

Topology-Audited Computational Screening of Single-Atom Heteroatom Core Hops in Drug-Like Scaffolds

A preliminary reproducibility study of structural filtering, formal atom-transfer reactions, conformer sampling, solvent sensitivity, and electronic-structure validation

Hassan Raza

henryraza786@gmail.com

Karachi

Research Article

Keywords: core hopping, skeletal editing, GFN2-xTB, Gibbs free energy, conformer validation, ALPB solvation, density functional theory, cheminformatics, reproducibility

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10706694/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Topology-Audited Computational Screening of Single-Atom Heteroatom Core Hops in Drug-Like Scaffolds

A preliminary reproducibility study of structural filtering, formal atom-transfer reactions, conformer sampling, solvent sensitivity, and electronic-structure validation

Hassan Raza¹

¹Karachi, Sindh, Pakistan. (ORCID: 0009-0009-3848-9220).

Structured Abstract

Context: Single-atom heteroatom core hops are useful structure-generation hypotheses, but a graph edit is not itself a balanced reaction, stationary point, or thermodynamic result. This study audits a reproducible workflow that screened phase-IV ChEMBL molecules, generated C→N, N→O, and O→S core-restricted daughters, and tested balanced atom-transfer reference reactions. The release contains 3,417 phase-IV records, 2,080 validated pairs, 1,528 synthetic-accessibility-filtered pairs, 583 pairs with operational therapeutic evidence, and 10 thermochemistry candidates. The original parser-level ChEMBL2 O→S result of $-76.53 \text{ kcal mol}^{-1}$ was invalidated after topology auditing, leaving zero reportable reaction-level Gibbs energies. ALPB single-point sensitivities were $-89.43 \text{ kcal mol}^{-1}$ in water and $-82.26 \text{ kcal mol}^{-1}$ in acetonitrile, but neither is a solution-phase free energy. Expanded unconstrained conformer calculations also failed topology integrity, while constrained calculations were diagnostic only. Large-state density functional theory produced no complete states. **Methods:** RDKit and ChEMBL/PubChem records supported structural screening; GFN2-xTB supplied gas-phase calculations; analytical linearized Poisson–Boltzmann corrections supplied solvent sensitivity; density functional theory was attempted as an independent benchmark; and NIST frequencies supported a small-reference comparison. The result is a transparent preliminary computational audit and reproducibility package, not evidence of synthetic feasibility or definitive thermochemistry.

Keywords: *core hopping; skeletal editing; GFN2-xTB; Gibbs free energy; conformer validation; ALPB solvation; density functional theory; cheminformatics; reproducibility.*

1. Introduction

1.1 Background

Single-atom skeletal editing changes the identity of an atom embedded in a molecular framework while attempting to retain the broader scaffold topology. In medicinal chemistry, such edits can generate analogue hypotheses involving polarity, hydrogen bonding, lipophilicity, molecular weight, and accessibility. A structural edit, however, is not automatically a chemically defined reaction. When the elemental composition changes, a parent–daughter energy difference requires an explicit atom balance and reference species. A thermochemical claim additionally requires stationary molecular states, vibrational analysis, thermal corrections, phase and standard-state definitions, and a method appropriate to the target property [1, 2].

The present work arose from a structure-screening workflow using the ChEMBL bioactivity database and open-source cheminformatics. ChEMBL is a manually curated resource that combines chemical, bioactivity, and genomic information for drug discovery [3, 9]. PubChem was used as an independent identity crosswalk [6], and synthetic accessibility was treated as a ranking heuristic based on molecular complexity and fragment contributions [5]. The workflow was expanded into a failure-aware thermochemical audit involving gas-phase tight-binding calculations, bounded conformer ensembles, implicit-solvent sensitivity, independent DFT attempts, and machine-readable quality assurance.

1.2 Research problem

The research problem is the potential conflation of structural prioritization with thermodynamic or synthetic evidence. An unbalanced parent–daughter energy difference can have no defensible reaction interpretation. An unconstrained geometry optimization can change the molecular graph. A single-point solvent correction on a gas-phase geometry does not establish a fully optimized solution-phase free energy. Incomplete conformer or DFT calculations must not be silently replaced by fabricated values. The study therefore applies a missingness policy in which failed and incomplete states remain visible but are excluded from headline quantitative claims.

1.3 Research aim

The aim was to construct, execute, audit, and consolidate a reproducible workflow for single-atom heteroatom core hops that separates structural screening from formal thermochemical reference reactions, quantifies conformer and solvent sensitivity where technically supported, attempts independent DFT validation, and reports all exclusions, failures, and remaining requirements transparently.

1.4 Research questions

1. Can the selected core-hop pairs be represented by exact atom-balanced reference reactions?
2. Which molecular states pass geometry, gradient, frequency, topology, and missingness gates?
3. Does bounded or expanded conformer sampling produce a reportable reaction-level free energy?
4. How sensitive are the calculations to implicit-solvent corrections in water and acetonitrile?
5. Can an independent DFT protocol produce a complete four-state reaction-level result?
6. Which conclusions remain defensible after invalidated, incomplete, and sensitivity-only results are excluded?

1.5 Scope and conservative methodological contribution

The contribution is restricted to workflow integration and auditability. The study does not claim that single-atom skeletal editing, GFN2-xTB, RDKit, or any individual validation test is novel. The contribution is the integration of corrected core-restricted graph edits, explicit atom balance, independent identity evidence, state-level numerical and topology gates, conformer and solvent diagnostics, DFT failure documentation, and a complete machine-readable release. The formal reactions are reference reactions for atom transfer; they are not proposed laboratory mechanisms, transition-state models, kinetic pathways, or synthetic procedures.

2. Materials and Methods

2.1 Data acquisition and provenance

The structural source was a snapshot of ChEMBL molecule records with maximum development phase equal to four. Raw JSON responses, request metadata, ChEMBL identifiers, canonical SMILES, source URLs, and intermediate tables were retained. The final structural funnel contained 3,417 phase-IV records with canonical SMILES, 2,080 validated core-restricted parent–daughter pairs, 1,528 pairs after a daughter synthetic-accessibility threshold of ≤ 4.5 , and 583 pairs with explicit ChEMBL indication evidence matching an operational antiviral or oncology-associated lexical screen. ChEMBL phase was not treated as proof of regulatory approval.

2.2 Core-restricted graph editing and structural validation

Validated SMARTS patterns represented the intended heterocyclic core families. Candidate daughters were retained only after RDKit sanitization, canonicalization, exact atom-count comparison, atomic-number-difference checking, and duplicate removal. The principal descriptors were molecular weight, calculated LogP, ring count, hydrogen-bond donors, hydrogen-bond acceptors, topological polar surface area, and the Ertl synthetic-accessibility score [5, 7].

2.3 Candidate selection and independent compound evidence

Ten candidates were advanced to formal reference-reaction thermochemistry. Seven unique records were crosswalked to preferred names: prazosin, pirfenidone, furosemide, lapatinib, fruquintinib, methoxsalen, and niacin. OpenFDA Drugs@FDA active-ingredient searches returned records for five terms and no match for 8-methoxypsoralen and nicotinic acid. These results were treated as evidence-search outputs only; exact salt or product-form identity and current marketing status were not manually adjudicated.

2.4 Formal atom-transfer reference reactions

Table M1. Formal reference reactions

Table M1 (about here).

Edit class	Formal reference reaction	Interpretation
C→N	$P + \text{NH}_3 \rightarrow D + \text{CH}_4$	Atom-transfer reference only
N→O	$P + \text{H}_2\text{O} \rightarrow D + \text{NH}_3$	Atom-transfer reference only
O→S	$P + \text{H}_2\text{S} \rightarrow D + \text{H}_2\text{O}$	Atom-transfer reference only

Every retained reaction had zero elemental-balance residual. These equations do not define mechanism, transition state, rate, or synthetic yield.

2.5 Gas-phase GFN2-xTB thermochemistry

GFN2-xTB version 6.5.1 was used for gas-phase geometry optimization, numerical Hessians, and ideal-gas rigid-rotor harmonic-oscillator thermochemistry at 298.15 K and 1 atm [1]. The original numerical state gate required normal termination, a parsed free energy, no significant imaginary frequencies, and a gradient norm $\leq 5 \times 10^{-4}$ Eh Å⁻¹. The corrected posthoc gate additionally required preservation of the intended molecular topology and rejected any optimized intended bond longer than 2.2 Å. Failed values were left missing rather than imputed.

2.6 Conformer ensembles and topology-preserving diagnostics

The bounded workflow generated twelve RDKit ETKDG conformers per large state and tested five pre-ranked conformers. A broader unconstrained follow-up generated 85 parent and 86 daughter conformers, with 20 tested per large state. Although

11/20 parent and 17/20 daughter conformers passed the original numerical xTB gates, their optimized geometries changed the inferred topology. A graph-preserving follow-up attempted 40 conformers using an all-bonds harmonic restraint; five of 20 parent tests and three of 20 daughter tests passed the numerical gate. These calculations are sensitivity-only because the restraint potential changes the Hamiltonian [8].

2.7 Solvent sensitivity and standard states

Fully optimized ALPB calculations were attempted in water and acetonitrile after large-state self-consistent-field convergence problems. Only two of eight solvent-optimized states passed. Separately, ALPB single-point corrections were calculated for all eight states on validated gas-phase geometries. For $P + H_2S \rightarrow D + H_2O$, the stoichiometric change in the number of gas-phase molecules is zero, so the 1 atm-to-1 M correction cancels for that formal equation. The resulting values retain gas-phase geometries and gas-phase thermal corrections and are not solution-phase free energies [2, 4].

2.8 Independent DFT and small-reference benchmark

Large-state DFT attempts included wB97X-D/def2-TZVP optimization and resource-bounded wB97X-D/def2-SVP tests. The available 4 GB memory envelope terminated the large optimization and single-point attempts, and no complete four-state DFT reaction result was produced. A separate H_2O/H_2S benchmark completed B3LYP/def2-SVP and wB97X-D/def2-SVP frequencies and compared them with NIST Chemistry WebBook data [2, 10]. B3LYP had a mean absolute frequency error of 75.49 cm^{-1} and wB97X-D had 135.97 cm^{-1} ; this benchmark does not validate the large reaction.

2.9 Quality assurance and missingness policy

Quality assurance was performed at structural, balance, state, topology, conformer, solvent, DFT, evidence, and release levels. The original numerical reconciliation passed its checks, but the strengthened topology audit superseded the earlier reportability decision. Failed and invalidated states were retained in raw logs and excluded from headline calculations. Independent human computational-chemist review remains outstanding.

3. Results

3.1 Structural screening and evidence yield

Table 1. Dataset-flow counts

Table 1 (about here).

