Artificial intelligence-enhanced quantum chemical method with broad applicability

Peikun Zheng[1], Roman Zubatyuk[2], Wei Wu[1], Olexandr Isayev [2✉] & Pavlo O. Dral [1✉]

High-level quantum mechanical (QM) calculations are indispensable for accurate explanation of natural phenomena on the atomistic level. Their staggering computational cost, however, poses great limitations, which luckily can be lifted to a great extent by exploiting advances in artificial intelligence (AI). Here we introduce the general-purpose, highly transferable artificial intelligence–quantum mechanical method 1 (AIQM1). It approaches the accuracy of the gold-standard coupled cluster QM method with high computational speed of the approximate low-level semiempirical QM methods for the neutral, closed-shell species in the ground state. AIQM1 can provide accurate ground-state energies for diverse organic compounds as well as geometries for even challenging systems such as large conjugated compounds (fullerene C_{60}) close to experiment. This opens an opportunity to investigate chemical compounds with previously unattainable speed and accuracy as we demonstrate by determining geometries of polyyne molecules—the task difficult for both experiment and theory. Noteworthy, our method's accuracy is also good for ions and excited-state properties, although the neural network part of AIQM1 was never fitted to these properties.

[1] State Key Laboratory of Physical Chemistry of Solid Surfaces, Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Department of Chemistry, and College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China. [2] Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213, USA. ✉email: olexandr@olexandrisayev.com; dral@xmu.edu.cn

Quantum mechanical (QM) methods used in chemistry are invaluable for today's modern science as they allow insights into electronic structure at an atomistic level, which are experimentally unattainable. This in turn helps to find answers to fundamental scientific questions in chemistry and related fields, such as chemical physics and biology, and assists applied science in designing better materials and discover new medicines.

The usefulness of QM methods in practical applications is determined by their accuracy and computational cost. The trade-off between these two factors guides the choice of the QM method. On the one side, we have very accurate, but slow high-level ab initio QM methods such as coupled cluster with single, double, and perturbative triple excitations, CCSD(T)[1], which has established itself as the gold standard in most applications, particularly, for closed-shell molecules[2–4]. On the other side, we have very fast semiempirical QM (SQM) methods that have rather limited accuracy[5]. The sweet spot of moderate computational cost and often sufficient accuracy is occupied by density functional theory (DFT) that has become a workhorse in the investigation of medium-sized systems (Fig. 1a)[6]. The efforts for developing faster and more accurate QM methods is an active research field, but it is clear that traditional approaches to QM method development require years of hard human work and typically yield only relatively modest improvements.

Advances in artificial intelligence (AI) bring chemistry research to a radically new level and provide a much-needed alternative to the traditional QM method development[7,8]. AI allows to perform calculations with both high accuracy and very low computational cost that was previously unattainable with the traditional QM methods. Nevertheless, most of the applications of AI to quantum chemistry are either proof-of-principle or limited to specific applications. Developing general-purpose AI approaches with transferability of QM methods remains a big challenge. A significant step towards transferable accurate AI approaches is the family of ANI potentials[9–12] that can describe energies and forces of compounds of different size and composition in equilibrium and non-equilibrium configurations with accuracy approaching DFT (i.e., the ANI-1 potential trained on 20M energies of the H, C, N, and O-containing compounds at ωB97X/6-31G(d), ANI-1x trained on 5M energies at ωB97X/6-31G(d) selected by active learning, and ANI-2x extension of ANI-1x to S, F, Cl elements)[9–11], or even coupled cluster QM level (ANI-1ccx[12] trained on 0.5M at CCSD(T)*/CBS energies using transfer learning; CCSD(T)* is an approximation to CCSD(T) based on multi-step calculations with domain-based local-pair natural-orbital-CCSD(T)[13] and CBS is an extrapolation to complete basis set, for the complete description of the technical details behind CCSD(T)*/CBS see refs. [12,14]) (Fig. 1a). The ANI potentials are transferable to much larger systems than those included in the training data set, because the total energy is calculated within the local approximation by the sum of the atomic contributions with each atom feeling the environment only within some cutoff.

While impressive, ANI potentials are however much less transferable than general-purpose QM methods, because they are limited to closed-shell, neutral organic compounds, and the use of the local approximation imposes further limitations on their transferability, e.g., to large, highly conjugated systems (Fig. 1b, c). A rational approach is to exploit synergies of AI and QM methods by merging them[7] as well as improving AI-based methods by including the effects of dispersion and long-range interactions[7,15–19]. This approach has already given rise to an increasing number of hybrid AI/QM methods[7,8,20–22], although most of them are either proof-of-principle or based on relatively slow DFT or trained on data of
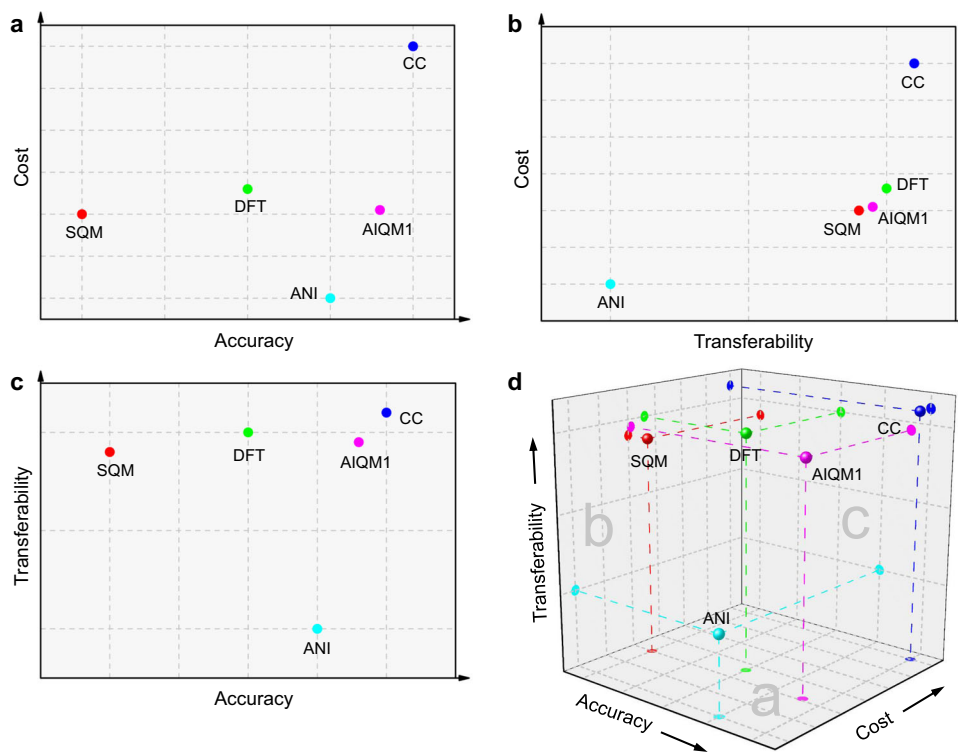

