Supplementary Information for the paper: Understanding the structure-band gap relationship in $SrZrS_3$ at elevated temperatures: A detailed NPT MD study.

Namrata Jaykhedkar,[a,b] Roman Bystrický,[a,c] Milan Sýkora,*[a] and Tomáš Bučko*[b,c]
[a] *Laboratory for Advanced Materials, Faculty of Natural Sciences, Comenius University, Ilkovičova 6, 841 04 Bratislava, Slovakia.* [b] *Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University, Ilkovičova 6, 841 04 Bratislava, Slovakia.* [c] *Institute of Inorganic Chemistry, Slovak Academy of Sciences, Dúbravská Cesta 9, 842 36, Bratislava, Slovakia*
*e-mail: *tomas.bucko@uniba.sk, sykoram@uniba.sk*

SI. VALIDATION OF THE MLFF APPROACH

In this section we demonstrate that the MLFF [S1, S2] approach used in this work reproduces accurately full ab initio molecular dynamics (AIMD) results. To this end, we performed reference AIMD calculations with the same setting as described in section 2 of the main text except of the k-points sampling that was reduced from 2×2×2 to a single point (Γ). Since 2×4×1 and 2×2×2 supercells have been used in MD simulations, the reduced k-points grid is perfectly consistent with the grids of 2×4×1 (NL) and 2×2×2 (DP/C) points employed in calculations with conventional unit cells. This compromise allowed us to produce trajectories of the length of 60 ps in a reasonable time. As shown in Tab. S1, the k-points grid reduction has only a modest effect on the cell geometry of all three phases of $SrZrS_3$ discussed in this work.

TABLE S1. Optimized lengths of the lattice vectors (a, b, c), volumes (V) of conventional cells, and bulk moduli (B_0) of needle-like (NL), distorted perovskite (DP) and cubic (C) phases of $SrZrS_3$ obtained from static calculations with different k-points grids. The values in the parentheses correspond to cubic cell expressed in the setting of distorted perovskite structure. Note that, unlike in MD, conventional unit cells were used in static calculations.

k-points	a (Å)	b (Å)	c (Å)	V (Å^3)	B_0 (GPa)
			NL phase		
4×8×2	8.649	3.840	14.015	465.4	50.3
2×4×1	8.607	3.849	14.032	467.3	48.8
			DP phase		
4×4×4	7.169	9.829	6.787	478.2	69.7
2×2×2	7.198	9.803	6.800	479.9	65.9
			C phase		
4×4×4	5.007			125.6	73.3
	(7.082)	(10.015)	(7.082)	(502.3)	
2×2×2	5.016			126.2	73.2
	(7.094)	(10.033)	(7.094)	(504.9)	

In our first test we performed MLFF calculations trained for the NL and DP/C phases at 300 K and 1500 K. Unlike the MLFF results presented in the main text, the same reduced k-points sampling as in our reference AIMD calculations was used to train the MLFF for the purposes of this test. Besides the k-points sampling, the same training procedure as described in Sec. 2 of the main text was employed.

As shown in Tab S2, the MLFF yields average lattice geometries that are in a very good agreement with the uncompromised DFT calculations. Notably, both the AIMD and MLFF simulations initialized from the DP phase at 1500 K converged to a (quasi-)C phase, consistent with the results presented in the main text. Also the internal structure of $SrZrS_3$ is very well reproduced at all temperatures, as evident from a virtually perfect overlap of radial distribution functions computed using DFT and MLFF (Figs. S1 and S2).

TABLE S2. Comparison of the structural parameters obtained from full DFT and MLFF MD simulations in *NPT* ensemble for the NL and DP phases at 300 K and 1500 K.

T (K)		a (Å)	b (Å)	c (Å)	V (Å^3)
		NL phase			
300	Full DFT	8.637±0.010	3.867±0.000	14.116±0.014	471.4±0.4
	MLFF	8.635±0.000	3.866±0.001	14.113±0.000	471.2±0.0
1500	Full DFT	8.850±0.008	3.962±0.004	14.489±0.012	507.7±0.6
	MLFF	8.862±0.000	3.960±0.000	14.469±0.000	507.4±0.1
		DP/C phase			
300	Full DFT	7.202±0.000	9.859±0.004	6.825±0.002	484.6±0.2
	MLFF	7.201± 0.000	9.861± 0.000	6.821± 0.000	484.3± 0.0
1500	Full DFT	7.166±0.014	10.100±0.014	7.160±0.004	518.0±1.0
	MLFF	7.148±0.001	10.111±0.000	7.148±0.001	516.3±0.1

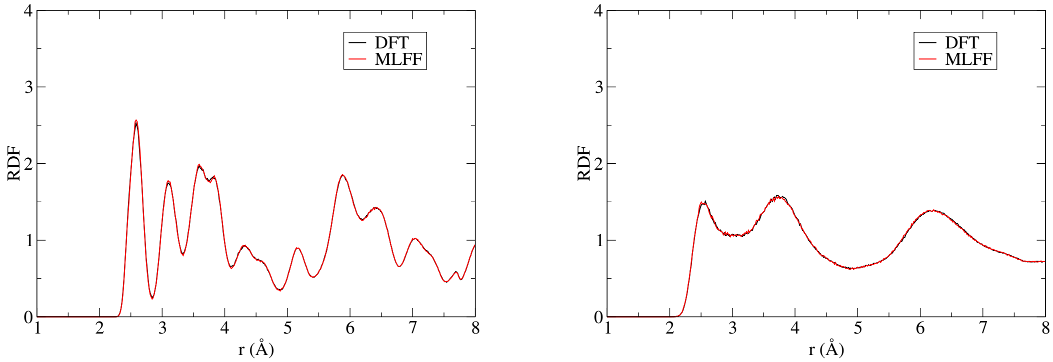

FIG. S1. Comparison of radial distribution functions obtained from *NPT* MD simulations of NL phase of $SrZrS_3$ at 300 K (left) and 1500 K (right) using DFT and MLFF.

