Phaeon Molecular

Liquid Water Simulation in a Conserving Energy Graph

Marcel Langjahr

Pantareon
Pantareon Foundations

langjahr@pantareon.io · https://pantareon.io/

July 2026

Abstract

We present *Phaeon Molecular*, the molecular mode of the Emergent Graph Theory (EGT) engine: the same conserving operator grammar that produces baryonic structure in the cosmological mode and linguistic structure in the language mode, applied to condensed-phase matter. A water molecule is not a rigid body with tabulated geometry — it is a bonded subgraph whose shape is a measured output. The H–O–H angle is emergent, not imposed.

The resulting liquid reaches TIP4P/2005-class quality *in structure and dynamics* at a single validated operating point: tetrahedral order $q_{\mathrm{tet}} = 0.556$ against a reference of 0.55, coordination number 4.8 against 4.5, first g_{OO} peak at 2.62 Å against 2.78, self-diffusion $3.0 \times 10^{-9}\,\mathrm{m}^2/\mathrm{s}$ against an experimental 2.3×10^{-9}, and a virial pressure statistically indistinguishable from zero at real liquid density — i.e. NPT self-consistency. The electrostatics is full Ewald with a buried TIP4P M-site, accelerated by a Smooth Particle Mesh Ewald implementation that is gated against the direct sum as an *exact oracle* in both energy and per-site forces.

The engine's distinguishing property is verification as a construction principle. Every force term is finite-difference tested against $-\nabla V$; the neighbour list is gated by exact pair-set equality against brute force; momentum closure, charge conservation and seeded bit-reproducibility are automated gates. Thirty-four such gates run on every build. This layer earned its keep: it exposed a systematic bias in the project's own potential-of-mean-force estimator — an omitted entropic Jacobian term of +0.83 to +1.00 kcal/mol, precisely the magnitude of the effect being measured — which had propagated silently through four measurement campaigns and had produced a published negative result. That result is retracted here, and the corrected re-analysis of the identical raw data returns contact minima of −0.68 and −0.67 kcal/mol with the canonical solvation curve shape.

The programme's binding constraint is quantified rather than asserted. The hydrophobicity campaign sampled 0.72 *pico*seconds per umbrella window against a field standard of 1–10 nanoseconds — a shortfall of three to four orders of magnitude, which alone accounts for every unresolved observable reported to date. Direct one-body measurement confirms the driving force is present: the water charges +3.72 kcal/mol of cavity work against a solute at its physical exclusion radius. Removing the throughput wall is engineering with named mechanisms; 5.63× has been taken at $N = 512$ (scaling exponent $2.04 \to 0.84$), and a further factor of order 30 is identified — narrowed, rather than widened, by a parallelisation attempt that returned 1.06× and relocated the bottleneck from arithmetic to per-step allocation. What remains open is compute, not architecture.

Availability of Source Code: The engine (Rust core with Python orchestration) is proprietary to Pantareon and is not released as open source. Access can be granted upon reasonable request for technical due diligence or scientific verification, subject to a non-disclosure agreement. Licensing, acquisition, and funded co-development of the technology are open to discussion.

Scope of this Paper: This report presents the molecular mode of the EGT engine and its measured results to date. It is a technical status report intended for partners and investors with scientific standards; it is not submitted for peer review. All numbers are from deterministic, reproducible runs, quoted at a single stated operating point, and compared against published reference values for liquid water. Where the model does not reach a reference, that is stated explicitly in §6. The theoretical frame-

work is presented in the companion paper *Emergent Graph Theory (EGT)*; the engine shares its operator grammar and conservation discipline with the cosmological and language modes reported elsewhere.

1 Introduction

Molecular dynamics of condensed-phase matter is a mature field with excellent tools. Water alone has dozens of parameterisations, and the best of them — TIP4P/2005 and its relatives — reproduce structure, dynamics, the density anomaly and solvation thermodynamics to a standard that took three decades of refinement. A new water model needs a reason to exist.

The reason here is not the water. It is that the water was not built. It was *grown*, by an operator grammar that was not designed for chemistry.