Fig. 1 Simplified scheme of quantum chemistry approximations. Traditional quantum mechanical approaches such as the gold-standard coupled cluster (CC, blue), work-horse density functional theory (DFT, green), fast and approximate semiempirical quantum mechanical (SQM, red) methods, artificial intelligence-based ANI (cyan), and the new artificial intelligence–quantum mechanical method 1 (AIQM1, magenta). They are compared with respect to **a** cost and accuracy, **b** cost and transferability, **c** accuracy and transferability, and **d** cost, accuracy, and transferability. The accuracy of ANI is with respect to what it is applicable to. The scheme is an authors' qualitative interpretation supported by the results of this work.
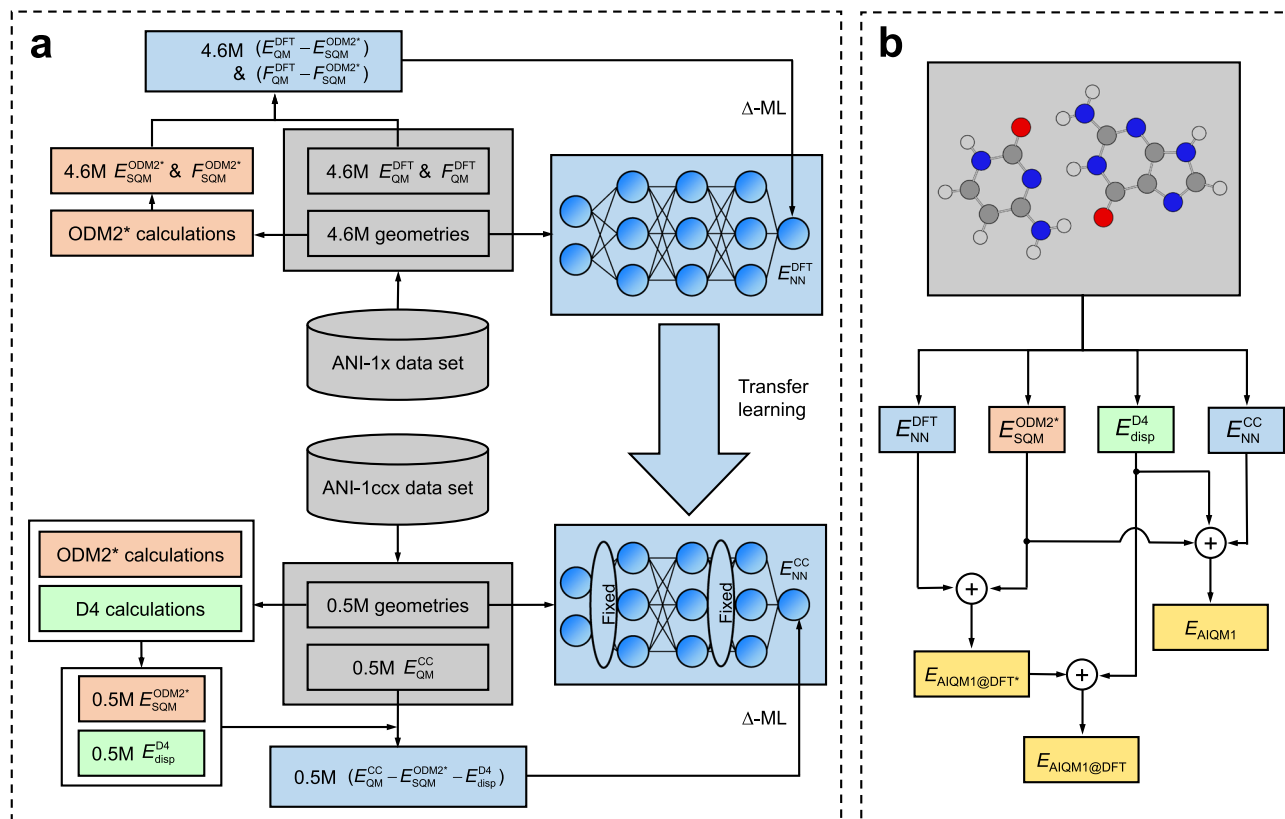
Fig. 2 The design of the artificial intelligence–quantum mechanical method 1 (AIQM1). a Flowchart of training the neural network (NN) part of the AIQM1, AIQM1@DFT*, and AIQM1@DFT methods. **b** Schematic representation of the components of the AIQM1, AIQM1@DFT*, and AIQM1@DFT methods (yellow). DFT denotes density functional theory data at ωB97X/def2-TZVPP, CC— approximation to coupled cluster with single, double, and perturbative triple excitations with complete basis set extrapolation scheme (CCSD(T)*/CBS), E energies, F forces, ODM2*—semiempirical quantum mechanical (SQM) orthogonalization- and dispersion-corrected method 2 (light brown), D4—the fourth generation of dispersion (subscript disp) corrections (green), NN – neural networks of ANI-type (blue; corrections to learn also blue). Molecular data sets are in gray.

limited quantity and quality potentially restricting their transferability and accuracy.

Here we describe the general-purpose artificial intelligence–quantum mechanical method 1 (AIQM1) that approaches the coupled cluster accuracy with transferability of the QM methods and computational speed of the SQM methods (Fig. 1d). AIQM1 stands on the shoulders of decades-long method development in SQM methods[5] as well as more recent advances in developing NN potentials[15] and combining QM with AI such as Δ-learning[23], leveraging the power of transfer learning for exploiting limited amount of high-level reference data[12], extensive developments in treating dispersion corrections[24,25], efforts in accelerating high-level approaches[13], and hard work and lots of resources invested in generating highly-accurate diverse reference data[14]. 1 in AIQM1 stands for the first iteration of the method as we envision that AIQM approaches will be further refined by using better reference data for training and making changes to the methodology, which is currently the topic of ongoing work in our labs.

Results

Method structure. The AIQM1 method consists of three main parts (Fig. 2): (1) SQM Hamiltonian, (2) neural network (NN) correction to the potential, and (3) dispersion corrections. The AIQM1 total energy E_{AIQM1} is the sum of the contributions from these three parts, E_{SQM}, E_{NN}, E_{disp}, respectively:

$$E_{AIQM1} = E_{SQM} + E_{NN} + E_{disp}. \quad (1)$$

For the first part, we have chosen the orthogonalization- and dispersion-corrected method 2 (ODM2) Hamiltonian[26], which provides the most consistent and accurate predictions across different properties (from ground-state to excited-state and noncovalent interactions) among other SQM methods, particularly those based on neglect of diatomic differential overlap (NDDO) approximation. We remove the original D3-based dispersion corrections from the ODM2 approach and denote the modified approach as ODM2*. Instead, we add the state-of-the-art D4-dispersion corrections[24,25] including Axilrod–Teller–Muto three-body contributions[27,28]—the third part of the AIQM1 method. Dispersion corrections are essential to properly describe dispersion terms in noncovalent interactions as they are poorly described by both SQM[5] and local NN approaches such as ANI-1ccx[29], and these corrections are therefore often added to AI approaches[7,15–19]. For the second part, we took the ANI-type of NN potentials. We preserved the NN-architecture of ANI-1x that predicts E_{NN} by summing over N_{atoms} atomic contributions E_A:[11]

$$E_{NN} = \sum_A^{N_{atoms}} E_A. \quad (2)$$

We made only two minor modifications to NN model based on ANI-1x. First, we changed the activation function to GELU (Gaussian error linear unit) instead of CELU (continuously-differentiable exponential linear unit), because GELU is infinitely differentiable. This is important for applications where higher derivatives are required, e.g., geometry optimization and frequency

calculations. Second, we increased the angular cutoff to 4 Å to assist with a better description of long-range interactions. Note that within ANI framework, the scalar values to learn are centered before fitting NN, i.e., the atomic contributions also include element-dependent terms obtained by linear fitting to the reference scalar values.