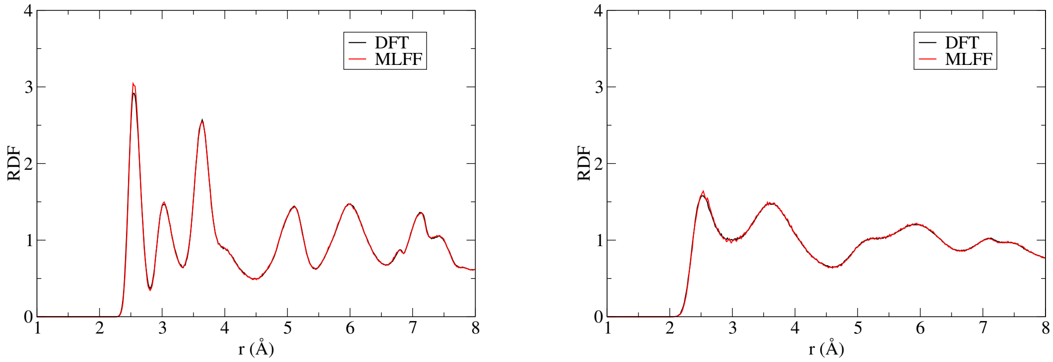

FIG. S2. Comparison of radial distribution functions obtained from *NPT* MD simulations of DP phase of $SrZrS_3$ at 300 K (left) and 1500 K (right) using DFT and MLFF.

Next, we compare our finite temperature results presented in the main text (note that, in this case, the MLFF has been trained using extended k-points mesh ($2\times2\times2$)) with the reference AIMD (performed with reduced k-points mesh, as discussed above). The results for the lattice parameters are presented in Figs. S3 and S4. As can be seen, the AIMD and MLFF results are consistent in predicting the trends and the differences are to be attributed to the different k-points sampling (cf. Tab. S1). We note that the trends obtained from our MD (MLFF or AIMD) results are consistent with our experimental measurements, although the theoretical predictions and experimental data differ somewhat in the absolute values of the lattice parameters (Fig. S4).

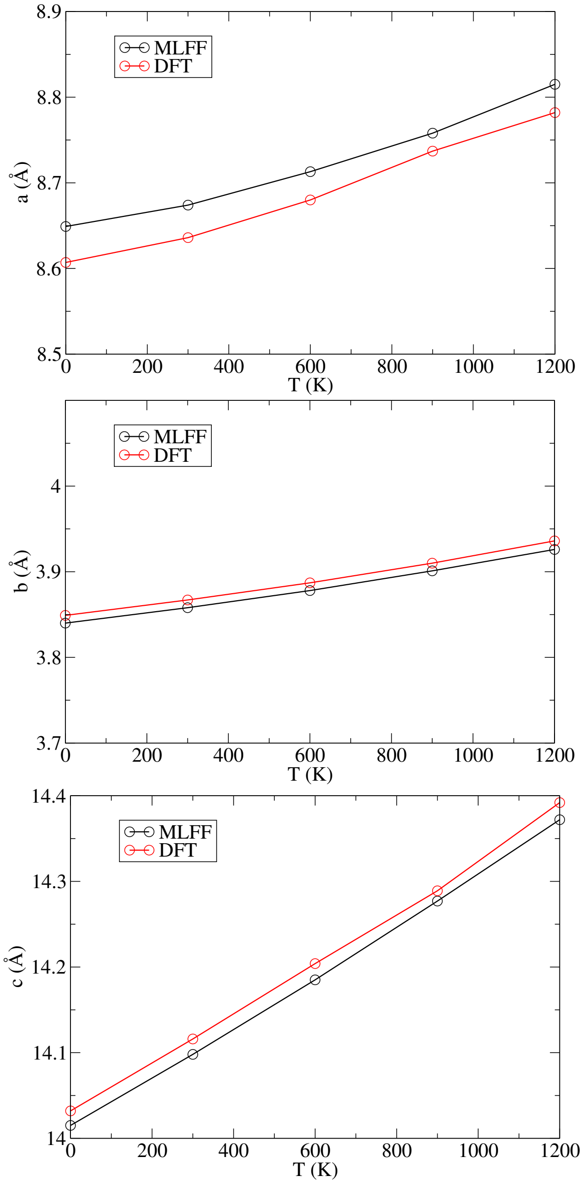

FIG. S3. Variation in length of lattice vectors in conventional cells of NL phase as a function of temperature. Results obtained from DFT with reduced k-points sampling and MLFF with extended k-points sampling are compared.

