Supplementary Information

Equivariant Diffusion Solution for Inorganic Crystal Structure Determination from Powder X-ray Diffraction Data

Dongfang Yu[a,b], Zhewen Zhu[a,b], Fucheng Leng[c], Yizhou Zhu[b*]

[a] School of Materials Science and Engineering, Zhejiang University, Hangzhou 310058, China

[b] Department of Materials Science and Engineering, Westlake University, Hangzhou 310030, China

[c] School of Engineering and Architecture, University College Cork, Cork, Ireland

*Corresponding authors: Yizhou Zhu

E-mail: zhuyizhou@westlake.edu.cn.

Supplementary Table 1. Statistics analysis of inorganic entries in the ICDD PDF-4+ database. The diffraction quality for all entries is categorized as "Star", "Good", or "Indexed", with a status listed as "Primary".

Source Prefix	Source	Entries (Total)	Entries (without atomic coordinates)	Percentage (without atomic coordinates)
00	ICDD	44812	20741	46.28%
01	ICSD	34037	13342	39.20%

Supplementary Table 2. The statistics of solved entries in ICSD.

Structures	Number
Ordered structures in the ICSD database	49283
Potentially incorrect structures with unit cell ≤ 20 atoms (E_{hull} > 0.1 eV/atom)	3122
XRDSol proposed structures that differ from the original ICSD structures	1429
XRDSol proposed structures that show good XRD match (R_{cos} > 0.9)	588
XRDSol proposed structures that show low energy (E_{hull} < 0.1 eV/atom)	39

Supplementary Table 3. Comparison of the E_{hull} between 39 pairs the original ICSD structures and solved structures by XRDSol.

No.	Name	ICSD ID	Original structures (E_{hull} (eV/atom))	Solved structures (E_{hull} (eV/atom))
1	$CsIO_3$	33665	1.966	0.021
2	$BaTiO_3$	161341	1.807	0.110
3	$U(CoP)_2$	78578	0.811	0.049
4	AsH_3	24499	0.727	0.012
5	NbSnRh	105220	0.704	0.000
6	$ZrAlPt_2$	609167	0.678	0.034
7	TiNiSn	105368	0.664	0.059
8	$TlSbSe_2$	20374	0.650	0.030
9	Pb_2Br_5N	26662	0.641	0.056
10	NbP	76027	0.600	0.000
11	ZrSnPd	105705	0.532	0.019
12	HfSnPt	104261	0.494	0.000
13	HfSbRh	53036	0.484	0.000
14	Pu_5Pb_3	648410	0.468	0.094
15	TiNiSb	646450	0.457	0.074
16	SrBiAu	106271	0.333	0.079
17	$SmCo_3NiB$	613218	0.331	0.020
18	$CaHg_2$	619359	0.280	0.103
19	HoSbPd	108547	0.276	0.000
20	TmSbPd	54328	0.272	0.000
21	$Cr(GaS_2)_2$	626045	0.266	0.089
22	MgCuSn	103054	0.262	0.020
23	$FeBiO_3$	162834	0.257	0.056
24	DyAgGe	86731	0.238	0.055
25	ErAgGe	86733	0.234	0.050
26	Eu_3BrO_4	56818	0.206	0.066

27	$Fe(Mo_3S_4)_2$	632656	0.206	0.086
28	NdSbPd	80274	0.197	0.008
29	$Cs(SeO_3)_2$	192325	0.179	0.001
30	Ca_2FeWO_6	81204	0.171	0.002
31	TbAgGe	86730	0.169	0.000
32	HoAgGe	86732	0.165	0.000
33	$SrPbF_6$	25629	0.162	0.000
34	$TlFeI_3$	26423	0.147	0.046
35	Tl_5Te_3	604388	0.143	0.038
36	$BaEu_2PdO_5$	83232	0.135	0.046
37	$K_2MoO_2F_5$	56840	0.130	0.037
38	GeI_2	27674	0.120	0.020
39	$CsSb_2F_7$	24741	0.114	0.042

GeI_2:The structure of GeI_2 was first reported in the literature in 1938[18], and collected into the ICSD database (ICSD-27674) with a hexagonal lattice and spacegroup *P*6*m*2. The calculated E_{hull} of this structure is as high as 0.12 eV/atom. In contrast, the crystal structure proposed by XRDSol has a significantly lower E_{hull} of 0.02 eV/atom and a different spacegroup *P*-3*m*1. In the original structure from ICSD, iodine occupies 2h position with coordinates (1/3, 2/3, 0.224). In contrast, iodine locates at the 2d position with coordinates (1/3, 2/3, 0.242) in our solution.

Supplementary Table 4. Comparison of the E_{hull} between the origin incomplete structure in ICSD and solved structures by XRDSol.

No.	Name	ICSD ID	Original structures (E_{hull} (eV/atom) from MP)	Solved structures (E_{hull} (eV/atom))
1	$Fe(OH)_2$	53992	0.28	0.005
2	SrHN	158968	0.43	0.04
3	$Tm(OH)_3$	200489	0.46	0.06

Supplementary Table 5. Solved structures of natural minerals by XRDSol.

No.	Name	ICDD ID	Solved structures (E_{hull} (eV/atom))
1	Ni	00-061-0762	0.000
2	TlCl	00-058-0499	0.047
3	LaOF	00-057-0608	0.063
4	PdCu	00-057-0606	0.000
5	$CaMg_2As_2(H_2O_5)_2$	00-053-0882	0.020
6	$Zn_2As_2H_4PbO_{10}$	00-051-1444	0.040
7	$ZnFeAs_2H_2PbO_{10}$	00-051-1443	0.055
8	$Cu_2H_3ClO_3$	00-050-1560	0.045
9	$CeTiO_3$	00-049-1808	0.079
10	Bi_4TeS_2	00-047-1747	0.026
11	Cu_6SnMoS_8	00-046-1272	0.082
12	$Pb_2S(O_2F)_2$	00-045-1455	0.097
13	MnO_2	00-042-1316	0.009
14	$CaMn(SiO_3)_2$	00-038-0413	0.078
15	$NaSrPO_4$	00-033-1212	0.012

