Aim Materials: Physics-Aware Crystal Generation and Staged DFT Validation with an E(3)-Equivariant Graph Neural Network Variational Autoencoder

Sunwoo Lee The Ohio State University Corresponding author: lee.11539@buckeyemail.osu.edu

Abstract

Generative crystal pipelines can produce many structurally plausible candidates, but many fail under strict first-principles elastic validation, limiting useful screening yield. We present Aim Materials, a reproducible workflow that couples E(3)-equivariant crystal generation [6] with Voigt-space elastic constraint screening and staged Quantum ESPRESSO validation with recoverable run-state logging [3-5]. The workflow is designed to filter mechanically implausible candidates earlier, preserve stage-resolved failure information, and keep generated-scale telemetry separate from completed DFT evidence. In the refreshed snapshot used here, public records contain 271 generated structures and 271 MPContribs contributions (271 with structures) [1], while the active DFT screening cohort contains N=294 candidates with pending=246, relax_done=0, scf_done=3, elastic_ready=10, and output_present_not_converged=35. All 271 generated structures are parseable and structure-unique under the current serialization and StructureMatcher settings, and the formula-level novelty rate versus the 6056-entry training corpus is 0.9742. In the de-duplicated elastic-ready subset (n=10), 7 candidates satisfy the fit criterion, 6 satisfy positive-definiteness, and 5 satisfy both (rate=0.5000, Wilson 95% CI [0.2366, 0.7634]). The strongest case is GeCl3 (2el/015_gen_00073_GeCl3), with B_H=27.69 GPa, G_H=11.04 GPa, E_H=29.23 GPa, and nu=0.324. A QE-to-Materials Project elemental-reference alignment yields indicative candidate-level thermodynamic estimates for the seven QE-backed elastic-ready entries, and none falls within 0.1 eV/atom of the convex hull. Three legacy non-QE elastic-ready entries remain thermodynamically unavailable under the same reference workflow. We therefore present Aim Materials as a method-centered, compute-aware screening framework with explicit evidence boundaries. Current conclusions remain limited by incomplete DFT coverage, a still modest fully validated subset, and the absence of a reviewer-facing control baseline.

Keywords

E(3)-equivariant graph neural network variational autoencoder; crystal generation; elastic tensor constraints; Quantum ESPRESSO; high-throughput DFT screening; materials informatics

1. Introduction

AI-assisted materials discovery is increasingly moving from isolated prediction tasks to integrated workflows that generate, rank, and validate candidate structures under practical computational constraints. In this setting, structural plausibility alone is insufficient. Many generated crystals that satisfy geometric or compositional checks fail under strict first-principles evaluation of their elastic response. This gap is important because the scientific value of a generative pipeline depends not only on how many candidates it proposes, but also on how many survive physically grounded downstream screening.

A recurring weakness in generative materials workflows is that elastic plausibility is treated as a late-stage filter rather than as an early design objective. When this happens, computational effort is spent on candidates that are unlikely to survive detailed validation. Later-stage throughput declines, the validated subset remains small, and the resulting scientific interpretation becomes difficult to support. For Computational Materials Science, that mismatch is especially problematic because the main contribution is expected to be computationally rigorous, reproducible, and physically interpretable.

Recent crystal-generation studies have used latent-variable and diffusion formulations to improve structural validity, diversity, and conditional control, including CDVAE [17], DiffCSP [18], and symmetry-guided generative approaches [19]. Earlier graph-based materials models such as CGCNN [7] and MEGNet [8] established strong crystal-property prediction baselines, while benchmark suites such as Matbench [10] made reproducible comparison central to the field. Geometry-aware downstream surrogates such as CHGNet [9] and E(3)-equivariant interatomic models in the NequIP family [12] further broadened the practical screening toolkit available after generation. Those studies establish that learned crystal generators can sample chemically plausible structures at scale, but they also leave an important practical question open: how well do the generated candidates survive downstream first-principles screening when elastic plausibility is treated as a design target rather than as an after-the-fact filter? Aim Materials is positioned against that question. The contribution claimed here is therefore narrower than a general-purpose generation benchmark; it is a physics-aware screening workflow that couples generation, elastic triage, staged DFT validation, and parser-backed evidence regeneration under one reproducible campaign.

Aim Materials was developed to address this problem by coupling crystal generation and elastic plausibility more tightly. The workflow combines an E(3)-equivariant graph neural network variational autoencoder with Voigt-space elastic constraint screening and staged Quantum ESPRESSO validation. In parallel, the study enforces explicit separation between public project-scale telemetry and completed DFT evidence. Public records report what has been generated and published to MPContribs [1,11], while manuscript claims are restricted to the smaller subset that has actually completed later validation stages. This scope separation is treated here as part of the computational method rather than as a reporting convenience.

The second distinguishing feature of Aim Materials is workflow instrumentation. Long-running relax, SCF, and elastic calculations frequently fail because of numerical instability, restart conditions, memory limits, or queue interruptions. In many campaigns, those failures are difficult to audit after the fact. Aim Materials instead records stage-resolved state so that failure modes, partial outputs, and rerun priorities remain visible. The workflow is therefore designed not simply to generate candidates, but to produce a more transparent and recoverable path from candidate generation to first-principles evidence.

The present manuscript is written as a methods paper. Its aim is to show that physically informed candidate screening, conservative evidence accounting, and auditable DFT staging can be integrated into a single reproducible workflow. It does not attempt to claim broad validated-materials discovery at the current snapshot. Instead, it evaluates the current evidence ceiling and discusses what the available validation depth can and cannot support.

