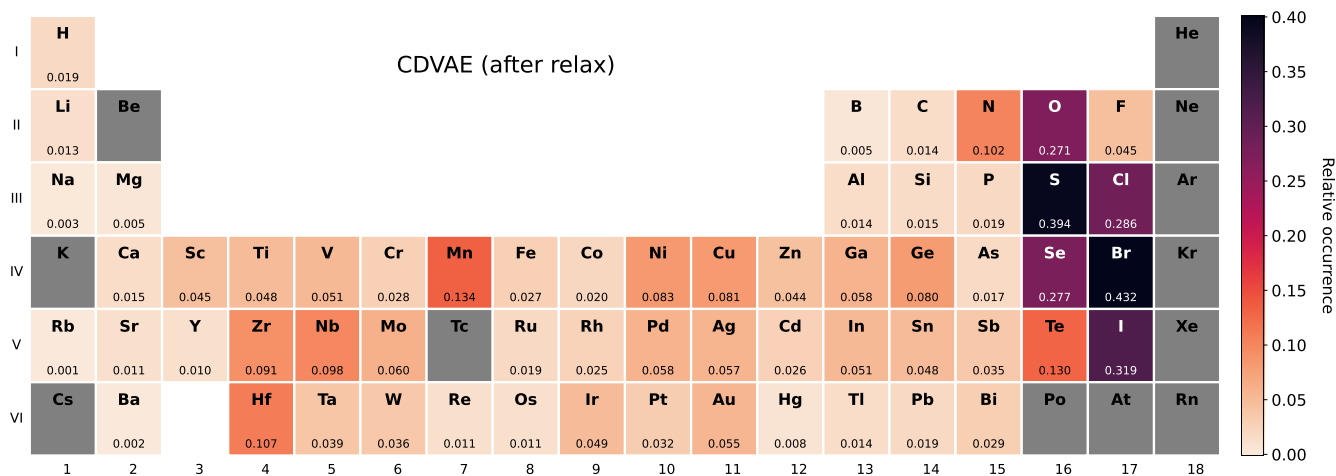


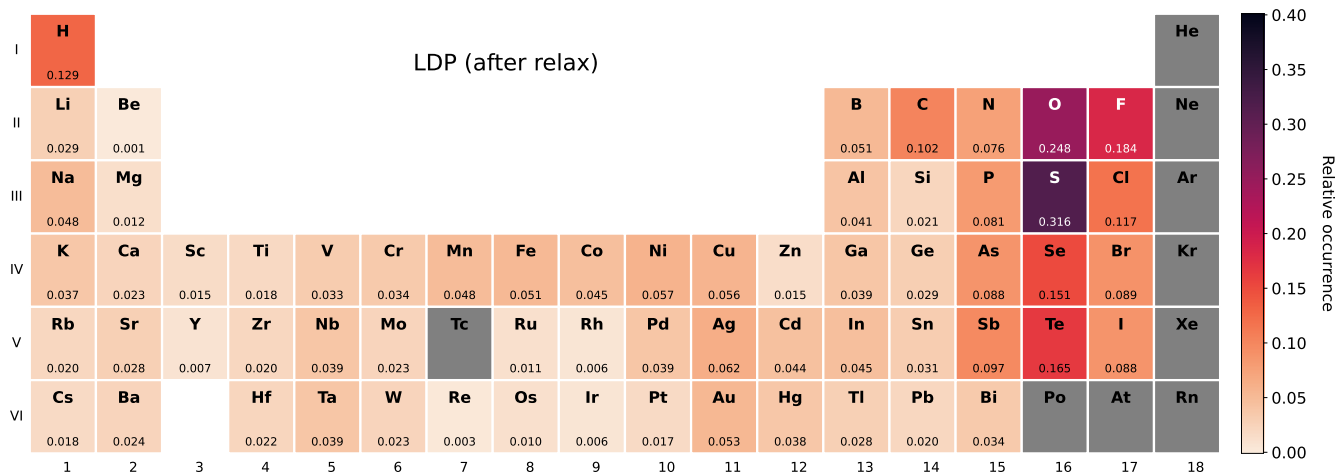
Supplementary Information: Data-driven discovery of 2D materials by deep generative models

Peder Lyngby and Kristian Sommer Thygesen
*Computational Atomic-scale Materials Design (CAMD), Department of Physics,
 Technical University of Denmark, 2800 Kgs. Lyngby Denmark*

I. SUPPLEMENTARY FIGURES



(a)



(b)

Supplementary Figure 1: Heat map of the relative occurrence of each element in the CDVAE and LDP generated and relaxed data set. The bottom number is the relative occurrence.

II. SUPPLEMENTARY TABLES

Supplementary Table 1: New occupied Wyckoff positions in the CDVAE dataset compared to the C2DB training set

Space group number	New occupied Wyckoff positions (CDVAE)
2	aci, bfgi, cdi, fghi, fgi, fhi
3	bce, bcd, acde, cde, abcde, abcd, acd, ade, abce, ab, abc, bcde
5	c
6	a, b
10	admn, hmn, chmn, bm, bdmn, cfmno, abjmno, ghkmno, befh, fgm, deimno, befo, ehi
12	bij, abeij, dfi, dij, beij, cdfij, abij, abi, cdi, cfij, bgij, dhij, cij
21	cdhl, cdk
25	ah, bcd, dg, bfi, adh, acd, ach, ad, abcdg, bdg, abd, abcdgh, ab, cd, abcdh, bc, bh
35	abe, abcde, abcd, abce, abc
38	adf, bef, af, b
47	cgs, abet, cdgt, aes, aet, ars, dghq, chs, cgt, bfq, dhqrt, ch, efq, bfr, dhr, aeqs
51	fk, ef
65	abim, cdjr, fh, abn, cdjo, cdfo, il, abio, abin, abgn
67	ag
99	abcd, ab
111	aden, adn, bcn
115	ceg, cdg, beg, abg
123	bdeh, bdg, bdi, beh, bdh
149	bdf, dfgl
156	abcd
187	cgi, ce, bdg, bd, ac, egjn
191	h, bde, bgh

Only the space groups where CDVAE has one or more different occupied Wyckoff position compared to the training set are reported.

Supplementary Table 2: New occupied Wyckoff positions in the LDP dataset compared to the C2DB seed set

Space group number	New occupied Wyckoff positions (LDP)
2	di, fgi, cei, hi, ci, ghi, aei, adi
5	c
6	b
10	aimo, chmn, gimo, cflmno, ekno, fgmn, eimno
11	bef
12	cdhi, fij, abgi, aij, gi, cij
13	aeg, bef, ef
25	bdg, aceg, abeg
28	bcd
47	amsu, chr
51	ef, dfk, cdeik
53	eh, dg, bgh
55	eg
59	abfg, abcef, abdef, ae, abg
65	cfkl, dfkl, cdjo, dlm, abio, dl
67	dg, agmn, g
113	ace
123	bgms, d, bi, de
129	bcj
143	abd, acd, bcd
149	dfgl
156	abcd
164	cdi, cdfi
174	ikl, bfk, ijl
187	dghi, cgi, cegi, bfgm, dgi
191	dgm

Only the space groups where the LDP dataset has one or more different occupied Wyckoff position compared to the seed set are reported.

III. ACKNOWLEDGEMENTS

We acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program Grant No. 773122 (LIMA) and Grant agreement No. 951786 (NOMAD CoE). K. S. T. is a Villum Investigator supported by VILLUM FONDEN (grant no. 37789).