Supporting Information: NN-Dx: Chemically adaptive dispersion physics for density-functional theory

Yufan Xia,[1] Adrian R. Pearce,[1] and Giuseppe Barca[2,3,*]

[1]*Faculty of Engineering and Information Technology, University of Melbourne, Melbourne, Victoria, Australia*

[2]*Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia*

[3]*School of Computing, Australian National University, Canberra, ACT, Australia*

(Dated: August 3, 2026)

SUPPLEMENTARY NOTE 1. NN-DX MODEL AND TRAINING HYPERPARAMETERS

Table S1 lists the complete set of architecture and training hyperparameters for the NN-Dx model, complementing the description in the Methods section of the main paper. All models in this work share these settings; only the functional-specific baseline damping and scaling parameters $\{a_1, a_2, s_6, s_8, s_9\}_0$ differ between functionals, while the element-specific baseline polarizability spectra $\alpha^0(i\omega)$ are shared across functionals.

SUPPLEMENTARY NOTE 2. DATASET COMPOSITION AND TRAIN/VALIDATION/TEST SPLITS

The unified training corpus used in this work combines the D4-original benchmark sets with three complementary benchmark subsets and four train-only augmentation sets. Table S2 reports per-subset counts and reference levels of theory; Table S3 summarizes the training / validation / test partitioning protocol.

The corpus includes four train-only augmentation sets (Tables S2 and S3) that target regimes or local chemical environments underrepresented in equilibrium dimer benchmarks. All four are assigned entirely to the training partition; their held-out analogues—the eclipsed cofacial scans and the large polycyclic and cage complexes—serve as the corresponding evaluation probes.

a. Cofacial-stacking augmentation set. To broaden coverage of the cofacial (π-stacked sandwich) motif—which the benchmark subsets above sample only in T-shaped and parallel-displaced orientations—the corpus is augmented with 40 face-to-face benzene·benzene and pyridine·pyridine dimers. For each dimer the ring-plane separation is scanned over 13–14 points from 3.0 to 8.0 Å with rigid monomers. To keep these training configurations geometrically distinct from the eclipsed cofacial sandwich scans used for held-out PES validation (the main-text PES figure; per-functional versions in Supplementary Note 4), the stacked monomers are rotated about the stacking axis: benzene to a $\varphi = 30°$ staggered registry, and pyridine to parallel and antiparallel dipole orientations. Every augmentation geometry differs from its nearest validation scan point by an all-atom RMSD of at least 0.3 Å (per-element Hungarian assignment with Kabsch alignment), so those cofacial scans remain a genuine held-out probe of the motif. Reference interaction energies are CCSD(T)/CBS from a focal-point scheme: an MP2/CBS extrapolation [aug-cc-pV(TQ)Z, Helgaker $1/X^3$] augmented with a δCCSD(T)/aug-cc-pVDZ correction, counterpoise-corrected with a frozen core. All 40 points are assigned to the training partition (Table S3); held-out evaluation of the cofacial motif is provided by these PES scans. This is a pilot augmentation, restricted to homodimers of rigid monomers; a broader grid—spanning the sandwich-to-parallel-displaced bridge, twist-angle and lateral-offset variation, and heteroaromatic stacks—is left to future work.

b. Long-range tail-zero set. A further 1048 train-only configurations per functional place the two monomers at large separation, where the interaction energy—and hence the dispersion correction—is essentially zero. The target residual is approximately zero by construction (the analytic long-range asymptote), so these points anchor the correct vanishing tail of the correction and suppress any spurious constant offset as the intermonomer distance grows.

c. PAH-interior set. To cover an aromatic environment absent from the dimer benchmarks—the peri-fused *interior* sp^2 carbon of extended polycyclic aromatics (three aromatic-carbon neighbours and no hydrogen), as opposed to the edge aromatic carbon sampled by S22, S66 and X40—we add 27 train-only points per functional from cofacial staggered (30°) homodimers of pyrene ($C_{16}H_{10}$) and perylene ($C_{20}H_{12}$), scanned over centre-of-mass separations of 3.0–8.0 Å. Reference interaction energies use DLPNO-CCSD(T)/CBS [aug-cc-pVTZ/QZ, TightPNO, counterpoise], the level adopted for the large-complex references, because canonical CCSD(T) is infeasible at the resulting 52–64-atom dimer size. Coronene and circumcoronene, the interior-carbon members of the large-complex test set, are held out, and pyrene and perylene appear in no evaluation set, so the large-system extrapolation reported in the main text remains a genuine small-to-large transfer test.

d. Ureido-interior set. To cover the sp^3 aminal carbon flanked by two ureido nitrogens (the N–C–N bridgehead of glycoluril and related cage bridges), likewise

* giuseppe.barca@monash.edu

TABLE S1. Architecture and training hyperparameters of NN-Dx. The equivariant encoder uses two MACE message-passing layers; the prediction heads are per-parameter multilayer perceptrons operating on its scalar node and pooled descriptors.

