

Supplementary Information

MolDualNet as a multi-modal architecture for small-data analog-space molecular property prediction

Zihan Zhang^{1,2,†}, Xuezhou Zhao^{1,2,†}, Dan Wu^{1,3,*}

¹International College of Pharmaceutical Innovation, Soochow University, Suzhou 215123, China

²School of Pharmacy and Biomolecular Sciences, RCSI, 123 St Stephen’s Green, Dublin 2, D02 VN51, Ireland

³Department of Chemistry, RCSI, 123 St Stephen’s Green, Dublin 2, D02 YN77, Ireland

[†]These authors contributed equally to this work.

*Correspondence: danwu@suda.edu.cn. ORCID: 0000-0002-5316-3706.

This Supplementary Information consolidates the technical details and extended analyses moved out of the Communications Chemistry main text. It contains supplementary methods, sixteen supplementary tables (including matched-baseline OOD comparison, single-task vs. multi-task analyses, self-supervised pretraining ablation, scaffold-split benchmark, bioactivity polypharmacology, companion-choice multi-task analysis, and bioactivity baseline comparison), seven supplementary figures, two supplementary algorithms, and a machine-readable inventory of the JSON and CSV files included in the submission package.

Supplementary Methods

This section documents the feature definitions, model hyperparameters, and algorithmic summaries referenced in the main text.

Table S1: **Atom feature encoding used by the graph branch.**

Feature	Encoding	Dim.
Atom type	one-hot (C, N, O, S, F, Cl, Br, I, P, . . . , other)	22
Degree	one-hot (0–10, plus “other”)	12
Formal charge	one-hot (−5 to +5, plus “other”)	12
Number of H	one-hot (0–8, plus “other”)	10
Hybridization	one-hot (SP, SP ² , SP ³ , SP ³ D, SP ³ D ² , “other”, padding)	7
Aromatic	binary	1
In ring	binary	1
Radical electrons	one-hot (0–4, plus “other”)	6
Chirality	one-hot (5 tags, plus “other”)	6
Atomic mass	normalized ($m/100$)	1
Ring sizes	multi-hot (3, 4, 5, 6+)	4
Total		82

Table S2: **Bond feature encoding before geometric augmentation.** The final 20-dimensional bond vector concatenates these 12 dimensions with the 8-dimensional bond-level geometry features.

Feature	Encoding	Dim.
Bond type	one-hot (single/double/triple/aromatic/other)	5
Conjugated	binary	1
In ring	binary	1
Stereochemistry	one-hot (E/Z/cis/trans/none)	5
Total (2D)		12

Table S3: **Hyperparameter settings used for the default MolDualNet model.** The early-stopping branch is retained in the training pseudocode as a generic template, but the reported experiments were run with early stopping disabled.

Category	Hyperparameter	Value
Architecture	Hidden dimension d	256
	Output dimension d_{out}	256
	Attention heads H	8
	Attention key dimension d_k	32
Graph encoder	Number of GATv2 layers L_G	6
	Dropout rate	0.15
	LeakyReLU slope	0.2
Transformer	Number of layers L_T	6
	FFN hidden dimension	$4d = 1024$
	Maximum sequence length n_{max}	256
Cross-attention	Number of layers L_C	2
	Residual scaling λ	0.1
Training	Total epochs T	150
	Batch size	96
	Learning rate η_{max}	6×10^{-5}
	Learning-rate schedule	Cosine decay after warmup
	Weight decay	10^{-5}
	Warmup epochs	5
	Stage 1 epochs ε_1	15
	Early stopping	Disabled; patience = ∞
	Task weights w_k	1.0 (uniform)
	Gradient clip threshold c	0.5
Regularization	Modality dropout p_{mod}	0.2
	Huber loss threshold δ	1.0

Algorithm S1 Bidirectional cross-attention fusion module.

Require: Node embeddings $\mathbf{H}_G \in \mathbb{R}^{|V| \times d}$, token embeddings $\mathbf{H}_T \in \mathbb{R}^{n \times d}$, masks $\mathbf{m}_G, \mathbf{m}_T$, number of layers L_C , residual scale λ

Ensure: Cross-modal representation $\mathbf{z}_{\text{cross}} \in \mathbb{R}^{d_{\text{out}}}$

