Electronic Supplementary Information for "Element similarity in high-dimensional materials" representations

Anthony Onwuli,[1] Ashish V. Hegde,[1] Kevin Nguyen,[1] Keith T. Butler,[2, a] and Aron Walsh[1, 3, b]

[1] *Department of Materials, Imperial College London, London SW7 2AZ, UK*

[2] *School of Engineering and Materials Science, Queen Mary University of London, London E1 4NS, UK*

[3] *Department of Physics, Ewha Womans University, Seoul 03760, Korea*

(Dated: 24 August 2023)

---

[a] Electronic mail: k.butler@qmul.ac.uk

[b] Electronic mail: a.walsh@imperial.ac.uk

Supplementary Note 1: Similarity measures Within the main body of the text, only the Euclidean and Manhattan distances are shown for four of the similarity measures. Here, we show the distance measures currently available in ELEMENTEMBEDDINGS for seven of the representation schemes. The distance measures mentioned in the main body of the text, Euclidean, Manhattan and Chebyshev are depicted in Figures S1, S2 and S3, respectively.

The cosine distance is also included as a distance measure within the package. It is the complement of the cosine similarity:

$$d_{cosine} = 1 - cos(\theta) = 1 - \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \tag{S1}$$

The heatmaps associated with this distance measure are shown in Figure S4.

Another metric included in the package is the Wasserstein distance which can be defined as the minimum amount of work required to transform distribution u into v. The first Wasserstein distance for distributions u and v is:

$$W_1(u,v) = \inf_{\pi \in \Gamma(u,v)} \int_{\mathbb{R} \times \mathbb{R}} |x-y| \mathrm{d}\pi(x,y) \tag{S2}$$

where $\Gamma(u,v)$ is the set of distributions on $\mathbb{R} \times \mathbb{R}$ whose marginals are u and v on the first and second factors, respectively. The heatmap associated with this distance measure is shown in Figure S5.

The Pearson correlation and cosine similarity measures which appear in the main text are also extended to seven representation schemes. Their maps are shown in Figures S6 and S8. An additional correlation coefficient included in the package is the Spearman's rank correlation coefficient. This correlation measure assesses the monotonic relationship between two variables.

$$r_s = \rho_{R(\mathbf{A}), R(\mathbf{B})} = \frac{\mathrm{cov}(R(\mathbf{A}), R(\mathbf{B}))}{\sigma_{R(\mathbf{A})} \sigma_{R(\mathbf{B})}} \tag{S3}$$

Figure S7 shows the heatmaps associated with this measure.

Supplementary Note 2: Similarity distributions Figures S9 and S10 feature the distribution of the Pearson correlation coefficient and cosine similarity values between the element vectors, respectively.

Supplementary Note 3: Two-dimensional projections The PCA and t-SNE plots which are depicted in the main body of the text are shown in Figures S11 and S12 for seven of the representations. Addtionally, a plot for UMAP is also provided in Figure S13. For the skipatom representation in Figure S11, the two-dimensional projection is heavily compressed due to the anomalous Kr vector.

