Supporting Information

Accurate quantum chemical property prediction with transferable electron-density representations

Xiang et al.

Table of Contents

Supplementary Section 1	**Drug-relevant molecular classification datasets**	**23**
Supplementary Section 2	**Mechanism-relevant motif attribution in Mpro/3CLpro and CYP3A4**	**23**
Supplementary Section 3	**Retrospective DrugBank screening for Mpro/3CLpro**	**23**
Supplementary Section 4	**External evaluation of literature-supported KRAS G12D-targeting compounds**	**24**
Supplementary Section 5	**Full attention statistics across chemical categories**	**24**
Supplementary Section 6	**Correlation between learned attention and local electron density**	**25**
Supplementary Section 7	**Retrieval validation of electronic memory**	**26**
Supplementary Section 8	**Computational efficiency**	**26**
Supplementary Section 9	**Definition of the electron-probe descriptor**	**27**
Supplementary Section 10	**Training details**	**27**
Supplementary Section 11	**Baseline reproduction and transfer settings**	**28**
Supplementary Section 12	**Supplementary Tables**	**31**

Supplementary Section 1. Drug-relevant molecular classification datasets

Compound–target activity records were collected from ChEMBL for SARS-CoV-2 main protease (Mpro/3CLpro), KRAS G12D, BTK and CYP3A4. Only records reporting IC_{50} values were retained. Molecular structures were standardized by canonicalizing SMILES, removing salts, eliminating duplicate structures and applying a consistent molecular standardization procedure. Optimized three-dimensional conformations were generated with RDKit when three-dimensional structures were unavailable. Compounds with $IC_{50} \leq 100\,\text{nM}$ were labelled active ($y = 1$), whereas compounds with $IC_{50} > 100\,\text{nM}$ were labelled inactive ($y = 0$). Each dataset was divided into training, validation and test sets using a scaffold split with a ratio of 8:1:1. eMol, MotiL and JMM were trained and evaluated using the same dataset splits, and classification performance was assessed using AUROC and AUPRC. The resulting datasets contain between 1,370 and 11,893 compounds, with positive ratios ranging from 3.55% to 69.18% and imbalance ratios ranging from 1.96:1 to 27.15:1 (**Supplementary Table 1**).

Supplementary Section 2. Mechanism-relevant motif attribution in Mpro/3CLpro and CYP3A4

We selected Mpro/3CLpro and CYP3A4 as two complementary settings for evaluating whether model attribution reflects mechanism-relevant local electronic environments. Mpro/3CLpro represents target inhibition, for which electrophilic carbonyl warheads and neighbouring polar recognition fragments can contribute to ligand binding and covalent reactivity. CYP3A4 represents metabolism-related molecular recognition, for which heteroatom-rich, ionizable and oxidation-prone motifs influence enzyme binding, inhibition and metabolic transformation. These two tasks therefore examine attribution under distinct mechanisms involving target engagement and metabolic-enzyme recognition.

For Mpro/3CLpro, we analysed piperazine, piperidine, acrylamide and three-carbon linker motifs. The nitrogen-containing heterocycles and linkers represent polar or spatial recognition environments, whereas acrylamide contains an electrophilic α,β-unsaturated carbonyl that can act as a Michael acceptor. For CYP3A4, we analysed methoxy/alkoxy groups, tertiary amines, morpholine and a broader metabolism-relevant motif category. These motifs capture heteroatom-rich, ionizable and potentially oxidation-sensitive local environments commonly involved in CYP3A4 recognition. For each motif, enrichment was calculated across all unique molecules containing the corresponding substructure. The sample size n denotes the number of unique motif-containing molecules. Values are reported as mean enrichment, followed in parentheses by the lower and upper bounds of the 95% confidence interval obtained from 5,000 bootstrap resamples. Under the adopted normalization, 1.0 represents the reference enrichment level.

eMol achieved the highest mean enrichment for all four Mpro/3CLpro motifs, with values ranging from 1.232 to 1.411 and all 95% confidence intervals remaining above 1.0 (**Supplementary Table 2**). Notably, its enrichment for acrylamide was 1.237 (1.044–1.476), whereas the confidence intervals for MotiL and JMM included or remained below the reference level. eMol also achieved the highest enrichment for all four CYP3A4 motif categories, with values ranging from 1.137 to 1.415 and all confidence intervals above 1.0 (**Supplementary Table 4**). The strongest effect was observed for methoxy/alkoxy groups, for which eMol reached 1.415 (1.364–1.466).

Supplementary Section 3. Retrospective DrugBank screening for Mpro/3CLpro

Mpro/3CLpro was selected for retrospective virtual screening because it is a well-established antiviral target with experimentally characterized inhibitors, providing a suitable setting in which to assess whether molecular activity prediction can translate into practical candidate prioritization beyond held-out classification benchmarks. DrugBank was used as the screening library to evaluate whether the trained models could recover compounds with independently reported evidence of Mpro/3CLpro inhibition or anti-SARS-CoV-2 activity.

To prevent overlap with model-development data, DrugBank compounds present in the ChEMBL-derived datasets were excluded before screening. The remaining compounds were scored independently by the Mpro/3CLpro classifiers trained with eMol, MotiL and JMM and ranked according to their respective prediction scores. Because the objective was to assess early enrichment, experimental evidence was examined for the ten highest-ranked candidates from each model. Literature evidence was classified into four categories. Direct Mpro/3CLpro enzymatic inhibition was denoted by E. Evidence of Mpro/3CLpro binding together with cell-based antiviral activity, but without direct enzymatic inhibition measurements, was denoted by B/C. Cell-based anti-SARS-CoV-2 activity without direct evidence of Mpro/3CLpro engagement was denoted by C. Compounds for which no relevant evidence was identified in the targeted literature search were denoted by N. An N designation indicates only that no supporting evidence was identified under the adopted search procedure and does not establish that the compound is inactive.

Using these criteria, eMol recovered two literature-supported compounds among its top ten candidates: Brilacidin, which showed cell-based anti-SARS-CoV-2 activity, and Pomotrelvir, which directly inhibited Mpro/3CLpro. MotiL recovered one supported compound, Vorapaxar, for which Mpro/3CLpro binding and cell-based antiviral activity had been reported, whereas no supported compound was identified among the JMM top ten candidates (**Supplementary Table 5**). These results indicate that the eMol ranking was more strongly enriched at the top for compounds with experimental evidence relevant to Mpro/3CLpro inhibition or antiviral activity.

Supplementary Section 4. External evaluation of literature-supported KRAS G12D-targeting compounds

KRAS G12D was selected for external evaluation because it is a therapeutically important oncogenic driver for which several chemically and mechanistically distinct agents have recently been reported. In contrast to the retrospective DrugBank screening for Mpro/3CLpro, this analysis was designed as a focused positive-compound recovery test. It examined whether the trained classifiers assigned high activity scores to recently reported KRAS G12D-targeting compounds spanning mutation-selective inhibition, targeted protein degradation and broader pan-KRAS inhibition.

Candidate compounds were identified through a targeted Google Scholar search using the query `KRAS G12D inhibitor source:"Nature" OR source:"Science" OR source:"Cell"`. All five result pages available at the time of analysis, comprising approximately 50 records, were manually reviewed. Titles and abstracts were screened, and a compound was retained only if it was explicitly identified by name, a defined chemical structure was available and the corresponding study reported wet-lab evidence of KRAS G12D inhibition, degradation or pharmacological activity in a KRAS G12D model. Reviews, perspectives, news articles, purely computational studies, unnamed compounds and studies lacking explicit experimental evidence were excluded. Canonical SMILES and reported potency measurements were subsequently collected for the retained compounds.