Our NN corrections only depend on structural parameters calculated for atoms within a cutoff. Thus, these corrections have the same limitations as the ANI models and the increased transferability of AIQM1 comes from the SQM Hamiltonian and dispersion corrections. For example, NN corrections are exactly the same for the same molecular geometry, regardless of the molecular charge or electronic state. In the future, when accurate reference data exists for diverse charged and/or electronically excited species, NN corrections can be improved by taking into account charge/electronic state as, e.g., was recently done in ref. [30].

Method training and validation. We fitted NN using the ANI-1x and ANI-1ccx data sets[14], which contain small neutral, closed-shell molecules in ground state with up to 8 non-hydrogen atoms and only considers molecules with the H, C, N, and O elements. Most of the molecules are drug-like and oligopeptides. The data sets cover not only equilibrium geometries, but also conformational space by using various sampling techniques, such as normal mode sampling and dynamics. The ANI-1x data set contains 5M geometries in total, for which ωB97X/6-31G* energies and forces were calculated (these energies were used to fit ANI-1x model). For 4.6M geometries of the ANI-1x data set, ωB97X/def2-TZVPP energies and forces are available (used previously to fit another successful general-purpose NN potential AIMNet[31]). Only energies for 0.5M geometries are available at CCSD(T)*/CBS in the ANI-1ccx data set (used to fit the ANI-1ccx potential). Such a choice of the data sets ensures high accuracy and transferability, but because of its composition, the best accuracy of AIQM1 is expected to be for ground-states energies and forces of neutral, closed-shell molecules, and it is only applicable to species with elements H, C, N, and O.

The NN weights were obtained in two steps (Fig. 2). In the first step, we fitted NN weights on the differences between the ground-state potentials calculated at DFT ωB97X/def2-TZVPP and ODM2* (see Fig. 3a for the distribution of the learned differences). This step is based on the Δ-learning[23] approach introduced by one of us and used here to correct the low-level SQM method to the target accuracy of the higher-level DFT method with comparatively small additional computational cost. (Calculations for the entire ANI-1x data set on a single CPU are ca. 10 times faster with a single ANI-type network NN compared to SQM calculations, but the difference should become larger for bigger systems and parallel computing.) The loss function L in this step is the geometric mean of the loss functions for energy differences between DFT and ODM2* (L_E, scalar values) and differences in forces (L_F, energy gradients $\partial E_{NN}/\partial R$ taken with opposite sign, vector values):

$$L = \sqrt{L_E L_F}, \qquad (3)$$

with L_E and L_F defined analogously to the loss functions for energies and forces used in ANI-2x[9].

In this way we trained an ensemble of eight NNs (Fig. 2), which provides better accuracy than a single NN[12] (see Methods). The method obtained in this first step is denoted by AIQM1@DFT* and it approaches DFT accuracy at the SQM cost for the hold-out test set as its mean absolute deviation (MAD) is only 0.7 kcal/mol for energies and 1.6 kcal/mol/Å for forces (Fig. 3b).

Since AIQM1@DFT* has no explicit dispersion corrections, we add the D4-dispersion corrections fitted[25] for the DFT functional ωB97X and denote the resulting method as AIQM1@DFT.

In the second step of NN fitting (Fig. 2), we used transfer learning[32] to reach coupled cluster accuracy using the 0.5M data points of the ANI-1ccx data set as was done for creating ANI-1ccx method[12]. Transfer learning is a powerful technique allowing to leverage more abundant training data for a related task to obtain the model for the target task using much fewer training points. For developing the AIQM1 method, we fixed the weights of the first and third hidden layers of NN from the first step to only minimize the loss function L_E for differences between the ground-state energies at CCSD(T)*/CBS and ODM2* with D4 corrections (forces are not available at CCSD(T)*/CBS and thus not included for training; see Fig. 3c for the distribution of the learned differences in energies). The resulting approach is our final AIQM1 method and it closely approaches coupled cluster level for the hold-out test set as its MAD for energies is 0.8 kcal/mol (Fig. 3d). Note that although forces and Hessians are not available at CCSD(T)*/CBS, both forces and Hessians can be easily calculated with AIQM1 with little computational cost as first- and second-order derivatives are implemented for all AIQM1 components (ODM2*, NN, and D4), which, as we will see later, is of great significance.

In the following we perform validation of our method AIQM1 and its parent variants AIQM1@DFT and AIQM1@DFT* on the independent test sets not used for fitting their NN parts. Wherever possible, we compare their performance for a range of established methods such as ODM2 (as one of the best SQM methods), B3LYP/6-31G* (because of its popularity), ωB97X/6-31G* (because it was used for generating reference data for early ANI-1 and ANI-1x models), ωB97X-D/6-31G* (as a popular representative of range-corrected DFT methods), ωB97X/def2-TZVPP (because it was used for generating reference data for AIQM1@DFT and AIQM1@DFT*), ωB97X-D4/def2-TZVPP (to test the effect of the D4 corrections), CCSD(T)*/CBS (because it was used for generating reference data for AIQM1), ANI-1ccx (best representative of the general-purpose NN potentials), and, occasionally, other relevant methods.

We cannot compare AIQM1 to ANI-1ccx for ions, radicals, and excited states, as ANI-1ccx is not transferable to such cases and they were excluded from statistics; in addition, there is no implementation for heats of formation at ANI-1ccx. No comparisons to CCSD(T)*/CBS-optimized geometries were done, because of prohibitive computational cost for such calculations and absence of implementation of analytical derivatives. This method cannot be used for excited states either. To prevent the paper from becoming unwieldy, we only focus in this text on the most important benchmarks, while the summary of calculations with all aforementioned methods can be found in Fig. 4 and Supplementary Data 1 sheet S1 and details (overview of data sets, list of compounds, reference data, and data calculated with above methods, etc.) are provided in the Supplementary Data 1 sheets S3–S20.