Processing stage	Count
Phase-IV ChEMBL records with canonical SMILES	3417
Validated core-restricted parent–daughter pairs	2080
Pairs with daughter SA score ≤ 4.5	1528
Pairs with antiviral/oncology indication evidence and SA ≤ 4.5	583
Unique ChEMBL records in therapeutic subset	171
Unique PubChem CIDs matched	2022
Staged parent/daughter candidate pairs	10

Purpose: quantify structural attrition from ChEMBL phase-IV records to thermochemistry candidates. Finding: structural filtering produced a large auditable screening set, but structural filtering does not establish thermodynamic validity.

Fig. 1 (about here).

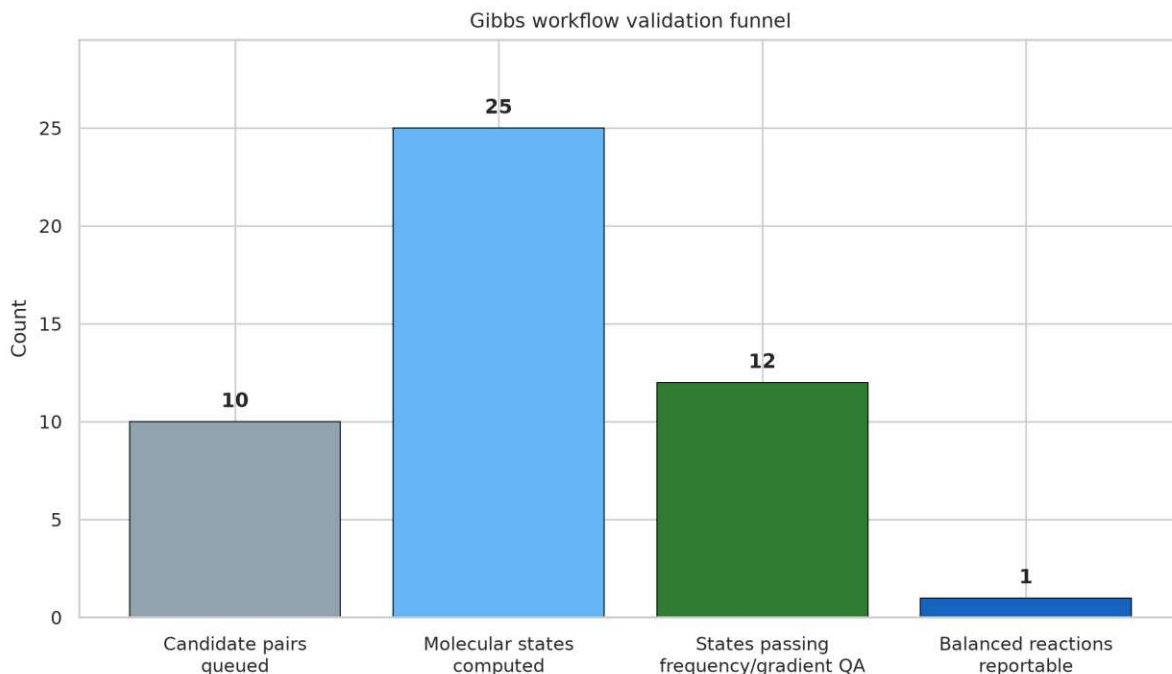


Figure 1. Gibbs validation funnel. Purpose: quantify attrition from queued candidates to reportable balanced reactions. Finding: 10 candidates produced 25 molecular states and 12 passed the original parser-level state gate, but the strengthened topology audit leaves 0 current reportable reactions.

Table 2. Mutation-class composition

Table 2 (about here).

Dataset	Mutation type	Pairs
All validated pairs	C to N	2045
All validated pairs	N to O	0
All validated pairs	O to S	35
SA-filtered pairs	C to N	1497
SA-filtered pairs	N to O	0
SA-filtered pairs	O to S	31
Therapeutic-evidence subset	C to N	574
Therapeutic-evidence subset	N to O	0
Therapeutic-evidence subset	O to S	9

Purpose: report the distribution of C→N, N→O, and O→S edits. Finding: the final thermochemistry queue was dominated by candidates whose reaction-level protocol remained incomplete or failed topology QA.

Table 3. Therapeutic-evidence screen

Table 3 (about here).

Evidence class	Unique ChEMBL molecules	SA-filtered pairs	Unique indication terms
oncology	3866	467	256
antiviral	1289	237	57

Purpose: show operational indication-evidence yield. Finding: these labels describe database evidence under a lexical rule and are not clinical indication or approval determinations.

Table 4. Descriptor statistics

Table 4 (about here).

Mutation type	Metric	N	Mean	SD	Median	Min	Max
C to N	Daughter SA Score	574	3.175271426933414	0.5673651023905164	3.0437586791490263	1.6914394318459838	4.4804582961104735
C_to_N	Delta_LogP	574	-0.6050000000000005	8.861659190909622e-16	-0.6050000000000004	-0.6050000000000004	-0.6049999999999999
C_to_N	Delta_TPSA	574	12.889999999999995	1.5460116986547974e-14	12.89	12.889999999999993	12.890000000000001
C to N	Delta HBA	574	1.0	0.0	1.0	1.0	1.0
C to N	Delta HBD	574	0.0	0.0	0.0	0.0	0.0
C_to_N	Delta_Molecular_Weight	574	0.9879999999999984	3.936993299994934e-14	0.9879999999999995	0.9879999999999829	0.98800000000001132
C to N	Delta_Ring_Strain_Proxy	574	0.0	0.0	0.0	0.0	0.0
O to S	Daughter SA Score	9	2.759501935478016	0.3841151665539936	2.699982201632828	2.2882347135495777	3.407046569794055
O_to_S	Delta_LogP	9	0.4685000000000001	6.203231439246562e-16	0.4685000000000003	0.4684999999999988	0.4685000000000008
O_to_S	Delta_TPSA	9	-13.14	7.562566735475685e-15	-13.14	-13.140000000000017	-13.139999999999986
O to S	Delta HBA	9	0.0	0.0	0.0	0.0	0.0
O to S	Delta HBD	9	0.0	0.0	0.0	0.0	0.0
O_to_S	Delta_Molecular_Weight	9	16.068	6.356528578195824e-14	16.067999999999984	16.067999999999987	16.068000000000097
O to S	Delta_Ring_Strain_Proxy	9	0.0	0.0	0.0	0.0	0.0

Purpose: summarize parent–daughter descriptor changes and accessibility distributions. Finding: core hops alter several descriptors, but descriptor shifts are not evidence of favorable reaction free energy.

Fig. 2 (about here).

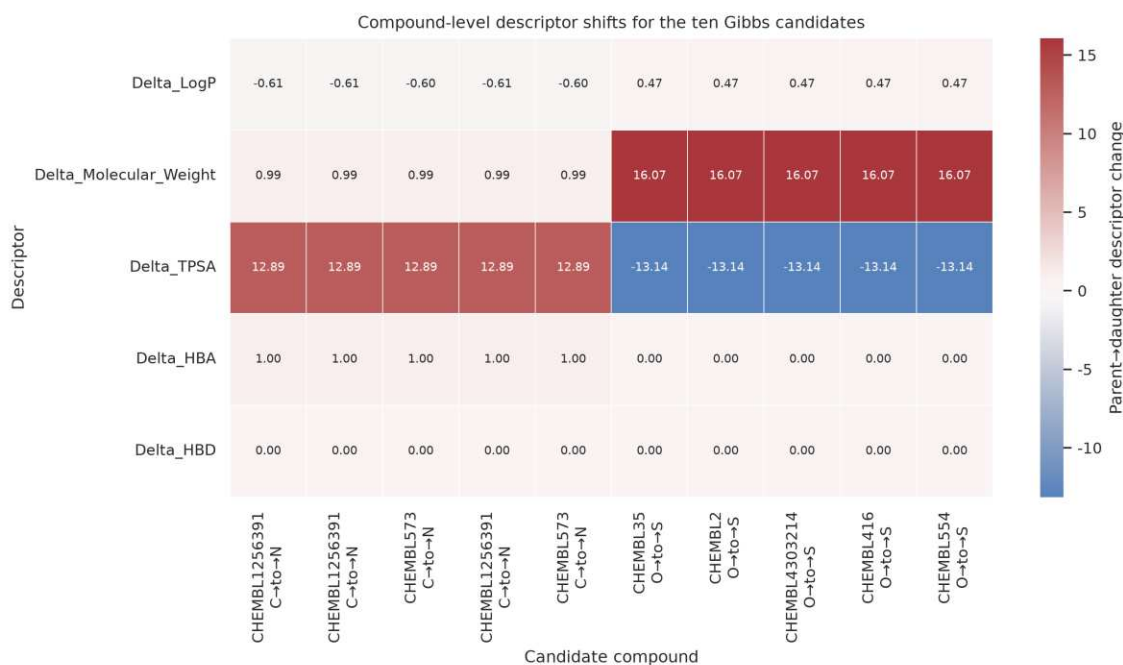


Figure 2. Compound-level descriptor shifts. Purpose: show parent-to-daughter changes in accessibility, lipophilicity, mass, polarity, and hydrogen-bond descriptors. Finding: changes are heterogeneous and not thermodynamic evidence.

3.2 Candidate-level thermochemistry

Ten candidates were queued for balanced reference-reaction calculations. The initial numerical workflow evaluated 25 molecular states and identified 12 parser-level numerical passes. The corrected topology gate eliminated the only historical reaction-level value. The complete candidate table retains every candidate and makes missingness explicit.

Table 5. Complete Gibbs-energy candidate results

Table 5 (about here).