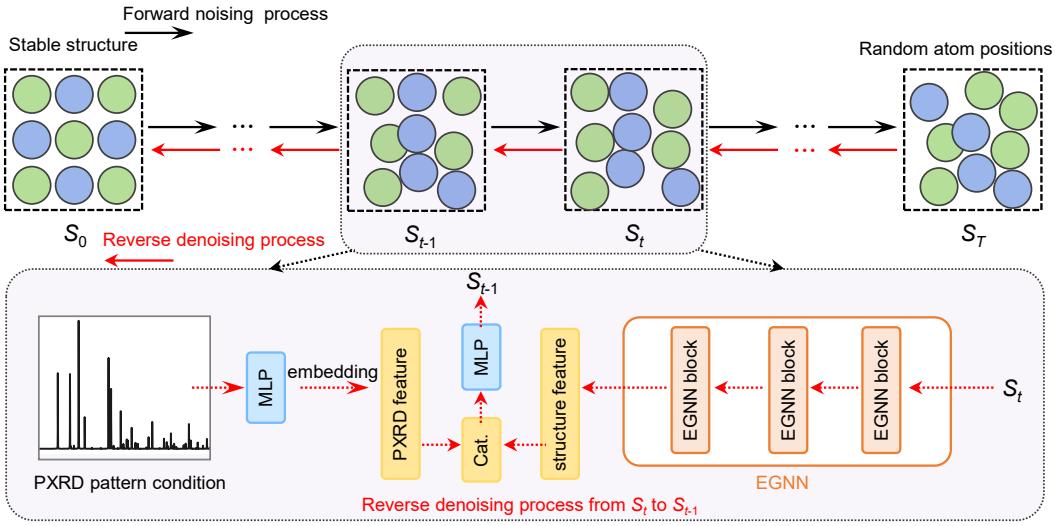

Figure S1. Schematics workflow of the nosing and denoising process in the XRDSol. The training of XRDSol follows a conditional diffusion framework, which consists of a forward noising process and a reverse denoising process. The reverse denoising process aims to reconstruct the original structure S_0 from S_T with random atomic coordinates. This is achieved using an equivariant graph neural network (EGNN). The EGNN predicts the noise to construct the desired samples. The EGNN includes 3 EGNN blocks, each composed of equivariant convolutional layers. The noisy structure S_t is passed through the EGNN to extract structure features. And the structure features are concatenated with the PXRD features which are derived from the target PXRD pattern by using a neural network. Next, the concatenated features are fed into a neural network to predict the atomic coordinates of S_{t-1}. This process is repeated for 1000 steps and gradually restores the atomic coordinates from S_T to S_0 with the guidance of target PXRD patterns.

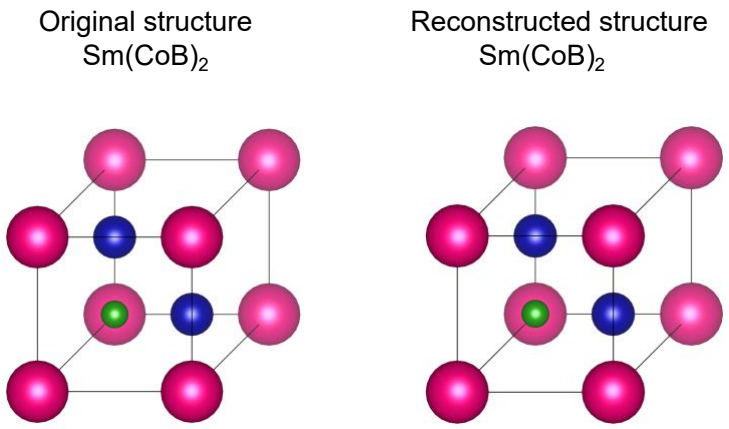

Figure S2. Visual verification of reconstructed crystal structure $Sm(CoB)_2$ compared to its original counterpart. All crystal structures were visualized using VESTA[1]. Sm atoms are pink, Co atoms are blue, and B atoms are green.

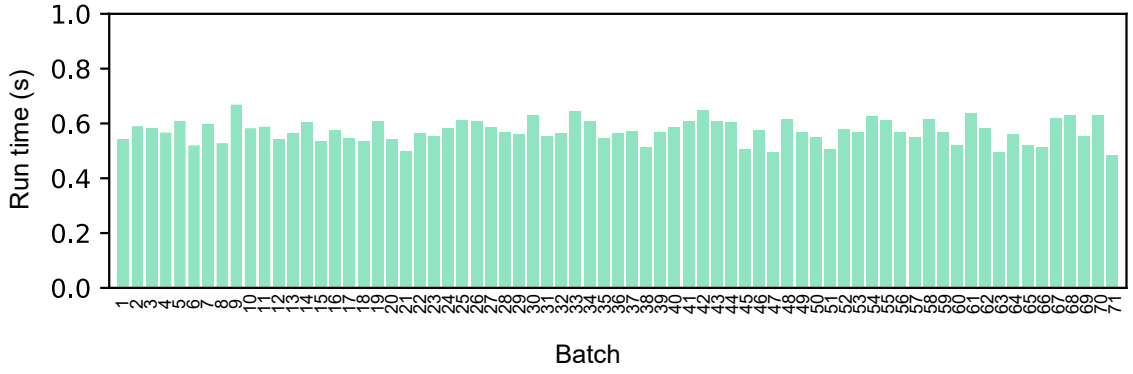

Figure S3. Distribution of run times in 71 batches (MP-20 test dataset (9046 samples), batch size = 128) for XRDSol. The average solution time is about 0.6 seconds per crystal structure with a batch size of 128.

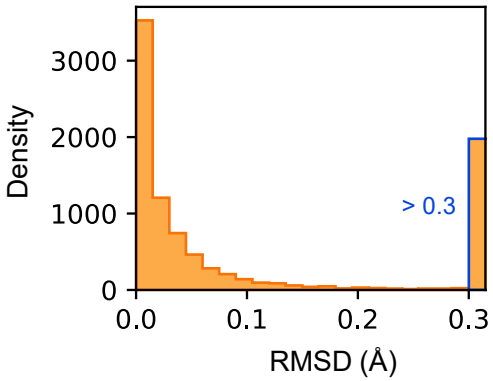

Figure S4. Distribution of root mean square deviation (RMSD) for 9046 structures solved by XRDSol in the MP-20 test dataset, reflecting the reconstruction quality of the crystal structures.

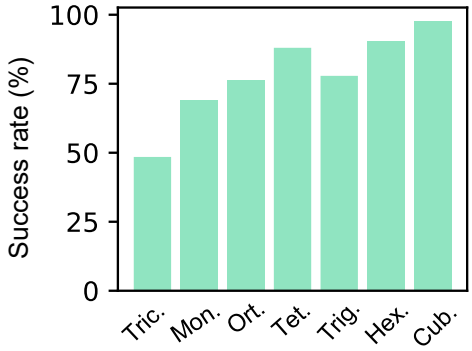

Figure S5. Success rate across the 7 crystal systems, including triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, and cubic.

