S6. D10 descriptor values.

AA	n_{CS}	n_C	n_{ON}	q_{std}	n_{heavy}	β_{br}	q_{signed}	n_{ring}	Aromatic	n_S
A	1	1	0	0.0	1	0	0	0	0	0
C	2	1	0	0.0	2	0	0	0	0	1
D	2	2	2	1.0	4	0	-1	0	0	0
E	3	3	2	1.0	5	0	-1	0	0	0
F	7	7	0	0.0	7	0	0	1	1	0

AA	n_{CS}	n_C	n_{ON}	q_{std}	n_{heavy}	β_{br}	q_{signed}	n_{ring}	Aromatic	n_S
G	0	0	0	0.0	0	0	0	0	0	0
H	4	4	2	0.0	6	0	0	1	1	0
I	4	4	0	0.0	4	1	0	0	0	0
K	4	4	1	1.0	5	0	1	0	0	0
L	4	4	0	0.0	4	0	0	0	0	0
M	4	3	0	0.0	4	0	0	0	0	1
N	2	2	2	0.0	4	0	0	0	0	0
P	3	3	0	0.0	3	0	0	1	0	0
Q	3	3	2	0.0	5	0	0	0	0	0
R	4	4	3	1.0	7	0	1	0	0	0
S	1	1	1	0.0	2	0	0	0	0	0
T	2	2	1	0.0	3	1	0	0	0	0
V	3	3	0	0.0	3	1	0	0	0	0
W	9	9	1	0.0	10	0	0	2	1	0
Y	7	7	1	0.0	8	0	0	1	1	0

The D10 grammar contains exact engineered redundancies. Under affine OLS on A20, 1023 nonempty syntax subsets reduce to 415 semantic column-space classes. Removing both engineered redundancy controls gives 255 syntax subsets and 255 semantic classes. The low-capacity budget $d_{eff} \leq 4$ leaves 262 reference semantic states before the validation-depth gate.

S7.1 Finite injectivity does not identify a unique encoding

As a response-blind negative control, all 1023 nonempty D10 coordinate subsets were exhaustively tested for injectivity on the frozen canonical A20 rows. No subset with fewer than five coordinates is injective. At five coordinates, **16 distinct subsets** are injective. The full list is recorded in reproducibility/predictive_compression_audit/results/D10_MINIMAL_INJECTIVE_ENCODINGS.csv and the corresponding JSON file.

This certificate establishes a useful boundary: finite injectivity can prove that a declared encoding separates the observed panel, but it cannot identify a unique molecular geometry. The physical status of q_2 therefore derives from its declared primitive family and witness-driven refinement path; it is not inferred from the mere fact that five coordinates happen to separate A20.

S8 Validation-depth sufficiency theorem and finite certificates

Let $A = [1, X]$ be full-column-rank and $T(V)$ the training domains produced by validation protocol V . Protocol safety is

$$\text{rank}(A_T) = \text{rank}(A) \forall T \in T(V). \quad (\text{S5})$$

For deletion set D , rank loss after deletion is equivalent to the existence of a nonzero c such that $A c$ has support contained in D . Proof: $\text{rank}(A_{-D}) < \text{rank}(A)$ iff $A_{-D} c = 0$ for some $c \neq 0$; hence $A c$ can be nonzero only on D . The converse is immediate.

Define the standard cospark invariant (manuscript reference 33) by

$$\text{cospark}(A) = \min_{c \neq 0} \|A c\|_0. \quad (\text{S6})$$

The support characterization gives the exact corollary

$$\text{rank}(A_{-D}) = \text{rank}(A) \forall |D| \leq d \Leftrightarrow \text{cospark}(A) > d. \quad (\text{S7})$$

Thus $d_{\text{worst}}(A) = \text{cospark}(A) - 1$ is the maximum number of **arbitrary** row deletions against which affine rank is guaranteed. For the present outer-LOO/inner-LOO protocol, every unordered two-row deletion is generated, so nested safety is exactly equivalent to $\text{cospark}(A) > 2$. The response-blind certificate records 131 of 262 low-capacity reference semantic classes satisfying this condition in `reproducibility/predictive_compression_audit/results/VALIDATION_DEPTH_COSPARK.json`. For structured scaffold/group/cluster/chemotype split families, the general protocol-specific condition over $T(V)$ remains primary; cospark does not replace the typed deletion family.

For outer-LOO/inner-LOO, all unordered two-row deletions are generated. The finite certified contractions are:

Finite class counts.

Table S7. Semantic capacity and validation safety.

Profile	Ambient semantic classes	Low-capacity classes	Nested-safe classes	Maximum A20 rows separated	Safe classes at maximum	Safe injective classes
Reference D10	415	262	131	19	4	0
Observer-safe D8	103	93	47	17	7	0
Moon pH 3.8 D10	207	161	89	17	9	0

Physical-coordinate preservation by the safe family.

Table S8. Affine preservation by safe classes.

Profile	Preserve Born affinely	Preserve sulfur affinely	Preserve all q-structure affinely	Pairs never separated by safe family
Reference D10	0	0	0	none
Observer-safe D8	0	0	0	D/N; E/Q
Moon pH 3.8 D10	0	0	0	D/N; E/Q

Physical-span count distribution across safe classes.

Table S9. Structural core preservation counts.

Profile	Span count 0	Span count 1	Span count 2	Span count 3
Reference D10	35	59	32	5
Observer-safe D8	11	21	12	3
Moon pH 3.8 D10	23	40	22	4

No target values are used to determine these counts.

S9 Fully nested H1 results and dual nulls

Table S10. Nested H1 performance and dual nulls.

Endpoint	Grammar	Safe	Q^2	MAE	RMSE	Holm pipeline p	Fixed-class stress p
WW Octanol	Reference D10	131	0.825	0.512	0.654	8e-05	2e-05
WW Interface	Reference D10	131	0.529	0.442	0.573	0.00116	0.00266
Hessa	Observer-safe D8	47	0.586	0.600	0.821	0.00066	0.00072
Moon-Fleming	Moon pH 3.8 D10	89	0.213	1.463	2.001	0.0229	0.113

MAE and RMSE in Table S10 are in kcal/mol. Holm pipeline p is adjusted across the four endpoint tests; fixed-class stress p is unadjusted.

The pipeline-exact null permutes the response and reruns the complete outer-LOO/inner-LOO adaptive selector for every replicate. The auxiliary fixed-class stress null instead records the best full-LOO statistic among admissible fixed classes for each permutation. It is an additional stress family, not an analytically assumed upper bound for the adaptive nested statistic.

Moon-Fleming’s terminal qualification is driven jointly by its fixed-class stress-null $p=0.1127$, seven selected outer-fold classes, large post-selection modal-vs-nested gap, and open row-level uncertainty provenance.

The predictive coefficient is $Q^2 = 1 - \sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2$, using outer-fold predictions $\hat{y}_i$ and the mean observed response $\bar{y}$ over the evaluation panel.

