Accelerating Materials Discovery of Novel Europium(II)-Activated Phosphors through Machine Learning Classification of Europium Valences

Yukinori Koyama, [*, a] *Yukako Kohriki,* [b] *Masamichi Harada,* [b] *Naoto Hirosaki,* [b] *Takashi Takeda* [b]

[a] Center for Basic Research on Materials, National Institute for Materials Science, Tsukuba, Ibaraki 305-0047, Japan

[b] Research Center for Electronic and Optical Materials, National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan.

S1. Feature importance of the machine learning model

This section discusses the importance of the features used in the machine learning model. Figure S1 illustrates the absolute values of the logistic regression coefficients in descending order for the 90 features selected in the preprocessing steps. The top three regression coefficients were significant, and the following six features had coefficients of similar size. The coefficient of the tenth feature was less than one-third of that of the top feature. After that, the coefficients gradually decreased.

The top nine features and their coefficients are presented in Table S1. The top two features were both related to electronegativity, indicating that the electronegativity of the constituent elements significantly affects the oxidation state of the Eu ions in the host. The signs of the coefficients were inverse between the arithmetic mean of the top feature and the geometric mean of the second one, suggesting a nonlinear relationship between the electronegativity and the oxidation states of Eu ions. This study used arithmetic, geometric, and harmonic means to average the elemental features. Items with smaller values contribute more in this order. Therefore, the negative coefficient for the arithmetic mean and the positive coefficient for the geometric mean indicate that there was a negative correlation between the electronegativity and the Eu^{2+}-paper ratio for electronegative elements, namely anions and a positive correlation for less electronegative elements, namely cations. The negative correlation for anions corresponds to the small Eu^{2+}-paper ratio in the O- and F-containing compounds. On the other hand, we have not identified specific trends regarding the positive correlation for cations.

Similarly, looking at the third and sixth features, which are related to the Mendeleev numbers (MN), there was a positive correlation between the MN and Eu^{2+}-paper ratio for elements of large MN located on the right side of the periodic table, and a negative correlation for elements of small

MN located on the left side of the periodic table. The former positive correlation corresponds to the large Eu^{2+}-paper ratio of compounds containing halogens (Cl, Br, and I), and the latter negative correlation corresponds to the small Eu^{2+}-paper ratio of compounds containing rare earth and Group-5 and 6 metals. Because the number of unoccupied valence p states is inversely correlated with the MN, the trend of the Eu^{2+}-paper ratio regarding the number of unoccupied valence p states is almost the same as that regarding the MN.

Because this study used general-purpose descriptors, it is impossible to directly measure the effect of elemental content. Thus, the above discussion is only an interpretation of the feature importance. Furthermore, atomic number, ionization energy, and group number appear only once in this table, but the averages of different types are included in the group below the tenth feature. It is necessary to interpret the effect of elemental content with consideration of the related features.

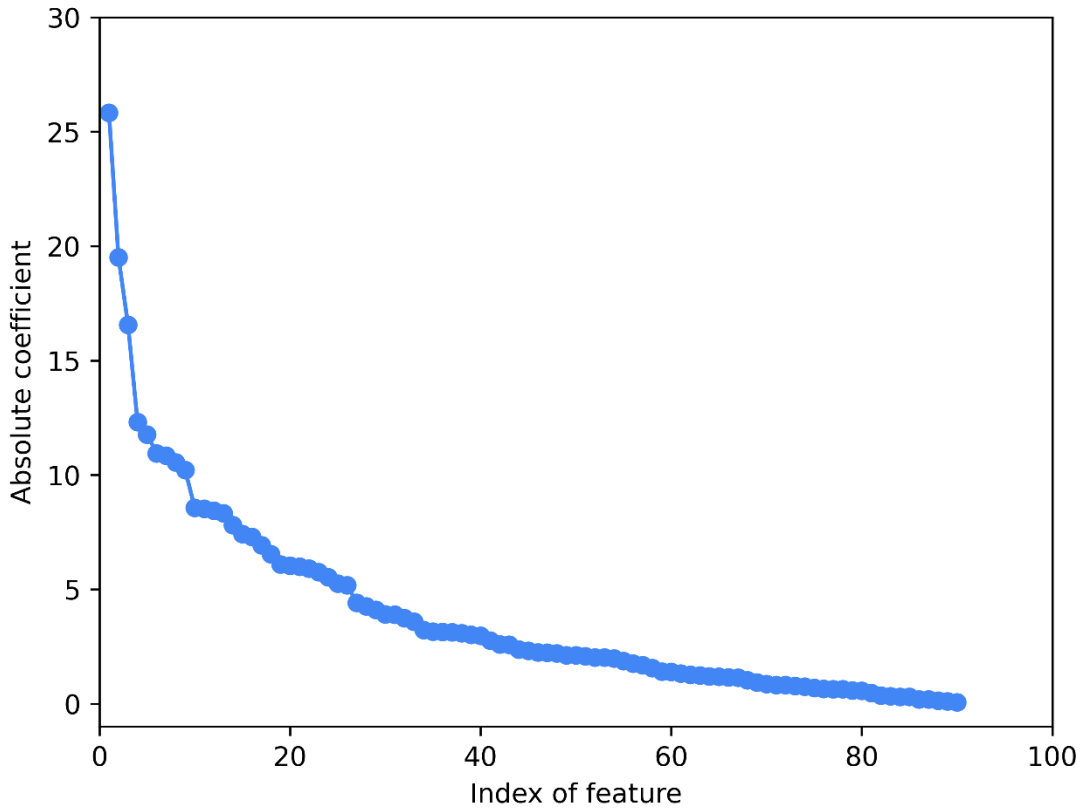

Figure S1. Absolute values of the logistic regression coefficients in descending order.