To investigate the potential bias on certain symmetry or space groups, we evaluated XRDSol's performance on the MP-20 benchmark test dataset. This dataset contains 9046 diverse structures, including structures with low-symmetry and structures from various space groups. We first evaluated XRDSol's performance across different crystal systems, which exhibit different degrees of symmetry. For example, the cubic system has the highest symmetry with multiple high-order rotation axes and the triclinic system has the lowest symmetry. As shown in Fig. S5, the success rate varies across the seven crystal systems with different symmetries (cubic 97.72%, hexagonal 90.33%, trigonal 77.86%, tetragonal 88.02%, orthorhombic 76.32%, monoclinic 68.91%, and triclinic 48.48%). Among these, the cubic crystal system, which has the highest symmetry, shows the best performance. In contrast, the triclinic system exhibits the poorest accuracy. The model's performance decreases as the symmetry decreases. It indicates that structures with higher symmetry are easier to solve accurately, while lower-symmetry structures pose greater challenges. We further analyzed model accuracy across different space groups. In this work, we exclude those space groups with 10 or fewer samples in the MP-20 test dataset, as they would lack statistical significance. As shown in Fig. S6, the accuracy varies across different space groups. Generally, space groups with higher symmetry tend to exhibit higher estimation accuracy. Specifically, 26 space groups achieved an accuracy of 100%, and 41 space groups had accuracy greater than 90%. However, there are 6 space groups with accuracy lower than 50%, namely space groups 2, 19, 46, 51, 55, and 156. This suggests that symmetry plays a significant role in the model's ability to predict crystal structures accurately.

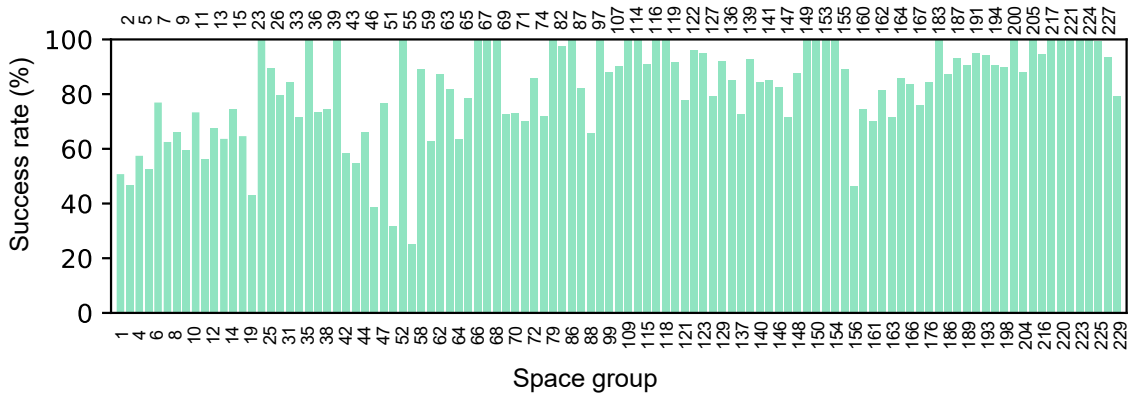

Figure S6. Success rate across the 107 space groups.

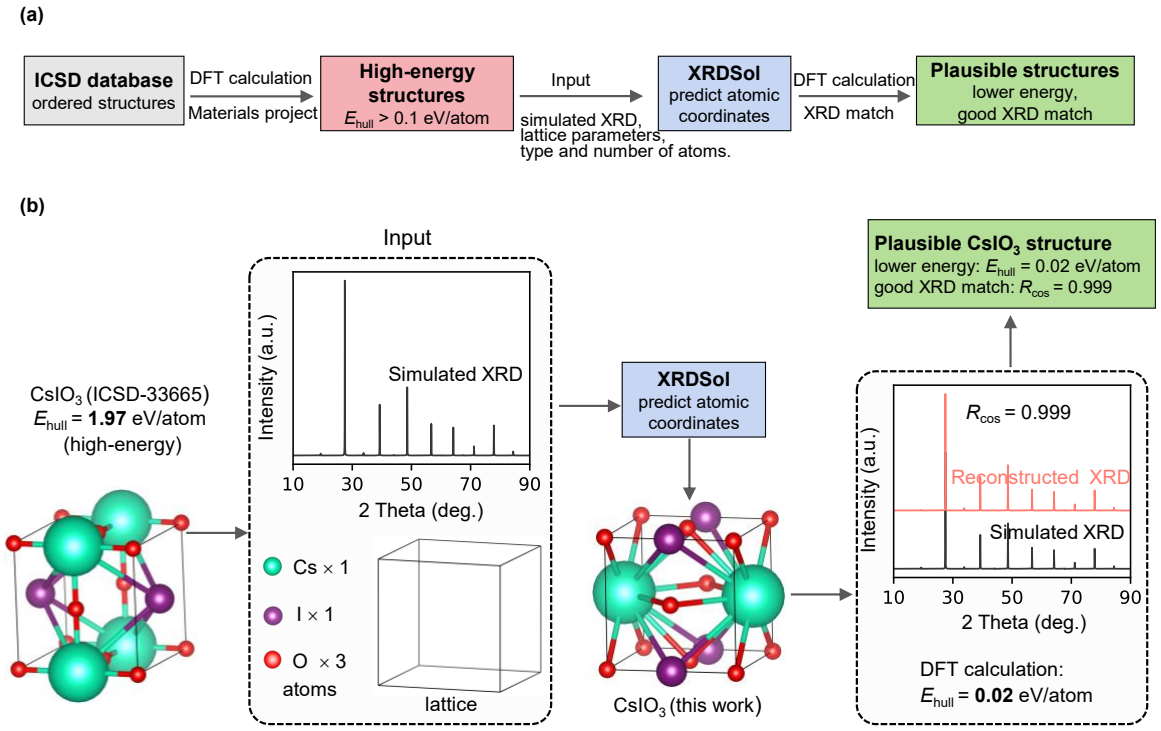

Figure S7. Illustration of revisiting the ICSD high-energy structures. (a) Workflow of XRDSol as a quality control step for entries in ICSD. (b) Illustration using $CsIO_3$ (ICSD-33665) from the ICSD database as an example. All crystal structures were visualized using VESTA[1].

