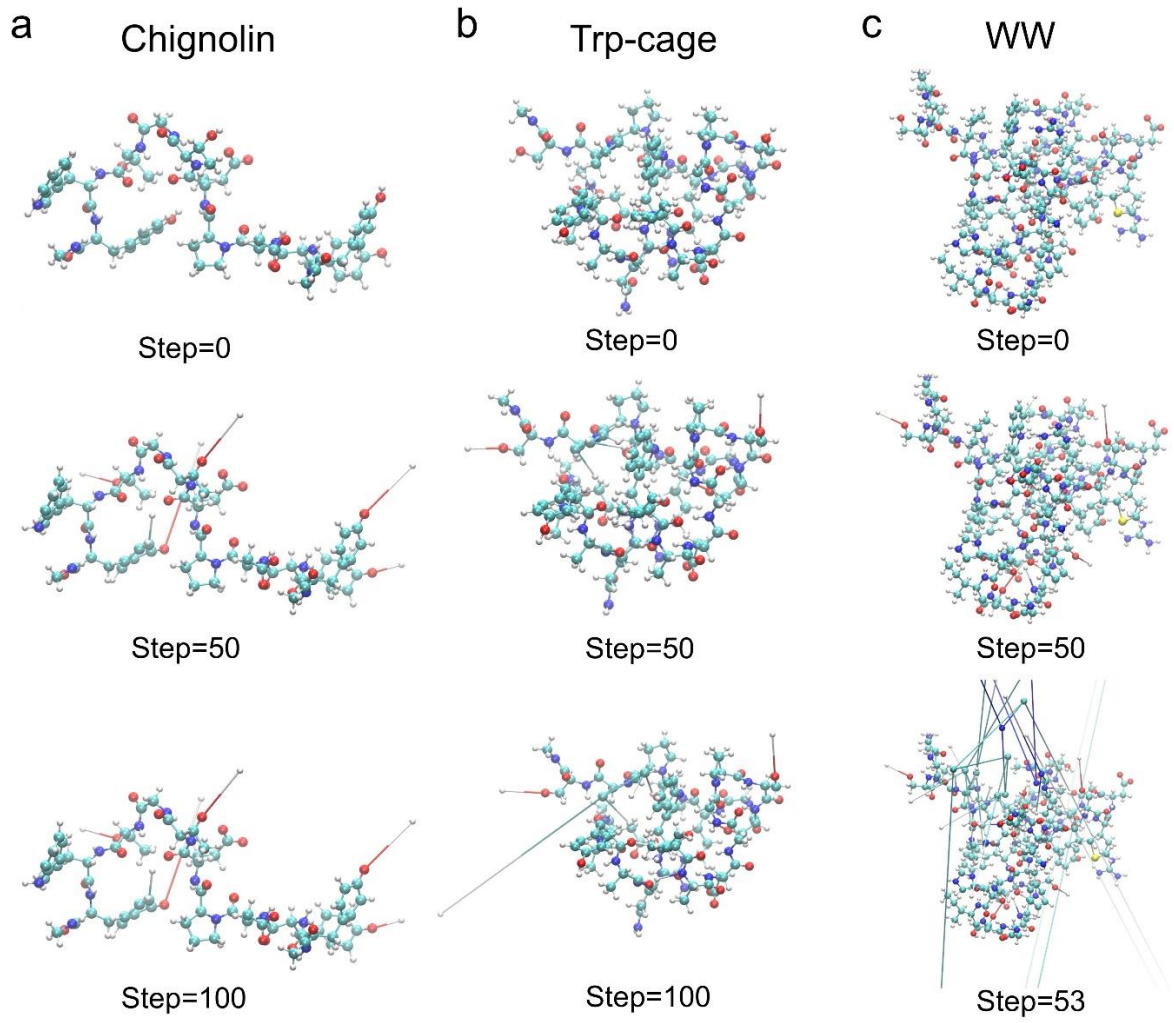
nature portfolio

https://doi.org/10.1038/s41586-024-08127-z

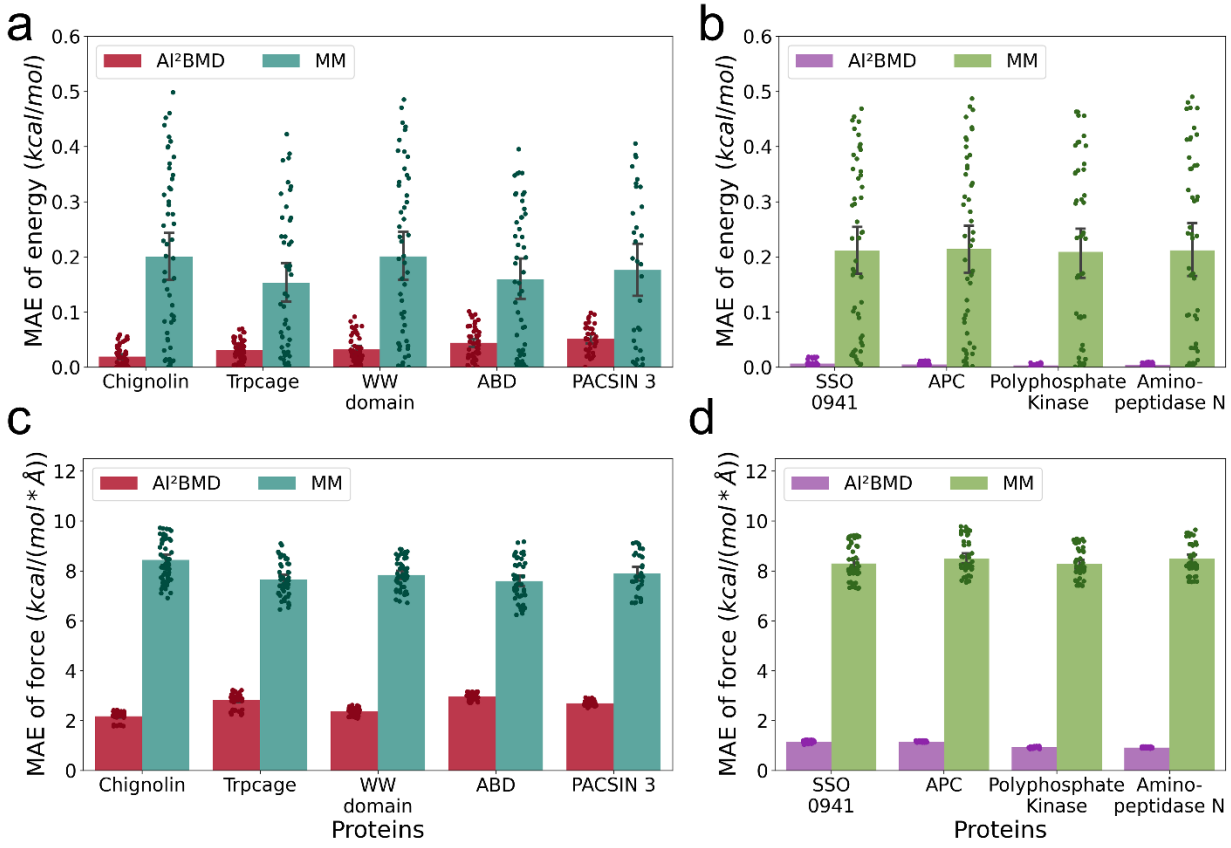
Supplementary information

Ab initio characterization of protein molecular dynamics with AI^{2}BMD

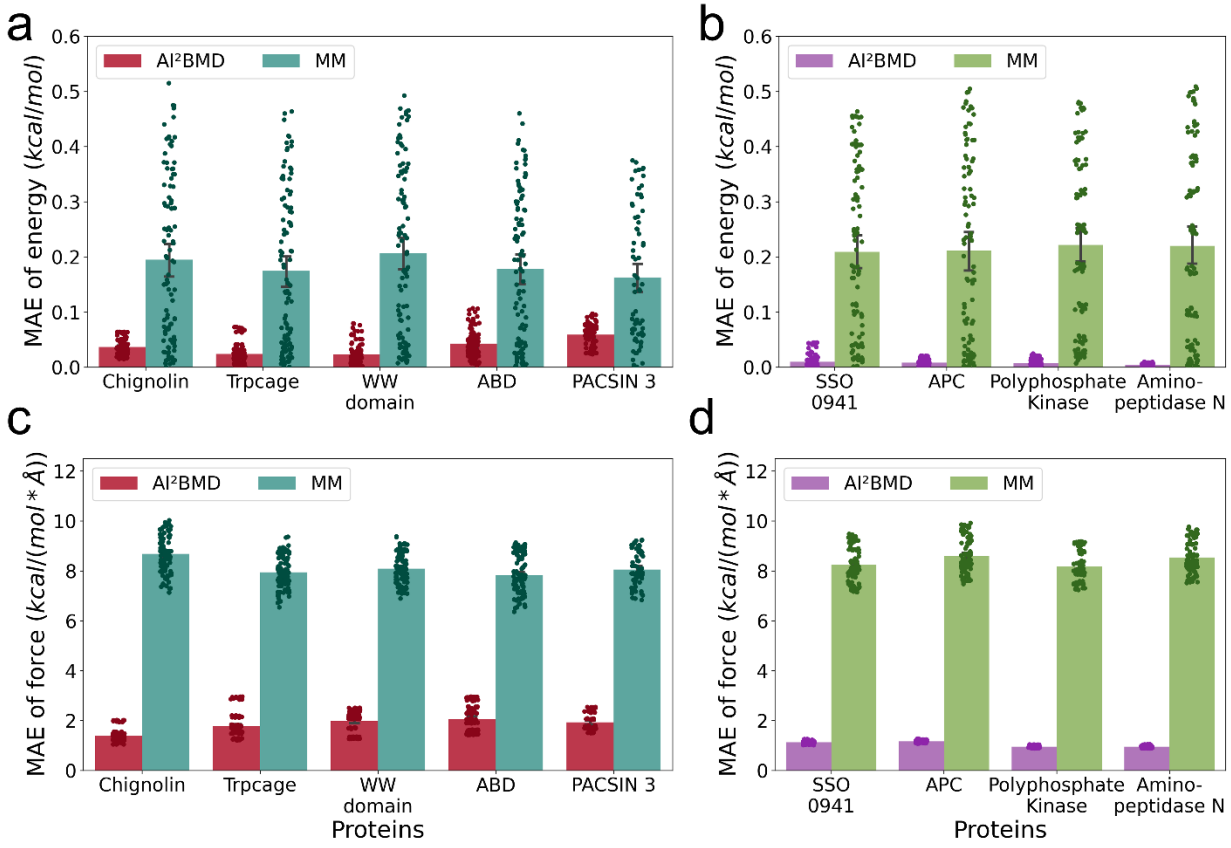
In the format provided by the authors and unedited



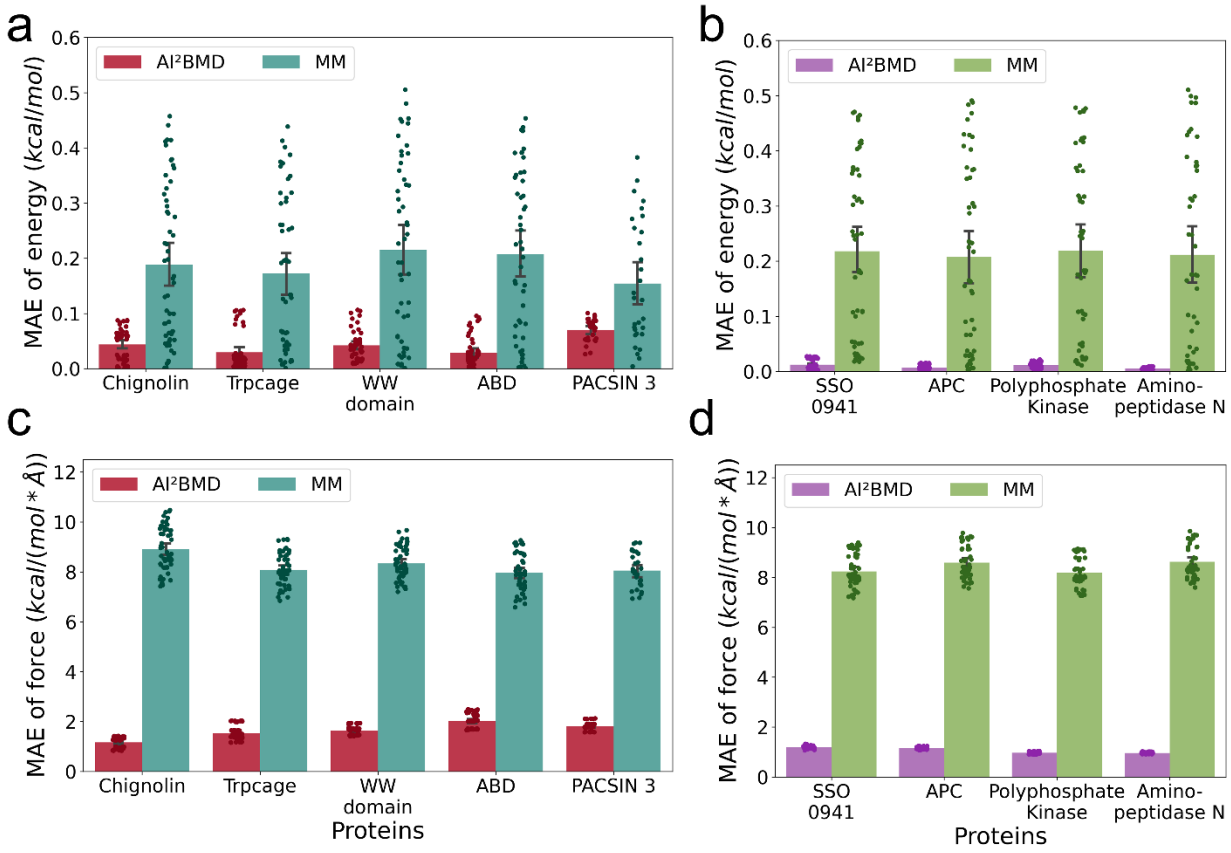
Supplementary Figure 1. Simulation collapse during Chignolin, Trp-cage and WW domain simulations driven by a machine learning potential trained on the tetrapeptide Ac-Ala3-NHMe on MD22 dataset. During the simulation, initial structure, step 50 and step 100 (step 53 for WW domain due to the simulation system crashed at that step) are visualized. The problematic positions of several atoms are visible by the unphysical bonds.