2. Computational methodology

2.1 Workflow overview

The Aim Materials workflow consists of four connected stages: candidate generation, elastic screening, staged DFT validation, and parser-backed evidence regeneration. First, an E(3)-equivariant graph neural network variational autoencoder generates candidate periodic crystal structures from composition-aware inputs [6]. Second, a physics-aware elastic screening stage evaluates candidates in Voigt form and suppresses mechanically implausible outputs before expensive first-principles calculations are launched. Third, shortlisted structures enter a staged Quantum ESPRESSO workflow consisting of structural relaxation, SCF convergence, and elastic calculations [3-5]. Finally, parser and snapshot scripts regenerate the manuscript-facing counts, tables, and derived artifacts directly from the campaign outputs.

This workflow was designed to solve two related problems. Operationally, it supports triage, restart, and rerun prioritization during a live campaign. Scientifically, it ensures that manuscript values are tied to versioned parser outputs rather than to hand-updated spreadsheets. The paper therefore depends on a frozen snapshot policy rather than on manually maintained reporting.

2.2 E(3)-equivariant generation and elastic constraint screening

Each crystal is represented as a periodic graph in which nodes encode elemental identity and edges encode local geometric relationships under periodic boundary conditions. The E(3)-equivariant representation is used so that rotations, translations, and related geometric changes are handled consistently within the model. This is important because elastic behavior depends strongly on geometry, symmetry, and local environment rather than on composition alone. Graph-level latent descriptors are assembled with motif pooling, used here in the concrete architectural sense of concatenating mean pooling, max pooling, and an attention-weighted sum over node embeddings for each periodic graph before decoding and property prediction.

The elastic screening stage operates in Voigt form and applies explicit constraints intended to reduce physically inadmissible predictions. The purpose of these constraints is not to guarantee that every surviving candidate will pass DFT validation. The purpose is to reduce the number of candidates that are mechanically inconsistent before they enter the most computationally expensive part of the workflow. In practice, the model aims to improve candidate triage rather than to replace first-principles validation.

Training combines generative reconstruction and latent regularization with property-aware and physical-consistency terms. In compact form, the training objective can be written as `L = L_recon + beta_KL L_KL + sum_i lambda_i L_i^phys`, where the physical terms include Voigt-tensor reconstruction, derived-scalar consistency, equivariance, positive-definiteness, symmetry masking, tensor-scalar consistency, and auxiliary class penalties. This balances candidate diversity against constraint compliance. If the constraints are too weak, downstream false positives increase. If they are too strong, generative diversity can be over-regularized. The present study reports the resulting workflow behavior and uses archived ablation signals to interpret these tradeoffs rather than to claim exhaustive benchmark dominance.

Table 1. Canonical model and training settings used for the publication-anchor retraining run discussed in this manuscript. These values are taken directly from the saved run manifests and training logs rather than from retrospective note taking.

Parameter	Value
Training corpus	6056 filtered crystal graphs assembled into the curated training dataset
Split protocol	Leakage-aware group split using `(num_elements, crystal_system, num_nodes, num_edges)` as the proxy key
Train/validation/test split	4774 / 673 / 609 entries across 1123 groups
Core architecture	EGNN-based graph variational autoencoder with periodic graph inputs and motif pooling
Message-passing layers	4
Hidden dimension	128
Latent dimension	64
Maximum atoms / species	64 atoms / 92 species
Optimizer	AdamW, weight decay = `1e-4`
Initial learning rate	`2e-4`
Batch size	8
Checkpoint metric	`val_voigt_balanced`
Canonical retraining run	80 epochs, best checkpoint at epoch 71
Random seeds	global seed = 42, split seed = 42
Loss profile	Manual tensor-focused profile: `w_pred=1.0`, `w_voigt=0.35`, `w_voigt_derived=1.1`, `w_voigt_eq=1.8`, `w_voigt_pd=1.0`, `w_voigt_sym=0.9`, `w_consistency=0.25`, `w_scalar_pos=0.1`, `w_tensor_scalar_cons=0.8`, `w_class=0.1`, `w_kld=0.05` with linear KL annealing over the first 30 epochs

2.3 Candidate ranking and staged DFT validation

Generated candidates are ranked using a composite screening logic that combines model-side plausibility signals with downstream feasibility considerations. The objective is to spend limited DFT budget on candidates with a higher probability of yielding interpretable elastic results. Candidate ranking is therefore part of the computational method itself rather than a cosmetic post-processing step.

For the current local QE campaign builder, the primary selection stage required `strict_pass = True`, `formula_novel_vs_train = True`, `scalar_std_mean <= 0.05`, `voigt_std_mean <= 0.05`, and `chgnet_force_max <= 0.25 eV/Angstrom` when a CHGNet force estimate was available. Candidates passing those gates were ranked by the deterministic heuristic `S = q + 0.15 d_min - 40 sigma_scalar - 30 sigma_voigt - 0.8 F_max`, where `q` is `quality_score`, `d_min` is the minimum interatomic distance in angstrom, `sigma_scalar` and `sigma_voigt` are the ensemble uncertainty summaries for scalar and Voigt outputs, and `F_max` is the maximum CHGNet force estimate. The CHGNet-based force gate is used here only as a low-cost feasibility signal rather than as a replacement for DFT validation [9]. The current rerun builder then selected the top 24 entries per element-count subset with a cap of two entries per reduced formula. Because the publication snapshot merges local reruns with archived results, not every historical row was chosen by this exact rule, but this explicit heuristic defines the ranking logic for the present local QE campaign and the current rerun priorities.