This procedure yielded four compounds with distinct pharmacological mechanisms. HRS-4642 and RMC-9805 (Zoldonrasib) are mutation-selective KRAS G12D inhibitors, whereas ASP3082 is a selective KRAS G12D-targeted protein degrader. MCB-294 was included as a dual-state pan-KRAS inhibitor with experimentally demonstrated activity in KRAS G12D models. These compounds were treated as positive reference examples and independently scored using the KRAS G12D classifiers trained with eMol, JMM and MotiL. Positive predictions were defined using the same classification threshold of 0.5 applied in the benchmark evaluation.

eMol assigned scores greater than or equal to 0.5 to HRS-4642, RMC-9805 and MCB-294, recovering three of the four literature-supported compounds, compared with two for JMM and none for MotiL (**Supplementary Table 6**). Notably, eMol was the only model to classify RMC-9805 as positive and successfully recognized compounds spanning mutation-selective and pan-KRAS inhibition. These results demonstrate that eMol generalizes more effectively to recently reported KRAS G12D-targeting chemical matter, supporting its utility for identifying and prioritizing emerging candidates in drug discovery.

Supplementary Section 5. Full attention statistics across chemical categories

We further report the full atom-, bond- and motif-level attention statistics across all analysed chemical categories in **Supplementary Tables 7–9**. All statistics were computed using the QM9-trained models and the same attention aggregation protocol as in the main analysis. Here, $\mu_{\mathrm{w}/\rho}$ and $\mu_{\mathrm{w/o}\rho}$ denote the mean attention of eMol w/ ρ and the atom-only counterpart w/o ρ, respectively. The absolute and relative attention changes are defined as

$$\Delta_{\mathrm{abs}} = \mu_{\mathrm{w}/\rho} - \mu_{\mathrm{w/o}\rho},$$

and

$$\Delta_{\mathrm{rel}}(\%) = 100 \times \frac{\mu_{\mathrm{w}/\rho} - \mu_{\mathrm{w/o}\rho}}{\mu_{\mathrm{w/o}\rho}}.$$

At the atom level, eMol shows a clear redistribution of attention from H to more electronically informative elements. The mean attention on H decreases from 0.345383 to 0.150114 ($\Delta_{\mathrm{abs}} = -0.195269$, $\Delta_{\mathrm{rel}} = -56.5\%$), whereas attention increases on O ($\Delta_{\mathrm{abs}} = 0.044474$, $\Delta_{\mathrm{rel}} = 15.3\%$), N ($\Delta_{\mathrm{abs}} = 0.214909$, $\Delta_{\mathrm{rel}} = 198.2\%$) and F ($\Delta_{\mathrm{abs}} = 0.248434$, $\Delta_{\mathrm{rel}} = 215.9\%$). C shows a smaller increase ($\Delta_{\mathrm{abs}} = 0.034056$, $\Delta_{\mathrm{rel}} = 9.9\%$), indicating that the dominant atom-level effect is not a uniform increase across all heavy atoms, but a shift towards elements associated with heterogeneous local electronic environments.

At the bond level, the redistribution is selective and context dependent. C–H bonds show a marked decrease in attention ($\Delta_{\mathrm{abs}} = -0.091852$, $\Delta_{\mathrm{rel}} = -26.7\%$), whereas C–C bonds remain essentially unchanged ($\Delta_{\mathrm{abs}} = 0.000004$, $\Delta_{\mathrm{rel}} = 0.0\%$). In contrast, several heteroatom-containing bonds receive higher attention with electron density, including C–N ($\Delta_{\mathrm{rel}} = 75.9\%$),

C=N (62.7%), C≡N (104.2%), C–F (115.2%), N–H (65.2%), N–O (27.1%), N–N (120.7%) and N=N (117.2%). Some polar bonds, such as C=O and O–H, do not show increased aggregate bond-level attention, indicating that electron density does not simply amplify predefined chemical heuristics. Instead, it selectively reallocates attention towards bonding environments that are more informative in the learned electronic context.

At the motif level, eMol decreases attention on common hydrocarbon motifs, including methyl ($\Delta_{\mathrm{rel}} = -36.2\%$), methylene ($-30.1\%$) and alkene ($-15.1\%$). In contrast, attention increases on several chemically distinctive motifs, including amine (58.1%), amide (14.7%), nitrile (95.8%), imine (55.7%), fluorocarbon (86.2%), aromatic ring (5.6%) and five-membered ring (10.3%). The full motif-level statistics also show that not every polar or heteroatom-containing motif is uniformly enhanced, as hydroxyl groups remain nearly unchanged ($\Delta_{\mathrm{rel}} = -0.1\%$), while carboxyl and nitro groups show small decreases (-3.3% and -3.7%, respectively). These results further support that electron-density-derived probes induce a selective, context-dependent attention redistribution from frequent hydrocarbon structural cues towards chemically informative local electronic environments.

Supplementary Section 6. Correlation between learned attention and local electron density

To determine whether the atom-level attention learned by eMol is associated with local electronic structure, we analysed 500 randomly sampled molecules from the QM9-DFT test set. Attention weights were extracted from the eMol model trained for HOMO–LUMO gap prediction. For each molecule, we examined whether atoms receiving greater attention were associated with higher local electron density or stronger local variation in the electron-density field.

For each atom, local electron density was calculated by integrating $\rho(\mathbf{r})$ within an atom-centred spherical neighborhood. The atomic covalent radius plus 0.5 Å was used as the primary neighborhood definition, while fixed radii of 1.0 and 1.5 Å were included as robustness checks. Local variation in the electron-density field was quantified using the density-weighted gradient magnitude, $\rho(\mathbf{r})|\nabla\rho(\mathbf{r})|$, within the same neighborhoods. Spearman rank correlations between atom-level attention and each electron-density descriptor were calculated across atoms within each molecule.

We evaluated four settings to distinguish learned electronic associations from random, architectural and element-dependent effects:

- **Trained eMol:** atom-level attention extracted from the eMol model trained for HOMO–LUMO gap prediction.
- **Untrained eMol:** attention extracted from the same eMol architecture before training, controlling for correlations introduced by the architecture and random initialization.
- **Shuffled attention:** trained eMol attention values randomly permuted across atoms within each molecule, preserving the molecule-level attention distribution while disrupting the correspondence between attention and individual atoms.
- **Trained eMol (element-controlled):** trained eMol attention analysed after removing the average differences among H, C, N, O and F atoms from both attention and the corresponding electron-density descriptor. This control assesses whether attention remains associated with variation in local electronic structure after accounting for elemental identity.

The element-controlled analysis accounts for the intrinsic dependence of local electron density on elemental identity. For example, oxygen generally has greater local electron density than carbon or hydrogen. If eMol systematically assigns greater attention to O or N atoms, a positive attention–density correlation could therefore arise from elemental differences alone. Removing the average differences among H, C, N, O and F allows us to test whether attention remains associated with variations in local electronic structure among atoms of the same element.

For each setting, we report the mean and standard deviation of the molecule-level Spearman correlations and the 95% bootstrap confidence interval for the mean. Statistical significance was evaluated using a two-sided significance test (sign test) of whether the median molecule-level correlation differed from zero (**Supplementary Table 10**).