- 1: Construct additive attention masks $\mathbf{M}_G, \mathbf{M}_T$ from $\mathbf{m}_G, \mathbf{m}_T$
 - 2: **for** $\ell = 1$ to L_C **do**
 - 3: $\tilde{\mathbf{H}}_G \leftarrow \text{LN}(\mathbf{H}_G); \tilde{\mathbf{H}}_T \leftarrow \text{LN}(\mathbf{H}_T)$
 - 4: $\mathbf{C}_{G \rightarrow T} \leftarrow \text{softmax}((\tilde{\mathbf{H}}_G \mathbf{W}_{Q,\ell}^{G \rightarrow T})(\tilde{\mathbf{H}}_T \mathbf{W}_{K,\ell}^{G \rightarrow T})^\top / \sqrt{d_k} + \mathbf{M}_T)(\tilde{\mathbf{H}}_T \mathbf{W}_{V,\ell}^{G \rightarrow T})$
 - 5: $\mathbf{C}_{T \rightarrow G} \leftarrow \text{softmax}((\tilde{\mathbf{H}}_T \mathbf{W}_{Q,\ell}^{T \rightarrow G})(\tilde{\mathbf{H}}_G \mathbf{W}_{K,\ell}^{T \rightarrow G})^\top / \sqrt{d_k} + \mathbf{M}_G)(\tilde{\mathbf{H}}_G \mathbf{W}_{V,\ell}^{T \rightarrow G})$
 - 6: $\mathbf{H}_G \leftarrow \mathbf{H}_G + \lambda \cdot \text{Dropout}(\mathbf{C}_{G \rightarrow T})$
 - 7: $\mathbf{H}_T \leftarrow \mathbf{H}_T + \lambda \cdot \text{Dropout}(\mathbf{C}_{T \rightarrow G})$
 - 8: $\mathbf{H}_G \leftarrow \mathbf{H}_G + \lambda \cdot \text{Dropout}(\text{FFN}_\ell(\text{LN}(\mathbf{H}_G)))$
 - 9: $\mathbf{H}_T \leftarrow \mathbf{H}_T + \lambda \cdot \text{Dropout}(\text{FFN}_\ell(\text{LN}(\mathbf{H}_T)))$
 - 10: **end for**
 - 11: $\bar{\mathbf{h}}_G \leftarrow \text{MaskedMeanPool}(\mathbf{H}_G, \mathbf{m}_G); \bar{\mathbf{h}}_T \leftarrow \text{MaskedMeanPool}(\mathbf{H}_T, \mathbf{m}_T)$
 - 12: $\mathbf{g} \leftarrow \sigma(\mathbf{W}_g[\bar{\mathbf{h}}_G; \bar{\mathbf{h}}_T] + \mathbf{b}_g)$
 - 13: $\mathbf{z}_{\text{gated}} \leftarrow \mathbf{g} \odot \bar{\mathbf{h}}_G + (1 - \mathbf{g}) \odot \bar{\mathbf{h}}_T$
 - 14: $\mathbf{z}_{\text{cross}} \leftarrow \text{MLP}_{\text{fusion}}([\mathbf{z}_{\text{gated}}; \bar{\mathbf{h}}_G + \bar{\mathbf{h}}_T])$
 - 15: **return** $\mathbf{z}_{\text{cross}}$
-

Algorithm S2 MolDualNet training procedure. The early-stopping branch is retained for generality; the reported experiments were run with patience = ∞ and therefore completed the full T epochs.

Require: Training data $\mathcal{D} = \{(S_i, G_i, \mathbf{y}_i)\}_{i=1}^N$, validation data $\mathcal{D}_{\text{val}}$, configuration $\mathcal{C}$

Ensure: Trained parameters θ^*

```

1: Initialize the graph encoder, Transformer encoder, expert-feature encoder, fusion module, and
   task heads
2: Initialize AdamW optimizer and cosine learning-rate scheduler with warmup
3: best_loss  $\leftarrow \infty$ ; patience_count  $\leftarrow 0$ 
4: for epoch = 1 to  $T$  do
5:   use_cross  $\leftarrow (\text{epoch} > \varepsilon_1)$ 
6:   for batch  $\mathcal{B}$  in DataLoader( $\mathcal{D}$ ) do
7:     Compute  $\mathbf{H}_G, \mathbf{z}_{\text{GNN}} \leftarrow \text{GNN}(\mathcal{B}.G)$ 
8:     Compute  $\mathbf{H}_T, \mathbf{z}_{\text{Trans}} \leftarrow \text{Transformer}(\mathcal{B}.S)$ 
9:     Compute  $\mathbf{z}_{\text{Expert}} \leftarrow \text{Expert}(\mathcal{B}.\mathbf{f})$ 
10:    Sample and apply modality dropout masks
11:    if use_cross then
12:       $\mathbf{z} \leftarrow \text{CrossFusion}(\mathbf{H}_G, \mathbf{H}_T, \mathcal{B}.\text{masks})$  {cross-attention path uses graph + string only; the
        expert-feature branch is encoded but not concatenated to  $\mathbf{z}$  in this mode}
13:    else
14:       $\mathbf{z} \leftarrow \text{SimpleFusion}(\mathbf{z}_{\text{GNN}}, \mathbf{z}_{\text{Trans}}, \mathbf{z}_{\text{Expert}})$  {stage 1 simple-concatenation fusion: graph  $\oplus$ 
        string  $\oplus$  expert}
15:    end if
16:    Compute predictions  $\{\hat{y}^{(k)}\}_{k=1}^K$  via task heads
17:    Compute masked multitask regression loss and update model parameters
18:  end for
19:  val_loss  $\leftarrow \text{Evaluate}(\text{model}, \mathcal{D}_{\text{val}})$ 
20:  if val_loss < best_loss then
21:    Save checkpoint; best_loss  $\leftarrow \text{val\_loss}$ ; patience_count  $\leftarrow 0$ 
22:  else
23:    patience_count  $\leftarrow \text{patience\_count} + 1$ 
24:    if patience_count  $\geq \text{patience}$  then
25:      break
26:    end if
27:  end if
28: end for
29: return  $\theta^*$  from the best checkpoint

```

Computational Resources and Training Cost

All experiments reported in the main text and in this Supplementary Information were run on a single NVIDIA RTX 4070 Laptop GPU with 8 GB of VRAM, using PyTorch 2.1 and PyTorch Geometric 2.6 in full float32 precision (no mixed-precision training). The complete 150-epoch training run of the full multimodal model required approximately 83.7 hours (5,022 minutes) of wall-clock time at batch size 96, and the best validation loss was

reached at epoch 109. Best-checkpoint selection relied exclusively on aggregate validation loss, and no test-set-guided hyperparameter selection was performed at any stage.