We illustrate this process using $CsIO_3$ (ICSD-33665) reported in 1928 from ICSD database as an example (Figure. S7). In this early study, $CsIO_3$ was reported to be a covalent perovskite material with a cubic perovskite structure in the $Pm\bar{3}m$ space group, where Cs occupies the six-coordinate B-site. The calculated E_{hull} of this original structure (ICSD-33665) is extremely high at 1.97 eV/atom, which strongly suggesting an incorrect atomic arrangement. We simulated corresponding XRD pattern and extracted the lattice parameters, number and type of atoms based on the originally reported $CsIO_3$ structure. We used these as inputs to XRDSol. 5 atoms (1 Cs, 1 I, and 3 O atoms) with random atomic coordinates were initialized in a primitive cell ($a = b = c =$ 4.67 Å, $\alpha = \beta = \gamma = 90°$). XRDSol then predicted atomic coordinates to reconstruct $CsIO_3$ crystal structure. In the $CsIO_3$ structure proposed by XRDSol, the larger Cs cations are moved to the twelve-coordinate A-sites. Subsequent DFT calculations confirmed the thermodynamic stability of this solution with a significantly lower E_{hull} of 0.02 eV/atom. Moreover, the structure solution is very close to the experimentally resolved single-crystal structure of $CsIO_3$[2]. The study clarified that the earlier assignment of $Pm\bar{3}m$ space group in 1928 using powder X-ray diffraction was inaccurate. Thus, this case demonstrated XRDSol's capability to identify and correct some suspicious structure reports.

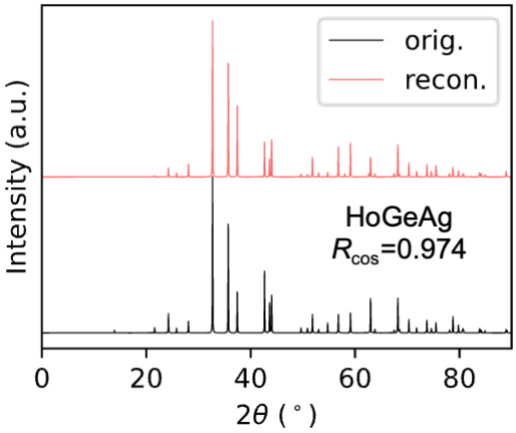

Figure S8. The original (black) and reconstructed (red) XRD patterns of HoGeAg.

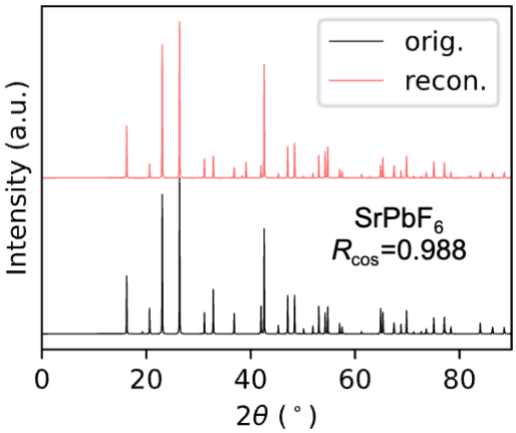

Figure S9. The original (black) and reconstructed (red) XRD patterns of $SrPbF_6$.

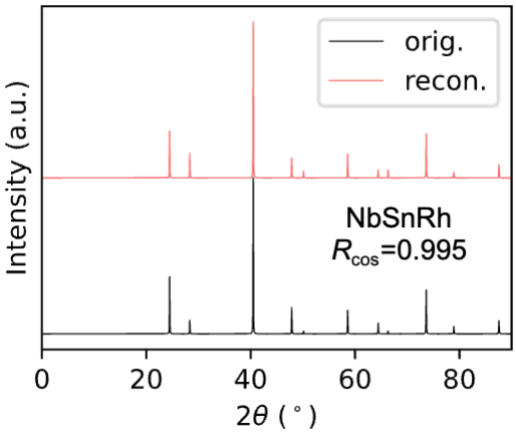

Figure S10. The original (black) and reconstructed (red) XRD patterns of NbSnRh.

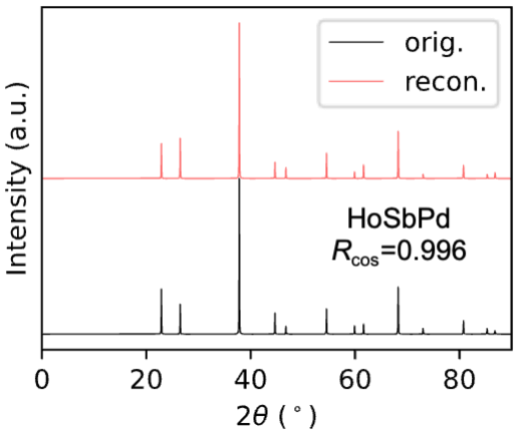

Figure S11. The original (black) and reconstructed (red) XRD patterns of HoSbPd.

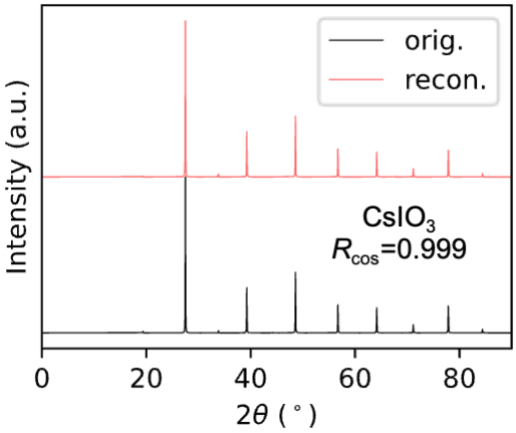

Figure S12. The original (black) and reconstructed (red) XRD patterns of $CsIO_3$.

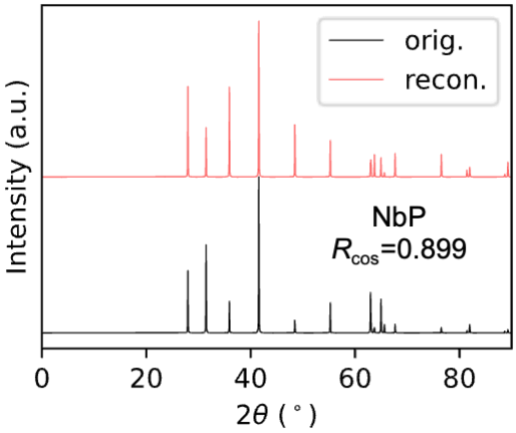

Figure S13. The original (black) and reconstructed (red) XRD patterns of NbP.