DFT validation follows a staged policy. Candidates first undergo structural relaxation, then SCF convergence, and then elastic calculations if the earlier stages complete successfully. Non-converged outputs are retained and classified rather than silently removed. This retention matters because unresolved calculations still contain methodological information: they identify where throughput is lost, where restart policy matters, and which candidates remain plausible rerun targets.

2.4 DFT campaign and snapshot protocol

All publication-facing values in this manuscript are regenerated from parser outputs and synchronized campaign files. The refreshed snapshot used here was regenerated on 2026-03-09 from the merged local and archived campaign outputs. Within that snapshot, project-scale telemetry and DFT-stage evidence are reported separately. This separation is essential because public counts describe workflow scale, whereas elastic-ready outcomes describe actual validated evidence depth.

The active DFT cohort for the manuscript contains 294 candidates. Stage counters in the refreshed snapshot report 246 pending entries, 3 scf_done entries, 10 elastic_ready entries, and 35 entries with output_present_not_converged. These stage-resolved counts, together with the de-duplicated elastic-ready subset, define the current evidence ceiling. The underlying scripts and structured artifacts are part of the reproducibility baseline for the submission.

Because the campaign merges successful local reruns and archived historical outputs, the QE inputs are not represented by one single parameter file. Table 2 therefore reports representative successful settings observed in current SCF and elastic inputs rather than claiming that every archived run used identical numerical parameters.

Table 2. Representative Quantum ESPRESSO settings observed in successful SCF and elastic inputs from the current campaign archives.

Parameter	Representative setting
DFT engine	Quantum ESPRESSO `pw.x` [3-5]
Exchange-correlation functional	PBE-GGA [14]
Pseudopotentials	Legacy local UPF pseudopotential set applied consistently by element within each run; the current QE-backed thermo and elastic results use A. Dal Corso espresso-distribution PAW files for `Ge`, `H`, `I`, and `O`, and a Vanderbilt ultrasoft file for `Cl`; candidate/reference thermodynamic offsets were computed with the same element-specific pseudopotential choices used in the corresponding candidate calculations
Occupations / smearing	Smearing with Marzari-Vanderbilt (`mv`), `degauss = 0.02`
Wavefunction cutoff	`ecutwfc = 50 Ry` in representative successful SCF and elastic inputs
Charge-density cutoff	`ecutrho = 400 Ry` in representative successful SCF and elastic inputs
Electronic convergence threshold	`conv_thr = 1e-9` to `1e-10`
Mixing parameter	`mixing_beta = 0.15` to `0.40`, depending on candidate and stage
K-point policy	Automatic Monkhorst-Pack meshes [13]; representative successful GeI3 and GeCl3 inputs used `6 x 6 x 6` meshes
Maximum SCF steps	`electron_maxstep = 300`
Elastic protocol	Finite-strain SCF evaluations staged under `03_elastic/strain_*`; representative strain jobs disable symmetry reduction with `nosym = .TRUE.` and `noinv = .TRUE.`
Relaxation thresholds	Representative successful QE XML outputs record `etot_conv_thr = 5e-5`, `forc_conv_thr = 5e-4`, and `press_conv_thr = 0.5`
Frozen manuscript snapshot	`2026-03-09T15:46:27.366756+09:00`

3. Results and discussion

3.1 Generated candidate statistics

The public generation layer of the workflow currently contains 271 saved structures and 271 corresponding MPContribs entries [1,11]. To convert that generation count into manuscript-facing evidence, the structures were reloaded from their serialized pymatgen dictionaries and grouped with StructureMatcher using the same tolerance policy applied during manuscript regeneration. Under those settings, all 271 generated structures remain parseable and all 271 remain structure-unique. The generated set spans 231 reduced formulas, 228 of which are absent from the reduced-formula set assembled for the 6056-entry training corpus. The resulting formula-level novelty rate is therefore 0.9742.

Table 3 reports the generated-candidate metrics used in the manuscript. These values should be read as parseability, deduplication, and training-corpus novelty diagnostics rather than as thermodynamic or application-specific novelty claims. In particular, the reported novelty is formula-level novelty versus the reduced-formula set of the training corpus, not polymorph-level novelty relative to the full Materials Project.

Table 3. Generated-candidate statistics derived from the 271 published structures and the 6056-entry training corpus used for the canonical retraining run.

Metric	Value	Definition
Generated structures	271	Total records in the published Aim Materials candidate set
Parseable structures	271	Structures successfully reconstructed from the saved pymatgen dictionaries
Validity rate	1.0000	Parseable structures / generated structures
Unique structures	271	StructureMatcher groups with one representative per group
Uniqueness rate	1.0000	Unique structures / parseable structures
Unique formulas	231	Distinct reduced formulas across parseable generated structures
Formula-novel structures	264	Generated structures whose reduced formulas are absent from the 6056-entry training formula set
Formula-level novelty rate	0.9742	Formula-level novelty versus the reduced-formula set of the training corpus
Formula-novel unique formulas	228	Distinct reduced formulas not present in the training formula set

3.2 Screening cohort and stage-resolved outcomes

The central result of the present study is a transparent separation between public project activity and completed DFT evidence. At the public telemetry level, the project has produced 271 generated structures and 271 MPContribs contributions [1,11]. At the active DFT level, the manuscript is restricted to the 294-candidate screening cohort described above. This separation prevents generation-scale and validation-scale counts from being merged into a single, misleading narrative.

