Supporting Information

Exploration of elastic moduli of molecular crystals via database screening by pretrained neural network potential

Takuya Taniguchi*[1]

[1] Center for Data Science, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan

* Correspondence to takuya.taniguchi@aoni.waseda.jp

Contents

Supporting Tables

Table S1. Cell volume change of the optimized structure from the experimental structure.

Compound	Literature		This work		
	Refcode	$\Delta V_{S\text{-}HF\text{-}3c}$ (%)	Refcode	ΔV_{PFP} (%)	ΔV_{CHGNet} (%)
1,2,4,5-tetrabromobenzene	TETBBZ01	-3.1	TETBBZ03*	1.62	135.89
1,3-dinitrobenzene	DNBENZ10	-4.2	DNBENZ10	-1.22	8.30
1,3,5-trichlorobenzene	TCHLBZ	-0.2	TCHLBZ	2.70	49.45
3-nitroaniline	MNIANL05	-5.8	MNIANL05	-0.21	5.35
acenaphthene	ACENAP03	-9.3	ACENAP03	-0.97	24.32
anthracene	ANTCEN14	-10.2	ANTCEN14	2.81	18.98
aspirin	ACSALA07	-6.5	ACSALA07	-2.71	9.53
benzene	BENZEN	-12.8	BENZEN	-1.85	26.84
benzil	BENZIL02	-8.9	BENZIL02	-0.77	12.12
benzophenone	BPHENO10	-11.2	BPHENO17*	-0.94	16.20
beta-HMX	OCHTET15	3.9	OCHTET15	1.74	8.61
carbamazepine	CBMZPN01	-8.1	CBMZPN01	1.45	24.25
citric acid	CITRAC10	-4.9	CITRAC10	-1.22	6.33
citric acid monohydrate	CITARC02	-2.7	CITARC02	-2.54	3.45
DC apohost	QQQESP01	-6.5	QQQESP01	-1.33	- **
durene	DURENE01	-11.6	DURENE01	-6.15	106.28
epsilon CL-20	PUBMUU02	2.4	PUBMUU02	5.88	10.95
hippuric acid	HIPPAC	-5.0	HIPPAC	0.56	10.63
HMT	HXMTAM08	-8.2	HXMTAM03*	-0.90	3.65
iodic acid	ICSD-66643	-13.2	ICSD-66643	-0.77	15.28
maleic acid	MALIAC11	-5.2	MALIAC11	-0.81	11.52
malonic acid	MALNAC06	-3.5	MALNAC06	2.22	10.37
melamine	MELAMI02	-6.2	MELAMI02	0.069	10.61
naphthalene	NAPHTA36	-12.2	NAPHTA36	4.19	20.79
oxalic acid dihydrate	OXACDH04	-1.7	OXACDH04	-6.12	6.40
pentaerythritol	PERYTO03	-3.7	PERYTO03	-1.93	32.53
PETN	PERYTN10	-0.5	PERYTN11*	3.08	11.48
phenothiazine	PHESAZ07	-6.0	PHESAZ07	-1.25	17.83
phthalic acid	PHTHAC02	-7.9	PHTHAC02	-0.89	11.53
POM	MNPYDO	-1.9	MNPYDO	2.94	12.85
pyrazine	PYRAZI	-9.3	PYRAZI	5.74	7.65
RDX	CTMTNA12	0.9	CTMTNA12	2.66	11.02
Alpha-resorcinol	RESORA19	-6.6	RESORA19	-0.27	3.63
Beta-resorcinol	RESORA24	-7.5	RESORA24	2.72	13.76
rubrene	QQQCIG11	-8.3	QQQCIG11	-0.98	21.70
succinic acid	SUCACB10	-7.2	SUCACB12*	-1.18	12.35
succinimide	SUCCIN01	-6.0	SUCCIN01	-2.30	13.78
sulfamic acid	ICSD-94	-10.0	ICSD-94	-2.85	5.33
sulfur	ICSD-200454	-1.5	ICSD-200454	-0.13	49.12
tartaric acid	TARTAC01	-2.8	TARTAC01	-1.88	11.63
taurine	TAURIN03	-8.5	TAURIN03	-1.50	11.73
TCNE	TCYETY01	-9.6	TCYETY01	-2.75	14.58
thiourea	THIOUR01	6.6	THIOUR01	-6.89	38.96
tolane	DPHACT02	-11.1	DPHACT03*	0.76	16.06
urea	UREAXX22	-7.1	UREAXX22	-2.50	1.14

	Mean Error	-5.7		-0.33	19.88
	MAE	6.5		2.15	19.88

* The original CIF file did not have the hydrogen atom positions assigned, so they were changed to this CIF file.

** QQQESP01, which has the largest volume in the dataset, was not calculated using CHGNet due to the shortage of computation memory.

Table S2. Cell volume change of X23 dataset after structure optimization by NNPs.