S9.1 Matched-grammar baselines

To distinguish the contribution of validation-safe quotient construction from algorithm choice, Ridge, PCR, and PLS were evaluated on response-blind linearly independent bases selected in the frozen descriptor order derived from the same frozen endpoint grammar. Standardization, hyperparameter/component selection, and fitting were performed strictly inside the nested training folds.

Table S11. Matched grammar predictive baselines.

Endpoint	H1 Q^2	Ridge Q^2	PCR Q^2	PLS Q^2
WW Octanol	0.825	0.738	0.583	0.721
WW Interface	0.529	0.612	0.575	0.708
Hessa	0.586	0.374	0.416	0.367
Moon-Fleming	0.213	0.432	0.185	0.295

H1 exceeds all three matched baselines for WW octanol and Hessa, but not for WW interface or Moon-Fleming. The result therefore supports validation-safe state-space construction and auditable compression, not universal predictive superiority. Full MAE/RMSE values, selected hyperparameters, and outer predictions are stored in `results/MATCHED_GRAMMAR_BASELINES.json`.

S10 H2a shared prediction and H2b identifiability

Table S12. Shared prediction and alignment nulls.

Campaign	Endpoints	Candidate classes	Aggregate Q^2	Alignment exceedances / 10,000	Plus-one p
Four endpoints	WW_Octanol, WW_Interface, Hessa, Moon_Fleming	47	0.5491	0	9.999e-05
Three endpoints	WW_Octanol, Hessa, Moon_Fleming	47	0.5756	0	9.999e-05

For each alignment campaign, NumPy `default_rng` is initialized with seed 2877662319691443303. Draw order is replicate, endpoint in the table order, then `rng.permutation` of its 20 response entries. The RNG is restarted for each campaign. Every draw repeats nested shared selection over the 47 observer-safe classes; endpoint scaling uses only the outer training set. Independent within-endpoint permutations test this specific null and do not preserve the empirical WW cross-endpoint dependence. Dependency weighting is not a replacement for a dependence-preserving null.

A complementary joint-profile null uses a common permutation π for all endpoints: $Y_{i,e}^\pi = Y_{\pi(i),e}$. It preserves the complete multivariate profiles, marginal distributions and cross-endpoint covariance. NumPy PCG64 via `default_rng(2026091601)` is restarted for each campaign; each replicate draws `rng.permutation(20)` once. We run 10,000 draws per campaign with the unchanged weights, 47 classes, full nested shared selection, fold-local normalization and deterministic tie rules. The null asserts exchangeability of whole profiles relative to fixed descriptors. It neither preserves chemistry-conditioned association nor adjusts prior manual pipeline design.

Both campaigns have zero exceedances, $p_+ = 1 / 10001$ and Holm-adjusted $p = 2 / 10001$ over the two synchronized campaigns. The two-sided 95% Clopper–Pearson interval for each unknown permutation exceedance probability is $[0, 0.00036882]$; this interval concerns Monte Carlo precision, not population prediction uncertainty. The four-endpoint and three-endpoint null 99th percentiles are **0.195034** and **0.206407**, respectively. The broader synchronized tails show why preserving dependence is substantively relevant despite the same zero-exceedance count. Two fixed joint permutations per campaign were independently checked by explicit inner and outer least-squares refits, with absolute discrepancies below 10^{-10} ; covariance preservation was checked to 10^{-12} . Synchronized resampling has established methodological precedent [34]. This tests profile alignment, not a unique generator or superiority of shared over separate models.

Figure S1 displays the synchronized null distributions.

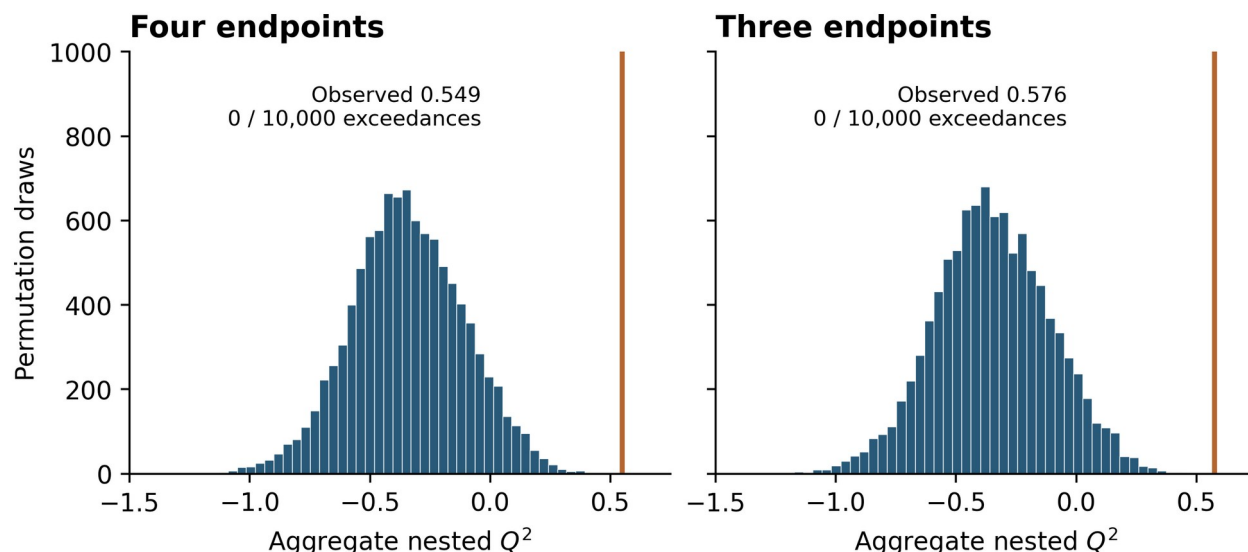


Figure S1. Synchronized H2 alignment nulls. Histograms show the complete 10,000-draw distributions for each campaign; vertical lines mark the observed aggregate nested Q^2 . Whole endpoint profiles are permuted together. The zero exceedances in both campaigns do not imply a zero tail probability.

The H2 local sensitivity statistic is $\Delta_{share} = \text{mean}(Q_{separate}^2) - Q_{shared}^2$. It is an illustrative comparison of predictive procedures, not a physical identifiability test. Use the three dependency-reduced endpoints and the same 47-class family for both procedures. Shared selection minimizes endpoint-normalized inner SSE; separate selection uses the H1 MAE/RMSE tie rule. Let $B = (n_{CS}, n_{ON}, \beta_{br})$, $N = (n_C, n_{ON}, \beta_{br})$, $F_1 = (n_C, n_{heavy}, \beta_{br}, n_{ring})$ and $F_2 = (n_C, \beta_{br}, aromatic_flag)$. The shared, near-shared and separate scenarios project the three observed endpoints onto (B, B, B) , (B, N, B) and (B, F_1, F_2) , respectively, with an intercept. Add independent Gaussian noise of standard deviation $SD(\mu)\sqrt{(1-R^2)/R^2}$, using population SD and no per-draw normalization.