Trained eMol produced the highest mean correlation across all six combinations of electron-density descriptor and neighborhood definition. Its correlation with local electron density remained consistently high across the three neighborhoods ($\overline{r_s} = 0.660$–0.717). The correlation with the density-weighted gradient was strongest within the most local neighborhood and decreased as the radius increased ($\overline{r_s} = 0.546$, 0.359 and 0.210), indicating that attention was most strongly associated with localized density variation. In all six settings, the 95% bootstrap confidence interval for trained eMol excluded zero, with two-sided significance-test p values ranging from 9.51×10^{-111} to 1.84×10^{-32}.

Shuffled attention remained close to zero, showing that the observed correlations did not arise from the attention distribution alone. The untrained model exhibited weaker but non-negligible correlations, indicating contributions from architectural and elemental biases. Controlling for elemental identity substantially reduced the correlations, confirming that average differences among H, C, N, O and F explain part of the raw association, while residual positive correlations remained in several settings.

Overall, trained eMol consistently showed the strongest association across electron-density descriptors and neighborhood definitions, demonstrating that training aligns atom-level attention with local electronic structure beyond random correspondence and element-level differences.

Supplementary Section 7. Retrieval validation of electronic memory

Retrieval setting. We evaluated whether the electron density retrieved from electronic memory is consistent with the molecule-specific electron density computed for the query molecule. For each query, eMol retrieved a reference molecule from electronic memory using structural information only. No query electron density, target property, energy label or force label was used during retrieval. The retrieved molecule was aligned to the query through MCS-guided atom correspondence, and its stored density was transformed into the query coordinate frame. The memory-retrieved density was then compared with the computed electron density of the query molecule. This experiment therefore directly evaluates the quality of density transfer from electronic memory, rather than downstream prediction accuracy.

Electron-density similarity scores. Electron-density consistency was quantified using two complementary scores. Local electron-density similarity, S^{ρ}_{loc}, compares atom-centred local electronic environments. For each atom, nearby density samples were weighted by a Gaussian kernel within a local radius of 2.0 Å. The normalized weights were used to compute local density statistics, including density magnitude, density dispersion, weighted density mass, radial distance statistics, density–distance covariance and entropy. These atom-level descriptors were pooled into a molecular descriptor using histogram pooling, and S^{ρ}_{loc} was computed as the cosine similarity between the pooled descriptors of the query and retrieved molecules.

Radial electron-density similarity, S^{ρ}_{rad}, compares atom-centred radial density distributions. Each density sample was assigned to its nearest atom. For each atom, a density-weighted radial histogram was computed using 32 radial bins up to 4.0 Å. The atom-level histograms were normalized, pooled into a molecular descriptor and compared between the query and retrieved molecules using cosine similarity. Thus, S^{ρ}_{loc} measures similarity in local density environments, whereas S^{ρ}_{rad} measures similarity in radial density profiles around atoms. Larger values indicate greater electron-density consistency. Both scores were used only for retrieval validation and were not used during electronic-memory retrieval.

Binning and correlation analysis. To examine how density consistency changes with conventional molecular similarity, eMol-retrieved molecular pairs were grouped by MCS, ECFP4 and MACCS similarity. For each structural metric, pairs were assigned to five intervals: $[0.0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.6)$, $[0.6, 0.8)$ and $[0.8, 1.0]$. Each bin contained $5,000$ retrieved pairs. The mean values of S^{ρ}_{loc} and S^{ρ}_{rad} were computed within each bin. We further computed Pearson and Spearman correlation coefficients between structural similarity and the two electron-density similarity scores across the retrieved pairs. Pearson correlation measures linear association, whereas Spearman correlation measures rank-level association. Statistical significance was computed for each correlation, and all reported values satisfied $P \leq 1 \times 10^{-7}$.

Random-pair control. As a control, we constructed random molecular pairs under the same MCS-bin setting. For each MCS interval, random molecular pairs were sampled until $5,000$ pairs were obtained, and the same S^{ρ}_{loc} and S^{ρ}_{rad} scores were computed. This control used the same bin definitions, sample counts and density-similarity calculations as the eMol retrieval evaluation. The only difference was the pair-selection procedure: eMol pairs were selected by electronic-memory retrieval, whereas control pairs were selected randomly within each MCS interval.

Supplementary Section 8. Computational efficiency

We evaluated the inference efficiency of eMol, QuinNet and ViSNet using the same 1,000 molecules from the QM9 test set (**Supplementary Table 14**). All measurements used a batch size of 32 and were repeated five times on a single NVIDIA A800-SXM4-80GB GPU, with the mean values across the five runs reported. QuinNet was evaluated using the official QM9-gap architecture with a fixed random initialization and reports architecture-level neural-network forward latency. QuinNet and ViSNet received atom-based inputs, whereas eMol additionally incorporated electronic-density probes.

For eMol, electronic-memory retrieval and neural-network inference were measured separately. The eMol-memory stage measures only the time required to retrieve the electronic-density probes corresponding to each molecular structure from the electronic-memory records after they had been loaded into CPU memory. The eMol-Inference stage measures only the neural-network forward pass using the atomic structures and retrieved density probes. The complete eMol measurement combines these two stages and therefore reports the total retrieval-plus-forward cost.

The complete eMol pipeline processed approximately 455 molecules per second with 1.245 GB of GPU memory. Electronic-memory retrieval required only 0.000762 ms per molecule, contributing less than 0.04% of the total retrieval-plus-forward latency. Thus, the computational cost of eMol was dominated by the density-conditioned neural-network forward pass, while retrieval of the corresponding electronic-density information introduced negligible runtime overhead.

Supplementary Section 9. Definition of the electron-probe descriptor

For completeness, we give the exact descriptor used to initialize each electron probe. Let

$$d_{mj} = \left\| \mathbf{r}_j^{\rho} - \mathbf{r}_m^{e} \right\|, \qquad \bar{\omega}_{mj} = \frac{\omega_{mj}}{\sum_k \omega_{mk} + \varepsilon}, \tag{27}$$

where ω_{mj} is the smooth kernel defined in the main Methods and $\varepsilon = 10^{-8}$. The weighted density mean, variance and standard deviation are

$$\mu_{\rho,m} = \sum_j \bar{\omega}_{mj} \rho_j, \qquad v_{\rho,m} = \sum_j \bar{\omega}_{mj} \left(\rho_j - \mu_{\rho,m}\right)^2, \qquad s_{\rho,m} = \sqrt{v_{\rho,m} + \varepsilon}. \tag{28}$$

The density-magnitude terms are the weighted first and second norms

$$L_{1,m} = \sum_j \bar{\omega}_{mj} |\rho_j|, \qquad L_{2,m} = \sqrt{\sum_j \bar{\omega}_{mj} \rho_j^2 + \varepsilon}. \tag{29}$$

To measure concentration, non-negative density mass is normalized as

$$p_{mj} = \frac{\omega_{mj} \max(\rho_j, 0)}{\sum_k \omega_{mk} \max(\rho_k, 0) + \varepsilon}, \qquad H_m = -\sum_j p_{mj} \log(p_{mj} + \varepsilon). \tag{30}$$

The effective support size, logarithmic kernel mass and unnormalized weighted density mass are

$$N_{\mathrm{eff},m} = \frac{1}{\sum_j \bar{\omega}_{mj}^2 + \varepsilon}, \qquad M_{\omega,m} = \log\left(1 + \sum_j \omega_{mj}\right), \qquad M_{\rho,m} = \sum_j \omega_{mj} \rho_j. \tag{31}$$