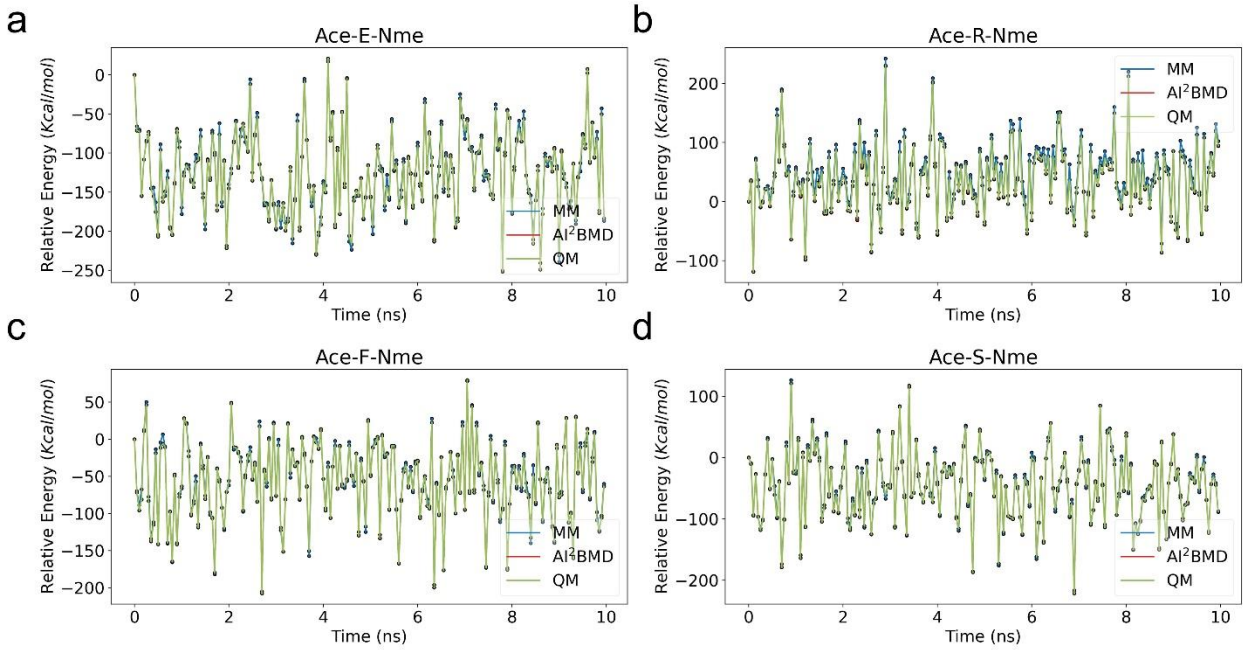
Supplementary Figure 2. Energy and force calculations by AI^{2}BMD and molecular mechanics (MM) for various proteins in the folded state. a-b, The mean absolute error (MAE) of potential energy compared with DFT as the reference values for proteins within 1,040 atoms (a) or fragmented DFT as the reference values for larger proteins (b). c-d, The mean absolute error (MAE) of atomic forces compared with DFT as reference values for proteins within 1,040 atoms (c) or fragmented DFT as the reference values for larger proteins (d).