Compound	CSD identifier	ΔV_{PFP} (%)	ΔV_{CHGNet} (%)
vdW-bonded systems			
1,4-Cyclohexanedione	CYHEXO	-0.51	14.78
Adamantane	ADAMAN08	1.39	197.83
Anthracene	ANTCEN09	7.07	23.82
Benzene	BENZEN01	2.05	31.90
Carbon dioxide	SACBAA02	12.04	22.38
Hexamine	QAMCAM	-1.36	7.08
Naphthalene	NAPHTA23	10.31	31.04
Pyrazine	PYRAZI01	9.11	11.15
Pyrazole	PYRZOL05	12.78	42.10
s-Triazine	TRIZIN	-9.50	-4.80
s-Trioxane	TROXAN11	0.59	15.63
H-bonded systems			
Acetic acid	ACETAC07	-0.52	9.09
Ammonia	-	-9.64	5.43
Cyanamide	CYANAM01	-0.70	22.10
Cytosine	CYTSIN01	3.01	19.65
Ethyl carbamate	ECARBM01	-0.26	7.68
Formamide	FORMAM02	1.28	18.76
Imidazole	IMAZOL04	6.63	36.22
Oxalic acid α	OXALAC03	-0.43	20.92
Oxalic acid β	OXALAC04	1.81	20.16
Succinic acid	SUCACB10	-6.22	-1.29
Uracil	URACIL	-0.06	11.34
Urea	UREAXX02	0.86	4.18
	MAE (total)	4.27	25.19
	MAE (vdW bonding)	6.06	36.59
	MAE (Hydrogen bonding)	2.62	14.74

Table S3. Validation data, whose paper has been published in 2023.

| CCDC number	ΔV_{PFP} (%)	ΔV_{CHGNet} (%)
2045405	3.49	21.53
2107786	-1.26	-*
2191502	-1.78	-*
2191503	0.40	13.89
2191504	-1.10	-*
2191505	-1.32	31.27
2191506	0.069	-*
2191507	-0.58	71.32
2192139	0.58	47.45
2206389	9.08	30.12
2208671	0.59	39.43
2208672	0.43	21.93
2216257	-2.43	17.65
2216734	2.80	24.93
2216735	0.13	129.92
2216737	2.83	25.51
2219715	2.25	12.90
2219716	5.28	27.16
2242898	0.23	5.50
2242899	0.90	7.45
MAE	1.88	33.00

* The crystal structure was not optimized using CHGNet due to computation memory limit.

Table S4. Nanoindentation dataset of experimental Young's modulus and predicted values by PFP.

CSD refcode	Name	Indentation direction	E_{exp} (GPa)	E_{pred} (GPa)
AMBNAC06	pABA-α	[200]	7.17	3.054
AMBNAC04	pABA-β	[10−1]	11.38	7.503
NBZOAC03	pNBA-I	[002]	9.19	3.394
NBZOAC14	pNBA-II	[002]	5.88	2.251
SUTHAZ28	STZ-I	[100]	10.01	3.518
SUTHAZ24	STZ-II	[100]	20.44	20.878
SUTHAZ25	STZ-III	[100]	16.42	15.063
SUTHAZ26	STZ-IV	[10−1]	17.31	13.116
BINMEQ02	CCM-I	[001]	11.15	8.664
BINMEQ08	CCM-II	[100]	5.68	8.668
BINMEQ07	CCM-III	[001]	5.6	9.008
VAPMEH	FlA-I	[001]	8.36	4.55
VAPMEH01	FlA-II	[001]	6.17	8.658
FOGVIG04	FAM-A	[100]	22.6	11.264
FOGVIG04	FAM-A	[001]	20.1	11.206
FOGVIG05	FAM-B	[-101]	19.5	11.512
HIQQAB	FEB-Q	[011]*	2.5	5.352
HIQQAB02	FEB-H1	[010]*	4.5	9.96
DONTIJ	FEL-I	[100]	12.8	10.389
DONTIJ02	FEL-III	[10-1]	10.5	12.11
			MAE	4.28

* The indentation direction was not assigned in the literature. Thus, it is assumed that indentation was conducted on the major face based on Bravais-Friedel Donnay-Harker (BFDH) morphology.

