scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A database of computed Raman spectra of inorganic compounds with accurate hybrid functionals

Yuheng Li[1], Damien K. J. Lee[1], Pengfei Cai[1], Ziyi Zhang[1], Prashun Gorai[2] & Pieremanuele Canepa[1,3,4] ✉

Raman spectroscopy is widely applied in identifying local structures in materials, but the interpretation of Raman spectra is non-trivial. An accurate computational database of reference spectra calculated with a consistent level of theory can significantly aid in interpreting measured Raman spectra. Here, we present a database of Raman spectra of inorganic compounds calculated with accurate hybrid functionals in density functional theory. Raman spectra were obtained by calculating dynamical matrices and polarizability tensors for structures from the Inorganic Crystal Structure Database. The calculated Raman spectra and other phonon properties (e.g., infrared spectra) are stored in a MongoDB database publicly shared through a web application. We assess the accuracy of our Raman calculations by statistically comparing ~80 calculated spectra with an existing experimental Raman database. To date, the database contains 161 compounds and is continuously growing as we add more materials computed with our automated workflow.

Background & Summary

Raman scattering is the inelastic scattering of light by either absorbing or creating lattice vibrations (phonons)[1]. Raman spectroscopy provides local structure fingerprints of materials by detecting changes in the frequency of incident monochromatic radiation due to the scattering[2,3]. Materials identification with Raman spectroscopy holds advantages including high chemical sensitivity, tolerance to long-range disorder, and high spatial resolution[4,5]. The non-invasive and non-destructive Raman measurement sees applications in *in-situ* and *operando* characterization modes[6]. Compared to other phonon measurement techniques, such as neutron scattering, Raman spectroscopy offers better frequency resolution, easy sample preparation, relatively inexpensive equipment and modularity of lasers, and its fast operation[3–5]. Thanks to these advantages, Raman spectroscopy is widely used in the characterization of biological and organic molecules, as well as inorganic materials in various fields including energy storage and conversion[6–11], catalysis[12–14], low-dimensional materials[15,16], biomedical applications[17–19], and others. However, the interpretation of experimental Raman spectra may be complicated requiring significant time and effort. This urges the need for accurate computational references to improve the speed and accuracy of Raman spectra's interpretation.

First-principles calculations of lattice dynamics and phonons have gained considerable attention and are pushing forward the sub-field of computational Raman spectroscopy[20–22]. In 2015, Togo and Tanaka built the well-known first-principles phonon database that contains the full phonon dispersion of a large number of inorganic and ordered materials[23]. This database showcases the power of phonopy -a software package developed to predict the phonon properties of ordered crystals in combination with the accuracy enabled by first-principles calculations[24]. More recently, Liang and co-workers reported a database of computed Raman spectra of approximately 55 compounds derived from density-functional perturbation theory (DFPT) calculations and finite difference derivative of the dielectric tensors[25]. Taghizadeh *et al.* applied time-dependent third-order perturbation theory to calculate Raman spectra of 733 two-dimensional (2D) monolayer materials and used the calculated spectra to identify materials from experimental Raman spectra[26]. Popov and co-workers proposed a new average

[1]Department of Materials Science and Engineering, National University of Singapore, 9 Engineering Drive 1, 117575, Singapore, Singapore. [2]Department of Metallurgical and Materials Engineering, Colorado School of Mines, Golden, Colorado, 80401, USA. [3]Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, 117585, Singapore, Singapore. [4]Department of Electrical and Computer Engineering, University of Houston, Houston, Texas, 77204, USA. ✉e-mail: pcanepa@central.uh.edu

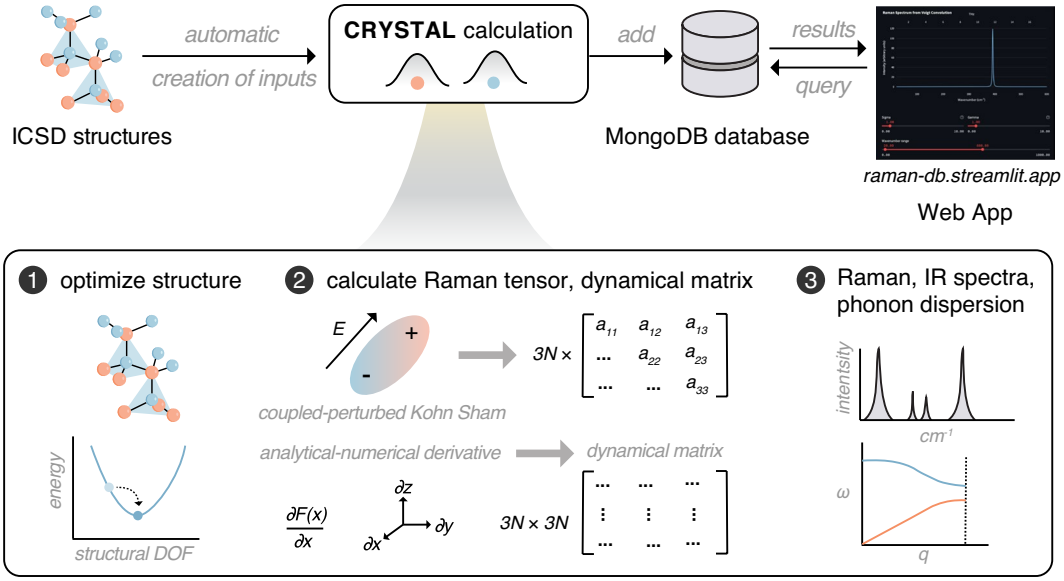

Fig. 1 Computational workflow to build the Raman spectra database: structures of inorganic compounds are from the ICSD experimental database; the structures are automatically converted to CRYSTAL input files; CRYSTAL calculations include structure optimization, coupled-perturbed Kohn-Sham for Raman tensors, and numerical differentiation of analytical first derivative of energy versus displacement of each atom to get dynamical matrices; the resulting vibration frequencies and Raman intensities yield Raman spectra; the calculations can also produce phonon dispersion and density of states, as well as Infrared (IR) spectra; after CRYSTAL calculations, the processed outputs are stored in the MongoDB database and interfaced through the web application (https://raman-db.streamlit.app/).

method for computing Raman spectra of polar polycrystalline materials from DFPT[27]. Most recently, utilizing force constant matrices from the Togo phonon database, Bagheri *et al*. built a Raman database by only calculating Raman tensors by finite differences[28].

This important progress represents the "bleeding edge" of computational Raman spectroscopy for the purpose of spectra interpretation and further data-driven development. However, one common shortcoming of all these databases is that the predictions are limited to rather low levels of theory, namely the generalized gradient approximation (GGA) exchange-correlation functionals[29,30], while more accurate hybrid functionals (beyond "plain vanilla" GGA) have not been applied in calculating Raman databases. In particular, the 2D Raman library by Taghizadeh *et al*. is based on the Perdew–Burke–Ernzerhof (PBE) functional[26], Liang and co-workers' database was calculated using PBE + U[25,31,32], and the database by Bagheri *et al*. are based on the revised PBE for solids (PBEsol)[28,33]. All these predictions are plagued to a different degree by the pernicious self-interaction error of the semi-local GGA functionals. However, the use of such low-level theories is due to the intrinsic complexity and high computational cost of phonon calculations with higher-level theories, and in particular, their implementation in plane-wave and pseudo-potential codes. Therefore, an efficient and accurate hybrid-functional computational Raman database is in high demand.

In the present data records, we compute Raman spectra together with phonon properties at a consistently high level of theory, namely the hybrid functional PBE0[34], in a linear combination of atomic orbitals, expanded in terms of triple-ζ valence with polarization (pob-TZVP-rev2) Gaussian basis sets[35,36]. These approximations are realized using the CRYSTAL code at a relatively low computation cost thanks to its exploitation of lattice symmetry during frequency calculations[37–39]. Here, the computation of Raman spectra is automated such that experimentally measured structures are "digested" into the CRYSTAL code, followed by the calculation of Raman spectra and other phonon properties, which are subsequently stored into a MongoDB database. The database is interfaced with an interactive user-friendly web application.

Methods

Theory. A small fraction of an incident light beam is scattered when it passes through a substance, or more precisely a material. The inelastic part of the scattering is known as Raman scattering, which changes its frequency by creating or absorbing phonons. From group-theory analysis of crystal symmetry, only particular phonon modes are allowed to take part in Raman scattering. The intensity of the scattering is dependent on the polarizability of materials because the scattering originates from the polarization fluctuations induced by the electric field of the incident electromagnetic radiation. Frequencies of the Raman-active phonon modes and the scattering intensities are calculated to get Raman spectra for inorganic compounds.