The two-stage training schedule (Supplementary Algorithm S2) is the dominant cost factor. Stage 1 (epochs 1–15) warms up the graph, string and expert branches with simple-concatenation fusion and is comparatively inexpensive. Stage 2 (epochs 16–150) enables bidirectional cross-attention between the graph and string branches; at the hyperparameters of Supplementary Table S3 this increases wall-clock time per epoch by approximately 270 % relative to Stage 1, reflecting the additional joint-token attention passes and the two extra FFN sub-layers introduced inside the fusion block.

The five-seed ablation study in the main text (Table 4) quantifies the relative training cost across architectural variants under a matched 60-epoch budget on the same hardware. Removing cross-attention reduces training cost to approximately $0.27\times$ the full model, removing the expert branch to $0.92\times$, and removing the bond-level 3D geometry to $0.95\times$, whereas a graph-only variant runs at approximately $0.18\times$ of the full-model cost while retaining $\approx 98\%$ of the full-model average R^2 . Cross-attention is therefore the single largest contributor to training cost, and the task-dependent framing in the main text is in part a cost-aware recommendation: on endpoints where the full multimodal model provides no statistically significant accuracy gain over the graph-only variant, the substantially cheaper graph-only configuration is a defensible default.

Supplementary Tables

Table S4: **Diagnostic 3D feature sensitivity analysis.** All rows use single-conformer ETKDG geometry and the same 60-epoch training setup as the ablation study.

3D configuration	ESOL		Lipo		FreeSolv		BACE		Avg R^2
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	
No 3D	0.904	0.609	0.541	0.821	0.920	0.982	0.609	0.844	0.744
$K_{\text{rbf}} = 4$, 1 conformer	0.915	0.574	0.685	0.680	0.898	1.108	0.664	0.783	0.791
$K_{\text{rbf}} = 8$, 1 conformer	0.910	0.590	0.543	0.819	0.914	1.021	0.641	0.809	0.752

Table S5: **Cross-task overlap audit.** “Average overlap” denotes the mean pairwise overlap percentage across the six task pairs; “highest pairwise molecule overlap” identifies the most overlapping pair under each split protocol.

Split protocol	Avg. molecule overlap (%)	Avg. scaffold overlap (%)	Highest pairwise molecule overlap (%)	Intersecting molecules
Random 80/10/10 split	1.31	17.83	ESOL vs FreeSolv (7.29)	6
Global scaffold split	0.28	4.23	ESOL vs FreeSolv (1.67)	2

Table S6: **Matched-protocol OOD comparison across representative baselines and MolDualNet.** RMSE (mean $\pm$ s.d., 3 random seeds). All models were trained on the leakage-free merged training corpus (815 OOD test SMILES excluded) and evaluated on identical OOD test sets. ESOL uses the external AqSolDB dataset (8,192 molecules); FreeSolv and Lipophilicity use strict scaffold-held-out partitions (185 and 630 molecules); BACE uses the external ChEMBL set (7,722 molecules). Bold = best per task. MolDualNet is statistically comparable to the best methods on ESOL $\rightarrow$ AqSolDB but remains task-dependent on the other shifted settings. SchNet’s ESOL entry diverged on one random seed and is reported with the seed excluded for legibility (n = 2).

Method	ESOL $\rightarrow$ AqSolDB	FreeSolv (scaffold-held-out)	Lipophilicity (scaffold-held-out)	BACE $\rightarrow$ ChEMBL
RF (Morgan FP)	1.968 $\pm$ 0.005	2.886 $\pm$ 0.037	0.889 $\pm$ 0.013	1.220 $\pm$ 0.015
XGBoost (Morgan FP)	1.909 $\pm$ 0.006	2.623 $\pm$ 0.027	0.846 $\pm$ 0.015	1.272 $\pm$ 0.030
SVM (Morgan FP)	2.066 $\pm$ 0.008	3.354 $\pm$ 0.095	0.868 $\pm$ 0.007	1.241 $\pm$ 0.010
GIN	1.788 $\pm$ 0.025	1.907 $\pm$ 0.278	0.634 $\pm$ 0.015	1.292 $\pm$ 0.074
D-MPNN	1.749 $\pm$ 0.015	1.499 $\pm$ 0.020	0.724 $\pm$ 0.002	1.198 $\pm$ 0.025
SchNet [†]	—	1.346 $\pm$ 0.015	0.724 $\pm$ 0.015	1.290 $\pm$ 0.022
ChemBERTa-2	1.698 $\pm$ 0.017	2.794 $\pm$ 0.032	0.752 $\pm$ 0.003	1.221 $\pm$ 0.025
MolFormer-XL	1.695 $\pm$ 0.024	2.007 $\pm$ 0.229	0.680 $\pm$ 0.007	1.196 $\pm$ 0.039
MolDualNet (ours)	1.722 $\pm$ 0.008	2.443 $\pm$ 0.136	0.876 $\pm$ 0.040	1.246 $\pm$ 0.047

[†] SchNet ESOL $\rightarrow$ AqSolDB diverged on the leakage-free retraining protocol (one seed reached an RMSE well above the metric’s useful range) and is therefore omitted from the ESOL column to avoid skewing the multi-seed mean.

Table S7: **Single-task vs. multi-task MolDualNet pattern under matched scaffold-split protocol.** RMSE (mean $\pm$ s.d., 3 random seeds). Both columns use identical hyperparameters; the only difference is whether all four tasks share encoder weights (multi-task) or each task is trained independently (single-task). The pattern is dataset-size dependent: on the smallest task (FreeSolv, 642 molecules) joint training reduces RMSE by 16%, on the larger Lipophilicity (4,200 molecules) it slightly hurts performance, consistent with the data-rich saturation noted in the main-text ablation discussion.