There are 1,000 replicates for each scenario at each signal $R^2 = 0.3, 0.5, 0.7$. One continuous NumPy RNG stream uses seed 20260915 and order signal level, scenario as listed, replicate, endpoint, residue; batches contain 50 replicates. The table reports exceedance of the empirical 95th percentile of the corresponding shared-scenario statistic. The 5% shared value is calibration by construction, not held-out calibration accuracy.

Table S13. Local H2 sensitivity under declared scenarios.

Signal R^2	Shared threshold	Shared	Near-shared	Separate
0.3	0.156324	50/1000	42/1000	112/1000
0.5	0.109891	50/1000	54/1000	111/1000
0.7	0.071671	50/1000	49/1000	129/1000

The limited separation is specific to these response-conditioned signals, noise model and selection rules. Low simulated discrimination does not prove structural nonidentifiability. It illustrates why the observed shared predictive score cannot establish a unique microscopic geometry. All simulation choices and outputs are included in the executable package.

S11 Physical-to-predictive bridge

Finite exact facts:

- A20 row partition under full D10: 20.
- A20 row partition under q2: 20.
- Full D10 affine dimension: 8.
- q2 affine dimension: 5.
- $\text{rank}([1, D10]) = 9$.
- $\text{rank}([1, D10, b_{Born}]) = 10$; therefore b_{Born} is outside the full D10 affine span.
- Sulfur support in the declared A20 D10 ledger is C/M only.
- Reference nested-safe classes: 131; observer-safe Hessa: 47; Moon pH3.8: 89.
- No nested-safe reference class is A20-injective; the maximum row partition is 19.
- No reference nested-safe class preserves b_{Born} affinely.
- No reference nested-safe class preserves sulfur presence affinely.
- Five reference nested-safe classes preserve $(n_{CS}, n_{ON}, \beta_{br})$ affinely.

These facts establish a lossy bridge, not an exact factorization of q2 through one predictive safe class.

S12 Retrospective geometry-concordance audit

The fixed historical WWO basis gives:

- D3 distance: 0.0176931.
- Mean recovered D: 2.9992233.
- D coefficient-of-variation diagnostic: 0.011612.
- B2 relative error to $1/\sqrt{2\pi}$: 0.9676%, computed directly from the supplied historical coefficient record.

This is a retrospective fixed-basis record with response-exposed descriptor and constant selection. The historical Cys hydrogen-bond coding differs from the exact N/O count. S22 supplies the complete conditional permutation test and the Cys correction control; neither is a test of intrinsic physical dimension.

S13 Mechanistic gate matrix

Table S14. Mechanistic tests and evidence status.

Candidate	Gate	Status	Reason
Formal charge only	response-blind state	PASS	formal charge fixed by state contract
Formal charge only	charged-sector exact fiber sufficiency	FAIL	K/R have the same +1 formal charge while WWO differs by 0.99 kcal/mol
Additive Born reference	response-blind state	PASS	n_{CS} , n_{ON} , and b_{Born} fixed independently of endpoint score
Additive Born reference	global exact A20 fiber closure	FAIL	I/L/M and P/V remain
Additive Born reference	first-order empirical adequacy	PARTIAL SUPPORT	declared downstream reference $Q^2 = 0.917248$
Additive Born reference	first-principles coefficient derivation	OPEN	coefficients are not derived from a microscopic Hamiltonian
Conserved-support response	support topology	EXACT PASS	$W_m = 1^\perp$, $\dim W_m = m - 1$

Candidate	Gate	Status	Reason
Conserved-support response	carboxylate adversarial control	EXACT PASS	$m = 2$ gives S^0 , not continuous S^1
Conserved-support response	guanidinium robust-rank carrier	OPEN	required physical-carrier evidence is not yet available
Conserved-support response	independent four-support carrier	OPEN	required physical-carrier evidence is not yet available
Conserved-support response	branch-balance bridge	OPEN	branch balance has not been empirically established
Conserved-support response	metric-equality bridge	OPEN	the required metric relation has not been empirically established
Conserved-support response	solvent-transfer factorization	OPEN	the required transfer factorization has not been empirically established
Resonance only	support-geometry discrimination	FAIL	resonance does not determine m or S^{m-2}
Historical $D \approx 3$ coefficient lock	response-blind eligibility	PROVENANCE FAIL	response-exposed historical selector
Mechanistic survivor set	unique mechanism identified	NO	Born and conserved-support candidates survive with distinct domains and open bridges

S13.1 Conserved-support theorem

For fixed total charge on m supports, admissible internal redistributions satisfy $1^T \chi = 0$, so $W_m = 1^\perp$ and $\dim W_m = m - 1$. At fixed nonzero norm, the direction space is S^{m-2} for $m \geq 2$. This theorem is exact inside the declared support model; it does not establish the physical carrier or solvent-transfer mapping.

S14 Adversarial boundary map

The eight declared adversarial axes yield the following terminal boundaries:

- formal charge alone: no-go as complete charged-sector quotient;
- resonance alone: no-go as support-geometry selector;
- protonation-agnostic state quotient: no-go;
- tautomer sufficiency: nonidentifiable;
- static conformer-complete mechanism: no-go;
- residue-independent WWO-to-WWI environment offset: no-go;
- broad unseen-chemotype extrapolation: no-go under current stress audit;
- observer deletion: valid only with explicit information loss.

For the closest environment perturbation, WWI-WWO has mean -0.3075 kcal/mol, range 3.22, population SD 0.8584, and maximum absolute centered deviation 2.1025 kcal/mol at D. Raw WWO-WWI Pearson correlation is 0.9194. High correlation therefore does not eliminate the mixed interaction.

S15 Applicability and uncertainty

Leave-one-chemotype-out metrics are:

Table S15. Chemotype holdout performance.

Endpoint	Q^2	MAE	RMSE
WW Octanol	-1.977	1.595	2.694
WW Interface	-1.356	0.911	1.283
Hessa	0.277	0.946	1.085
Moon-Fleming	0.289	1.509	1.902

WW octanol acidic+basic holdouts contribute 95.29% of stress SSE. A charge-free observer-safe stress gives WWO $Q^2=0.232$ and WWI $Q^2=-0.239$.

WW reported-error sensitivity uses 10,000 independent bounded row perturbations as a sensitivity analysis, not a confidence distribution because endpoint covariance is unavailable.

Table S16. WW bounded error perturbation sensitivity.

Endpoint	Central Q^2	2.5% perturbation quantile	Median perturbation Q^2	97.5% perturbation quantile	Central modal class retained
WW Octanol	0.8245	0.7958	0.8231	0.8505	34.73%
WW Interface	0.5294	0.4567	0.5564	0.6786	56.88%

The perturbation ranges show stable predictive performance but substantially less stable selected-class identity. These quantiles are sensitivity summaries, not confidence intervals. Hessa residue-wise uncertainty is UNKNOWN; Moon-Fleming row-level source/uncertainty provenance remains open.