Ran k	ChEMBL_ID	Mutation_T ype	reaction_str ing	balanc ed	Reaction_Status	Reaction_DeltaG_kc al mol	Reaction_DeltaH_kc al mol	Thermochemistry_reporting status
1	CHEMBL1256391	C_to_N	Parent + NH3 -> Daughter + CH4	True	not_reportable_failed_state_or _balance	NR	NR	Not reportable; retained as failed/incomplete
2	CHEMBL1256391	C_to_N	Parent + NH3 -> Daughter + CH4	True	not_reportable_failed_state_or _balance	NR	NR	Not reportable; retained as failed/incomplete
3	CHEMBL573	C_to_N	Parent + NH3 -> Daughter + CH4	True	not_reportable_failed_state_or _balance	NR	NR	Not reportable; retained as failed/incomplete
4	CHEMBL1256391	C_to_N	Parent + NH3 -> Daughter + CH4	True	not_reportable_failed_state_or _balance	NR	NR	Not reportable; retained as failed/incomplete
5	CHEMBL573	C_to_N	Parent + NH3 -> Daughter + CH4	True	not_reportable_failed_state_or _balance	NR	NR	Not reportable; retained as failed/incomplete
6	CHEMBL35	O_to_S	Parent + H2S -> Daughter + H2O	True	not_reportable_failed_state_or _balance	NR	NR	Not reportable; retained as failed/incomplete
7	CHEMBL2	O_to_S	Parent + H2S -> Daughter + H2O	True	Historical value invalidated after topology audit	-76.53440757328326	-79.65863861555727	Historical only; not reportable
8	CHEMBL4303214	O_to_S	Parent + H2S -> Daughter + H2O	True	not_reportable_failed_state_or _balance	NR	NR	Not reportable; retained as failed/incomplete
9	CHEMBL416	O_to_S	Parent + H2S -> Daughter + H2O	True	not_reportable_failed_state_or _balance	NR	NR	Not reportable; retained as failed/incomplete
10	CHEMBL554	O_to_S	Parent + H2S ->	True	not_reportable_failed_state_or _balance	NR	NR	Not reportable; retained as failed/incomplete

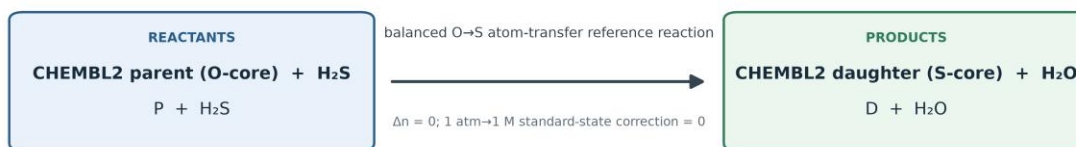
			Daughter + H2O					
--	--	--	-------------------	--	--	--	--	--

Purpose: preserve all ten candidates and distinguish valid, invalidated, failed, and incomplete outcomes. Finding: zero candidates currently have reportable reaction-level ΔG or ΔH ; ChEMBL2 is historical and invalidated.

Fig. 3 (about here).

Gibbs free-energy diagram for the validated ChEMBL2 O→S core hop

Balanced gas-phase reference reaction at 298.15 K and 1 atm | GFN2-xTB RRHO thermochemistry



Interpretation: the negative ΔG applies only to this balanced gas-phase reference reaction and method; it is not a transition-state barrier, synthetic-yield prediction, or solvent-optimized solution free energy.

Figure 3. Gibbs thermodynamic reaction diagram for ChEMBL2 O→S. Purpose: show the balanced formal reaction and $\Delta G = \Delta H - T\Delta S$ decomposition. Finding: historical values are retained for audit traceability only and are invalidated by topology QA.

Fig. 4 (about here).

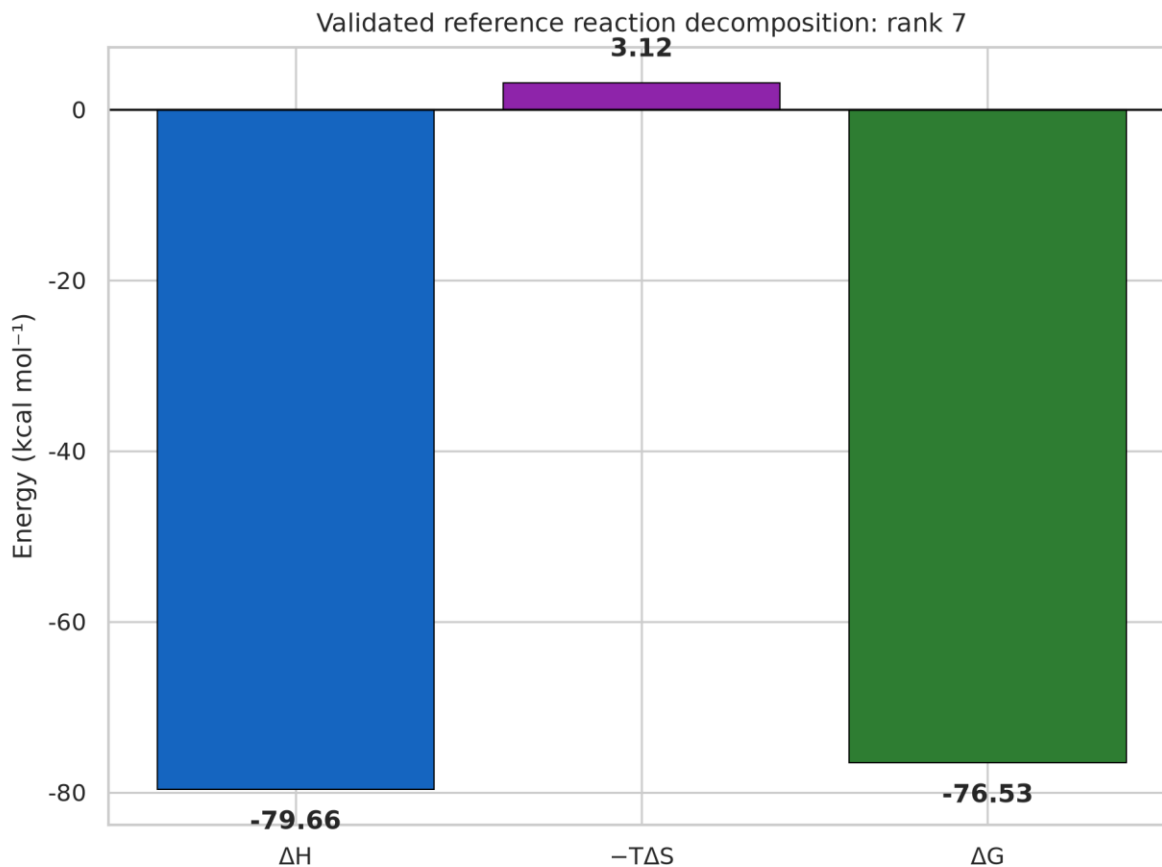


Figure 4. Historical thermodynamic decomposition. Purpose: separate enthalpic and entropic terms. Finding: the decomposition is not a current validated thermochemical claim.

Fig. 5 (about here).

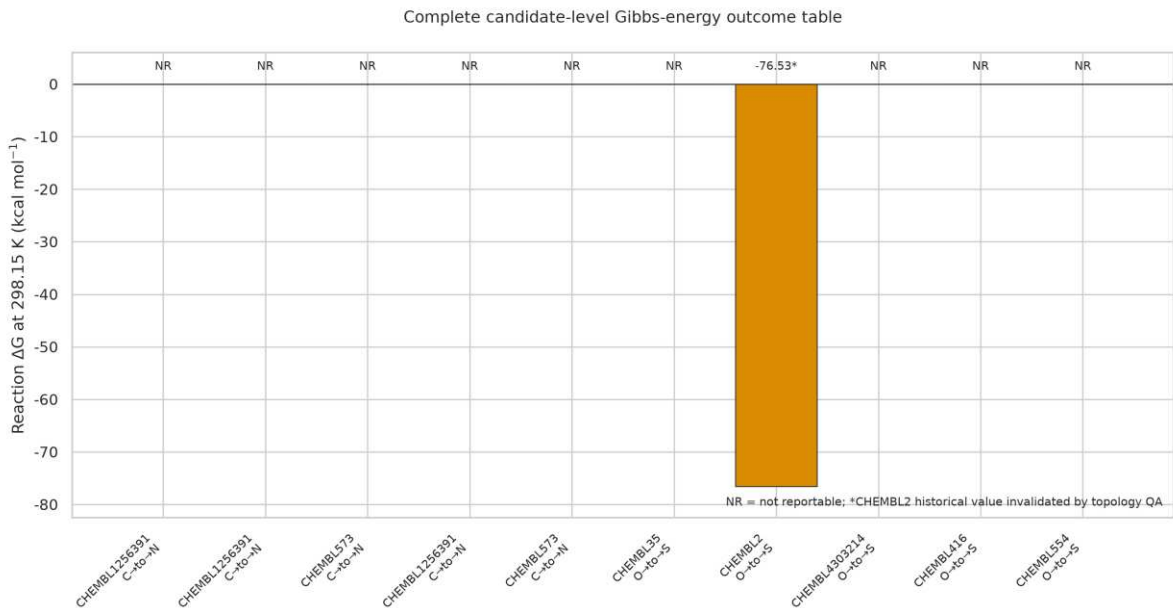


Figure 5. Complete candidate-level Gibbs outcomes. Purpose: display every queued candidate. Finding: one historical numerical row is invalidated and nine candidates are not reportable.

Table 6. Complete compound-level candidate results

Table 6 (about here).

Ra nk	ChEMBL_I D	Mutation_ Type	Target_Core_ Family	Therapeutic_Class Evidence	Parent_SA_S core	Daughter_SA Score	Delta_LogP	Delta_Molecular Weight	Delta_TPSA	Gibbs_Sta tus
1	CHEMBL12 56391	C_to_N	pyridine	antiviral;oncology	1.7299975087 690598	1.6914394318 459838	- 0.6050000000 000004	0.987999999999 9995	12.89	Not reportable; retained as failed/inco mplete
2	CHEMBL12 56391	C_to_N	pyridine	antiviral;oncology	1.7299975087 690598	1.9955044318 459847	- 0.6050000000 000004	0.987999999999 9712	12.89	Not reportable; retained as failed/inco mplete
3	CHEMBL57 3	C_to_N	pyridine	antiviral;oncology	1.5852360072 255571	2.0291547764 56325	-0.605	0.988000000000 028	12.89	Not reportable; retained as failed/inco mplete
4	CHEMBL12 56391	C_to_N	pyridine	antiviral;oncology	1.7299975087 690598	2.0440213549 22908	- 0.6050000000 000004	0.987999999999 9995	12.89	Not reportable; retained as failed/inco mplete
5	CHEMBL57 3	C_to_N	pyridine	antiviral;oncology	1.5852360072 255571	2.1032316995 332483	-0.605	0.988000000000 028	12.89	Not reportable; retained as failed/inco mplete
6	CHEMBL35	O_to_S	furan	oncology	2.2774055637 520085	2.2882347135 495777	0.4685000000 000003	16.06799999999 9984	-13.14	Not reportable; retained as failed/inco mplete
7	CHEMBL2	O_to_S	furan	antiviral	2.3020663908 165098	2.2970436985 088174	0.4685000000 000003	16.06800000000 004	-13.14	Historical parser- level value; invalidated
8	CHEMBL43 03214	O_to_S	benzofuran	oncology	2.4511514130 78449	2.5466780547 294814	0.4685000000 000006	16.06800000000 004	- 13.139999999 999986	Not reportable; retained as failed/inco mplete
9	CHEMBL41 6	O_to_S	benzofuran	oncology	2.5214605154 821097	2.5989929569 53684	0.4685000000 000006	16.06800000000 004	-13.14	Not reportable; retained as failed/inco mplete
10	CHEMBL55 4	O_to_S	furan	oncology	2.6734662150 10756	2.6999822016 32828	0.4684999999 999997	16.06799999999 9984	-13.14	Not reportable; retained as failed/inco mplete

Purpose: show identities, edit classes, core families, evidence labels, SA values, descriptor shifts, and Gibbs status. Finding: descriptor changes cannot replace validated thermochemistry.