Our calculations of Raman spectra are performed using the *ab initio* computational program CRYSTAL[37–39]. The calculations with CRYSTAL have several unique features that facilitate efficient and accurate computation of lattice dynamics: (i) In contrast to most commonly used plane–wave basis sets, the crystalline orbitals are

Mode No.	Mode Symmetry	PBE0	HSE06	R^{2}SCAN	PBE0-D3	PBEsol	PBE	Exp.[10]
4–5	E	25	25	31	33	28	27	32
6	A1	71	71	79	77	70	67	70
7–8	E	72	72	80	92	80	71	84
9	B2	87	86	100	111	97	84	84
10–11	E	97	97	102	114	99	94	—
12	B1	100	100	105	116	100	96	—
13–14	E	120	120	125	127	117	116	—
15–16	E	123	123	131	154	137	120	—
18	A1	140	139	146	165	149	135	—
19–20	E	152	152	158	171	158	148	—
21–22	E	156	156	161	173	161	151	157
23	B1	157	156	162	188	164	152	157
24	B2	169	170	175	198	177	165	—
26	B1	175	175	180	207	184	170	—
27–28	E	196	196	200	218	198	188	—
29–30	E	198	197	203	224	203	190	—
31	A1	210	210	211	227	208	202	214
33	B1	227	226	227	239	219	216	226
34	B2	229	229	229	242	221	218	226
35	B1	272	271	269	280	261	258	276
36	B2	272	271	271	283	261	258	276
37–38	E	283	282	282	293	272	268	283
39–40	E	284	283	283	296	273	270	293
42	A1	415	413	404	421	392	387	413
43	B2	538	535	517	546	507	497	537
44–45	E	539	536	517	547	507	498	537
46	B1	569	567	552	579	541	531	567
47–48	E	571	568	553	580	542	533	570
Mean Absolute Error		3 [2]	3 [2]	8 [9]	11 [12]	14 [16]	17 [21]	—
Maximum Absolute Error		12 [9]	12 [10]	20 [20]	31 [31]	30 [30]	40 [40]	—
Mean Absolute Percentage Error		3% [1%]	3% [1%]	4% [2%]	7% [5%]	5% [4%]	6% [6%]	—

Table 1. Computational Raman frequencies (in cm^{-1}) for Na_3PS_4 ($P\bar{4}2_1c$, ICSD No. 121566) calculated with different levels of theory compared to experimental Raman frequencies measured at 100 K. Mode number and mode symmetry are from calculations on the PBE0 level. The statistics in square brackets are for the frequencies $>100\,cm^{-1}$.

expanded as linear combination of atom centered atomic–orbitals, (ii) the hybrid functional PBE0, with 25% Hartree–Fock exchange[34], is used rather than the semi-local GGA functionals, such as PBE[30], and (iii) point group symmetry of lattice is utilized at multiple levels (i.e., atomic positions, wave functions, etc.) during the calculations of vibration frequencies, which means the calculations are only performed on symmetrically irreducible atoms.

Vibration frequencies are determined in CRYSTAL by calculating and diagonalizing mass-weighted dynamical matrices (bottom box in Fig. 1)[40,41]. The dynamical matrix is a $3n \times 3n$ ($n =$ number of atoms, $3n$ is the number of vibrational normal modes) matrix with components being the second partial derivatives of the DFT total energies versus positional displacements, which are calculated by numerical differentiation of the analytical gradient of the DFT total energy with respect to the atomic positions. Diagonalization of an appropriately symmetrized mass-weighted dynamical matrix then gives the normal modes (eigenvectors) and vibration frequencies (eigenvalues).

In CRYSTAL, Raman intensities are calculated directly through the Raman polarizability tensors obtained via perturbation theory, i.e. the coupled-perturbed Kohn-Sham (CPKS) approach (middle box in Fig. 1)[42–46]. In this approach, a static electric field is applied along different directions to calculate the first-, the second-, and the third-order electric susceptibility tensors, which can be converted into polarizability and dielectric tensors, the first hyperpolarizability tensor, and second hyperpolarizability tensor, respectively. From the electric susceptibility tensors, a 3×3 Raman tensors are calculated for each of the $3n$ normal modes. The Raman intensities can then be calculated using the following equations:

$$I_{i,j}(n) = VX(n, i, j)^2, \tag{1}$$

Mode No.	Mode Symmetry	PBE0	HSE06	R^{2}SCAN	PBE0-D3	PBEsol	PBE	Exp.[56]
4–5	E	127	126	134	135	125	120	128
6	A1	212	211	240	247	228	206	206
7–8	E	254	254	261	264	246	242	263
9	A1	334	334	329	332	309	314	354
11–12	E	372	372	370	373	351	353	394
13–14	E	428	427	432	440	406	400	402
15	A1	450	450	460	469	440	428	464
17–18	E	686	685	689	702	669	654	697
20–21	E	787	786	789	803	770	753	797
—	(E)	—	—	—	—	—	—	809*
22–23	E	1072	1071	1054	1071	1023	1010	1066*
25	A1	1090	1089	1074	1092	1044	1031	1083
26–27	E	1157	1156	1133	1150	1100	1096	1161
—	(E)	—	—	—	—	—	—	1232*
Mean Absolute Error		11	12	16	14	30	34	—
Maximum Absolute Error		26	26	35	41	61	64	—
Mean Absolute Percentage Error		3%	3%	4%	4%	6%	6%	—

Table 2. Computational Raman frequencies (in cm^{-1}) for α-quartz SiO_2 ($P3_221$, ICSD No. 156197) calculated with different levels of theory and compared to experimental Raman frequencies. Mode number and mode symmetry are from calculations at the PBE0 level. The experimental frequencies marked with * show barely identifiable intensities as reported in the original spectrum[56]. Their mode symmetries assigned in the experiment are indicated in parentheses when our calculations do not show the mode. Furthermore, the vibration at ~809 cm^{-1} belongs to the degenerate symmetry mode E and can be matched with modes 20–21 indexed by the same symmetry.

where $I_{ij}(n)$ is the directional Raman intensity for single crystals, $X(n, i, j)$ is the i, j-th component of the Raman tensor expressed in the basis of normal mode n, and V is cell volume. Placzek rotation invariants[5] are used for averaging the mode Raman tensors of single crystals with Eqs. 2–4:

$$G_0(n) = \frac{1}{3}\left[\sum_i X(n, i, i)\right]^2, \tag{2}$$

$$G_1(n) = \frac{1}{2}\sum_{i,j}[X(n, i, j) + X(n, j, i)]^2, \tag{3}$$

$$G_2(n) = \frac{1}{3}\sum_{i,j}[X(n, i, i) - X(n, j, j)]^2. \tag{4}$$

The parallel, perpendicular, and total Raman intensities of powders of polycrystalline materials are computed using Eqs. 5–7:

$$I_{\text{par}}(n) = V\,[15G_0(n) + 6(G_1(n) + G_2(n))], \tag{5}$$

$$I_{\text{perp}}(n) = \frac{9}{2}V\,[G_1(n) + G_2(n)], \tag{6}$$

$$I_{\text{tot}}(n) = I_{\text{par}}(n) + I_{\text{perp}}(n). \tag{7}$$

Note that intensities can be conveniently recalculated from available Raman tensors with different measurement conditions including temperature and the incident laser frequency of the Raman instrument. Intensities for infrared (IR) absorption are also calculated in the CPKS process and conveniently stored in our database. Symmetry and group theory are applied in the calculations to differentiate Raman- and IR-active vibrational modes.

Workflow. The development of the present database consisted of three major phases: I. Collecting crystal structures and automatically generating input files for CRYSTAL, II. performing first-principles CRYSTAL calculations, troubleshooting their successful completion followed by data collection, and III. curation of the data with a MongoDB database interfaced by a user-friendly website (Fig. 1).

Key	Datatype	Size	Description
natoms	scalar	—	Number of atoms
energetics	array	5	Total DFT energy, zero-point energy, pV, TS, and heat capacity
dielectric	array	3×3	Dielectric tensor (First-order electric susceptibility or polarizability)
electric_sus_2	array	$3 \times 3 \times 3$	Second-order electric susceptibility tensor (or first hyperpolarizability)
electric_sus_3	array	$3 \times 3 \times 3 \times 3$	Third-order electric susceptibility tensor (or second hyperpolarizability)
tens_ir	array	3natoms × 3	Infrared (IR) tensor
tens_raman	array	3natoms × 6	Raman tensor (six independent tensor components for each mode)
intens_raman_single	array	3natoms × 3 × 3	Raman intensities for single crystal, which is a 3×3 tensor for each mode
intens_raman_multi	array	1D vector	Raman intensities for polycrystal
born	array	natoms × 3 × 3	Born charge matrix for each atom
born_trace	array	natoms	Born charge trace for each atom
born_mode	array	3natoms × 3	Born charge in the basis of normal modes
dyn_matrix	array	3natoms × 3natoms	Symmetric dynamical matrix
vib_freq	array	3natoms	Vibration frequencies (eigenvalues of dyn_matrix)
mode_symmetry	string	—	Symmetry of vibrational modes (irreducible representation)
spectra_raman	array	1D vector	Raman spectra processed via a Voigt convolution
incident_laser	scalar	—	Incident laser wavelength (in nm) used in the calculation
measure_temp	scalar	—	Measurement temperature (in K) used in the calculation
q_points	dict	—	Q-points and coordinates in reciprocal lattice
phonon_dispersion	array	vector	Frequency versus wave vector
phonon_tdos	array	1D vector	Total phonon density of states versus energy/frequency
phonon_pdos	array	1D vector	Projected phonon density of states versus energy/frequency
spectra_ins	array	1D vector	Inelastic neutron scattering spectra, intensity versus energy transfer

Table 3. Description of the name key, data type, and size for the computational properties as stored in the present database using a nested JSON structure.