Table S1. Top nine features and those coefficients in the logistic regression.

Rank	Elemental feature	Statistics	Coefficient
1	Electronegativity	arithmetic mean	–25.829
2	Electronegativity	geometric mean	19.505
3	Mendeleev number	geometric mean	–16.567
4	Atomic number	geometric mean	12.311
5	Ionization energy	geometric mean	–11.775
6	Mendeleev number	arithmetic mean	10.947
7	Number of unoccupied valence p states	geometric mean	10.840
8	Number of unoccupied valence p states	harmonic mean	–10.552
9	Group in the periodic table	geometric mean	10.212

S2. Machine learning classification with gradient boosted trees method

A machine learning classification model with the gradient boosted trees method was developed to predict the oxidation states of Eu ions in the hosts regarding luminescence based on their host compositions. The machine learning pipeline was the same as that using logistic regression, explained in the main text, and the gradient-boosted-trees method was used as the classifier in the RFE and final steps. The parameters shown in Table S2 were determined by the Bayesian optimization method using the scikit-optimize package [Ref. S1] to maximize the F-score for the validation data of the 10-fold cross-validation. Accuracy, precision, recall, and F-score with the optimum parameters evaluated for the training and validation dataset in 10-fold cross-validation are summarized in Table S3.

Table S2. Parameter ranges and optimum values of the machine learning classification model using the gradient-boosted-trees method.

Parameter	**Candidates**	**Optimum**
Number of selected features in RFE	10, 20, …, 100	100
Learning rate	10^{-2} – 0.5 (log-scale)	0.037
Maximum depth of trees	1, 2, 3, 4, 5	3
Number of trees	100, 200, …, 1000	700

Table S3. Classification metrics of the machine learning model using the gradient-boosted-trees method with the optimum parameters evaluated for the training and validation datasets in 10-fold cross-validation.

Metric	Training	Validation
Accuracy	94.7% (0.1%)	86.9% (1.8%)
Precision	92.9% (0.2%)	84.0% (3.7%)
Recall	93.6% (0.4%)	81.7% (2.7%)
F-score	93.2% (0.2%)	82.8% (2.6%)

The scores were averaged among the folds. Standard deviations over folds are shown in parentheses.

References

[S1] Scikit-optimize. https://scikit-optimize.github.io/.

S3. Synthesis and powder X-ray diffraction analysis

Synthesis conditions and powder X-ray diffraction (XRD) analysis results of the samples for which photoluminescence was observed are summarized in this section. The XRD patterns of the powder samples are illustrated with simulated patterns of target and impurity phases. Some samples showed hollow peaks at 10-20 degrees and high backgrounds owing to using an airtight sample holder to protect air-sensitive samples. The XRD patterns were analyzed to identify phases and fractions in the products using the whole powder pattern fitting method (Rigaku, PDXL2) with card peak patterns provided by the Joint Committee on Powder Diffraction Standards (JCPDS) and simulated peak patterns from crystal structure data obtained from the Inorganic Crystal Structure Database (ICSD).

S3.1. $Rb_3Ca_2Cl_7$

Table S4. Starting materials and synthesis conditions of $Rb_3Ca_2Cl_7$.

Starting materials	RbCl, $CaCl_2$, $EuCl_3$
Synthesis conditions	600 °C, 5 h, H_2(5%)/N_2

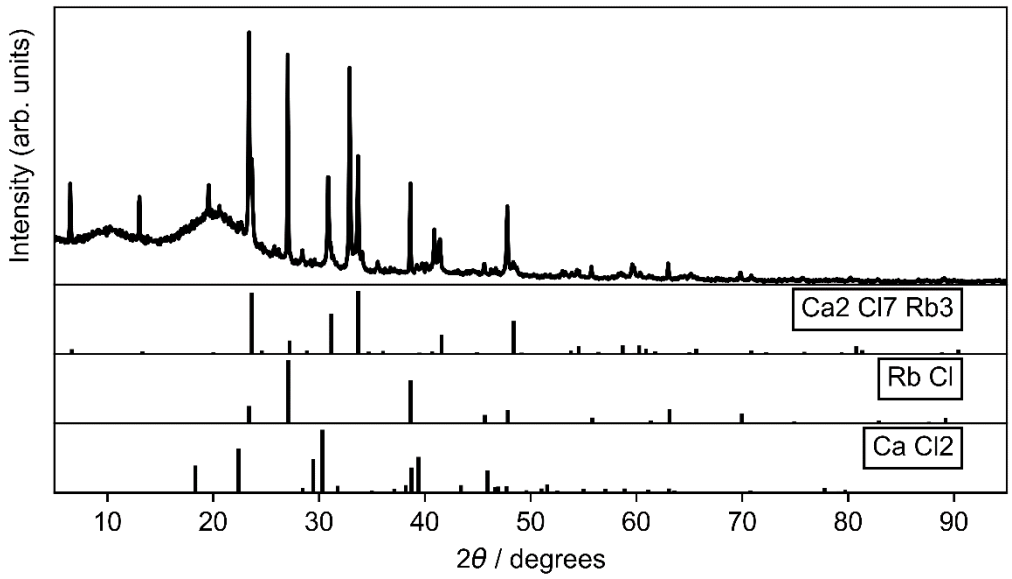

Figure S2. XRD pattern of the $Rb_3Ca_2Cl_7$ sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S5. XRD analysis results of the $Rb_3Ca_2Cl_7$ sample.