Component / hyperparameter	**Value**
Equivariant encoder (MACE)	
Message-passing layers T	2
Max spherical-harmonic order $l_{\max}$	3
Correlation order $\nu_{\max}$	3
Hidden irreps	64x0e+64x1o+64x2e+64x3o
Channels	64
Cutoff radius $r_{\max}$	5.0 Å
Radial basis (Bessel)	8
Cutoff envelope basis	5
Radial MLP	$[64, 64, 64]$
Prediction heads	
MLP depth / width	3 layers / 128
Activation	SiLU
Bounded clamp (c, b)	$(10^{-4}, 0.5)$
Element-parameter outputs	25
Global-parameter outputs	6 (disabled)
Pair-parameter outputs	5
Correction parametrization	
Baseline spectra $\alpha^0(i\omega)$	isolated-atom (element-specific) reference
Polarizability correction	additive, $\alpha^0 + \Delta\alpha$, shift $\in [-0.5, 0.5]$
Damping / scaling corrections	multiplicative, $p_0[1 + \frac{1}{2}(\Delta_A + \Delta_B)]$, factor $\in [0.5, 1.5]$
Training	
Optimizer	AdamW
Weight decay	10^{-5}
Initial learning rate	10^{-3}
LR schedule	reduce-on-plateau ($\times 0.5$, patience 10)
Early-stop LR floor	10^{-9}
Loss	energy MSE
Precision	float64
Hardware	single NVIDIA V100 (<1 day/functional)

TABLE S2. Composition of the unified benchmark corpus. The reference column indicates the high-level method used to generate the target interaction energy; system size is the number of atoms in the dimer complex.

Subset	**Count**	**System size**	**Reference**	**Source**
S22×5	108	6–34	CCSD(T)/CBS	D4 original
S66×8	528	6–34	CCSD(T)/CBS	Řezáč 2011
NCIBLIND10	80	14–32	CCSD(T)/CBS	D4 original
DES15K_ion	4 388	2–33	CCSD(T)/CBS	DES370K subset [1]
DES15K_neutral	9 539	4–34	CCSD(T)/CBS	DES370K subset [1]
Others (A24/IDISP/X40×10)	430	2–68	CCSD(T)/CBS	[2–5]
Cofacial sandwich	40	22–24	CCSD(T)/CBS	This work
Tail-zero	1 048	—	≈ 0 (analytic)	This work
PAH-interior	27	52–64	DLPNO-CCSD(T)/CBS	This work
Ureido-interior	34	32	CCSD(T)/CBS	This work
Total	**16 222**	2–68		

absent from the training dimers, we add 34 train-only points per functional from glycoluril ($C_4H_6N_4O_2$) homodimers scanned over 3.0–8.0 Å, with counterpoise CCSD(T)/CBS references. Larger cucurbituril cages built from the same aminal motif are excluded from training.

SUPPLEMENTARY NOTE 3. ML-XDM COMPARISON: EVALUATION PROTOCOL AND CONVENTIONS

This section documents the evaluation protocol, the method implementations, and the counterpoise convention underlying the NN-Dx versus ML-XDM comparison

TABLE S3. Train / validation / test split sizes per subset for the nine DFT functionals modeled in this work (counts shown for PBE0; other functionals differ slightly, mainly in DES15K availability). Splits are stratified by molecule/dimer identity to prevent leakage between partitions; the final column reports the realized train/validation/test percentages.

Subset	Train	Validation	Test	Total	Train/val/test (%)
S22×5	92	5	11	108	85/5/10
S66×8	449	26	53	528	85/5/10
NCIBLIND10	68	4	8	80	85/5/10
DES15K_ion	2 336	1 016	1 036	4 388	53/23/23
DES15K_neutral	5 666	1 872	2 001	9 539	59/19/20
Others (A24/IDISP/X40×10)	345	42	43	430	80/9/10
Cofacial sandwich	40	0	0	40	100/0/0
Tail-zero	1 048	0	0	1 048	100/0/0
PAH-interior	27	0	0	27	100/0/0
Ureido-interior	34	0	0	34	100/0/0
Total	**10 105**	**2 965**	**3 152**	**16 222**	62/18/19

reported in the main paper.

e. Evaluation protocol. The 500-sample HCNO subset is drawn from the DES15K_neutral test partition (Supplementary Note 2) by applying a fixed random permutation to the lexicographically sorted task identifiers and retaining the first 500 tasks whose every frame contains only H, C, N, and O; a nested first-200 subset was also examined and gave consistent results. Each task comprises three frames—the dimer AB and its two monomers A and B—with stoichiometry $[1, -1, -1]$, so that the interaction energy is $E_{\text{int}} = E_{AB} - E_A - E_B$; frames span 2–34 atoms and are all neutral closed-shell singlets (charge 0, multiplicity 1). Energies are converted to kcal/mol with the factor 627.5094740631, and the ML-XDM outputs (reported in eV) are converted to Hartree using the `ase` unit definition. NN-Dx uses the PBE0 model; ML-XDM-1x and ML-XDM-2x are evaluated in dispersion-only mode from the `torchanipbe0 MLXDM` (ANI-1x) and `MLXDM_2x` (ANI-2x) checkpoints respectively; and the reference empirical correction is the default `dxtb` D4 with PBE0 damping.

f. Counterpoise convention. The dispersion-residual target is $E_{\text{disp}} = E_{\text{QM}} - E_{\text{DFT}}$, where E_{QM} is the counterpoise-corrected CCSD(T)/CBS interaction energy and E_{DFT} is the *non*-counterpoise-corrected supermolecular PBE0 interaction energy. This is a deliberate mixed-counterpoise definition: by referencing the correction to an uncorrected supermolecular DFT calculation, the target folds the basis-set superposition error of a standard (non-counterpoise) PBE0 calculation into the learned correction. An end user adding NN-Dx to a routine PBE0 supermolecular calculation therefore recovers the counterpoise-corrected CCSD(T)/CBS interaction energy without performing a separate counterpoise calculation. The same convention is applied uniformly to all four methods in the comparison.