S16 Declared prospective contracts

S16.1 Primary no-refit Born predictor

$$\widehat{\Delta G} = 0.771447 - 0.357128 n_{CS} + 0.372280 n_{ON} + 16.778421 b_{Born}. \quad (S8)$$

Same-observer panel requirements: new noncanonical guests; panel fixed before responses; $N \geq 12$; no coefficient or descriptor refit. PASS requires external $Q^2 > 0$ and exact one-sided Spearman permutation $p \leq 0.01$. FAIL requires $Q^2 \leq 0$ or $p > 0.05$; otherwise INCONCLUSIVE.

S16.2 Counts-only ablation

$$\widehat{\Delta G}_0 = 0.8663722885 - 0.3863271738 n_{CS} + 0.9671738459 n_{ON}. \quad (S9)$$

b_{Born} passes the incremental gate only if its no-refit MAE is lower and the paired absolute-error advantage has exact one-sided random-sign $p \leq 0.05$.

S16.3 Conserved-support physical gates

Let M be the measured or computed carrier matrix and suppose the error has spectral norm at most ϵ_M . Rank at least r is robustly certified when $\sigma_r(M) > \epsilon_M$, equivalently $\lambda_r(M^T M) > \epsilon_M^2$, with eigenvalues in decreasing order. Positive definiteness of $M^T M - \epsilon_M^2 I_r$ is equivalent only when M has exactly r columns in an independently declared internal-mode basis. Known conserved null directions must first be removed in that basis; full-space positive definiteness must not be required when such directions remain. The guanidinium carrier test requires robust rank at least two and the independent four-support test at least three; equality additionally uses the declared support-model rank ceiling. No physical verdict has been opened.

S16.4 Crossed Born/conserved-support design

A future panel must include (i) same support count with at least 25% Born-amplitude separation and (ii) different support count with Born amplitudes matched within 10%. These are prespecified design tolerances, not claimed physical constants. If no chemically admissible panel satisfies the design, the terminal is **OPEN (panel feasibility)** rather than post hoc threshold relaxation.

S16.5 Planning requirements before new measurements

The primary frozen predictor and original decision rule in S16.1 remain unchanged. The counts-only MAE comparison in S16.2 and chemical-family coverage were already part of the archived contract. Before any new panel is measured or unblinded, record exact molecular identities, state conventions, coverage counts and matched contrasts for branching, sulfur and polar/charged structure. Repeated near-identical analogues must not stand in for broad chemical coverage. The Born/support crossed experiment requires contrasts at approximately fixed competing coordinates.

A meaningful effect-size threshold requires the intended application and assay precision, neither of which is supplied by A20 alone. A definitive augmented protocol must fix that threshold and perform power/sensitivity calculations for $N = 12, 16, 20$ using the actual proposed structures, their frozen predictions, comparator predictions and an explicit measurement-error model. That design-specific calculation remains open until those inputs exist; no power number or practically meaningful cutoff is invented here. Calibration slope/intercept may be reported diagnostically without refitting the frozen predictions. Any new decision rule must be versioned before new responses are exposed and reported alongside, rather than substituted for, the original contract.

S17 Claim-boundary ledger

The claim-boundary ledger preserves each scientific statement together with its scope and promotion guard. The evidence vocabulary is **THEOREM, CERTIFIED, EMPIRICAL, CONDITIONAL, MODEL, NONIDENTIFIABLE, NO-GO, and OPEN**.

C01 — CERTIFIED

Domain: space. **Proposition:** Declared finite slices are not identified with their ambient molecular/endpoint/descriptor possibility spaces. **Scope:** finite-slice typing. **Promotion guard:** No promotion to universal molecular space.

C02 — THEOREM

Domain: thermodynamics. **Proposition:** For a declared equilibrium ensemble, free-energy contrasts remove the additive absolute-free-energy gauge. **Scope:** declared equilibrium ensemble. **Promotion guard:** The theorem is observer-geometry, not a microscopic Hamiltonian identification.

C03 — THEOREM

Domain: thermodynamics. **Proposition:** For a phase x molecular-variant free-energy table, the exact mixed interaction component is $P_m F P_n$ and double differences depend only on it. **Scope:** matched phase/variant table. **Promotion guard:** Does not identify a microscopic channel basis.

C04 — CERTIFIED

Domain: physical_state. **Proposition:** $q_2 = (n_{CS}, n_{ON}, b_{Born}, \beta_{br}, s)$ is injective on the declared A20 WWO state slice. **Scope:** declared A20 WWO state slice. **Promotion guard:** Injectivity is not a global microstate mechanism.

C05 — CERTIFIED

Domain: physical_state. **Proposition:** Every proper subset of the five q_2 primitive coordinates fails exact WWO point-estimate sufficiency on declared A20. **Scope:** declared q_2 coordinate family on A20. **Promotion guard:** Minimal only within the declared primitive family.

C06 — CERTIFIED

Domain: physical_state. **Proposition:** The q_2 structural matrix has exact column and affine rank 5 on the declared rationalized A20 ledger. **Scope:** declared finite ledger. **Promotion guard:** Structural rank is not intrinsic response rank.

C07 — NONIDENTIFIABLE

Domain: identifiability. **Proposition:** The intrinsic latent response rank is identified by the scalar WWO record. **Scope:** current scalar observer. **Promotion guard:** No $D=3$, $D=5$, or other latent-rank promotion is allowed.

C08 — NO-GO

Domain: observer_transfer. **Proposition:** A common physical observer/protocol is established across WWO, WWI, Hessa, Moon-Fleming, Fauchere-Pliska, and Eisenberg. **Scope:** current source contracts. **Promotion guard:** Shared labels or algebra do not override NO_TRANSFER/PARTIAL/NONIDENTIFIABLE entries.

C09 — NO-GO

Domain: bridge. **Proposition:** The complete declared q_2 state factors affinely through a single nested-safe D10 predictive class. **Scope:** current D10/nested-LOO protocol. **Promotion guard:** Physical sufficiency and predictive sufficiency remain distinct.

C10 — THEOREM

Domain: validation. **Proposition:** Validation-depth sufficiency is the generic rule that admissible predictive states preserve affine rank on every training domain induced by the full resampling protocol. **Scope:** declared affine predictive design/resampling protocol. **Promotion guard:** Dataset counts are separate certified instances.

C11 — CERTIFIED

Domain: validation. **Proposition:** For the current outer-LOO/inner-LOO audit, the low-capacity safe-class counts are 131 (reference D10), 47 (observer-safe D8), and 89 (Moon pH3.8). **Scope:** declared A20 ledgers and protocol. **Promotion guard:** Counts are finite-study certificates, not universal capacities.