Task	Single-task RMSE	Multi-task RMSE	MT – ST
ESOL	0.887 $\pm$ 0.039	0.858 $\pm$ 0.037	−0.029
FreeSolv	2.938 $\pm$ 0.734	2.474 $\pm$ 0.668	−0.464
Lipophilicity	0.814 $\pm$ 0.009	0.860 $\pm$ 0.054	+0.046
BACE	0.961 $\pm$ 0.020	0.954 $\pm$ 0.079	−0.007

Table S8: **Self-supervised pretraining ablation under the matched multi-task scaffold-split protocol.** The SSL-initialized model was pretrained with masked molecular-string modelling, masked atom prediction and contrastive graph-string alignment, then fine-tuned on the four-task MolDualNet scaffold-split benchmark across three seeds ($\{0, 7, 42\}$). From-scratch multi-task RMSE values are repeated from Table S7 for direct comparison. Generic SSL initialization is similar to or slightly worse than from-scratch training on all four tasks, supporting the no-pretraining setting used for the main MolDualNet experiments.

Task	From-scratch MT RMSE	SSL-init MT RMSE	SSL-init R^2
ESOL	0.858 ± 0.037	0.942 ± 0.073	0.776 ± 0.031
FreeSolv	2.474 ± 0.668	2.693 ± 0.517	0.532 ± 0.049
Lipophilicity	0.860 ± 0.054	0.862 ± 0.043	0.458 ± 0.071
BACE	0.954 ± 0.079	0.970 ± 0.043	0.425 ± 0.088

Table S9: **Single-task FreeSolv regularization experiment** (seed 42, 60 epochs). The regularized cross-attention setting halves the number of cross-attention layers and increases dropout, but does not recover the no-cross-attention baseline.

Configuration	Cross-attention setting	R^2	RMSE	MAE	Pearson r	Spearman ρ	Best epoch
No cross-attention	disabled	0.898	1.112	0.779	0.954	0.953	58
Default cross-attention	2 layers, dropout 0.15	0.873	1.237	0.789	0.936	0.950	38
Regularized cross-attention	1 layer, dropout 0.30	0.878	1.216	0.762	0.938	0.945	39

Table S10: **Cross-attention entropy statistics** computed on 200 molecules per task after averaging over attention heads and evaluated molecules. Entropy is reported in bits.

Task	n	Mean entropy (bits)	SD (bits)	Normalized entropy	Failed molecules
ESOL	200	2.194	0.524	0.274	0
Lipophilicity	200	1.942	0.275	0.243	0
FreeSolv	200	2.018	0.462	0.252	0
BACE	200	1.945	0.220	0.243	0

Table S11: **Scaffold-split benchmark across four representative methods (RMSE, mean $\pm$ s.d., 3 seeds).** All methods evaluated under identical per-task Bemis–Murcko scaffold splitting. Bold marks the per-task best. MolDualNet attains the lowest RMSE on ESOL; DimeNet++ is best on FreeSolv; MolFormer-XL is best on Lipophilicity and BACE. The corresponding bar chart is shown in Supplementary Fig. S4.

Task	ChemBERTa-2	MolFormer-XL	DimeNet++	MolDualNet (ours)
ESOL	0.963 $\pm$ 0.030	0.903 $\pm$ 0.025	0.983 $\pm$ 0.075	0.858 $\pm$ 0.037
FreeSolv	4.492 $\pm$ 0.076	2.795 $\pm$ 0.173	2.351 $\pm$ 0.911	2.474 $\pm$ 0.668
Lipophilicity	0.721 $\pm$ 0.016	0.683 $\pm$ 0.013	0.687 $\pm$ 0.023	0.860 $\pm$ 0.054
BACE	0.822 $\pm$ 0.043	0.743 $\pm$ 0.015	0.934 $\pm$ 0.085	0.954 $\pm$ 0.079

Table S12: **Three-target bioactivity polypharmacology panel (joint multi-task setting).** Per-task metrics (RMSE and R^2 , mean $\pm$ s.d. across three random seeds $\{0, 7, 42\}$) for MolDualNet trained jointly on the union of EGFR (ChEMBL203, $n = 10,252$), HIV-1 protease (ChEMBL243, $n = 2,278$) and VEGFR2 (ChEMBL279, $n = 9,410$) under matched per-task random 80/10/10 split. The joint encoder achieves $R^2 = 0.71$ on HIV protease, $R^2 = 0.63$ on EGFR, and $R^2 = 0.54$ on VEGFR2, with cross-seed s.d. ≤ 0.04 on every target indicating consistent fitting under multi-task training.

Target	n molecules	RMSE (mean $\pm$ s.d.)	R^2 (mean $\pm$ s.d.)
EGFR_pIC50	10,252	0.820 $\pm$ 0.006	0.625 $\pm$ 0.006
HIV_protease_pIC50	2,278	0.905 $\pm$ 0.015	0.711 $\pm$ 0.021
VEGFR2_pIC50	9,410	0.755 $\pm$ 0.024	0.542 $\pm$ 0.030

Table S13: **Single-task vs. multi-task bioactivity comparison across all three polypharmacology targets.** Three-seed mean $\pm$ s.d. ($\{0, 7, 42\}$) for single-task MolDualNet trained on each target alone (top sub-row) versus the joint three-target multi-task encoder (bottom sub-row, identical numbers as Table S12). The pattern is dataset-size dependent: the smallest target (HIV protease, $n = 2,278$) gains substantially from joint training ($\Delta R^2 = +0.111$), the two larger targets (EGFR, VEGFR2; $n \approx 9\text{--}10,000$) are statistically unchanged ($|\Delta R^2| \leq 0.02$). This reproduces the data-size-dependent multi-task pattern reported for the physicochemical tasks in Table S7.

