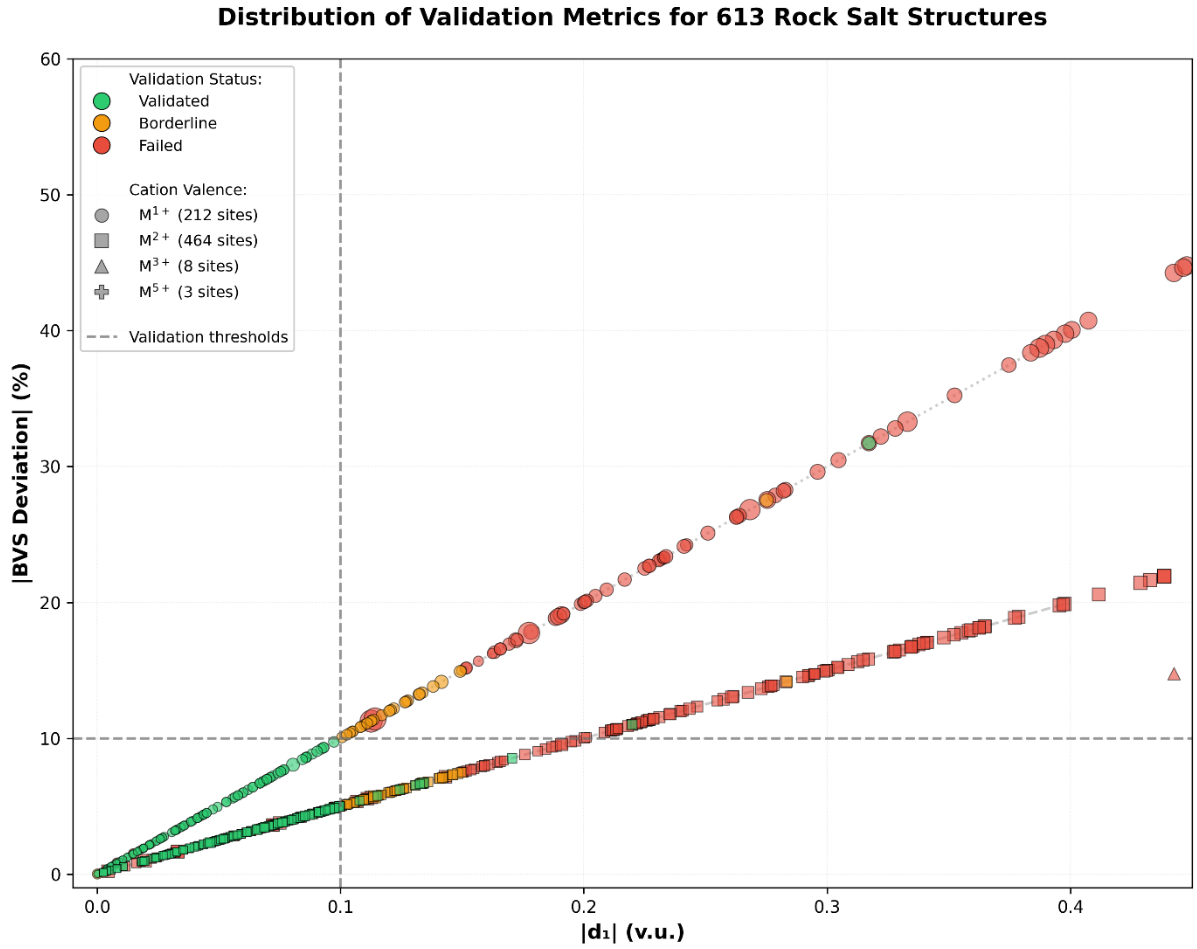
Supplementary Information

Figure S1. Distribution of validation metrics for 613 assessable rock salt structures.

