

This dataset was generated to train machine-learning interatomic potentials (MLIPs) for peptoid conformational energetics as part of **PEPPER (Peptoid Energies & Pairwise Property Evaluation Reference)**. The configurations were generated using a systematic conformational sampling and selection workflow was used for 70 peptoid residue identities derived from the STEPs peptoid force-field development effort. The PEPPER structures were represented in a capped dipeptoid context to retain the local backbone environment. Representative low-energy configurations were selected from replica-exchange molecular dynamics (REMD) trajectories by separating the ensemble into *cis* and *trans* backbone states, identifying basins in the conformational space, clustering side-chain rotamers within each basin, and selecting representative basin-rotamer configurations. The selected structures were subsequently relaxed with GFN2-xTB.

To provide coverage beyond low-energy conformational basins, additional higher-energy structures associated with the peptoid *cis/trans* transition region were included. The minima-only dataset (`MIN.extxyz`) contains **3,366 configurations**, while the transition-augmented dataset (`ALL.extxyz`) contains these structures together with **2,787 transition-region configurations**, for a total of **6,153 configurations**. A separate hydrogen-capped dataset (`HCAPPED.extxyz`) was also generated to examine the effect of removing the complete peptoid backbone environment.

For the selected configurations, reference electronic-structure energies and atomic forces were calculated with **ORCA 6.0.0** using the **ωB97M-D3(BJ)/def2-TZVPPD** level of theory, following the electronic-structure protocol used by SPICE. The resulting energies and forces constitute the reference labels used to train the custom MLIPs contained in this dataset.

Below we provide some examples of the dataset.

#####

MIN.extxyz: the **minima-only PEPPER training dataset**. It contains **3,366 representative peptoid conformations** selected from the low-energy conformational basins sampled in the REMD workflow. Here are the two first snapshots:

25

```
Properties=species:S:1:pos:R:3:REF_forces:R:3 REF_energy=-14570.338879004823 pbc="F F F"
C 15.76000110 26.70000115 27.88999914 -0.31479916 0.34516453 2.15561219
H 16.11000108 27.29000114 27.04000312 -0.24973101 -0.68384628 -0.30391878
H 16.17000210 27.26000013 28.74000216 -0.36869030 -0.13082900 -0.03495094
H 14.67000109 26.67000014 28.01000214 0.36157807 0.35030201 -0.44068879
C 16.21000107 25.24000212 27.88999914 1.91719175 2.78216958 -1.45393642
O 15.56000106 24.48000109 27.24000215 -0.86380700 -2.23650116 -0.35582621
N 17.52000010 25.04000113 28.25000015 1.66549912 -0.33029364 1.91279386
C 18.33000007 23.93000013 27.70000315 -1.20848064 -0.06442418 -0.55127373
H 18.59000008 24.22000114 26.67000014 0.48092620 -0.23808284 0.43178892
H 17.76000208 23.00000213 27.59000012 -0.21683232 0.38133195 0.18679065
H 19.18000007 23.67000213 28.33000216 -0.00183907 0.49890232 -0.17486947
C 18.28000108 26.03000311 29.13000116 -2.28712019 0.48662804 -0.90692134
H 18.23000108 26.96000115 28.54999912 0.50313676 -0.01549557 0.61942008
H 19.33000207 25.75000010 29.27000012 -0.50482725 -0.22177039 -0.11461521
C 17.68000006 26.37000111 30.57000215 2.20371385 -1.95049334 0.31137540
O 17.86000107 27.45000115 31.08000013 -0.51432962 1.11476230 -0.50725504
N 17.37000107 25.30000315 31.35000217 -0.18178537 -3.07956297 1.17185114
C 17.42000207 23.80000113 30.90000214 -0.55978837 1.76133529 0.24044469
H 18.41000008 23.46000313 30.58000217 -0.11814413 0.41660888 0.02636340
H 16.69000110 23.64000112 30.09000016 0.43560133 -0.04714468 0.54250657
H 17.15000009 23.22000110 31.79000116 0.01502544 0.18812548 -0.51821417
C 17.16000011 25.41000210 32.86000119 -0.93271279 0.60644812 -1.01775619
H 16.78000107 24.44000112 33.20999915 0.11282367 0.53804416 -0.20873334
H 16.40000209 26.18000014 33.06000117 0.14207655 -0.34229727 -0.21626356
H 17.99000207 25.70000111 33.51000215 0.48531449 -0.12908133 -0.79372372
```