Supplementary Note 4: Crystal structure prediction SMACT was used to perform crystal structure prediction via structure substitutions of known materials from the Materials Project. Tables S1 and S2 show the target formulas and for each representation, the template material which the target formula was substituted in order to assign the structure to the target.
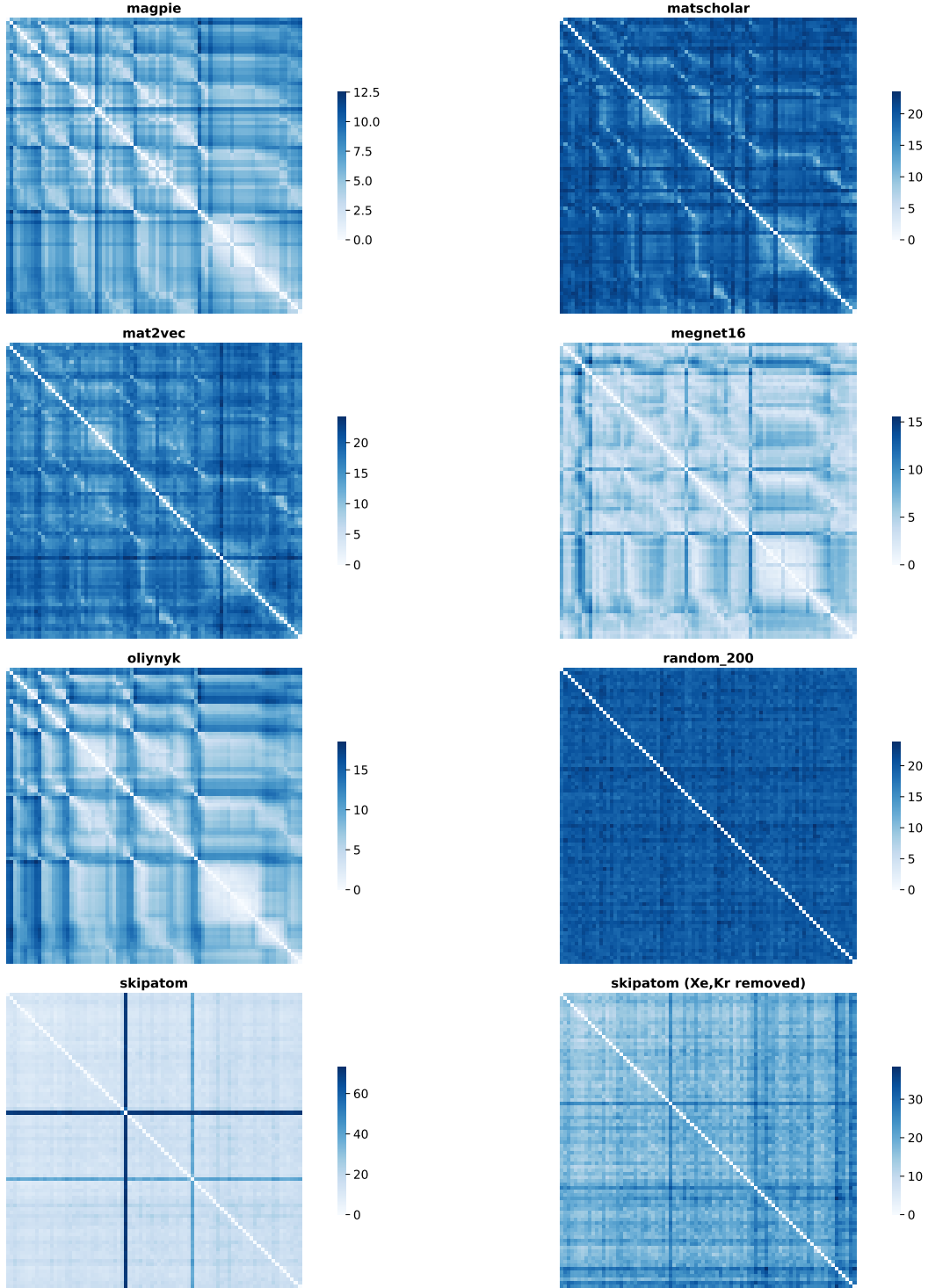
FIG. S1. Map of the pairwise Euclidean distance between element vectors for seven representation schemes. The elements are ordered in increasing atomic number along the axes. The pronounced vertical and horizontal stripes in the skipatom representation, and features in others, correspond to the noble gases Kr and Xe.
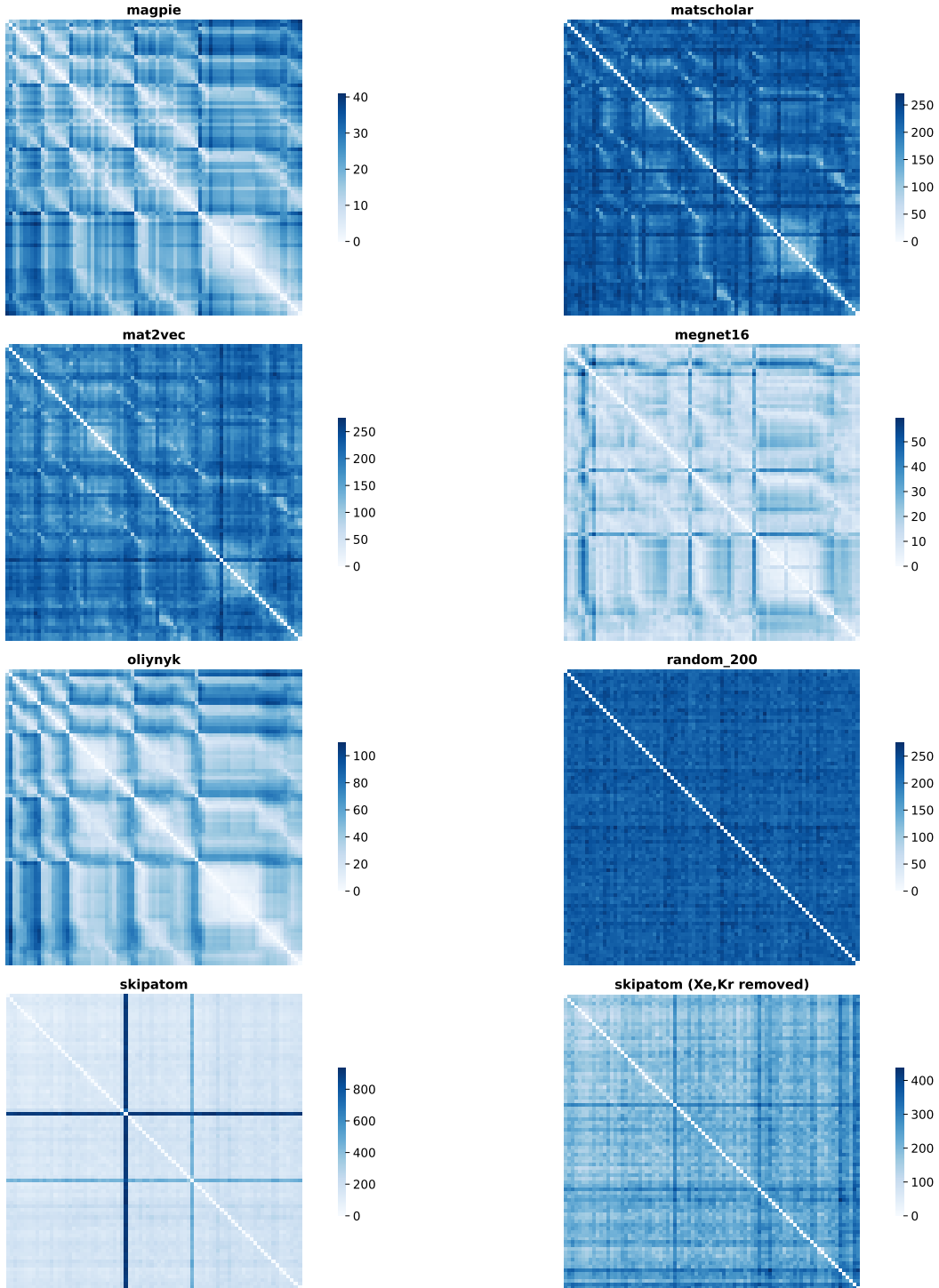
FIG. S2. Map of the pairwise Manhattan distance between element vectors for seven representation schemes. The elements are ordered in increasing atomic number along the axes.
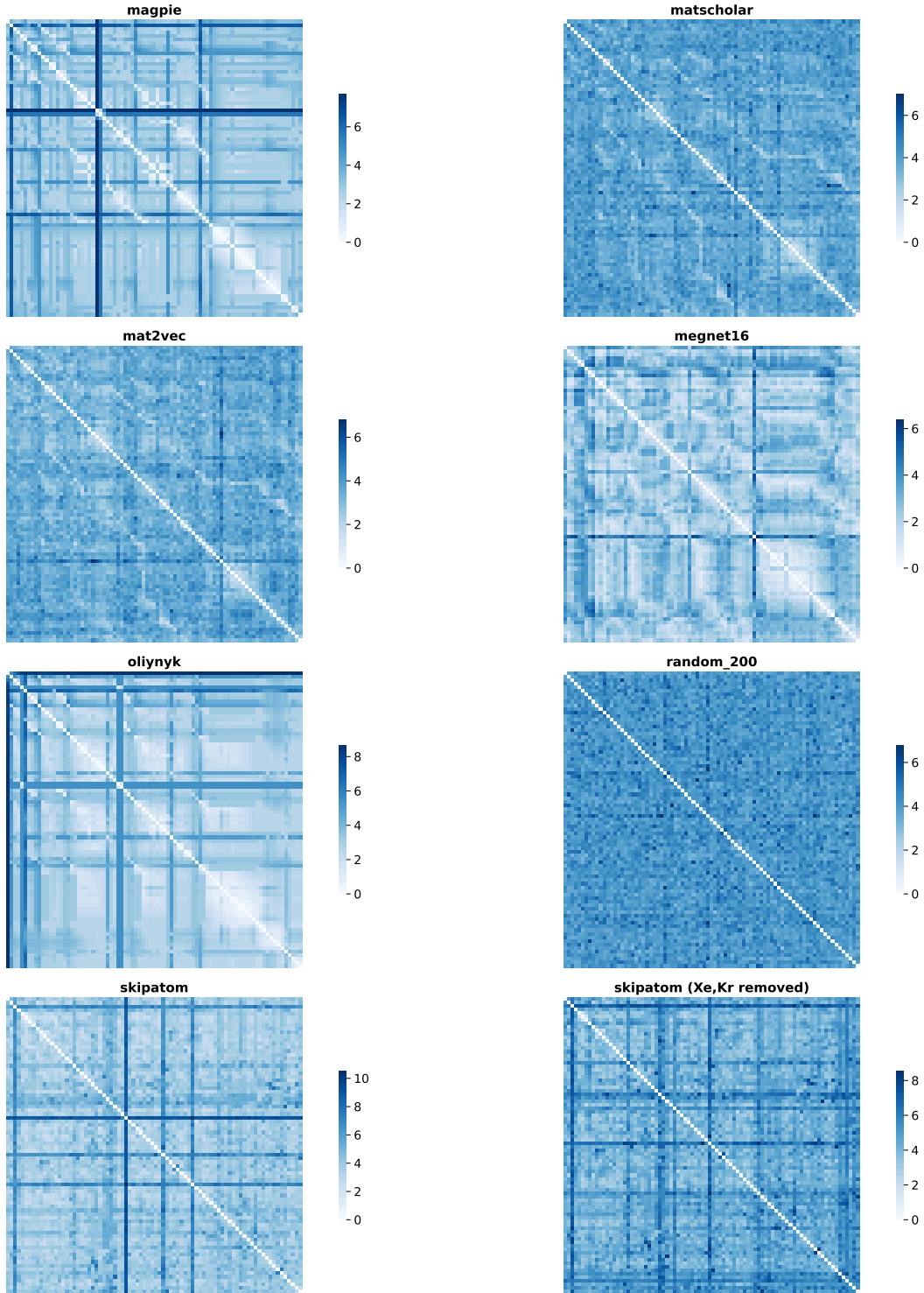
FIG. S3. Map of the pairwise Chebyshev distance between element vectors for seven representation schemes. The elements are ordered in increasing atomic number along the axes.
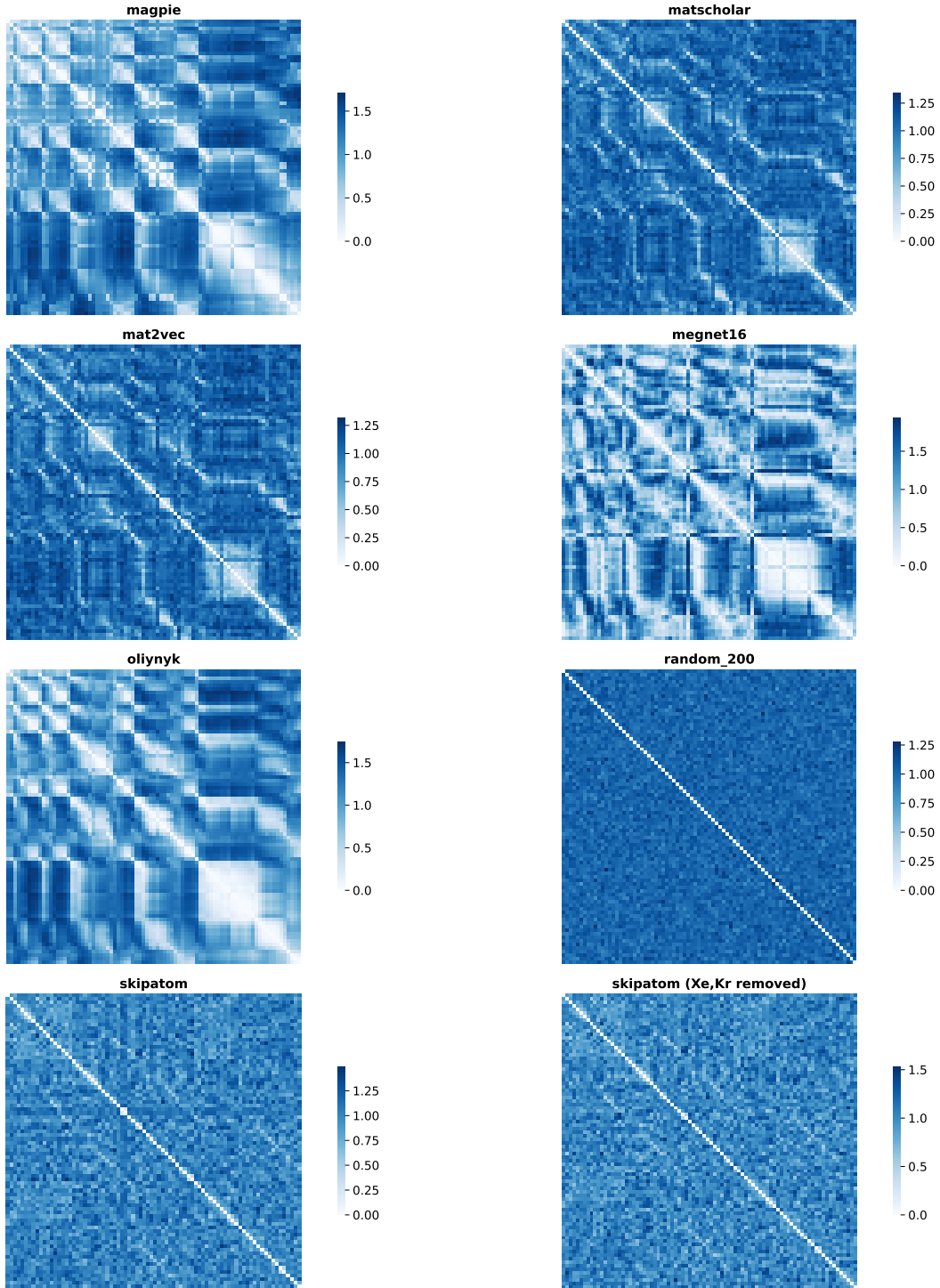
FIG. S4. Map of the pairwise cosine distance between element vectors for seven representation schemes. The elements are ordered in increasing atomic number along the axes.