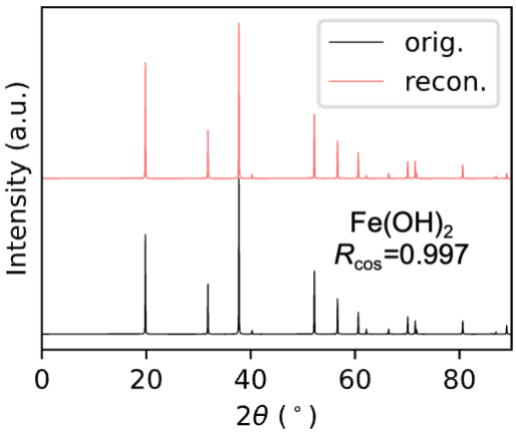

Figure S14. The original (black) and reconstructed (red) XRD patterns of $Fe(OH)_2$.

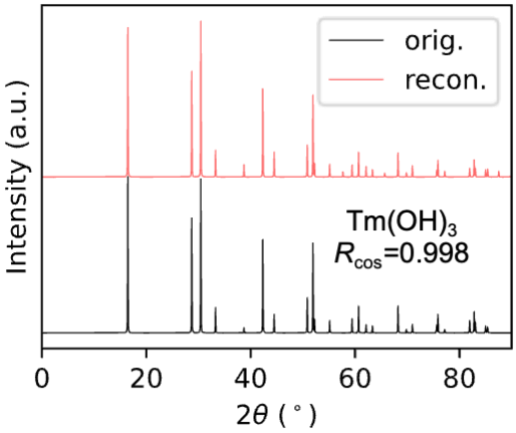

Figure S15. The original (black) and reconstructed (red) XRD patterns of $Tm(OH)_3$.

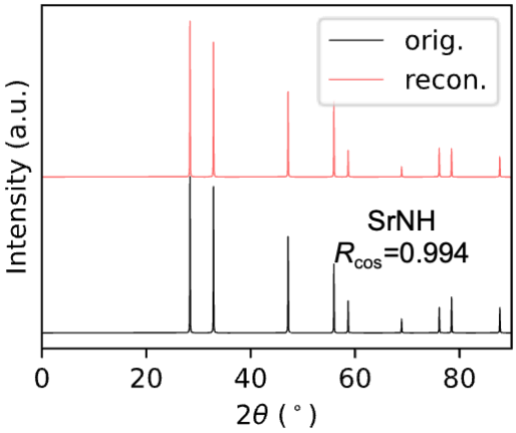

Figure S16. The original (black) and reconstructed (red) XRD patterns of SrNH.

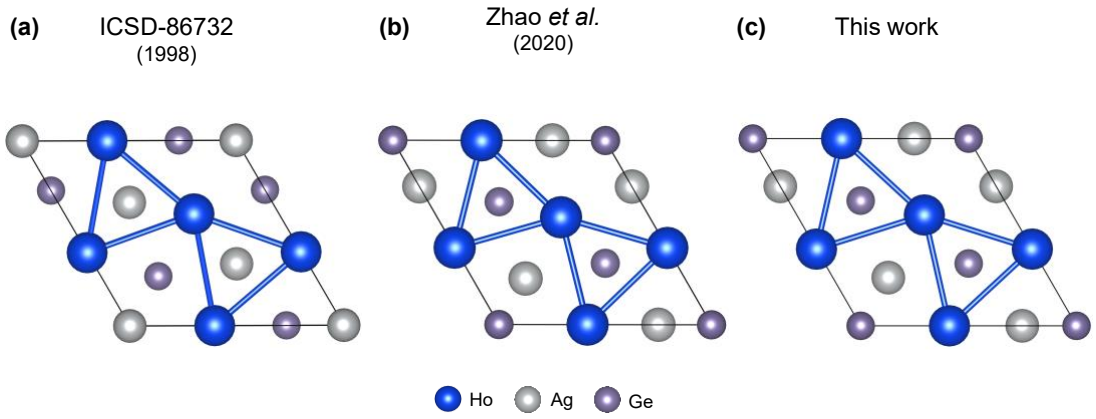

Figure S17. Structural comparison of HoGeAg. (a) ICSD-reported structure[3] (ICSD-86732); (b) Recently reported experimental structure[4]; (c) Structure proposed by this work, consistent with (b). All crystal structures were visualized using VESTA[1].

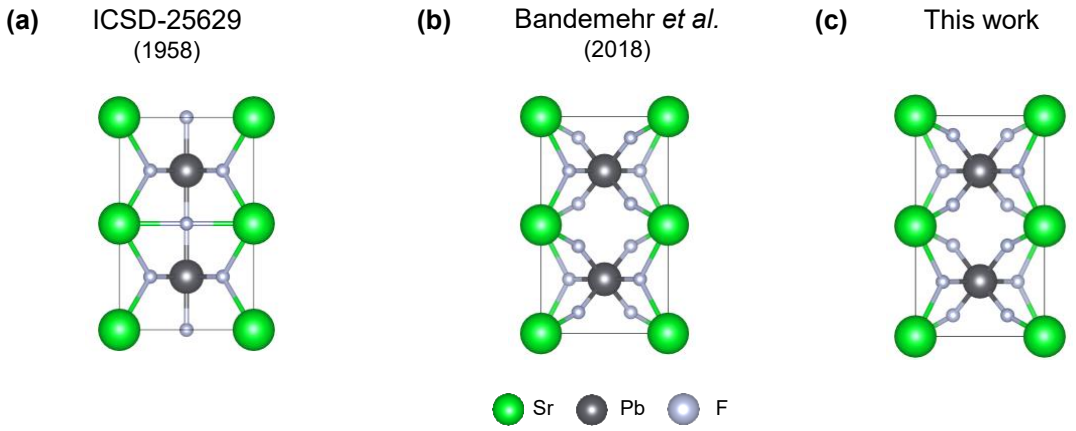

Figure S18. Structural comparison of $SrPbF_6$. (a) ICSD-reported structure (ICSD-25626)[5]; (b) Revisited structure by Bandemehr *et al.*[6]; (c) Structure proposed by this work, consistent with (b). All crystal structures were visualized using VESTA[1].