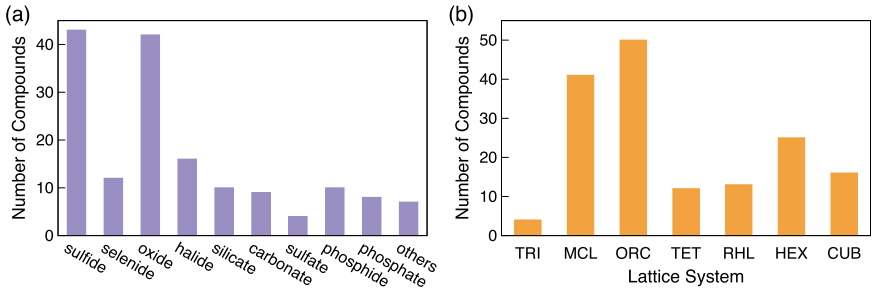

Fig. 2 Database statistics. (**a**) Number of compounds for different chemistries; the selenide bar also includes telluride-based compounds, and the phosphate bar also includes thiophosphate-based compounds; "others" include nitride, nitrate, borate, and carbide. (**b**) Number of compounds for the seven lattice systems: triclinic (TRI), monoclinic (MCL), orthorhombic (ORC), tetragonal (TET), rhombohedral (RHL), hexagonal (HEX), and cubic (CUB).

In Phase I, we first select high-quality experimental structures characterized at low temperature (or room temperature) and ambient pressure from the Inorganic Crystal Structure Database (ICSD)[47]. Then, the experimental structures (i.e., lattice constants, atomic positions, and space groups) are automatically converted into CRYSTAL inputs by symmetry analysis. Specifically, space group and irreducible-atom information needed for CRYSTAL is extracted using the `spglib` interfaced with the `pymatgen` library[48,49]. Exceptions are the monoclinic and orthorhombic structures, which may present alternative definitions of unique axes or origins, and they are processed separately. Rhombohedral structures from ICSD usually take their hexagonal conventional cells, and they are all converted to standard rhombohedral primitive cells for consistent input[50].

In Phase II of our workflow, CRYSTAL calculations are performed for Raman and IR spectra and other phonon-related properties (e.g., phonon dispersion and density of states, etc.). The calculation includes three parts, namely a full structure optimization (i.e., atomic position, lattice parameter, and volume), calculation of vibration frequencies (IR and Raman), and CPKS calculation of Raman intensities. A quasi-Newton algorithm

Compound [ICSD, S.G.]	Raman Table	Compound [ICSD, S.G.]	Raman Table
Al_2SiO_5 [84613, 58]	Al2SiO5_58_icsd84613	P_2S_3 [423037, 11]	P2S3_11_icsd423037
$BaTi(SiO_3)_3$ [18100, 188]	BaTi(SiO3)3_188_icsd18100	$Ca_2ZnSi_2O_7$ [186944, 113]	Ca2ZnSi2O7_113_icsd186944
Sb_2O_3 [240206, 227]	Sb2O3_227_icsd240206	P_4S_3 [406048, 62]	P4S3_62_icsd406048
Li_3PO_4 [257440, 62]	Li3PO4_62_icsd257440	P_2S_5 [409061, 2]	P2S5_2_icsd409061
Sb_2O_3 [414463, 19]	Sb2O3_19_icsd414463	CoS_2 [86351, 205]	CoS2_205_icsd86351
NaS [73173, 194]	NaS_194_icsd73173	HgS [81923, 152]	HgS_152_icsd81923
Na_2O_2 [134060, 189]	Na2O2_189_icsd134060	$MgTiO_3$ [55285, 148]	MgTiO3_148_icsd55285
$BeAl_2O_4$ [72416, 62]	BeAl2O4_62_icsd72416	NaP_5 [99177, 62]	NaP5_62_icsd99177
$Mg_3(BO_3)_2$ [31385, 58]	Mg3(BO3)2_58_icsd31385	NaP_{15} [239084, 2]	NaP15_2_icsd239084
P_2O_3 [24407, 11]	P2O3_11_icsd24407	CoAsS [69129, 29]	CoAsS_29_icsd69129
$Na_2Mg(CO_3)_2$ [100482, 148]	Na2Mg(CO3)2_148_icsd100482	As_4S_5 [16107, 11]	As4S5_11_icsd16107
FeAsS [185809, 14]	FeAsS_14_icsd185809	$P2S_7$ [423062, 2]	P2S7_2_icsd423062
P_2Se_5 [74546, 14]	P2Se5_14_icsd74546	$Si_3(BiO_3)_4$ [69430, 220]	Si3(BiO3)4_220_icsd69430
BeO [15620, 186]	BeO_186_icsd15620	ZnO_2 [38979, 205]	ZnO2_205_icsd38979
$Li_4P_2O_7$ [130427, 153]	Li4P2O7_153_icsd130427	Na_3PS_4 [121567, 114]	Na 3 PS 4 _114_icsd121567
TeO_2 [26844, 61]	TeO2_61_icsd26844	LiP [100465, 14]	LiP_14_icsd100465
As_2S_3 [15239, 14]	As2S3_14_icsd15239	$CuSbS_2$ [171051, 62]	CuSbS2_62_icsd171051
PS [1703, 15]	PS_15_icsd1703	$BaBe_2Si_2O_7$ [100030, 62]	BaBe2Si2O7_62_icsd100030
Sb_2Te [69557, 164]	Sb2Te_164_icsd69557	P_2S_7 [423061, 14]	P2S7_14_icsd423061
Te_3As_2 [18208, 12]	Te3As2_12_icsd18208	AsSe [2056, 14]	AsSe_14_icsd2056
FeSbS [24161, 14]	FeSbS_14_icsd24161	TiO_2 [9161, 136]	TiO2_136_icsd9161
AsS [86228, 15]	AsS_15_icsd86228	Li_3P [240861, 194]	Li3P_194_icsd240861
ZnS [67453, 186]	ZnS_186_icsd67453	ZrO_2 [86692, 14]	ZrO2_14_icsd86692
As_2O_3 [238641, 14]	As2O3_14_icsd238641	Li_5Ge_2 [121533, 166]	Li5Ge2_166_icsd121533
TiO_2 [9852, 141]	TiO2_141_icsd9852	HgO [14124, 62]	HgO_62_icsd14124
FeS_2 [42415, 58]	FeS2_58_icsd42415	Li_4GeS_4 [95649, 62]	Li4GeS4_62_icsd95649
As_2S_3 [185819, 2]	As2S3_2_icsd185819	PbO [15466, 129]	PbO_129_icsd15466
GeS_2 [44, 7]	GeS2_7_icsd44	$Zn(AsO_2)_2$ [202249, 14]	Zn(AsO2)2_14_icsd202249
LiI [414242, 186]	LiI_186_icsd414242	$PbSeO_3$ [98376, 11]	PbSeO3_11_icsd98376
Na_3PO_4 [97205, 114]	Na3PO4_114_icsd97205	$SbCl_3$ [8258, 62]	SbCl3_62_icsd8258
CdS [154186, 186]	CdS_186_icsd154186	ZnSe [77091, 216]	ZnSe_216_icsd77091
ZnO [44477, 186]	ZnO_186_icsd44477	As_4S_3 [16105, 62]	As4S3_62_icsd16105
CuAgS [66580, 36]	CuAgS_36_icsd66580	Na_3PS_4 [121566, 114]	Na 3 PS 4 _114_icsd121566
NaS_2 [2586, 122]	NaS2_122_icsd2586	As_8S_9 [98792, 13]	As8S9_13_icsd98792
Al_2ZnO_4 [185709, 227]	Al2ZnO4_227_icsd185709	HgS [38471, 154]	HgS_154_icsd38471
AsS_2 [424590, 4]	AsS2_4_icsd424590	$Ba(NO_3)_2$ [35495, 205]	Ba(NO3)2_205_icsd35495
$PbCO_3$ [166089, 62]	PbCO3_62_icsd166089	ZnTe [77072, 216]	ZnTe_216_icsd77072
Sb_2O_3 [2033, 56]	Sb2O3_56_icsd2033	$P4S_5$ [1995, 11]	P4S5_11_icsd1995
PO_2 [42777, 15]	PO2_15_icsd42777	Li_3PO_4 [77095, 62]	Li3PO4_62_icsd77095
SbO_2 [153156, 15]	SbO2_15_icsd153156	$BaCO_3$ [158378, 62]	BaCO3_62_icsd158378
$CaTa_2O6$ [24091, 62]	CaTa2O6_62_icsd24091	Li_2O [257372, 225]	Li2O_225_icsd257372
Ca_3Si_2O7 [2282, 14]	Ca3Si2O7_14_icsd2282	Sb_2Se_3 [30973, 62]	Sb2Se3_62_icsd30973
$Al_2Si(O_2F)_2$ [59410, 62]	Al2Si(O2F)2_62_icsd59410	TiO_2 [36408, 61]	TiO2_61_icsd36408
ZnS [136503, 216]	ZnS_216_icsd136503	Ag_2HgS_2 [201713, 14]	Ag2HgS2_14_icsd201713
PbClF [39165, 129]	PbClF_129_icsd39165	GeP_3 [16294, 166]	GeP3_166_icsd16294