FIG. S5. Map of the pairwise Wasserstein distance between element vectors for seven representation schemes. The elements are ordered in increasing atomic number along the axes.
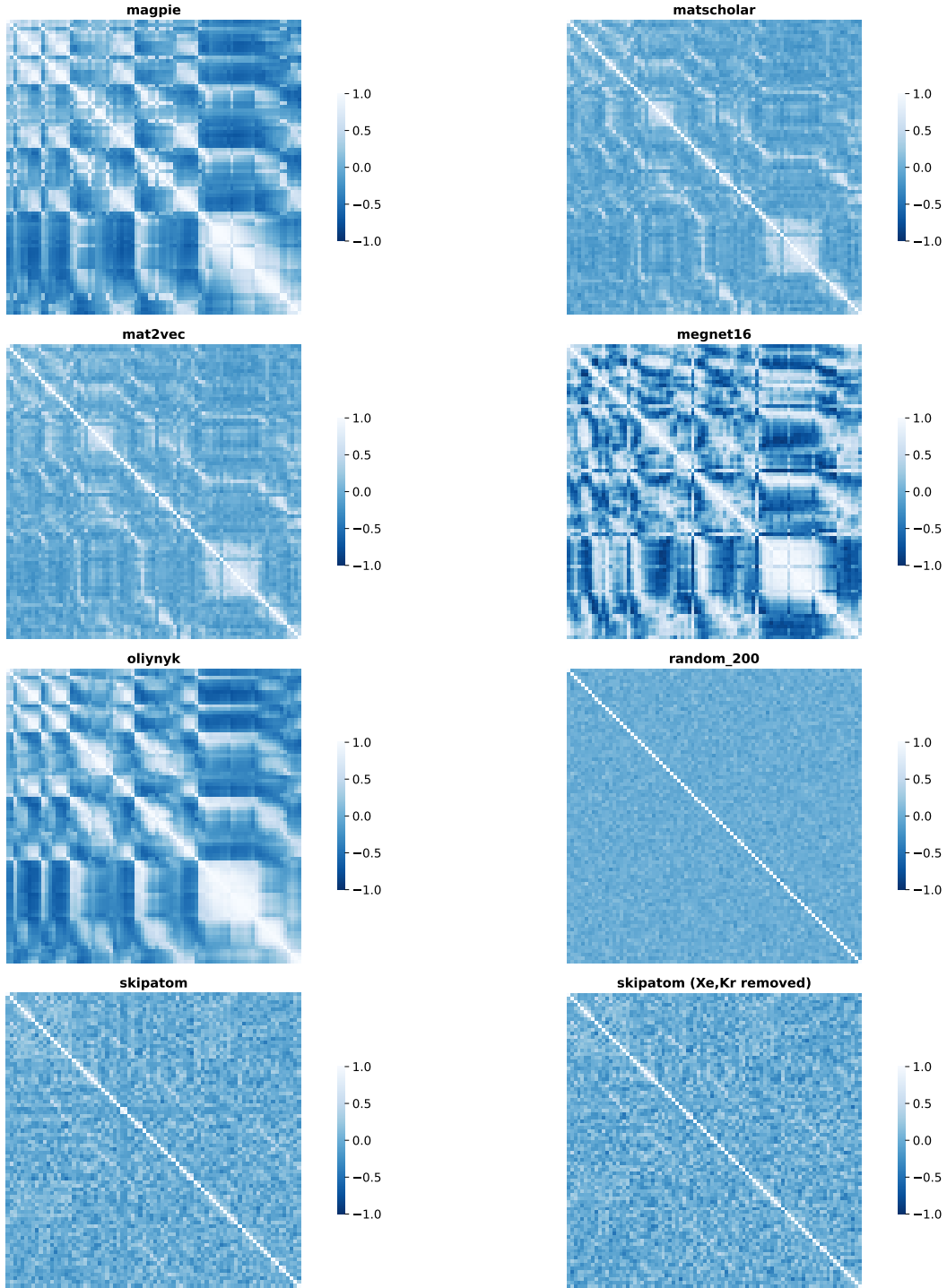
FIG. S6. Map of the Pearson correlation coefficient between element vectors for seven representation schemes. The elements are ordered in increasing atomic number along the axes.
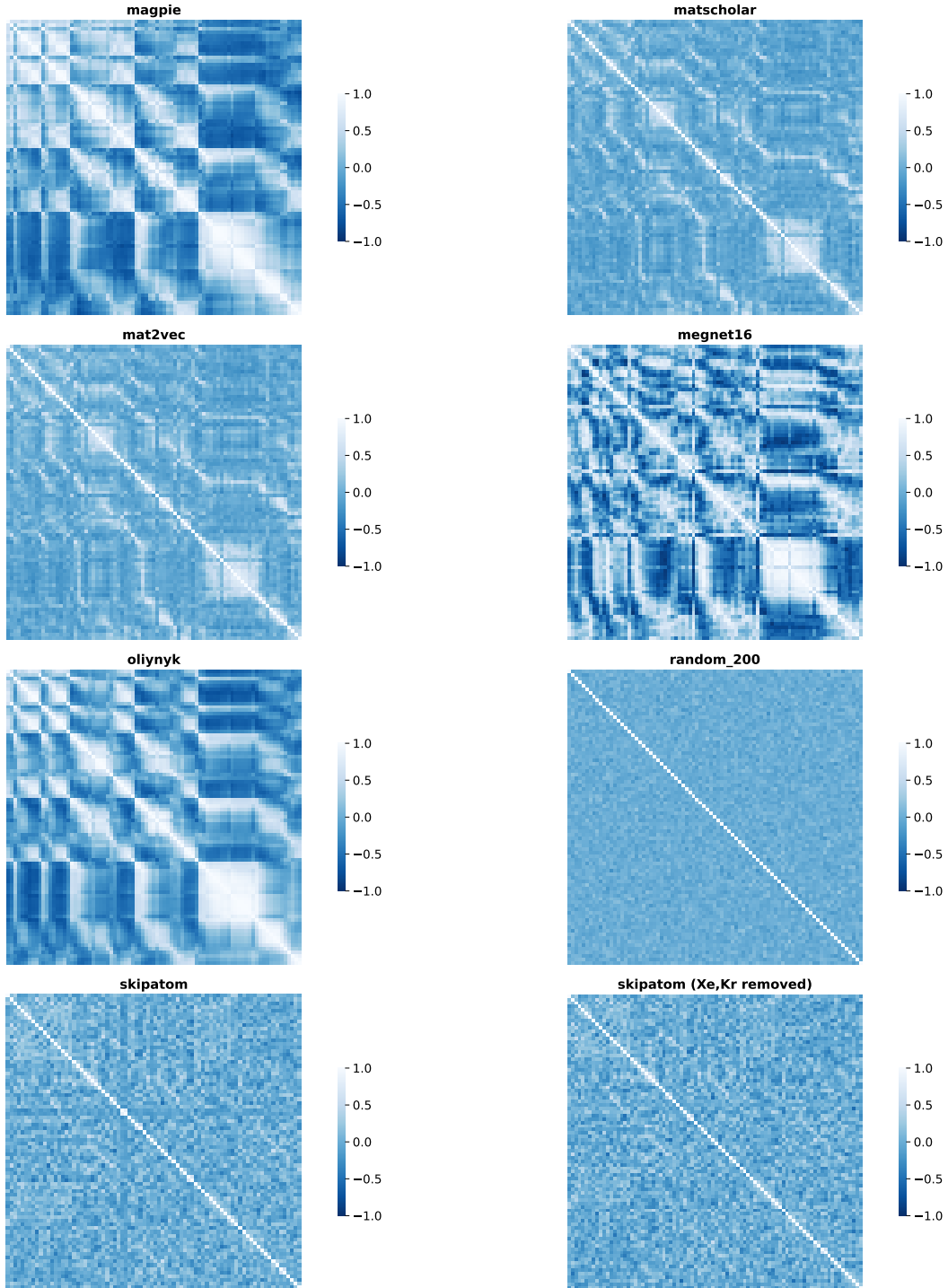
FIG. S7. Map of the Spearman correlation between element vectors for seven representation schemes. The elements are ordered in increasing atomic number along the axes.
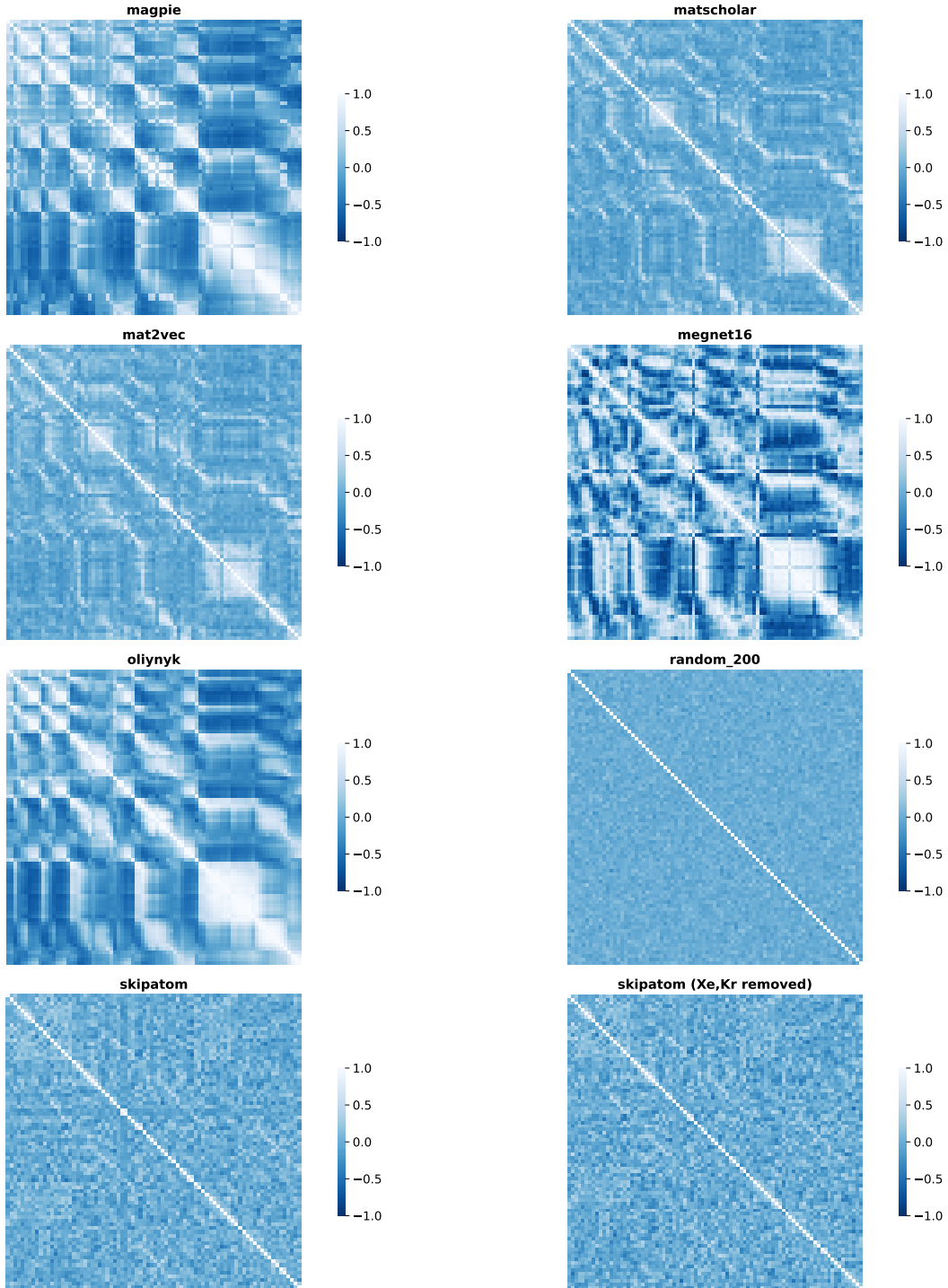
FIG. S8. Map of the cosine similarity between element vectors for seven representation schemes. The elements are ordered in increasing atomic number along the axes.