SUPPLEMENTARY NOTE 4. PER-FUNCTIONAL POTENTIAL ENERGY SURFACES

The main text presents the six-dimer binding-curve scans on a PBE0 base. Figures S1–S8 repeat the identical scans for the other eight functionals. Panels (a)–(f), the dimer set, and the curve conventions match the main-text PES figure, except that ML-XDM is omitted (its published checkpoints support only a PBE0 base); each panel therefore compares the bare functional, the default D4 correction, and NN-Dx against the CCSD(T)/CBS reference, with the per-dimer MAE (kcal/mol) annotated in the inset. Note that the water-dimer panel (a) uses the S66 Water–Water scan from the base run (available for every functional) rather than the S22×7 water scan of the main text (which exists only for PBE0); its absolute energies are thus not directly comparable to the main-text water panel, although the relative ordering of methods is preserved.

SUPPLEMENTARY NOTE 5. FORCE DIFFERENTIABILITY AND MOLECULAR DYNAMICS

The exact differentiability of NN-Dx and its behaviour in molecular dynamics (main-text force-conservation figure) were assessed with the settings in Table S4. The derivative test compares the analytic reverse-mode (automatic-differentiation) gradient and Hessian of the NN-Dx dispersion energy against central finite differences of the same energy; the dynamics test propagates microcanonical (NVE) trajectories on the PBE0/def2-TZVP + NN-Dx surface for six dimers (water, formic acid, methane, ethene, benzene·CH_4 and the parallel-displaced benzene dimer).

The small per-panel total-energy slopes quoted in the main-text figure are not long-term drift: on sub-picosecond windows they are dominated by slow intermolecular soft modes (30–90 cm^{-1}, periods 370–1100 fs), change sign with the fit window, and do not scale with the integration timestep, whereas genuine numerical drift would. The residual total-energy fluctuation instead scales as $\Delta t^{1.95}$, confirming that it originates from the symplectic velocity-Verlet integrator rather than from the neural potential.

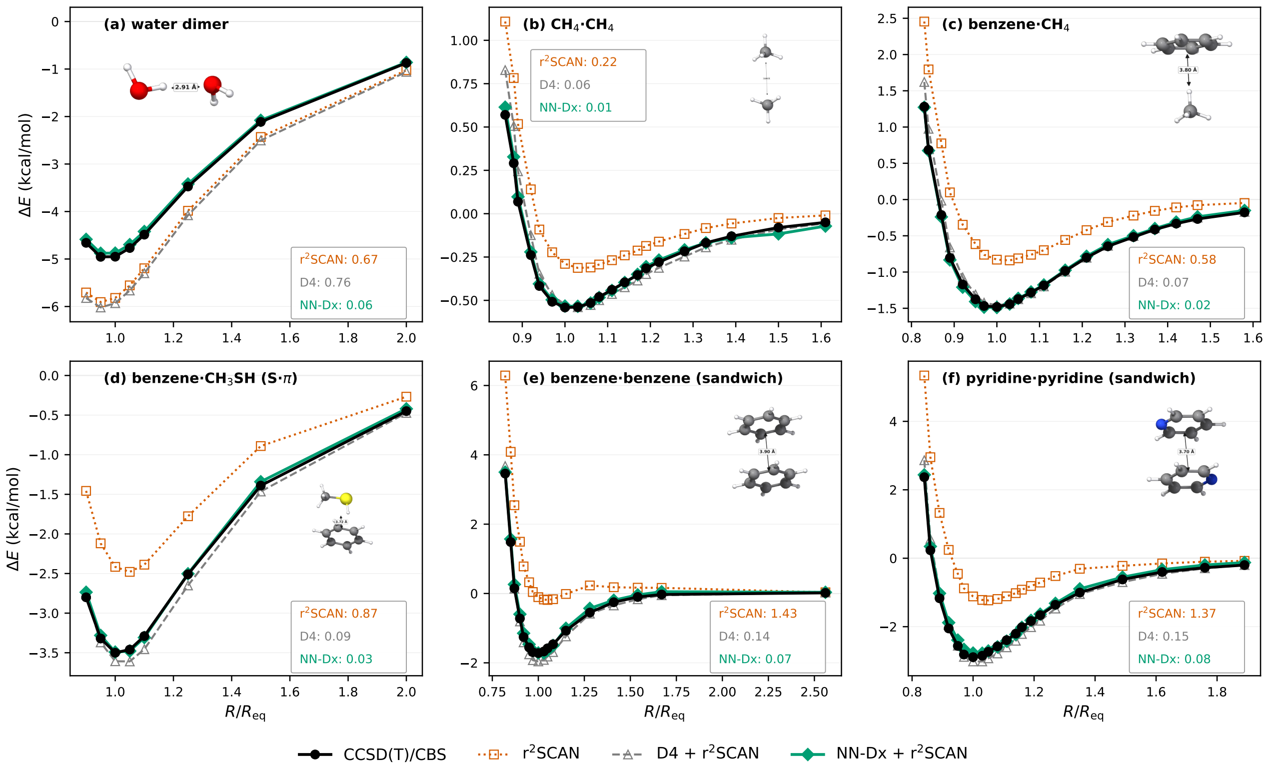

FIG. S1. Per-dimer binding-curve scans with an r^2SCAN base; conventions as in Supplementary Note 4.