25

```
Properties=species:S:1:pos:R:3:REF_forces:R:3 REF_energy=-14569.603270121786 pbc="F F F"
C 16.20000110 29.02000215 31.93000018 1.00534915 1.61983780 0.65976298
H 16.30000109 29.60000016 31.01000015 -0.52964452 -0.63285089 -0.23178093
H 16.54000106 27.98000113 31.82000218 -0.32277276 0.36938885 -0.18600794
H 16.95000110 29.55000112 32.54000117 -0.44664096 -0.41091838 0.32275406
C 14.74000107 29.03000117 32.45000114 5.78550262 -1.78511359 3.83121659
O 13.93000004 28.82000216 31.65000215 -4.52057781 0.13189002 -2.48047348
N 14.64000108 28.78000113 33.86000117 -2.99104871 0.45728155 -1.43717641
C 13.30000008 28.19000214 34.26000215 0.31306442 0.60042300 0.86331858
H 12.61000104 27.99000014 33.43000019 1.08970309 -0.47351661 -0.16795245
H 12.61000104 28.79000115 34.88000115 0.92675551 0.65473689 0.02750752
H 13.52000106 27.34000214 34.92000218 -0.24271419 -0.22245595 -0.67964061
C 15.78000109 28.62000313 34.77000415 -2.22903003 -0.67159920 -0.36190156
H 16.44000106 28.02000015 34.13000115 1.01093525 0.67102186 1.07889372
H 15.37000109 27.95000112 35.54000118 0.55279406 0.10906436 -0.13026456
C 16.24000208 29.95000115 35.42000215 1.56375537 -0.62973444 -0.53985148
O 15.97000105 30.99000217 34.87000319 0.04610558 0.22698424 0.19757504
```

N	17.15000009	29.85000216	36.48000316	-0.50368641	0.15606220	-0.64910817
C	17.05000110	28.70000112	37.43000421	0.12846830	1.73272767	-0.68779974
H	17.44000110	28.97000115	38.42000217	0.53110629	-0.02023079	-0.59236570
H	17.61000108	27.90000213	36.93000017	-0.74767954	-0.48095118	0.91702735
H	15.98000007	28.64000116	37.66000421	0.28113960	-1.12425858	-0.45071622
C	18.05000109	30.98000014	36.67000217	-0.02317329	0.23780817	2.66366524
H	18.40000011	30.98000014	37.71000320	-0.44568891	-0.41106069	0.29565090
H	17.48000007	31.92000217	36.69000216	0.58125565	-0.20537608	-1.04259671
H	19.06000109	30.95000113	36.24000219	-0.81327774	0.10083976	-1.21973603

#####

ALL.extxyz is the **transition-augmented PEPPER training dataset**. It contains all **3,366 configurations from MIN.extxyz**, together with **2,787 additional higher-energy configurations sampling the cis/trans transition region**, giving **6,153 structures total**.

25

```
Properties=species:S:1:pos:R:3:REF_forces:R:3 REF_energy=-14571.00918306878 pbc="F F F"
C   -1.47346161    1.09686619    0.45485062   -0.12744272    0.07629817   -0.13021392
H   -2.48231272    0.76657722    0.69024150   -0.01492809    0.00467262    0.02101757
H   -1.55772869    2.05279466   -0.05866638   -0.02995883    0.00720542   -0.00087249
H   -0.89832150    1.20890248    1.36961196   -0.08570821    0.03967530   -0.04855171
C   -0.84848105    0.07062796   -0.46695470    0.39570225    1.08293229   -0.02536101
O   -1.35454179   -0.29739172   -1.49212643   -0.34931442   -0.50736621    0.28807231
N    0.42355291   -0.33622978   -0.00863241    0.10616013   -0.98496386   -0.46917642
C    1.45522156    0.67057881   -0.19349169    0.01086558    0.34334275    0.45690855
H    1.03693031    1.65789072   -0.00717787    0.04246899   -0.02017914    0.02118045
H    1.85706921    0.65401117   -1.20908388   -0.01680020    0.07734464    0.04843879
H    2.26341608    0.51093642    0.52144646   -0.00832825    0.06733423   -0.12378143
C    0.81095698   -1.68707071   -0.43637831   -0.20304468   -0.24728441    0.06759790
H   -0.11312819   -2.22463062   -0.66834082    0.32792586    0.07444967    0.03658580
H    1.29358438   -2.18127201    0.41225884   -0.06579380    0.05964093   -0.18138063
C    1.67515491   -1.79189469   -1.69822545   -0.10975446   -0.00750519    0.33712887
O    1.17492729   -1.93268906   -2.79559188   -0.17614327    0.02883978   -0.17074152
N    3.02700043   -1.71557072   -1.56459175    0.26036221   -0.01114247   -0.20371008
C    3.79559287   -1.79371098   -0.34600188    0.04988052   -0.08295426    0.21284906
H    3.22353198   -1.46203059    0.51505543    0.07510198   -0.07340167   -0.08947410
H    4.67105370   -1.14787020   -0.44071598   -0.10112484   -0.05125397    0.09163121
H    4.14353900   -2.81850244   -0.16994727   -0.00314896    0.12714411    0.07425286
C    3.77961897   -1.89006325   -2.79458824   -0.09006643    0.02725933   -0.29748122
H    3.42819376   -1.17156866   -3.53677139    0.09430820   -0.15514166    0.03104178
H    3.63167795   -2.89334847   -3.20423940    0.05828294    0.12251793    0.00635706
H    4.83689355   -1.72773486   -2.60329120   -0.03950152    0.00253567    0.04768232
```