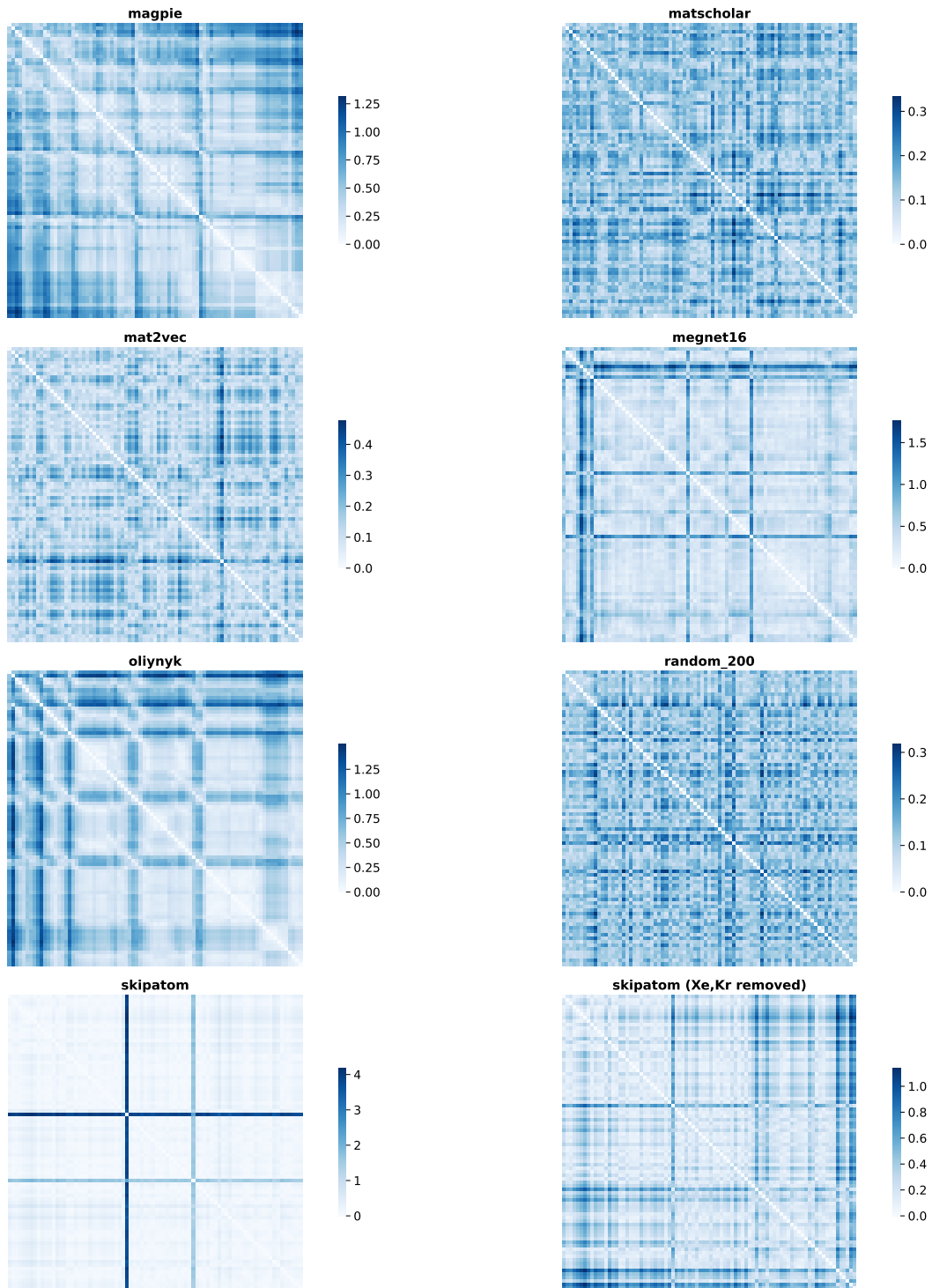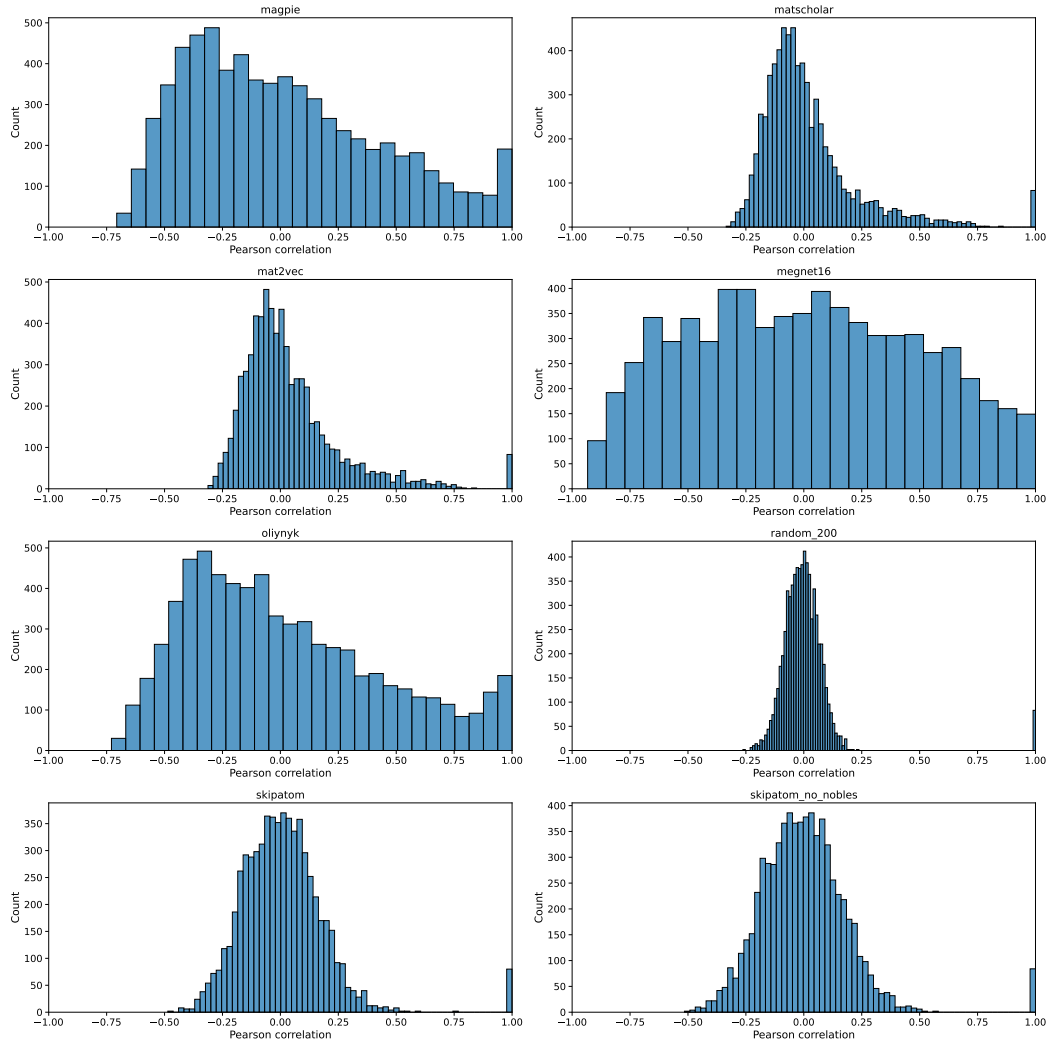FIG. S9. Distribution of the Pearson correlation between the element vectors for the four representation schemes.
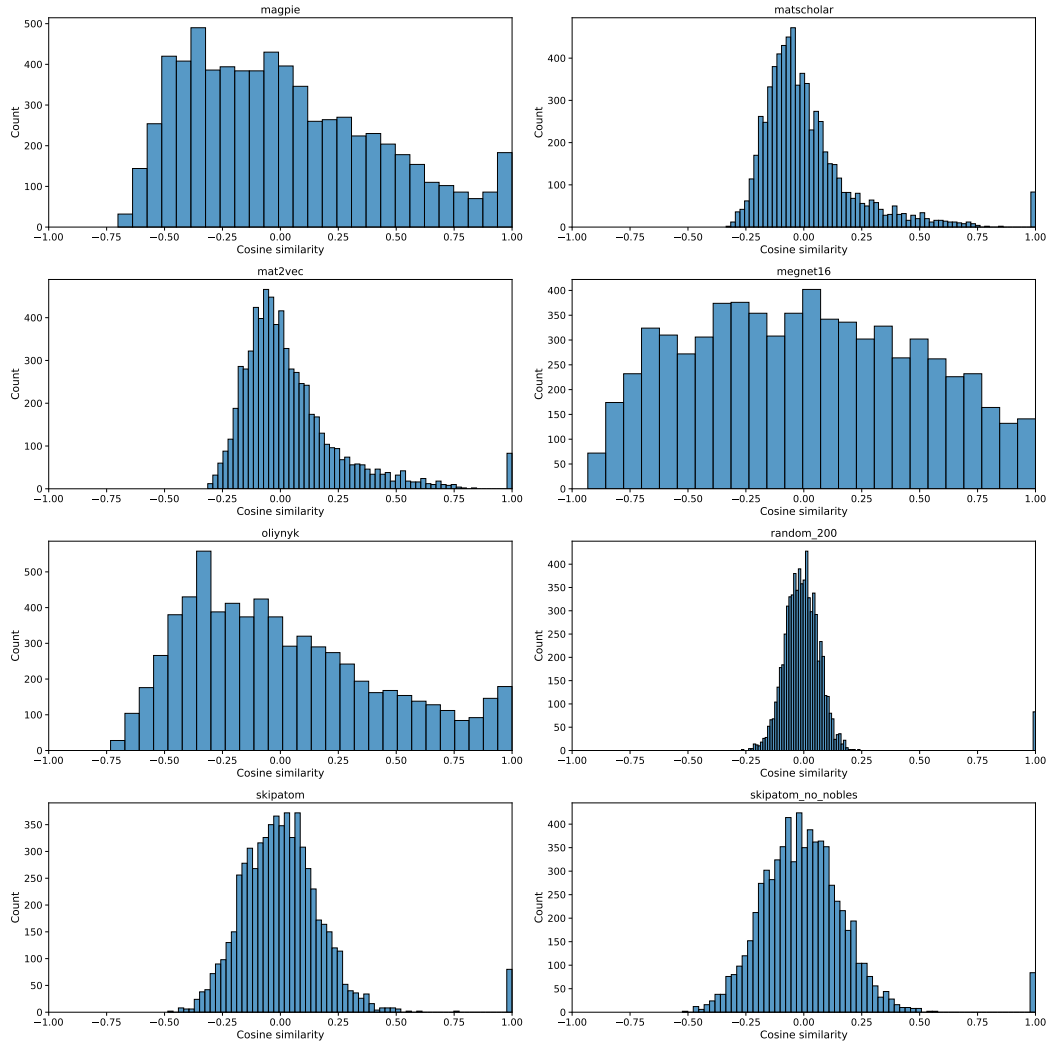
FIG. S10. Distribution of the cosine similarity between the element vectors for the four represen-