3.3 Conformer validation

Table 7. Conformer ensemble results

Table 7 (about here).

label	generated_conformers	xtb_tested_conformers	xtb_validated_conformers	ensemble_status	lowest_conformer_G_kcal_mol	ensemble_G_kcal_mol	ensemble_minus_lowest_kcal_mol	population_sum	method
top_07_parent	12	5	3	pass	-48046.30190819308	-48046.30190819308	0.0	1.0	GFN2-xTB verytight optimized numerical Hessian; ideal-gas RRHO; 298.15 K; 1 atm
top_07_daughter	12	5	5	pass	-47552.606083811945	-47552.606083811945	0.0	1.0	GFN2-xTB verytight optimized numerical Hessian; ideal-gas RRHO; 298.15 K; 1 atm
ref_H2S	1	1	1	pass	-2503.829640228937	-2503.829640228937	0.0	1.0	GFN2-xTB verytight optimized numerical Hessian; ideal-gas RRHO; 298.15 K; 1 atm
ref_H2O	1	1	1	pass	-3136.819405833752	-3136.819405833752	0.0	1.0	GFN2-xTB verytight optimized numerical Hessian; ideal-gas RRHO; 298.15 K; 1 atm

Purpose: show generated, tested, and numerically validated conformers. Finding: the bounded and expanded calculations do not provide a reportable reaction-level free energy after topology and Hamiltonian checks.

Fig. 6 (about here).

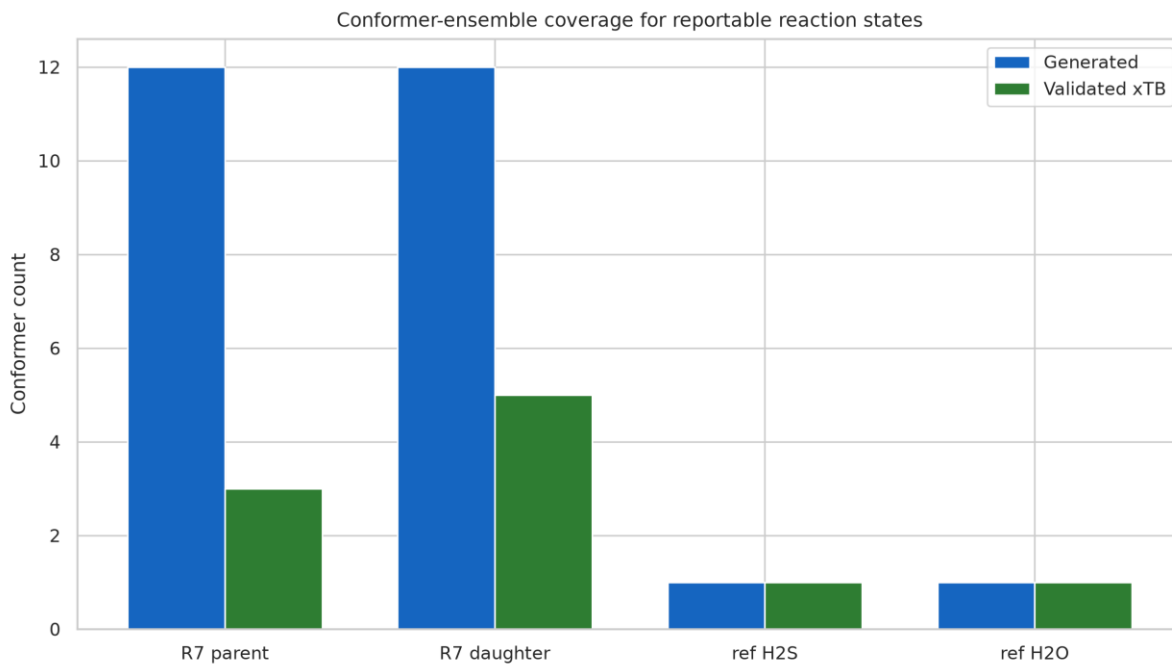


Figure 6. Conformer coverage. Purpose: compare generated, tested, and validated conformers. Finding: unconstrained numerical passes can coexist with topology failure, while constrained passes are sensitivity-only.

3.4 Solvent sensitivity

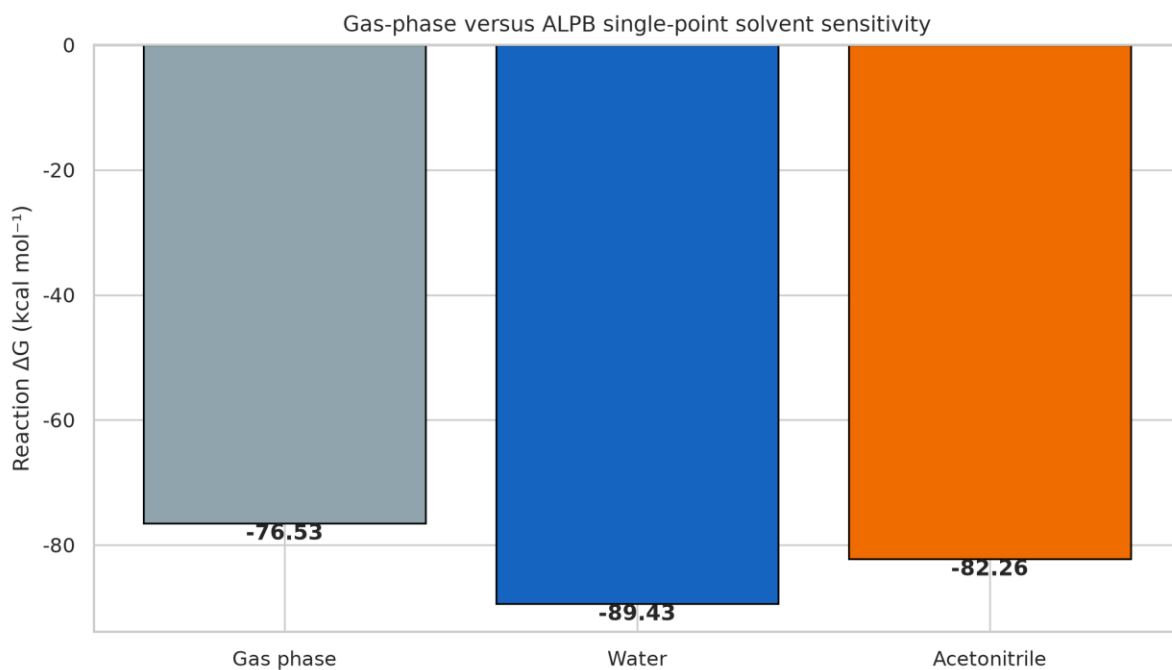
Table 8. Solvent sensitivity values

Table 8 (about here).

Model	Reaction ΔG (kcal mol ⁻¹)	Claim status
Historical parser-level gas-phase ChEMBL2 O→S	-76.53	Historical only; invalidated by topology QA
ALPB water single-point	-89.43	Sensitivity only
ALPB acetonitrile single-point	-82.26	Sensitivity only

Purpose: show method-dependent shifts. Finding: neither solvent value is a fully optimized solution-phase free energy.

Fig. 7 (about here).



Solvent values retain gas-phase RRHO thermal corrections and are not solvent-optimized free energies.

Figure 7. Gas-phase and ALPB single-point sensitivity. Purpose: show numerical solvent shifts. Finding: the shifts are material but retain gas-phase geometry and thermal corrections.

3.5 DFT and small-reference benchmark

Table 9. Large-state DFT queue

Table 9 (about here).

Rank	ChEMBL_ID	Mutation_Type	Target_Core_Family	Therapeutic_Class_Evidence	Daughter_SA_Score	DFT_Status	Confidence
1	CHEMBL1256391	C_to_N	pyridine	antiviral;oncology	1.6914394318459838	Input staged only; calculation not executed	High for structure; Low for thermodynamics
2	CHEMBL1256391	C_to_N	pyridine	antiviral;oncology	1.9955044318459847	Input staged only; calculation not executed	High for structure; Low for thermodynamics
3	CHEMBL573	C_to_N	pyridine	antiviral;oncology	2.029154776456325	Input staged only; calculation not executed	High for structure; Low for thermodynamics
4	CHEMBL1256391	C_to_N	pyridine	antiviral;oncology	2.044021354922908	Input staged only; calculation not executed	High for structure; Low for thermodynamics
5	CHEMBL573	C_to_N	pyridine	antiviral;oncology	2.1032316995332483	Input staged only; calculation not executed	High for structure; Low for thermodynamics
6	CHEMBL35	O_to_S	furan	oncology	2.2882347135495777	Input staged only; calculation not executed	High for structure; Low for thermodynamics
7	CHEMBL2	O_to_S	furan	antiviral	2.2970436985088174	Input staged only; calculation not executed	High for structure; Low for thermodynamics
8	CHEMBL4303214	O_to_S	benzofuran	oncology	2.5466780547294814	Input staged only; calculation not executed	High for structure; Low for thermodynamics
9	CHEMBL416	O_to_S	benzofuran	oncology	2.598992956953684	Input staged only; calculation not executed	High for structure; Low for thermodynamics
10	CHEMBL554	O_to_S	furan	oncology	2.699982201632828	Input staged only; calculation not executed	High for structure; Low for thermodynamics

Purpose: document staged DFT inputs and completion status. Finding: zero large-state DFT reaction results are reportable.

Table 10. Small-reference DFT frequency comparison

Table 10 (about here).

species	functional	mode	NIST_cm-1	DFT_cm-1	absolute_error_cm-1	percent_error
H2O	b3lyp	1	1594.59	1638.8797527004538	44.289752700453846	2.7775009689295587
H2O	b3lyp	2	3656.65	3791.082628902979	134.4326289029791	3.676387647244858
H2O	b3lyp	3	3755.79	3886.290687765359	130.50068776535863	3.474653475443479
H2O	wb97xd	1	1594.59	1636.2596920111648	41.6696920111649	2.6131916048115755
H2O	wb97xd	2	3656.65	3894.416038674002	237.76603867400215	6.502291405357423
H2O	wb97xd	3	3755.79	3994.141966344034	238.35196634403383	6.34625381994291
H2S	b3lyp	1	1182.7	1196.4555417847755	13.755541784775687	1.163062635053326
H2S	b3lyp	2	2614.6	2675.436463964903	60.83646396490303	2.326798132215369
H2S	b3lyp	3	2626.0	2695.1154530143936	69.11545301439355	2.63196698455421
H2S	wb97xd	1	1182.7	1211.8710370158765	29.17103701587644	2.4664781445739776
H2S	wb97xd	2	2614.6	2745.6005827655586	131.00058276555865	5.010348916299191
H2S	wb97xd	3	2626.0	2763.881206981247	137.88120698124703	5.2506171736956215

Purpose: compare B3LYP and wB97X-D/def2-SVP against NIST modes. Finding: B3LYP had the lower small-reference MAE, but this does not validate the 49-atom reaction.