Percentage deviation from formal valence (|BVS deviation|, y-axis) versus absolute discrepancy factor ($|d_1|$, x-axis) for all cation sites analysed. Marker colours indicate structure-level validation status (green: validated, amber: borderline, red: failed); marker shapes indicate cation oxidation state (circles: M^{1+}, squares: M^{2+}, triangles: M^{3+}, plus symbols: M^{5+}). Point size scales with Global Instability Index (GII), with larger markers indicating higher structural strain. Grey dashed lines mark the dual validation thresholds (horizontal: 10% deviation; vertical: $|d_1|$ = 0.10 v.u.); structures must satisfy both criteria simultaneously for automated validation. Subtle grey trend lines (dotted: M^{1+}, dashed: M^{2+}) show the mathematical relationship between absolute and relative deviations for each valence state, demonstrating that the dual-threshold approach ensures consistent validation standards across different oxidation states. Each point represents a cation site from the Eir analysis; structures with mixed cation occupancy contribute multiple points. The clear separation between validated (lower-left), borderline (intermediate), and failed (upper-right) regions confirms that the combined criteria successfully partition structures according to validation reliability.



SECTION S2: SUPPLEMENTARY DATA FILES

The following CSV files are provided as separate downloadable files:

Data File S1: All_Assessable_Structures.csv (613 structures)

Complete listing with validation status, measurement conditions, failure mode classifications. Enables full reproduction of manuscript statistics.

Data File S2: Excluded_No_Parameters.csv (154 structures)

Structures lacking BVS parameters. Identifies priority areas for future parameter development.

Data File S3: Type1_Parameter_Inadequacy.csv (107 structures)

Systematic parameter inadequacy - highest priority for revision. Includes all alkaline earth oxides with r >1.0 Å.

Data File S4: Types3-5_Methodological_Limitations.csv (95 structures)

Structures inappropriate for conventional BVS analysis due to disorder effects.

Data File S5: Type6_Database_Quality.csv (12 structures)

Database quality issues identified by BVS screening. Includes complete references for verification.

Data File S6: Borderline_Structures.csv (93 structures)

Intermediate confidence category that required manual review.

DATA FILE S1: All_Assessable_Structures.csv

Complete listing of 613 rock salt structures assessed by automated bond valence sum analysis.

Columns:

- ICSD Code: Unique identifier from Inorganic Crystal Structure Database
- Formula: Normalised chemical formula (alphabetical element order)
- Chemical Formula: Full stoichiometric formula as reported
- Space Group: $Fm\bar{3}m$ (all structures are rock salt type)
- Validation Status: Validated / Borderline / Failed
- Measurement Conditions: Ambient / High_PT / Elevated
- Compound Class: Oxide / Halide / Sulphide / Other
- Year: Publication/measurement year
- Journal: Source journal (abbreviated)
- GII: Global Instability Index (structure-wide BVS metric)
- Failure Mode: Classification for failed structures (Type 1, Type 3-6, or blank)
- Disorder Type: pure / homovalent / heterovalent / mixed_anion

Validation criteria: "Validated" (<10% deviation, |d1| <0.10 v.u.), "Borderline" (10-15% deviation, requires manual review), "Failed" (>15% deviation or methodological limitations).

Measurement conditions: "Ambient" (T <350 K, P <0.5 GPa), "High_PT" (elevated pressure and/or temperature >10 GPa or >1000 K), "Elevated" (T >350 K at ambient pressure).

Usage: Enables complete reproduction of validation statistics in manuscript Tables 1-3. Researchers building computational datasets can filter by Validation Status and Measurement Conditions to select appropriate reference structures.

File size: ~165 KB, 613 rows (plus header)

DATA FILE S2: Excluded_No_Parameters.csv

Rock salt structures excluded from BVS analysis due to incomplete parameter coverage in the Gagné & Hawthorne (2015) and Brown (2020) compilations.