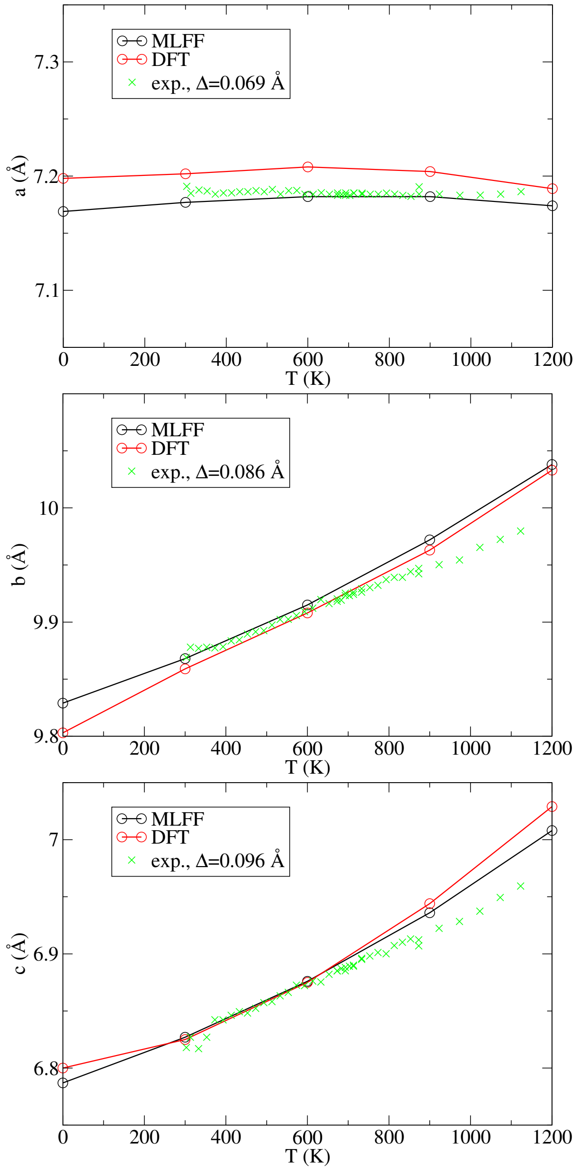

FIG. S4. Variation in length of lattice vectors in conventional cells of DP phase as a function of temperature. Results obtained from DFT with reduced k-points sampling and MLFF with extended k-points sampling are compared with the temperature-dependent x-ray diffraction data obtained here for the temperature range 300-1100 K. Note that the experimental data were shifted by a constant value Δ accounting for the difference between MLFF value at 300 K and the experimental results from Ref. [S3]

SII. CONVERGENCE OF THE MD RESULTS WITH THE SUPERCELL SIZE

The supercells used in the MD simulations must be large enough so that all the essential vibrational modes contributing to the temperature induced structural changes can be described by the atomic motions of atoms within the cell. In this section, we test the convergence of our structural predictions obtained from NPT MD simulations accelerated by MLFF with the size of the supercell. To this end, MD simulations of length of 300 ps have been performed with various sizes of supercell whereby the same simulation setting as used in the MD calculations presented in the main text has been employed. For the DP/C structures at 300 K and 1500 K, the supercells chosen as the 1×1×1 (20 atoms), 2×2×2 (160 atoms), and 3×3×3 (540 atoms) multiples of the primitive cell have been tested, while for the NL phase at 300 K and 1200 K the 2×4×1 (160 atoms) and 3×5×2 (600 atoms) supercells were compared. It is evident from the results presented in Tab. S3 and Fig. S5 and S6 that increasing the cell size beyond the 2×2×2 (DP/C) or 2×4×1 (NL) multiples of the primitive cells does not significantly alter predictions of the lattice parameters nor the internal structures of the crystal. This result therefore justifies the simulation setting used to produce the results presented in the main text of this work.

TABLE S3. Comparison of the lattice parameters of $SrZrS_3$ in the DP/C phase at 300 K and 1500 K and of the NL phase at 300 K and 1200 K as obtained from MLFF NPT MD simulations using supercells of different sizes.

supercell	a Å	b Å	c Å	V Å^3
		DP 300 K		
1×1×1	7.178	9.869	6.839	484.2
2×2×2	7.177	9.868	6.827	483.5
3×3×3	7.177	9.868	6.826	483.4
		DP/C 1500 K		
1×1×1	7.167	10.129	7.169	518.6
2×2×2	7.142	10.103	7.143	515.2
3×3×3	7.143	10.113	7.147	516.2
		NL 300 K		
2×4×1	8.674	3.858	14.098	471.7
3×5×2	8.673	3.858	14.094	471.6
		NL 1200 K		
2×4×1	8.815	3.926	14.372	497.2
3×5×2	8.808	3.925	14.365	496.6