Performance for energies. AIQM1 has an excellent accuracy in energies for a broad range of data sets not used for fitting its NN part (Fig. 4). A very important energy-based property is heat (enthalpy) of formation—a fundamental thermochemical quantity, which is notoriously difficult to accurately predict with quantum chemistry. Typically, only very computationally expensive QM methods are able to achieve the desired chemical accuracy for heats of formation ΔH_f (errors below 1 kcal/mol). Thus, AI was suggested as a potent approach to specifically target accurate and cost-efficient predictions of heats of formation by improving upon predictions made by the low-cost QM methods
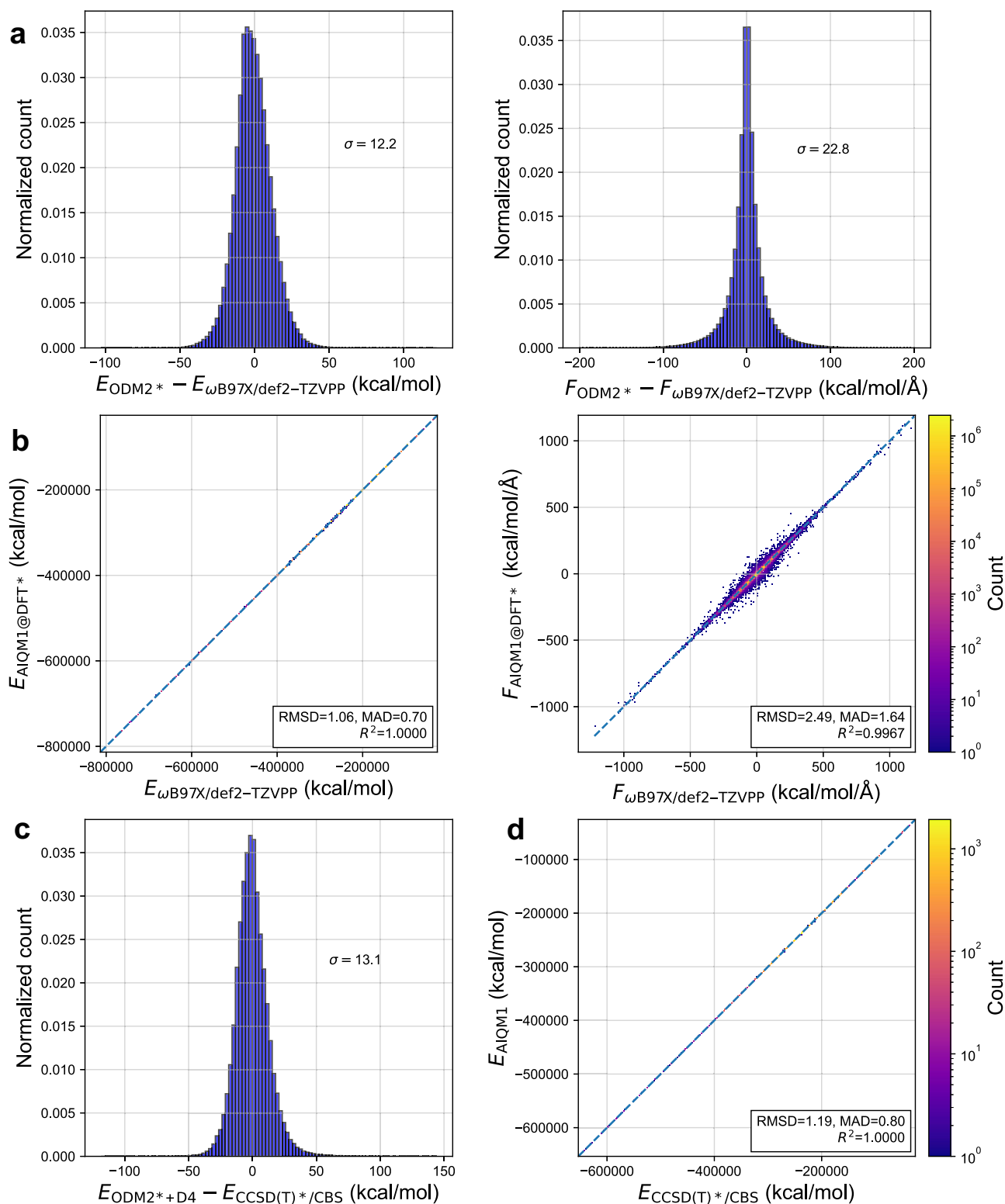
Fig. 3 Correlation between the artificial intelligence–quantum mechanical method 1 (AIQM1) variants and reference methods for the hold-out test set.
a Histogram of learned centered differences between ODM2* and ωB97X/def2-TZVPP energies (E, left) and forces (F, right). **b** Correlation between AIQM1@DFT* and ωB97X/def2-TZVPP energies and forces. **c** Histogram of learned centered differences between ODM2* + D4 and CCSD(T)*/CBS energies. **d** Correlation between AIQM1 and CCSD(T)*/CBS energies. Root-mean-squared deviations (RMSDs), standard deviations (σ), mean absolute deviations (MADs), and squared correlation coefficients R^2 are also shown on the plots. Errors for energies are in kcal/mol, for forces—in kcal/mol/Å.

	ODM2	B3LYP/6-31G*	ωB97X/6-31G*	ωB97X-D/6-31G*	ωB97X-D/def2-TZVPP	ωB97X-D4/def2-TZVPP	ANI-1ccx	AIQM1@DFT*	AIQM1@DFT	AIQM1	CCSD(T)*/CBS
energies, kcal/mol											
CHNO	2.64	6.71	4.10	3.84	3.21	2.76	—	2.49	2.12	0.87	—
G3/99	3.04	8.53	3.46	3.22	4.18	3.20	—	2.83	2.06	0.88	—
ISOMERS44 (ΔH_f)	1.16	8.08	3.57	3.53	4.52	3.78	—	3.00	2.27	0.42	—
ISOMERS44 (ΔH_r)	0.70	2.29	1.45	1.31	1.19	1.10	1.68	0.95	0.89	0.50	—
IsoL6/11	1.48	5.26	3.83	3.36	1.75	1.64	1.46	1.65	1.55	0.62	0.47
HC7/11	5.37	6.44	16.9	13.98	6.83	7.10	2.53	8.89	9.16	1.43	1.57
Torsion	0.74	0.55	0.30	0.29	0.20	0.19	0.23	0.23	0.23	0.19	0.05
S66x8	0.75	1.70	0.97	1.22	0.49	0.49	2.20	0.69	0.52	0.56	0.09
WATER27	4.45	36.37	41.28	36.95	14.07	15.26	—	11.59	12.77	2.07	2.57
PA	16.62	8.04	7.33	8.38	3.75	3.72	—	16.04	16.02	10.53	0.32
G21IP	10.17	6.67	7.03	6.98	4.57	4.56	—	8.69	8.68	8.82	3.26
G21EA	13.93	28.51	28.41	27.72	8.42	8.41	—	12.94	12.93	13.97	8.09
excitation energies, eV											
Thiel's set	0.35	0.32	0.45	0.36	0.36	0.36	—	0.35	0.35	0.35	—
bond lengths, Å											
CHNO	0.015	0.006	0.008	0.007	0.010	0.010	0.011	0.010	0.010	0.007	—
MGHBL9	0.023	0.007	0.006	0.005	0.002	0.002	0.047	0.011	0.011	0.004	—
MGNHBL11	0.026	0.006	0.003	0.002	0.008	0.008	0.004	0.008	0.008	0.002	—
CATIONS41	0.023	0.011	0.009	0.010	0.011	0.011	—	0.015	0.015	0.017	—
bond angles, °											
CHNO	2.04	0.70	0.68	0.64	0.68	0.68	1.00	0.77	0.77	0.70	—
CATIONS41	2.21	0.52	0.37	0.44	0.46	0.46	—	1.48	1.48	1.26	—
dihedral angles, °											
CHNO	4.07	5.20	4.68	6.10	7.12	7.11	5.86	2.14	2.14	2.31	—
CATIONS41	2.49	0.31	0.32	0.40	0.19	0.19	—	0.76	0.76	0.72	—

Fig. 4 Performance of tested methods for diverse benchmarks. Mean absolute deviations (MADs) are provided for all methods and benchmark sets used in this study. MADs are color coded for easier visual analysis—red values have the highest MAD, green values—the lowest. "——" is used where (all) calculations were not possible. Data set abbreviation are their established names, i.e., CHNO—compounds with H, C, N, and O elements, G Gaussian, IsoL isomerization of larger molecules, HC hydrocarbons, IP ionization potentials, EA electron affinities, MGHBL main-group hydrogenic X–H bond lengths, MGNHBL11 main-group non-hydrogenic X–Y bond lengths; the numbers in data set names usually indicate either the original number of data points in the data set or a year it was assembled.

(DFT[33–35] and SQM[36] methods). In contrast, in our approach we did not fit NN part to better reproduce the heats of formation; we merely had to offset the bias in AIQM1 heats of formations at 298 K with respect to the experimental reference data in the CHNO data set[37] by just fitting four parameters—atomic energies of H, C, N, and O elements, which we treat as energies of free atoms in the most stable electronic configuration at AIQM1 (Methods). The CHNO data set includes carefully curated 138 heats of formation of various molecules ranging from inorganic (H_2, O_2, H_2O, NH_3, etc.) to diverse classes of organic compounds (alkanes, alkenes, alkynes, linear and cyclic compounds, molecules with different functional groups, e.g., alcohols, amines, acids), which allowed development of general-purpose, transferable SQM methods[26,37]. This set consists of compounds with only H, C, N, and O elements, hence the name.