tation schemes.
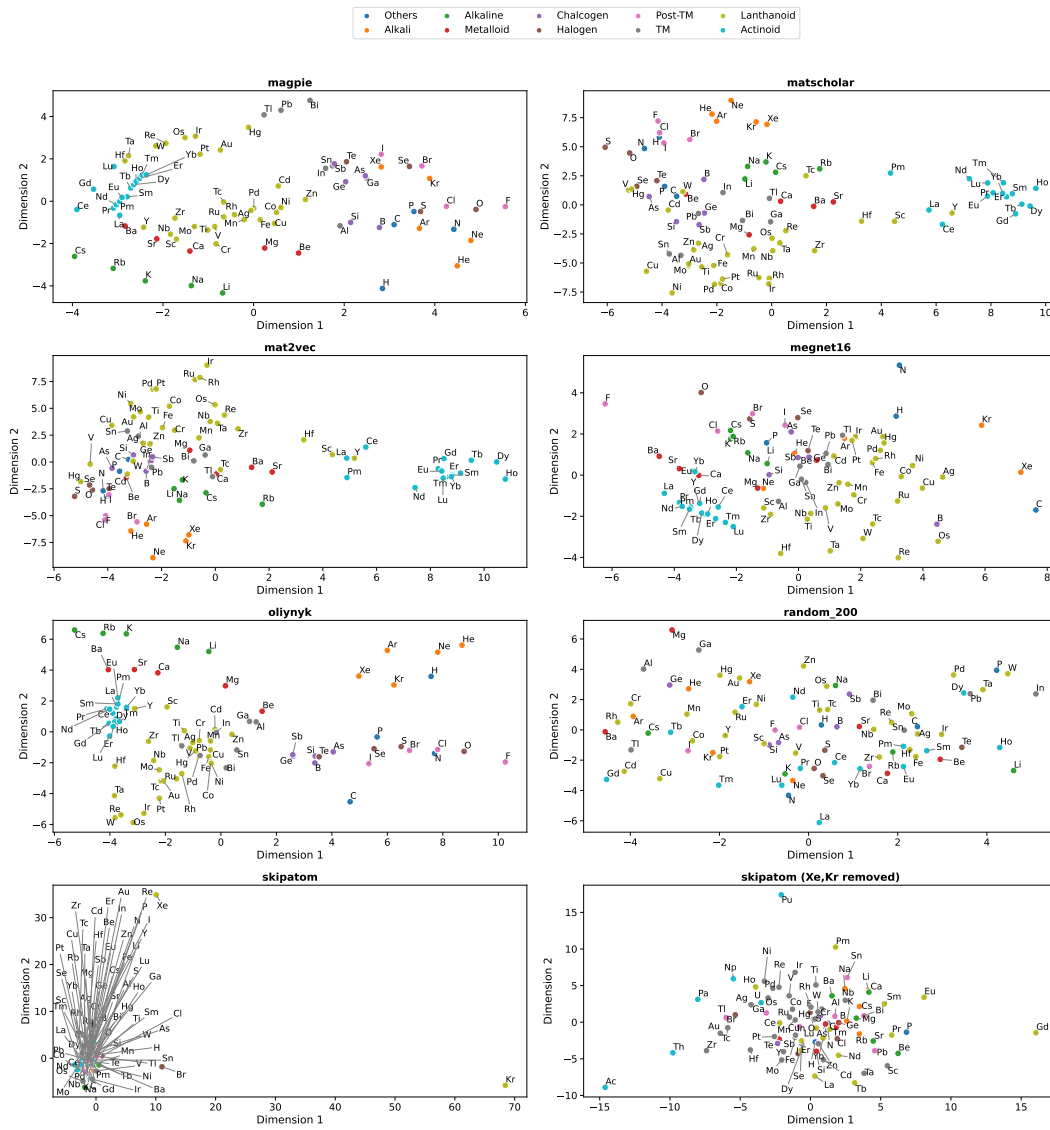
FIG. S11. Two-dimensional projection of seven element representations using principal component analysis (PCA).
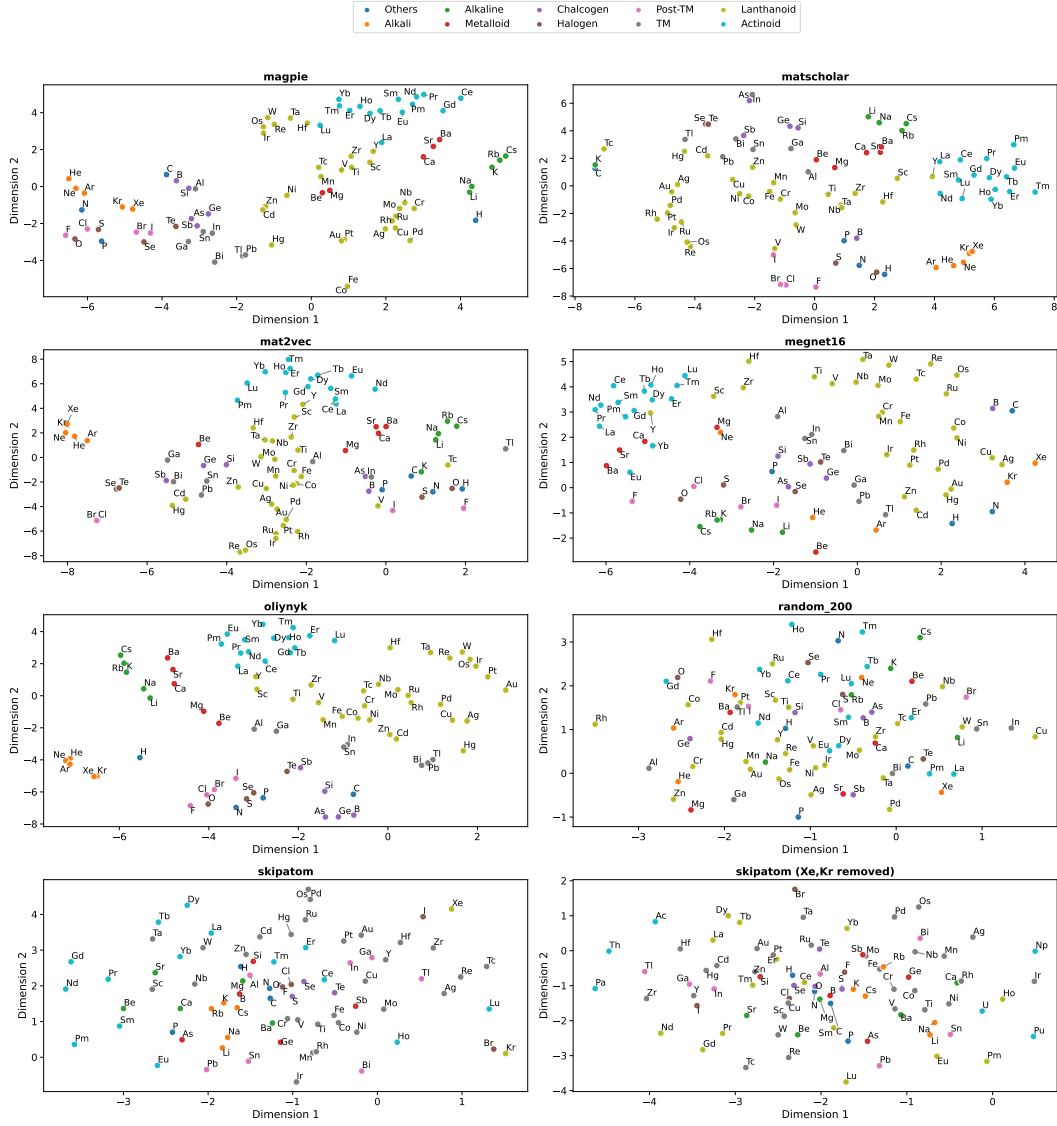
FIG. S12. Two-dimensional projection of seven element representations using t-distributed stochastic neighbours (t-SNE).