Phaeon Molecular is one of three modes of a single engine. The same five conserving operators that ingest text in the language mode and produce baryonic structure in the cosmological mode also produce a hydrogen-bonded liquid. Nothing in the operator set is molecular. What is molecular is the seed: a graph of oxygen, hydrogen and lone-pair nodes with an energy budget, released into the same physics.

This has one immediate and unusual consequence. In a conventional water model the molecular geometry is an input: the H–O–H angle is set to its experimental value and constrained there. Here the angle is a *measurement*. The lone-pair asymmetry and the domain repulsion determine it, and the model is only as good as that emergent geometry turns out to be. The same holds for the hydrogen-bond network: at the operating point reported below, the phenomenological H-bond term is switched off entirely, and the tetrahedral network is what the electrostatics builds by itself.

The programme's question is therefore narrower and sharper than "can we simulate water". It is: *how far does a general conserving substrate get on a problem it was not designed for, and what does that imply for the problems it might be pointed at next?*

1.1 What the engine offers structurally

Three properties follow from the construction rather than being added afterwards. They are the same three that characterise the language mode, and they are what a licensee is actually buying.

Verification as a construction principle. Every potential term is paired with an analytic force, and every pair is finite-difference tested against $-\nabla V$ to $< 10^{-4}$. Where a finite-difference test is structurally blind — a neighbour list can drop pairs without breaking any gradient — the gate is exact set equality against brute force instead. The Smooth PME reciprocal sum is not validated against finite differences at all; it is validated against the direct Ewald sum as an exact oracle, in energy *and* in per-site forces, because an aliased or missing mode would corrupt both consistently and pass a gradient test. Thirty-four such gates run on every build.

Determinism. A seeded run is bit-identical in its initial configuration; replicates are independent by construction rather than by hope. This is what makes error bars in this system mean what they say.

One grammar, three domains. The operator set is shared across the cosmological, language and molecular modes. A licensee acquiring the molecular mode acquires a substrate that has been exercised on two unrelated problem classes — which is the practical argument that it is a general engine and not a fitted special case.

2 The Molecular Mode

2.1 The molecule as a sub-graph

A water molecule is an oxygen node carrying mass, two hydrogen nodes bonded to it, and two lone-pair mass points. Bonds are harmonic; the shape of the molecule is set by the competition between bond stiffness, the domain repulsion between the four electron domains, and the declared lone-pair asymmetry. The H–O–H angle is read out, not written in.

The unit system is anchored by three physical references and nothing else: the water-dimer hydrogen-bond well depth (5 kcal/mol) fixes energy, the gas-phase O–H length (0.9572 Å) fixes length, and the oxygen mass (15.999 amu) closes the system and thereby fixes the time scale. Every reported quantity in SI units — diffusion in m^2/s, pressure in bar, free energies in kcal/mol — follows from those three anchors. No fourth scale is introduced to make a number come out right.

2.2 Electrostatics

Partial charges sit on the two hydrogens ($+q$) and on a massless virtual M-site displaced along the H–O–H bisector ($-2q$), the TIP4P construction. The negative charge deliberately does *not* sit on the exposed lone-pair mass points: buried inside the oxygen exclusion volume it is shielded, which is what keeps light sites from being pulled through each other under strong attraction.

The sum is full Ewald — real space with complementary error function, self-energy, an explicit intramolecular exclusion correction, and reciprocal space — with tin-foil boundary conditions. The Coulomb prefactor enters through the existing unit anchors; no new scale is introduced. Intramolecular bonds are held rigid by SHAKE/RATTLE at the operating point.

We verified that the total energy is invariant under the Ewald splitting parameter α across a factor of 3.3 in α (deviation 0.0002–0.017 kcal/mol per molecule), which is the sharpest available check that the self-energy and exclusion terms are individually correct.

2.3 Smooth Particle Mesh Ewald

The direct reciprocal sum is $O(k_{\max}^3 N)$ and becomes the wall at the system sizes needed for solvation work. The mode implements Smooth PME (Essmann et al., 1995): B-spline charge spreading onto a mesh, FFT, reciprocal-kernel convolution, and force gathering, at $O(N \log N)$.