Figure 1 summarizes the stage-resolved outcomes by subset in the refreshed snapshot. The dominant pattern is attrition between generated-scale activity and elastic-ready evidence, with most unresolved entries remaining in `pending` or `output_present_not_converged` states rather than in the final elastic-

ready pool. The refreshed snapshot contains 10 elastic-ready entries, 35 output-present-but-not-converged entries, and 246 still-pending entries. This visualization makes clear that the current bottleneck is downstream first-principles completion rather than the generation count itself.

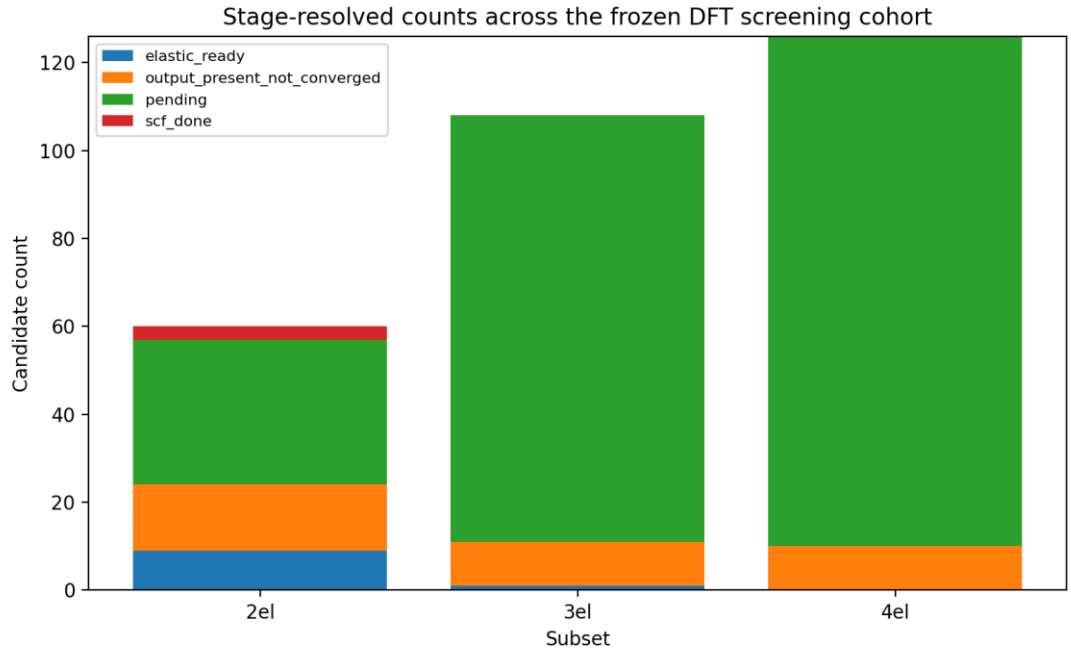

Figure 1. Stage-resolved counts across the refreshed DFT screening cohort, grouped by subset. The figure highlights the large attrition between public-scale candidate generation and the much smaller elastic-ready evidence set used in the manuscript.

The current evidence should therefore be read as stage-limited. The workflow is operational across generation, ranking, validation, and parsing, but the number of fully validated elastic cases remains modest. This does not mean the generative pipeline failed. It means that convergence economics, rerun burden, and compute limits still dominate the size of the final validated subset. That distinction is central to the present manuscript.

3.3 De-duplicated elastic-ready subset

The de-duplicated elastic-ready subset now contains 10 entries. De-duplication uses a combined `campaign_dir + candidate_relpath` key inside the parser artifacts to avoid relpath collisions across merged campaign sources, thereby preventing the same candidate from being counted twice through rerun archives. Among these 10 entries, 7 satisfy the elastic fit criterion, 6 satisfy positive-definiteness, and 5 satisfy both. The resulting pass_both rate is 0.5000 with a Wilson 95% confidence interval of [0.2366, 0.7634]. The interval remains broad, but it is narrower than the earlier seven-entry snapshot and now reflects a larger locally completed subset.

Table 4. De-duplicated elastic-ready subset reported with structural descriptors, elastic outputs, and thermodynamic values where a uniform QE-backed reference workflow is available. `N/A`

thermodynamic entries correspond to legacy non-QE rows that cannot yet be placed on the same energy scale.

Candidate	SG	Density (g/cm3)	Fit/PD	B_H (GPa)	G_H (GPa)	E_H (GPa)	nu_H	E_form (eV/atom)	E_hull (eV/atom)
Pd3N5 (1)	P1	2.91862	No / No	-26.3341	16.4691	62.4194	0.895048	N/A	N/A
Cs3N5 (2)	Cm	4.37448	No / No	24.2267	-4.94035	-15.9020	0.609397	N/A	N/A
Ge3O5 (6)	P1	3.81602	No / Yes	130.5730	65.3072	167.9250	0.285656	-0.9344	1.0129
GeI3 (11)	P1	2.68419	Yes / Yes	10.8552	4.84571	12.6542	0.305712	-0.0830	0.5062
GeI3 (13)	P1	2.93646	Yes / Yes	12.0781	4.51045	12.0334	0.333950	-0.1395	0.4498
H3I5 (14)	P1	2.56729	Yes / No	9.77405	-10.0813	-46.0901	1.285930	0.1300	0.3555
GeCl3 (15)	P1	1.72197	Yes / Yes	27.6907	11.0371	29.2281	0.324080	-0.7742	0.5573
GeCl3 (16)	P1	1.82537	Yes / Yes	25.9882	12.8185	33.0257	0.288201	-0.7660	0.5655
HCl (18)	P1	1.46623	Yes / Yes	26.2317	12.1349	31.5409	0.299601	-0.4811	0.3347
ReI5N2 (4)	P1	7.06890	Yes / No	47.3509	-0.646506	-1.94839	0.506858	N/A	N/A