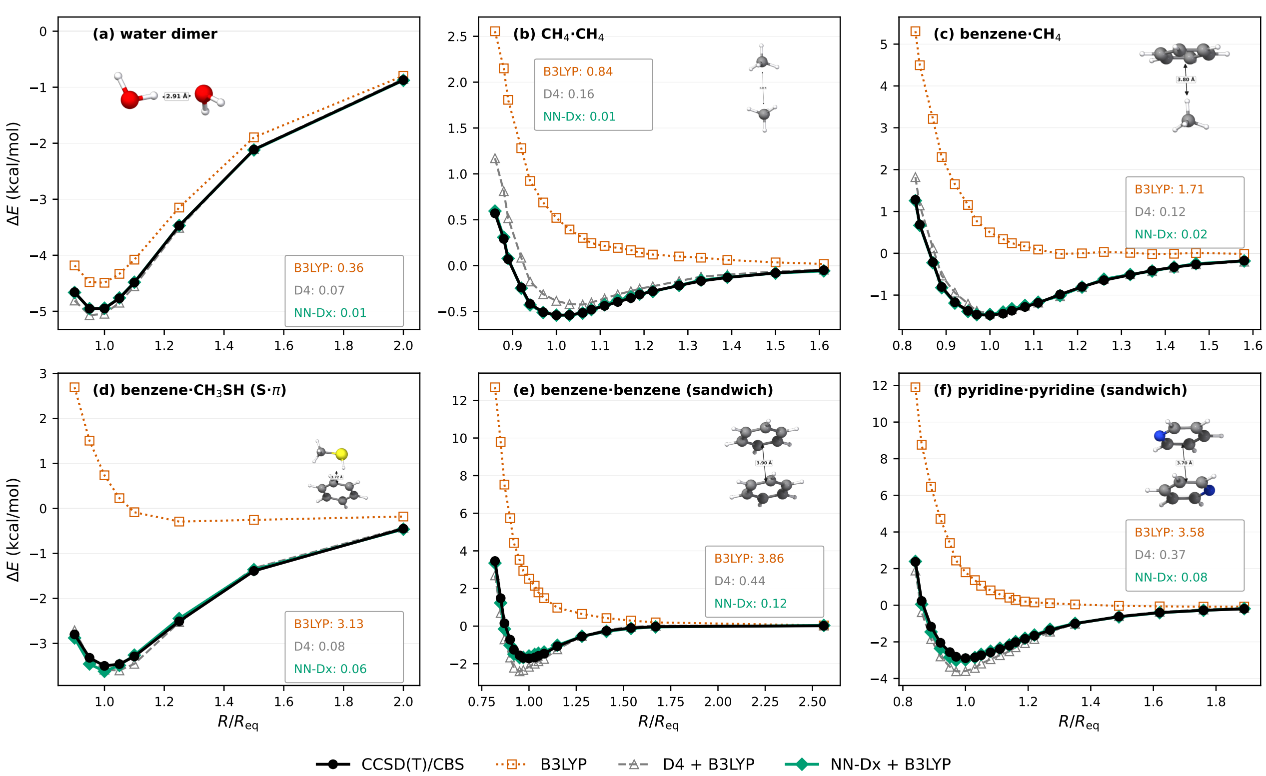

FIG. S2. Per-dimer binding-curve scans with a B3LYP base; conventions as in Supplementary Note 4.