The radial mean, radial standard deviation and density–distance covariance are

$$\mu_{d,m} = \sum_j \bar{\omega}_{mj} d_{mj}, \qquad s_{d,m} = \sqrt{\sum_j \bar{\omega}_{mj} (d_{mj} - \mu_{d,m})^2 + \varepsilon}, \tag{32}$$

$$C_{\rho d,m} = \sum_j \bar{\omega}_{mj} (\rho_j - \mu_{\rho,m})(d_{mj} - \mu_{d,m}). \tag{33}$$

Spatial asymmetry around the probe is represented by the three Cartesian components of the weighted centroid offset,

$$\delta_m = \sum_j \bar{\omega}_{mj} \left(\mathbf{r}_j^{\rho} - \mathbf{r}_m^{e}\right) \in \mathbb{R}^3, \tag{34}$$

together with the second radial moment

$$q_{d,m} = \sum_j \bar{\omega}_{mj} d_{mj}^2. \tag{35}$$

Finally, the standardized third and fourth density moments are

$$\gamma_{\rho,m} = \frac{\sum_j \bar{\omega}_{mj} (\rho_j - \mu_{\rho,m})^3}{s_{\rho,m}^3 + \varepsilon}, \qquad \kappa_{\rho,m} = \frac{\sum_j \bar{\omega}_{mj} (\rho_j - \mu_{\rho,m})^4}{v_{\rho,m}^2 + \varepsilon}. \tag{36}$$

The complete probe descriptor is ordered as

$$\phi_m = \left[\mu_{\rho,m}, s_{\rho,m}, L_{1,m}, L_{2,m}, H_m, N_{\mathrm{eff},m}, M_{\omega,m}, M_{\rho,m}, \mu_{d,m}, s_{d,m}, C_{\rho d,m}, \delta_m^{\top}, q_{d,m}, \gamma_{\rho,m}, \kappa_{\rho,m}\right]^{\top} \in \mathbb{R}^{17}. \tag{37}$$

When $\sum_j \omega_{mj} \le 10^{-6}$, all 17 components are set to zero. Otherwise, the descriptor is mapped to the hidden dimension d by

$$\mathbf{g}_m^0 = \mathrm{SiLU}\left(\mathbf{W}_e \phi_m + \mathbf{b}_e\right), \qquad \mathbf{W}_e \in \mathbb{R}^{d \times 17}. \tag{38}$$

Supplementary Section 10. Training details

Default optimization setting. Unless otherwise specified, models were trained with AdamW. The initial learning rate was set to 2×10^{-4}, the batch size was 4 and the maximum number of training epochs was 3000. The weight decay was set to 0. The learning rate was linearly warmed up for the first 1000 steps and then reduced on validation-loss plateaus with a decay factor of 0.8, a patience of 30 epochs and a minimum learning rate of 1×10^{-7}. Training was stopped when the validation loss did not improve for 500 epochs.

Molecular dynamics benchmarks. For molecular dynamics benchmarks, eMol was trained to jointly predict molecular energies and atomic forces using a weighted energy–force objective with $\lambda_E = 0.05$ and $\lambda_F = 0.95$. The model used 9 interaction layers, 256 hidden channels, 32 radial basis functions and an atom cutoff of 5.0 Å. The local electron radius was set to 2.0 Å, each atom was assigned $K = 3$ electron probes and the maximum number of atom neighbours was 32.

QM9 property prediction. For QM9 property prediction, eMol used the same electron-density molecular representation with a target-specific scalar readout. We evaluated 12 targets, including dipole moment, isotropic polarizability, frontier orbital energies, HOMO–LUMO gap, electronic spatial extent, zero-point vibrational energy, thermodynamic energies and heat capacity. The data splitting and evaluation metrics were identical to those of ViSNet[19].

QM9-DFT dataset construction. For the QM9-DFT experiment, molecule-specific electron-density fields were computed following the DFT protocol used in EDBench[34]. The computed-density records were matched to QM9 molecules by elemental composition and three-dimensional structure. Ambiguous or duplicate matches were excluded. The final dataset contained 105,909 uniquely matched molecules and was split at the molecule level into 95,318 training, 5,295 validation and 5,296 test molecules.

Molecular dynamics simulations. For molecular dynamics simulations, the trained eMol model was incorporated into the Atomic Simulation Environment. Simulations were performed on the seven MD17 molecules with a time step of $\tau = 0.5$ fs under a Berendsen thermostat, following the simulation protocol of ViSNet[19]. The thermostat temperature, coupling constant, initial structures, simulation length and remaining simulation settings were kept consistent with the corresponding MD17 reference trajectories and the ViSNet evaluation setting. The generated trajectories were compared with reference DFT-level ab initio molecular dynamics trajectories using the interatomic distance distribution $h(r)$, which summarizes the ensemble-averaged distribution of atoms at a distance r from each atom during simulation. The binning and normalization of $h(r)$ followed the ViSNet protocol so that learned-potential trajectories and reference DFT trajectories were evaluated under the same statistical procedure.

Chignolin evaluation. For Chignolin, the evaluation followed the ViSNet Chignolin protocol[19]. The dataset consists of conformations of the 166-atom mini-protein Chignolin sampled by replica-exchange molecular dynamics and labelled at the DFT level using Gaussian 16. The dataset was split into training, validation and test sets at a ratio of 8:1:1, and energy and force labels from DFT calculations were used as the ground truth. In our evaluation, the test set contained 9,543 conformations with 166 atoms per conformation. Each structure was projected onto $d_{\mathrm{D3-G7}}$ and $d_{\mathrm{E5-T8}}$, defined by the D3 carbonyl O–G7 backbone N and E5 carbonyl O–T8 backbone N distances, respectively. A two-dimensional occupancy histogram was converted to conformational free energy using $F = -k_{\mathrm{B}}T \ln P$ at 300 K and shifted to a zero minimum. Energy and force errors were evaluated by mean absolute error. LiTEN and MGNN were trained and evaluated under the same data split, DFT labels and evaluation metrics as eMol to ensure a fair comparison.

Supplementary Section 11. Baseline reproduction and transfer settings

To extend the benchmark comparison beyond previously reported results, we reproduced or transferred LiTEN, MGNN, QuinNet, Euclidean Fast Attention (EFA) and MACE using their public implementations and model configurations. Each method was run in an independent software environment following the dependencies of its official repository. Where an original study provided a configuration for the corresponding benchmark, we followed that setting as closely as possible. For tasks not reported in the original work, we retained the published backbone architecture and used the transfer settings described below. No pretrained weights were used in these additional experiments unless otherwise stated. For all reproduced and transferred baselines, model selection was based on validation-set performance.

Benchmark splits. MD17 used 1,000 training configurations and 500 validation configurations for each molecular system, with the remaining configurations used for testing. For rMD17, each molecule used the fixed set of 2,000 ordered configurations, comprising 950 training, 50 validation and 1,000 test configurations. QM9 contained 130,831 molecules and used a fixed split of 100,000 training, 17,748 validation and 13,083 test molecules. MD22 used the benchmark-specific split sizes with random sampling using seed 0. QM9-DFT contained 105,909 molecules, with 95,318, 5,295 and 5,296 molecules assigned to the training, validation and test sets, respectively. The same split was used across models within each benchmark.