25

```
Properties=species:S:1:pos:R:3:REF_forces:R:3 REF_energy=-14571.00918306878 pbc="F F F"
C   -1.47346161    1.09686619    0.45485062   -0.12744272    0.07629817   -0.13021392
H   -2.48231272    0.76657722    0.69024150   -0.01492809    0.00467262    0.02101757
H   -1.55772869    2.05279466   -0.05866638   -0.02995883    0.00720542   -0.00087249
H   -0.89832150    1.20890248    1.36961196   -0.08570821    0.03967530   -0.04855171
C   -0.84848105    0.07062796   -0.46695470    0.39570225    1.08293229   -0.02536101
O   -1.35454179   -0.29739172   -1.49212643   -0.34931442   -0.50736621    0.28807231
N    0.42355291   -0.33622978   -0.00863241    0.10616013   -0.98496386   -0.46917642
C    1.45522156    0.67057881   -0.19349169    0.01086558    0.34334275    0.45690855
H    1.03693031    1.65789072   -0.00717787    0.04246899   -0.02017914    0.02118045
H    1.85706921    0.65401117   -1.20908388   -0.01680020    0.07734464    0.04843879
H    2.26341608    0.51093642    0.52144646   -0.00832825    0.06733423   -0.12378143
C    0.81095698   -1.68707071   -0.43637831   -0.20304468   -0.24728441    0.06759790
H   -0.11312819   -2.22463062   -0.66834082    0.32792586    0.07444967    0.03658580
H    1.29358438   -2.18127201    0.41225884   -0.06579380    0.05964093   -0.18138063
C    1.67515491   -1.79189469   -1.69822545   -0.10975446   -0.00750519    0.33712887
O    1.17492729   -1.93268906   -2.79559188   -0.17614327    0.02883978   -0.17074152
```

N	3.02700043	-1.71557072	-1.56459175	0.26036221	-0.01114247	-0.20371008
C	3.79559287	-1.79371098	-0.34600188	0.04988052	-0.08295426	0.21284906
H	3.22353198	-1.46203059	0.51505543	0.07510198	-0.07340167	-0.08947410
H	4.67105370	-1.14787020	-0.44071598	-0.10112484	-0.05125397	0.09163121
H	4.14353900	-2.81850244	-0.16994727	-0.00314896	0.12714411	0.07425286
C	3.77961897	-1.89006325	-2.79458824	-0.09006643	0.02725933	-0.29748122
H	3.42819376	-1.17156866	-3.53677139	0.09430820	-0.15514166	0.03104178
H	3.63167795	-2.89334847	-3.20423940	0.05828294	0.12251793	0.00635706
H	4.83689355	-1.72773486	-2.60329120	-0.03950152	0.00253567	0.04768232

#####

HCAPPED.extxyz is the **hydrogen-capped, sidechain-focused PEPPER training dataset**. Its purpose is different from both MIN.extxyz and ALL.extxyz: it was constructed specifically to test what is lost when the chemically realistic peptoid backbone environment is removed.