C12 — EMPIRICAL

Domain: prediction. **Proposition:** WW Octanol admits useful nested low-dimensional predictive compression under the declared audit. **Scope:** A20 WW Octanol; declared grammar/protocol. **Promotion guard:** Does not imply universal superiority or mechanism.

C13 — EMPIRICAL

Domain: prediction. **Proposition:** WW Interface admits useful nested predictive compression under the declared audit. **Scope:** A20 WW Interface; WW dependency family. **Promotion guard:** Dependency on WW family must be retained.

C14 — EMPIRICAL

Domain: prediction. **Proposition:** Hessa admits useful observer-safe nested predictive compression under the declared audit. **Scope:** A20 Hessa observer-safe grammar. **Promotion guard:** Residue-wise uncertainty remains open.

C15 — CONDITIONAL

Domain: prediction. **Proposition:** Moon-Fleming has strong endpoint-level compression evidence equivalent to the stronger H1 endpoints. **Scope:** pH3.8-conditioned Moon-Fleming slice. **Promotion guard:** High selection dispersion and open row-level uncertainty block stronger promotion.

C16 — EMPIRICAL

Domain: shared_prediction. **Proposition:** One observer-safe structural class supports useful endpoint-specific prediction across the declared four-endpoint campaign. **Scope:** declared campaign with endpoint-specific coefficients. **Promotion guard:** Shared prediction is not shared physical geometry.

C17 — NONIDENTIFIABLE

Domain: shared_geometry. **Proposition:** A unique common structural geometry is identified across endpoints. **Scope:** N=20 current design. **Promotion guard:** A useful shared predictor does not identify a unique generator.

C18 — NO-GO

Domain: prediction. **Proposition:** The compression grammar is universally predictively superior to matched or broader representations. **Scope:** current endpoints/comparators. **Promotion guard:** Comparator wins vary by endpoint; the formal-charge magnitude $|z_{formal}|$ closes much of the WWO Generic11 gap.

C19 — NO-GO

Domain: generalization. **Proposition:** The current predictive grammar certifies broad unseen-chemotype generalization. **Scope:** current leave-one-chemotype-out audit. **Promotion guard:** Acidic/basic WWO holdouts dominate stress; no broad transfer certificate.

C20 — EMPIRICAL

Domain: geometry. **Proposition:** The historical WWO coefficient fingerprint and its fixed-basis response-permutation stress statistic are independently reproducible on the recovered historical design. **Scope:** fixed historical WWO coding, including Cys = 1 and effective Arg = 0.5. **Promotion guard:** Retrospective, response-exposed design; the permutation calculation does not repeat historical model/constant selection and does not identify physical dimension.

C21 — NONIDENTIFIABLE

Domain: geometry. **Proposition:** $D=3$ is the intrinsic protein response rank/dimension. **Scope:** current scalar/multi-endpoint evidence. **Promotion guard:** No coefficient concordance can repair rank nonidentifiability.

C22 — NO-GO

Domain: geometry. **Proposition:** The same absolute $D=3$ coefficient lock transfers across all six endpoints.

Scope: current endpoint transfer contracts. **Promotion guard:** Observer typing remains upstream.

C23 — NO-GO

Domain: mechanism. **Proposition:** The old zero-parameter/canonical coefficient formula is a first-principles microscopic derivation. **Scope:** historical fitted basis. **Promotion guard:** Numerical concordance is not derivation.

C24 — MODEL

Domain: mechanism. **Proposition:** The additive Born reference is retained as a response-blind first-order energetic/structural lift. **Scope:** typed state and environment slice. **Promotion guard:** It is globally incomplete and coefficients are empirical.

C25 — NO-GO

Domain: mechanism. **Proposition:** The additive Born reference is an exact global A20/microstate mechanism.

Scope: current finite and adversarial tests. **Promotion guard:** I/L/M, P/V, protonation, conformation, and environment adversaries prevent promotion.

C26 — THEOREM

Domain: mechanism. **Proposition:** For fixed total charge on m declared supports, $W_m = 1^\perp$ has dimension $m-1$ and unit shell S^{m-2} when $m \geq 2$. **Scope:** declared conserved-support model. **Promotion guard:** Mathematical theorem inside the declared support model; physical carrier bridge remains open.

C27 — OPEN

Domain: mechanism. **Proposition:** The guanidinium and independent four-support robust-rank physical-carrier gates are resolved. **Scope:** declared response-blind producer tests. **Promotion guard:** No $1/\pi$, $4/3$, or $4/(3\pi)$ promotion before these and downstream gates.

C28 — OPEN

Domain: mechanism. **Proposition:** Born-like amplitude and conserved-support topology are physically equivalent/coupled by an established theorem. **Scope:** current survivor set. **Promotion guard:** Treat as complementary candidates, not equivalent mechanisms.

C29 — NONIDENTIFIABLE

Domain: mechanism. **Proposition:** A unique complete microscopic mechanism for the observed amino-acid response has been identified. **Scope:** current evidence. **Promotion guard:** At least two partial survivors remain and neither closes to the observable.

C30 — NO-GO

Domain: chemistry. **Proposition:** Protonation state can be globally quotiented out without changing the response representation. **Scope:** current chemistry/environment audit. **Promotion guard:** State typing is mandatory.

C31 — NO-GO

Domain: chemistry. **Proposition:** Static q_2 /Born/conserved-support coordinates form a conformer-complete microscopic mechanism. **Scope:** current adversarial audit. **Promotion guard:** Conformational ensemble dependence remains irreducible.

C32 — NO-GO

Domain: environment. **Proposition:** WWO to WWI differs only by a residue-independent environment offset. **Scope:** WW source family. **Promotion guard:** Environment x residue interaction cannot be quotiented to a common offset.

C33 — CERTIFIED

Domain: prospective. **Proposition:** The primary no-refit prospective Born predictor and success/failure criteria were declared before new blind data. **Scope:** future same-observer blind panel. **Promotion guard:** Prespecification is not a positive prospective result.

C34 — OPEN

Domain: prospective. **Proposition:** The additive Born lift has new untouched prospective independent support beyond A20. **Scope:** declared prospective panel. **Promotion guard:** No historical dataset may be relabeled prospective.

C35 — NO-GO

Domain: prospective. **Proposition:** Previously exposed WWI/Hessa/Moon/FP/Eisenberg or historical Radzicka-Wolfenden data constitute prospective confirmation. **Scope:** prospective evidence class. **Promotion guard:** They remain retrospective/context evidence only.

C36 — OPEN

Domain: uncertainty. **Proposition:** Residue-wise Hessa uncertainty is closed under the current source contract. **Scope:** Hessa. **Promotion guard:** No uncertainty distribution is imputed.

C37 — OPEN

Domain: uncertainty. **Proposition:** Moon-Fleming row-level uncertainty provenance is closed. **Scope:** Moon-Fleming. **Promotion guard:** No perturbation claim is made until source provenance closes.