Phase	ICDD card number	Space group	Fraction (wt%)
Ca2 Cl7 Rb3	04-022-6840	I 4/m m m (139)	72.1(15)
Rb Cl	04-005-4309	F m -3 m (225)	14.9(5)
Ca Cl2	01-071-5407	P b c n (60)	13.0(17)

S3.2. $RbSrCl_3$

Table S6. Starting materials and synthesis conditions of $RbSrCl_3$.

Starting materials	RbCl, $SrCl_2$, $EuCl_3$
Synthesis conditions	600 °C, 5 h, H_2(5%)/N_2

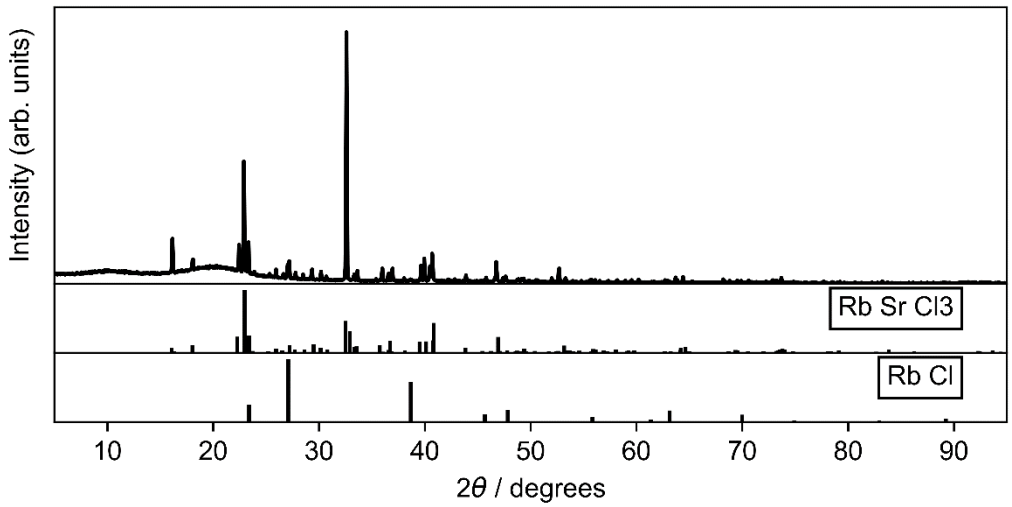

Figure S3. XRD pattern of the $RbSrCl_3$ sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S7. XRD analysis results of the $RbSrCl_3$ sample.

Phase	**ICDD card number**	**Space group**	**Fraction (wt%)**
Rb Sr Cl3	ICSD 138559	P n m a (62)	98.3(8)
Rb Cl	04-005-4309	F m -3 m (225)	1.7(8)

S3.3. Cs_2CaBr_4

Table S8. Starting materials and synthesis conditions of Cs_2CaBr_4.

Starting materials	CsBr, $CaBr_2$, $EuBr_2$
Synthesis conditions	550 °C, 5 h, H_2(5%)/N_2

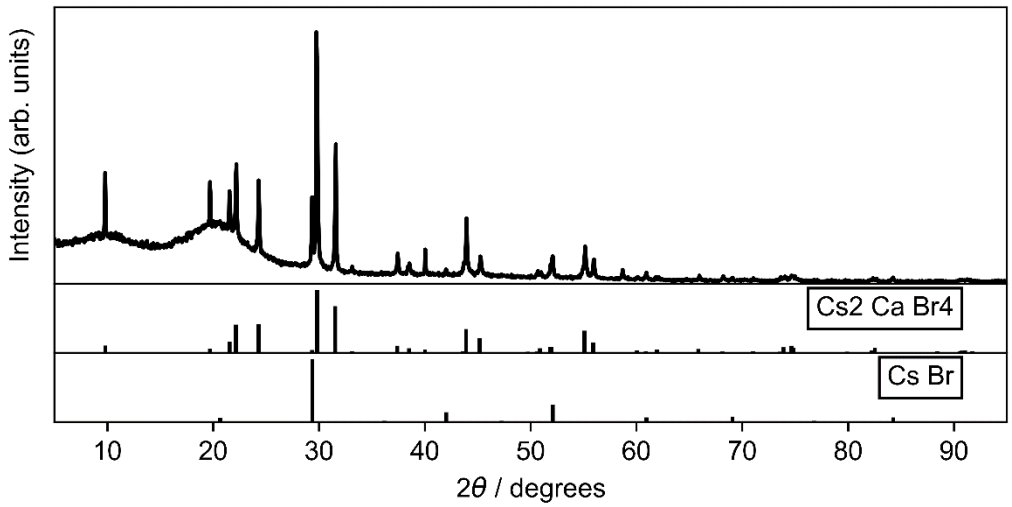

Figure S4. XRD pattern of the Cs_2CaBr_4 sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S9. XRD analysis results of the Cs_2CaBr_4 sample.