LiTEN

LiTEN experiments were based on the public implementation and configurations of the original model. The backbone contained 6 message-passing layers, 256 hidden channels, 8 attention heads and 32 radial basis functions.

For MD17, the cutoff radius was 5 Å. Models were optimized using Adam with an initial learning rate of 1×10^{-4} and a batch size of 2 for up to 3,000 epochs. Energy and force MAE losses were weighted by 1 and 99, respectively. The learning

rate was adjusted according to validation loss using ReduceLROnPlateau with a decay factor of 0.9, a patience of 30 epochs and a minimum learning rate of 1×10^{-6}.

For QM9 and QM9-DFT, the same LiTEN backbone was transferred to molecular scalar prediction, with the force-prediction branch disabled. Each of the 12 properties was trained independently. Targets were normalized using the training-set mean and mean absolute deviation, and models were optimized using an L1 loss and Adam with cosine learning-rate decay. The initial and minimum learning rates were 5×10^{-4} and 1×10^{-6}, respectively, with a batch size of 128 and a maximum of 1,000 epochs. These scalar-prediction settings constitute transfer experiments for tasks not covered by the original LiTEN configuration.

MGNN

MGNN experiments followed the architecture and training settings of the original implementation where available. The molecular-potential model used 2 interaction layers, 512 feature channels, a radial expansion parameter of $\beta = 8$ and a cutoff radius of 4.5 Å.

For MD17 and MD22, models were optimized using Adam with an initial learning rate of 3×10^{-5} and a batch size of 4. The learning rate was warmed up during the first 10 epochs and subsequently adjusted using ReduceLROnPlateau with a decay factor of 0.8, a patience of 100 epochs and a minimum learning rate of 1×10^{-8}. Energy and force predictions were jointly optimized using a Huber loss with $\delta = 0.5$. For MD17, the force-loss weights followed the molecule-specific settings reported in the original work. For MD22, the transferred setting used energy and force weights of 0.001 and 0.999, respectively. Exponential moving averaging of model parameters was applied with a decay rate of 0.99.

For QM9-DFT, we followed the MGNN QM9 architecture as the basis for transfer. The model contained 3 interaction layers, 512 channels and a cutoff radius of 5 Å. Each property was trained independently using Adam and a Huber loss with $\delta = 0.5$. The initial learning rate was 1×10^{-5} with a batch size of 16. Training used a 5-epoch warm-up followed by ReduceLROnPlateau with a decay factor of 0.5 and a patience of 30 epochs.

QuinNet

QuinNet experiments were based on the public implementation and official example configuration. The transferred model used 5 message-passing layers, 256 hidden channels, 8 attention heads, 32 radial basis functions, second-order spherical harmonic features and a cutoff radius of 5 Å.

For QM9-DFT, each of the 12 molecular properties was trained independently and the force-prediction branch was disabled. Training targets were normalized using statistics from the training set. Models were optimized using an L1 loss and Adam with cosine learning-rate decay, an initial learning rate of 5×10^{-4}, a minimum learning rate of 1×10^{-6}, a batch size of 128 and a maximum of 1,000 epochs. These optimization settings were used as a transfer protocol because the original QuinNet configuration did not provide the corresponding QM9-DFT experiments.

Euclidean Fast Attention

EFA experiments used the public Euclidean Fast Attention implementation. The base model contained 3 message-passing layers with 128 feature channels and a local interaction cutoff of 5 Å, with Euclidean Fast Attention applied during the first two message-passing stages. Radial basis, spherical quadrature and attention-channel settings followed the public base-model configuration.

For MD17, rMD17 and MD22, EFA was trained jointly on total energies and atomic forces. Energy and force mean-squared-error losses were weighted by 0.001 and 0.999, respectively. Models were optimized using Adam with an initial learning rate of 1×10^{-3}, exponential learning-rate decay and gradient-norm clipping at 1.5. Training targeted approximately 10^6 optimization steps. MD17 and rMD17 used a batch size of 4, whereas MD22 used batch sizes of 1 or 8 depending on system size.

Because neither the EFA publication nor its public repository provides a QM9 benchmark configuration, we transferred the official EFA energy model to the 12 QM9 scalar-property prediction tasks. Each target was trained independently. Labels were normalized using the training-set mean and mean absolute deviation, and models were optimized using an L1 loss and Adam with cosine learning-rate decay from 5×10^{-4} to 1×10^{-6}. The original atomwise-summation readout of the EFA energy model was retained. For QM9, we transferred the official EFA energy model to the 12 molecular scalar-property prediction tasks. Each target was trained independently. Labels were normalized using the training-set mean and mean absolute deviation, and models were optimized using an L1 loss and Adam with cosine learning-rate decay from 5×10^{-4} to 1×10^{-6}. The original atomwise-summation readout of the EFA energy model was retained.

All EFA models were trained from random initialization without pretrained weights.

MACE

MACE experiments used the `ScaleShiftMACE` architecture with the model-size configuration of MACE-OFF24. The model used a cutoff radius of 6 Å, 2 interaction blocks, 128 channels, max_L = 1, max_ell = 3, correlation order 3, 8 radial basis functions and 5 cutoff basis functions. The first interaction layer used `RealAgnosticInteractionBlock`, followed by

`RealAgnosticResidualInteractionBlock`. Only the architecture and principal hyperparameters of MACE-OFF24 were adopted; no MACE-OFF24 pretrained weights or SPICE training data were used.

For MD17, rMD17 and MD22, models were trained from random initialization in float64 precision and jointly optimized for energies and forces. The first training stage used energy and force loss weights of 40 and 1,000, respectively. Optimization used Adam with AMSGrad, an initial learning rate of 0.01, weight decay of 5×10^{-10} and gradient-norm clipping at 1.0. Models were trained for up to 3,500 epochs using ReduceLROnPlateau, exponential moving averaging and stochastic weight averaging (SWA). SWA was started at epoch 2,000 with a learning rate of 2.5×10^{-4}. MD17 and rMD17 primarily used a batch size of 5, whereas MD22 used batch sizes of 1, 2 or 4 according to molecular size.

For QM9, each of the 12 molecular properties was trained independently using the same fixed QM9 split. Targets were normalized using the training-set mean and mean absolute deviation and optimized using an L1 loss. Training used Adam with cosine annealing, an initial learning rate of 5×10^{-4}, a minimum learning rate of 1×10^{-6}, a batch size of 128 and up to 1,000 epochs.

All MACE models were trained from random initialization without pretrained weights.

QuiNet 5-body@I control for eMol-DFT

QuiNet introduced three forms of 5-body interaction, denoted 5-body@I, 5-body@II and 5-body@III[43]. Briefly, 5-body@I constructs node-level higher-order geometry from the squared norm of local geometric features, 5-body@II aggregates transformed neighbouring geometric features with edge-dependent modulation, and 5-body@III captures cross-node geometric correlations through inner products between transformed local features.

To assess whether the original 5-body formulation could be directly transferred to the heterogeneous atom–electron representation, we incorporated the QuiNet 5-body@I term into eMol-DFT without the adapted 5-body interactions used in the final model. This control used the same QM9-DFT data split and training protocol as eMol-DFT, so that the comparison isolates the effect of the 5-body formulation.

Supplementary Section 12. Supplementary Tables

Supplementary Table 1: Overall distribution of the four classification datasets in the drug discovery task. The imbalance ratio is calculated as the number of majority-class samples divided by the number of minority-class samples.