25

```
Properties=species:S:1:pos:R:3:REF_forces:R:3 REF_energy=-14570.338879004823 pbc="F F F"
C 15.76000110 26.70000115 27.88999914 -0.31479916 0.34516453 2.15561219
H 16.11000108 27.29000114 27.04000312 -0.24973101 -0.68384628 -0.30391878
H 16.17000210 27.26000013 28.74000216 -0.36869030 -0.13082900 -0.03495094
H 14.67000109 26.67000014 28.01000214 0.36157807 0.35030201 -0.44068879
C 16.21000107 25.24000212 27.88999914 1.91719175 2.78216958 -1.45393642
O 15.56000106 24.48000109 27.24000215 -0.86380700 -2.23650116 -0.35582621
N 17.52000010 25.04000113 28.25000015 1.66549912 -0.33029364 1.91279386
C 18.33000007 23.93000013 27.70000315 -1.20848064 -0.06442418 -0.55127373
H 18.59000008 24.22000114 26.67000014 0.48092620 -0.23808284 0.43178892
H 17.76000208 23.00000213 27.59000012 -0.21683232 0.38133195 0.18679065
H 19.18000007 23.67000213 28.33000216 -0.00183907 0.49890232 -0.17486947
C 18.28000108 26.03000311 29.13000116 -2.28712019 0.48662804 -0.90692134
H 18.23000108 26.96000115 28.54999912 0.50313676 -0.01549557 0.61942008
H 19.33000207 25.75000010 29.27000012 -0.50482725 -0.22177039 -0.11461521
C 17.68000006 26.37000111 30.57000215 2.20371385 -1.95049334 0.31137540
O 17.86000107 27.45000115 31.08000013 -0.51432962 1.11476230 -0.50725504
N 17.37000107 25.30000315 31.35000217 -0.18178537 -3.07956297 1.17185114
C 17.42000207 23.80000113 30.90000214 -0.55978837 1.76133529 0.24044469
H 18.41000008 23.46000313 30.58000217 -0.11814413 0.41660888 0.02636340
H 16.69000110 23.64000112 30.09000016 0.43560133 -0.04714468 0.54250657
H 17.15000009 23.22000110 31.79000116 0.01502544 0.18812548 -0.51821417
C 17.16000011 25.41000210 32.86000119 -0.93271279 0.60644812 -1.01775619
H 16.78000107 24.44000112 33.20999915 0.11282367 0.53804416 -0.20873334
H 16.40000209 26.18000014 33.06000117 0.14207655 -0.34229727 -0.21626356
H 17.99000207 25.70000111 33.51000215 0.48531449 -0.12908133 -0.79372372
```

25

```
Properties=species:S:1:pos:R:3:REF_forces:R:3 REF_energy=-14569.603270121786 pbc="F F F"
C 16.20000110 29.02000215 31.93000018 1.00534915 1.61983780 0.65976298
H 16.30000109 29.60000016 31.01000015 -0.52964452 -0.63285089 -0.23178093
H 16.54000106 27.98000113 31.82000218 -0.32277276 0.36938885 -0.18600794
H 16.95000110 29.55000112 32.54000117 -0.44664096 -0.41091838 0.32275406
C 14.74000107 29.03000117 32.45000114 5.78550262 -1.78511359 3.83121659
O 13.93000004 28.82000216 31.65000215 -4.52057781 0.13189002 -2.48047348
N 14.64000108 28.78000113 33.86000117 -2.99104871 0.45728155 -1.43717641
C 13.30000008 28.19000214 34.26000215 0.31306442 0.60042300 0.86331858
H 12.61000104 27.99000014 33.43000019 1.08970309 -0.47351661 -0.16795245
H 12.61000104 28.79000115 34.88000115 0.92675551 0.65473689 0.02750752
H 13.52000106 27.34000214 34.92000218 -0.24271419 -0.22245595 -0.67964061
C 15.78000109 28.62000313 34.77000415 -2.22903003 -0.67159920 -0.36190156
H 16.44000106 28.02000015 34.13000115 1.01093525 0.67102186 1.07889372
H 15.37000109 27.95000112 35.54000118 0.55279406 0.10906436 -0.13026456
C 16.24000208 29.95000115 35.42000215 1.56375537 -0.62973444 -0.53985148
O 15.97000105 30.99000217 34.87000319 0.04610558 0.22698424 0.19757504
```

N	17.15000009	29.85000216	36.48000316	-0.50368641	0.15606220	-0.64910817
C	17.05000110	28.70000112	37.43000421	0.12846830	1.73272767	-0.68779974
H	17.44000110	28.97000115	38.42000217	0.53110629	-0.02023079	-0.59236570
H	17.61000108	27.90000213	36.93000017	-0.74767954	-0.48095118	0.91702735
H	15.98000007	28.64000116	37.66000421	0.28113960	-1.12425858	-0.45071622
C	18.05000109	30.98000014	36.67000217	-0.02317329	0.23780817	2.66366524
H	18.40000011	30.98000014	37.71000320	-0.44568891	-0.41106069	0.29565090
H	17.48000007	31.92000217	36.69000216	0.58125565	-0.20537608	-1.04259671
H	19.06000109	30.95000113	36.24000219	-0.81327774	0.10083976	-1.21973603