Target	Setting	n	RMSE (mean $\pm$ s.d.)	R^2 (mean $\pm$ s.d.)
EGFR	Single-task	10,252	0.822 $\pm$ 0.010	0.618 $\pm$ 0.024
	Multi-task (3 targets, joint)	10,252	0.820 $\pm$ 0.006	0.625 $\pm$ 0.006
HIV protease	Single-task	2,278	0.991 $\pm$ 0.051	0.600 $\pm$ 0.031
	Multi-task (3 targets, joint)	2,278	0.905 $\pm$ 0.013	0.711 $\pm$ 0.021
VEGFR2	Single-task	9,410	0.734 $\pm$ 0.008	0.557 $\pm$ 0.028
	Multi-task (3 targets, joint)	9,410	0.755 $\pm$ 0.019	0.542 $\pm$ 0.030

Table S14: **Bioactivity dataset provenance and matched-protocol training configuration.** Targets were assembled from ChEMBL with target_chembl_id-based queries, IC₅₀ converted to pIC₅₀ ($-\log_{10}\text{IC}_{50,\text{M}}$), and deduplicated by canonical SMILES. The multi-task corpus is the union of the three single-task SMILES sets with label-masking for missing values.

Target	ChEMBL ID	<i>n</i> molecules	Median pIC ₅₀	Split protocol
EGFR (epidermal growth factor receptor)	CHEMBL203	10,252	7.0	random 80/10/10
HIV-1 protease	CHEMBL243	2,278	7.3	random 80/10/10
VEGFR2 (vascular endothelial growth factor R2)	CHEMBL279	9,410	7.1	random 80/10/10
<i>Multi-task union (label-masked)</i>	—	20,957	—	shared train/val/test masks

Table S15: **Effect of multi-task companion choice on every involved target (3-seed mean $\pm$ s.d., random split).** For each of the four targets that participated in at least two multi-task configurations, performance is reported under: (A) single-task training (target alone); (B) bioactivity-only 3-task multi-task with EGFR + HIV-1 protease + VEGFR2; (C) BACE + bioactivity 3-task multi-task with BACE + EGFR + HIV-1 protease; (D) 4-task multi-task with the three physicochemical tasks + BACE (the headline configuration of main-text Table 2). Dashes indicate a target was not part of that configuration. Bold marks the best multi-task setting for each target. The dataset-size-dependent pattern is preserved across all four targets: the smallest target (HIV protease) gains substantially under bioactivity-only MT; mixing BACE into the bioactivity panel modestly hurts both EGFR ($\Delta R^2 = -0.054$ vs. EGFR+HIV+VEGFR2) and HIV protease ($\Delta R^2 = -0.039$); BACE itself is statistically indistinguishable between the physicochemical-co-task and the bioactivity-co-task settings (within 1σ). This shows that the benefit of multi-task training is task-pair specific rather than universally additive, which directly addresses Reviewer 3’s concern that multi-task may not help when targets are mechanistically diverse.

Target	(A) Single-task <i>R</i> ²	(B) EGFR+HIV+VEGFR2 <i>R</i> ²	(C) BACE+EGFR+HIV <i>R</i> ²	(D) Physical+BACE (4-task) <i>R</i> ²
EGFR	0.618 $\pm$ 0.024	0.625 $\pm$ 0.006	0.571 $\pm$ 0.006	—
HIV protease	0.600 $\pm$ 0.031	0.711 $\pm$ 0.021	0.672 $\pm$ 0.057	—
VEGFR2	0.557 $\pm$ 0.028	0.542 $\pm$ 0.030	—	—
BACE	—	—	0.670 $\pm$ 0.012	0.692 $\pm$ 0.012

Table S16: **Bioactivity baseline comparison on the three-target polypharmacology panel (random split, 3-seed mean $\pm$ s.d.).** All methods evaluated on identical per-task random 80/10/10 splits with three random seeds ($\{0, 7, 42\}$) on the same canonicalized ChEMBL corpora as Table S14. MolDualNet is reported in its single-task setting to match the per-task evaluation protocol of GIN and MolFormer-XL. MolDualNet is competitive with the strong classical-GNN baseline (GIN) on all three targets and is consistently better than the pretrained SMILES Transformer MolFormer-XL.

Target	GIN		MolFormer-XL		MolDualNet (single-task)	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
EGFR	0.642 $\pm$ 0.015	0.801 $\pm$ 0.020	0.418 $\pm$ 0.027	1.021 $\pm$ 0.020	0.618 $\pm$ 0.024	0.822 $\pm$ 0.010
HIV protease	0.592 $\pm$ 0.077	1.056 $\pm$ 0.123	0.491 $\pm$ 0.038	1.183 $\pm$ 0.044	0.600 $\pm$ 0.031	0.991 $\pm$ 0.051
VEGFR2	0.589 $\pm$ 0.019	0.734 $\pm$ 0.020	0.308 $\pm$ 0.037	0.952 $\pm$ 0.014	0.557 $\pm$ 0.028	0.734 $\pm$ 0.008