AIQM1 performance is remarkable for heats of formation as it easily reaches chemical accuracy for the CHNO data set (MAD of 0.9 kcal/mol), even though this property was not included in the training set of its NN part. It is the first time that a QM method with semiempirical speed has broken this threshold as, e.g., ODM2 method with the best-reported accuracy among semiempirical methods to date has three times higher MAD of 2.6 kcal/mol. Similarly, AIQM1 has MAD of 0.9 kcal/mol in heats of formation for the CHNO subset[38] of the independent G3/99 test set (G stands for Gaussian)[39]. This subset contains species only with H, C, N, and O elements and includes 47 experimental heats of formation of medium-sized organic species (e.g., piperidine, acetal, azulene, phenyl radical etc.). The full G3/99 set formed a backbone for developing and testing many QM methods such as popular, but very computationally expensive composite approaches Gaussian-4 (G4)[40] and G4MP2[41]

(approximation of G4 for faster calculations) targeting the coveted chemical accuracy. AIQM1 accuracy for both the CHNO and G3/99 sets is on par with G4 and G4MP2 (their MADs are in the range of 0.65–0.90 kcal/mol, see the Supplementary Data 1 sheets S4 and S5) and thus AIQM1 can be used as a computationally-efficient alternative to such composite methods. Noteworthy, AIQM1 is clearly better than DFT approaches tested here (Fig. 4).

It is important to point out the limitations of the AIQM1 as well. For example, analysis of heats of formation shows that AIQM1 has relatively large error of −2.9 kcal/mol for the H_2 molecule, which is similar to DFT approaches (error up to 3.7 kcal/mol at ωB97X/6-31G*), but much larger than errors at G4 (−0.3 kcal/mol) and G4MP2 (−1.0 kcal/mol) (see Supplementary Data 1 sheet S4). Possible cause for such a large error is the lack of H_2 in the ANI-1x data set used for fitting NN part of AIQM1. This example shows that AIQM1 accuracy may deteriorate significantly for cases underrepresented in its training set, regardless whether molecular structures are simple or not. On the other hand, it also shows the path to overcome such problems—by including more such cases in the training set in the future.

In chemistry, we often have to deal with such relative energies as isomerization energies, reaction energies, and enthalpies as well as relative energies between conformers, because relative energies determine the outcome of reactions and 3D structures of molecules in thermal equilibrium. AIQM1 not only has good accuracy for heats of formation, but also faithfully reproduces other types of relative energies. One example is the heats of formation ΔH_f and isomerization enthalpies ΔH_r at 298 K of organic compounds in the ISOMERS44 data set[38,42], for which AIQM1 has MAD of 0.4 and 0.5 kcal/mol, respectively. The
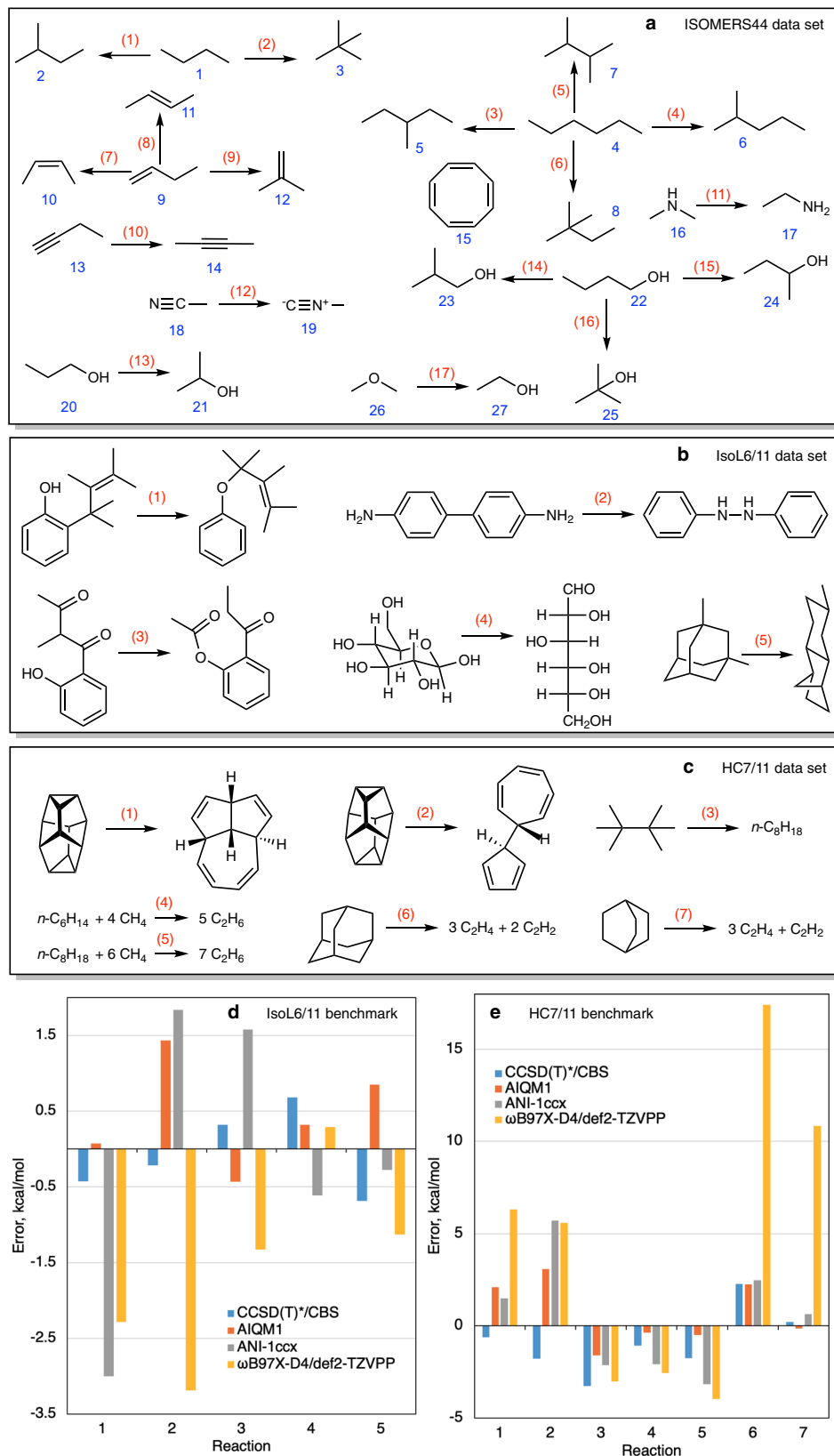
Fig. 5 Selection of data sets for testing performance of the artificial intelligence–quantum mechanical method 1 (AIQM1) for ground-state energies.
a ISOMERS44 set for testing accuracy in predicting heats of formation for 27 molecules (blue numbers) and 17 reaction enthalpies (red numbers),
b reactions in IsoL6/11 set used to test accuracy in isomerization energies, **c** reactions in HC7/11 set used to test relative energies. Comparison between
errors of CCSD(T)*/CBS, AIQM1, ANI-1ccx, and ωB97X-D4/def2-TZVPP for the reaction energies in the **d** IsoL6/11 and **e** HC7/11 benchmark sets.
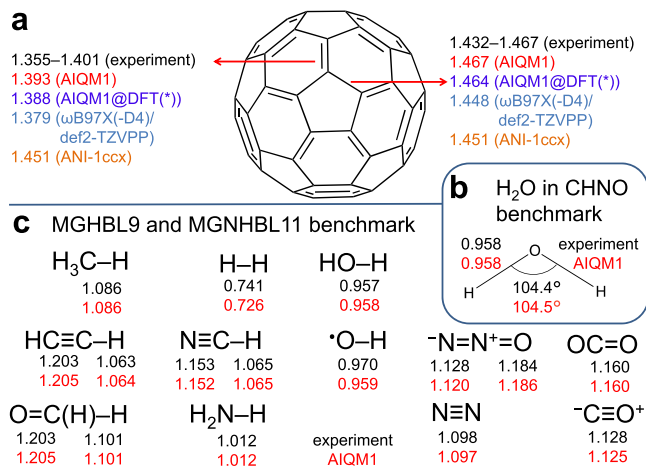
a