C38 — NO-GO

Domain: translation. **Proposition:** The present work establishes a translational/clinical efficacy claim. **Scope:** current evidence. **Promotion guard:** Potential relevance is downstream; no diagnostic/therapeutic efficacy claim is established.

C39 — OPEN

Domain: translation. **Proposition:** A future translational application based on truth-preserving interaction-response representation is admissible as a research direction. **Scope:** future independently validated domain. **Promotion guard:** Requires evidence appropriate to translational level.

C40 — NO-GO

Domain: closure. **Proposition:** All current scientific questions are resolved by the present claim-boundary analysis. **Scope:** current evidence base. **Promotion guard:** Claim typing does not close the open science; prospective and mechanistic gates remain open.

C41 — THEOREM

Domain: validation. **Proposition:** For arbitrary deletion sets of size at most d , affine-rank preservation is equivalent to $\text{cospark}(A) > d$; for the declared outer-LOO/inner-LOO protocol, nested safety is therefore equivalent to $\text{cospark}(A) > 2$. **Scope:** full-column-rank affine designs; arbitrary row deletions and the current nested-LOO protocol. **Promotion guard:** Structured split families remain governed by their explicit protocol-induced deletion sets and are not reduced to cardinality alone.

C42 — CERTIFIED

Domain: representation. **Proposition:** In the frozen response-blind D10 ledger, the minimum injective A20 encoding uses five coordinates, and 16 distinct five-coordinate D10 subsets are injective. **Scope:** canonical A20 rows in the D10 ledger. **Promotion guard:** Finite injectivity does not identify a unique physical geometry and does not privilege q_2 by injectivity alone.

C43 — THEOREM

Domain: quantitative sufficiency. **Proposition:** Within-fiber averaging minimizes finite-record Frobenius loss, and nested refinement obeys the orthogonal squared-loss decomposition proved in S19. **Scope:** Declared finite response matrix, units and equal state weights. **Promotion guard:** Zero loss on an injective finite record does not imply unseen-state prediction.

C44 — THEOREM

Domain: validation identifiability. **Proposition:** A held-out row has the same prediction for every OLS minimizer exactly when it lies in the training row space; S20 supplies a singleton-Arg counterexample after rank loss. **Scope:** Unregularized affine model and declared training domain. **Promotion guard:** A regularizer defines an additional predictor convention; it does not restore quotient invariance automatically.

C45 — THEOREM

Domain: incomplete interaction records. **Proposition:** The interaction quotient of a bipartite observation graph has dimension equal to edges minus vertices plus connected components, as proved in S21. **Scope:** Observed additive context-by-residue design, including isolated vertices. **Promotion guard:** Missing observations are not zeros or demonstrated physical inadmissibility.

C46 — CERTIFIED

Domain: historical coding sensitivity. **Proposition:** Changing only the Cys historical hydrogen-bond code from one to the exact N/O count zero changes the coefficient-template distance from 0.01769313 to 0.07742298. **Scope:** Two explicitly supplied retrospective designs on the same WW octanol response. **Promotion guard:** The historical permutation p-value is not transferred to the corrected design or interpreted as physical dimension.

C47 — CERTIFIED

Domain: validation. **Proposition:** The recorded q2 affine design has rank six and cospark two; all single deletions pass and exactly the Cys/Met pair fails. **Scope:** Exact recorded A20 decimal ledger; ordinary versus nested LOO. **Promotion guard:** Finite physical injectivity does not imply affine interpolation or nested predictive identifiability.

C48 — EMPIRICAL

Domain: prediction. **Proposition:** Both shared H2 campaigns exceed all 10,000 joint-profile-preserving permutation draws. **Scope:** Retrospective 47-class selector; exchangeable whole-profile alignment null. **Promotion guard:** Not unique geometry, shared-over-separate superiority, chemistry-conditioned exchangeability or correction of human pipeline discovery.

C49 — CERTIFIED

Domain: validation. **Proposition:** FreeSolv count designs preserve LOO rank but lose rank under element-family deletion; carbon-only controls pass. **Scope:** 642 recorded structures and numeric-ID-selected 40-molecule subset; declared group splits. **Promotion guard:** No endpoint prediction, prevalence or general performance-improvement claim.

S18 Reproducibility materials

The accompanying reproducibility archive (Additional file 2) is organized around declared scientific inputs, analysis code, and replayable outputs. It contains:

1. source endpoint and descriptor tables with provenance records;
2. exact finite-state, interaction-quotient, minimality, rank, and observer-transfer certificates;
3. the predictive compression audit with validation-safe class registries, the cospark/nested-safety corollary certificate, the finite-injective-encoding negative control, deterministic H1/H2 replays, comparator and chemotype analyses, and uncertainty-sensitivity outputs;
4. endpoint-isolated 50,000-permutation algorithm-matched H1 null replays and the 50,000-permutation fixed-class stress replay;
5. mechanistic and adversarial boundary records plus the declared prospective protocols;
6. executable verification scripts, environment requirements, and a SHA-256 manifest.

The deterministic analyses and structural mutation guards are replayable from the archive. The null ledgers are independently reproducible endpoint by endpoint. Integrity verification checks file hashes and agreement between reported values and replay outputs. Only materials required to reproduce and audit the reported scientific results are included in the publication-facing archive.

S19 Quantitative fiber sufficiency

Let $R \in \mathbb{R}^{n \times k}$ contain the declared response profiles, with states in rows. Fix a representation q and let Π_q average rows within its fibers, using equal state weights. Its image consists precisely of matrices whose rows are constant on each fiber. Since Π_q is a symmetric idempotent matrix, it is the orthogonal projector onto that space. Consequently

$$\Delta_q^2(R) := \left\| R - \Pi_q R \right\|_F^2 = \min_g \left\| R - g(q) \right\|_F^2. \quad (\text{S10})$$

The minimum is over arbitrary finite-record functions of q , not merely affine regressions. Thus $\Delta_q=0$ is equivalent to exact factorization of this response table through q .

If q' refines q , then the two projection ranges are nested and $\Pi_q \Pi_{q'} = \Pi_q$. Orthogonal decomposition gives

$$\Delta_q^2(R) = \Delta_{q'}^2(R) + \left\| (\Pi_{q'} - \Pi_q) R \right\|_F^2. \quad (\text{S11})$$

The final term measures the exact reduction in squared reconstruction loss. An injective representation has zero loss automatically; this does not certify unseen-state prediction. The norm depends on the declared units and weighting of responses. Physically different observables must not be pooled into an unqualified energy norm.

For a complementary coordinatewise uniform loss, choose the midpoint of each response coordinate's range on each fiber. The triangle inequality supplies the matching lower bound, proving

$$\min_g \max_{i,c} |R_{ic} - g_c(q_i)| = \frac{1}{2} \max_{q_i=q_j,c} |R_{ic} - R_{jc}|. \quad (\text{S12})$$

This is a statement about a finite point-estimate record. Measurement error requires a separate declared tolerance or uncertainty model; a pairwise tolerance relation is not automatically transitive.