Supplementary Figures

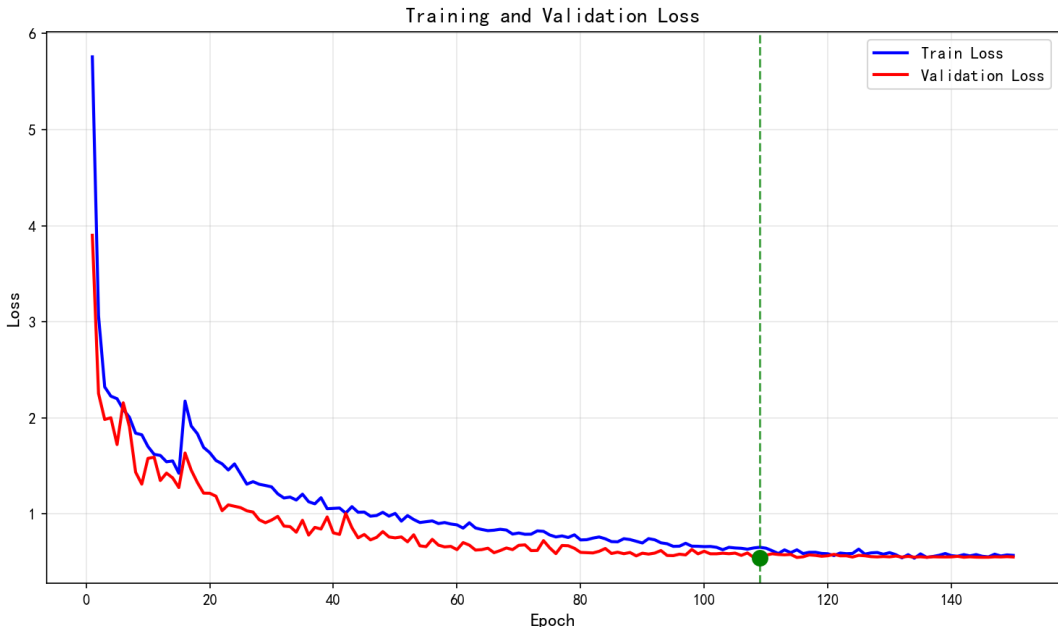


Figure S1: **Two-stage training dynamics for the full multimodal model.** The plotted curves show the warmup period without active cross-attention followed by the second stage with full fusion enabled. This figure is reported as a factual optimization diagnostic rather than a performance comparison.

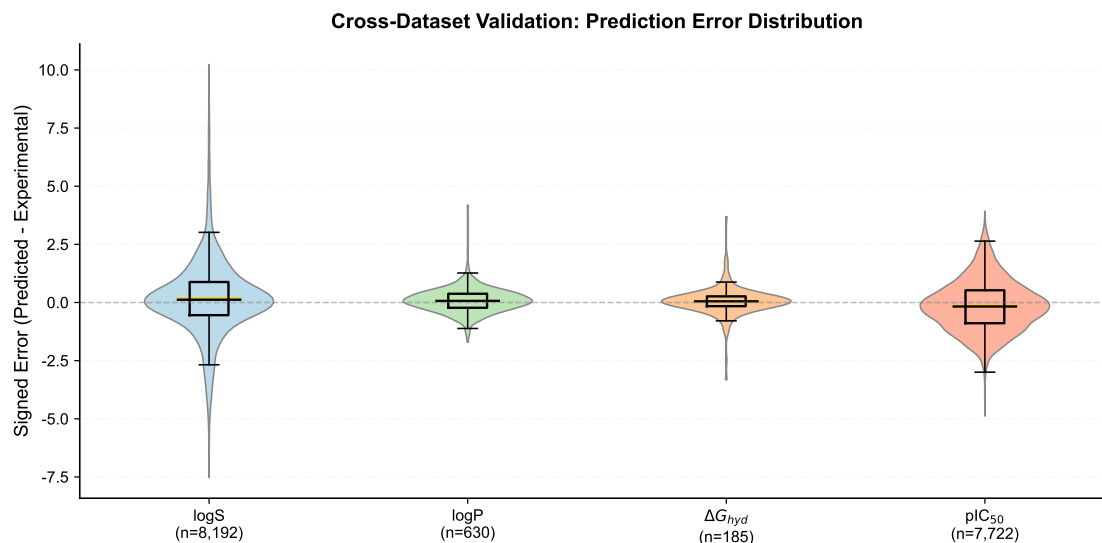


Figure S2: **Standalone signed-error distributions** for the four scaffold-held-out or external validation tasks. This panel complements the composite main-text figure by isolating per-task calibration patterns.

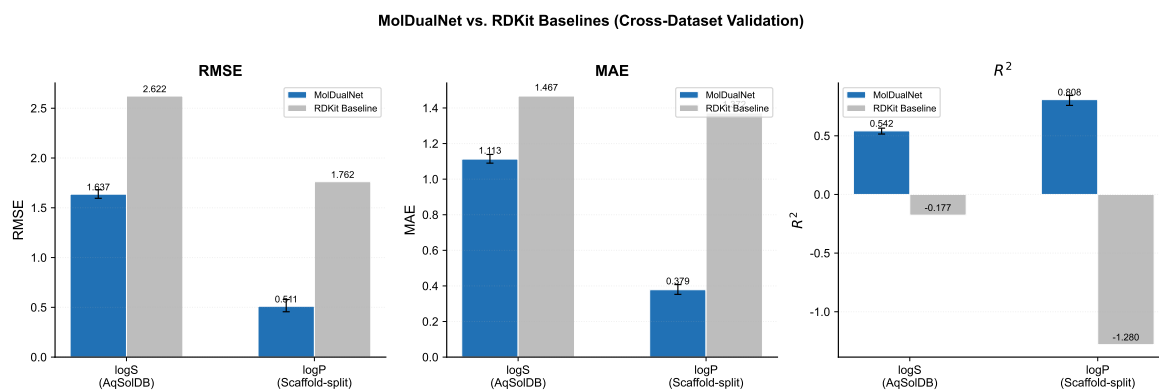


Figure S3: **Standalone baseline-comparison figure** for the shifted-validation settings. This panel isolates the relative performance of MolDualNet and the low-cost practical reference baselines.