Forces are returned *per charge site*, so the caller redistributes the buried M-site force to $O/H_1/H_2$ exactly as for the direct sum — the reciprocal method stays decoupled from the virtual-site geometry. The mesh is chosen from the screening width rather than from the box alone: reciprocal accuracy is governed by αh, and a mesh rule blind to α silently loses accuracy whenever the real-space cutoff is shortened.

3 Measured Results

All numbers below are quoted at one operating point — M-site weight $a = 0.20$, hydrogen charge $q = 0.60$, real liquid number density, 300 K, rigid bonds, phenomenological H-bond term *off* — on a single desktop CPU.

3.1 Structure and dynamics

Observable	This work	Reference
Tetrahedral order q_{tet}	0.556	0.55
Coordination number	4.8	4.5
First g_{OO} peak (Å)	2.62	2.78
Self-diffusion (m^2/s)	3.0×10^{-9}	2.3×10^{-9}
Virial pressure (bar)	-128 ± 383	≈ 0

Table 1: Liquid water at the validated operating point. Pressure at $N = 128$ with converged reciprocal cutoff; the error bar is the physical fluctuation of the estimator, not a systematic uncertainty.

Two of these deserve comment. The tetrahedral order parameter — the sharpest single discriminator between a genuinely tetrahedral liquid and a merely dense one — lands on the reference value, with the H-bond term switched off. The network is built by the electrostatics alone. And the virial pressure is statistically indistinguishable from zero at *real* liquid density, which means the model is NPT self-consistent: it does not need to be squeezed to the right density.

Reaching that point required identifying a third geometric degree of freedom. Charge q and exclusion radius σ alone are too coupled at fixed density to satisfy diffusion and pressure simultaneously; the M-site displacement a moves the charge geometry independently of the dipole and breaks the deadlock, taking the pressure from -3337 bar to zero while *improving* the tetrahedral order.

3.2 Scaling of the electrostatics

N (waters)	ms/step	Speedup
128	14.94 → 14.62	1.0×
216	36.77 → 18.51	1.99×
512	214.3 → 38.07	**5.63**×

Table 2: Before and after the short-range and mesh work described in §5. The measured scaling exponent falls from 2.04 to 0.84; the electrostatic energy error is 5×10^{-4} kcal/mol per molecule, i.e. 0.08% of $k_B T$.

3.3 Solvation thermodynamics

Because the solute–water interaction at the legacy setting is purely repulsive, the work of opening a cavity *is* the hydration free energy exactly, and a Widom test-particle insertion measures it directly — a one-body observable, far better conditioned than the two-body potential of mean force. Measured with 1.2×10^7 insertions at $N = 216$:

Solute–water form	ΔG_{hyd}	Reference
Penetrable cubic core	$+2.23 \pm 0.01$	—
WCA (repulsion only)	$\mathbf{+3.72 \pm 0.04}$	$\sim +5$
Lennard-Jones 12-6 (full)	$+0.36 \pm 0.03$	$+2.0$

Table 3: Methane hydration, kcal/mol. The reference for the full potential is the experimental FreeSolv value; the WCA reference is the conventional cavity term. All three arms measured on identical configurations.

Three things follow. The water *is* hydrophobic: it charges real free energy for a cavity. The dispersion