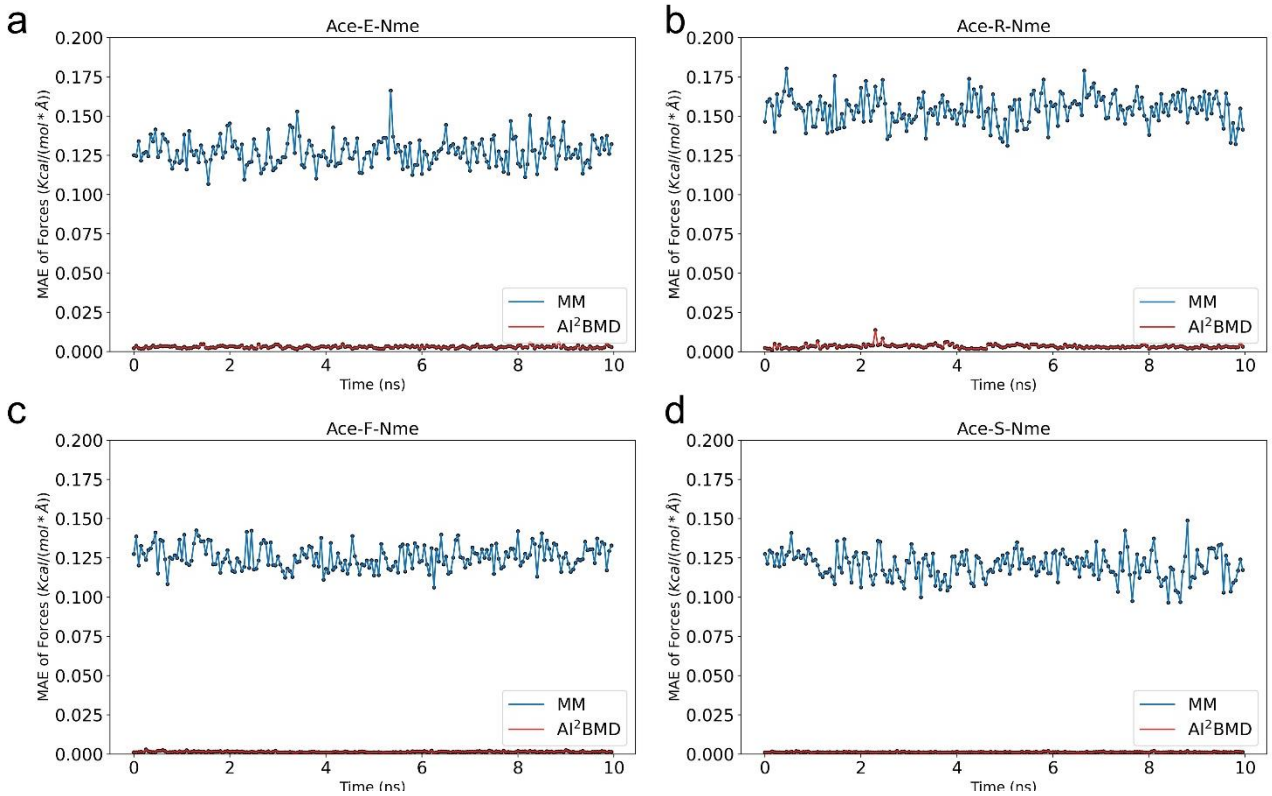
Supplementary Figure 3. Energy and force calculations by AI^{2}BMD and molecular mechanics (MM) for various proteins in the intermediate states. a-b, The mean absolute error (MAE) of potential energy compared with DFT as the reference values for proteins within 1,040 atoms (a) or fragmented DFT as the reference values for larger proteins (b). c-d, The mean absolute error (MAE) of atomic forces compared with DFT as the reference values for proteins within 1,040 atoms (c) or fragmented DFT as the reference values for larger proteins (d).



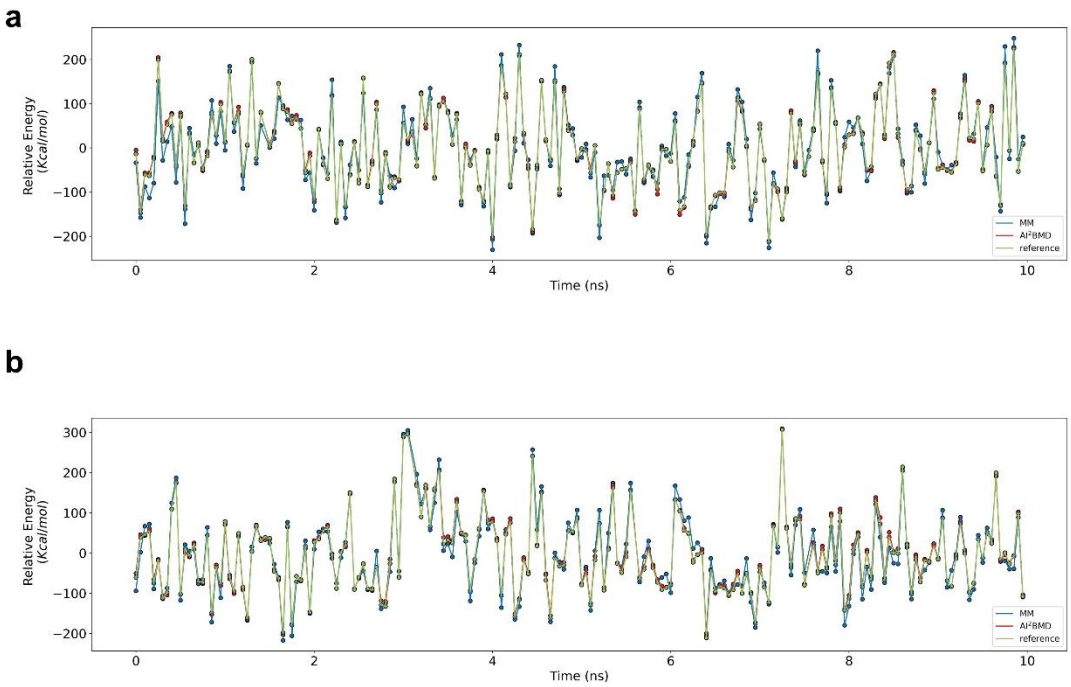
Supplementary Figure 4. Energy and force calculations by AI^{2}BMD and molecular mechanics (MM) for various proteins in the unfolded state. a-b, The mean absolute error (MAE) of potential energy compared with DFT as the reference values for proteins within 1,040 atoms (a) or fragmented DFT as the reference values for larger proteins (b). c-d, The mean absolute error (MAE) of atomic forces compared with DFT as the reference values for proteins within 1,040 atoms (c) or fragmented DFT as the reference values for larger proteins (d).



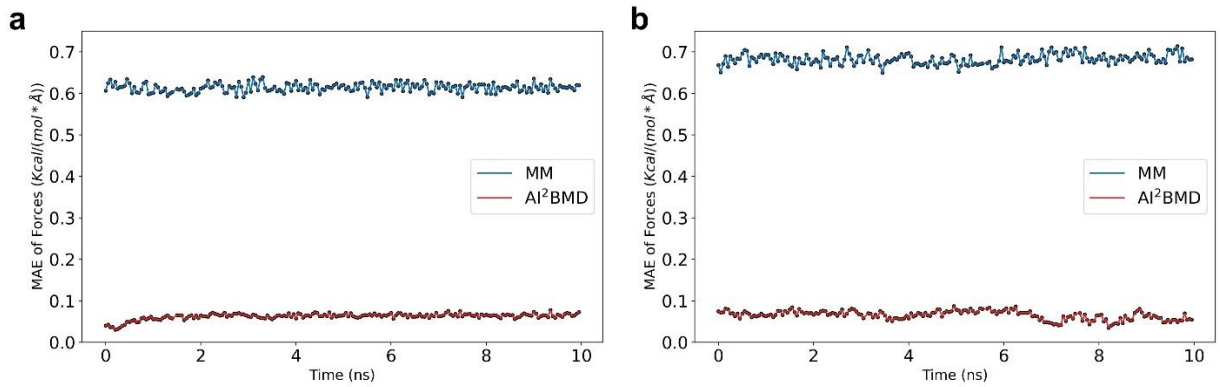
Supplementary Figure 5. The relative potential energies of the entire system during the 10 ns simulations for negatively charged Ace-E-Nme (a), positively charged Ace-R-Nme (b), aromatic Ace-F-Nme (c) and polar Ace-S-Nme (d). QM calculation for the whole protein and Amoeba force field calculation for the remaining solvent part act as the reference value. The relative energy is defined as the potential energy of each structure subtracts that of a reference structure.