Phase	**ICDD card number**	**Space group**	**Fraction (wt%)**
Cs2 Ca Br4	04-010-0063	I 4/m m m (139)	90.8(7)
Cs Br	01-082-9638	P m -3 m (221)	9.2(7)

S3.4. Rb_4CaBr_6

Table S10. Starting materials and synthesis conditions of Rb_4CaBr_6.

Starting materials	RbBr, $CaBr_2$, $EuBr_2$
Synthesis conditions	550 °C, 5 h, H_2(5%)/N_2

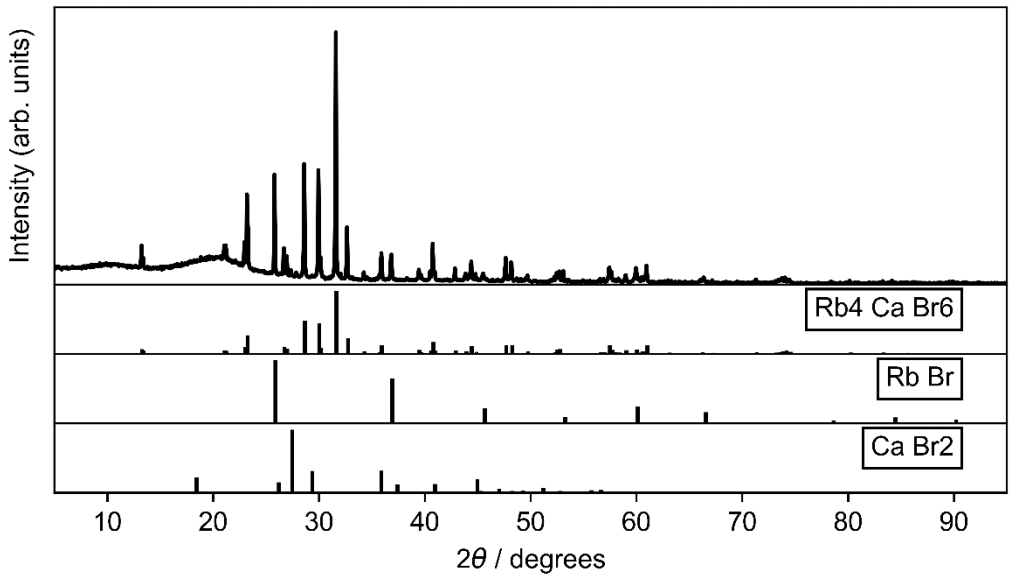

Figure S5. XRD pattern of the Rb_4CaBr_6 sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S11. XRD analysis results of the Rb_4CaBr_6 sample.

Phase	ICDD card number	Space group	Fraction (wt%)
Rb4 Ca Br6	04-010-0064	R -3 c (167)	85.4(3)
Rb Br	01-073-0384	F m -3 m (225)	13.1(3)
Ca Br2	01-071-5406	P 4_2/m n m (136)	1.45(2)

S3.5. BaB_2S_4

Table S12. Starting materials and synthesis conditions of BaB_2S_4.

Starting materials	BaS, B, S, EuS
Synthesis conditions	450 °C, 2 h + 700 °C, 6 h, vacuum-sealed quartz ampoule

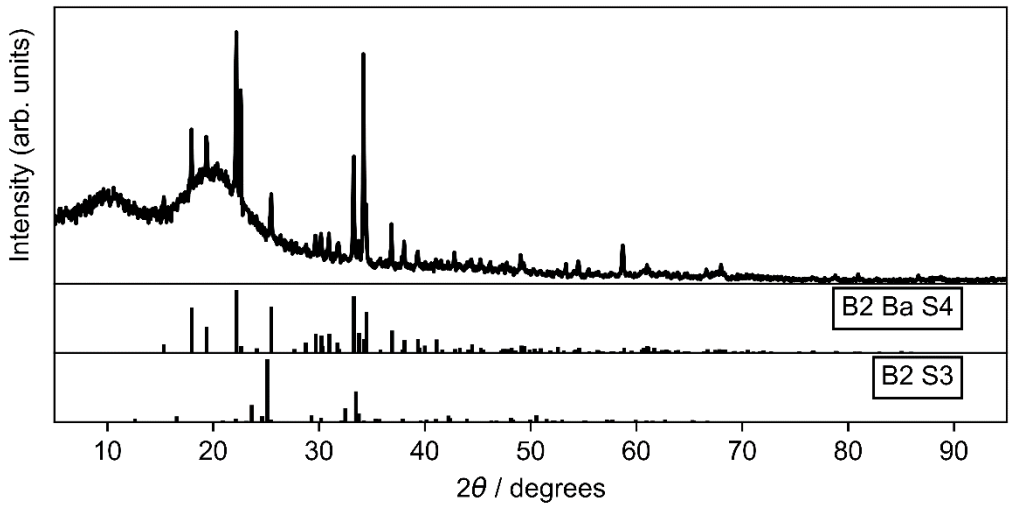

Figure S6. XRD pattern of the BaB_2S_4 sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S13. XRD analysis results of the BaB_2S_4 sample.