Columns:

- ICSD Code: Unique identifier from ICSD
- Formula: Normalised chemical formula
- Chemical Formula: Full stoichiometric formula
- Space Group: $Fm\bar{3}m$
- Compound Class: Oxide / Halide / Sulphide / Other
- Year: Publication/measurement year
- Journal: Source journal
- Elements Present: Comma-separated list of elements
- Missing Parameters: Description of parameter gap

Key statistics: 154 structures (20.1% of total dataset). Lanthanide Sulphides dominate exclusions (116 structures, 75.3%), followed by early transition metal oxides (Ti^{2+}, Nb^{2+}, Sc^{2+}), silver halides (Ag^{+}–Br^{-}, Ag^{+}–I^{-}), and 4d metal Sulphides.

Usage: Identifies priority areas for future bond valence parameter development. These structures represent legitimate crystallographic determinations but cannot be assessed via conventional BVS until parameter compilations are extended.

File size: ~35 KB, 154 rows

DATA FILE S3: Type1_Parameter_Inadequacy.csv

Systematic parameter inadequacy failures - pure phases with ≥3 independent measurements showing reproducible validation failures.

Columns:

- Formula: Chemical formula
- ICSD Code: Unique structure identifier
- n (total): Number of independent measurements for this formula
- n (failed): Number that failed validation
- Year: Measurement year
- Mean Deviation (%): Average BVS deviation from formal valence
- GII: Global Instability Index
- Compound Class: Oxide / Sulphide / etc.

Key compounds with 100% failure rates:

- CdO: 17/17 structures
- CaO: 15/15 structures
- Nd_1S_1: 13/13 structures
- EuO: 12/12 structures
- SrO: 11/11 structures
- VO: 9/9 structures
- BaO: 5/5 structures

Diagnostic criteria: Reproducible failures across multiple laboratories and decades with exceptionally low inter-structure variance (<5% relative to mean deviation) distinguish genuine parameter inadequacy from experimental scatter.

Usage: Highest priority targets for bond valence parameter revision. Researchers using these compounds should note that current M^{2+}–O^{2-} and Ln^{2+}–S^{2-} parameters exhibit systematic inadequacies for large cations (r >1.0 Å).

File size: ~15 KB, 107 rows

DATA FILE S4: Types3-5_Methodological_Limitations.csv

Structures where diffraction-averaged geometries prove fundamentally inappropriate for BVS analysis regardless of parameter quality.

Columns:

- Type: Classification (3, 4, or 5)
- Formula: Chemical formula
- ICSD Code: Unique identifier
- Year: Measurement year
- Mean Deviation (%): BVS deviation
- GII: Global Instability Index
- Details: Type-specific information (ΔV for heterovalent, etc.)

Type classifications:

- Type 3 (14 structures): Vacancy disorder in nonstoichiometric phases
- Type 4 (11 structures): Heterovalent disorder, $\Delta V \geq 2$ (Bosi-type artifacts)
- Type 5 (45 structures): Homovalent and mixed anion disorder

Type 4 structures ($LiFeO_2$, $LiCoO_2$, Li_3TaO_4, Li_3NbO_4) exhibit bidirectional deviations characteristic of mathematical artifacts from diffraction-averaged bond lengths. Type 5 includes compositionally disordered solid solutions ($Na_{1-x}K_xCl$, $Ni_{1-x}Zn_xO$) where deviations scale with size mismatch.

Usage: These structures should be flagged in databases as requiring complementary characterization (spectroscopy for Type 4, occupancy-weighted BVS for Type 3) rather than conventional validation. Inappropriate as computational reference structures for bond-length-sensitive applications.

File size: ~13 KB, 95 rows

DATA FILE S5: Type6_Database_Quality.csv

Database quality issues identified through BVS screening - structures with metadata inadequacies or measurement contexts rendering them inappropriate as bulk reference structures.