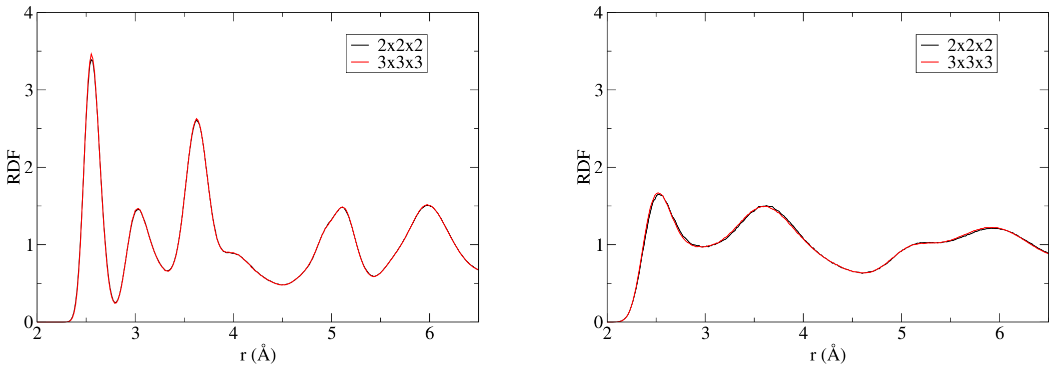

FIG. S5. Radial distribution functions for the Zr-S atomic pairs in the DP/C phase at 300 K (left) and 1500 K (right) obtained using MLFF NPT MD simulations performed with supercells of different sizes (2×2×2 and 3×3×3 multiples of the primitive cell of DP). Notice that the two sets of data generated for each temperature overlap almost perfectly.

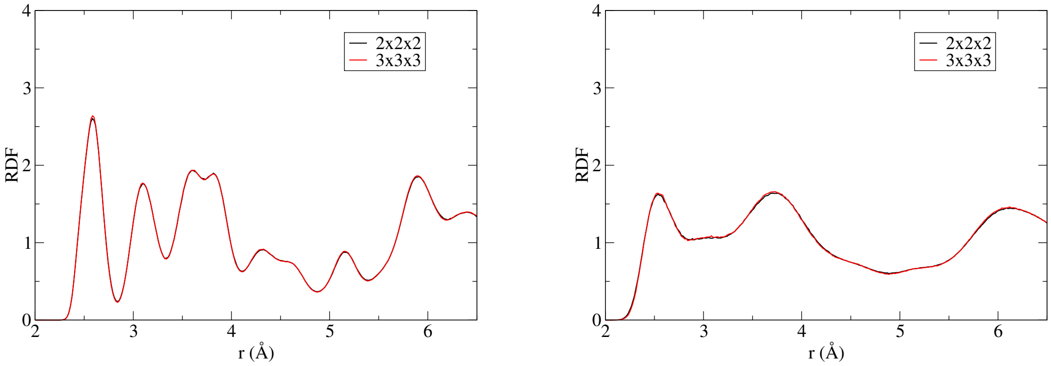

FIG. S6. Radial distribution functions for the Zr-S atomic pairs in the NL phase at 300 K (left) and 1200 K (right) obtained using MLFF NPT MD simulations performed with supercells of different sizes (2×4×1 and 3×5×2 multiples of the primitive cell of DP). Notice that the two sets of data generated for each temperature overlap almost perfectly.

SIII. FURTHER INFORMATION ON ZERO TEMPERATURE CALCULATIONS

TABLE S4. The Γ-point vibrational wave numbers ($\tilde{\nu}$) computed for the relaxed cubic structure of $SrZrS_3$. Note that the frequencies of modes 1 and 2 are imaginary, as indicated by the symbol $i = \sqrt{-1}$. All modes are three-fold degenerate.

mode	$\tilde{\nu}$ (cm^{-1})
1	79.4 i
2	58.4 i
3	31.7
4	336.0

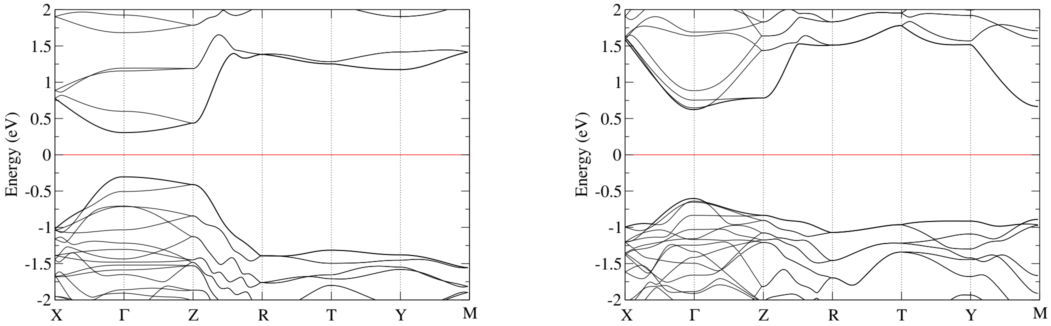

FIG. S7. Band structure near the Fermi level (red line) determined for the relaxed configurations of NL (left) and DP (right) phases of $SrZrS_3$.

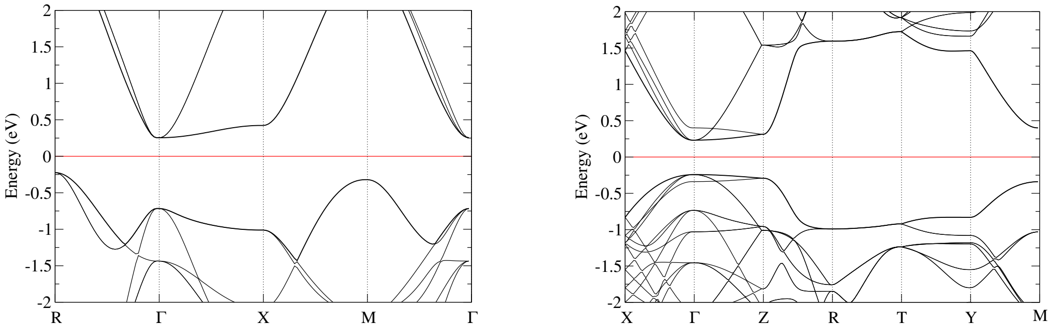

FIG. S8. Band structure near the Fermi level (red line) determined for the C phase of $SrZrS_3$. As a result of zone folding [S4, S5], the indirect band gap present in the band structure computed using the primitive cell (left) appears as a direct gap at the Γ-point when a supercell compatible with the DP cell is used in the calculation (right).