Phase	ICDD card number	Space group	Fraction (wt%)
B2 Ba S4	04-010-8537	C c (9)	87(6)
B2 S3	03-065-1287	P 2_1/c (14)	12(5)

S3.6. $BaLa_2ZnS_5$

Table S14. Starting materials and synthesis conditions of $BaLa_2ZnS_5$.

Starting materials	BaS, La_2S_3, ZnS, EuS
Synthesis conditions	900 °C, 10 h, vacuum-sealed quartz ampoule.

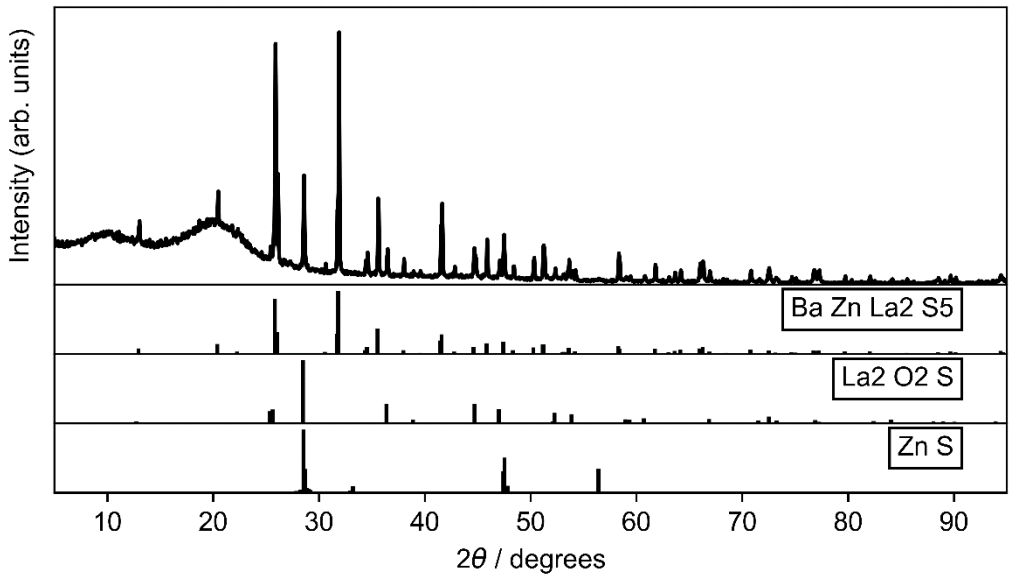

Figure S7. XRD pattern of the $BaLa_2ZnS_5$ sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S15. XRD analysis results of the $BaLa_2ZnS_5$ sample.

Phase	**ICDD card number**	**Space group**	**Fraction (wt%)**
Ba Zn La2 S5	04-009-1670	I 4/m c m (140)	76.8(8)
La2 O2 S	01-078-7509	P -3 m 1 (164)	18.4(6)
Zn S	01-072-9271	P 3 m 1 (156)	4.8(6)

S3.7. $NaLi_3BaB_6O_{12}$

Table S16. Starting materials and synthesis conditions of $NaLi_3BaB_6O_{12}$.

Starting materials	$BaCO_3$, Na_2CO_3, Li_2CO_3, H_3BO_3, Eu_2O_3
Synthesis conditions	300 °C, 4 h, air + 600 °C, 8 h, air + 600 °C, 5h, H_2(4%)/Ar

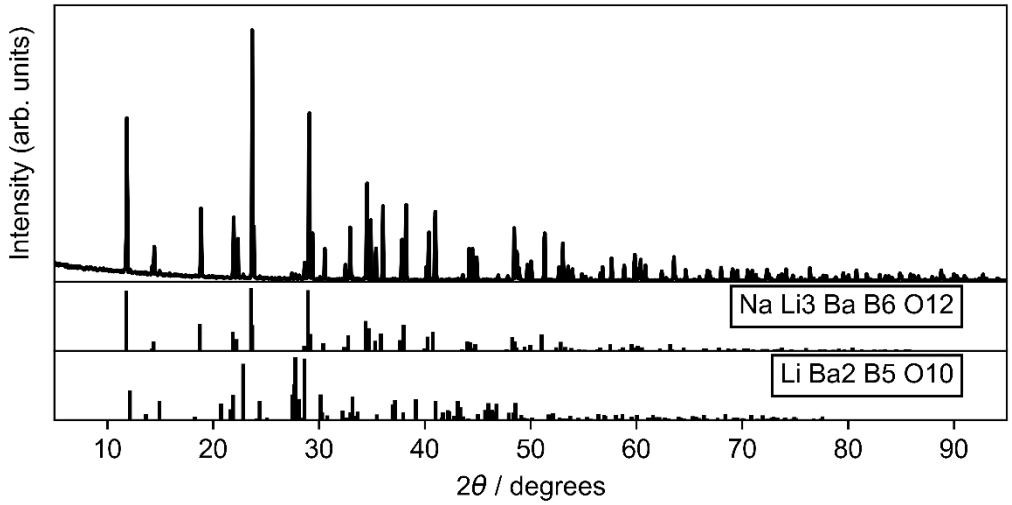

Figure S8. XRD pattern of the $NaLi_3BaB_6O_{12}$ sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S17. XRD analysis results of the $NaLi_3BaB_6O_{12}$ sample.