Table 4. List of computed compounds (No. 1–90) in the database and their Raman information, such as Raman-active frequencies and Raman intensities. The database currently contains 161 compounds and constantly grows. ICSD codes and space group (S.G.) numbers, of compounds are provided.

is used for the optimization of atomic positions, cell parameters, and volumes. Consistent and high-accuracy settings are used for all calculations. The DFT total energy was converged to 10^{-11} Hartree/cell (~2.7×10^{-10} eV/cell) for the self-consistent field (SCF) procedure. The DFT total energy was integrated with the Pack-Monkhorst sampling scheme over large and symmetrized $8 \times 8 \times 8$ k-points grids (`SHRINKING = 8`). Tolerances for Coulomb and exchange integral series were set to 10^{-7} Hartree for both Coulomb overlap and penetration, 10^{-7} Hartree for exchange overlap, and 10^{-9} and 10^{-30} Hartree for exchange penetration. Crystalline orbitals were expanded as a linear combination of atomic orbitals, which are described by Gaussian valence triple-ζ with

Compound [ICSD, S.G.]	Raman Table	Compound [ICSD, S.G.]	Raman Table
ZnF_2 [280605, 136]	ZnF2_136_icsd280605	AsS [360, 14]	AsS_14_icsd360
SiO_2 [90145, 152]	SiO2_152_icsd90145	LiP_5 [88710, 33]	LiP5_33_icsd88710
Li_3P_7 [60774, 19]	Li3P7_19_icsd60774	$CaMgCO_{32}$ [52149, 148]	CaMgCO32_148_icsd52149
Bi_2Se_3 [60205, 62]	Bi2Se3_62_icsd60205	$K_3Na(SO_4)_2$ [26014, 164]	K3Na(SO4)2_164_icsd26014
ZnP_4 [40428, 92]	ZnP4_92_icsd40428	$KNbO_3$ [190923, 160]	KNbO3_160_icsd190923
$NaHCO_3$ [18183, 14]	NaHCO3_14_icsd18183	$NaMgF_3$ [72318, 62]	NaMgF3_62_icsd72318
ZnAs [427612, 61]	ZnAs_61_icsd427612	$Mg_3(PO_4)_2$ [31005, 14]	Mg3(PO4)2_14_icsd31005
As_2Se_3 [29535, 14]	As2Se3_14_icsd29535	SiC [156190, 186]	SiC_186_icsd156190
$CaTiO_3$ [62149, 62]	CaTiO3_62_icsd62149	Ag_3AsS_3 [36352, 15]	Ag3AsS3_15_icsd36352
$CaCO_3$ [15194, 62]	CaCO3_62_icsd15194	As_2O_3 [238665, 14]	As2O3_14_icsd238665
Li_3N [156894, 191]	Li3N_191_icsd156894	$BaSO_4$ [33730, 62]	BaSO4_62_icsd33730
NaP [14009, 19]	NaP_19_icsd14009	Li_2O_2 [50658, 194]	Li2O2_194_icsd50658
SbI_3 [30906, 14]	SbI3_14_icsd30906	CoAsS [610107, 198]	CoAsS_198_icsd610107
AsO_2 [10436, 62]	AsO2_62_icsd10436	FeS_2 [316, 205]	FeS2_205_icsd316
$CuBiS_2$ [34936, 62]	CuBiS2_62_icsd34936	Sb_2O_5 [1422, 15]	Sb2O5_15_icsd1422
$P2O_5$ [79698, 62]	P2O5_62_icsd79698	$CaMg_3(CO_3)_4$ [201729, 155]	CaMg3(CO3)4_155_icsd201729
$AsBr_3$ [26774, 19]	AsBr3_19_icsd26774	$LiCaAlF_6$ [150332, 163]	LiCaAlF6_163_icsd150332
Sb_2PbO_6 [81387, 162]	Sb2PbO6_162_icsd81387	As_2Se_3 [291473, 12]	As2Se3_12_icsd291473
$Ca_2CO_3F_2$ [100607, 60]	Ca2CO3F2_60_icsd100607	SiO_2 [403365, 92]	SiO2_92_icsd403365
$PbSO_4$ [16916, 62]	PbSO4_62_icsd16916	$NaBe_4SbO_7$ [27599, 186]	NaBe4SbO7_186_icsd27599
$KAlSiO_4$ [83449, 159]	KAlSiO4_159_icsd83449	RuS_2 [68472, 205]	RuS2_205_icsd68472
$BaBe_2Si_2O_7$ [20409, 31]	BaBe2Si2O7_31_icsd20409	$BaSi_2O_5$ [100313, 62]	BaSi2O5_62_icsd100313
Ag_2S [262636, 14]	Ag2S_14_icsd262636	$Na_2Si_2O_5$ [34688, 14]	Na2Si2O5_14_icsd34688
SbO_2 [153154, 33]	SbO2_33_icsd153154	PSe [74878, 14]	PSe_14_icsd74878
$TlAsS_2$ [81094, 14]	TlAsS2_14_icsd81094	$CaCO_3$ [18166, 167]	CaCO3_167_icsd18166
Al_2SiO_5 [24275, 58]	Al2SiO5_58_icsd24275	ZnN_6 [430428, 14]	ZnN6_14_icsd430428
$SbCl_5$ [250363, 194]	SbCl5_194_icsd250363	AsI_3 [23003, 148]	AsI3_148_icsd23003
As_2O_3 [238612, 227]	As2O3_227_icsd238612	$ZnCl_2$ [2459, 33]	ZnCl2_33_icsd2459
Li_2S [657596, 225]	Li2S_225_icsd657596	P_4O_9 [300205, 167]	P4O9_167_icsd300205
KBF_4 [9875, 62]	KBF4_62_icsd9875	$MgCO_3$ [80870, 167]	MgCO3_167_icsd80870
SbI_3 [26082, 148]	SbI3_148_icsd26082	Na_2S_5 [430333, 32]	Na2S5_32_icsd430333
Na_2S_5 [38349, 62]	Na2S5_62_icsd38349	Na_3P [433838, 185]	Na3P_185_icsd433838
P_4O_7 [16452, 14]	P4O7_14_icsd16452	$SbBr_3$ [14217, 19]	SbBr3_19_icsd14217
Cu_2O [172174, 224]	Cu2O_224_icsd172174	Sb_2Te_3 [2084, 166]	Sb2Te3_166_icsd2084
WS_2 [202366, 194]	WS2_194_icsd202366	$Na_2Ca(SO_4)_2$ [16901, 15]	Na2Ca(SO4)2_15_icsd16901
Li_3PO_4 [257439, 31]	Li3PO4_31_icsd257439		

Table 5. List of computed compounds (No. 91–161) in the database and their Raman information, such as Raman-active frequencies and Raman intensities. The database currently contains 161 compounds and constantly grows. ICSD codes and space group (S.G.) numbers, of compounds are provided.

polarization (pob-TZVP-rev2) basis sets[35,36]. In DFT the unknown exchange and correlation functional was approximated with the PBE0 hybrid functional, which provides excellent predictions of the experimental Raman spectra of Na_3PS_4 and quartz-SiO_2 (Tables 1, 2)[34].

At the Γ-point, $3n+1$ (with $n=$ number of atoms) total energy and gradient calculations are required for computing the IR and Raman frequencies. The first derivatives of the DFT total energies versus atomic displacements, i.e., the total energy gradients to displacements, are computed with a single ionic displacement (0.003 Å) for each coordinate with respect to their equilibrium position. Born charges are also obtained during the frequency calculations. Based on the input generation in Phase I, all calculations are performed in an automatic and high-throughput manner.

In Phase III, this workflow deals with the post-processing of CRYSTAL outputs and subsequent data curation. Output files are parsed to extract relevant data, such as IR and Raman frequencies, and their intensities. The post-processing also includes a convolution routine to generate Raman spectra based on the Voigt model with adjustable percentages of the Gaussian and the Lorentzian shapes. The computed data can be found in Table 3. All data is stored and organized using a MongoDB database currently hosted in the MongoDB Atlas cloud service. An user-friendly and publicly available web app is built for users to search for compounds and plot their Raman spectra interactively (https://raman-db.streamlit.app/).