SIV. FURTHER INFORMATION ON THE FINITE TEMPERATURE CALCULATIONS

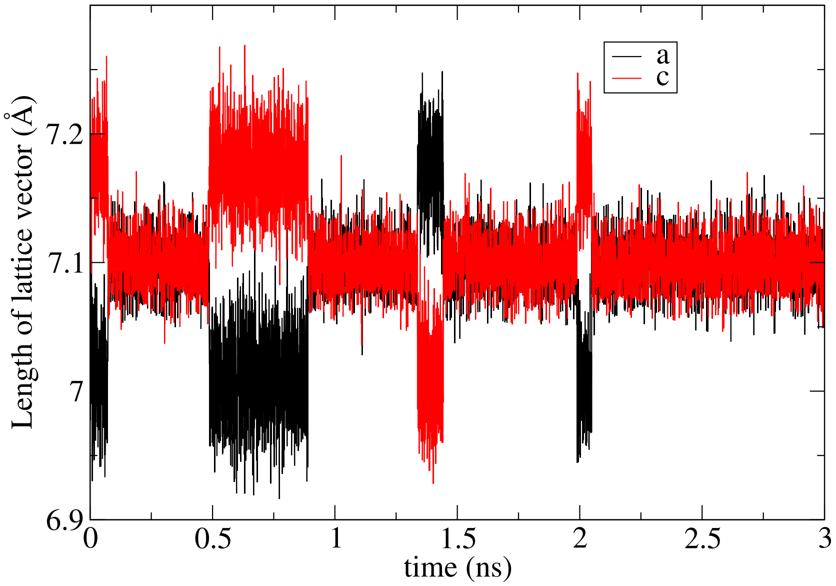

FIG. S9. Time evolution of the lengths of lattice vectors **a** and **c** of DP/C phase of $SrZrS_3$ at 1200 K. The sequences with $a \approx c$ and $|a - c| > 0$ correspond to the (quasi-)C and DP phases, respectively. Note that to make the differences in different regimes of dynamics of the lattice better visible, running averaging with a block size of 100 points has been applied on both data sets presented in this plot.

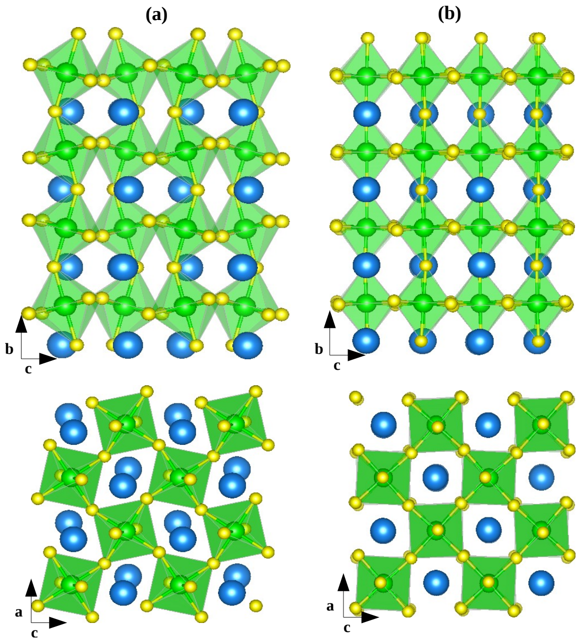

FIG. S10. Two views of structures obtained upon averaging of atomic positions and lattice vectors components over all configurations generated by NPT MD simulations of the DP/C phase at (a) 300 K and (b) 1500 K.

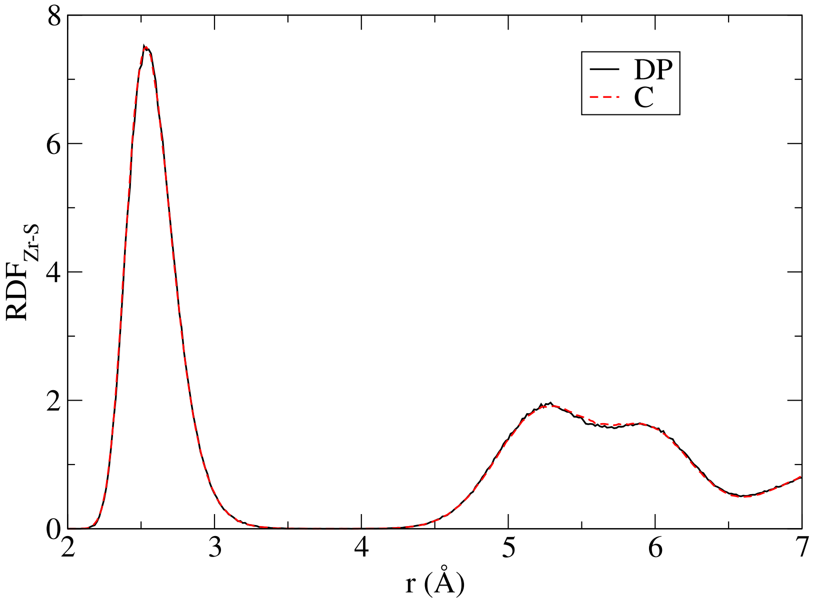

FIG. S11. Comparison of radial distribution functions for the Zr-S pairs (RDF$_{Zr-S}$) in the DP and (quasi-)C phases of $SrZrS_3$ at 1200 K.