Phase	**ICDD card number**	**Space group**	**Fraction (wt%)**
Na Li3 Ba B6 O12	04-019-4572	R -3 (148)	97.1(2)
Li Ba2 B5 O10	04-009-4377	P 2_1/m (11)	2.9(2)

S3.8. $KNaCaMg_5Si_8O_{22}F_2$

Table S18. Starting materials and synthesis conditions of $KNaCaMg_5Si_8O_{22}F_2$.

Starting materials	Na_2CO_3, K_2CO_3, SiO_2, MgO, CaF_2, Eu_2O_3
Synthesis conditions	1000 °C, 5 h, H_2(5%)/N_2

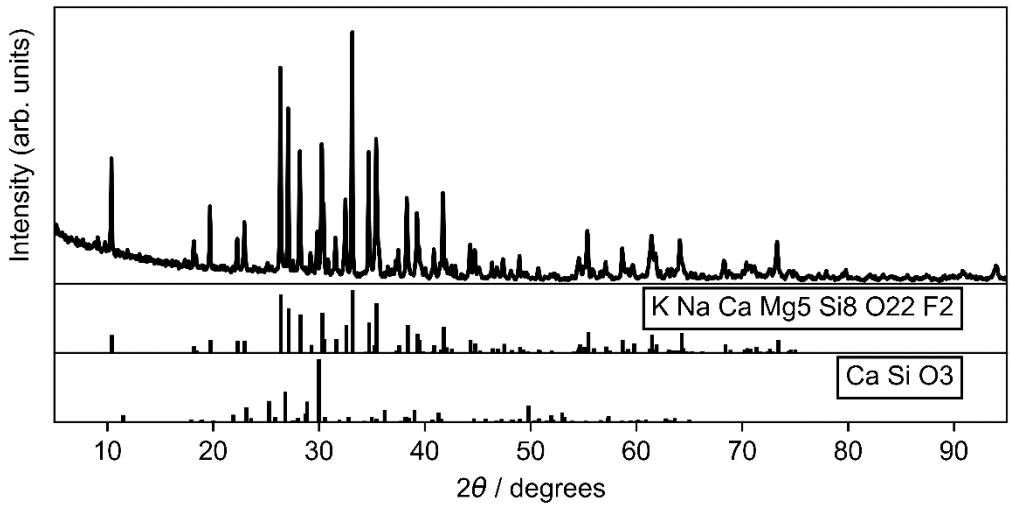

Figure S9. XRD pattern of the $KNaCaMg_5Si_8O_{22}F_2$ sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S19. XRD analysis results of the $KNaCaMg_5Si_8O_{22}F_2$ sample.

Phase	ICDD card number	Space group	Fraction (wt%)
K Na Ca Mg5 Si8 O22 F2	04-013-2169	C 2/m (12)	95.1(6)
Ca Si O3	04-016-5569	P -1 (2)	4.3(5)

S3.9. Sr_2OI_2

Table S20. Starting materials and synthesis conditions of Sr_2OI_2.

Starting materials	SrO, SrI_2, Eu_2O_3
Synthesis conditions	675 °C, 5 h, H_2(5%)/N_2

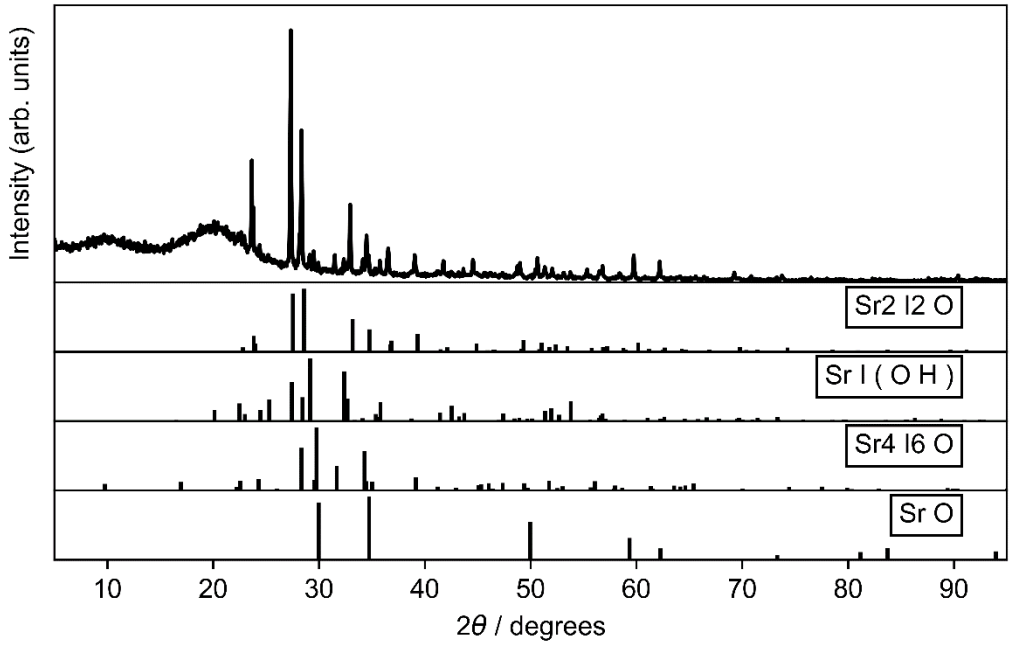

Figure S10. XRD pattern of the Sr_2OI_2 sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S21. XRD analysis results of the Sr_2OI_2 sample.