Supporting Figures
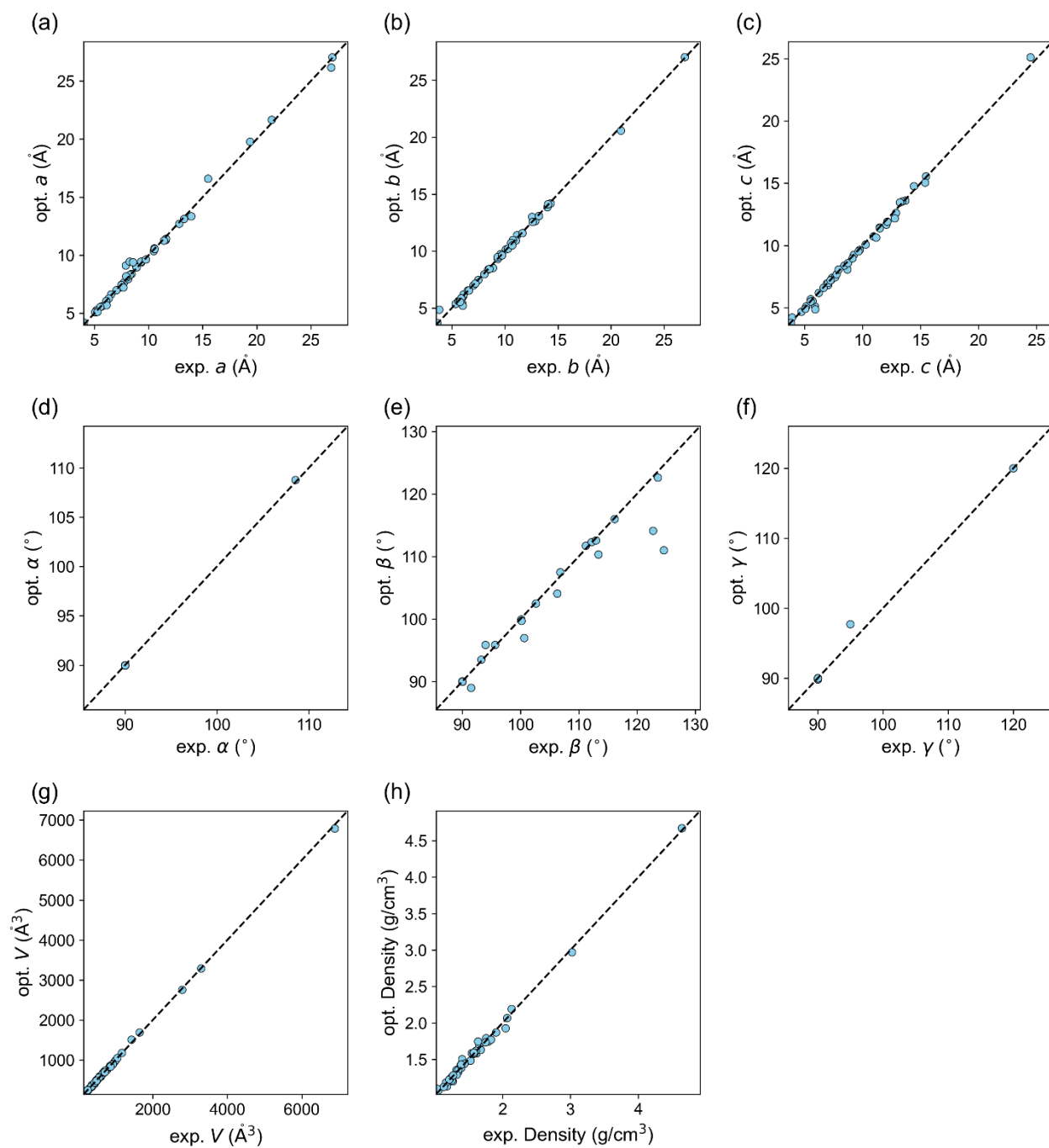
Figure S1. The lattice changes before and after the cell optimization using PFP. Elasticity dataset was used for evaluation.
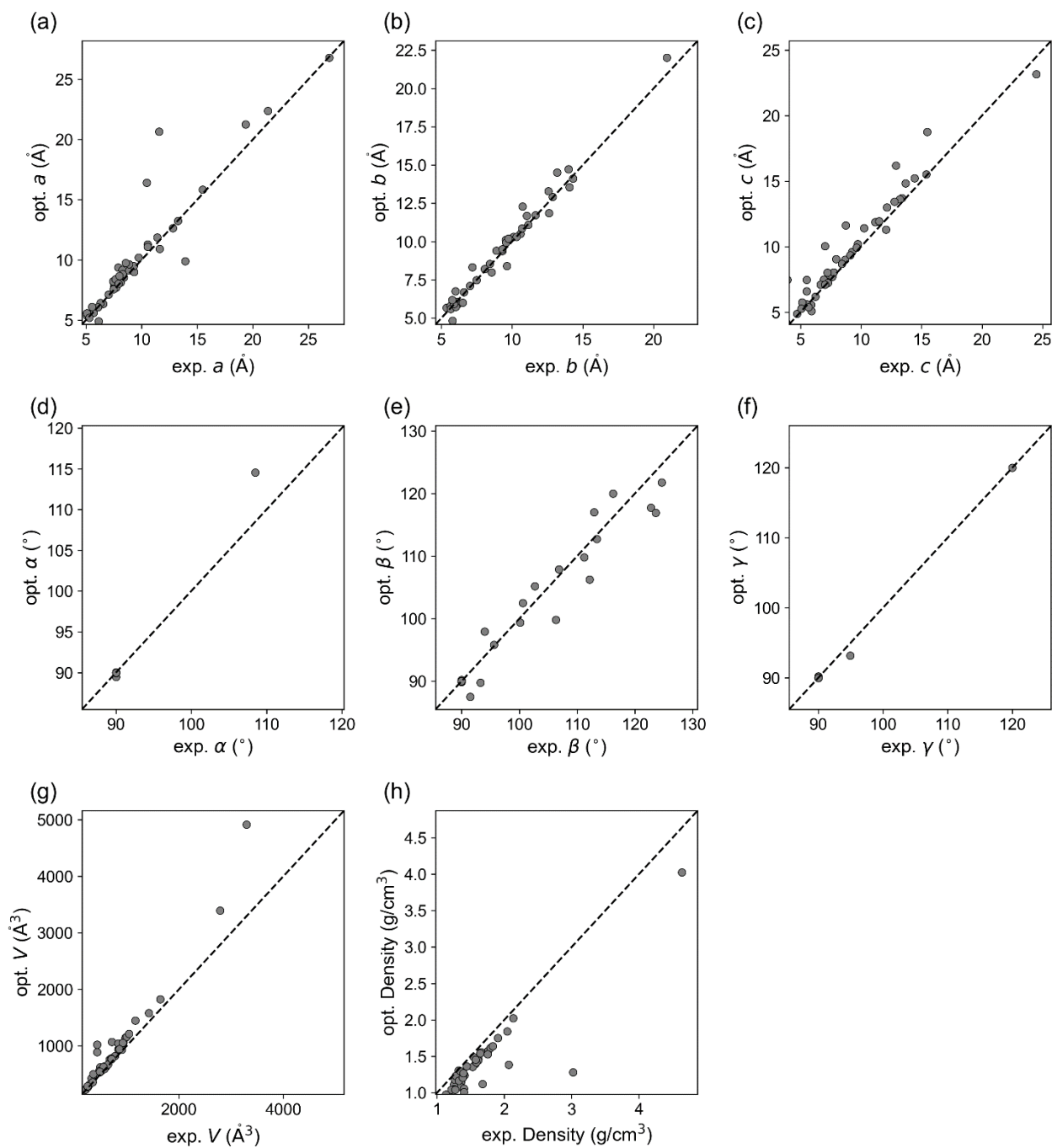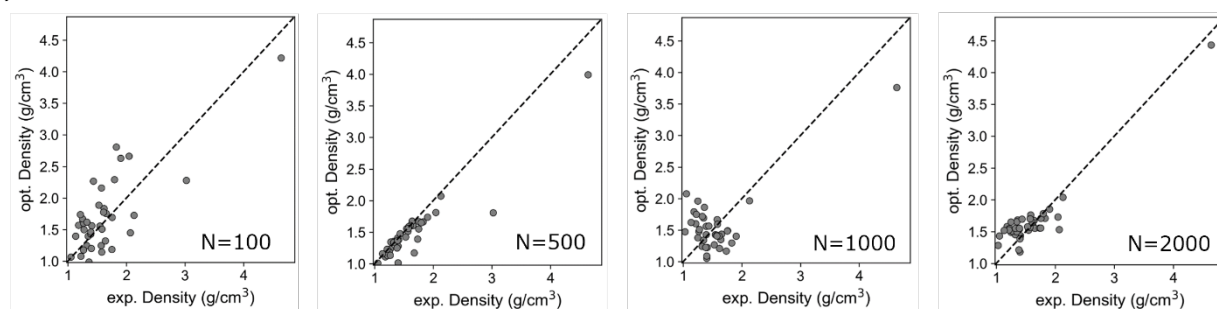