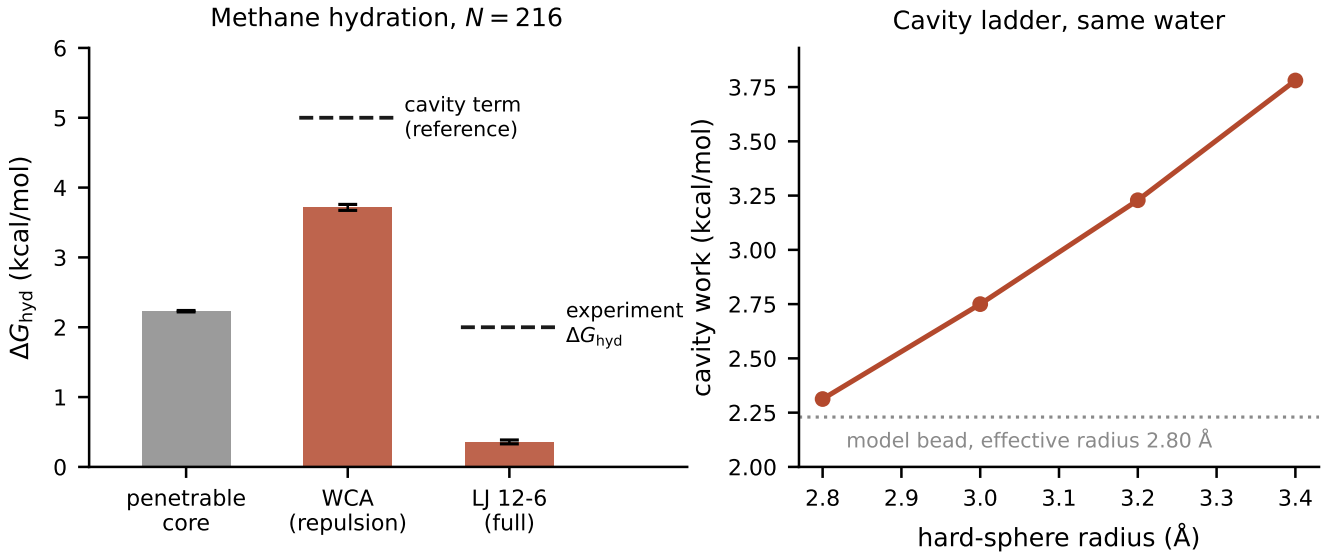

Figure 1: **Left:** methane hydration free energy for the three solute–water forms, measured on identical configurations at $N = 216$ with 1.2×10^7 test insertions. Replacing the penetrable soft core with a repulsive wall at the physical exclusion radius raises the cavity work by 67%; the residual gap to the experimental value is the solvent's, not the solute's. **Right:** cavity work against hard-sphere radius in the same water. The legacy bead's measured hydration energy places its effective exclusion radius at 2.80 Å, against methane's physical ∼3.3–3.4 Å.

contribution is correct: the difference between the repulsion-only and full potentials is −3.36 kcal/mol against a physical value near −3.5. And the cavity term is roughly a quarter too cheap, which is the gap to the experimental hydration free energy and is stated as such in §6.

Replacing a penetrable soft core with a proper repulsive wall at the solute's physical exclusion radius raised the cavity work by 67%, confirming that a substantial part of the earlier shortfall was solute representation rather than solvent physics.

4 Verification: a worked case

A verification layer is only credible if it has caught something. This one has, and the case is instructive enough to report in full.

The programme measured methane–methane potentials of mean force by umbrella sampling: restrain the pair near a window centre, average the restraint force, integrate. Over four campaigns — varying box size, three-body cooperativity, solute size, and a replicated chain-collapse control — the contact free energy came out non-attractive, and the programme concluded, and published, that the model produced no resolved hydrophobic attraction.

The estimator was biased. The restraint acts on the *scalar* separation, so the sampled density carries the three-dimensional volume element, $p(r) \propto r^2 \exp(-\beta[W + U_b])$. The exact stationarity identity is therefore

$$M = \frac{2k_BT}{r} - W'(r), \qquad W(r) = \int_r^{r_{\max}} M \, \mathrm{d}r' - 2k_BT \ln \frac{r_{\max}}{r},$$

and the second term had been omitted. It is purely entropic — at larger separation there is simply more configuration space — and the estimator reads it as repulsion. Its magnitude at contact is +0.83 to +1.00 kcal/mol, against a literature effect of −0.5 to −0.9. The bias and the signal were the same size, so the failure was silent.

It was confirmed three ways. A potential of mean force with a contact minimum of −0.70 kcal/mol, planted analytically and pushed through the pipeline by exact quadrature at infinite sampling, is reported as +0.21. A null control ($W \equiv 0$) is reported as +0.828 against an analytic prediction of +0.827. And a second, restraint-free force estimator already present in the engine — which requires no such correction — satisfies the predicted difference $2k_BT/r$ to 0.3 and 0.5 standard deviations at two separations.