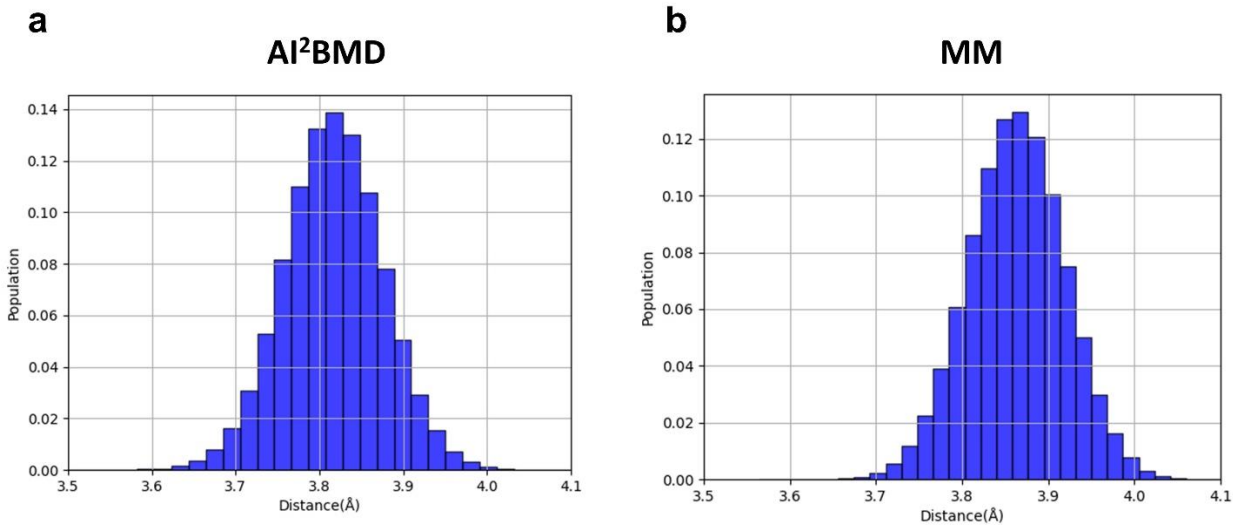
Supplementary Figure 6. The MAE of atomic forces of the entire system during the 10 ns simulations for negatively charged Ace-E-Nme (a), positively charged Ace-R-Nme (b), aromatic Ace-F-Nme (c) and polar Ace-S-Nme (d). Calculations by QM for the whole protein and Amoeba force field for the solvent part act as the reference values, and those calculated by MM are shown for comparison.



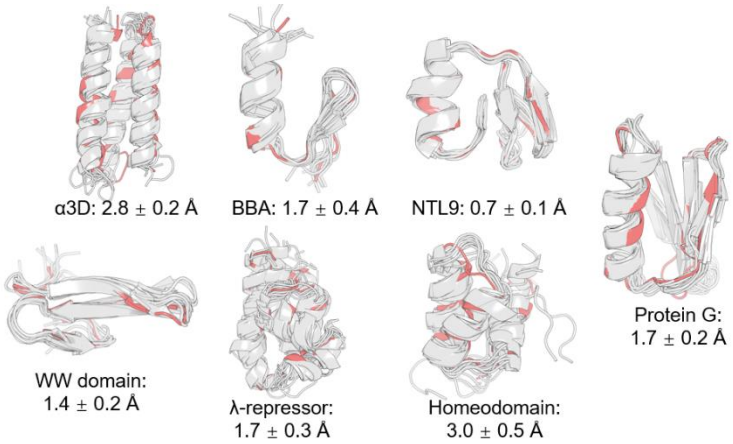
Supplementary Figure 7. The relative potential energies of the entire system during the 10 ns simulations for Chignolin starting from an unfolded (a) and a folded structure (b). QM calculation for the whole protein and Amoeba force field calculation for the remaining solvent part act as the reference value. The relative energy is defined as the potential energy of each structure subtracts that of a reference structure.



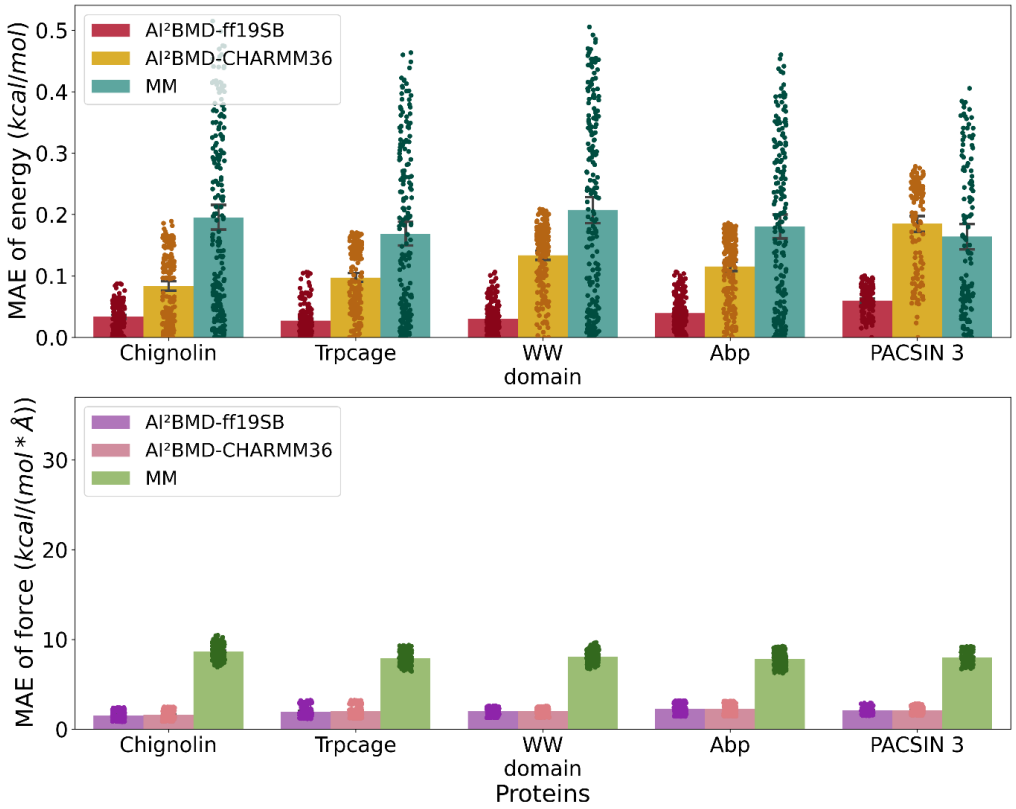
Supplementary Figure 8. The MAE of atomic forces of the entire system of solvated Chignolin during the simulations for two different initial structures. The initial and final structures in (a) and (b) are the same of those in Fig. 4 (a)-(b). Calculations by QM/MM act as the reference values and those calculated by MM are shown for comparison.