Dataset	Total	Positive ($y = 1$)	Negative ($y = 0$)	Positive ratio	Imbalance ratio	Active threshold
3CLpro (Mpro)	2,904	980	1,924	33.75%	1.96:1	100 nM
KRAS G12D	1,370	219	1,151	15.99%	5.26:1	100 nM
BTK	11,893	8,228	3,665	69.18%	2.25:1	100 nM
CYP3A4	9,907	352	9,555	3.55%	27.15:1	100 nM

Supplementary Table 2: Attention enrichment for mechanism-relevant molecular motifs in the Mpro/3CLpro dataset. *n* denotes the number of unique molecules containing each motif. Values are reported as mean enrichment, with the lower and upper bounds of the 95% bootstrap confidence interval shown in parentheses. Confidence intervals were obtained from 5,000 bootstrap resamples of the motif-containing molecules. The highest mean enrichment for each motif is shown in bold.

Motif	*n*	JMM	MotiL	eMol
Piperazine	13	0.961 (0.765–1.171)	1.060 (0.924–1.193)	**1.411 (1.241–1.586)**
Piperidine	46	0.985 (0.884–1.091)	1.081 (0.975–1.193)	**1.269 (1.157–1.391)**
Acrylamide	10	0.445 (0.234–0.711)	0.934 (0.747–1.144)	**1.237 (1.044–1.476)**
Three-carbon linker	79	0.823 (0.730–0.923)	0.734 (0.671–0.802)	**1.232 (1.125–1.349)**

Supplementary Table 3: Representative compounds used for atom-level attribution visualization in the Mpro/3CLpro and CYP3A4 tasks.

Task	Case	Compound ID	Canonical SMILES
Mpro/3CLpro	1	cmpd_002802	O=c1c2ccccc2[se]n1C1CCN(S(=O)(=O)C2CC2)CC1
Mpro/3CLpro	2	cmpd_002650	O=C[C@H](C[C@@H]1CCCNC1=O)NC(=O)[C@@H](NC(=O)/C=C/c1ccccc1)c1ccccc1
Mpro/3CLpro	3	cmpd_002595	O=C1CC2(CCN(C(=O)C3CN(c4ccc(Cl)c(Cl)c4)CCN3C(=O)c3cc(=O)[nH]c4ccccc34)CC2)C1
CYP3A4	1	cmpd_004900	COc1nc(-c2ccc(Cl)cc2)ccc1[C@@H](c1ccccc1)[C@@](O)(CCN(C)C)c1cccc2ccccc12
CYP3A4	2	cmpd_008467	O=C(CC1([C@H]2CO[C@H]3CCC[C@H]3N2S(=O)(=O)c2ccc(Cl)cc2)CC1)N1CCC(N2CCCCC2)CC1
CYP3A4	3	cmpd_008747	O=C(N[C@@H](c1ccccn1)C1CCCC1)c1ccc2[nH]nc(-c3ccc(N4[C@H]5COC[C@H]4CC(O)C5)cc3)c2c1

Supplementary Table 4: Attention enrichment for metabolism-relevant molecular motifs in the CYP3A4 dataset. *n* denotes the number of unique molecules containing each motif. Values are reported as mean enrichment, with the lower and upper bounds of the 95% bootstrap confidence interval shown in parentheses. Confidence intervals were obtained from 5,000 bootstrap resamples of the motif-containing molecules. The highest mean enrichment for each motif is shown in bold.

Motif	*n*	JMM	MotiL	eMol
Methoxy/alkoxy	458	0.786 (0.746–0.825)	1.105 (1.057–1.156)	**1.415 (1.364–1.466)**
Tertiary amine	326	1.008 (0.976–1.041)	0.789 (0.750–0.830)	**1.137 (1.094–1.184)**
Metabolism-relevant motifs	828	0.971 (0.951–0.992)	1.026 (0.998–1.053)	**1.172 (1.146–1.197)**
Morpholine	64	1.055 (0.997–1.115)	0.858 (0.794–0.928)	**1.194 (1.098–1.291)**

Supplementary Table 5: Top-ranked DrugBank candidates prioritized by eMol, MotiL and JMM for Mpro/3CLpro inhibition.

	eMol			MotiL			JMM		
Rank	DrugBank ID	Score	Evid.	DrugBank ID	Score	Evid.	DrugBank ID	Score	Evid.
1	DB16823	0.999890	N	DB01106	0.950434	N	DB16766	0.997418	N
2	DB08539	0.999873	N	DB12973	0.937651	N	DB09211	0.993732	N
3	DB18062	0.999786	N	DB18968	0.931071	N	DB14137	0.992443	N
4	DB13113	0.999776	N	DB11993	0.927559	N	DB02341	0.990880	N
5	**DB12997**	0.999699	**C**[a]	DB12427	0.923167	N	DB05403	0.986364	N
6	**DB18975**	0.999561	**E**[b]	DB07741	0.921327	N	DB14641	0.984893	N
7	DB18811	0.999559	N	**DB09030**	0.915462	**B/C**[c]	DB14002	0.982877	N
8	DB02647	0.999534	N	DB11487	0.914158	N	DB05514	0.981556	N
9	DB11899	0.999374	N	DB17171	0.912474	N	DB01669	0.980048	N
10	DB01996	0.999321	N	DB18126	0.912317	N	DB06521	0.979971	N
Evidence-supported		2/10			1/10			0/10	

Note: Scores are reported on their original, model-specific scales and should not be compared directly across methods. Literature evidence was searched only for the top 10 candidates prioritized by each method. Evidence categories are defined as follows: E, direct Mpro/3CLpro enzymatic inhibition; B/C, Mpro/3CLpro binding together with cell-based antiviral activity but without direct enzymatic inhibition data; C, cell-based anti-SARS-CoV-2 activity only; and N, no relevant evidence identified in the targeted search. An N designation does not indicate that the compound is inactive.

[a] Brilacidin showed cell-based anti-SARS-CoV-2 activity with an IC_{50} of 0.565 μM and an IC_{90} of 2.63 μM; no direct Mpro/3CLpro enzymatic IC_{50} was identified[48].

[b] Pomotrelvir directly inhibited wild-type SARS-CoV-2 Mpro/3CLpro with an IC_{50} of 24 nM[49].

[c] Vorapaxar bound to Mpro/3CLpro with a K_d of 27 μM and inhibited viral replication in infected cells with an EC_{50} of 11 μM; no direct enzymatic IC_{50} was reported[51].

Supplementary Table 6: Literature-supported KRAS G12D-targeting compounds and prediction scores assigned by eMol, JMM and MotiL. Compounds were treated as positive reference examples based on reported wet-lab evidence of KRAS G12D inhibition, degradation or pharmacological activity in KRAS G12D models.