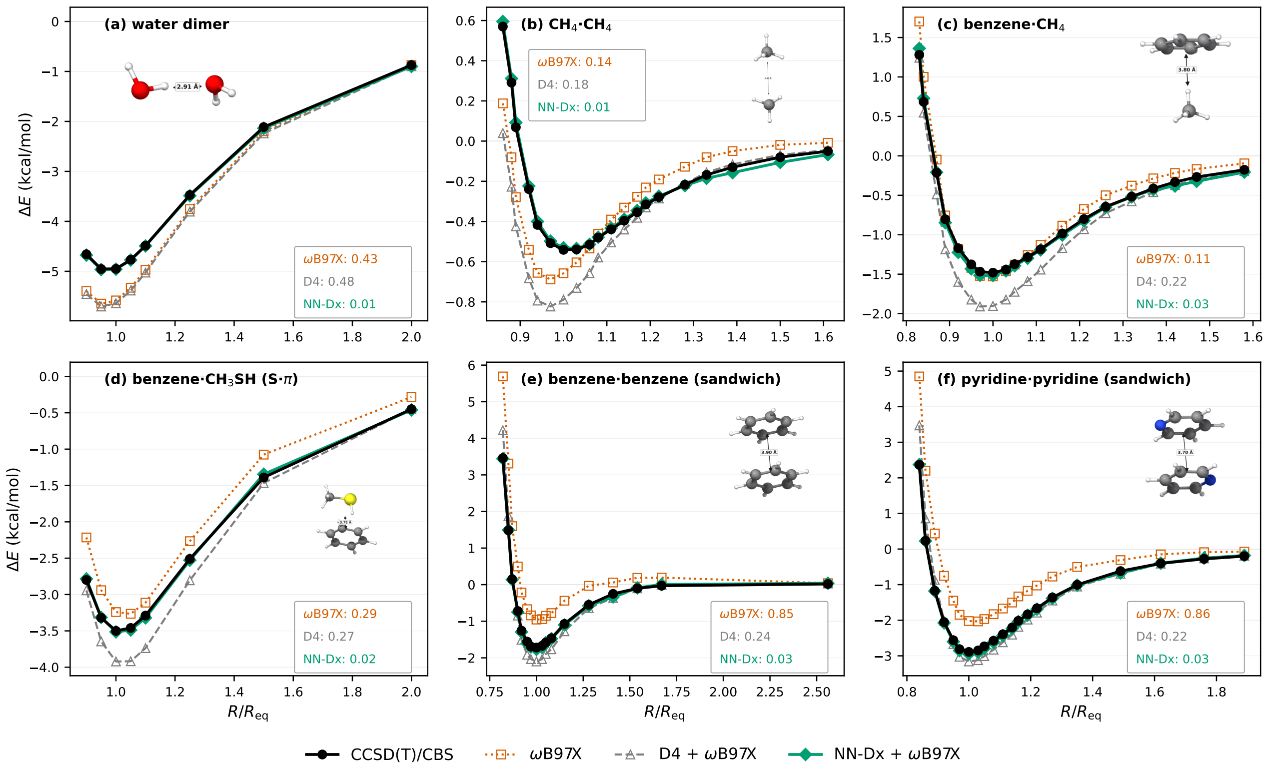

FIG. S3. Per-dimer binding-curve scans with an ωB97X base; conventions as in Supplementary Note 4.

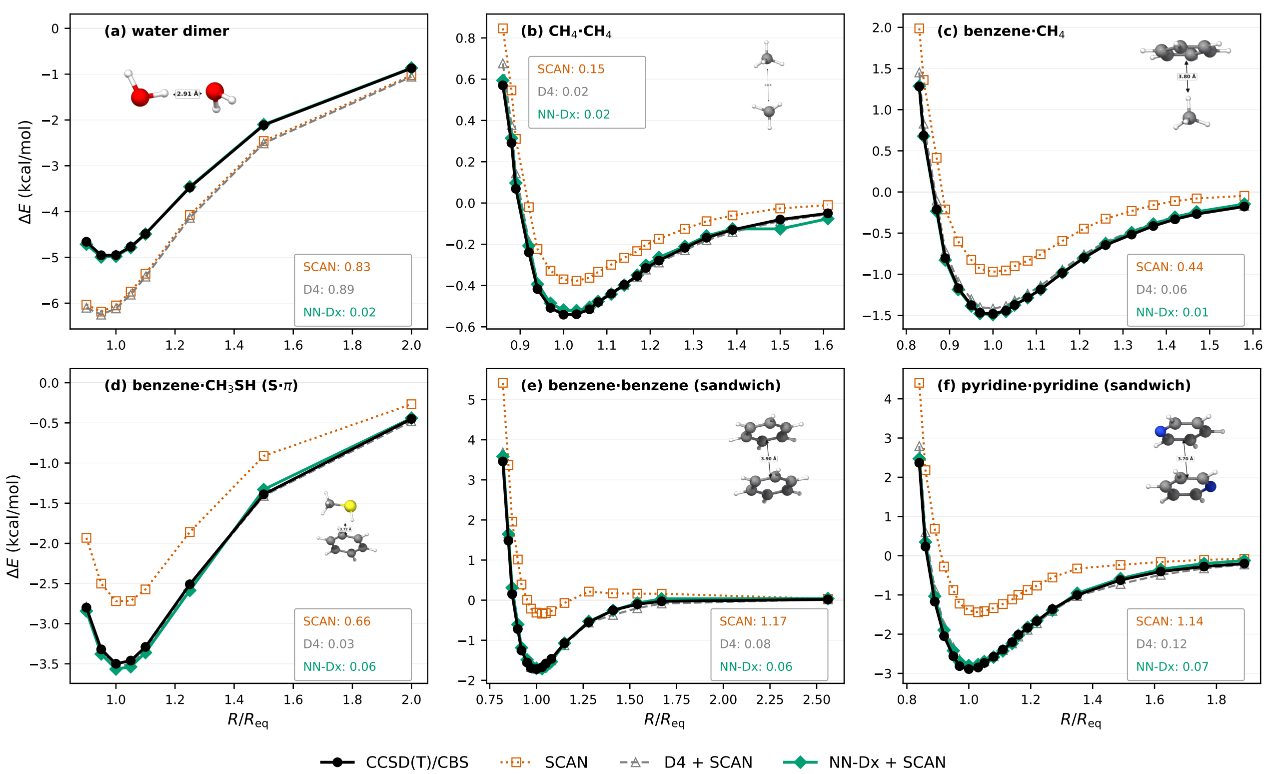

FIG. S4. Per-dimer binding-curve scans with a SCAN base; conventions as in Supplementary Note 4.