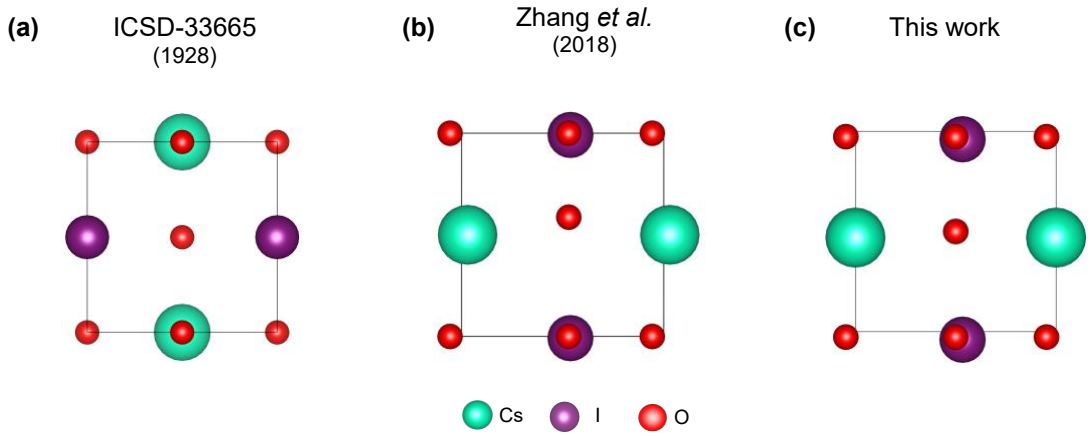

Figure S19. Structural comparison of $CsIO_3$. (a) ICSD-reported structure (ICSD-33665)[7]; (b) Structure recently reported by Zhang *et al.* characterized based on a single crystal sample[2]; (c) Structure proposed by this work, consistent with (b). All crystal structures were visualized using VESTA[1].

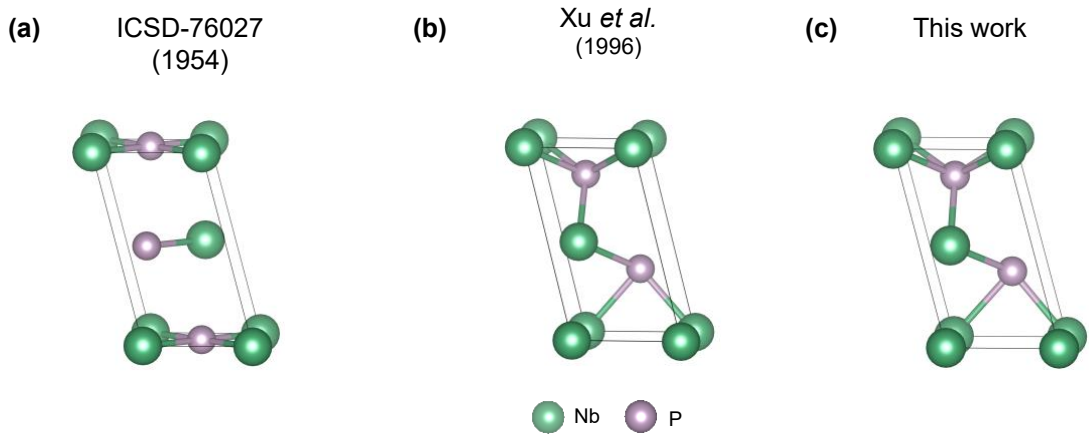

Figure S20. Structural comparison of NbP. (a) ICSD-reported structure (ICSD-33665); (b) Previously reported experimental structure from a single crystal sample[8]; (c) Structure proposed by this work, consistent with (b). All crystal structures were visualized using VESTA[1].

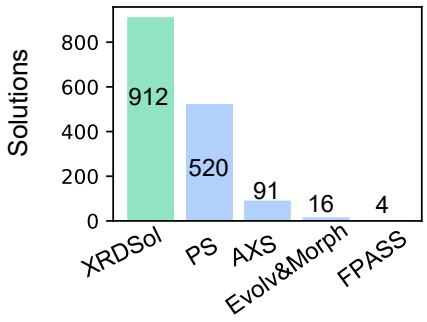

Figure S21. Comparison between solutions obtained from XRDSol and several baseline models, including Prototype search (PS)[9], AXS[10], Evolv&Morph[11], and FPASS[12].

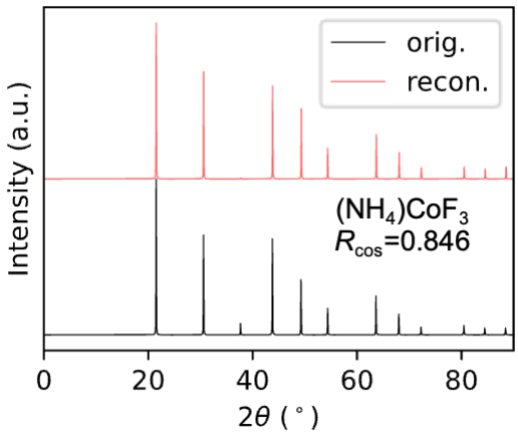

Figure S22. The original (black) and reconstructed (red) XRD patterns of $(NH_4)CoF_3$.

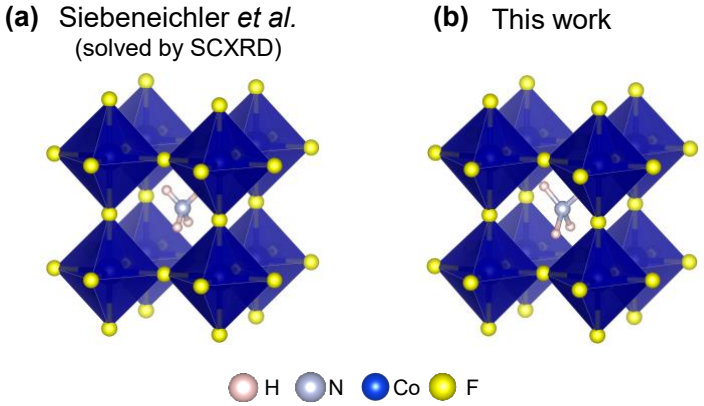

Figure S23. Structural comparison of $(NH_4)CoF_3$. (a) Previously reported experimental structure from a single crystal[13]; (b) Structure proposed by this work, consistent with (a). All crystal structures were visualized using VESTA[1].

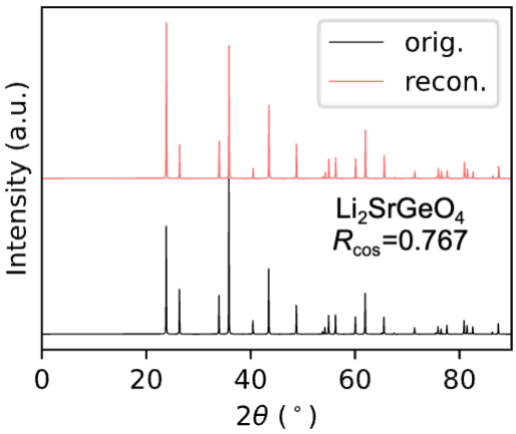

Figure S24. The original (black) and reconstructed (red) XRD patterns of Li_2SrGeO_4.