Figure S2. The lattice changes before and after the cell optimization using CHGNet. Elasticity dataset was used for evaluation. Note that the data of QQQESP01, which has the largest cell volume, was not processed due to the limitation of CPU memory.
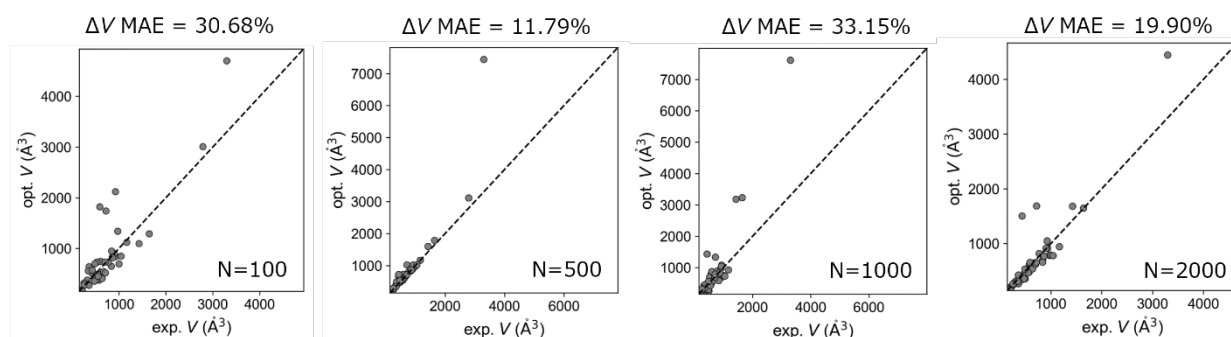