1.355–1.401 (experiment)
1.393 (AIQM1)
1.388 (AIQM1@DFT(*))
1.379 (ωB97X(-D4)/
def2-TZVPP)
1.451 (ANI-1ccx)

1.432–1.467 (experiment)
1.467 (AIQM1)
1.464 (AIQM1@DFT(*))
1.448 (ωB97X(-D4)/
def2-TZVPP)
1.451 (ANI-1ccx)

b H_2O in CHNO benchmark

0.958   experiment
0.958   O   AIQM1
H   104.4°   H
                104.5°

c MGHBL9 and MGNHBL11 benchmark

$H_3C–H$
1.086
1.086

H–H
0.741
0.726

HO–H
0.957
0.958

$HC≡C–H$
1.203   1.063
1.205   1.064

$N≡C–H$
1.153   1.065
1.152   1.065

$•O–H$
0.970
0.959

$^-N=N^+=O$
1.128   1.184
1.120   1.186

$OC=O$
1.160
1.160

$O=C(H)–H$
1.203   1.101
1.205   1.101

$H_2N–H$
1.012
1.012

experiment
AIQM1

$N≡N$
1.098
1.097

$^-C≡O^+$
1.128
1.125

Fig. 6 Performance of the artificial intelligence–quantum mechanical method 1 (AIQM1) for finding ground-state minimum geometries. a Short and long bond lengths in C_{60} as calculated at different levels of theory and compared to experimental values[46–49]. Note that geometries of C_{60} are the same at both AIQM1@DFT and AIQM1@DFT* as well as at both ωB97X and ωB97X-D4/def2-TZVPP. **b** All hydrogenic bond lengths in the MGHBL9 data set[50] and non-hydrogenic bond lengths in the MGNHBL11 data set[50, 51]. **c** Geometry of a water molecule (one of the molecules in the CHNO benchmark; note that experimental reference is slightly different from that in MGHBL9 set in panel **b**)[37]. Bond lengths are in Å. Experimental values in black, AIQM1 values in red, other methods marked with other colors.

ISOMERS44 set contains 27 experimental heats of formation of several different classes of compounds (hydrocarbons, alcohols, amines etc.) and 17 isomerization enthalpies derived from these heats of formations (Fig. 5a). The performance of AIQM1 for the ISOMERS44 set are therefore much better than performance of the DFT methods tested here (Fig. 4).

Other types of relative energies, such as zero-point energy-excluded reaction energies at 0 K are also reproduced by AIQM1 very well. For example, isomerization energies in the subset of the IsoL6/11 data set[43] with five reactions of compounds containing only H, C, N, and O elements are reproduced by AIQM1 with chemical accuracy (MAD 0.6 kcal/mol, Fig. 5d), while the MADs of all other methods tested here are equal or larger than 1.5 kcal/mol, except for CCSD(T)*/CBS with MAD of 0.5 kcal/mol (Fig. 4). IsoL6/11 is an acronym for a data set consisting of six isomerization energies of large organic compounds. Reference energies were calculated at CCSD(T)-F12a/aug-cc-pVDZ (see Fig. 5b for isomerization reaction schemes); the whole data set is often used for testing QM methods.

Similarly, for another data set, reaction energies in the HC7/11 set[44], AIQM1 accuracy is also very close to that[12] of CCSD(T)*/CBS (MADs of 1.4 and 1.6 kcal/mol, respectively) and clearly outperforms all other methods having MADs from 2.5 kcal/mol (ANI-1ccx) to 16.9 kcal/mol (ωB97X/6-31G*) (Fig. 4 and Fig. 5e). HC7/11 is widely used for testing QM methods and it consists of seven difficult cases for DFT including isomerization and isodesmic energies of hydrocarbon compounds; reference energies in HC7/11 are either zero-point energy-excluded experimental values or CCSD(T)/6-311 + G(d,p) (see Fig. 5c for reaction schemes).

Relative energies of the configurations of the same molecule are also important as they define, among others, what rotational conformers are more stable, which in turn is crucial for determining 3D structures of flexible molecules. AIQM1 confidently handles this task as its median MAD for the popular torsion benchmark set[45] is only 0.19 kcal/mol, which is the same as for much more expensive ωB97X-D4/def2-TZVPP and lower than other methods tested here

(median MADs range from 0.20 to 0.74 kcal/mol, Fig. 4). We used the subset of the torsion benchmark set with only H, C, N, O-containing compounds; it consists of test cases with torsion scans for 45 fragments grouped into alkyl, aryl, aryl-amide, and bi-aryl cases with torsion profiles calculated at CCSD(T)/CBS. AIQM1 is only inferior to CCSD(T)*/CBS and MP2/CBS (median MAD 0.11 kcal/mol)[45], which are however much slower than DFT methods tested here. Now we can turn into investigating the performance of AIQM1 for predicting geometries themselves.

Performance for geometries. Theoretical prediction of molecular geometries is one of the most common applications of quantum chemistry, which is essential for chemical research as conclusive geometries are not always available from experiment. Geometry optimization is an iterative procedure requiring forces (and often Hessians), which makes it much more computationally expensive than energy calculations for a single geometry. SQM methods are much less accurate for geometries than common DFT methods and general-purpose NN potentials fail to deal with subtle conjugation effects, e.g., ANI-1ccx predicts that all bond lengths in C_{60} are equal to 1.451 Å, while it is known from experiment[46–49] that bond length between two adjacent hexagon rings is shorter than bond length between pentagon and hexagon rings (Fig. 6a).

Optimization with AIQM1 forces successfully distinguishes these two bond types in C_{60} and predicts short and long bond lengths to be 1.393 and 1.467 Å, respectively (Fig. 6a). For this molecule, we cannot compare AIQM1 predictions with CCSD(T)*/CBS due to the staggering cost of this coupled cluster approach (single-point energy calculations take 69 hours on 15 CPU cores), while experimental data are not conclusive as they range from 1.355 to 1.401 Å for short bond length and from 1.432 to 1.467 Å for long bond length depending on measurement conditions[46–49]. Instead, we compare AIQM1@DFT predicting 1.388 and 1.464 Å to ωB97X-D4/def2-TZVPP predictions of 1.379 and 1.449 Å, which are in acceptable agreement (Fig. 6a), while the cost of geometry optimization with AIQM1@DFT* is 14 s on a single CPU core vs. 31 min on 32 CPU cores at DFT.