Columns:

- Type: Subclassification (6A / 6B / 6C / 6D)
- ICSD Code: Unique identifier
- Formula: Chemical formula
- Year: Measurement/publication year
- Deviation (%): BVS deviation magnitude
- Issue: Description of quality concern
- Reference: Source publication (abbreviated)

Type classifications:

- Type 6A (6 structures): Pre-modern techniques (1920-1970), film-based diffraction
- Type 6B (1 structure): Thin film determination (substrate effects, metastability)
- Type 6C (4 structures): Multi-phase Rietveld refinement contamination
- Type 6D (1 structure): Miscategorized measurement conditions (high-pressure synthesis)

Notable case: ICSD 22171 ($K_{0.3}Rb_{0.7}Cl$) includes metadata stating "composition has largest deviation from Vegard's Law", indicating structure was deposited due to anomalous behavior.

Usage: Demonstrates that BVS screening identifies fitness-for-purpose issues. These structures are appropriate for phase diagram compilation but unsuitable for applications requiring high-precision bulk references (machine learning training, computational benchmarking). Complete references enable verification of synthesis contexts.

File size: ~4 KB, 12 rows

DATA FILE S6: Borderline_Structures.csv

Structures requiring manual review to distinguish parameter inadequacy from expected geometric effects.

Columns:

- ICSD Code: Unique identifier
- Formula: Chemical formula
- Chemical Formula: Full stoichiometric formula
- Space Group: $Fm\bar{3}m$
- Compound Class: Oxide / Halide / Sulphide
- Year: Measurement year
- Mean Deviation (%): BVS deviation (10-15% range)
- Mean |d1| (v.u.): Absolute discrepancy factor
- GII: Global Instability Index
- Disorder Type: pure / homovalent / mixed_anion
- Review Reason: Classification rationale

Criteria: Deviations 10-15% or modest absolute discrepancy $0.10 < |d1| < 0.15$ v.u., falling between automated acceptance and definitive failure.

Distribution: Predominantly oxides (64 structures, 68.8%); chlorides (25 structures, 26.9%), sulphides (3 structures, 3.2%), fluorides (1 structure, 1.1%)

Manual review workflow:

1. Verify space group and coordination number assignments
2. Check source publication for synthesis anomalies or non-ambient conditions
3. Compare to related compounds in chemical series
4. For homovalent solid solutions, verify deviation scales with composition

Usage: Intermediate confidence category where automated tools flag potential issues but chemical expertise required for classification. Researchers should consult source publications before using these structures as computational references.

File size: ~14 KB, 93 rows

TOTAL DATA ACCOUNTING

Assessable structures (Data File S1):	613
└ Validated:	286 (46.7%)
└ Borderline (Data File S6):	93 (15.2%)
└ Failed - Type 1 (Data File S3):	107 (17.5%)
└ Failed - Types 3-5 (Data File S4):	95 (15.5%)
└ Failed - Type 6 (Data File S5):	12 (2.0%)
└ Other failures (n<3, unclassified):	20 (3.3%)
Excluded - no parameters (Data File S2):	154

TOTAL UNIQUE STRUCTURES:	766
TOTAL MEASUREMENTS (incl. replicates):	841

NOTE: Data Files S3, S4, and S5 contain 9 overlapping structures (6 appearing in both S3 and S5, 3 in both S4 and S5). The sum of individual file counts (107 + 95 + 12 = 214) therefore exceeds the unique structure count (205) by 9. Overlapping structures represent cases exhibiting multiple failure modes simultaneously.

FILE FORMAT NOTES

All files are provided as comma-separated values (CSV) with UTF-8 encoding.

Import into Python:

```
  import pandas as pd
  df = pd.read_csv('All_Assessable_Structures.csv')
```

Import into R:

```
  df <- read.csv('All_Assessable_Structures.csv')
```

Import into Excel:

Data → Get External Data → From Text/CSV → Select file

Some ICSD codes may display in scientific notation in Excel. To fix:

Select column → Format Cells → Number → 0 decimal places

Missing values are represented as blank cells or "Unknown" for text fields.