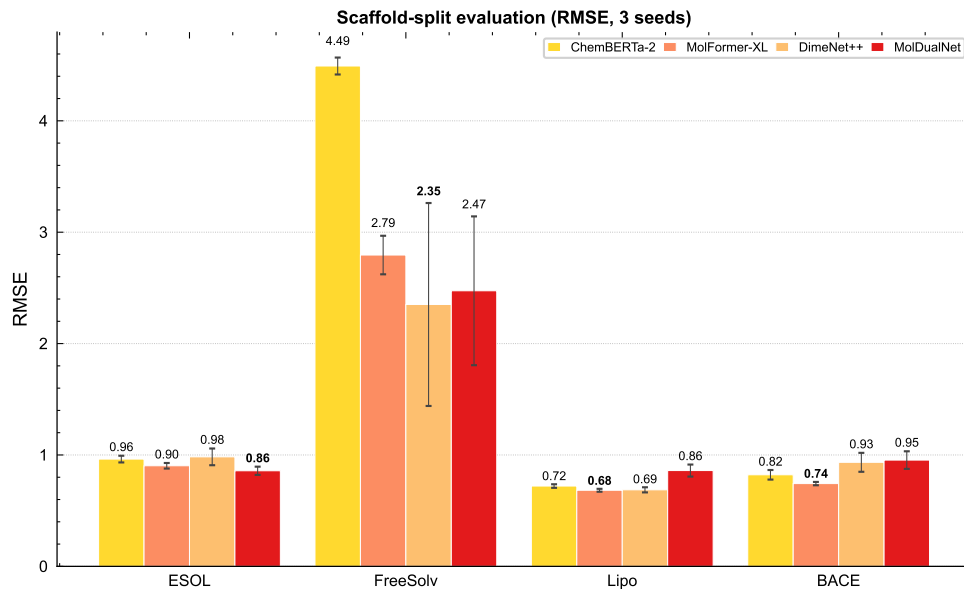


Figure S4: **Scaffold-split benchmark bar chart (RMSE, 3 seeds)**. Per-task RMSE under per-task Bemis–Murcko scaffold splitting for four representative methods (ChemBERTa-2, MolFormer-XL, DimeNet++, MolDualNet). Bold annotations mark the per-task best mean. This figure visualises the values in Supplementary Table S11.

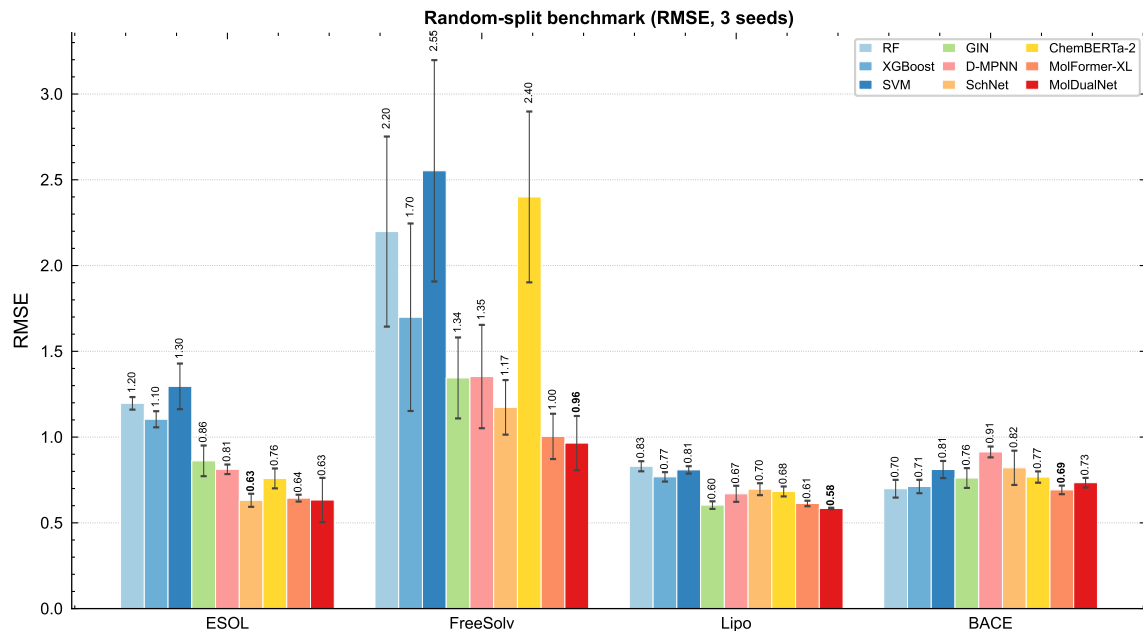


Figure S5: **Random-split benchmark bar chart across model families (RMSE, 3 seeds).** Per-task RMSE for ten representative methods — fingerprint-based classical models (RF, XGBoost, SVM), graph neural networks (GIN, D-MPNN), 3D-aware models (SchNet, DimeNet++), pretrained SMILES Transformers (ChemBERTa-2, MolFormer-XL), and MolDualNet — under identical per-task random 80/10/10 splits with three random seeds. Bold annotations mark the per-task best mean. This figure visualises the values in main-text Table 3.