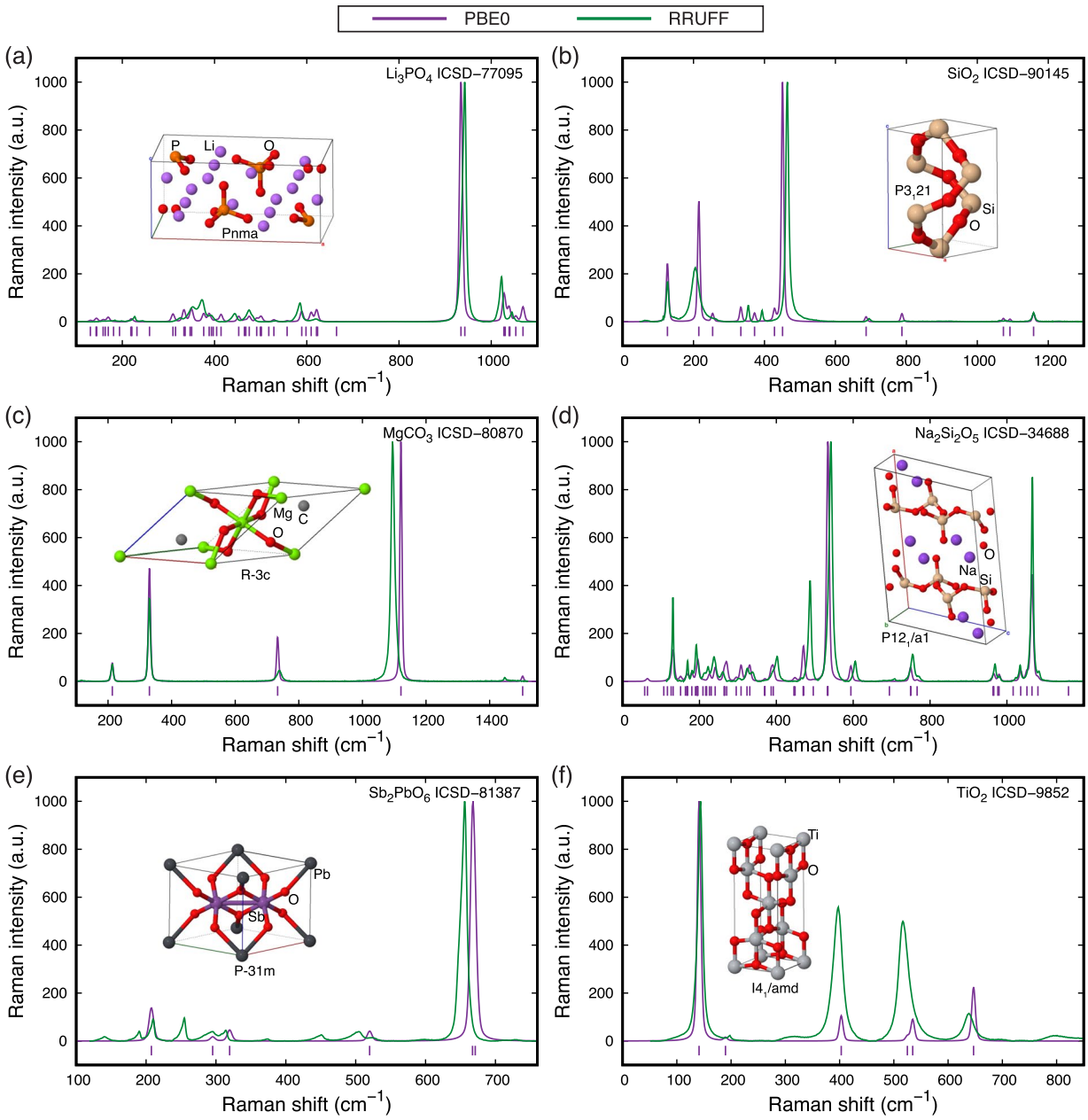

Fig. 3 Raman spectra calculated using our workflow with the hybrid functional PBE0 (purple) in comparison with RRUFF experimental spectra (green) for six inorganic compounds: (**a**) Li_3PO_4 with identified ICSD No. 77095 and space group *Pnma*, (**b**) SiO_2 (ICSD No. 90145 and $P3_121$), (**c**) $MgCO_3$ (ICSD No. 80870 and $R{-}3c$), (**d**) $Na_2Si_2O_5$ (ICSD No. 34688 and $P12_1/a1$), (**e**) Sb_2PbO_6 (ICSD No. 81387 and $P{-}31\,m$), and (**f**) TiO_2 (ICSD No. 9852 and $I4_1/amd$). Raman intensities are in arbitrary units (a.u.) and the maximum peak in each spectrum is normalized to 1000. Raman-active vibration modes are marked as purple ticks at the bottom of each spectrum. All calculated spectra are plotted using a Voigt line shape with 50% Lorentzian and 50% Gaussian. Insets show optimized crystal structures from our calculations. A complete spectra comparison of all the 78 calculated RRUFF compounds can be found in Section S3 of Supporting Information.

Data Records

The computational data is organized in JSON format and stored in a MongoDB database, which is available in our public GitHub repository (https://github.com/caneparesearch/project Raman) and on Zenodo[51]. The computed IR and Raman frequencies, Raman intensities, and other phonon-related properties for any calculated compounds can be accessed directly from the repository. The names and quantities for the computed properties are stored as key-value pairs using a nested JSON structure, and they are elaborated in Table 3. For each computed compound, the database includes the structure from ICSD and the DFT-optimized structures, the DFT total energies, the vibrational entropy, the heat capacity, the electric susceptibilities up to the third order, the Raman and the IR tensors, the Raman intensities, the Born charges, the dynamical matrices, the vibration

frequencies, and their modes' symmetries. The convoluted Raman and IR spectra obtained from the computations, as well as the simulated measurement temperature and incident laser wavelength, are also included. While the current database includes only Γ-point phonons, the database is designed to be compatible with calculations at different Q points other than the Γ-point, phonon dispersions, phonon density of states (DOS), and inelastic neutron scattering (INS) spectra.

The database currently contains 161 calculated inorganic compounds. Figure 2 shows the statistics of the database: There are 43 sulfides, 12 selenides and telluriums, 42 oxides, and peroxides, 16 halides, 10 silicates, 9 carbonates, 4 sulfates, 10 phosphides, 8 phosphates, and thiophosphates, and the remailing 7 compounds are nitrides, nitrates, borates, and carbides, respectively. Most compounds belong to the monoclinic (41) and orthorhombic (50) lattice systems, and there are also 4 triclinic, 12 tetragonal, 13 rhombohedral, 25 hexagonal, and 16 cubic systems, respectively. The 161 compounds computed are listed in Tables 4, 5 with their ICSD codes, space group symbols, and a link to their detailed Raman information (computed outputs), including the calculated Raman- (and IR-) active vibrational modes and intensities.

Technical Validation

To validate our simulated Raman spectra obtained with the hybrid functional and local basis sets, we compare these predictions with experimental spectra directly available in the literature or in the RRUFF database -an integrated database of Raman spectra data for minerals[52].

Computational Raman spectra of two compounds (Na_3PS_4 and α-quartz SiO_2) were benchmarked against experimental Raman spectra from the literature to demonstrate the high accuracy of the hybrid functional PBE0 approximation compared to other functionals. Table 1 shows the computed frequencies of Raman-active vibrational modes of Na_3PS_4 ($P\bar{4}2_1c$, ICSD No. 121566) calculated by different approximations of the unknown exchange and correlation functional, and compared with experimental frequencies[10]. The experimentally observed Raman-active modes are matched to computational ones according to their mode symmetries. Errors of predicted frequencies are calculated with respect to the matched experimental frequencies, and they are listed in the last three rows of Table 1. Among all the exchange and correlation functionals tested, the hybrid functional PBE0 used for this dataset shows the smallest mean absolute error (MAE), maximum absolute error (MaxAE), and mean absolute percentage error (MAPE), of only 3 cm^{-1}, 12 cm^{-1}, and 2.9%, respectively. In contrast, the range–separated hybrid functional by Heyd, Scuseria and Ernzerhof HSE06[53,54] shows comparable accuracy as PBE0 in determining Raman frequencies, but calculations with HSE06 in CRYSTAL are almost twice more expensive than those with PBE0 (e.g., computation time in seconds is 31,492 vs. 12,059 on 128 cores –2 × AMD EPYC 7742– for Na_3PS_4). Notably, PBE0 and HSE06 perform especially well for the prediction of "high"-frequency modes (>100 cm^{-1}). The GGA functionals PBE and PBEsol show the worst performance with MAE, MaxAE, and MAPE of up to 17 cm^{-1}, 40 cm^{-1}, and 6%, respectively. These inaccuracies are not acceptable if computation serves to fingerprint Raman spectra of ordered materials. Meta-GGA functionals in the flavor of R^{2}SCAN show relatively small absolute errors compared to GGA, but the inaccuracies compared to experimental data are still approximately twice (2×) that of PBE0. With the van der Waals correction (Grimmes's D3)[55], PBE0-D3 shows worse accuracy in frequencies than PBE0 alone, with an 11 cm^{-1} MAE. The worse accuracy when including D3 is partially attributed to the ionic character of the Na^+–PS_4^{3-} bonds in Na_3PS_4. As shown by the error values in square brackets of Table 1, PBE0 and HSE06 show decreased MAE for the high-frequency modes, while the high-frequency MAE of all the other functionals increases. A decrease in MaxAE from 12 to 9 cm^{-1} from all frequencies to high frequencies for PBE0 also indicates that the max error ≥10 cm^{-1} only appears at low frequencies (<100 cm^{-1}). This further encourages Raman calculations using PBE0 because high-frequency vibrational modes appear less noisy than soft modes in experimental measurements, and accurate computational reference of the high-frequency modes are more important in facilitating the interpretation of experimental spectra.