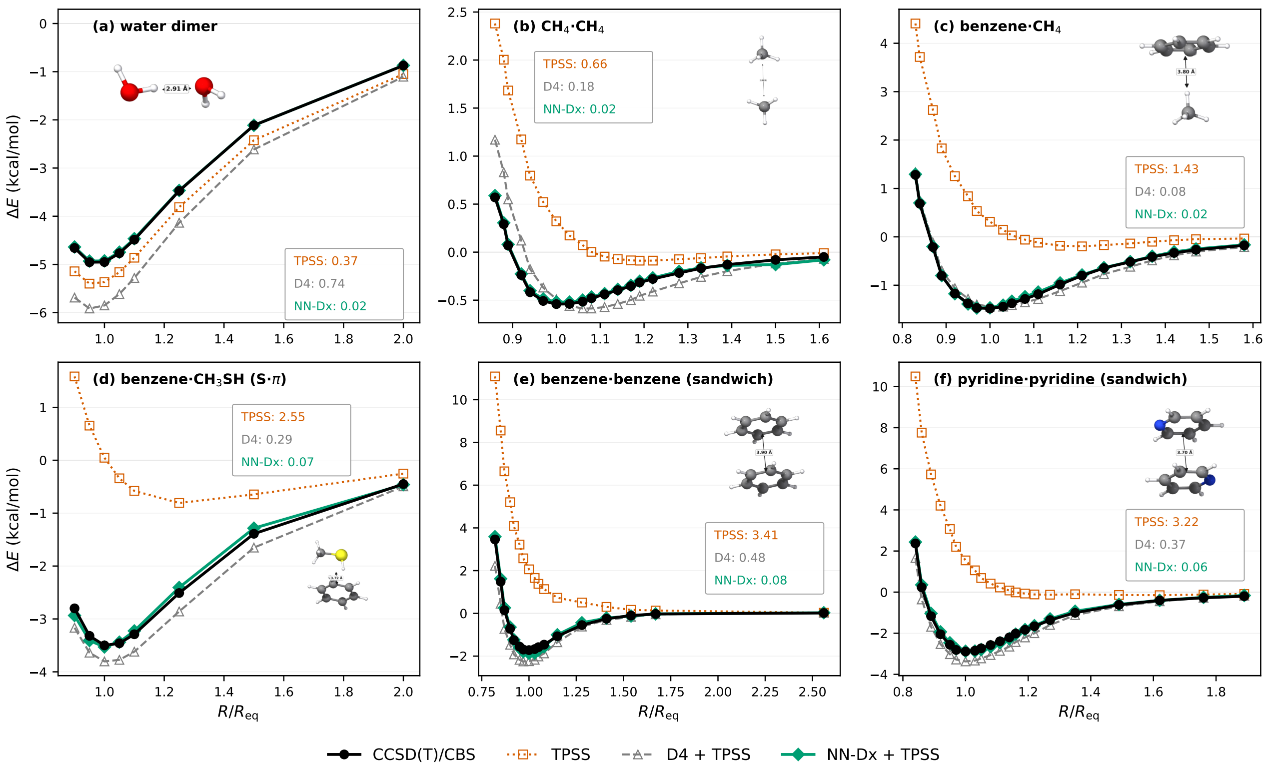

FIG. S5. Per-dimer binding-curve scans with a TPSS base; conventions as in Supplementary Note 4.