Fig. 8 (about here).

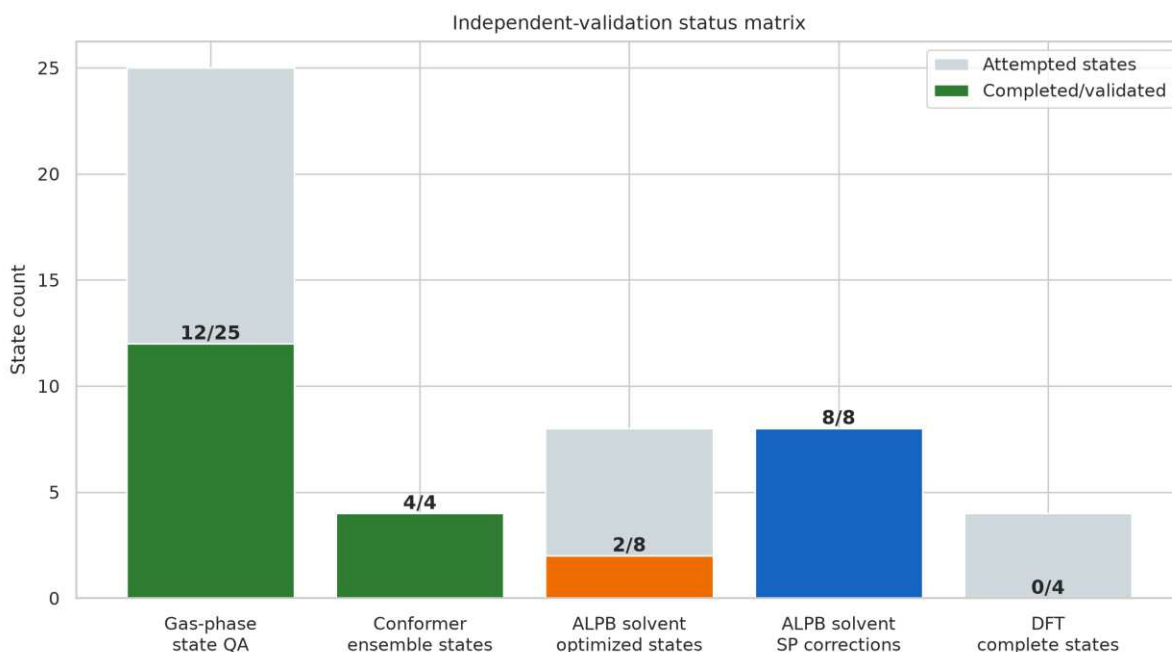


Figure 8. Independent-validation status matrix. Purpose: separate completed analyses from incomplete attempts. Finding: full solvent optimization and large-state DFT were not reportable.

3.6 Validation matrix and independent evidence

Table 11. Final validation matrix

Table 11 (about here).

Validation layer	Observed result	Use in headline claims	Interpretation
Structural screening	3,417 → 2,080 → 1,528 → 583	Yes	Reportable screening yield; therapeutic evidence is operational, not clinical
Formal reaction balance	10/10 retained reactions balanced	Formal reference only	Atom-transfer accounting, not mechanism or synthetic feasibility
Original GFN2-xTB thermochemistry	25 states; 12 numerical passes	No	CHEMBL2 value invalidated by topology QA
Topology audit	38 intended bonds >2.2 Å in original CHEMBL2 parent and daughter	No	Inferred graph changed
Connectivity-preserving xTB	52/52 intended bonds retained	Diagnostic only	Constraints alter the Hamiltonian
Constrained Hessian sensitivity	80–82 imaginary modes in unconstrained Hessian; +100.44 kcal mol ⁻¹ constrained sensitivity	Sensitivity only	Not standard thermochemistry
Expanded unconstrained ensemble	85 parent/86 daughter generated; 11/20 and 17/20 numerical passes	Diagnostic only	Topology changed after optimization
Expanded constrained ensemble	40 attempts; 5/20 parent and 3/20 daughter numerical passes	Sensitivity only	Not standard unconstrained thermochemistry
ALPB solvent-optimized calculations	2/8 states passed	No	No complete solution-phase thermochemistry
ALPB single-point corrections	8/8 state corrections completed	Sensitivity only	Gas geometries and gas RRHO corrections retained
Large-state DFT	0/4 complete states	No	Jobs terminated by 4 GB memory envelope
Small-reference DFT	H ₂ O/H ₂ S B3LYP and wB97X-D def2-SVP completed	Benchmark only	Does not validate the large reaction
Compound evidence	7 unique ChEMBL/PubChem records; 5/7 openFDA terms matched	Evidence only	Exact form/status not adjudicated

Purpose: separate reportable screening, formal balance, diagnostics, sensitivity-only values, incomplete calculations, and outstanding review. Finding: zero reaction-level Gibbs energies survive topology QA.

Table 12. Independent compound-evidence crosswalk

Table 12 (about here).

Preferred-name record	openFDA search term	Records returned	Search outcome	Interpretation
PRazosin	Prazosin	30	Record found	Exact product/form and current status not adjudicated
PIRfenidone	Pirfenidone	26	Record found	Exact product/form and current status not adjudicated
FUROsemide	Furosemide	73	Record found	Exact product/form and current status not adjudicated
LAPatinib	Lapatinib	3	Record found	Exact product/form and current status not adjudicated
FRUquintinib	Fruquintinib	1	Record found	Exact product/form and current status not adjudicated
METHoxSALEN	8-Methoxypsoralen	0	No match	No openFDA Drugs@FDA active-ingredient record returned
NIACIN	Nicotinic Acid	0	No match	No openFDA Drugs@FDA active-ingredient record returned

Purpose: report seven preferred-name records and openFDA Drugs@FDA search outcomes. Finding: five terms returned records and two did not; exact form and current approval status were not adjudicated.

3.7 Structural and visual evidence

Fig. 9 (about here).

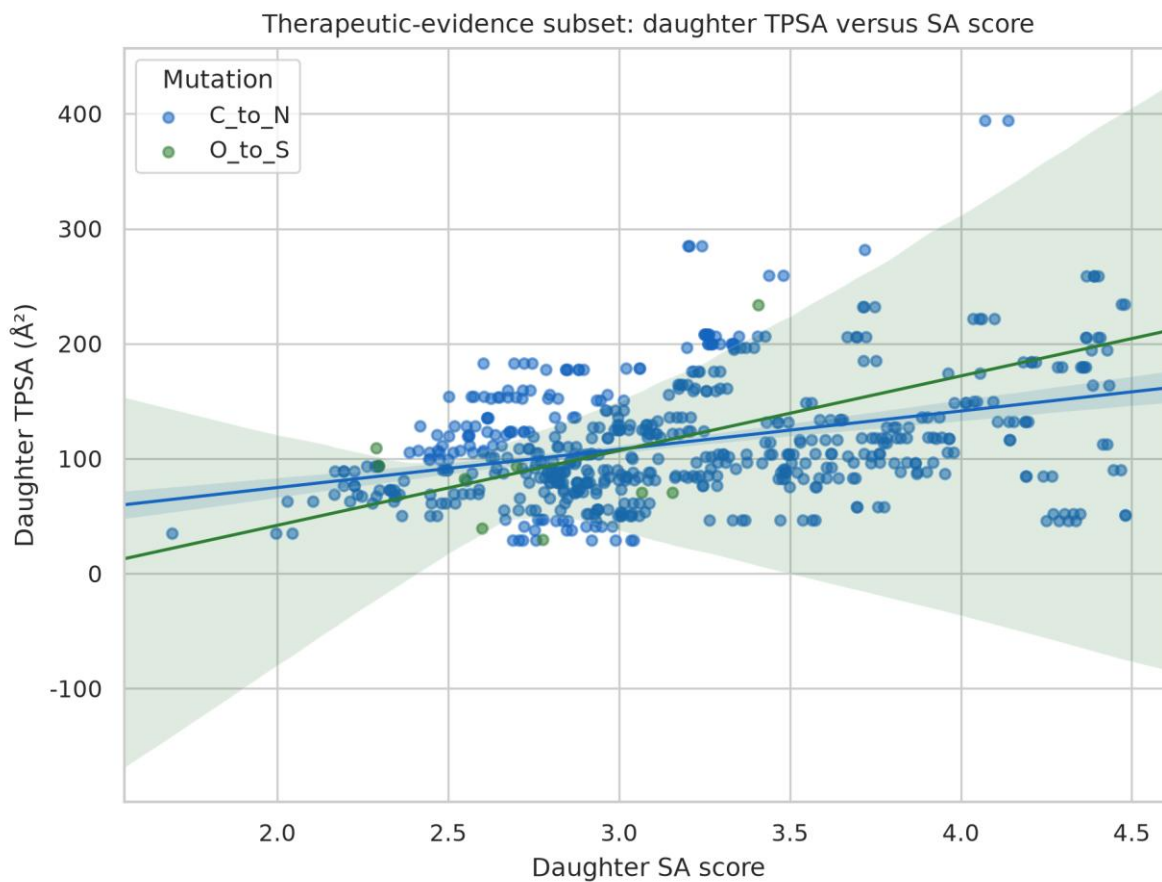


Figure 9. Descriptor shift versus synthetic accessibility. Purpose: display descriptor/accessibility relationships. Finding: descriptive only; not a ΔG model.

Fig. 10 (about here).



Figure 10. Descriptor-shift distributions. Purpose: summarize parent–daughter changes. Finding: multiple descriptors change across edits.

Fig. 11 (about here).