Phase	ICDD card number	Space group	Fraction (wt%)
Sr2 I2 O	04-014-0889	I b a m (72)	73(2)
Sr I (O H)	04-011-6332	P n m a (62)	13.5(10)
Sr4 I6 O	04-022-6011	P 6_3 m c (186)	7.9(11)
Sr O	01-075-6979	F m -3 m (225)	6(2)

S3.10. Sr_4OI_6

Table S22. Starting materials and synthesis conditions of Sr_4OI_6.

Starting materials	SrO, SrI_2, Eu_2O_3
Synthesis conditions	650 °C, 8 h, H_2(5%)/N_2

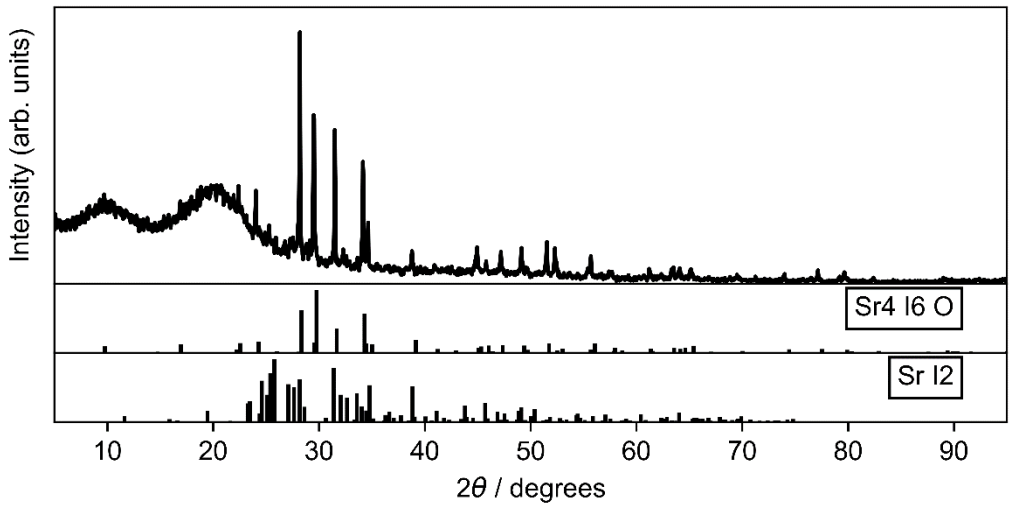

Figure S11. XRD pattern of the Sr_4OI_6 sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S23. XRD analysis results of the Sr_4OI_6 sample.

Phase	ICDD card number	Space group	Fraction (wt%)
Sr4 I6 O	04-022-6011	P 6_3 m c (186)	74.5(7)
Sr I2	04-007-1983	P b c a (61)	25.5(7)

S3.11. Ba_4OI_6

Table S24. Starting materials and synthesis conditions of Ba_4OI_6.

Starting materials	BaO, BaI_2, Eu_2O_3
Synthesis conditions	650 °C, 5 h, H_2(5%)/N_2 + 650 °C, 2 h, H_2(5%)/N_2

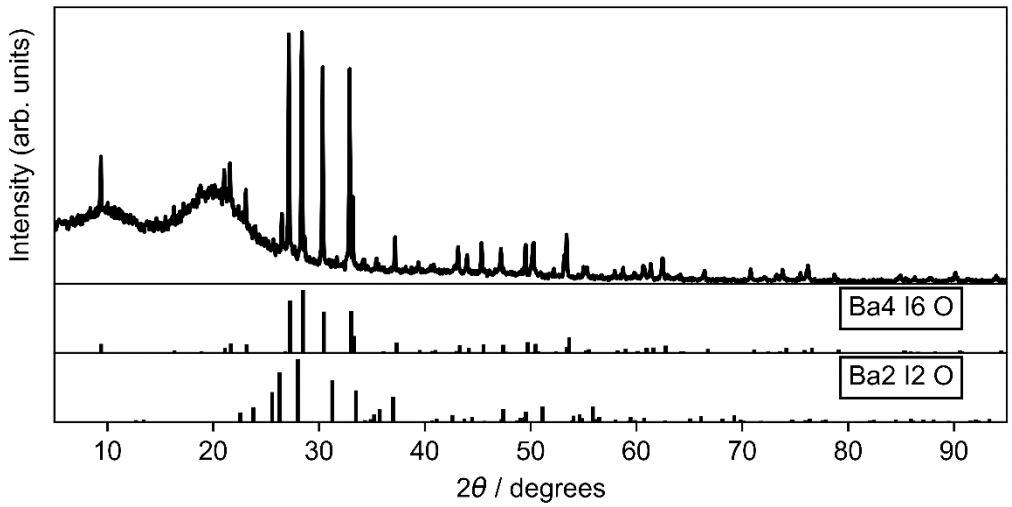

Figure S12. XRD pattern of the Ba_4OI_6 sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S25. XRD analysis results of the Ba_4OI_6 sample.