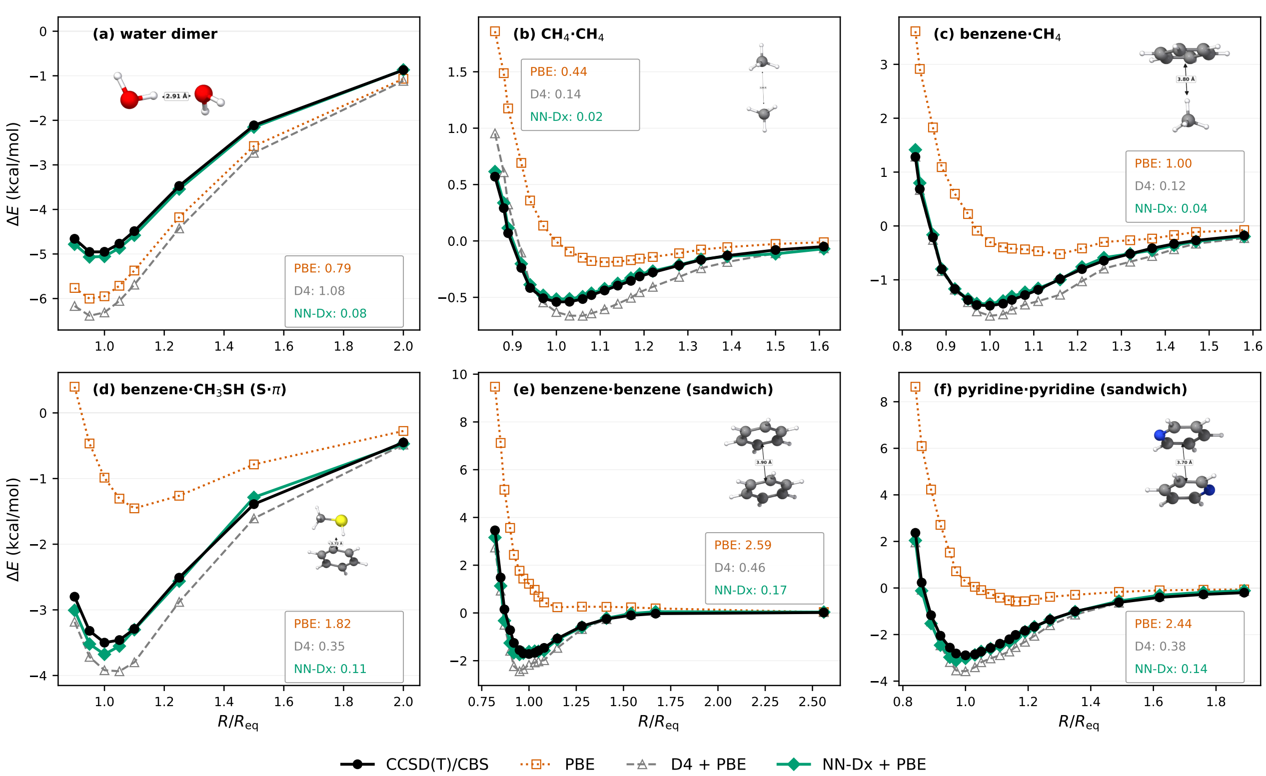

FIG. S6. Per-dimer binding-curve scans with a PBE base; conventions as in Supplementary Note 4.

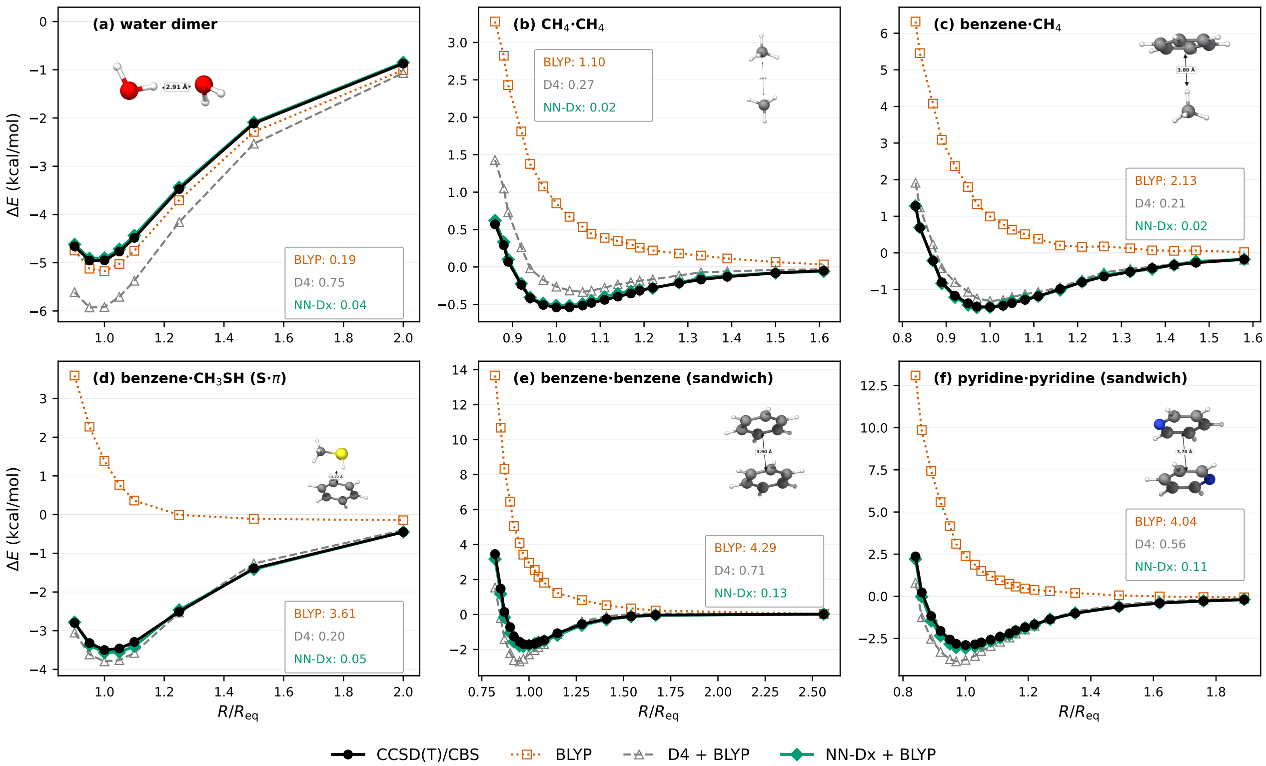

FIG. S7. Per-dimer binding-curve scans with a BLYP base; conventions as in Supplementary Note 4.