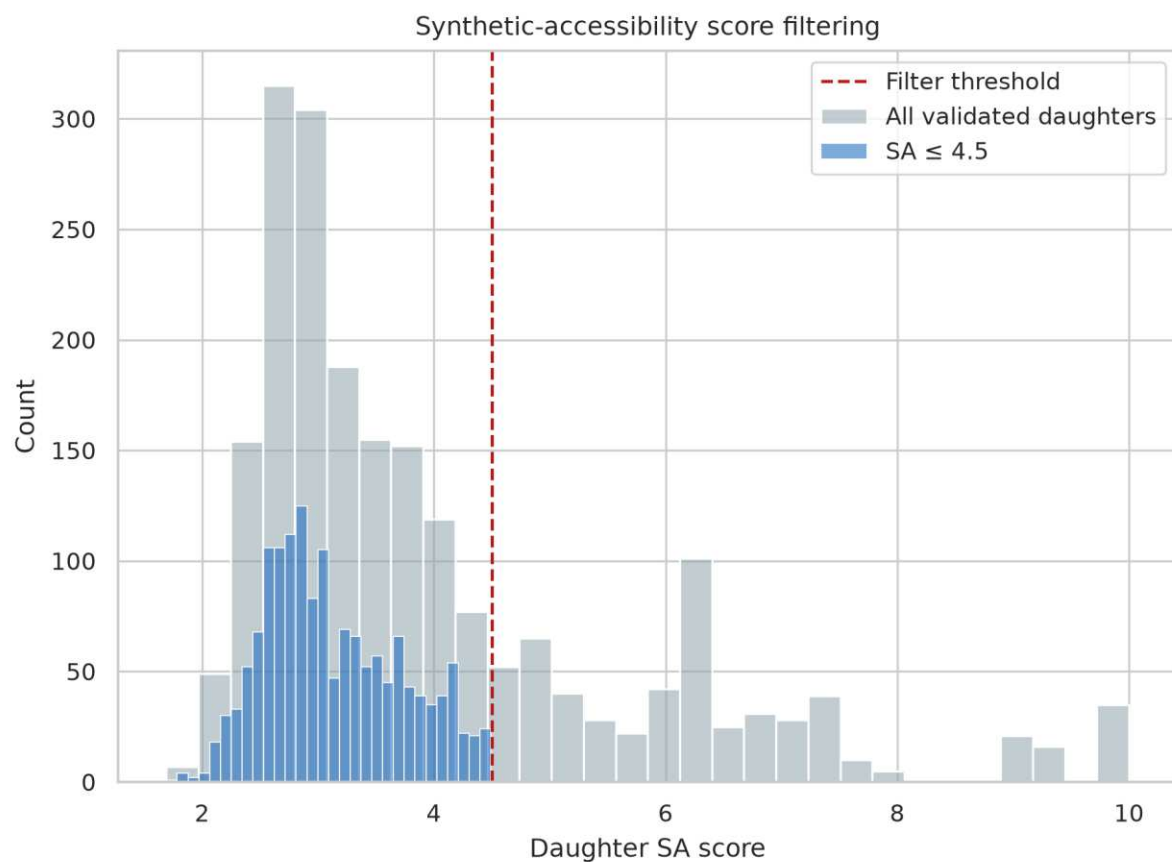


Figure 11. Synthetic-accessibility filtering. Purpose: show the $SA \leq 4.5$ screen. Finding: a screening rule, not a synthesis probability.

Fig. 12 (about here).

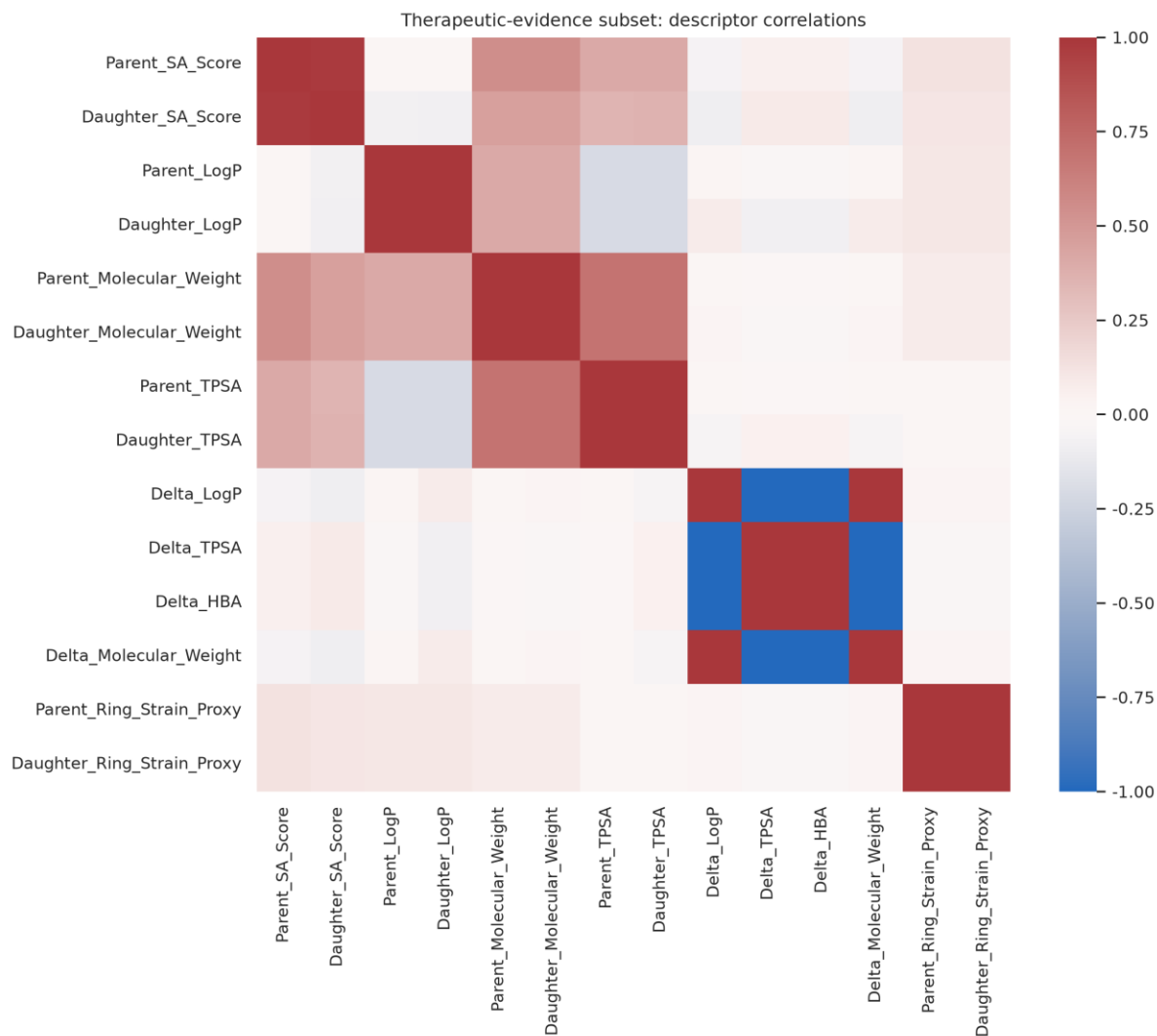


Figure 12. Descriptor correlation heatmap. Purpose: display descriptor dependence. Finding: correlations inform interpretation but do not establish thermodynamic causality.

Fig. 13 (about here).

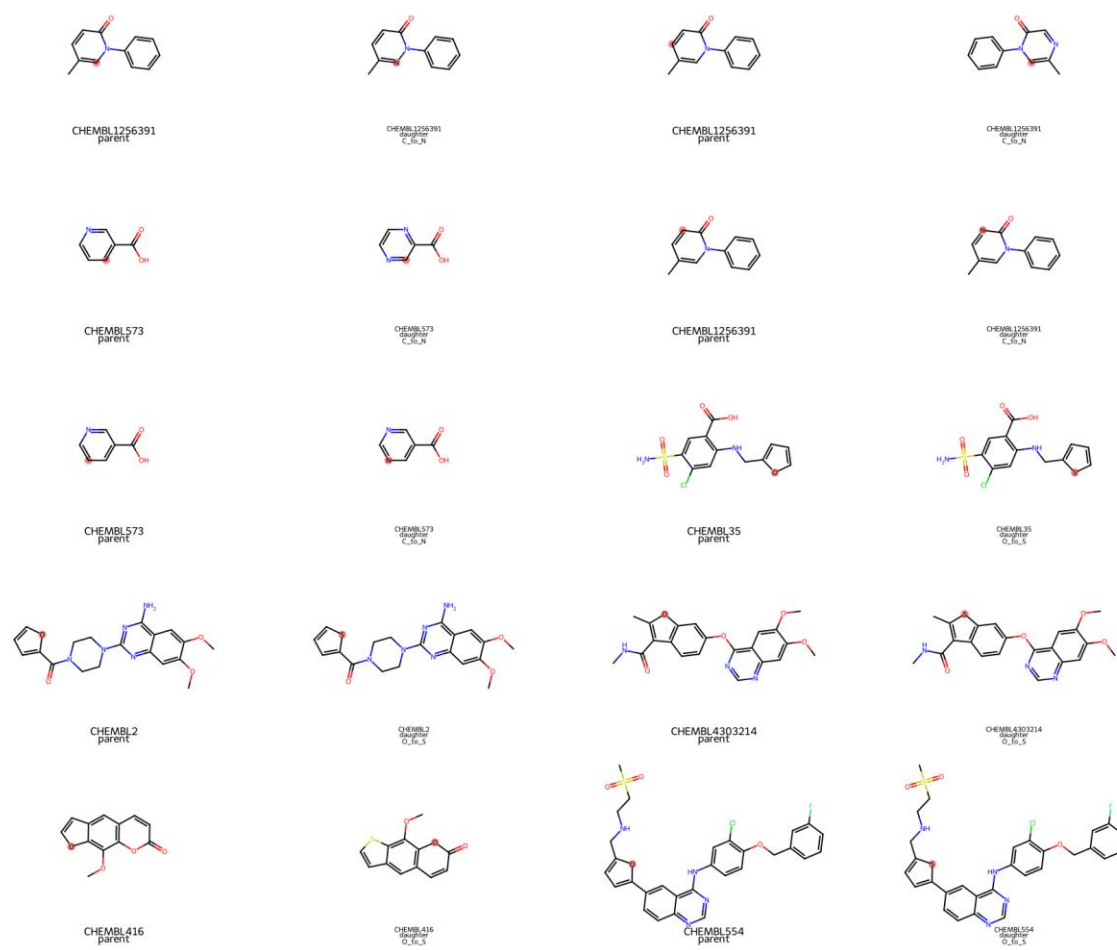


Figure 13. Parent–daughter two-dimensional structures. Purpose: document the candidate queue. Finding: structural rendering, not a three-dimensional optimized overlay.

Fig. 14 (about here).