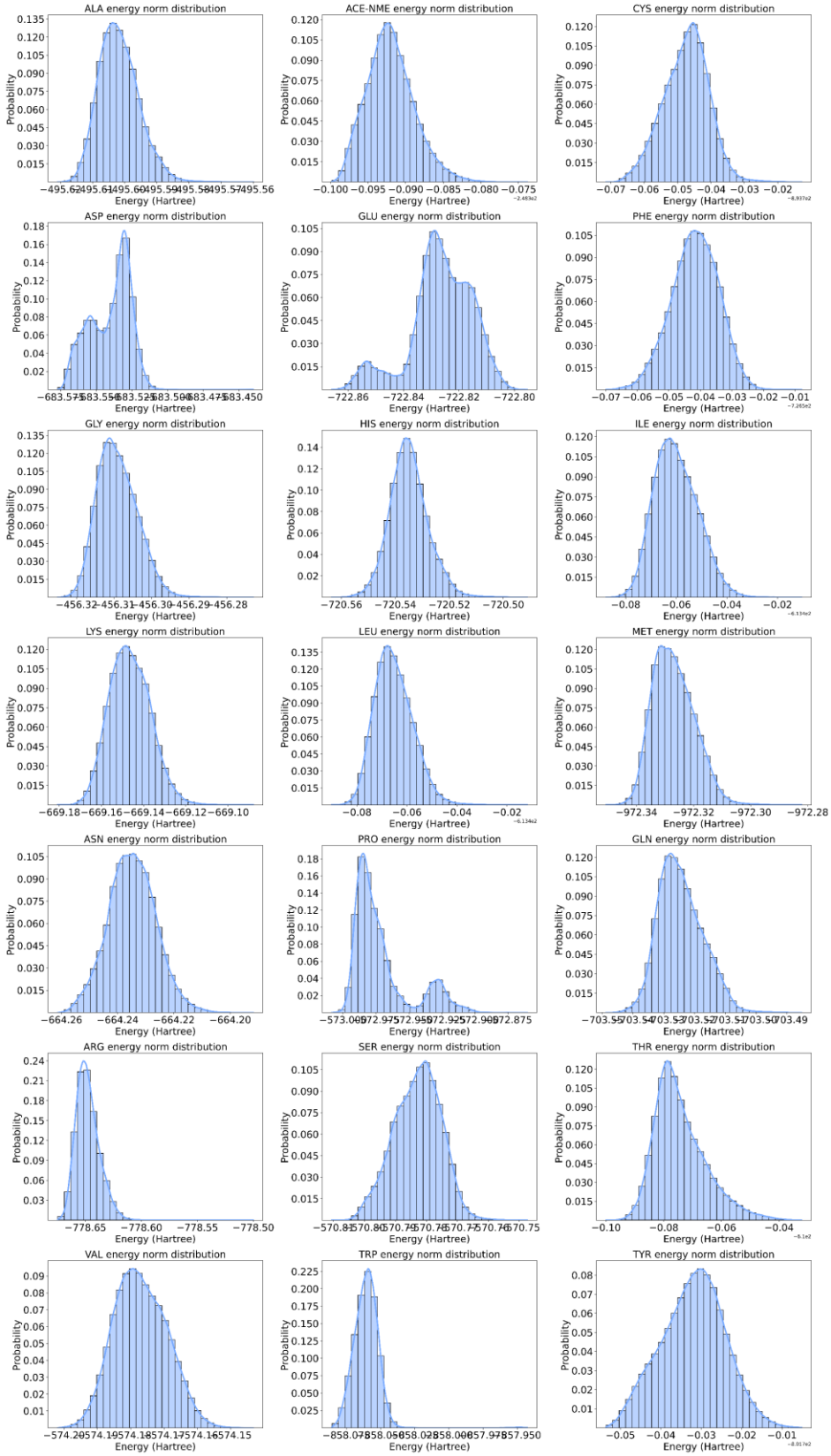
Supplementary Figure 9. The distance distribution of adjacent Cα atoms during the 10 ns Chignolin simulations performed by AI^{2}BMD and MM, respectively.