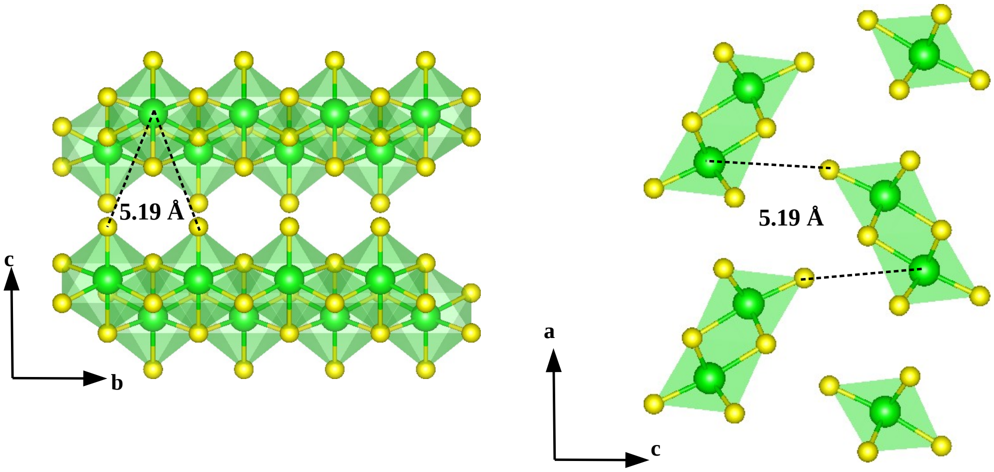

FIG. S12. Inter-chain non-bonding Zr-S distance in the NL phase that corresponds to the RDF peak located at r=5.2 Å in Fig. 5 of the main text.

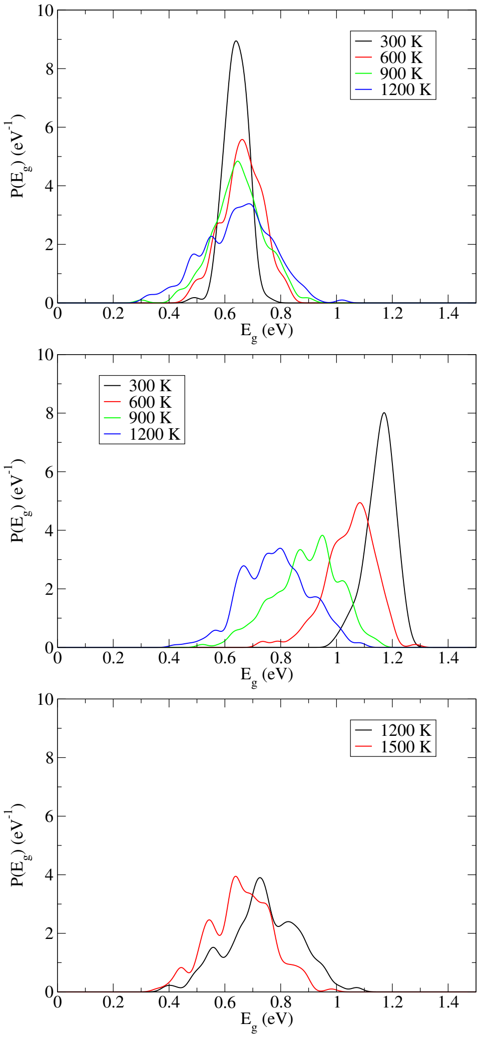

FIG. S13. Probability distribution functions of PBE band gap ($P(E_g)$) for the NL (top), DP (middle) and C (bottom) phases of SrZrS$_3$ at various temperatures.