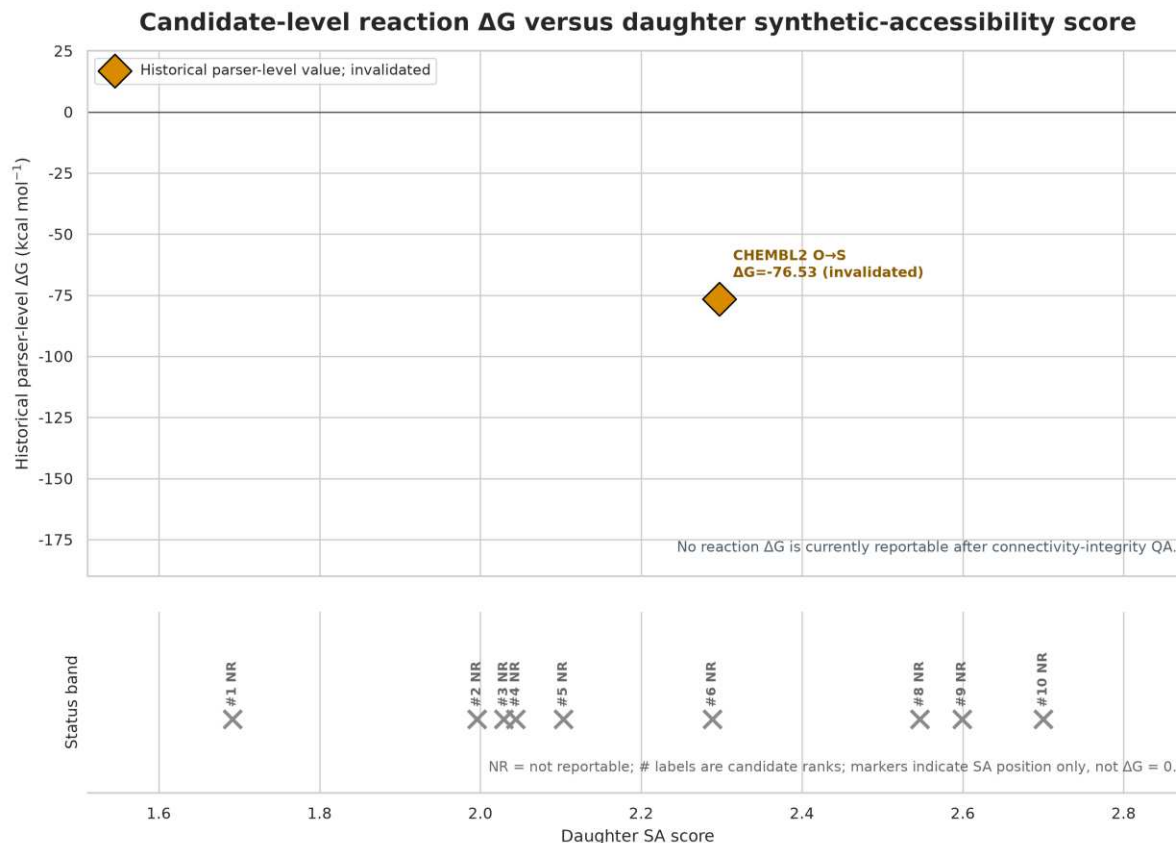


Figure 14. Transparent candidate-level ΔG versus daughter synthetic accessibility. Purpose: show one historical invalidated point and nine not-reportable candidates in a separate status band. Finding: no candidate currently has a reportable reaction ΔG ; not-reportable markers are not zero-energy values.

4. Discussion

The principal finding is methodological rather than a positive thermochemical signal. Structural screening and formal reaction balancing were successful, but graph integrity became the decisive validation layer. The original ChEMBL2 optimized structures changed the inferred molecular graph, invalidating the historical -76.53 kcal mol⁻¹ value. A parsed free energy, even when apparently favorable, cannot be promoted without structure and stationary-point validation.

The all-bonds-constrained xTB protocol preserved the intended graph, but its restraint changes the Hamiltonian. The $+100.44$ kcal mol⁻¹ constrained reaction sensitivity is therefore diagnostic rather than standard thermochemistry. The same distinction applies to the constrained conformer ensemble. The expanded unconstrained ensemble demonstrates that broader sampling alone cannot cure a topology-changing optimization workflow.

The solvent layer reinforces the same claim boundary. ALPB single-point corrections shifted the historical gas-phase value to -89.43 kcal mol⁻¹ in water and -82.26 kcal mol⁻¹ in acetonitrile, but retained gas-phase geometries and gas-phase RRHO corrections. The fully solvent-optimized calculation passed only 2/8 states, so a solution-phase free energy would be unsupported.

The DFT layer is a documented negative result. Large-state jobs were terminated by the available memory envelope, and the small H₂O/H₂S benchmark cannot substitute for a large-reaction validation. The decision to exclude incomplete DFT values is consistent with best-practice protocol selection and uncertainty reporting [2].

The compound-evidence crosswalk improves provenance but does not establish clinical or regulatory claims. ChEMBL, PubChem, and openFDA records should be interpreted as database evidence, with exact product forms and current status manually adjudicated before any approval statement.

5. Strengths and limitations

5.1 Strengths

Strengths include explicit atom balance; separation of structural, thermochemical, solvent, and mechanistic claims; dual-source compound provenance; corrected SMARTS; duplicate removal; state-level gradient and frequency gates; topology QA; conformer and solvent diagnostics; retention of rejected states; machine-readable data; raster and vector figures; a complete release archive; and a clean-clone audit. The workflow is failure-aware: missing or failed calculations remain missing.

5.2 Limitations

The current study has zero reportable reaction-level Gibbs energies. The historical ChEMBL2 value is invalidated after topology QA. Constrained xTB and constrained Hessian results are diagnostic only. The expanded unconstrained conformer output fails connectivity QA, while the constrained ensemble changes the Hamiltonian. Fully optimized solvent thermochemistry is incomplete, with only 2/8 states passing. Large-state DFT produced zero complete states. The small-reference benchmark is not a large-reaction validation. ALPB single-point values retain gas-phase structures and thermal corrections. No transition-state searches, experimental mechanisms, explicit solvent clusters, higher-level large-reaction benchmark, or independent human computational-chemist review were completed. The therapeutic-evidence rule is lexical and not a clinical ontology. The FDA crosswalk is an evidence search rather than an adjudicated regulatory conclusion.

5.3 Future directions

Future work should use a larger-memory environment to complete unconstrained large-state DFT optimizations and frequencies; establish topology-preserving unconstrained stationary points; perform a CREST-scale or otherwise benchmarked conformer search; complete solvent-optimized geometries and frequencies with a consistent standard-state treatment; compare multiple functionals and basis sets; seek higher-level or experimental benchmarks; test explicit solvent where strong interactions are relevant; conduct transition-state calculations only if synthetic-mechanistic claims are intended; and obtain independent review of structures, mappings, charges, multiplicities, balances, convergence, and claims.

6. Conclusions

This study provides a complete, transparent computational audit of single-atom heteroatom core-hop screening from ChEMBL phase-IV records through formal reference-reaction analysis. Structural screening and atom-balance layers are complete. The original ChEMBL2 O→S Gibbs value is invalidated by a posthoc topology audit, and the current reportable reaction count is zero. Constrained xTB, expanded conformer, ALPB single-point, solvent-optimized, and DFT results are retained according to their evidentiary status rather than promoted beyond their validation. The release is suitable as a preliminary computational methods, reproducibility, or negative-results study, not as a definitive thermochemistry or synthetic-feasibility paper.

Declarations

Funding. No external funding information was documented in the project records.

Conflict of interest. No conflict-of-interest information was documented in the project records.

Data availability. All raw and processed data, scripts, figures, tables, logs, machine-readable audit outputs, appendices, and checksums are included in the release archive and private repository at <https://github.com/Hassan786Raza-dot/audited-single-atom-core-hop-screening>, commit ec60aa5. The release archive SHA-256 checksum is 2055ae742bc79ea5288f181f7851043c11241a0085e24f90788be1b9266925fb.

Code availability. The workflow uses Python, RDKit, GFN2-xTB 6.5.1, PySCF, pandas, NumPy, Matplotlib, seaborn, python-docx, and requests. The repository contains executable scripts and raw logs needed to reproduce the audited analyses.

Ethics and regulatory status. No human or animal subjects were involved. ChEMBL phase, PubChem identity, and openFDA search records were not interpreted as clinical efficacy or definitive regulatory approval without manual adjudication.

References

1. Bannwarth, C., Ehlert, S., & Grimme, S. (2019). GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of Chemical Theory and Computation*, 15(3), 1652–1671. <https://doi.org/10.1021/acs.jctc.8b01176>
2. Bursch, M., Mewes, J.-M., Hansen, A., & Grimme, S. (2022). Best-practice DFT protocols for basic molecular computational chemistry. *Angewandte Chemie International Edition*, 61(42), e202205735. <https://doi.org/10.1002/anie.202205735>

3. Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L., & Overington, J. P. (2015). ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Research*, 43(W1), W612–W620. <https://doi.org/10.1093/nar/gkv352>
4. Ehlert, S., Stahn, M., Spicher, S., & Grimme, S. (2021). Robust and efficient implicit solvation model for fast semiempirical methods. *Journal of Chemical Theory and Computation*, 17(7), 4250–4261. <https://doi.org/10.1021/acs.jctc.1c00471>
5. Ertl, P., & Schuffenhauer, A. (2009). Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*, 1(1), 8. <https://doi.org/10.1186/1758-2946-1-8>
6. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., Shi, W., Tan, G., Yu, B., Zhang, J., Bryant, S. H., et al. (2019). PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Research*, 47(D1), D1102–D1109. <https://doi.org/10.1093/nar/gky1033>
7. Landrum, G. (2026). RDKit: Open-source cheminformatics (version 2026.03.5). <https://www.rdkit.org/docs/Overview.html>; version DOI: <https://doi.org/10.5281/zenodo.591637>
8. Pracht, P., Bohle, F., & Grimme, S. (2020). Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics*, 22, 7169–7192. <https://doi.org/10.1039/C9CP06869D>
9. Zdrazil, B., Felix, E., Hunter, F., Manners, E. J., Blackshaw, J., Corbett, S., de Veij, M., Ioannidis, H., Mendez Lopez, D., Mosquera, J. F., Magarinos, M. P., Bosc, N., Arcila, R., Kizilören, T., Gaulton, A., Bento, A. P., Adasme, M. F., Monecke, P., Landrum, G. A., & Leach, A. R. (2023). The ChEMBL Database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research*, gkad1004. <https://doi.org/10.1093/nar/gkad1004>
10. National Institute of Standards and Technology. (2025). NIST Chemistry WebBook, SRD 69. <https://doi.org/10.18434/T4D303>
11. U.S. Food and Drug Administration. (2026). Approved drug products with therapeutic equivalence evaluations (Orange Book). <https://www.fda.gov/drugs/drug-approvals-and-databases/approved-drug-products-therapeutic-equivalence-evaluations>

Appendix A. Reaction-balance table

Table A1. Formal reaction-balance records

Table A1 (about here).

Rank	ChEMBL_ID	Mutation_Type	reaction_string	balanced
1	CHEMBL1256391	C_to_N	Parent + NH3 -> Daughter + CH4	True
2	CHEMBL1256391	C_to_N	Parent + NH3 -> Daughter + CH4	True
3	CHEMBL573	C_to_N	Parent + NH3 -> Daughter + CH4	True
4	CHEMBL1256391	C_to_N	Parent + NH3 -> Daughter + CH4	True
5	CHEMBL573	C_to_N	Parent + NH3 -> Daughter + CH4	True
6	CHEMBL35	O_to_S	Parent + H2S -> Daughter + H2O	True
7	CHEMBL2	O_to_S	Parent + H2S -> Daughter + H2O	True
8	CHEMBL4303214	O_to_S	Parent + H2S -> Daughter + H2O	True
9	CHEMBL416	O_to_S	Parent + H2S -> Daughter + H2O	True
10	CHEMBL554	O_to_S	Parent + H2S -> Daughter + H2O	True

Every retained reaction had zero elemental-balance residual.