The α-quartz SiO_2 ($P3_221$, ICSD No. 156197) is another widely studied compound with an experimental Raman spectrum[56]. Table 2 shows the calculated frequencies with different DFT exchange and correlation functionals and their errors with respect to the experimental reference. Interestingly, three of the reported experimental assignments do not show observable peaks in the Raman spectra from the same study[56], and two of them are also absent in our calculation results. The three non-observable frequencies are marked with * in the last column of Table 2, and the frequency errors were calculated only for the modes identified in our calculations.

As in the case of Na_3PS_4, frequencies calculated with PBE0 appear in good agreement with experimental observations, and thus very accurate. The PBE0 frequencies show the smallest MAE, MaxAE, and MAPE of 11 cm^{-1}, 26 cm^{-1}, and 3%. Except for the computationally more expensive HSE06, all the other exchange and correlation functionals do not show comparable accuracy to PBE0. Vibrational modes are assigned according to the selection rule as IR-active (change in dipole moment) and Raman-active (change in polarizability) modes.

Due to the lack of high-quality Raman data from consistent measurement in literature, we decided to further validate our computational approach by a systematic comparison of calculated spectra with the experimental Raman database RRUFF[52]. We applied our approach to calculating Raman spectra for 78 inorganic compounds available in the RRUFF database, and these spectra have also been added to our database. To generate input following our workflow, the RRUFF compounds were matched to ICSD structures according to their chemical formula and mineral names. The matching was subsequently verified using space group information of Materials Project[57] entries that were assigned to RRUFF compounds in Ref. [28].

Figure 3 shows the comparison of the calculated (purple) and RRUFF experimental (green) Raman spectra for six chosen inorganic compounds of the 78 inorganic compounds considered here: Li_3PO_4 (ICSD No. 77095), SiO_2 (ICSD No. 90145), $MgCO_3$ (ICSD No. 80870), $Na_2Si_2O_5$ (ICSD No. 34688), Sb_2PbO_6 (ICSD No. 81387), and TiO_2 (ICSD No. 9852). The maximum Raman intensities are normalized to 1,000 arbitrary units (a.u.) and

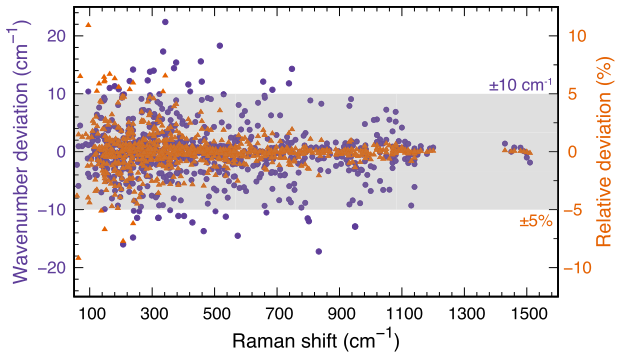

Fig. 4 Wavenumber deviation (purple) and relative deviation (orange) of calculated Raman frequencies compared to the RRUFF experimental Raman frequencies. The grey shaded area indicates a wavenumber-deviation range of $\pm 10\,cm^{-1}$ and a relative-deviation range of $\pm 5\%$. The frequency comparison is only conducted within the common frequency range between computational and experimental spectra. A total number of 63 PBE0-RRUFF spectra are used to extract the deviations (Supplementary Figures S2–S12 of Supporting Information). There are 15 RRUFF spectra found to have experimental artifacts and are not included in this comparison to the computed spectra (Supplementary Figures S13–S15 of Supporting Information).

plotted versus Raman frequencies in cm^{-1}. The calculated spectra are then convoluted with a Voigt line shape consisting of 50% Lorentzian and 50% Gaussian. Purple ticks at the bottom of each spectrum mark the calculated Raman frequencies. Insets show optimized crystal structures from our calculations, which are visualized with JS-ICE[58] together with their computed space groups. The spectra comparison for all the 78 calculated RRUFF compounds can be found in Supplementary Figures S2–S15 of Supporting Information (SI).

The calculated and experimental Raman spectra of the six compounds in Fig. 3 show very good agreement. As shown in the plots, the experimental peaks are all correctly predicted with very small deviations in the values of frequencies. Specifically, 87 out of the total 93 peaks show wavenumber deviations within $\pm 10\,cm^{-1}$ and 91 peaks within $\pm 15\,cm^{-1}$. The relative deviations are within $\pm 3\%$ for 88 out of the total 93 peaks, and within $\pm 5\%$ for 92 peaks. Predictions are especially accurate for the most intense peak in each spectrum, with a maximum deviation of only $\sim -9.3\,cm^{-1}$ in the case of $MgCO_3$ (Fig. 3c). Note that large deviations in intensities are only observed for relatively experimental weak peaks (Fig. 3f). Such accurate prediction of the most intense peak for a compound is very useful when the prediction is used to fingerprint phases from experimental Raman spectra. Notably, the six spectra (Fig. 3) do not represent the best computation–experiment matches among all the 78 spectra (Supplementary Figures S2–S15 of SI).

Besides the individual spectra comparison for each compound, a statistical comparison was also made between calculated and experimental Raman-active frequencies, to provide a more systematic evaluation of the overall computational accuracy. Among all the 78 RRUFF compounds we calculated, the spectra of 15 compounds show experimental artifacts that make it impossible to compare their frequencies with our calculated Raman-active frequencies (Supplementary Figures S13–S15 of SI), and the spectra of the rest 63 compounds (Supplementary Figures S2–S12 of SI) were used for the statistical comparison. Given a pair of computational and RRUFF experimental spectra for one of the 63 compounds, the common frequency range of the spectra was first identified. Then, within the common range, all local maxima of the Raman intensities and their frequencies were located. Finally, each computational intensity maximum (i.e., a peak in the computational Raman spectrum) is uniquely matched to an experimental intensity maximum (i.e., a peak in the experimental Raman spectrum), and the computational frequency deviation from the experiments is calculated for each pair of the matched Raman peaks.

Figure 4 shows the computational frequency deviations from the matched RRUFF Raman-active frequencies, plotted versus Raman shift. The wavenumber deviation is plotted using purple points, and the relative deviation is plotted using orange triangles. The grey shaded area indicates a wavenumber deviation range of $\pm 10\,cm^{-1}$ and a relative deviation range of $\pm 5\%$. There are a total number of 804 pairs of matched peaks from the 63 spectra used for comparison. A percentage of 94.7% (88.9%) wavenumber deviations are within the range of $\pm 10\,cm^{-1}$ ($\pm 8\,cm^{-1}$), and a percentage of 97.6% (92.9%) relative deviations are within the range of $\pm 5\%$ ($\pm 3\%$). The maximum absolute values of wavenumber and relative deviations are $22.4\,cm^{-1}$ and 10.9%, respectively. More data points are found at smaller Raman shift wavenumbers because the Raman-active modes are overall more frequently distributed at lower frequencies for the 63 compounds (Supplementary Figures S2–S12 of SI). Another trend shown in Fig. 4 is that the deviations are generally larger at lower frequencies. A plausible reason for this trend is that experimental measurements are usually more noisy for softer vibrational modes.