FIG. S13. Two-dimensional projection of seven element representations using uniform manifold approximation and projection (UMAP).

TABLE S1: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

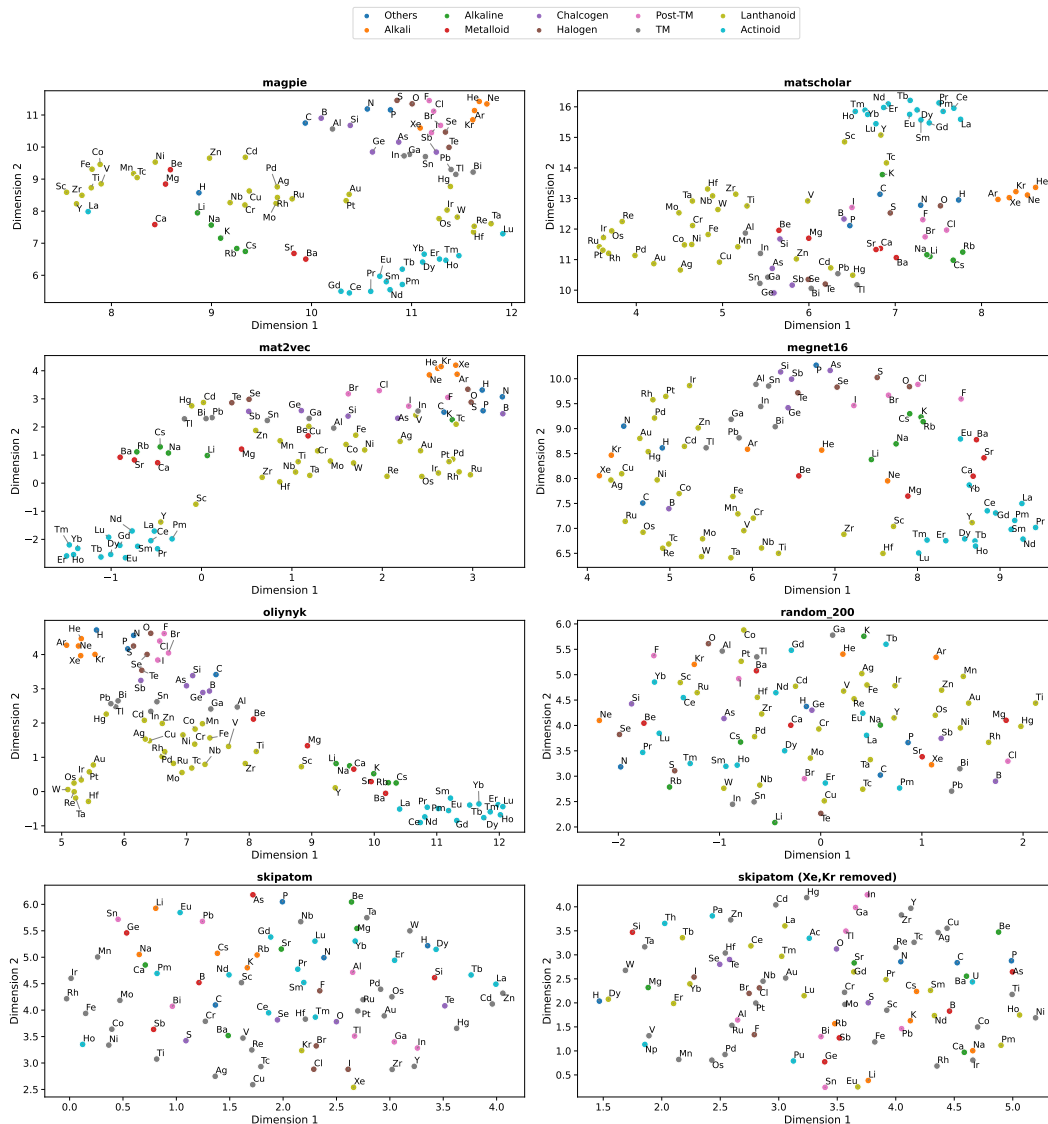| target_formula	oliynyk_template	matscholar_template	mat2vec_template	random_200_template
AgBr (mp-23231)	AgCl (mp-22922)	AgCl (mp-22922)	AgCl (mp-22922)	KBr (mp-23251)
AgCl (mp-22922)	AgBr (mp-23231)	AgBr (mp-23231)	AgBr (mp-23231)	KCl (mp-23193)
AgI (mp-22925)	AgBr (mp-23231)	AgBr (mp-23231)	AgBr (mp-23231)	KI (mp-22898)
AgO (mp-1079720)	AgCl (mp-22922)	AgCl (mp-22922)	AgCl (mp-22922)	AgI (mp-22925)
AlAs (mp-2172)	GaAs (mp-2534)	GaAs (mp-2534)	GaAs (mp-2534)	GaAs (mp-2534)
AlN (mp-661)	AlP (mp-1550)	GaN (mp-804)	AlP (mp-1550)	GaN (mp-804)
AlP (mp-1550)	AlN (mp-661)	AlN (mp-661)	AlN (mp-661)	GaP (mp-2490)
AlSb (mp-2624)	AlP (mp-1550)	AlAs (mp-2172)	AlAs (mp-2172)	AlAs (mp-2172)
BAs (mp-10044)	BP (mp-1479)	BP (mp-1479)	BP (mp-1479)	AlAs (mp-2172)
BP (mp-1479)	BAs (mp-10044)	BAs (mp-10044)	BAs (mp-10044)	AlP (mp-1550)
BaO (mp-1342)	SrO (mp-2472)	SrO (mp-2472)	SrO (mp-2472)	HgO (mp-1224)
BaS (mp-1500)	SrS (mp-1087)	SrS (mp-1087)	SrS (mp-1087)	HgS (mp-634)
BaSe (mp-1253)	SrSe (mp-2758)	SrSe (mp-2758)	SrSe (mp-2758)	HgSe (mp-1018722)
BaTe (mp-1000)	SrTe (mp-1958)	SrTe (mp-1958)	SrTe (mp-1958)	HgTe (mp-358)
BeO (mp-2542)	BeS (mp-422)	BeS (mp-422)	BeS (mp-422)	BeSe (mp-1541)
BeS (mp-422)	BeSe (mp-1541)	BeSe (mp-1541)	BeSe (mp-1541)	PbS (mp-21276)
BeSe (mp-1541)	BeS (mp-422)	BeTe (mp-252)	BeTe (mp-252)	PbSe (mp-2201)
BeTe (mp-252)	MgTe (mp-1039)	BeSe (mp-1541)	BeSe (mp-1541)	TePb (mp-19717)
CaO (mp-2605)	SrO (mp-2472)	SrO (mp-2472)	SrO (mp-2472)	YbO (mp-1216)
CaS (mp-1672)	SrS (mp-1087)	SrS (mp-1087)	SrS (mp-1087)	YbS (mp-1820)
CaSe (mp-1415)	SrSe (mp-2758)	SrSe (mp-2758)	SrSe (mp-2758)	YbSe (mp-286)
CaTe (mp-1519)	SrTe (mp-1958)	SrTe (mp-1958)	SrTe (mp-1958)	YbTe (mp-1779)
CdS (mp-672)	CdSe (mp-2691)	HgS (mp-634)	HgS (mp-634)	HgS (mp-634)
CdSe (mp-2691)	CdS (mp-672)	CdTe (mp-406)	CdTe (mp-406)	HgSe (mp-1018722)

TABLE S1: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

target_formula	oliynyk_template	matscholar_template	mat2vec_template	random_200_template
CdTe (mp-406)	ZnTe (mp-2176)	CdSe (mp-2691)	CdSe (mp-2691)	HgTe (mp-358)
CoO (mp-22408)	NiO (mp-19009)	NiO (mp-19009)	NiO (mp-19009)	MnO (mp-19006)
CsAu (mp-2667)	RbAu (mp-30373)	RbAu (mp-30373)	RbAu (mp-30373)	CsF (mp-1784)
CsBr (mp-571222)	RbBr (mp-22867)	CsCl (mp-573697)	CsCl (mp-573697)	CuBr (mp-22913)
CsCl (mp-573697)	RbCl (mp-23295)	CsBr (mp-571222)	CsBr (mp-571222)	CuCl (mp-22914)
CsF (mp-1784)	RbF (mp-2064)	RbF (mp-2064)	RbF (mp-2064)	NaF (mp-682)
CsH (mp-1057286)	RbH (mp-24721)	RbH (mp-24721)	RbH (mp-24721)	CuH (mp-24093)
CsI (mp-614603)	RbI (mp-22903)	RbI (mp-22903)	RbI (mp-22903)	NaI (mp-23268)
CsTe (mp-8361)	None	None	None	None
CuBr (mp-22913)	CuCl (mp-22914)	CuCl (mp-22914)	CuCl (mp-22914)	KBr (mp-23251)
CuCl (mp-22914)	CuBr (mp-22913)	CuBr (mp-22913)	CuBr (mp-22913)	KCl (mp-23193)
CuH (mp-24093)	CuCl (mp-22914)	CuCl (mp-22914)	CuCl (mp-22914)	KH (mp-24084)
DyN (mp-1410)	HoN (mp-883)	TbN (mp-2117)	TbN (mp-2117)	GaN (mp-804)
ErN (mp-19830)	TmN (mp-1975)	HoN (mp-883)	HoN (mp-883)	TmN (mp-1975)
GaAs (mp-2534)	AlAs (mp-2172)	AlAs (mp-2172)	AlAs (mp-2172)	AlAs (mp-2172)
GaN (mp-804)	GaP (mp-2490)	GaP (mp-2490)	GaP (mp-2490)	AlN (mp-661)
GaP (mp-2490)	GaN (mp-804)	GaN (mp-804)	GaN (mp-804)	AlP (mp-1550)
GeTe (mp-938)	BeTe (mp-252)	SnTe (mp-1883)	SnTe (mp-1883)	BeTe (mp-252)
HgO (mp-1224)	HgS (mp-634)	HgS (mp-634)	HgS (mp-634)	MgO (mp-1265)
HgS (mp-634)	HgSe (mp-1018722)	CdS (mp-672)	CdS (mp-672)	MgS (mp-1315)
HgSe (mp-1018722)	HgS (mp-634)	HgTe (mp-358)	HgTe (mp-358)	MgSe (mp-1018040)
HgTe (mp-358)	TePb (mp-19717)	HgSe (mp-1018722)	HgSe (mp-1018722)	MgTe (mp-1039)
HoN (mp-883)	DyN (mp-1410)	ErN (mp-19830)	ErN (mp-19830)	AlN (mp-661)
InP (mp-20351)	GaP (mp-2490)	BP (mp-1479)	BP (mp-1479)	GaP (mp-2490)