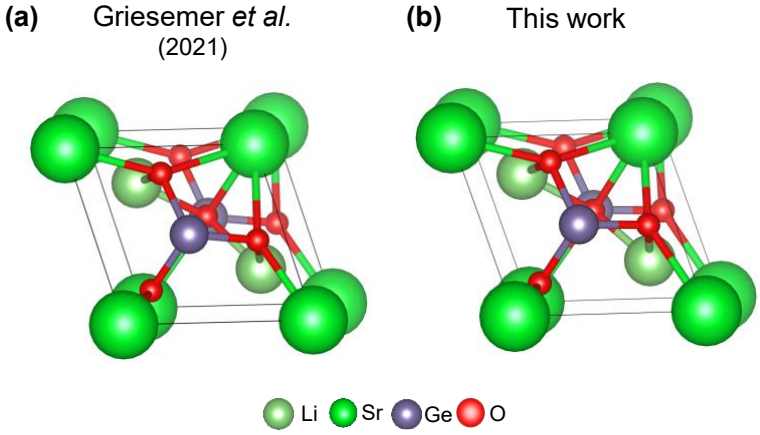

Figure S25. Structural comparison of Li_2SrGeO_4. (a) Previously reported structure[9]; (b) Structure proposed by this work, consistent with (a). All crystal structures were visualized using VESTA[1].

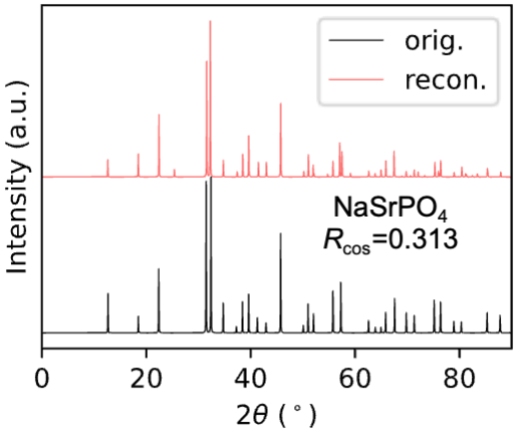

Figure S26. The original (black) and reconstructed (red) XRD patterns of $NaSrPO_4$.

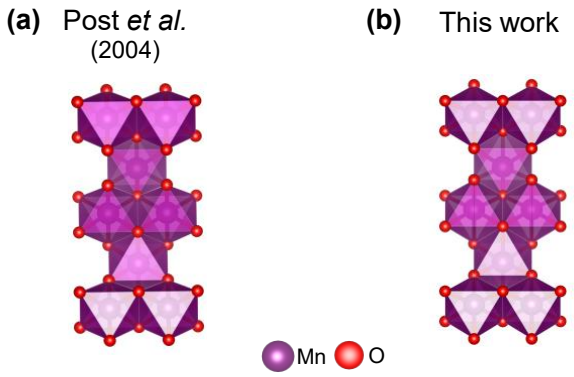

Figure S27. Structural comparison of MnO_2. (a) Previously reported experimental structure[14]; (b) Structure proposed by this work, consistent with (a). All crystal structures were visualized using VESTA[1].

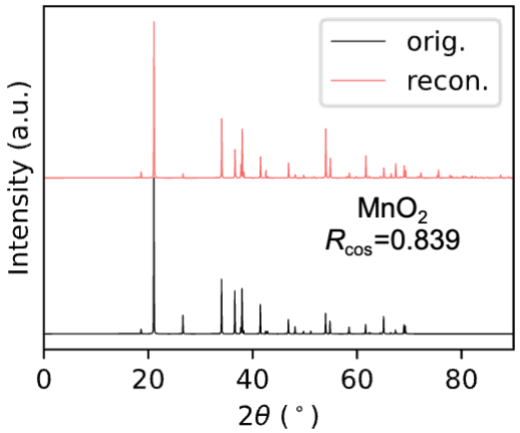

Figure S28. The original (black) and reconstructed (red) XRD patterns of MnO_2.

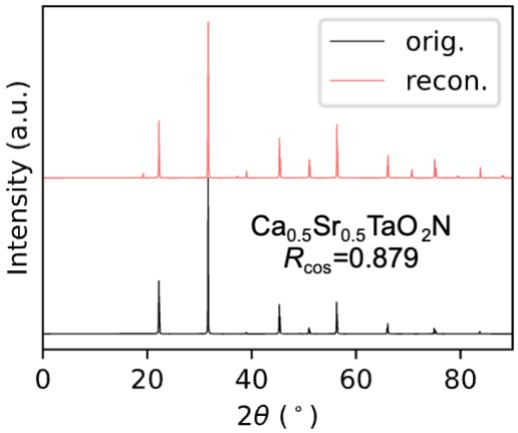

Figure S29. The original (black) and reconstructed (red) XRD patterns of $Ca_{0.5}Sr_{0.5}TaO_2N$.

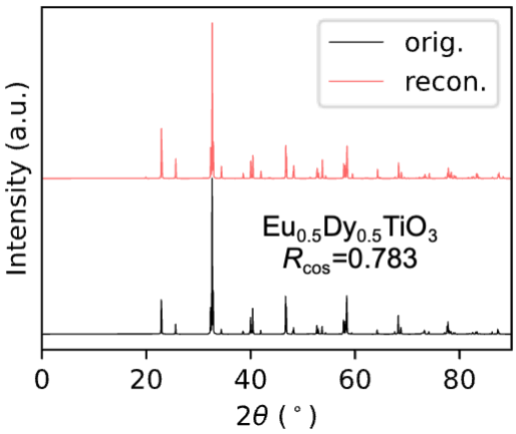

Figure S30. The original (black) and reconstructed (red) XRD patterns of $Eu_{0.5}Dy_{0.5}TiO_3$.

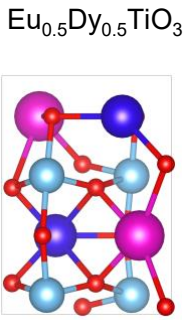

Figure S31. The solved $Eu_{0.5}Dy_{0.5}TiO_3$ structure. The crystal structure was visualized using VESTA[1].

Supplementary Table 6. Pseudo-code of XRDSol for crystal structure solutions using powder XRD data.