On the A20 WW-octanol record, q_0, q_1, q_2 have 17, 19 and 20 fibers. Their exact squared losses are 5479 / 15000, 841 / 5000 and zero, respectively, in $(kcal/mol)^2$. The refinement gains are 739 / 3750 and 841 / 5000, which sum exactly to 5479 / 15000. The relative Frobenius defects, normalized by $\|P_{20} R\|_F$, are 8.65636%, 5.87412% and zero. These percentages are ratios of norms, not percentages of variance. The unavoidable uniform errors are 0.30, 0.29 and zero kcal/mol. The second result makes clear that eliminating most collisions need not eliminate the worst remaining collision.

For the WW-interface-minus-octanol response, the same exact squared losses are 179 / 1875, 1 / 32 and zero. The full machine-readable table also contains the separate water-reference and reported dewetting blocks, with their units retained in the source ledger. No cross-observer causal equivalence is inferred.

S20 Held-out prediction identifiability and the singleton-Arg control

Let A_T be a training design and let a be a held-out row. The least-squares minimizers are $\hat{\beta} + \ker A_T$. Hence all minimizers predict the same $a\beta$ if and only if

$$\ker A_T \subseteq \ker a \Leftrightarrow a \in \text{rowspan}(A_T). \quad (\text{S13})$$

The forward implication follows by subtracting two minimizers; the reverse implication follows by annihilating every training-null direction. For every row of a declared full design A , the condition is equivalent to $\text{rank}(A_T) = \text{rank}(A)$. This connects the validation-depth criterion to identifiability of held-out predictions. A regularizer or specified pseudoinverse can still define a number after rank loss, but its additional convention is part of the predictor and is not an invariant of the unregularized affine quotient.

Consider $A = [1, n_{CS}, n_{ON}, f_{DEK}, f_R]$. The exact full rank is five; deleting Arg reduces it to four. The original-basis pseudoinverse LOO score is $Q^2 = 0.92416821454$. Replace the last column by $1 + f_R$, leaving all other columns unchanged. This invertible column transformation has determinant one and preserves the full column space. Yet the LOO score becomes 0.93813060522, and the held-out Arg prediction changes from 0.64963338672 to 0.99401050009 kcal/mol. The original numerical score therefore reproduces, while its

interpretation as validation-safe class separation fails. An otherwise analogous common-charge design passes the basis-invariance control.

This example does not alter the frozen 131-class nested H1 campaign: that campaign already rejects representations losing rank on its training domains. It explains why the apparently stronger unrestricted charge-family model cannot replace that gate.

S21 Context-relative equivalence and incomplete observations

For each declared context c , define a response r_c on its admissible molecular states. If admissibility itself is part of the target, extend r_c by a distinct symbol $\perp$ outside that domain. Define $x \sim_c y$ when $r_c(x) = r_c(y)$ for all $c \in C$. A representation is sufficient for this family exactly when every one of its fibers lies within a $\sim_c$ class. This follows by applying the usual factorization criterion to the vector of all responses.

If $C \subseteq C'$, equivalence for C' implies equivalence for C . Thus enlarging the retained response family can only refine the exact response partition. This is a logical statement about specified responses, not evidence that a finite observed panel covers every future environment. A missing observation is also different from physical inadmissibility: unknown cells are not assigned $\perp$ as if their feasibility had been disproved.

For an incomplete context-by-residue record, form the bipartite observation graph. Let E be its observed edges, $m+n$ its declared vertices and c its connected components, including isolated vertices. The additive design B has one row per observation, with a one in its context and residue columns. The equation $Bz=0$ requires opposite node offsets along each edge, leaving exactly one free constant per connected component. Therefore

$$\text{rank } B = m + n - c, \dim(\mathbb{R}^E / \text{im } B) = |E| - m - n + c. \quad (\text{S14})$$

The quotient is represented by cycle contrasts. A three-corner 2×2 table has no cycle contrast; observing the fourth corner creates one. Zero filling a missing corner can create a spurious interaction. Complete-grid double centering is recovered only when the full matched block is observed.

S22 Retrospective fixed-basis coefficient stress test

The entire 20-row design and WW octanol response are supplied in `reproducibility/historical_fixed_basis/HISTORICAL_INPUTS.csv`. The columns are an intercept, side-chain C+S count n_{nonpol} , historical hydrogen-bond coding n_{HB} and Q_{eff} , which is 1 for D/E/K, 0.5 for R and zero otherwise. The coding n_{HB} coincides with the N/O count except for Cys, assigned 1 instead of 0. This historical convention is confined to the present retrospective control and is not a chemical definition used in q_2 .

For fitted coefficients (A, B_1, B_2, B_3) , define

$$T_{D3} = \sqrt{\left(\frac{A - 2/3}{2/3}\right)^2 + \left(\frac{B_1 + 1/3}{1/3}\right)^2 + \left(\frac{B_3 - 3}{3}\right)^2}. \quad (\text{S15})$$

The coefficients are $(0.670456622280, -0.338774107047, 0.395082142529, 3.011348637875)$ and $T_{D3} = 0.017693128924$. The independent replay constructs the OLS map by exact rational elimination, then computes five million response permutations using NumPy `default_rng`, seed 20260906, batches of 20,000 and rowwise sorting of 20 random keys. No permutation gives $T_{D3} \leq T_{D3,obs}$, so

$$p_+ = \frac{0+1}{5,000,000+1} = 1.9999996 \times 10^{-7}.$$

(S16)

The script additionally checks six stored event counts, five observed statistics, three null minima and sixteen null quantiles. Replaying the same random stream checks computation; it does not double the number of independent trials. The plus-one value is not an upper confidence limit. With zero hits, the one-sided 95% binomial upper bound on the conditional hit probability is $1 - 0.05^{1/5,000,000} \approx 5.99146 \times 10^{-7}$.

The coefficient template, descriptor choice and effective-Arg convention were selected with prior response exposure. The permutation algorithm does not repeat that discovery process. Its p-value is consequently conditional on that selected design, not selection-adjusted evidence for a physical dimension.

Table S17. Historical coding and the Cys correction.

Fixed design	Cys second count	T_{D3}	Fitted R^2
Historical hydrogen-bond coding	1	0.01769313	0.96206031
Exact N/O count; all other entries retained	0	0.07742298	0.96578749

`reproduce_cys_contrast.py` recomputes both rows. A better unconstrained fit can be farther from the selected coefficient template. The historical p-value is not reused for the changed design. The matched contrasts D–N = 2.79 and E–Q = 2.86 kcal/mol cancel identical count coordinates on the same WW record; they are energetic contrasts, not independent data or dimension measurements.