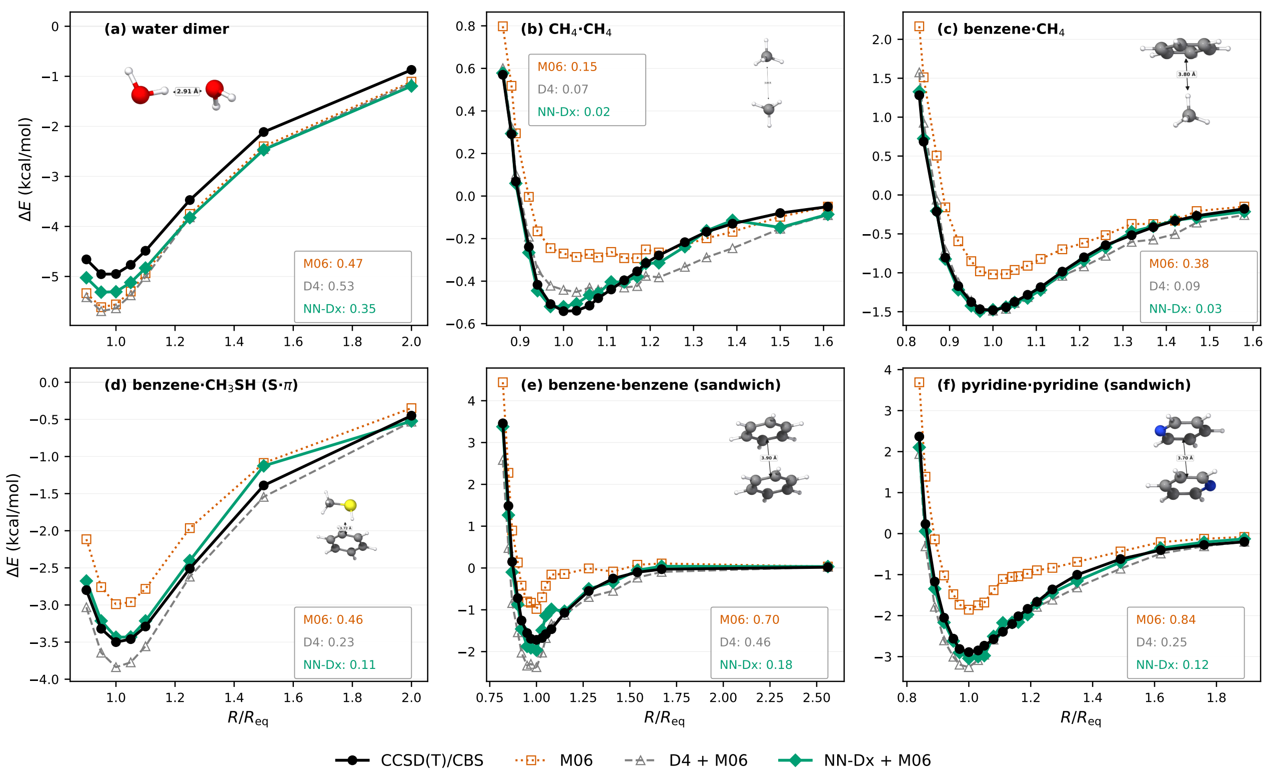

FIG. S8. Per-dimer binding-curve scans with an M06 base; conventions as in Supplementary Note 4.

TABLE S4. Settings for the force-differentiability and molecular-dynamics tests reported in the main text.

Setting	**Value**
Base method	PBE0/def2-TZVP (Q-Chem 6.3.0; `SCF_CONVERGENCE 11`, `THRESH 14`)
Integrator	velocity Verlet (ASE), microcanonical (NVE), $\Delta t = 0.25$ fs
Initial velocities	Maxwell–Boltzmann at 50 K, fixed seed, net momentum removed
Equilibration	first 100 fs discarded
Initial geometries	reference benchmark geometries ($r/R_{\mathrm{eq}} = 1.0$; not re-optimized)
Finite-difference step	10^{-4} Å (gradient); 2×10^{-4} Å (Hessian, differencing the analytic gradient)
Timestep-scaling test	$\Delta t = 0.5/0.25/0.125$ fs over a common 100–500 fs window

[1] A. G. Donchev, A. G. Taube, E. Decolvenaere, C. Hargus, R. T. McGibbon, K.-H. Law, B. A. Gregersen, J.-L. Li, K. Palmo, K. Siva, M. Bergdorf, J. L. Klepeis, and D. E. Shaw, Quantum chemical benchmark databases of gold-standard dimer interaction energies, Scientific Data **8**, 55 (2021).

[2] J. Řezáč and P. Hobza, Describing noncovalent interactions beyond the common approximations: How accurate is the "gold standard," CCSD(T) at the complete basis set limit?, Journal of Chemical Theory and Computation **9**, 2151 (2013).

[3] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, A look at the density functional theory zoo with the advanced gmtkn55 database for general main group thermochemistry, kinetics and noncovalent interactions, Physical Chemistry Chemical Physics **19**, 32184 (2017).

[4] S. Grimme, Seemingly simple stereoelectronic effects in alkane isomers and the implications for Kohn–Sham density functional theory, Angewandte Chemie International Edition **45**, 4460 (2006).

[5] J. Řezáč, K. E. Riley, and P. Hobza, Benchmark calculations of noncovalent interactions of halogenated molecules, Journal of Chemical Theory and Computation **8**, 4285 (2012).