Usage Notes

The present database is accessible through our web application (https://raman-db.streamlit.app/). In this web application (Fig. 5), users can search for chemical formulae, select the desired compound according to its ICSD identification number, view the crystal structure, and interactively plot the Raman and IR spectra with different convolution schemes (i.e., Gaussian shapes, Lorentzian shapes, and Voight shapes). The spectra plots can

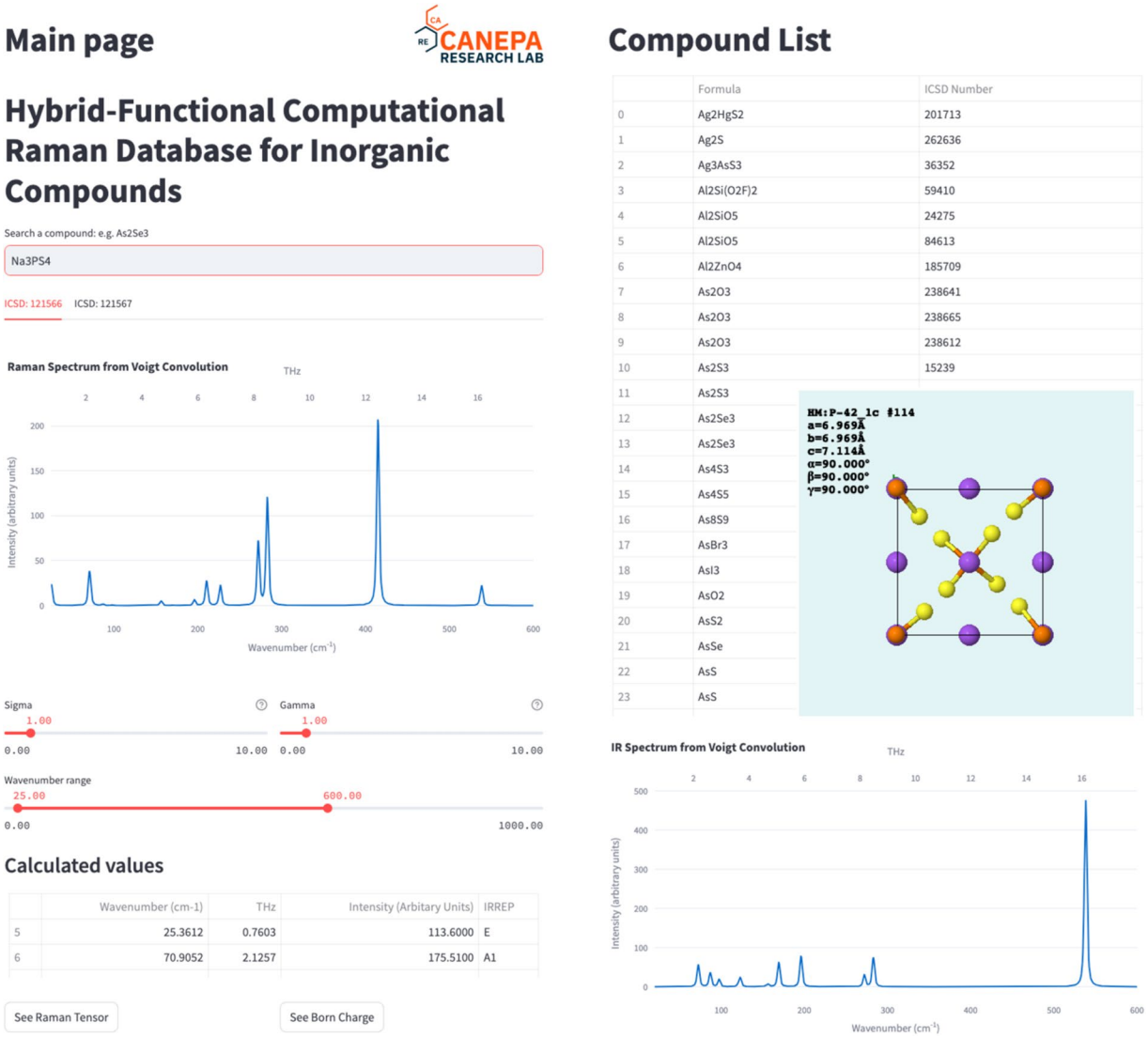

Fig. 5 A demo of the Raman-database web application (https://raman-db.streamlit.app/). The database of computed Raman properties is interfaced by this web application in the ways of searching for compounds, viewing crystal structures, interactively plotting Raman and IR spectra, and query for phonon properties (left and bottom right). A complete list of all available inorganic compounds in the database and their ICSD numbers (right). The list currently contains 161 compounds and is growing.

be downloaded as a PNG file, and the spectra data (frequencies and intensities for the Raman- and IR-active vibrational modes, together with their irreducible representations) can be downloaded as practical CSV files. Relevant quantities, such as Raman tensors and Born charges are also available in the application. A complete list of calculated compounds in the present database is constantly updated and shown on the web page.

Code availability

Data and custom code are available on GitHub repository for the automation of Raman predictions, post-processing, and database construction, under a CC BY 4.0 License[51]. The *ab initio* code CRYSTAL is commercially available, and the TZVP basis sets used (see Supplementary Section S1 of the SI) can be accessed from the CRYSTAL website.

Received: 18 September 2023; Accepted: 5 January 2024;
Published online: 22 January 2024

References

1. Raman, C. V. & Krishnan, K. S. A New Type of Secondary Radiation. *Nature* **121**, 501–502, https://doi.org/10.1038/121501c0 (1928).
2. Ganguly, A. K. & Birman, J. L. Theory of Lattice Raman Scattering in Insulators. *Physical Review* **162**, 806–816, https://doi.org/10.1103/PhysRev.162.806 (1967).
3. Dove, M. T. *Introduction to Lattice Dynamics*, 1 edn (Cambridge University Press, 1993).
4. Schrader, B. *Infrared and Raman Spectroscopy*, 1 edn (Wiley-VCH, Weinheim, 2008).