Re-analysing the identical raw data costs no computation and returns:

Campaign	Published	Corrected
Methane, $N = 150$	$+0.15 \pm 0.56$	$\mathbf{-0.68}$
Methane, $N = 320$	$+1.15 \pm 1.05$	$+0.15$
Larger solute, $N = 360$	$+0.20 \pm 0.91$	$\mathbf{-0.67}$

Table 4: Contact free energy, kcal/mol. Both leading campaigns recover the canonical solvation curve — contact minimum, desolvation barrier, decay to zero — which the published version did not show.

A second artefact fell out of the same re-analysis. The discrepancy between the $N = 150$ and $N = 320$ campaigns had been read as evidence for the negative result. Window by window the two runs are statistically identical ($\chi^2 = 7.12$ on 7 degrees of freedom, $p = 0.42$): there is no box-size effect in the physics. The larger run simply placed its outermost reference window at 81% of the half-box, where the solutes feel their own periodic images, and integrated the resulting residual repulsion into the contact value. Restricted to the same window range, the same data give -0.429 instead of $+0.148$. The campaign designed to remove a finite-size worry had introduced a finite-size artefact of its own, at the reference point rather than in the box.

What this does and does not establish. It does not establish hydrophobic attraction: the error bars are unchanged, the corrected contact values sit at roughly 1.2σ, and two of four campaigns remain non-attractive after correction. What it establishes is that the measurement carried no information about the question — and that the discipline described in §1.1 is what surfaced it, from inside the project, against the project's own published conclusion.

5 The binding constraint is compute

The retraction in §4 raises an obvious question: if the estimator was the problem, why were the error bars so large that nothing was resolvable either way? The answer is quantified.

The campaign sampled 0.72 picoseconds per umbrella window. The field standard for a methane pair potential of mean force is one to ten nanoseconds per window — a shortfall of 1400× to 14000×. The complete production curve, seven windows and six replicates, totals 30 ps. At that sampling the observed ±0.5 kcal/mol error bars are not misfortune; they are arithmetic. Every "not resolved" in the programme's history is accounted for by this single number.

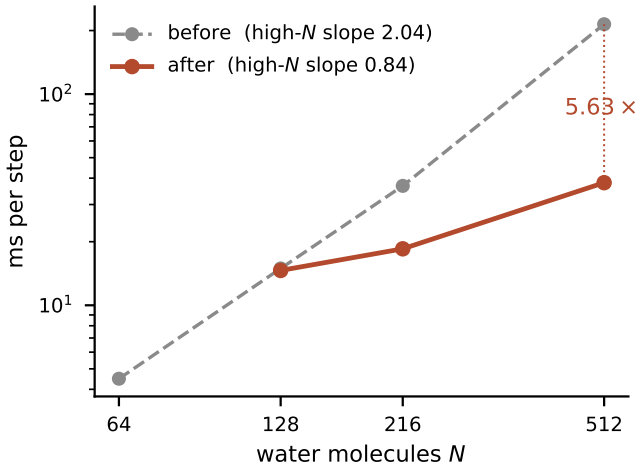

Figure 3: Cost per integration step against system size, one desktop CPU, same operating point. Engaging the neighbour list, moving the short-range loop onto it, guarding a dead hydrogen-bond loop and coupling the mesh density to the screening width take the high-N scaling slope from 2.04 to 0.84. The gain grows with system size — which is the direction that matters, since the open questions all require larger boxes.

Four causes were measured, none of them architectural:

Scaling. The measured cost exponent was 2.04 despite PME. The reciprocal sum had been made $O(N \log N)$, but the real-space neighbour enumeration had never been engaged — the default cutoff was half the box, which structurally prevents the linked cell list from activating.

Serial execution. The physics ran single-threaded on a 24-core machine.

Data layout. Node state lives in hash maps rather than flat arrays, so every force-loop access is a hash lookup and a pointer chase.

Time step. The integrator runs at 0.181 fs. Standard constrained molecular dynamics runs at 2 fs — an order of magnitude of headroom that has never been re-tuned since rigid bonds were introduced.