ID	Name	IC_{50}	eMol	JMM	MotiL
1	HRS-4642[a][52]	0.72 nM	0.9711	0.9826	0.4862
2	ASP3082[b][53]	19 nM	0.0459	0.0005	0.1425
3	Zoldonrasib[c][54]	N/A	0.6180	0.3084	0.1294
4	MCB-294[d][55]	N/A	0.7821	0.6236	0.0506
	Number of compounds with score > 0.5		**3**	**2**	**0**

Search and selection strategy. Candidate compounds were identified by manually screening all five pages returned by this Google Scholar search, comprising approximately 50 records in total. The query was `KRAS G12D inhibitor source:"Nature" OR source:"Science" OR source:"Cell"`, and was designed to retrieve studies on $KRAS^{G12D}$ inhibitors published in Nature, Science, Cell, or their associated journals. All records returned by the search were published in 2025–2026; therefore, no additional publication-year filter was applied. Titles and abstracts were manually screened. A compound was included only if it was explicitly identified by name, its abstract reported wet-lab experimental evidence supporting inhibition, degradation, or pharmacological targeting of $KRAS^{G12D}$, and a defined chemical structure was available. Reviews, perspectives, news articles, purely computational studies, unnamed compounds, and studies lacking explicit experimental validation of $KRAS^{G12D}$ activity were excluded. Canonical SMILES and reported IC_{50} values were subsequently collected for the retained compounds. ASP3082 was included as a selective $KRAS^{G12D}$ degrader, whereas MCB-294 was included as a pan-KRAS inhibitor with experimentally demonstrated activity in $KRAS^{G12D}$ models.

[a] **HRS-4642, canonical SMILES:** `CCc1c(F)ccc2cc(O)cc(-c3nc4c5c(nc(OC[C@@]67CCCN6C[C@H](F)C7)nc5c3F)N3C[C@H]5CC[C@H](N5)[C@H]3CO4)c12.`

[b] **ASP3082, canonical SMILES:** `Cc1ncsc1-c1ccc([C@H](CO)NC(=O)[C@@H]2C[C@@H](O)CN2C(=O)[C@H](C(C)C)n2cc(-c3ccc(COc4c(-c5c(C)c(F)cc6[nH]ncc56)c(C5CC5)cc5c(N6C[C@@H]7C[C@H]6CN7)nc(OC6CCOCC6)nc45)cc3)nn2)cc1.`

[c] **Zoldonrasib, canonical SMILES:** `O=C(N1CCC[C@@H]2N1)[C@H](C[C@@H](C3)OCCN3C4=CC5=C(N(CC(F)(F)F)[C@]([C@]6=C([C@H](C)OC)N=CC(N7CCN(C8CC8)CC7)=C6)=C5CC(C)(C)COC2=O)C=C4)NC([C@@H](N9C[C@]%10(CN(C([C@@H]%11N(C)[C@@H]%11C%12CC%12)=O)CC%10)CC9)C%13CCCC%13)=O.`

[d] **MCB-294, canonical SMILES:** `C[C@@]1(O)CCCN(C1)c2nc(OC[C@@]34CCCN3CC(C4)=C(F)F)nc5c(F)c(ncc25)c6cc(O)cc7ccc(F)c(C#C)c67.`

Supplementary Table 7: Atom-level attention statistics. Mean and standard deviation are reported for eMol w/ ρ and the atom-only counterpart w/o ρ. The absolute change is computed as $\Delta_{\text{abs}} = \mu_{\text{w}/\rho} - \mu_{\text{w/o}\rho}$, and the relative change is computed as $\Delta_{\text{rel}}(\%) = 100 \times \frac{\mu_{\text{w}/\rho} - \mu_{\text{w/o}\rho}}{\mu_{\text{w/o}\rho}}$.

Atom	n	eMol w/ ρ		eMol w/o ρ		Δ_{abs}	Δ_{rel} (%)
		Mean ($\mu_{\text{w}/\rho}$)	Std	Mean ($\mu_{\text{w/o}\rho}$)	Std		
H	108270	0.150114	0.068601	0.345383	0.089559	-0.195269	-56.54%
C	108316	0.376477	0.144185	0.342421	0.07611	0.034056	9.95%
O	93615	0.334492	0.15036	0.290018	0.110036	0.044474	15.33%
N	59633	0.323326	0.169122	0.108417	0.074999	0.214909	198.22%
F	1907	0.36353	0.184598	0.115096	0.07008	0.248434	215.85%

Supplementary Table 8: Bond-level attention statistics. Mean and standard deviation are reported for eMol w/ ρ and the atom-only counterpart w/o ρ. The absolute change is computed as $\Delta_{\text{abs}} = \mu_{\text{w}/\rho} - \mu_{\text{w/o}\rho}$, and the relative change is computed as $\Delta_{\text{rel}}(\%) = 100 \times \frac{\mu_{\text{w}/\rho} - \mu_{\text{w/o}\rho}}{\mu_{\text{w/o}\rho}}$.

Bond	n	eMol w/ ρ		eMol w/o ρ		Δ_{abs}	Δ_{rel} (%)
		Mean ($\mu_{\text{w}/\rho}$)	Std	Mean ($\mu_{\text{w/o}\rho}$)	Std		
C-C	103266	0.24842	0.129202	0.248416	0.103942	0.000004	0.00%
C=C	37684	0.077452	0.054654	0.091858	0.06326	-0.014406	-15.68%
C≡C	15581	0.062661	0.035487	0.063925	0.029018	-0.001264	-1.98%
C-N	44329	0.20706	0.105197	0.1177	0.0613	0.089360	75.92%
C=N	32005	0.166341	0.101131	0.102218	0.07336	0.064123	62.73%
C≡N	13919	0.129102	0.061451	0.063233	0.024948	0.065869	104.17%
C-O	75926	0.234822	0.121067	0.225083	0.097485	0.009739	4.33%
C=O	42909	0.13109	0.066743	0.133793	0.055486	-0.002703	-2.02%
C-F	1907	0.246429	0.152322	0.114497	0.073753	0.131932	115.23%
C-H	107828	0.252848	0.113619	0.3447	0.11004	-0.091852	-26.65%
N-H	38088	0.128722	0.073233	0.077902	0.049829	0.050820	65.24%
O-H	37527	0.105519	0.050632	0.112612	0.039562	-0.007093	-6.30%
N-O	5366	0.171602	0.063102	0.135012	0.051301	0.036590	27.10%
N-N	6341	0.187626	0.087366	0.085022	0.042422	0.102604	120.68%
N=N	1747	0.151614	0.049203	0.069805	0.021558	0.081809	117.20%
N=O	166	0.27212	0.077306	0.261761	0.062699	0.010359	3.96%

Supplementary Table 9: Motif-level attention statistics. Mean and standard deviation are reported for eMol w/ ρ and the atom-only counterpart w/o ρ. The absolute change is computed as $\Delta_{\text{abs}} = \mu_{\text{w}/\rho} - \mu_{\text{w/o}\rho}$, and the relative change is computed as $\Delta_{\text{rel}}(\%) = 100 \times \frac{\mu_{\text{w}/\rho} - \mu_{\text{w/o}\rho}}{\mu_{\text{w/o}\rho}}$.