Input: PXRD pattern C, initial structure $S_0 = \{L, A, P_0\}$, diffusion steps T, schedule $\{\alpha_t, \sigma_t\}$
$f_C \leftarrow \mathrm{MLP}(C)$
Noising process:
for t = 1 to T do
    $\varepsilon_t \sim \mathcal{N}_w\ (0,\ \mathrm{I})$
    $S_t \leftarrow\ \alpha_t \cdot S_{t-1} + \sigma_t \cdot \varepsilon_t$
end for
Denoising process:
$S_T \leftarrow \{L, A, P_T\}$
for t = T to 1 do
    $f_{struct} \leftarrow \mathrm{EGNN}\ (S_t)$
    $f_{cat}\ \leftarrow \mathrm{concat}(f_{struct}, f_C)$
    $\varepsilon_t^{\wedge} \leftarrow \varphi(f_{cat})$
    $S_{t-1} \leftarrow S_t - \sigma_t \cdot \varepsilon_t^{\wedge}$
end for
Output: $S_0 = \{L, A, P_0\}$

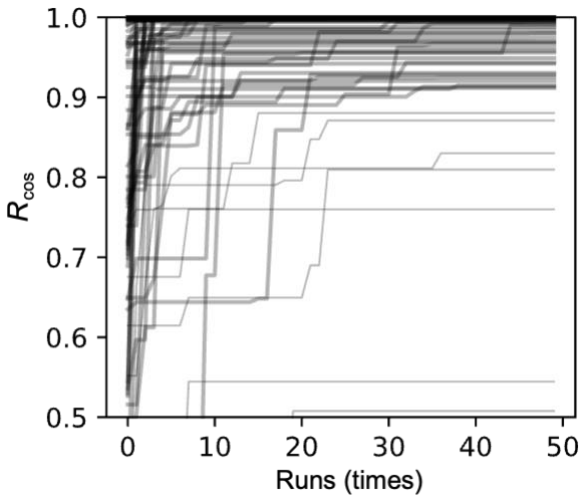

Figure S32. The R_{cos} of 128 samples (first batch) within 50 runs.

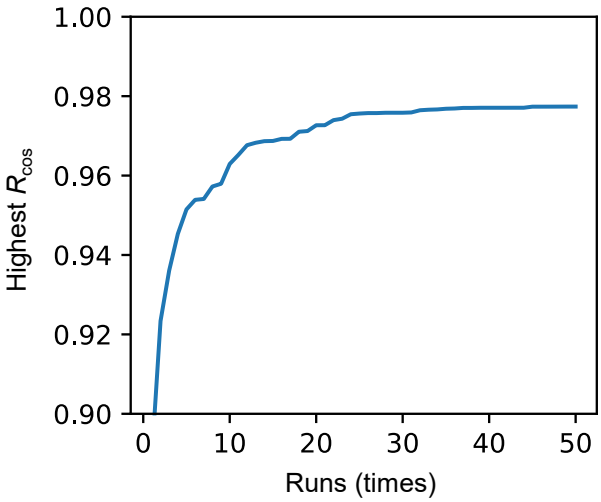

Figure S33. The highest R_{cos} within 50 runs.

Supplementary References

1 Momma, K. & Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *Journal of Applied Crystallography* **44**, 1272-1276, doi:10.1107/S0021889811038970 (2011).

2 Zhang, M., Hu, C., Abudouwufu, T., Yang, Z. & Pan, S. Functional Materials Design via Structural Regulation Originated from Ions Introduction: A Study Case in Cesium Iodate System. *Chemistry of Materials* **30**, 1136-1145, doi:10.1021/acs.chemmater.7b05252 (2018).

3 Baran, S. *et al.* Magnetic order in RAgGe (R=Gd–Er) intermetallic compounds. *Journal of Alloys and Compounds* **281**, 92-98, doi:10.1016/S0925-8388(98)00721-X (1998).

4 Zhao, K. *et al.* Realization of the kagome spin ice state in a frustrated intermetallic compound. *Science* **367**, 1218-1223, doi:10.1126/science.aaw1666 (2020).

5 Hoppe, R. & Blinne, K. Erdalkalihexafluoroplumbate(IV). *Zeitschrift für anorganische und allgemeine Chemie* **293**, 251-263, doi:10.1002/zaac.19582930504 (1958).

6 Bandemehr, J., Deubner, H. L., Sachs, M. & Kraus, F. Li_2PbF_6 and $SrPbF_6$ Revisited. *Zeitschrift für anorganische und allgemeine Chemie* **644**, 1721-1726, doi:10.1002/zaac.201800299 (2018).

7 Zachariasen, W. H. Untersuchungen über die Kristallstruktur von Sesquioxyden und Verbindungen ABO_3. **42**, 797-797, doi:10.1002/ange.19290423015 (1928).

8 Xu, J. *et al.* Crystal Structure, Electrical Transport, and Magnetic Properties of Niobium Monophosphide. *Inorganic Chemistry* **35**, 845-849, doi:10.1021/ic950826f (1996).

9 Griesemer, S. D., Ward, L. & Wolverton, C. High-throughput crystal structure solution using prototypes. *Physical Review Materials* **5**, 105003, doi:10.1103/PhysRevMaterials.5.105003 (2021).

10 Ling, H., Montoya, J., Hung, L. & Aykol, M. Solving inorganic crystal structures from X-ray powder diffraction using a generative first-principles framework. *Computational Materials Science* **214**, 111687, doi:10.1016/j.commatsci.2022.111687 (2022).

11 Lee, J., Oba, J., Ohba, N. & Kajita, S. Creation of crystal structure reproducing X-ray diffraction pattern without using database. *npj Computational Materials* **9**, doi:10.1038/s41524-023-01096-3 (2023).

12 Meredig, B. & Wolverton, C. A hybrid computational–experimental approach for automated crystal structure solution. *Nature Materials* **12**, 123-127, doi:10.1038/nmat3490 (2013).

13 Siebeneichler, S., Dorn, K. V., Smetana, V., Valldor, M. & Mudring, A.-V. A soft chemistry approach to the synthesis of single crystalline and highly pure $(NH_4)CoF_3$ for optical and magnetic investigations. *The Journal of Chemical Physics* **153**, 104501, doi:10.1063/5.0023343 (2020).

14 Post, J. E. & Heaney, P. J. Neutron and synchrotron X-ray diffraction study of the structures and dehydration behaviors of ramsdellite and “groutellite”. **89**, 969-975, doi:10.2138/am-2004-0706 (2004).