18

TABLE S1: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

target_formula	oliynyk_template	matscholar_template	mat2vec_template	random_200_template
KBr (mp-23251)	RbBr (mp-22867)	KCl (mp-23193)	KCl (mp-23193)	CuBr (mp-22913)
KCl (mp-23193)	RbCl (mp-23295)	KBr (mp-23251)	KBr (mp-23251)	CuCl (mp-22914)
KF (mp-463)	RbF (mp-2064)	KCl (mp-23193)	KCl (mp-23193)	TlF (mp-720)
KH (mp-24084)	RbH (mp-24721)	KF (mp-463)	KF (mp-463)	CuH (mp-24093)
KI (mp-22898)	RbI (mp-22903)	KBr (mp-23251)	KBr (mp-23251)	AgI (mp-22925)
LiBr (mp-23259)	NaBr (mp-22916)	LiCl (mp-22905)	LiCl (mp-22905)	TlBr (mp-568560)
LiCl (mp-22905)	NaCl (mp-22862)	LiBr (mp-23259)	LiBr (mp-23259)	TlCl (mp-569639)
LiF (mp-1138)	NaF (mp-682)	NaF (mp-682)	NaF (mp-682)	TlF (mp-720)
LiH (mp-23703)	NaH (mp-23870)	NaH (mp-23870)	NaH (mp-23870)	RbH (mp-24721)
LiI (mp-570935)	NaI (mp-23268)	NaI (mp-23268)	NaI (mp-23268)	TlI (mp-571102)
LuN (mp-1102)	ErN (mp-19830)	HoN (mp-883)	HoN (mp-883)	YN (mp-2114)
MgO (mp-1265)	MgS (mp-1315)	CaO (mp-2605)	CaO (mp-2605)	ZnO (mp-2133)
MgS (mp-1315)	MgSe (mp-1018040)	CaS (mp-1672)	CaS (mp-1672)	HgS (mp-634)
MgSe (mp-1018040)	MgS (mp-1315)	MgTe (mp-1039)	MgTe (mp-1039)	ZnSe (mp-1190)
MgTe (mp-1039)	CaTe (mp-1519)	MgSe (mp-1018040)	MgSe (mp-1018040)	ZnTe (mp-2176)
MnO (mp-19006)	MnSe (mp-2293)	ZnO (mp-2133)	ZnO (mp-2133)	MnSe (mp-2293)
MnSe (mp-2293)	MnO (mp-19006)	ZnSe (mp-1190)	ZnSe (mp-1190)	CdSe (mp-2691)
NaBr (mp-22916)	KBr (mp-23251)	NaCl (mp-22862)	NaCl (mp-22862)	CsBr (mp-571222)
NaCl (mp-22862)	KCl (mp-23193)	NaBr (mp-22916)	NaBr (mp-22916)	CsCl (mp-573697)
NaF (mp-682)	KF (mp-463)	LiF (mp-1138)	LiF (mp-1138)	CsF (mp-1784)
NaH (mp-23870)	KH (mp-24084)	LiH (mp-23703)	LiH (mp-23703)	CsH (mp-1057286)
NaI (mp-23268)	KI (mp-22898)	LiI (mp-570935)	LiI (mp-570935)	CsI (mp-614603)
NiO (mp-19009)	CoO (mp-22408)	CoO (mp-22408)	CoO (mp-22408)	SrO (mp-2472)
PbS (mp-21276)	PbSe (mp-2201)	CdS (mp-672)	CdS (mp-672)	HgS (mp-634)

19

TABLE S1: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

target_formula	oliynyk_template	matscholar_template	mat2vec_template	random_200_template
PbSe (mp-2201)	PbS (mp-21276)	TePb (mp-19717)	TePb (mp-19717)	HgSe (mp-1018722)
RbAu (mp-30373)	CsAu (mp-2667)	CsAu (mp-2667)	CsAu (mp-2667)	RbF (mp-2064)
RbBr (mp-22867)	CsBr (mp-571222)	RbCl (mp-23295)	RbCl (mp-23295)	LiBr (mp-23259)
RbCl (mp-23295)	CsCl (mp-573697)	RbBr (mp-22867)	RbBr (mp-22867)	LiCl (mp-22905)
RbF (mp-2064)	CsF (mp-1784)	CsF (mp-1784)	CsF (mp-1784)	LiF (mp-1138)
RbH (mp-24721)	CsH (mp-1057286)	CsH (mp-1057286)	CsH (mp-1057286)	LiH (mp-23703)
RbI (mp-22903)	CsI (mp-614603)	CsI (mp-614603)	CsI (mp-614603)	LiI (mp-570935)
SnTe (mp-1883)	CdTe (mp-406)	ZnTe (mp-2176)	ZnTe (mp-2176)	BeTe (mp-252)
SrO (mp-2472)	BaO (mp-1342)	CaO (mp-2605)	BaO (mp-1342)	MgO (mp-1265)
SrS (mp-1087)	BaS (mp-1500)	CaS (mp-1672)	BaS (mp-1500)	MgS (mp-1315)
SrSe (mp-2758)	BaSe (mp-1253)	CaSe (mp-1415)	BaSe (mp-1253)	MgSe (mp-1018040)
SrTe (mp-1958)	BaTe (mp-1000)	CaTe (mp-1519)	BaTe (mp-1000)	MgTe (mp-1039)
TbN (mp-2117)	DyN (mp-1410)	DyN (mp-1410)	DyN (mp-1410)	AlN (mp-661)
TePb (mp-19717)	HgTe (mp-358)	PbSe (mp-2201)	PbSe (mp-2201)	HgTe (mp-358)
TlBr (mp-568560)	TlCl (mp-569639)	TlCl (mp-569639)	TlCl (mp-569639)	LiBr (mp-23259)
TlCl (mp-569639)	TlBr (mp-568560)	TlBr (mp-568560)	TlBr (mp-568560)	LiCl (mp-22905)
TlF (mp-720)	TlCl (mp-569639)	TlCl (mp-569639)	TlCl (mp-569639)	LiF (mp-1138)
TlI (mp-571102)	TlBr (mp-568560)	CsI (mp-614603)	CsI (mp-614603)	LiI (mp-570935)
TmN (mp-1975)	HoN (mp-883)	HoN (mp-883)	HoN (mp-883)	ErN (mp-19830)
VO (mp-19184)	MnO (mp-19006)	HgO (mp-1224)	HgO (mp-1224)	YbO (mp-1216)
YN (mp-2114)	TbN (mp-2117)	HoN (mp-883)	HoN (mp-883)	LuN (mp-1102)
YbO (mp-1216)	YbS (mp-1820)	YbS (mp-1820)	YbS (mp-1820)	CaO (mp-2605)
YbS (mp-1820)	YbSe (mp-286)	YbSe (mp-286)	YbSe (mp-286)	CaS (mp-1672)
YbSe (mp-286)	YbS (mp-1820)	YbTe (mp-1779)	YbTe (mp-1779)	CaSe (mp-1415)

TABLE S1: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

target_formula	oliynyk_template	matscholar_template	mat2vec_template	random_200_template
YbTe (mp-1779)	YbSe (mp-286)	YbSe (mp-286)	YbSe (mp-286)	CaTe (mp-1519)
ZnO (mp-2133)	ZnSe (mp-1190)	MgO (mp-1265)	MgO (mp-1265)	MgO (mp-1265)
ZnSe (mp-1190)	CdSe (mp-2691)	ZnTe (mp-2176)	ZnTe (mp-2176)	MgSe (mp-1018040)
ZnTe (mp-2176)	CdTe (mp-406)	ZnSe (mp-1190)	ZnSe (mp-1190)	MgTe (mp-1039)

TABLE S2: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

target_formula	skipatom_template	magpie_template	megnet16_template	hautier_template
AgBr (mp-23231)	CuBr (mp-22913)	AgI (mp-22925)	AgO (mp-1079720)	AgI (mp-22925)
AgCl (mp-22922)	CuCl (mp-22914)	CuCl (mp-22914)	AgO (mp-1079720)	AgI (mp-22925)
AgI (mp-22925)	AgCl (mp-22922)	AgBr (mp-23231)	AgBr (mp-23231)	AgBr (mp-23231)
AgO (mp-1079720)	AgCl (mp-22922)	AgCl (mp-22922)	AgCl (mp-22922)	None
AlAs (mp-2172)	AlP (mp-1550)	AlSb (mp-2624)	AlP (mp-1550)	None
AlN (mp-661)	AlP (mp-1550)	AlP (mp-1550)	LuN (mp-1102)	GaN (mp-804)
AlP (mp-1550)	AlAs (mp-2172)	BP (mp-1479)	AlAs (mp-2172)	None
AlSb (mp-2624)	AlN (mp-661)	AlAs (mp-2172)	AlP (mp-1550)	None
BAs (mp-10044)	BP (mp-1479)	AlAs (mp-2172)	BP (mp-1479)	None
BP (mp-1479)	BAs (mp-10044)	AlP (mp-1550)	BAs (mp-10044)	None
BaO (mp-1342)	VO (mp-19184)	SrO (mp-2472)	SrO (mp-2472)	BaS (mp-1500)
BaS (mp-1500)	BaSe (mp-1253)	SrS (mp-1087)	SrS (mp-1087)	BaSe (mp-1253)
BaSe (mp-1253)	BaS (mp-1500)	SrSe (mp-2758)	SrSe (mp-2758)	BaS (mp-1500)
BaTe (mp-1000)	SrTe (mp-1958)	SrTe (mp-1958)	SrTe (mp-1958)	None