5. Long, D. A. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules* (Wiley, Chichester, 2002).
6. Sang, L., Haasch, R. T., Gewirth, A. A. & Nuzzo, R. G. Evolution at the Solid Electrolyte/Gold Electrode Interface during Lithium Deposition and Stripping. *Chemistry of Materials* **29**, 3029–3037, https://doi.org/10.1021/acs.chemmater.7b00034 (2017).
7. Rettenwander, D. *et al*. Interface Instability of Fe-Stabilized Li7La3Zr2O12 versus Li Metal. *The Journal of Physical Chemistry C* **122**, 3780–3785, https://doi.org/10.1021/acs.jpcc.7b12387 (2018).
8. Famprikis, T., Canepa, P., Dawson, J. A., Islam, M. S. & Masquelier, C. Fundamentals of inorganic solid-state electrolytes for batteries. *Nature Materials* **18**, 1278–1291, https://doi.org/10.1038/s41563-019-0431-3 (2019).
9. Li, W. *et al*. Unraveling the Origin of Moisture Stability of Halide Solid-State Electrolytes by *In Situ* and *Operando* Synchrotron X-ray Analytical Techniques. *Chemistry of Materials* **32**, 7019–7027, https://doi.org/10.1021/acs.chemmater.0c02419 (2020).
10. Famprikis, T. *et al*. Insights into the Rich Polymorphism of the Na+ Ion Conductor Na_3PS_4 from the Perspective of Variable-Temperature Diffraction and Spectroscopy. *Chemistry of Materials* **33**, 5652–5667, https://doi.org/10.1021/acs.chemmater.1c01113 (2021).
11. Chen, Y. *et al*. Strain engineering and epitaxial stabilization of halide perovskites. *Nature* **577**, 209–215, https://doi.org/10.1038/s41586-019-1868-x (2020).
12. Zhan, C. *et al*. Revealing the CO Coverage-Driven C–C Coupling Mechanism for Electrochemical CO 2 Reduction on Cu 2 O Nanocubes *via Operando* Raman Spectroscopy. *ACS Catalysis* **11**, 7694–7701, https://doi.org/10.1021/acscatal.1c01478 (2021).
13. Nam, D.-H. *et al*. Intermediate Binding Control Using Metal–Organic Frameworks Enhances Electrochemical CO 2 Reduction. *Journal of the American Chemical Society* **142**, 21513–21521, https://doi.org/10.1021/jacs.0c10774 (2020).
14. Zhao, Y. *et al*. Elucidating electrochemical CO 2 reduction reaction processes on Cu(*hkl*) single-crystal surfaces by *in situ* Raman spectroscopy. *Energy & Environmental Science* **15**, 3968–3977, https://doi.org/10.1039/D2EE01334G (2022).
15. Ferrari, A. C. & Basko, D. M. Raman spectroscopy as a versatile tool for studying the properties of graphene. *Nature Nanotechnology* **8**, 235–246, https://doi.org/10.1038/nnano.2013.46 (2013).
16. Zhang, S. *et al*. Spotting the differences in two-dimensional materials – the Raman scattering perspective. *Chemical Society Reviews* **47**, 3217–3240, https://doi.org/10.1039/C7CS00874K (2018).
17. Abramczyk, H. & Brozek-Pluska, B. Raman Imaging in Biochemical and Biomedical Applications. Diagnosis and Treatment of Breast Cancer. *Chemical Reviews* **113**, 5766–5781, https://doi.org/10.1021/cr300147r (2013).
18. Nicolson, F., Kircher, M. F., Stone, N. & Matousek, P. Spatially offset Raman spectroscopy for biomedical applications. *Chemical Society Reviews* **50**, 556–568, https://doi.org/10.1039/D0CS00855A (2021).
19. Cutshaw, G. *et al*. The Emerging Role of Raman Spectroscopy as an Omics Approach for Metabolic Profiling and Biomarker Detection toward Precision Medicine. *Chemical Reviews* acs.chemrev.2c00897, https://doi.org/10.1021/acs.chemrev.2c00897 (2023).
20. Gonze, X. First-principles responses of solids to atomic displacements and homogeneous electric fields: Implementation of a conjugate-gradient algorithm. *Physical Review B* **55**, 10337–10354, https://doi.org/10.1103/PhysRevB.55.10337 (1997).
21. Gonze, X. & Lee, C. Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Physical Review B* **55**, 10355–10368, https://doi.org/10.1103/PhysRevB.55.10355 (1997).
22. Baroni, S., De Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Reviews of Modern Physics* **73**, 515–562, https://doi.org/10.1103/RevModPhys.73.515 (2001).
23. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scripta Materialia* **108**, 1–5, https://doi.org/10.1016/j.scriptamat.2015.07.021 (2015).
24. Togo, A. First-principles Phonon Calculations with Phonopy and Phono3py. *Journal of the Physical Society of Japan* **92**, 012001, https://doi.org/10.7566/JPSJ.92.012001 (2023).
25. Liang, Q., Dwaraknath, S. & Persson, K. A. High-throughput computation and evaluation of raman spectra. *Scientific Data* **6**, 135, https://doi.org/10.1038/s41597-019-0138-y (2019).
26. Taghizadeh, A., Leffers, U., Pedersen, T. G. & Thygesen, K. S. A library of ab initio Raman spectra for automated identification of 2D materials. *Nature Communications* **11**, 3011, https://doi.org/10.1038/s41467-020-16529-6 (2020).
27. Popov, M. N. *et al*. Raman spectra of fine-grained materials from first principles. *npj Computational Materials* **6**, 121, https://doi.org/10.1038/s41524-020-00395-3 (2020).
28. Bagheri, M. & Komsa, H.-P. High-throughput computation of Raman spectra from first principles. *Scientific Data* **10**, 80, https://doi.org/10.1038/s41597-023-01988-5 (2023).
29. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **140**, A1133–A1138, https://doi.org/10.1103/PhysRev.140.A1133 (1965).
30. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **77**, 3865–3868, https://doi.org/10.1103/PhysRevLett.77.3865 (1996).
31. Anisimov, V. I., Zaanen, J. & Andersen, O. K. Band theory and Mott insulators: Hubbard *U* instead of Stoner *I*. *Physical Review B* **44**, 943–954, https://doi.org/10.1103/PhysRevB.44.943 (1991).
32. Dudarev, S. L., Botton, G. A., Savrasov, S. Y., Humphreys, C. J. & Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study. *Physical Review B* **57**, 1505–1509, https://doi.org/10.1103/PhysRevB.57.1505 (1998).
33. Perdew, J. P. *et al*. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Physical Review Letters* **100**, 136406, https://doi.org/10.1103/PhysRevLett.100.136406 (2008).
34. Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* **110**, 6158–6170, https://doi.org/10.1063/1.478522 (1999).
35. Vilela Oliveira, D., Laun, J., Peintinger, M. F. & Bredow, T. BSSE-correction scheme for consistent gaussian basis sets of double- and triple-zeta valence with polarization quality for solid-state calculations. *Journal of Computational Chemistry* **40**, 2364–2376, https://doi.org/10.1002/jcc.26013 (2019).
36. Seidler, L. M., Laun, J. & Bredow, T. BSSE-corrected consistent Gaussian basis sets of triple-zeta valence quality of the lanthanides La-Lu for solid-state calculations. *Journal of Computational Chemistry* **44**, 1418–1425, https://doi.org/10.1002/jcc.27097 (2023).
37. Dovesi, R. *et al*. Quantum-mechanical condensed matter simulations with CRYSTAL. *WIREs Computational Molecular Science* **8**, https://doi.org/10.1002/wcms.1360 (2018).
38. Erba, A. *et al*. CRYSTAL23: A Program for Computational Solid State Physics and Chemistry. *Journal of Chemical Theory and Computation* acs.jctc.2c00958, https://doi.org/10.1021/acs.jctc.2c00958 (2022).
39. Dovesi, R. *et al*. CRYSTAL23 User's Manual (2023).
40. Pascale, F. *et al*. The calculation of the vibrational frequencies of crystalline compounds and its implementation in the CRYSTAL code. *Journal of Computational Chemistry* **25**, 888–897, https://doi.org/10.1002/jcc.20019 (2004).
41. Zicovich-Wilson, C. M. *et al*. Calculation of the vibration frequencies of α-quartz: The effect of Hamiltonian and basis set. *Journal of Computational Chemistry* **25**, 1873–1881, https://doi.org/10.1002/jcc.20120 (2004).
42. Ferrero, M., Rérat, M., Orlando, R. & Dovesi, R. The calculation of static polarizabilities of 1-3D periodic compounds. the implementation in the crystal code. *Journal of Computational Chemistry* **29**, 1450–1459, https://doi.org/10.1002/jcc.20905 (2008).
43. Ferrero, M., Rérat, M., Orlando, R. & Dovesi, R. Coupled perturbed Hartree-Fock for periodic systems: The role of symmetry and related computational aspects. *The Journal of Chemical Physics* **128**, 014110, https://doi.org/10.1063/1.2817596 (2008).
44. Ferrero, M., Rérat, M., Kirtman, B. & Dovesi, R. Calculation of first and second static hyperpolarizabilities of one- to three-dimensional periodic compounds. Implementation in the CRYSTAL code. *The Journal of Chemical Physics* **129**, 244110, https://doi.org/10.1063/1.3043366 (2008).

45. Maschio, L., Kirtman, B., Rérat, M., Orlando, R. & Dovesi, R. *Ab Initio* analytical Raman intensities for periodic systems through a coupled perturbed Hartree-Fock/Kohn-Sham method in an atomic orbital basis. I. Theory. *The Journal of Chemical Physics* **139**, 164101, https://doi.org/10.1063/1.4824442 (2013).
46. Maschio, L., Kirtman, B., Rérat, M., Orlando, R. & Dovesi, R. *Ab Initio* analytical Raman intensities for periodic systems through a coupled perturbed Hartree-Fock/Kohn-Sham method in an atomic orbital basis. II. Validation and comparison with experiments. *The Journal of Chemical Physics* **139**, 164102, https://doi.org/10.1063/1.4824443 (2013).
47. Zagorac, D., Müller, H., Ruehl, S., Zagorac, J. & Rehme, S. Recent developments in the Inorganic Crystal Structure Database: Theoretical crystal structure data and related features. *Journal of Applied Crystallography* **52**, 918–925, https://doi.org/10.1107/S160057671900997X (2019).
48. Togo, A. & Tanaka, I. Spglib: A software library for crystal symmetry search. *arXiv* https://doi.org/10.48550/ARXIV.1808.01590 (2018).
49. Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319, https://doi.org/10.1016/j.commatsci.2012.10.028 (2013).
50. Setyawan, W. & Curtarolo, S. High-throughput electronic band structure calculations: Challenges and tools. *Computational Materials Science* **49**, 299–312, https://doi.org/10.1016/j.commatsci.2010.05.010 (2010).
51. Li, Y., Lee, D. K. J. & Cai, P. Caneparesearch/project_Raman: A Database of Computed Raman Spectra of Inorganic Compounds with Accurate Hybrid Functionals. *Zenodo* https://doi.org/10.5281/zenodo.8349322 (2023).
52. Lafuente, B., Downs, R. T., Yang, H. & Stone, N. 1. The power of databases: The RRUFF project. In Armbruster, T. & Danisi, R. M. (eds.) *Highlights in Mineralogical Crystallography*, 1–30, https://doi.org/10.1515/9783110417104-003 (DE GRUYTER, 2015).
53. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *The Journal of Chemical Physics* **118**, 8207–8215, https://doi.org/10.1063/1.1564060 (2003).
54. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Erratum: "Hybrid functionals based on a screened Coulomb potential" [J. Chem. Phys. 118, 8207 (2003)]. *The Journal of Chemical Physics* **124**, 219906, https://doi.org/10.1063/1.2204597 (2006).
55. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **132**, 154104, https://doi.org/10.1063/1.3382344 (2010).
56. Sato, R. K. & McMillan, P. F. An infrared and Raman study of the isotopic species of.alpha.-quartz. *The Journal of Physical Chemistry* **91**, 3494–3498, https://doi.org/10.1021/j100297a008 (1987).
57. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002, https://doi.org/10.1063/1.4812323 (2013).
58. Canepa, P., Hanson, R. M., Ugliengo, P. & Alfredsson, M. *J-ICE*: A new *Jmol* interface for handling and visualizing crystallographic and electronic properties. *Journal of Applied Crystallography* **44**, 225–229, https://doi.org/10.1107/S0021889810049411 (2011).

Acknowledgements

Y.L. and P.C. are grateful to the ANR-NRF NRF2019-NRF-ANR073 Na-MASTER. P.C. acknowledges funding from the National Research Foundation under his NRF Fellowship NRFF12-2020-0012. The computational work was performed on resources of the National Supercomputing Centre, Singapore. Y.L. acknowledges the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program.

Author contributions

P.C. conceived the main idea and defined the project, Y.L. and Z.Z. collected compounds for calculations, Y.L. wrote the input code and conducted all the calculations, D.L. and P.F.C. wrote the `Python` parser, `Mongo`, and streamlit code, Y.L., P.G. and P.C. analyzed the results, Y.L. wrote the manuscript. All authors reviewed and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-024-02924-x.

Correspondence and requests for materials should be addressed to P.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024