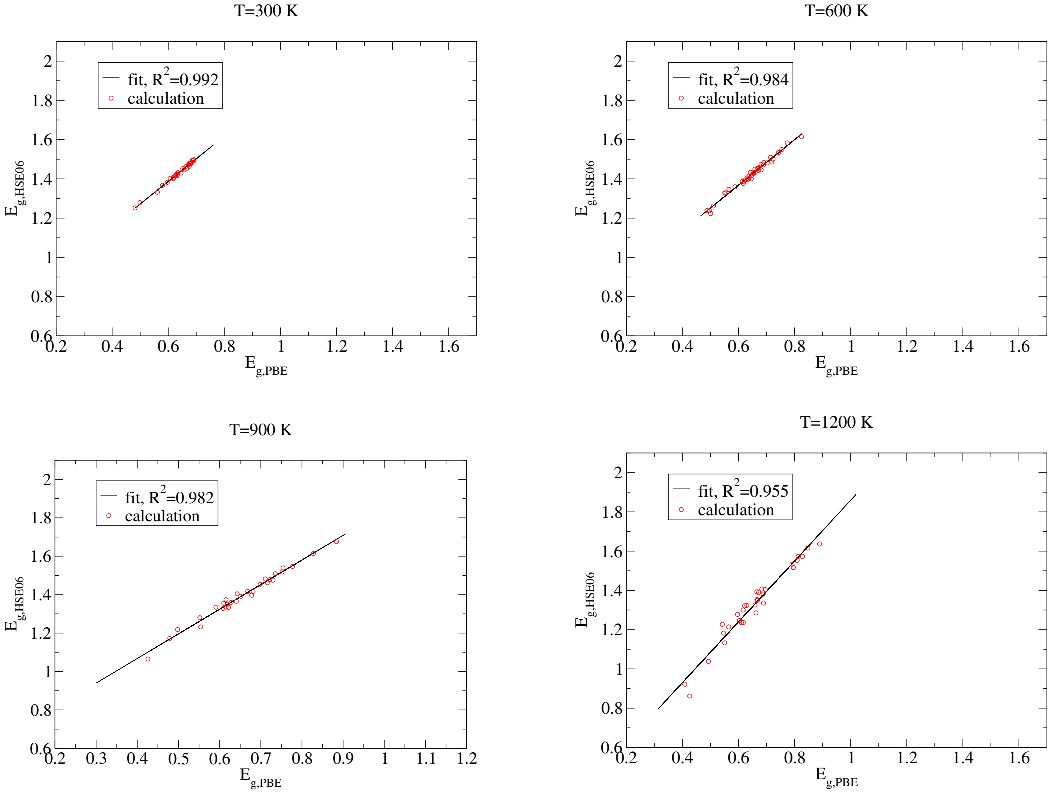

FIG. S14. Correlation between the HSE06 ($E_{g,HSE06}$) and the PBE ($E_{g,PBE}$) band gap values for configurations generated by MD for the NL phase of SrZrS$_3$ at various temperatures. Coefficients of determination (R^2) for the linear fit are also shown.

TABLE S5. Band gap values calculated as ensemble averages using PBE and HSE06 functional for the NL, DP and C phases of SrZrS$_3$ at different temperatures. The estimated statistical error is within ∼0.02 eV in all cases.

T (K)	PBE (eV)	HSE06 (eV)
	NL phase	
0	0.610	1.404
300	0.640	1.433
600	0.659	1.432
900	0.646	1.381
1200	0.644	1.306
	DP phase	
0	1.225	2.044
300	1.150	1.948
600	1.042	1.808
900	0.886	1.603
1200	0.778	1.420
	C phase	
0	0.473	1.205
1200	0.732	1.346
1500	0.655	1.155

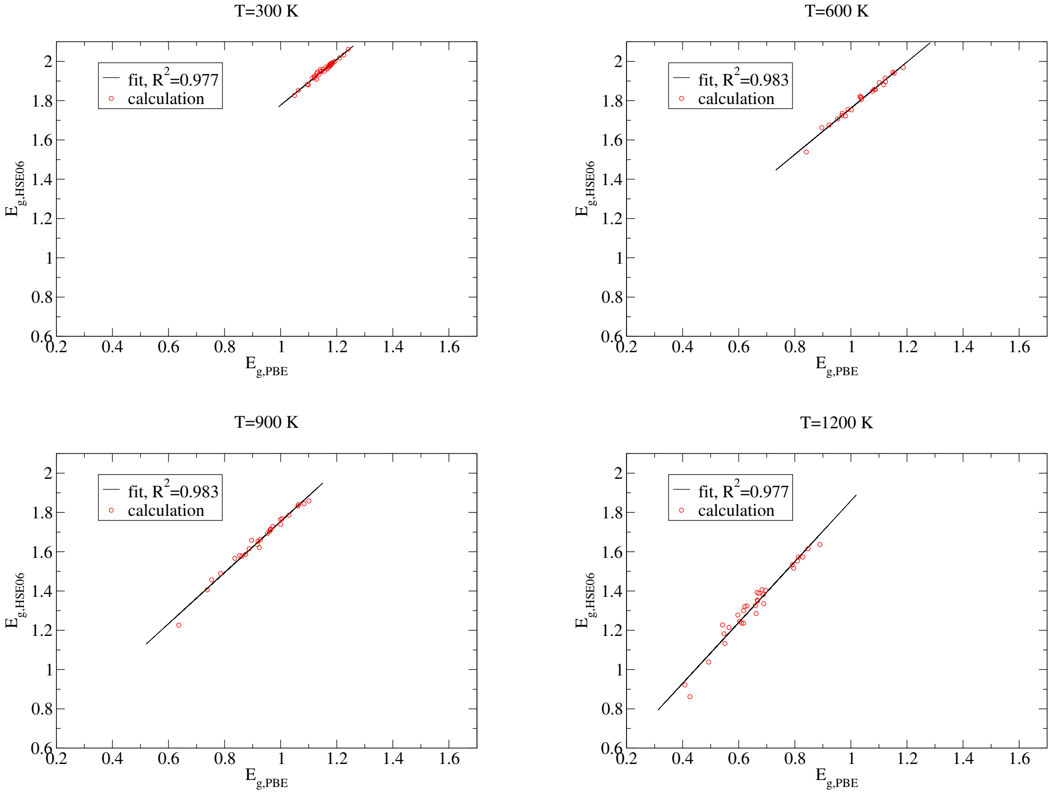

FIG. S15. Correlation between the HSE06 ($E_{g,HSE06}$) and the PBE ($E_{g,PBE}$) band gap values for configurations generated by MD for the DP phase of SrZrS$_3$ at various temperatures. Coefficients of determination (R^2) for the linear fit are also shown.