The refreshed elastic-ready subset is dominated by halide-rich low-symmetry candidates. Nine of the ten entries belong to the two-element subset, nine adopt triclinic `P1` symmetry, and one adopts monoclinic `Cm` symmetry. Densities span 1.47-7.07 g/cm3, and the current pass-both subset consists of two GeI3 entries, two GeCl3 entries, and one HCl entry. The GeI3 and GeCl3 duplicates indicate that the validated set is still chemically narrow, while the presence of HCl should be interpreted cautiously. HCl is a chemically simple known binary compound, not a novel materials discovery claim, and in the present manuscript it is better read as a recovered low-complexity sanity-check case than as evidence of broad exploration into new chemistry.

The strong `P1` dominance in the validated subset also needs to be interpreted carefully. The finite-strain elastic workflow disables symmetry reduction during the `03_elastic/strain_*` SCF jobs through `nosym = .TRUE.` and `noinv = .TRUE.`, and the manuscript tables report the final relaxed cells exactly as parsed rather than after a separate symmetrization pass. As a result, some entries that may be close to a higher-symmetry description are still represented conservatively as triclinic cells. This does not rule out a model-side bias toward low-symmetry outputs, but it does mean that the observed `P1` prevalence reflects both generation behavior and the deliberately symmetry-preserving downstream validation protocol.

Figure 2 complements Table 4 by plotting elastic fit RMS across the refreshed validated subset. The pass-both entries occupy the low-residual region, while Pd3N5 and Ge3O5 retain substantially larger residuals despite reaching elastic-ready status. At the same time, H3I5 shows why fit quality cannot be used as a standalone acceptance criterion: although its fit residual is below the current threshold, the derived shear and Young's moduli are negative and the candidate fails the positive-definiteness gate. In other words, the two-criterion filter is doing real work here, because fit-only screening would have retained a mechanically unstable case.

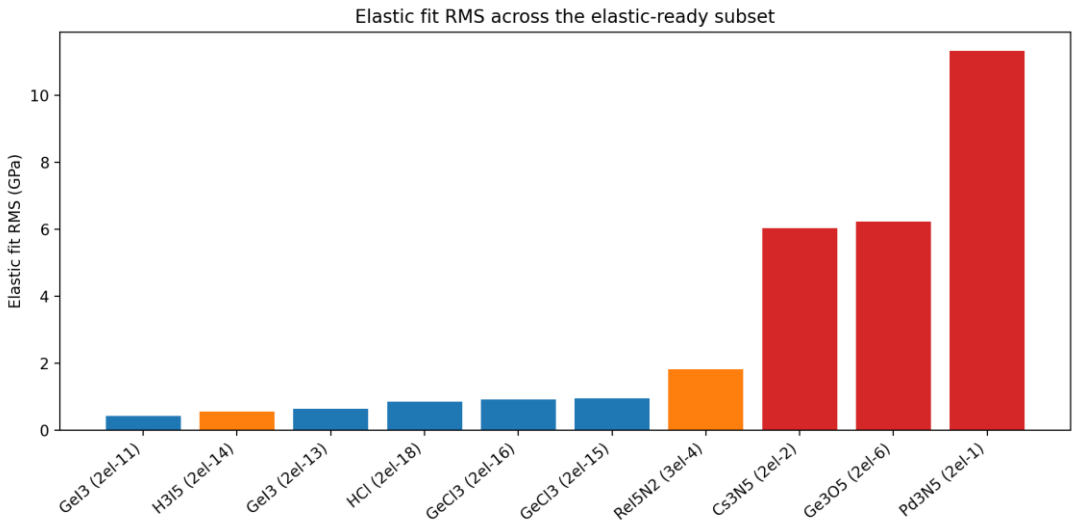

Figure 2. Elastic fit RMS across the de-duplicated elastic-ready subset. Lower values are associated with the entries that satisfy both fit-quality and positive-definiteness criteria.

Figure 3 shows representative structures from the current validated subset using a standardized crystallographic view, explicit element coloring, and bond overlays drawn only for short chemically plausible contacts. The panels are intended to illustrate the recurring low-symmetry halide-rich character of the present pass-both pool rather than to claim that every displayed structure is novel. In particular, the HCl panel is included as a recovered simple binary reference-like case, whereas the GeI3 and GeCl3 panels better represent the more distinctive halide-rich branch of the current validated subset.