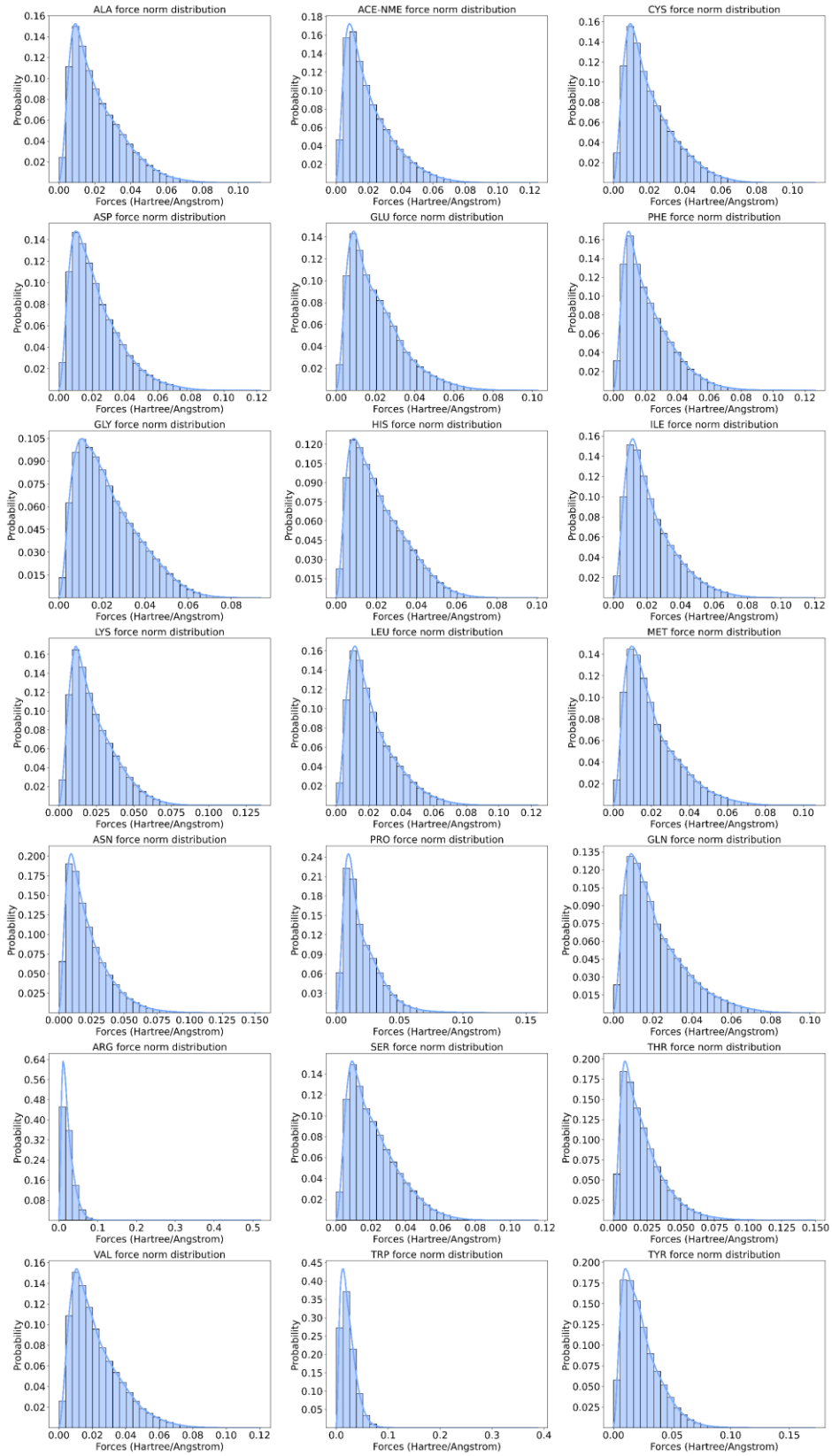
Supplementary Figure 10. The structures located at the center of the cluster of folded conformations (gray) compared with the corresponding folded crystal structures (salmon). Protein name and RMSD values are shown below.



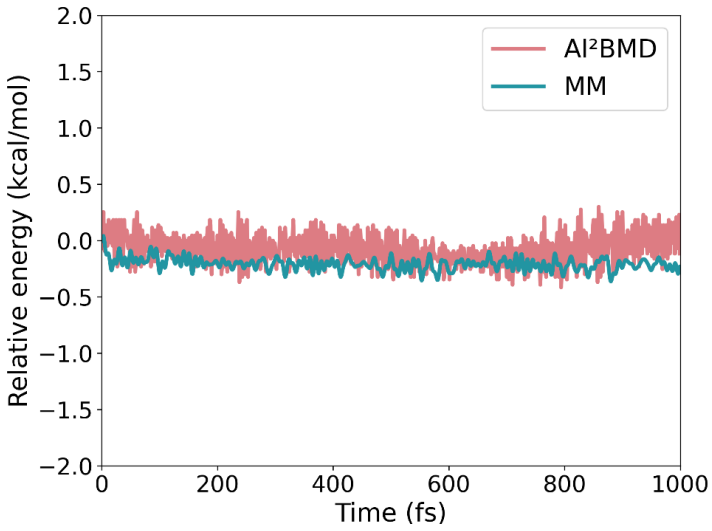
Supplementary Figure 11. The mean absolute error (MAE) of potential energy and atomic forces using the non-bonded coefficients derived from different force fields (FF19SB or Charmm36).



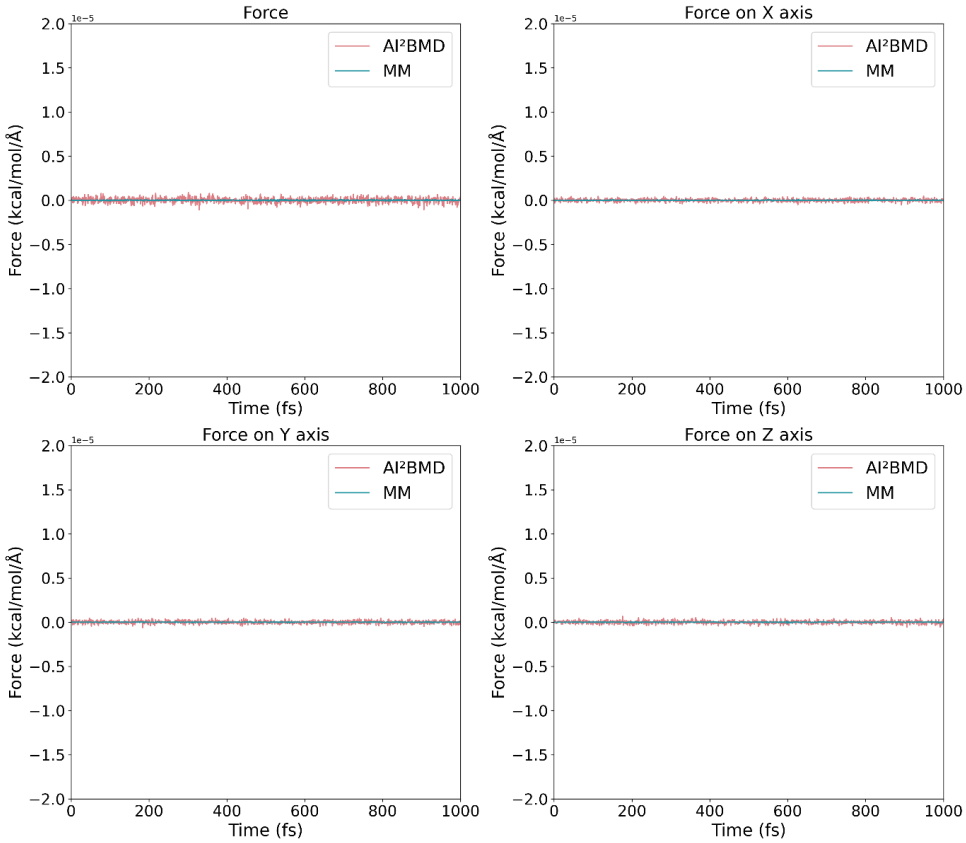
Supplementary Figure 12. The distribution of potential energy of each kind of protein unit.



Supplementary Figure 13. The distribution of the norm of atomic forces of each kind of protein unit.



Supplementary Figure 14. The time-course curves for the relative energy of the total system during 1 ps AI^{2}BMD and MM simulations for a dipeptide in a solvent environment. The relative energy is the difference between this snapshot and the initial structure.