Figure S3. (a) Cell volume and (b) density before and after the cell optimization using finetuned CHGNet. Elasticity dataset was used for evaluation. Note that the data of QQQESP01, which has the largest cell volume, was not processed due to the limitation of CPU memory. The number of molecular crystals used for training was written in the bottom right in each scatter plot.
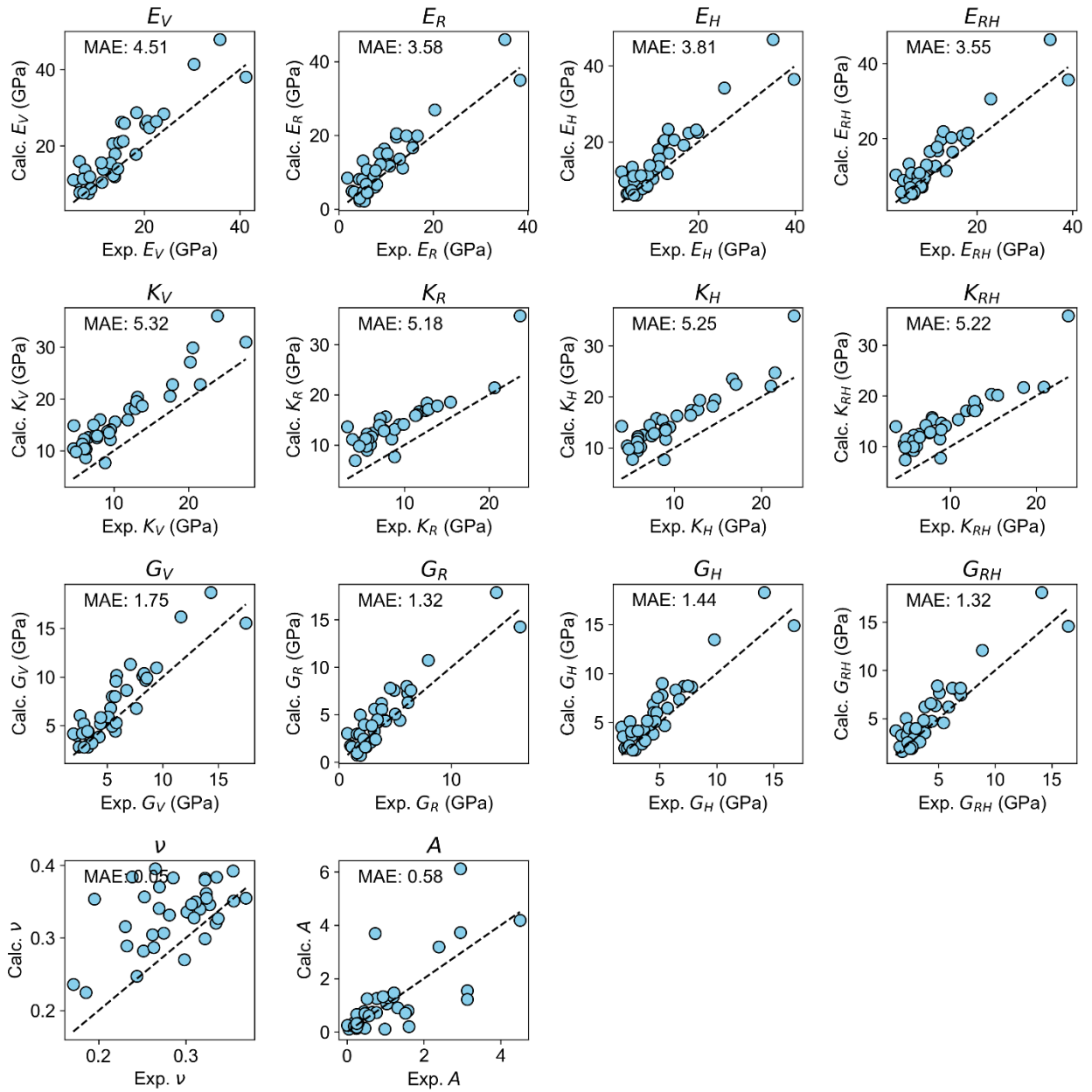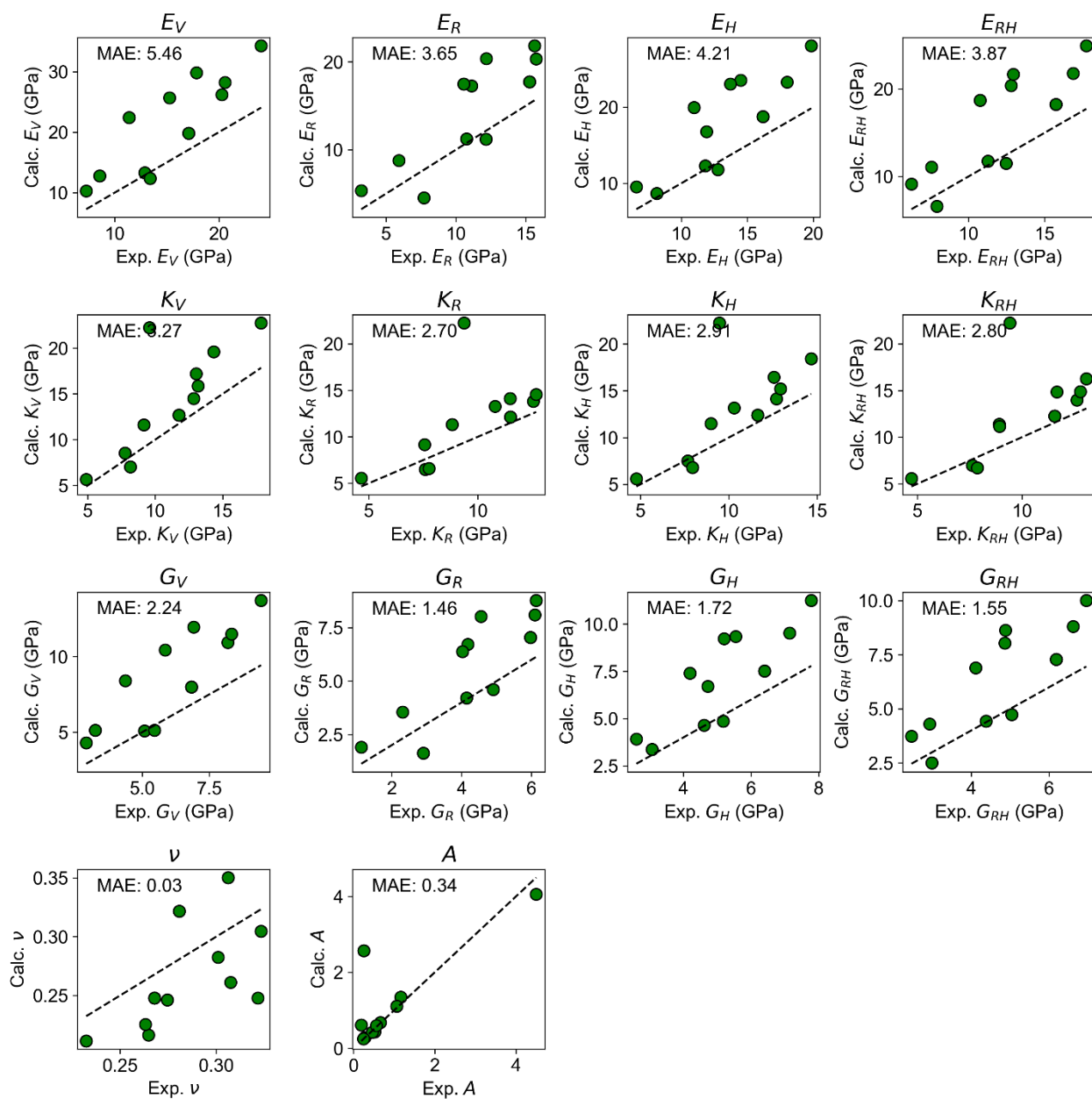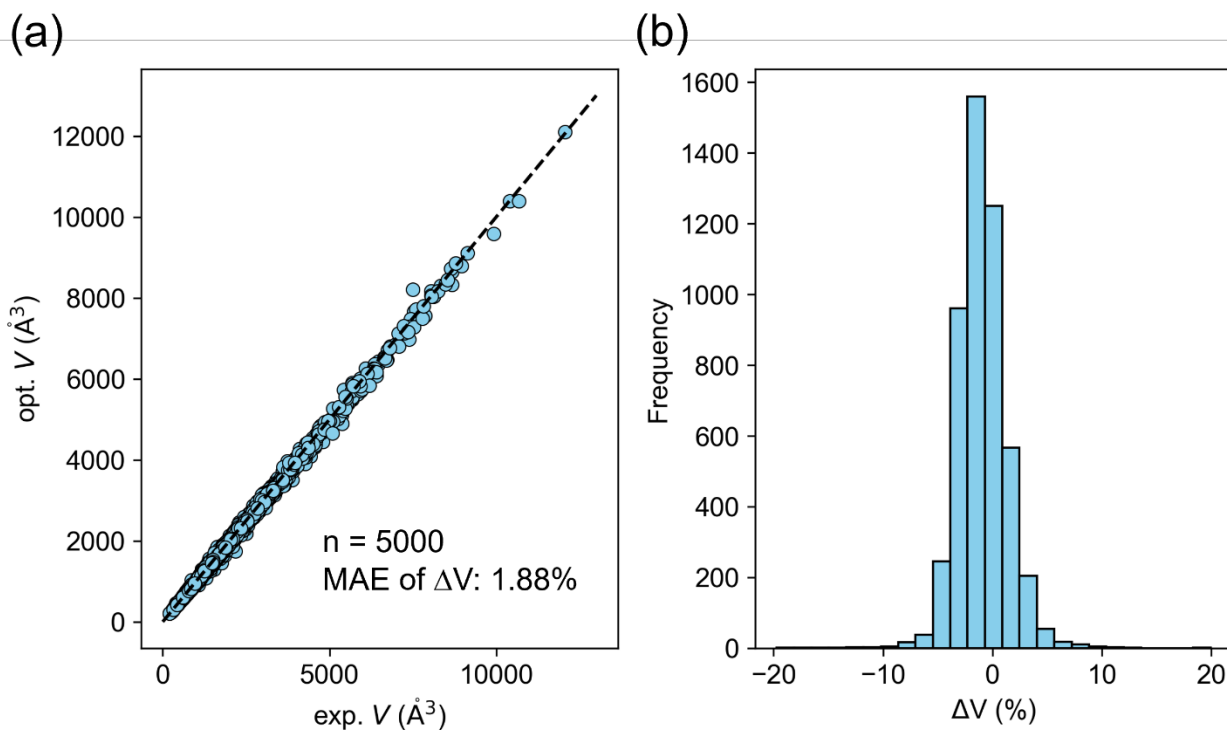