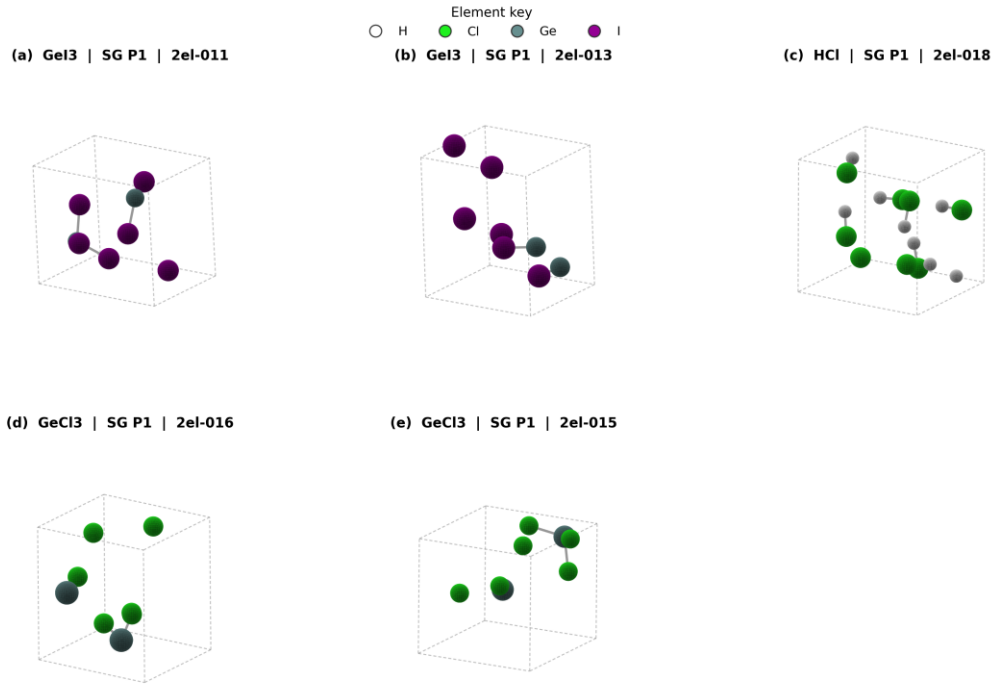

Figure 3. Representative structures from the current validated subset. Panels (a)-(e) use a common viewing orientation, a shared element key, and bond overlays limited to short chemically plausible contacts. The displayed set includes GeI3 and GeCl3 pass-both cases together with HCl as a recovered simple-binary reference-like example.

Figure 4 maps the elastic-ready subset in bulk-shear modulus space. The pass-both entries occupy a low-to-moderate modulus band dominated by GeI3, GeCl3, and HCl, whereas the entries that fail either fit quality or positive-definiteness are displaced toward negative or highly anisotropic regimes. Numbered markers are used in the plot and resolved through a compact side key so that the candidate identities remain readable without label overlap. This view complements the fit-RMS ranking by showing that the current validated pool is not only low-residual, but also mechanically clustered.

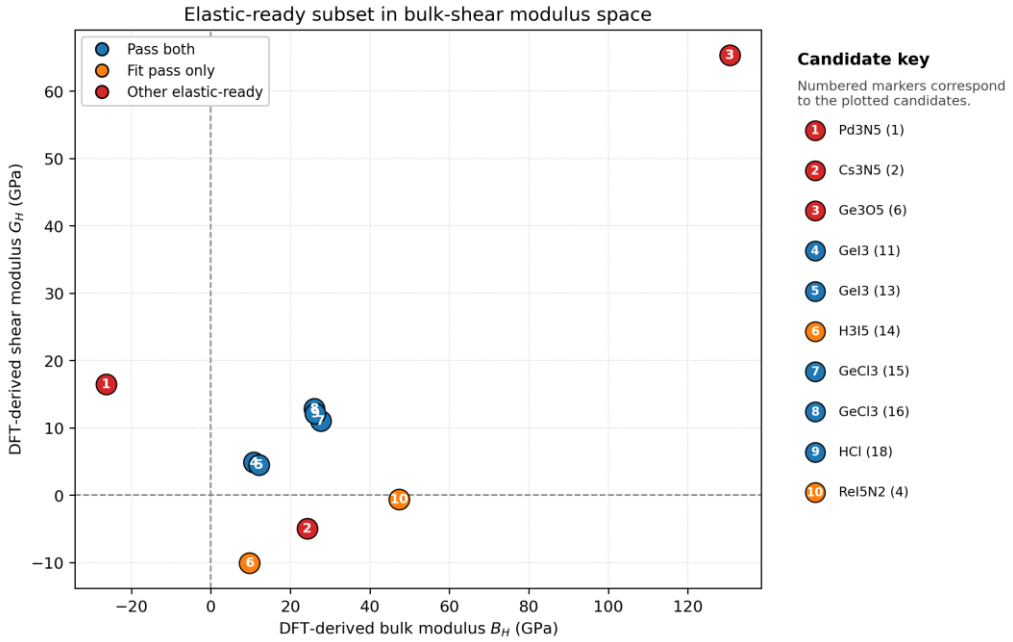

Figure 4. Elastic-ready candidates in bulk-shear modulus space. Numbered markers are resolved through the side key at right. Pass-both entries cluster in a low-to-moderate modulus region, whereas partial or failed elastic-ready entries are displaced toward negative or strongly unstable elastic responses.

3.4 Direct thermodynamic assessment of QE-backed validated candidates

Direct thermodynamic values are now available for the seven elastic-ready entries that retain a uniform QE-backed provenance chain. For these entries, stable elemental reference entries from the Materials Project were re-evaluated with the same element-specific pseudopotential choices and representative QE settings used in the campaign. The resulting per-element offsets were then used to align candidate total energies to the Materials Project elemental-reference scale before computing `E_form` and indicative `E_hull` values against the relevant phase diagrams. Three legacy elastic-ready rows (`Pd3N5`, `Cs3N5`, and `ReI5N2`) originate from older non-QE sources and therefore remain thermodynamically unavailable under the same compatibility workflow.