Supplementary Figure 15. The time-course curves for the force and its components on X, Y, and Z axis respectively during 1 ps AI^{2}BMD and MM simulations for a dipeptide in a solvent environment.

Supplementary Table 1. Evaluation of calculation accuracy of potential energy and atomic forces for protein units. For MAE of energy, the potential energy of each structure subtracts that of a reference structure first.

Method	AI^{2}BMD		MM	
Metrics	MAE of energy (kcal/mol)	MAE of force (kcal/(mol*Å))	MAE of energy (kcal/mol)	MAE of force (kcal/(mol*Å))
AA	0.0325	0.0585	2.578	7.817
HH1	0.0657	0.0852	3.696	9.034
RR	0.0854	0.0727	2.801	7.767
DD	0.0536	0.0852	6.553	9.164
CC	0.0501	0.0727	2.372	7.785
QQ	0.0485	0.0631	3.327	8.722
EE	0.0795	0.0831	4.941	8.733
II	0.0684	0.0860	2.895	7.522
GG	0.0292	0.0598	2.425	8.554
NN	0.0466	0.0738	3.481	9.165
LL	0.0599	0.0735	2.734	7.281
KK	0.1025	0.0680	3.223	8.125
MM	0.0428	0.0613	4.563	6.915
FF	0.0560	0.0723	3.266	7.842
PP	0.0687	0.1075	3.249	7.488
SS	0.0407	0.0720	3.031	8.037
TT	0.0575	0.0869	2.495	8.357
WW	0.0703	0.0828	3.045	8.418
YY	0.0553	0.0742	2.731	8.843
VV	0.0577	0.0758	2.533	7.645
ACE-NME	0.0138	0.0628	1.211	7.409

Supplementary Table 2. Summary of the thermodynamic parameters of 7 proteins

Protein	AI^{2}BMD ΔG (kcal/mol)	MM ΔG (kcal/mol)	Sim. T (K)	Exp. Tm (K)	AI^{2}BMD Tm (K)	MM Tm (K)
BBA	0.057	1.22	325	-	323.94±0.22	322.34±0.31
WW domain	0.41	-0.73	360	371±2	359.06±0.07	353.69±0.38
NTL9	-0.34	-1.54	355	355±2	351.84±0.11	349.47±0.35
Homeodomain	-0.18	-0.73	360	>372	359.61±0.14	359.60±0.13
Protein G	0.14	0.74	350	-	349.49±0.12	346.49±0.66
α3D	-0.098	1.33	370	>363	369.67±0.06	366.94±0.26
λ-repressor	0.79	1.09	350	347	349.55±0.21	349.48±0.21

Supplementary Table 3. The change of enthalpy and heat capacity for Barnase and CI2 calculated by MM, AI^{2}BMD and experiment, respectively.

Protein	Method	ΔH_{unf} (kcal/mol)	ΔCp_{unf} (kcal/mol·K)
Barnase	MM	110.4 ± 3.1	1.0 ± 0.1
	AI^{2}BMD	116.5 ± 0.43	1.4 ± 0.04
	Experiment	118.7 ± 4.9	1.4 ± 0.1
CI2	MM	57.1 ± 1.9	0.5 ± 0.1
	AI^{2}BMD	73.0 ± 0.97	0.7 ± 0.02
	Experiment	78.4 ± 1.4	0.8 ± 0.1

MM values are derived from J Chem Inf Model, 2023,63,24,7791-7806. The experimental values are derived from Eur J Biochem, 1994, 221,3, 1003-1012 and Biochem, 1991,30,43, 10428-10435.

Supplementary Table 4. Statistics of potential energy of each kind of protein unit.

Protein unit	Ave. value (Hartree)	Max. value (Hartree)	Min. value (Hartree)
ALA	-495.60	-495.56	-495.62
CYS	-893.75	-893.71	-893.77
ASP	-683.54	-683.44	-683.58
GLU	-722.83	-722.80	-722.87
PHE	-726.54	-726.51	-726.57
GLY	-456.31	-456.27	-456.32
HIS	-720.53	-720.49	-720.56
ILE	-613.46	-613.41	-613.49
LYS	-669.15	-669.09	-669.18
LEU	-613.47	-613.41	-613.49
MET	-972.33	-972.28	-972.35
ASN	-664.23	-664.19	-664.26
PRO	-572.97	-572.86	-573.01
GLN	-703.52	-703.49	-703.55
ARG	-778.65	-778.50	-778.67
SER	-570.78	-570.75	-570.81
THR	-610.07	-610.03	-610.10
VAL	-574.18	-574.14	-574.20
TRP	-858.06	-857.94	-858.09
TYR	-801.73	-801.71	-801.75
ACE-NME	-248.39	-248.37	-248.40

Supplementary Table 5. Statistics of the norm of atomic forces of each kind of protein unit.

Protein unit	Ave. value (Hartree/Å)	Max. value (Hartree/Å)	Min. value (Hartree/Å)
ALA	0.02152	0.11241	0.00002
CYS	0.02055	0.11248	0.00004
ASP	0.02263	0.12218	0.00006
GLU	0.02027	0.10274	0.00004
PHE	0.02195	0.12629	0.00002
GLY	0.02282	0.09356	0.00005
HIS	0.02100	0.10036	0.00003
ILE	0.02249	0.12020	0.00005
LYS	0.02299	0.13449	0.00004
LEU	0.02219	0.12394	0.00004
MET	0.02046	0.10614	0.00004
ASN	0.02189	0.15428	0.00004
PRO	0.02053	0.15938	0.00003
GLN	0.02144	0.10256	0.00002
ARG	0.02275	0.51981	0.00005
SER	0.02157	0.11596	0.00002
THR	0.02204	0.14952	0.00006
VAL	0.02233	0.12049	0.00007
TRP	0.02327	0.38795	0.00004
TYR	0.02396	0.17198	0.00004
ACE-NME	0.02077	0.12483	0.00006

Supplementary Table 6. Hyperparameters of AI^{2}BMD Potential.

Hyperparameters	Values
Init Learning Rate	0.0002
LR warmup (steps)	1000
LR decay patience (epochs)	15
LR decay factors	0.8
Batch Size / GPU	{64, 128}
Number of GPUs	16
Maximum number of epochs	6000
Early stopping patience (epochs)	150
Cutoff (Å)	5
Maximum number of neighbors	32
Forces/Energy weights	0.95/0.05
Number of Layers	6
Embedding Dimensions	128
Number of attention heads	8
Number of RBF features	32
L_{max}	2