Figure S4. The observed-predicted plot of elastic properties calculated by PFP model after the structural relaxation.
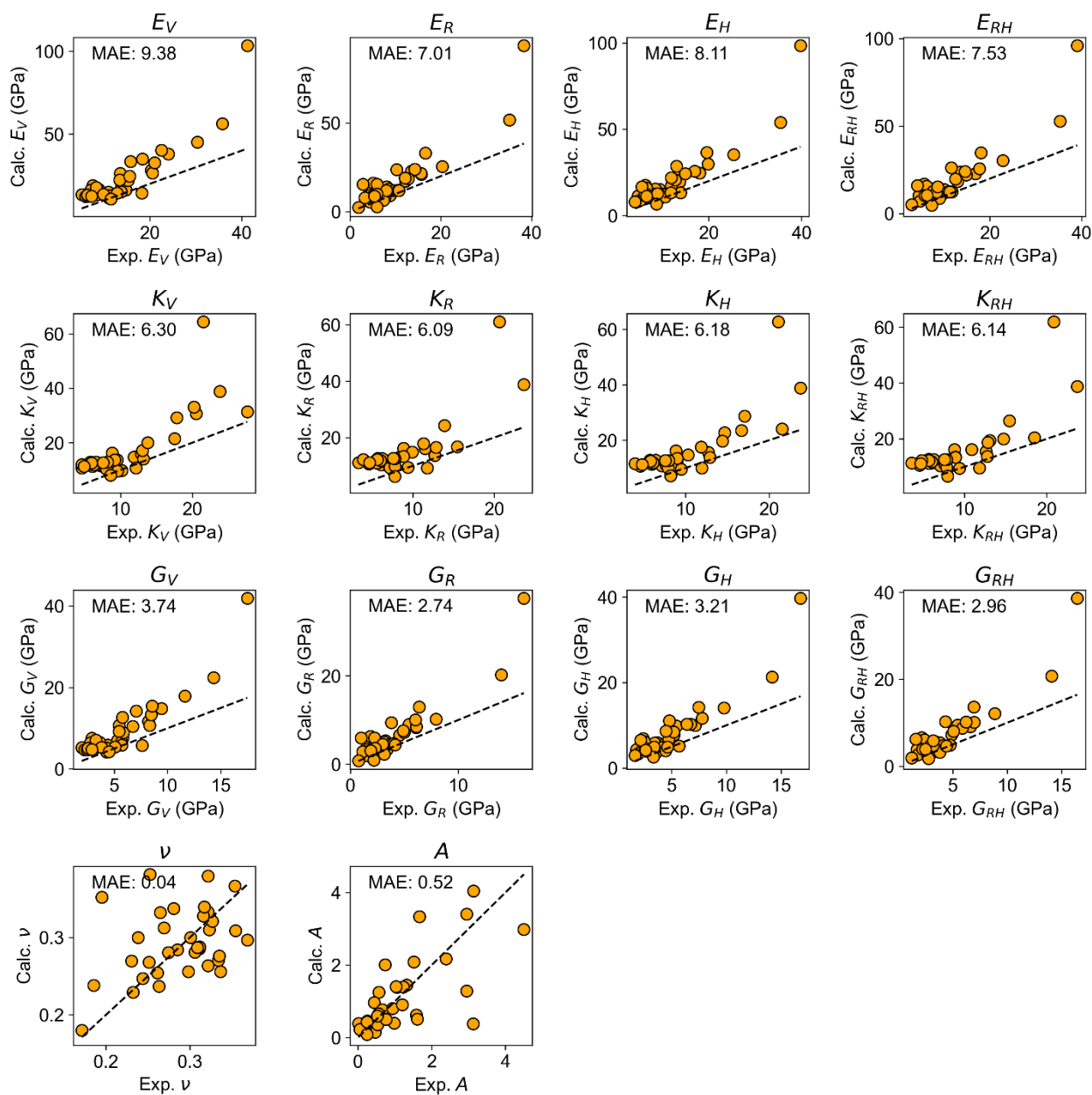
Figure S5. The observed-predicted plot of elastic properties calculated by S-HF-3c calculations summarized in the literature.[1]