Phase	ICDD card number	Space group	Fraction (wt%)
Ba4 I6 O	04-011-1893	P 6_3 m c (186)	95.8(8)
Ba2 I2 O	04-016-5407	I b a m (72)	4.2(8)

S3.12. Ba_2PO_4I

Table S26. Starting materials and synthesis conditions of Ba_2PO_4I.

Starting materials	$Ba_3(PO4)_2$, BaI_2, Eu_2O_3
Synthesis conditions	800 °C, 5 h, H_2(5%)/N_2

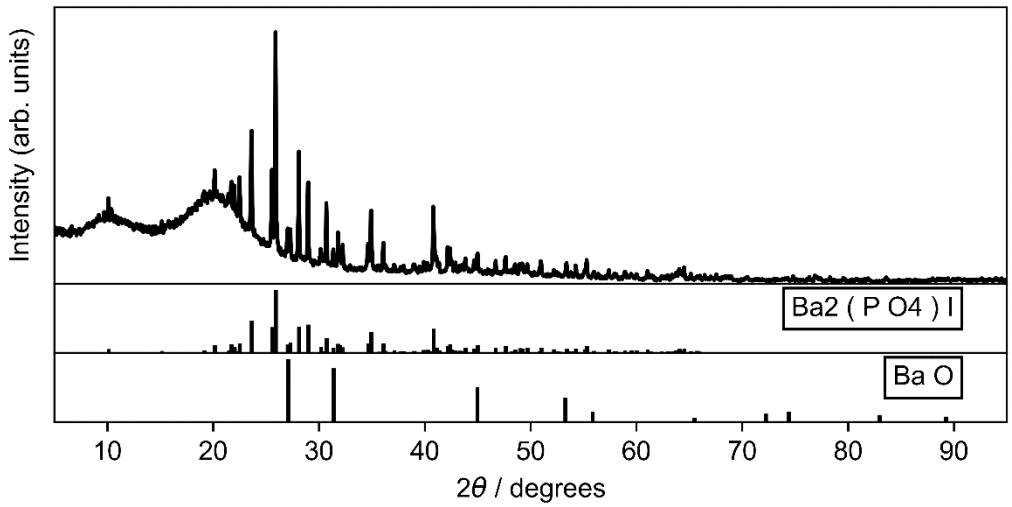

Figure S13. XRD pattern of the Ba_2PO_4I sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S27. XRD analysis results of the Ba_2PO_4I sample.

Phase	ICDD card number	Space group	Fraction (wt%)
Ba2 (P O4) I	04-021-7746	P 2_1/c (14)	94(2)
Ba O	01-080-3980	F m -3 m (225)	6(2)

S3.13. Sr_2PO_4I

Table S28. Starting materials and synthesis conditions of Sr_2PO_4I.

Starting materials	$Sr_3(PO4)_2$, SrI_2, Eu_2O_3
Synthesis conditions	800 °C, 5 h, H_2(5%)/N_2

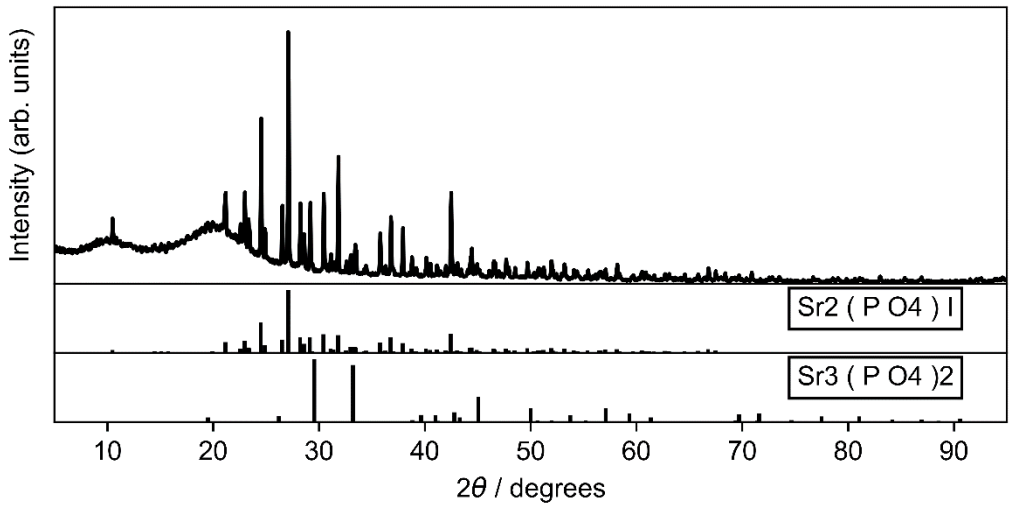

Figure S14. XRD pattern of the Sr_2PO_4I sample (top) and simulated patterns of the target and impurity phases (bottom).

Table S29. XRD analysis results of the Sr_2PO_4I sample.

Phase	ICDD card number	Space group	Fraction (wt%)
Sr2 (P O4) I	04-021-7748	P 2_1/c (14)	97(2)
Sr3 (P O4)2	01-073-4870	R -3 m (166)	3(2)