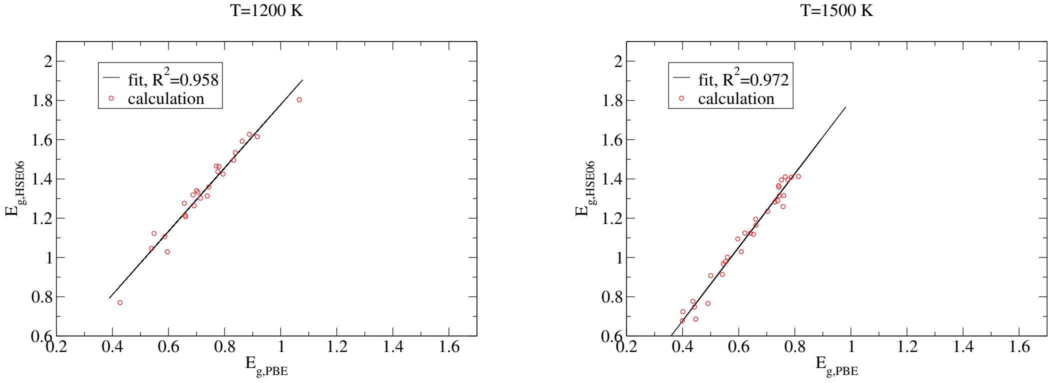

FIG. S16. Correlation between the HSE06 ($E_{g,HSE06}$) and the PBE ($E_{g,PBE}$) band gap values for configurations generated by MD for the (quasi-)C phase of SrZrS$_3$ at various temperatures. Coefficients of determination (R^2) for the linear fit are also shown.

SV. EXPLORATORY STATIC CALCULATIONS FOR SELECTED CHALCOGENIDE PEROVSKITES

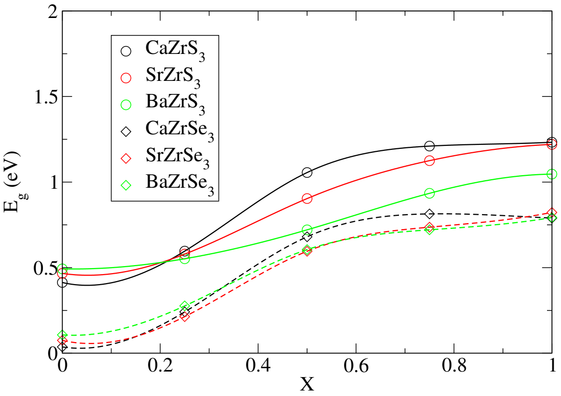

FIG. S17. Zero temperature variation of PBE band gap (E_g) of $CaZrS_3$, $SrZrS_3$, $BaZrS_3$, $CaZrSe_3$, $SrZrSe_3$, and $BaZrSe_3$ with distortion (X) of a cubic structure defined so that X=0 and X=1 correspond to relaxed C and DP phases, respectively.

TABLE S6. Relative energy (ΔE), cell volume per formula unit (V_1), bulk modulus (B_0) and band gap (E_g) for relaxed NL, DP and C phases of selected chalcogenide perovskites. The relative energy is referenced to the most stable phase of each compound. All results have been obtained using the PBE functional.

compound	phase	ΔE (eV/f.u.)	V_1 (Å/f.u.)	B_0 (GPa)	E_g (eV)
$CaZrS_3$	NL	0.043	110.4	52.1	0.71
	DP	0.000	112.3	68.0	1.23
	C	1.430	123.6	74.2	0.41
$SrZrS_3$	NL	0.000	116.4	50.3	0.61
	DP	0.002	119.6	69.7	1.23
	C	0.706	125.6	73.3	0.47
$BaZrS_3$	NL	0.027	124.7	60.1	0.78
	DP	0.000	127.1	70.6	1.03
	C	0.168	128.7	72.4	0.49
$CaZrSe_3$	NL	0.036	126.7	46.2	0.33
	DP	0.000	128.0	55.0	0.79
	C	1.594	142.5	60.3	0.04
$SrZrSe_3$	NL	0.000	132.2	40.8	0.23
	DP	0.038	136.2	55.5	0.83
	C	0.929	144.7	60.0	0.08
$BaZrSe_3$	NL	0.000	141.5	39.6	0.43
	DP	0.066	144.7	57.3	0.71
	C	0.371	147.6	59.3	0.11

[S1] R. Jinnouchi, F. Karsai and G. Kresse, Phys. Rev. B, 2019, **100**, 014105.
[S2] R. Jinnouchi, F. Karsai, C. Verdi, R. Asahi and G. Kresse, J. Chem. Phys., 2020, **152**, 234102.
[S3] C.-S. Lee, K. Kleinke and H. Kleinke, Solid State Sci., 2005, **7**, 1049–1054.
[S4] N. W. Ashcroft and N. D. Mermin, Solid State Physics, Harcourt College Publisher, 1976, pp. 151–174.
[S5] C. Kittel, Introduction to Solid State Physics, Wiley, 2004, pp. 223–254.