Figure S6. The observed-predicted plot of elastic properties calculated by DFT-D calculations summarized in the literature.[1]

Figure S7. The deviation of relaxed cell volumes of 5000 structures from CSD. (a) The observed-predicted plot of the relaxed cell volumes calculated by PFP. (b) Histogram of the relative deviation of the optimized cell volumes from the experimental volumes.
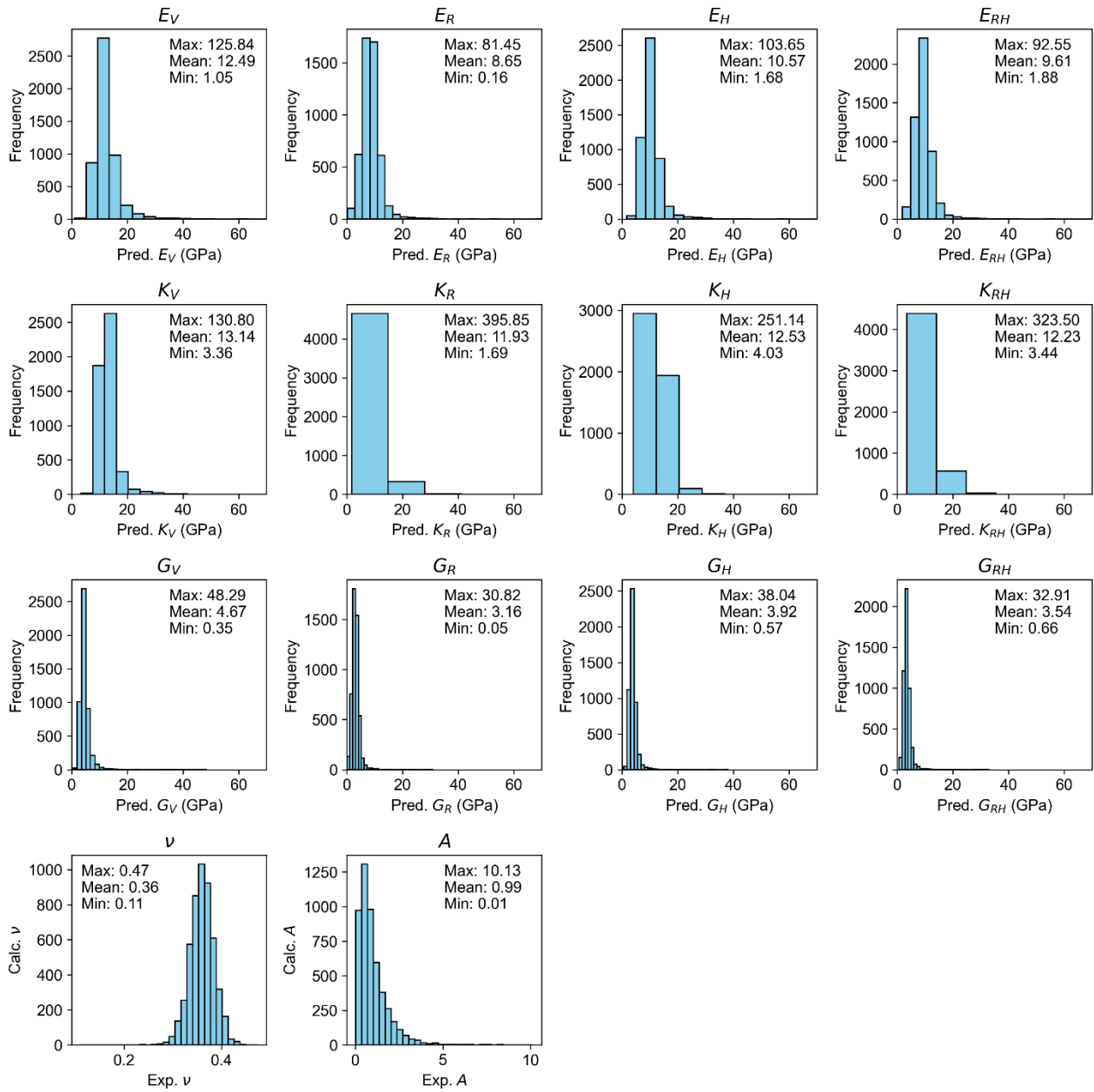
Figure S8. Histograms of elastic moduli obtained by large dataset inferred by PFP.