For smaller molecules, where reliable data is available, AIQM1 has very good accuracy, much better than, e.g., the accuracy of ODM2 or ANI-1ccx. AIQM1 is also more consistent than DFT methods, whose performance strongly depends on the functional and basis set (Fig. 4). For the CHNO data set[37] with experimental reference data, the MADs of AIQM1, ODM2, and ANI-1ccx are 0.007, 0.015, and 0.011 Å in bond lengths, 0.70°, 2.04°, and 1.00° in bond angles, and 2.31°, 4.07°, and 5.86° in dihedral angles, respectively (see, e.g., excellent prediction of water geometry, Fig. 6b). Similarly, for nine main-group hydrogenic X–H bond lengths (MGHBL9)[50] and 9 main-group non-hydrogenic X–Y bond lengths (MGNHBL11)[50,51] data sets with experimental data used to test DFT methods, MAD of AIQM1 in bond lengths is 0.004 and 0.002 Å (Fig. 6c), respectively, which is again much better than ODM2 (0.023 and 0.026 Å) or ANI-1ccx (0.047 and 0.004 Å). The latter two data sets contain small molecules H_2, CH_4, NH_3, H_2O, HF, C_2H_2, HCN, H_2CO, CO, N_2, F_2, CO_2, N_2O, and OH radical (investigated bond lengths are shown in Fig. 6c; OH radical was excluded from statistics of ANI-1ccx). We only excluded molecules Cl_2 and MgS from the full MGNHBL11 data set.

Predicting polyyne structures. AIQM1 opens the door for calculating geometries with previously unattainable accuracy and speed, which is crucial for compounds, whose structural determination is difficult both experimentally and theoretically. One such case is cyclo[18]carbon C_{18}, which was inspiring the imagination of
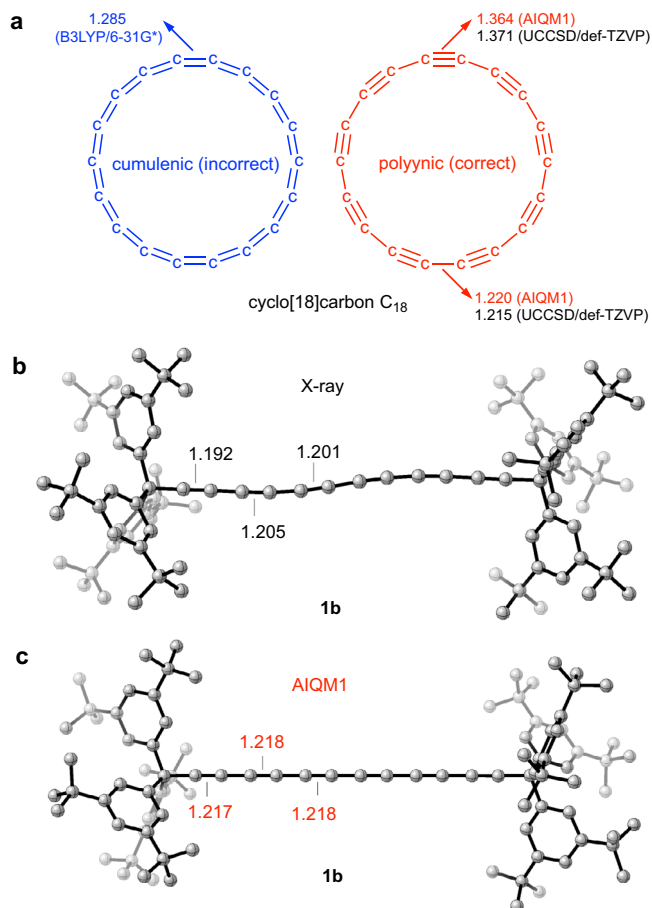
Fig. 7 Geometries of polyyne compounds. a Correct polyynic (red) geometry of C$_{18}$ predicted by AIQM1 (bond lengths in red) and UCCSD/def-TZVP (bond lengths in black from ref. [56]) and its incorrect alternative cumulenic (blue) geometry predicted by B3LYP/6-31G* (bond lengths in blue). **b** X-ray structure of **1b** compound from ref. [57] assigning too short triple C≡C bond lengths (shown in black). **c** AIQM1 geometry of **1b** with revised bond lengths (shown in red). Bond lengths are in Å. Other bond lengths and the coordinates of AIQM1 geometry are given in the Supplementary Data 1 sheet S23.

chemists from 1966[52], but whose accurate geometrical parameters are still unknown despite many efforts by both experimentalists and theoreticians[53–56]. Experiment only shows that this molecule has polyynic structure with alternating bond lengths rather than cumulenic structure with equal bond lengths[55] (Fig. 7a). C$_{18}$ is extremely challenging for QM methods too: some DFT methods predict wrong cumulenic structure (Fig. 7a), while others correct polyynic structure (Fig. 7)[53–56]. The best theoretical estimates reported to date are at coupled cluster with single and double excitations (CCSD), which neglect important triple excitations and use rather small basis sets (due to the very high computational cost of CCSD geometry optimizations, the largest basis set used was only def-TZVP)[53,56]. We optimized C$_{18}$ at AIQM1 without imposing any symmetry constraints in contrast to previous theoretical works (where such constraints were also necessary to reduce the computational cost) and report the revised best theoretical estimate of the geometry with short bond lengths of 1.220 Å and long bond lengths of 1.364 Å. These calculations only took 2 seconds on a single CPU. In retrospect, previous unrestricted CCSD (UCCSD/def-TZVP) calculations[56] (bong lengths of 1.215 Å and long bond length of 1.371 Å, Fig. 7a), are much closer to the AIQM1 result than DFT approaches benchmarked previously[56] against UCCSD/def-TZVP

(e.g., the best DFT method was found[56] to be ωB97X-D/def2-TZVP predicting 1.221 Å and 1.344 Å).

As a further demonstration of AIQM1 capabilities for polyynes, we report a structure of a free molecule of a model polyyne **1b** from ref. [57]. It has six C≡C bonds, but whopping 224 atoms due to bulky tris(3,5-di-t-butylphenyl)methyl (Tr*) end groups used for protection. As we have previously shown[58], electronic properties such as optical band gap of this molecule strongly depend on its geometry and therefore accurate optimizations of this class of compounds are of high importance. Optimization with coupled cluster methods is at the moment impossible due to their prohibitive cost for such a large number of atoms, while the truncation of structure will not fully capture the effect of the end groups. X-ray structure of **1b** is available[57]; nevertheless, it is well known[59–61] that the triple C≡C bond length determined by X-ray diffraction experiments are severely shortened due to high electron density in the middle of these bonds. AIQM1 revises the lengths of the triple C≡C bonds to be 0.013–0.025 Å longer than in previously reported X-ray structures (Fig. 7b, c). In addition, X-ray structures are significantly impacted by packing and vibrational effects depending on temperature and the measured structures have pronounced S-shaped bend, while a free standing **1b** molecule in vacuum is predicted by AIQM1 to be linear.

We hope that the future studies on these polyyne molecules with better experimental and theoretical methods can provide a conclusive, independent validation of our predictions with AIQM1. As indirect validation of AIQM1 serves its low MAD of 0.004 Å in seven triple C≡C bond lengths present in the CHNO set (see also Fig. 6c for an example of excellent accuracy of AIQM1 for the acetylene molecule in the MGNHBL11 set).