Motif	n	eMol w/ ρ		eMol w/o ρ		Δ_{abs}	Δ_{rel} (%)
		Mean ($\mu_{\text{w}/\rho}$)	Std	Mean ($\mu_{\text{w/o}\rho}$)	Std		
Carbonyl	23869	0.249064	0.148148	0.231216	0.115634	0.017848	7.72%
Carboxyl	730	0.414476	0.128729	0.428732	0.087477	-0.014256	-3.33%
Hydroxyl	36914	0.212076	0.145493	0.212221	0.114131	-0.000145	-0.07%
Ether	44752	0.263855	0.142317	0.24102	0.104826	0.022835	9.47%
Ester	7229	0.383556	0.158033	0.368788	0.118766	0.014768	4.00%
Amine	30106	0.174104	0.105642	0.110131	0.067072	0.063973	58.09%
Amide	15065	0.367681	0.163235	0.320561	0.133865	0.047120	14.70%
Nitrile	13919	0.241991	0.135434	0.12361	0.069923	0.118381	95.77%
Imine	32005	0.229918	0.136488	0.147688	0.105907	0.082230	55.68%
Nitro	166	0.368889	0.108932	0.383102	0.099083	-0.014213	-3.71%
Methyl	65652	0.131703	0.1153	0.206352	0.130895	-0.074649	-36.18%
Methylene	85977	0.158473	0.129464	0.226592	0.143287	-0.068119	-30.06%
Alkene	37684	0.110215	0.083697	0.129795	0.088866	-0.019580	-15.09%
Aromatic ring	10391	0.396835	0.140455	0.37577	0.123369	0.021065	5.61%
Five-membered ring	53531	0.502157	0.120694	0.455083	0.113143	0.047074	10.34%
Fluorocarbon	1907	0.34511	0.209254	0.185308	0.12983	0.159802	86.24%

Supplementary Table 10: Correlation between atom-level attention and local electron-density descriptors across three atomic neighborhoods for 500 QM9-DFT test molecules. Reported are the mean molecule-level Spearman rank correlation ($\overline{r_s}$), standard deviation (SD), 95% bootstrap confidence interval (CI) for the mean, and p value from a two-sided significance test (sign test) assessing whether the median molecule-level correlation differs from zero.

Radius	Method	Mean $\overline{r_s}$	SD	CI lower	CI upper	p-value
A. Local electron density						
Covalent radius +0.5 Å	**Trained eMol**	**0.7159**	0.0854	0.7087	0.7236	9.51×10^{-111}
	Untrained eMol	0.5503	0.1802	0.5342	0.5664	2.70×10^{-107}
	Trained eMol (element-controlled)	0.0420	0.2747	0.0175	0.0673	6.14×10^{-6}
	Shuffled attention	0.0116	0.2412	-0.0090	0.0324	0.655
Fixed radius 1.0 Å	**Trained eMol**	**0.6597**	0.1309	0.6488	0.6714	5.15×10^{-109}
	Untrained eMol	0.4709	0.2205	0.4515	0.4898	2.88×10^{-97}
	Trained eMol (element-controlled)	-2.38×10^{-5}	0.2789	-0.0255	0.0250	0.893
	Shuffled attention	-0.0208	0.2456	-0.0421	-4.98×10^{-5}	0.165
Fixed radius 1.5 Å	**Trained eMol**	**0.7168**	0.1550	0.7036	0.7304	9.51×10^{-111}
	Untrained eMol	0.5806	0.1869	0.5639	0.5974	3.74×10^{-108}
	Trained eMol (element-controlled)	0.2746	0.2810	0.2501	0.2984	2.48×10^{-48}
	Shuffled attention	-0.0164	0.2725	-0.0407	0.0067	0.227
B. Local electron-density gradient						
Covalent radius +0.5 Å	**Trained eMol**	**0.5464**	0.1664	0.5323	0.5610	7.02×10^{-110}
	Untrained eMol	0.3760	0.2310	0.3559	0.3951	3.37×10^{-86}
	Trained eMol (element-controlled)	-0.0246	0.2913	-0.0502	0.0021	0.622
	Shuffled attention	0.0037	0.2422	-0.0172	0.0241	0.858
Fixed radius 1.0 Å	**Trained eMol**	**0.3587**	0.2156	0.3398	0.3775	1.92×10^{-79}
	Untrained eMol	0.2422	0.2360	0.2212	0.2634	5.05×10^{-55}
	Trained eMol (element-controlled)	-0.0166	0.3019	-0.0435	0.0106	0.823
	Shuffled attention	-0.0224	0.2374	-0.0435	-0.0017	0.0283
Fixed radius 1.5 Å	**Trained eMol**	**0.2099**	0.2721	0.1853	0.2333	1.84×10^{-32}
	Untrained eMol	0.1036	0.2589	0.0812	0.1259	2.75×10^{-10}
	Trained eMol (element-controlled)	0.1201	0.2992	0.0931	0.1464	1.93×10^{-14}
	Shuffled attention	0.0074	0.2640	-0.0147	0.0300	0.754

Supplementary Table 11: Bin-wise electron-density similarity for eMol-retrieved molecular pairs. Each bin contains 5,000 retrieved pairs.

Structural metric	Similarity bin	Count	S^{ρ}_{loc}	S^{ρ}_{rad}
MCS	$[0.0, 0.2)$	5000	0.516	0.535
MCS	$[0.2, 0.4)$	5000	0.531	0.586
MCS	$[0.4, 0.6)$	5000	0.548	0.647
MCS	$[0.6, 0.8)$	5000	0.568	0.706
MCS	$[0.8, 1.0]$	5000	0.592	0.764
ECFP4	$[0.0, 0.2)$	5000	0.521	0.542
ECFP4	$[0.2, 0.4)$	5000	0.545	0.626
ECFP4	$[0.4, 0.6)$	5000	0.571	0.697
ECFP4	$[0.6, 0.8)$	5000	0.596	0.756
ECFP4	$[0.8, 1.0]$	5000	0.621	0.817
MACCS	$[0.0, 0.2)$	5000	0.515	0.535
MACCS	$[0.2, 0.4)$	5000	0.541	0.612
MACCS	$[0.4, 0.6)$	5000	0.568	0.686
MACCS	$[0.6, 0.8)$	5000	0.596	0.762
MACCS	$[0.8, 1.0]$	5000	0.625	0.832

Supplementary Table 12: Random-pair baseline under the same MCS similarity-bin setting. Each MCS interval contains 5,000 randomly sampled molecular pairs.

MCS interval	Sample count	S^{ρ}_{loc}	S^{ρ}_{rad}
$[0.0, 0.2)$	5000	0.156	0.337
$[0.2, 0.4)$	5000	0.204	0.364
$[0.4, 0.6)$	5000	0.261	0.397
$[0.6, 0.8)$	5000	0.323	0.432
$[0.8, 1.0]$	5000	0.402	0.474

Supplementary Table 13: Correlation between structural similarity metrics and electron-density similarity scores for eMol-retrieved molecular pairs.

Correlation	Structural metric	S^{ρ}_{loc}	P	S^{ρ}_{rad}	P
Pearson	MCS	0.32	1×10^{-7}	0.58	1×10^{-7}
Pearson	ECFP4	0.35	1×10^{-7}	0.62	1×10^{-7}
Pearson	MACCS	0.38	1×10^{-7}	0.65	1×10^{-7}
Spearman	MCS	0.33	1×10^{-7}	0.61	1×10^{-7}
Spearman	ECFP4	0.37	1×10^{-7}	0.67	1×10^{-7}
Spearman	MACCS	0.39	1×10^{-7}	0.68	1×10^{-7}

Supplementary Table 14: Inference efficiency on the QM9 test set.

Model	**Timed operation**	**Input**	**ms / molecule**	**molecules / s**	**GPU memory**
QuinNet	Neural-network forward	Atoms only	1.230901	812.413	0.619 GB
ViSNet	Neural-network forward	Atoms only	0.448865	2,227.855	0.622 GB
eMol-memory	In-memory retrieval	Retrieved probes	0.000762	1,397,878.584	N/A (CPU)
eMol-Inference	Neural-network forward	Atoms + density probes	2.195951	455.395	1.245 GB
eMol	Retrieval + forward	Atoms + retrieved probes	2.196713	455.237	1.245 GB