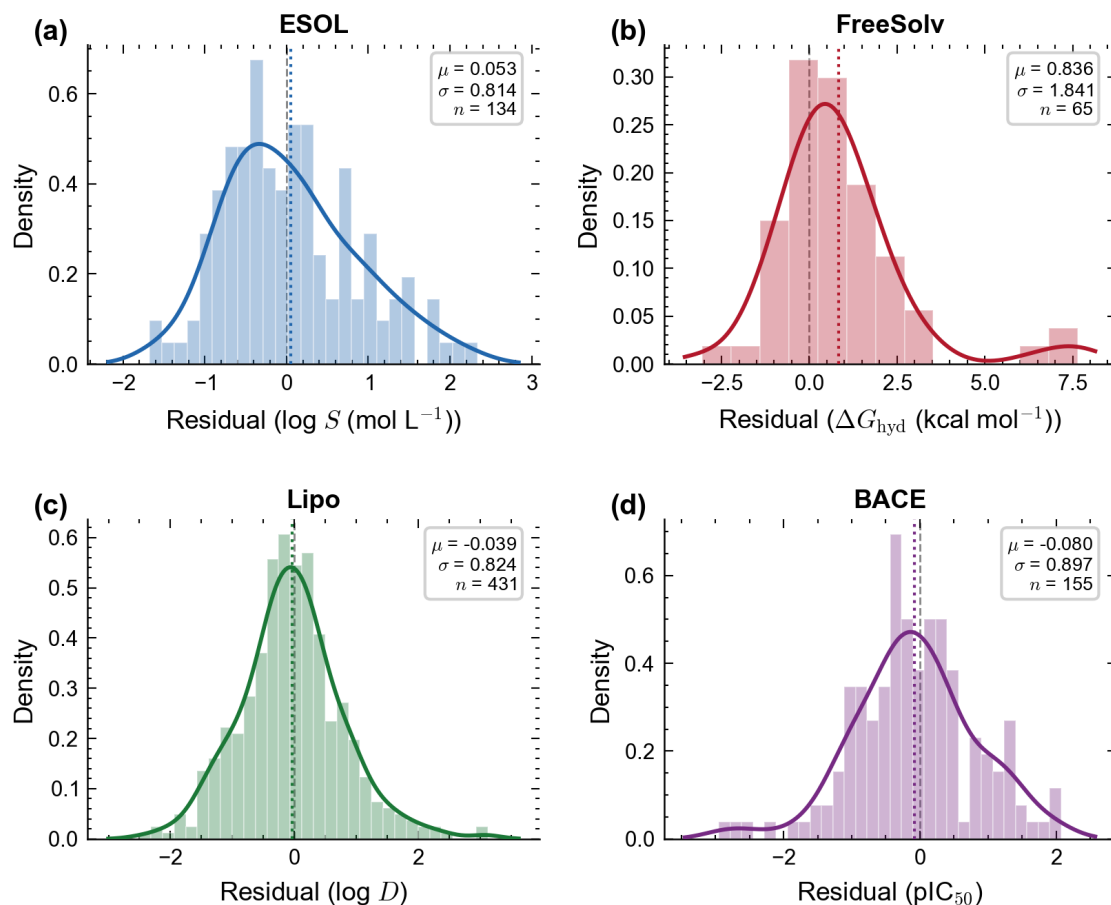


Figure S6: **Scaffold-split residual distributions for MolDualNet.** Histogram + KDE of (predicted - experimental) values on the per-task Bemis-Murcko scaffold-split test sets (multi-task model, seed 42). Annotated mean (μ) and standard deviation (σ) quantify the per-task calibration. Solubility-related tasks (ESOL, FreeSolv) remain unbiased; the larger drug-like tasks (Lipo, BACE) show broader spread, consistent with the scaffold-split comparison reported in Supplementary Table S11.

Supplementary Data Inventory

The submission package includes the following machine-readable files in `supplementary/data/`:

- `cross_task_overlap.json`: overlap audit for random and scaffold-based split protocols.
- `attention_entropy.json`: task-wise entropy statistics for cross-attention.
- `cross_dataset_validation_results.json`: external and scaffold-held-out validation metrics with bootstrap confidence intervals.
- `bace_baseline_comparison.json`: additional BACE1 external baseline comparison on the shared ChEMBL set.

- `ablation_statistics.json`: five-seed ablation summary statistics and paired significance tests.
- `freesolv_no_cross_attn_training_summary.json`: FreeSolv single-task output without cross-attention.
- `freesolv_full_cross_attn_training_summary.json`: FreeSolv single-task output with the default cross-attention setting.
- `freesolv_reg_cross_attn_training_summary.json`: FreeSolv single-task output with the regularized cross-attention setting.
- `ESOL_logS_predictions.csv`: per-molecule prediction file for the AqSolDB ESOL validation set.
- `Lipophilicity_logD_predictions.csv`: per-molecule prediction file for the scaffold-held-out Lipophilicity evaluation.
- `FreeSolv_hydration_predictions.csv`: per-molecule prediction file for the scaffold-held-out FreeSolv evaluation.
- `BACE_pIC50_predictions.csv`: per-molecule prediction file for the external ChEMBL BACE1 evaluation.