The thermodynamic results are more restrictive than the elastic results. None of the seven QE-backed elastic-ready entries lies within `0.1 eV/atom` of the convex hull, and the observed hull distances remain in a broad `0.3347-1.0129 eV/atom` range. The two `GeI3` entries yield the lowest-magnitude negative formation energies (`-0.0830` and `-0.1395 eV/atom`) but still sit `0.45-0.51 eV/atom` above hull. The two `GeCl3` pass-both entries likewise remain thermodynamically displaced (`0.5573` and `0.5655 eV/atom` above hull). `HCl` occupies the lowest-hull position inside the present QE-backed subset, but it should still be interpreted as a recovered simple-binary reference case rather than as a novel target chemistry. Because the Cl-containing calculations inherit a mixed PAW/ultrasoft pseudopotential setup from the legacy local library, the reported `E_form` values should be read as internally aligned workflow thermodynamics rather than as a uniform SSSP-benchmarked reference set [20]. The corresponding `E_hull` values are best treated as indicative rather than fully benchmarked, because the convex-hull

competing phases remain on the Materials Project VASP reference scale and single-element QE-to-Materials Project offsets cannot eliminate all phase-level cross-code and cross-pseudopotential compatibility errors. The current combined picture is that elastic plausibility and thermodynamic competitiveness are not yet aligned in the validated subset.

This outcome sharpens the paper's claim rather than weakening it. The present workflow can already enrich the candidate pool for mechanically interpretable elastic responses, but the same screen does not yet produce a near-hull thermodynamic subset. In practical terms, the validated halide-rich pass-both cases are better read as mechanically plausible screening successes than as immediately synthesizable materials predictions. That distinction is important for reviewer interpretation because it shows exactly where the current workflow adds value and where further ranking criteria are still needed.

3.5 Representative validated candidates

The strongest current elastic case is GeCl3 at `2el/015_gen_00073_GeCl3`, with B_H=27.69 GPa, G_H=11.04 GPa, E_H=29.23 GPa, and nu=0.324. The current pass-both subset now also includes two GeI3 entries and one HCl entry, all of which remain low-symmetry halide-rich structures. Combined with the thermodynamic results above, these cases show that the current validated pool is structurally interpretable but still materially short of near-hull stability under the present reference workflow. H3I5 remains the clearest counterexample in the same table: it satisfies the fit threshold but fails positive-definiteness because the derived shear and Young's moduli are negative. That contrast makes the present two-stage elastic acceptance logic easier to interpret than a fit-only ranking.

3.6 Implications and current limitations

At the current snapshot, the strongest defensible conclusion is that Aim Materials provides a reproducible and physically informed workflow that can produce elastic-ready candidates with interpretable mechanical properties while preserving transparent stage accounting. The current evidence does not support a broad validated-materials discovery claim. It does, however, support the workflow claim that coupling generation, elastic constraint screening, staged DFT validation, and parser-driven reporting produces a more auditable computational discovery pipeline.

The present evidence ceiling is set by four practical constraints. First, the DFT screening cohort is still incomplete, so the current manuscript reflects a snapshot rather than a terminal campaign state. Second, the fully validated subset remains small, which limits statistical sharpness and keeps the uncertainty interval around downstream retention broad even after the current local reruns. Third, direct thermodynamic values are currently available only for the seven QE-backed elastic-ready rows; three legacy elastic-ready entries still lack the provenance needed to place them on the same QE-to-Materials Project energy scale. Fourth, the current manuscript still does not include a reviewer-facing external generative baseline or a clean no-Voigt/no-equivariance control suitable as headline evidence of superiority. These limitations narrow the claims that can be defended, but they do not undermine the core methodological contribution.

The present pseudopotential choice is another interpretive constraint. The active QE-backed subset inherits a legacy local UPF library dominated by A. Dal Corso espresso-distribution files, with mixed PAW and Vanderbilt ultrasoft formalisms in Cl-containing chemistries. That setup was held consistent

within each candidate/reference thermodynamic alignment, so the reported formation energies remain internally comparable inside the current workflow. It is not, however, equivalent to a uniform SSSP-quality benchmark thermodynamic protocol [20]. In particular, the indicative hull distances inherit additional uncertainty because the candidate energies are mapped onto Materials Project phase diagrams whose competing phases remain on the VASP reference scale. A stronger thermodynamic study would rerun the validated subset with a single modern benchmarked library before making broader cross-chemistry stability claims.

For Computational Materials Science, this positioning is appropriate. The contribution is strongest where computation, workflow design, and reproducible evidence accounting intersect. The manuscript therefore emphasizes methodology, validation discipline, and explicit scope control rather than a large validated-candidate headline. The updated thermodynamic section also makes the current ranking gap explicit: elastic screening alone is not sufficient to isolate near-hull candidates, so future workflow gains should come from coupling the present elastic filter with a stronger thermodynamic or surrogate-energy ranking stage. Internal control studies performed during manuscript development also showed that raw-pool tensor/scalar consistency is highly sensitive to benchmark definition and calibration choice, so those exploratory comparisons are retained as supporting analysis rather than used as headline evidence of method superiority.