Appendix B. Molecular-state QA and topology audit

Table B1. Molecular-state and topology audit

Table B1 (about here).

Validation layer	Observed result	Use in headline claims	Interpretation
Structural screening	3,417 → 2,080 → 1,528 → 583	Yes	Reportable screening yield; therapeutic evidence is operational, not clinical
Formal reaction balance	10/10 retained reactions balanced	Formal reference only	Atom-transfer accounting, not mechanism or synthetic feasibility
Original GFN2-xTB thermochemistry	25 states; 12 numerical passes	No	CHEMBL2 value invalidated by topology QA
Topology audit	38 intended bonds >2.2 Å in original ChEMBL2 parent and daughter	No	Inferred graph changed
Connectivity-preserving xTB	52/52 intended bonds retained	Diagnostic only	Constraints alter the Hamiltonian
Constrained Hessian sensitivity	80–82 imaginary modes in unconstrained Hessian; +100.44 kcal mol ⁻¹ constrained sensitivity	Sensitivity only	Not standard thermochemistry
Expanded unconstrained ensemble	85 parent/86 daughter generated; 11/20 and 17/20 numerical passes	Diagnostic only	Topology changed after optimization
Expanded constrained ensemble	40 attempts; 5/20 parent and 3/20 daughter numerical passes	Sensitivity only	Not standard unconstrained thermochemistry
ALPB solvent-optimized calculations	2/8 states passed	No	No complete solution-phase thermochemistry
ALPB single-point corrections	8/8 state corrections completed	Sensitivity only	Gas geometries and gas RRHO corrections retained
Large-state DFT	0/4 complete states	No	Jobs terminated by 4 GB memory envelope
Small-reference DFT	H ₂ O/H ₂ S B3LYP and wB97X-D def2-SVP completed	Benchmark only	Does not validate the large reaction
Compound evidence	7 unique ChEMBL/PubChem records; 5/7 openFDA terms matched	Evidence only	Exact form/status not adjudicated

The corrected rule rejects any changed inferred graph or intended bond longer than 2.2 Å.

Appendix C. Conformer ensemble table

Table C1. Full conformer records

Table C1 (about here).

label	generated_conformers	xtb_tested_conformers	xtb_validated_conformers	ensemble_status	lowest_conformer_G_kcal_mol	ensemble_G_kcal_mol	ensemble_minus_lowest_kcal_mol	population_sum	method
top_07_parent	12	5	3	pass	-48046.30190819308	-48046.30190819308	0.0	1.0	GFN2-xTB verytight optimized numerical Hessian; ideal-gas RRHO; 298.15 K; 1 atm
top_07_daughter	12	5	5	pass	-47552.606083811945	-47552.606083811945	0.0	1.0	GFN2-xTB verytight optimized numerical Hessian; ideal-gas RRHO; 298.15 K; 1 atm
ref_H2S	1	1	1	pass	-2503.829640228937	-2503.829640228937	0.0	1.0	GFN2-xTB verytight optimized numerical Hessian; ideal-gas RRHO; 298.15 K; 1 atm
ref_H2O	1	1	1	pass	-3136.819405833752	-3136.819405833752	0.0	1.0	GFN2-xTB verytight optimized numerical Hessian; ideal-gas RRHO; 298.15 K; 1 atm

Appendix D. Solvent single-point table

Table D1. Solvent single-point corrections

Table D1 (about here).

state	solvent	status	solvation_electronic_correction_kcal_mol	single_point_solvent_corrected_G_kcal_mol
top_07_parent	water	pass	-38.95673586971798	-48074.21540257182
top_07_daughter	water	pass	-39.95126625692883	-47518.7545749275
ref_H2S	water	pass	0.4542788477894994	-2503.375361381148
ref_H2O	water	pass	-11.447437107094704	-3148.266842940847
top_07_parent	acetonitrile	pass	-33.956198538370884	-48069.21486524047
top_07_daughter	acetonitrile	pass	-36.96484075609623	-47515.768149426665
ref_H2S	acetonitrile	pass	-3.0997506702913777	-2506.929390899229
ref_H2O	acetonitrile	pass	-5.816081048967125	-3142.6354868827198

Appendix E. DFT status and small-reference benchmark

Table E1. DFT queue

Table E1 (about here).

Rank	ChEMBL_ID	Mutation_Type	Target_Core_Family	Therapeutic_Class_Evidence	Daughter_SA_Score	DFT_Status	Confidence
1	CHEMBL1256391	C_to_N	pyridine	antiviral;oncology	1.6914394318459838	Input staged only; calculation not executed	High for structure; Low for thermodynamics
2	CHEMBL1256391	C_to_N	pyridine	antiviral;oncology	1.9955044318459847	Input staged only; calculation not executed	High for structure; Low for thermodynamics
3	CHEMBL573	C_to_N	pyridine	antiviral;oncology	2.029154776456325	Input staged only; calculation not executed	High for structure; Low for thermodynamics
4	CHEMBL1256391	C_to_N	pyridine	antiviral;oncology	2.044021354922908	Input staged only; calculation not executed	High for structure; Low for thermodynamics
5	CHEMBL573	C_to_N	pyridine	antiviral;oncology	2.1032316995332483	Input staged only; calculation not executed	High for structure; Low for thermodynamics
6	CHEMBL35	O_to_S	furan	oncology	2.2882347135495777	Input staged only; calculation not executed	High for structure; Low for thermodynamics
7	CHEMBL2	O_to_S	furan	antiviral	2.2970436985088174	Input staged only; calculation not executed	High for structure; Low for thermodynamics
8	CHEMBL4303214	O_to_S	benzofuran	oncology	2.5466780547294814	Input staged only; calculation not executed	High for structure; Low for thermodynamics
9	CHEMBL416	O_to_S	benzofuran	oncology	2.598992956953684	Input staged only; calculation not executed	High for structure; Low for thermodynamics
10	CHEMBL554	O_to_S	furan	oncology	2.699982201632828	Input staged only; calculation not executed	High for structure; Low for thermodynamics

Table E2. H₂O/H₂S frequency benchmark

Table E2 (about here).

species	functional	mode	NIST_cm-1	DFT_cm-1	absolute_error_cm-1	percent_error
H2O	b3lyp	1	1594.59	1638.8797527004538	44.289752700453846	2.7775009689295587
H2O	b3lyp	2	3656.65	3791.082628902979	134.4326289029791	3.676387647244858
H2O	b3lyp	3	3755.79	3886.290687765359	130.50068776535863	3.474653475443479
H2O	wb97xd	1	1594.59	1636.2596920111648	41.6696920111649	2.6131916048115755
H2O	wb97xd	2	3656.65	3894.416038674002	237.76603867400215	6.502291405357423
H2O	wb97xd	3	3755.79	3994.141966344034	238.35196634403383	6.34625381994291
H2S	b3lyp	1	1182.7	1196.4555417847755	13.755541784775687	1.163062635053326
H2S	b3lyp	2	2614.6	2675.436463964903	60.83646396490303	2.326798132215369
H2S	b3lyp	3	2626.0	2695.1154530143936	69.11545301439355	2.63196698455421
H2S	wb97xd	1	1182.7	1211.8710370158765	29.17103701587644	2.4664781445739776
H2S	wb97xd	2	2614.6	2745.6005827655586	131.00058276555865	5.010348916299191
H2S	wb97xd	3	2626.0	2763.881206981247	137.88120698124703	5.2506171736956215

Appendix F. Compound-evidence crosswalk

Table F1. Seven-term evidence crosswalk

Table F1 (about here).

Preferred-name record	openFDA search term	Records returned	Search outcome	Interpretation
PRAZOSIN	Prazosin	30	Record found	Exact product/form and current status not adjudicated
PIRFENIDONE	Pirfenidone	26	Record found	Exact product/form and current status not adjudicated
FUROSEMIDE	Furosemide	73	Record found	Exact product/form and current status not adjudicated
LAPATINIB	Lapatinib	3	Record found	Exact product/form and current status not adjudicated
FRUQUINTINIB	Fruquintinib	1	Record found	Exact product/form and current status not adjudicated
METHOXSALEN	8-Methoxypsoralen	0	No match	No openFDA Drugs@FDA active-ingredient record returned
NIACIN	Nicotinic Acid	0	No match	No openFDA Drugs@FDA active-ingredient record returned

Preferred-name records: PRAZOSIN, PIRFENIDONE, FUROSEMIDE, LAPATINIB, FRUQUINTINIB, METHOXSALEN, and NIACIN. Five of seven openFDA searches returned records; exact product/form and current status were not adjudicated.

Appendix G. Reproducibility and release audit

The release contains the article, figures, tables, raw results, input structures, scripts, logs, evidence crosswalks, audit JSON files, and checksums. The repository commit is ec60aa5. The clean-clone and full-release audits passed.

Reproduction commands: `python3 -m py_compile scripts/*.py`; `python3 scripts/reconcile_final_findings.py`; `python3 scripts/generate_gibbs_compound_tables.py`; `python3 scripts/generate_jmm_submission.py`; `python3 scripts/audit_full_release.py`

Appendix H. Complete compound-level records

The full ten-row compound-level records are supplied in `results/final_compound_level_candidate_results.csv` and `.xlsx`, including parent and daughter SMILES, ChEMBL and PubChem links, descriptors, evidence fields, and Gibbs reporting status.

Appendix I. Complete Gibbs-energy records

The full ten-row Gibbs table is supplied in results/final_gibbs_energy_candidate_results.csv and .xlsx. Empty energy fields mean that no validated reportable reaction free energy was produced; they are not numerical zeros.

Appendix J. External-review checklist

Table J1. External scientific-review checklist

Table J1 (about here).

Review item	Current status
Parent/daughter structures and atom mapping	Package prepared; independent review outstanding
Charges and multiplicities	Inputs retained; independent review outstanding
Topology preservation and thresholds	Corrected rule documented; independent review outstanding
Reaction balance and formal interpretation	Exact balance completed; mechanism excluded
xTB convergence, frequencies, gradients, missingness	Machine-readable audit completed
Conformer sampling adequacy	Bounded/expanded diagnostics completed; CREST-scale validation outstanding
Solvent optimization and standard states	Full solvent thermochemistry outstanding
DFT protocol and large-state completion	Incomplete because of memory limit
Compound identity and regulatory evidence	Crosswalk completed; exact adjudication outstanding
Independent computational-chemist review	Outstanding

Appendix K. Figure captions and interpretation rules

The Word manuscript embeds 14 figures. Each caption states its purpose and finding. The transparent ΔG -versus-SA plot distinguishes one historical invalidated point from nine not-reportable candidates in a separate status band. Not-reportable markers are status markers, not zero-energy observations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix.docx](#)