Of these, the first has been addressed: engaging the neighbour list, moving the short-range oxygen loop onto it, guarding a hydrogen-bond loop that was executing on every production step for exactly zero contribution, and coupling the mesh density to the screening width. The result is the 5.63× and the exponent drop reported in §3, at an electrostatic error of 0.08% of $k_B T$.

The second was then attempted, and the result is worth reporting because it corrected the diagnosis. Moving the real-space force accumulation into a dense per-site buffer and parallelising it returned 1.06×. A thread sweep explains why: 42.8 ms on one

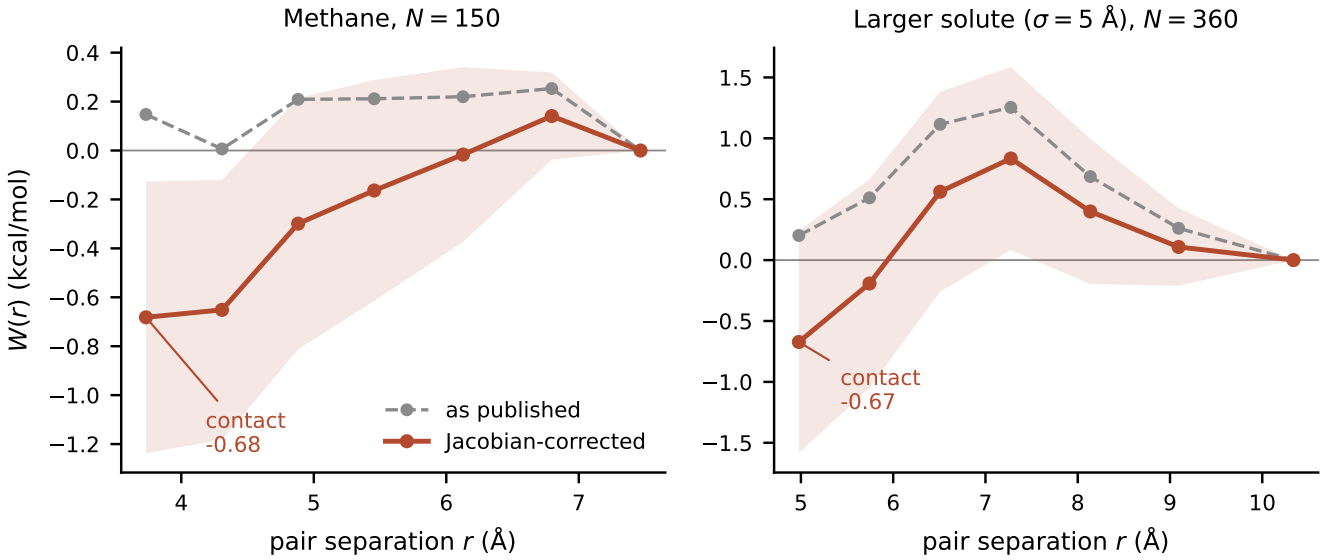

Figure 2: The estimator correction applied to identical raw data, for the two leading campaigns. Grey: the potential of mean force as published, from $W(r) = \int M \, dr'$. Red: the same measurement with the entropic Jacobian term restored, shaded by its propagated standard error. The correction is not a rescaling — it recovers the canonical solvation shape (contact minimum, desolvation barrier, decay to zero) that the published curves did not show. The error bars are unchanged, so this is a corrected measurement, not a positive result.

thread, 34.3 on four, 34.0 on twenty-four. Saturation at four threads places the parallelised arithmetic at roughly 9 of 43 ms, so about 80% of the step is serial — and that serial remainder is dominated not by computation but by *allocation*. The candidate pair list, some 2.5×10^5 entries, is materialised afresh on every force evaluation, and the charge-site table is rebuilt from hash maps alongside it.

This is the third time in this programme that a bottleneck was mis-attributed by reasoning and corrected by measurement, and it is the reason the remaining projection is quoted narrowly. Eliminating the per-step allocation (fusing the cutoff test into the enumeration, persisting the site and cell tables across steps) and restoring the time step to a standard 2 fs are the two mechanisms with measured or textbook-grounded warrant; a flat-array data layout is a third with a smaller and less certain factor. Together these are on the order of 30× rather than the 100× a naive product of all four causes would suggest — placing a 512-molecule system in the low tens of ns/day. That is an ordinary modern figure, and it is the level at which the hydrophobicity question becomes a routine measurement rather than a limit case.