4. Conclusions

Aim Materials provides a physics-aware and reproducible route from crystal generation to staged first-principles elastic validation. The workflow combines an E(3)-equivariant graph neural network variational autoencoder, Voigt-space elastic constraint screening, and recoverable Quantum ESPRESSO execution into a single parser-backed pipeline. The resulting manuscript values are generated from a frozen snapshot rather than from manually aggregated reporting.

Within the current evidence window, the paper supports a method-centered conclusion. The workflow can produce elastic-ready candidates with interpretable elastic properties, and the refreshed snapshot now contains 10 elastic-ready entries with 5 satisfying both fit-quality and positive-definiteness criteria. The manuscript also shows that the generated structure set is fully parseable, fully structure-unique under the current matcher settings, and highly formula-novel relative to the 6056-entry training corpus in the aggregate. A uniform QE-to-Materials Project thermodynamic alignment is now available for the seven QE-backed elastic-ready entries and shows that none of those entries is near-hull under the current reference workflow. At the same time, the manuscript explicitly acknowledges that the fully validated subset remains modest and that unresolved candidates continue to limit the breadth of downstream claims.

The main contribution of Aim Materials is therefore not a broad discovery count, but a computational workflow that treats physical plausibility, stage-resolved validation, and evidence integrity as first-class design goals. That is the result the present manuscript is intended to establish.

CRediT authorship contribution statement

Sunwoo Lee: Conceptualization, Methodology, Software, Formal analysis, Data curation, Investigation, Validation, Visualization, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The author declares no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The author acknowledges The Ohio State University for institutional support and local CPU/GPU workstation resources used for model training and DFT workflow execution.

Data and code availability

Public project data are available through the Aim Materials MPContribs project [11]. The structured artifacts used to generate the manuscript snapshot, tables, and figures are maintained in the associated repository outputs and can be regenerated from the included parsing and manuscript-generation scripts.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author used OpenAI Codex for drafting support, language refinement, and manuscript artifact automation. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

[1] A. Jain et al., Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Materials 1(1) (2013) 011002. https://doi.org/10.1063/1.4812323 [2] S.P. Ong et al., Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, Computational Materials Science 68 (2013) 314--319. https://doi.org/10.1016/j.commatsci.2012.10.028 [3] P. Giannozzi et al., QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, Journal of Physics: Condensed Matter 21(39) (2009) 395502. https://doi.org/10.1088/0953-8984/21/39/395502 [4] P. Giannozzi et al., Advanced capabilities for materials modelling with Quantum ESPRESSO, Journal of Physics: Condensed Matter 29(46) (2017) 465901. https://doi.org/10.1088/1361-648X/aa8f79 [5] P. Giannozzi et al., Quantum ESPRESSO toward the exascale, The Journal of Chemical Physics 152(15) (2020) 154105. https://doi.org/10.1063/5.0005082 [6] V.G. Satorras et al., E(n) Equivariant Graph Neural Networks, Proceedings of the 38th International Conference on Machine Learning, pp. 9323--9332, 2021. [7] T. Xie et al., Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties, Physical Review Letters 120(14) (2018) 145301. https://doi.org/10.1103/physrevlett.120.145301 [8] C. Chen et al., Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals, Chemistry of Materials 31(9) (2019) 3564--3572. https://doi.org/10.1021/acs.chemmater.9b01294 [9] B. Deng et al., CHGNet as a pretrained

universal neural network potential for charge-informed atomistic modelling, Nature Machine Intelligence 5(9) (2023) 1031--1041. https://doi.org/10.1038/s42256-023-00716-3 [10] A. Dunn et al., Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm, npj Computational Materials 6(1) (2020) 138. https://doi.org/10.1038/s41524-020-00406-3 [11] P. Huck et al., User applications driven by the community contribution framework MPContribs in the Materials Project, Concurrency and Computation: Practice and Experience 28(7) (2015) 1982--1993. https://doi.org/10.1002/cpe.3698 [12] S. Batzner et al., E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, Nature Communications 13(1) (2022) 2453. https://doi.org/10.1038/s41467-022-29939-5 [13] H.J. Monkhorst et al., Special points for Brillouin-zone integrations, Physical Review B 13(12) (1976) 5188--5192. https://doi.org/10.1103/physrevb.13.5188 [14] J.P. Perdew et al., Generalized Gradient Approximation Made Simple, Physical Review Letters 77(18) (1996) 3865--3868. https://doi.org/10.1103/physrevlett.77.3865 [15] A. Hjorth Larsen et al., The atomic simulation environment -- a Python library for working with atoms, Journal of Physics: Condensed Matter 29(27) (2017) 273002. https://doi.org/10.1088/1361-648x/aa680e [16] A.M. Ganose et al., Atomate2: modular workflows for materials science, Digital Discovery 4(7) (2025) 1944--1973. https://doi.org/10.1039/d5dd00019j [17] X. Xie et al., Crystal Diffusion Variational Autoencoder for Periodic Material Generation, arXiv:2110.06197 (2021). https://arxiv.org/abs/2110.06197 [18] Y. Jiao et al., Crystal Structure Prediction by Joint Equivariant Diffusion, arXiv:2309.04475 (2023). https://arxiv.org/abs/2309.04475 [19] T. Gong et al., Physics-guided deep learning for generative design of crystal materials with symmetry constraints, npj Computational Materials 10 (2024) 67. https://doi.org/10.1038/s41524-024-01285-0 [20] G. Prandini et al., Precision and efficiency in solid-state pseudopotential calculations, npj Computational Materials 4 (2018) 72. https://doi.org/10.1038/s41524-018-0127-2