TABLE S2: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

target_formula	skipatom_template	magpie_template	megnet16_template	hautier_template
BeO (mp-2542)	MgO (mp-1265)	BeS (mp-422)	BeS (mp-422)	BeS (mp-422)
BeS (mp-422)	BeSe (mp-1541)	MgS (mp-1315)	BeO (mp-2542)	BeSe (mp-1541)
BeSe (mp-1541)	BeS (mp-422)	MgSe (mp-1018040)	BeS (mp-422)	BeS (mp-422)
BeTe (mp-252)	MgTe (mp-1039)	MgTe (mp-1039)	BeSe (mp-1541)	None
CaO (mp-2605)	SrO (mp-2472)	SrO (mp-2472)	YbO (mp-1216)	CaS (mp-1672)
CaS (mp-1672)	CaSe (mp-1415)	SrS (mp-1087)	YbS (mp-1820)	CaSe (mp-1415)
CaSe (mp-1415)	CaS (mp-1672)	SrSe (mp-2758)	YbSe (mp-286)	CaS (mp-1672)
CaTe (mp-1519)	SrTe (mp-1958)	SrTe (mp-1958)	YbTe (mp-1779)	None
CdS (mp-672)	CdSe (mp-2691)	CdSe (mp-2691)	CdSe (mp-2691)	CdSe (mp-2691)
CdSe (mp-2691)	HgSe (mp-1018722)	ZnSe (mp-1190)	CdS (mp-672)	CdS (mp-672)
CdTe (mp-406)	HgTe (mp-358)	ZnTe (mp-2176)	ZnTe (mp-2176)	None
CoO (mp-22408)	NiO (mp-19009)	NiO (mp-19009)	NiO (mp-19009)	NiO (mp-19009)
CsAu (mp-2667)	RbAu (mp-30373)	RbAu (mp-30373)	RbAu (mp-30373)	None
CsBr (mp-571222)	RbBr (mp-22867)	RbBr (mp-22867)	RbBr (mp-22867)	CsI (mp-614603)
CsCl (mp-573697)	RbCl (mp-23295)	RbCl (mp-23295)	CsF (mp-1784)	CsI (mp-614603)
CsF (mp-1784)	RbF (mp-2064)	RbF (mp-2064)	CsCl (mp-573697)	CsCl (mp-573697)
CsH (mp-1057286)	RbH (mp-24721)	RbH (mp-24721)	RbH (mp-24721)	None
CsI (mp-614603)	RbI (mp-22903)	RbI (mp-22903)	RbI (mp-22903)	RbI (mp-22903)
CsTe (mp-8361)	None	None	None	None
CuBr (mp-22913)	AgBr (mp-23231)	AgBr (mp-23231)	AgBr (mp-23231)	AgBr (mp-23231)
CuCl (mp-22914)	AgCl (mp-22922)	AgCl (mp-22922)	AgCl (mp-22922)	CuBr (mp-22913)
CuH (mp-24093)	CuBr (mp-22913)	LiH (mp-23703)	CuBr (mp-22913)	None
DyN (mp-1410)	TbN (mp-2117)	TbN (mp-2117)	ErN (mp-19830)	ErN (mp-19830)
ErN (mp-19830)	TmN (mp-1975)	HoN (mp-883)	TmN (mp-1975)	DyN (mp-1410)

TABLE S2: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

target_formula	skipatom_template	magpie_template	megnet16_template	hautier_template
GaAs (mp-2534)	GaP (mp-2490)	GaP (mp-2490)	GaP (mp-2490)	None
GaN (mp-804)	YN (mp-2114)	GaP (mp-2490)	GaAs (mp-2534)	AlN (mp-661)
GaP (mp-2490)	GaAs (mp-2534)	InP (mp-20351)	GaAs (mp-2534)	None
GeTe (mp-938)	SnTe (mp-1883)	SnTe (mp-1883)	TePb (mp-19717)	None
HgO (mp-1224)	HgS (mp-634)	HgS (mp-634)	HgS (mp-634)	HgS (mp-634)
HgS (mp-634)	HgSe (mp-1018722)	HgO (mp-1224)	HgO (mp-1224)	HgSe (mp-1018722)
HgSe (mp-1018722)	CdSe (mp-2691)	HgTe (mp-358)	CdSe (mp-2691)	HgS (mp-634)
HgTe (mp-358)	CdTe (mp-406)	TePb (mp-19717)	CdTe (mp-406)	None
HoN (mp-883)	TmN (mp-1975)	DyN (mp-1410)	TbN (mp-2117)	TmN (mp-1975)
InP (mp-20351)	GaP (mp-2490)	GaP (mp-2490)	AlP (mp-1550)	None
KBr (mp-23251)	RbBr (mp-22867)	RbBr (mp-22867)	RbBr (mp-22867)	KI (mp-22898)
KCl (mp-23193)	RbCl (mp-23295)	RbCl (mp-23295)	RbCl (mp-23295)	KI (mp-22898)
KF (mp-463)	RbF (mp-2064)	RbF (mp-2064)	RbF (mp-2064)	KCl (mp-23193)
KH (mp-24084)	RbH (mp-24721)	RbH (mp-24721)	RbH (mp-24721)	None
KI (mp-22898)	RbI (mp-22903)	RbI (mp-22903)	RbI (mp-22903)	KBr (mp-23251)
LiBr (mp-23259)	NaBr (mp-22916)	NaBr (mp-22916)	NaBr (mp-22916)	LiI (mp-570935)
LiCl (mp-22905)	NaCl (mp-22862)	NaCl (mp-22862)	LiF (mp-1138)	LiI (mp-570935)
LiF (mp-1138)	NaF (mp-682)	NaF (mp-682)	LiCl (mp-22905)	LiCl (mp-22905)
LiH (mp-23703)	NaH (mp-23870)	NaH (mp-23870)	NaH (mp-23870)	None
LiI (mp-570935)	NaI (mp-23268)	NaI (mp-23268)	LiBr (mp-23259)	LiBr (mp-23259)
LuN (mp-1102)	ErN (mp-19830)	ErN (mp-19830)	TmN (mp-1975)	TmN (mp-1975)
MgO (mp-1265)	BeO (mp-2542)	MgS (mp-1315)	CaO (mp-2605)	MgS (mp-1315)
MgS (mp-1315)	MgSe (mp-1018040)	MgO (mp-1265)	CaS (mp-1672)	MgSe (mp-1018040)
MgSe (mp-1018040)	MgS (mp-1315)	MgTe (mp-1039)	CaSe (mp-1415)	MgS (mp-1315)

TABLE S2: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

target_formula	skipatom_template	magpie_template	megnet16_template	hautier_template
MgTe (mp-1039)	BeTe (mp-252)	MgSe (mp-1018040)	CaTe (mp-1519)	None
MnO (mp-19006)	NiO (mp-19009)	VO (mp-19184)	MnSe (mp-2293)	MnSe (mp-2293)
MnSe (mp-2293)	YbSe (mp-286)	MnO (mp-19006)	MnO (mp-19006)	CdSe (mp-2691)
NaBr (mp-22916)	LiBr (mp-23259)	KBr (mp-23251)	LiBr (mp-23259)	NaI (mp-23268)
NaCl (mp-22862)	LiCl (mp-22905)	KCl (mp-23193)	NaF (mp-682)	NaI (mp-23268)
NaF (mp-682)	LiF (mp-1138)	KF (mp-463)	NaCl (mp-22862)	NaCl (mp-22862)
NaH (mp-23870)	LiH (mp-23703)	KH (mp-24084)	LiH (mp-23703)	None
NaI (mp-23268)	LiI (mp-570935)	KI (mp-22898)	LiI (mp-570935)	NaBr (mp-22916)
NiO (mp-19009)	CoO (mp-22408)	CoO (mp-22408)	CoO (mp-22408)	VO (mp-19184)
PbS (mp-21276)	PbSe (mp-2201)	HgS (mp-634)	PbSe (mp-2201)	PbSe (mp-2201)
PbSe (mp-2201)	PbS (mp-21276)	HgSe (mp-1018722)	PbS (mp-21276)	PbS (mp-21276)
RbAu (mp-30373)	CsAu (mp-2667)	CsAu (mp-2667)	CsAu (mp-2667)	None
RbBr (mp-22867)	KBr (mp-23251)	CsBr (mp-571222)	KBr (mp-23251)	RbI (mp-22903)
RbCl (mp-23295)	KCl (mp-23193)	CsCl (mp-573697)	KCl (mp-23193)	RbI (mp-22903)
RbF (mp-2064)	KF (mp-463)	CsF (mp-1784)	KF (mp-463)	RbCl (mp-23295)
RbH (mp-24721)	KH (mp-24084)	CsH (mp-1057286)	KH (mp-24084)	None
RbI (mp-22903)	KI (mp-22898)	CsI (mp-614603)	KI (mp-22898)	RbBr (mp-22867)
SnTe (mp-1883)	GeTe (mp-938)	GeTe (mp-938)	ZnTe (mp-2176)	None
SrO (mp-2472)	CaO (mp-2605)	BaO (mp-1342)	CaO (mp-2605)	SrS (mp-1087)
SrS (mp-1087)	SrSe (mp-2758)	BaS (mp-1500)	CaS (mp-1672)	SrSe (mp-2758)
SrSe (mp-2758)	SrS (mp-1087)	BaSe (mp-1253)	CaSe (mp-1415)	SrS (mp-1087)
SrTe (mp-1958)	CaTe (mp-1519)	BaTe (mp-1000)	CaTe (mp-1519)	None
TbN (mp-2117)	DyN (mp-1410)	DyN (mp-1410)	DyN (mp-1410)	TmN (mp-1975)
TePb (mp-19717)	SnTe (mp-1883)	HgTe (mp-358)	CdTe (mp-406)	None

TABLE S2: Table of the target formulas from the Materials Project and the template materials used to predict the structure of the target material. The template materials can be considered the most similar under each representation.

target_formula	skipatom_template	magpie_template	megnet16_template	hautier_template
TlBr (mp-568560)	TlCl (mp-569639)	TlI (mp-571102)	TlI (mp-571102)	TlI (mp-571102)
TlCl (mp-569639)	TlI (mp-571102)	TlF (mp-720)	TlF (mp-720)	TlI (mp-571102)
TlF (mp-720)	KF (mp-463)	TlCl (mp-569639)	TlCl (mp-569639)	TlCl (mp-569639)
TlI (mp-571102)	TlCl (mp-569639)	TlBr (mp-568560)	TlBr (mp-568560)	TlBr (mp-568560)
TmN (mp-1975)	ErN (mp-19830)	ErN (mp-19830)	ErN (mp-19830)	TbN (mp-2117)
VO (mp-19184)	BaO (mp-1342)	MnO (mp-19006)	MnO (mp-19006)	NiO (mp-19009)
YN (mp-2114)	GaN (mp-804)	LuN (mp-1102)	DyN (mp-1410)	TbN (mp-2117)
YbO (mp-1216)	YbS (mp-1820)	YbS (mp-1820)	CaO (mp-2605)	YbS (mp-1820)
YbS (mp-1820)	YbSe (mp-286)	YbO (mp-1216)	CaS (mp-1672)	YbSe (mp-286)
YbSe (mp-286)	YbS (mp-1820)	YbTe (mp-1779)	CaSe (mp-1415)	YbS (mp-1820)
YbTe (mp-1779)	ZnTe (mp-2176)	YbSe (mp-286)	CaTe (mp-1519)	None
ZnO (mp-2133)	YbO (mp-1216)	ZnSe (mp-1190)	ZnSe (mp-1190)	ZnSe (mp-1190)
ZnSe (mp-1190)	CdSe (mp-2691)	CdSe (mp-2691)	ZnO (mp-2133)	CdSe (mp-2691)
ZnTe (mp-2176)	CdTe (mp-406)	CdTe (mp-406)	CdTe (mp-406)	None