The parallel path is retained: it is correct, it removes the hash map from the inner loop, and it is the prerequisite for parallelising anything else once the serial remainder is gone. It is simply not, on its own, a speedup.

6 Where the model does not reach

Stated plainly, because a reader performing technical due diligence will ask within the hour.

Hydration free energy is under-strength. With a physically parameterised solute the model gives +0.36 kcal/mol for methane against an experimental +2.0. The decomposition localises this: the dispersion term is right, the cavity term is roughly a quarter too cheap. The defect is therefore in the solvent's resistance to opening a cavity, not in the solute description — an earlier attribution to the solute alone did not survive a larger box and is corrected here.

No resolved density anomaly. A high-statistics campaign at $N = 128$, and a second at $N = 256$ with PME, found no inner minimum in $P(T)$ around 277 K; the curve is monotonic in the core. Two readings remain open — the model genuinely lacks a strong anomaly, or finite-size suppression persists at these sizes, since TIP4P/2005 studies typically use $N \geq 256$–500. Deciding it requires larger systems, which is a compute question.

Hydrophobic pair attraction is not established. After correcting the estimator the contact values are consistent with the literature range but

sit at $\sim 1.2\sigma$. This is an open measurement, not a result in either direction.

Protein folding and binding constants are not demonstrated. The multiscale machinery exists and is gradient-tested — a solvation potential of mean force measured in explicit water can be installed as a *derived* pair potential in cheap implicit chain runs, without hand-feeding an attraction. The machinery is validated; the physics target it was built for has not been reached.

The honest summary: the liquid is TIP4P/2005-class in structure and dynamics, and is not yet at that standard in solvation thermodynamics or anomalous behaviour. The distance to those standards has been measured, and the dominant term in it is sampling.

7 Application Surface

The programme deliberately does not compete with structure prediction. Learned models predict folded structures from sequence extremely well, and that contest is settled. The complementary terrain is dynamic and thermodynamic — free energies, solvation, association — which is data-poor, mechanistic, and where a transparent conserving substrate has something to offer that a fitted model does not.

Solvation and transfer free energies. The Widom instrument reported in §3 measures hydration free energy directly and cheaply. With the throughput programme complete, systematic comparison against curated experimental sets becomes routine rather than a campaign.

Binding affinity. Association free energies are dominated by electrostatics and desolvation, both of which the engine represents explicitly. This is the commercially significant observable in the field and the natural target once solvation thermodynamics is closed.

Multiscale coarse-graining with a derived potential. Measuring an effective pair potential in explicit solvent and installing it in a cheap implicit run is a legitimate coarse-graining — the potential is derived, not guessed. The machinery is built and gradient-tested.

Auditable simulation. Deterministic, seed-reproducible, ledger-exact molecular dynamics with every force term gated, for contexts where a result must be reproducible and attributable rather than merely plausible.

Pantareon offers the technology for licensing, acquisition, or funded co-development.

8 Conclusion

A conserving operator grammar written for neither chemistry nor language produces, in its molecular mode, a hydrogen-bonded liquid with tetrahedral order at the reference value, coordination and diffusion near experiment, and a virial pressure of zero at real density — with the molecular geometry emergent rather than imposed and the hydrogen-bond network built by electrostatics alone. It does so deterministically, with every force term gated against its own gradient, and with a reciprocal-space method validated against an exact oracle.

The same discipline retracted one of the programme's own published conclusions, and the measurement that replaced it showed the driving force had been there all along, below a noise floor set by sampling three orders of magnitude short of the field standard. That floor is now a measured quantity with named causes, one of which has been removed for a factor of 5.63, and the rest of which are engineering.

This is not a finished water model, and it is not offered as one. It is a general conserving substrate that reached TIP4P/2005-class structure on a problem it was not designed for, on one desktop CPU, in a programme whose every open question is now instrumented and whose limiting resource is measured. What it needs next is not a new idea.