S23 Executable scope of the additional certificates

`reproducibility/quantitative_certificates/reproduce_certificates.py` verifies 21 finite-record controls: nested averaging projectors, exact refinement-loss decompositions, singleton-Arg rank loss and basis dependence, a rank-safe positive control, incomplete-grid cycle dimension, and the printed dewetting interaction rank. The three data tables, source metadata and exact rational output are included. The dewetting table is Lobo et al. Table S1 [23]; its unit is kJ/mol per water molecule. The calculation concerns the printed values and does not reproduce molecular dynamics or supply unavailable cell uncertainties.

The separate historical module contains all inputs and code for S22. `code/reproduce_h2_simulations.py` within the predictive subpackage supplies both the 20,000 total H2 alignment draws and the 9,000 local sensitivity simulations described in S10. Each script specifies its random stream and writes a machine-readable result. Exact arithmetic, floating-point replay tolerances, stochastic frequency and physical inference are distinguished. No external project or unpublished companion work is required.

S24 Relation to recent work

This targeted review covers work published in 2021–2026, searched through 15 September 2026, with older foundational references retained. It addresses amino-acid scales, context-dependent energetics, molecular representation and validation. It is not an exhaustive systematic review of all protein modeling. Reference numbers refer to the manuscript bibliography. No head-to-head performance claim is made without a matched task, data split and fitting protocol.

Table S18. Intersections with recent literature.

Work	Established contribution or overlap	Distinction relevant to this article
Waibl et al. 2022 [22]	Scale and scoring choices affect antibody hydrophobicity prediction.	Observer and applicability are explicit; no new discovery of context dependence is claimed.

Work	Established contribution or overlap	Distinction relevant to this article
AAclust 2024 [9]	Clusters scales to reduce correlated redundancy.	Approximate scale similarity differs from exact equality of affine predictive spaces.
AAontology 2024 [10]	Classifies scales for interpretable representations.	Semantic organization does not by itself certify quotient sufficiency or fold-wise rank.
Grau et al. 2025 [11]	Assay distinguishes interfacial and transmembrane behavior.	Protocol-specific insertion is not identified with water-to-octanol transfer.
Najafi et al. 2025 [12]	Dewetting-based hydrophobicity depends on sequence context.	We use declared observer contracts; context dependence is established prior work.
Lobo et al. 2025 [23]	Compares hydrophobicity across disordered-protein force fields.	Only printed Table S1 is used for the finite loss example; dynamics are not rerun.
Rekhi and Mittal 2025 [24]	Resolves transfer energetics in condensate environments.	Carrier, medium and state conventions must accompany every transfer comparison.
Ok and Park 2026 [13]	Multiview spectral integration yields a useful one-dimensional ordering.	A useful scalar ordering does not identify latent physical rank or unique geometry.
Muhammedkutty and Zhou 2026 [14]	Computational membrane-insertion analysis resolves sequence effects.	Sequence-screening predictions are distinct from measured A20 transfer energies.
DataSAIL 2025 and addendum 2026 [25,26]	Builds similarity-aware data partitions; comparisons depend on benchmark.	Leakage control and protocol-specific rank preservation are complementary.
Guo et al. 2024 [27]	Scaffold splits can retain optimistic similarity patterns.	Split names alone do not certify extrapolation; the actual training domains matter.
ProQSAR 2026 [21]	Provides modular small-data modeling and applicability workflows.	Our pre-fit algebraic gate can accompany such workflows; no software superiority is tested.
ProtTrans 2022 and ESM 2023 [28,29]	Learn rich sequence representations and support downstream protein tasks.	A20 exact compression is a different task and supplies no ranking against foundation models.
ProteinGym 2023 [30]	Standardizes comparisons of mutation-effect predictors.	Assay-level fitness requires a defined adapter before transfer-energy claims can be tested there.
Double-mutant-cycle review 2021 [31]	Reviews established four-state coupling algebra.	Alternating differences are not new; their role in the full reduction-and-validation protocol is the contribution.

The retained methodological contribution is the integration of observer-conditioned thermodynamic quotients, quantitatively certified fiber refinement and validation-safe affine model spaces. Orthogonal projection, thermodynamic-cycle algebra and the cospark criterion are established mathematics; novelty is not assigned to those identities in isolation. No first-in-literature priority claim is based solely on this targeted review.

S25 External validation-rank demonstration

FreeSolv [35] supplies an independent molecular structure set. The archived source contains 642 records. The rank audit reads only identifiers and SMILES; measured and calculated hydration energies do not enter any selection or result. Atom tokens give counts of C, N, O, S, F, Cl, Br and I. The parser handles bracket atoms and aromatic atoms explicitly and rejects unrecognized tokens; it does not infer protonation or valence. The full source and license are retained, with a SHA-256 snapshot.

Before inspecting results, the local execution plan specified all records and a small subset consisting of the first 40 identifiers sorted by their numeric suffix. No subset search was conducted. Starting with the intercept, columns are retained in the fixed order only if they add exact rank. All nine columns survive in both panels. Rank is computed with rational elimination on integer Gram matrices.

Table S19. Response-free external rank audit.

Held-out element family	Full panel held out	Full train rank	Small panel held out	Small train rank
N	169	8	8	8
O	344	8	17	7
S	40	8	3	8
F	35	8	2	8
Cl	114	8	3	8
Br	25	8	2	8
I	12	8	2	8

Full ranks are nine. Every ordinary LOO fold preserves rank in both panels. Family holdouts overlap and are diagnostic extrapolation tests, not a partition into disjoint CV folds. Every held-out row in each family increases training rank when added back, certifying lack of estimability. The oxygen holdout in the small panel removes two independent directions. An intercept-plus-carbon-count control retains rank two on every same split.

For held-out design A_U , compare prediction operators $W = A_U A_T^+$ and $W' = (A_U B)(A_T B)^+$, where the invertible shear replaces each non-intercept coordinate by itself plus the intercept. These operators differ after rank loss, despite identical full column space. Their maximum entry differences range from 0.0774 to 0.6929 in the full panel and 0.2606 to 2.4835 in the small panel. The control differences remain below 10^{-10} . These are operator coefficients, not observed free-energy prediction errors. No target vector is required to establish the ambiguity. A regularizer or prior could define a different, fully specified extrapolator.

This is a deliberately chosen illustration of validation-induced loss of chemical support, not a representative prevalence study, software leaderboard or physical generalization test. It provides an external counterexample without claiming that rank filtering improves predictive accuracy.

S26 Execution and interpretation of the new checks

`reproducibility/revision_checks/ANALYSIS_PLAN.json` records the local analysis plan.
`reproduce_new_checks.py` recomputes the q2 rank audit and synchronized H2 campaigns;
`reproduce_external.py` recomputes the FreeSolv structural demonstration. The archive contains all inputs, complete new permutation arrays and machine-readable results. `verify_new_checks.py` verifies the new results and source provenance without requiring network access. Monte Carlo tail uncertainty, exact recorded-design statements, retrospective association and future physical validation remain separate evidence categories.