Performance for noncovalent interactions. AIQM1 is transferable to noncovalent interactions, which are very challenging even for the state-of-the-art QM methods and NN potentials. For the standard benchmark set S66x8 with CCSD(T)/CBS reference
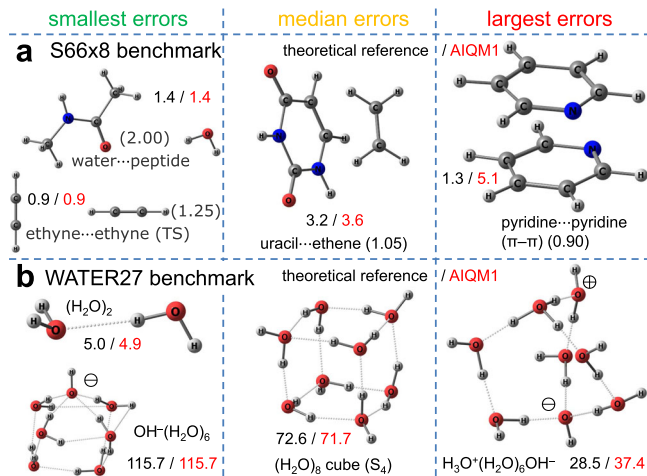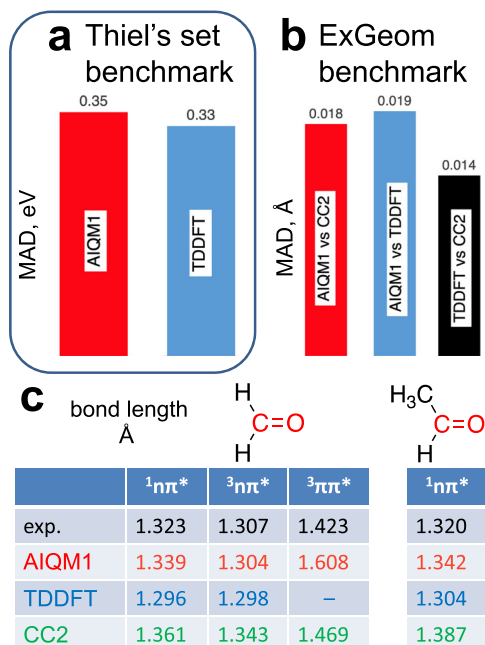

Fig. 8 Performance of the artificial intelligence-quantum mechanical method 1 (AIQM1) for noncovalent interactions. Selection of complexes with errors ranging from smallest to median to largest values for **a** the S66x8 benchmark and **b** the WATER27 benchmark. In the figure, interaction energies are shown in kcal/mol, black at reference and red at AIQM1. Other numbers such as in "(1.25) ethyne⋯ethyne (TS)" correspond to the scaling of the distance between monomers relative to equilibrium structures as used for construction of the S66x8 set.

Fig. 9 Performance of the artificial intelligence–quantum mechanical method 1 (AIQM1) for excited states. a Mean absolute error (MAD) in vertical excitation energies for Thiel's benchmark set[66] (AIQM1 in red, time-dependent density functional theory, TDDFT, in blue). **b** MAD in bond lengths for the ExGeom benchmark set[26, 66] (AIQM1 vs approximate coupled cluster singles-and-doubles method (CC2) in red, AIQM1 vs. TDDFT in blue, TDDFT vs. CC2 in black). **c** Lengths of C–O bonds in formaldehyde and acetaldehyde as compared to experiment (experiment in black, AIQM1 in red, TDDFT in blue, CC2 in green; one value for TDDFT is missing due to the failed geometry optimization at this level[66]). Values for TDDFT (B3LYP/TZVP), CC2, TDDFT vs. CC2, and experiment are taken from ref. [66].

noncovalent interaction energies[62], AIQM1 has rather good accuracy as its MAD is 0.6 kcal/mol, which is comparable to ODM2 (0.8 kcal/mol) and DFT, e.g., ωB97X-D/6-31G* (1.2 kcal/mol) and ωB97X-D4/def2-TZVPP (0.5 kcal/mol) (see Fig. 4 for MADs and Fig. 8 for selected structures). S66x8 set contains 66 noncovalent complexes in their equilibrium geometries and geometries with displaced monomers (in total 528 geometries) and it represents different types of interaction (electrostatic- and dispersion-dominated as well as mixed types). Hence, AIQM1 is a good cost-efficient alternative to many DFT methods.

The method performance is particularly good for hydrogen-bonded complexes. For 27 clusters of neutral water molecules $(H_2O)_n$, and charged clusters $H^+(H_2O)_n$ and $OH^-(H_2O)_n$ (WATER27 data set[63] with revised values[64] for $(H_2O)_{20}$ clusters), AIQM1 has MAD of only 2.1 kcal/mol compared to 4.5 of ODM2 (see Fig. 4 for MADs and Fig. 8 for selected structures). This makes the method competitive in terms of accuracy with popular dispersion-corrected DFT approaches, which have similar errors (see, e.g., ref. [64]), but are much slower. AIQM1 is therefore a promising method for simulating chemical processes in water solutions, essential for biological processes. It is noteworthy that this data set contains charged species, which can be adequately described neither by ANI-1ccx nor by the DFT methods tested here as the basis sets are not adequate for treating anionic species, which brings us to the next topic.

Beyond closed-shell, neutral molecules. AIQM1 is transferable beyond closed-shell, neutral species used for fitting its NN part

and even improves upon the ODM2 method (ANI potentials cannot be used at all for such simulations). We saw before that AIQM1 performs well for charged protonated and deprotonated water clusters. Other examples are proton affinities, where MAD is improved from 16.6 (ODM2) to 10.5 (AIQM1) kcal/mol for the proton affinities (PA) data set[63] and MAD in adiabatic ionization potentials (G21IP set)[63] from 10.2 to 8.8 kcal/mol (Fig. 4). Nevertheless, MAD in adiabatic electron affinities (G21EA set)[63] is practically the same for both ODM2 and AIQM1 (ca. 14.0 kcal/mol). All these data sets consist of experimental reference values for small compounds, and here we used only their subset with species containing at least two atoms and only H, C, N, O elements: PA has eight proton affinities of H_2, H_2O, NH_3, and five unsaturated hydrocarbons, IP21 and EA13 both have nine (albeit not the same) small organic and inorganic species (see Supplementary Data 1 sheets S13–S15 for the list of species, reference, and calculated data). In general, DFT outperforms AIQM1 (Fig. 4) for the benchmarked cationic species (PA and G21IP sets), but DFT performance has strong dependence on the basis set and, e.g., calculations with 6-31G* have similar or even larger errors than AIQM1, especially after removing the biggest outlier in AIQM1, which is the proton affinity of the H_2 molecule underestimated by −35.4 kcal/mol (see the Supplementary Data 1 sheet S13).

Anionic species (G21EA set) are even more difficult and require large, diffuse basis sets for proper QM treatment as is clear by comparing MAD of DFT approaches, which ranges from ca. 28 kcal/mol with the 6-31G* basis set to 8.4 kcal/mol with larger def2-TZVPP basis set; even CCSD(T)*/CBS has a large error of 8.09 kcal/mol (Fig. 4). Thus, a rather large error of AIQM1 (14.0 kcal/mol) is not surprising and the proper treatment of electron affinities remains a big challenge to be addressed in the future.

Interestingly, geometries are also improved for charged species as for the CATIONS41 data set[38,65], the MADs of AIQM1 and ODM2 are 0.017 and 0.023 Å in bond lengths, 1.26° and 2.21° in bond angles, and 0.72° and 2.49° in dihedral angles, respectively. The CATIONS41 data set consists of 75 bond lengths, 38 bond angles, and five dihedral angles, determined experimentally and by using high-level theoretical methods, of small organic (CH^+, $C_2H_3^+$, $C_2H_5^+$, propargyl cation, cyclopropenyl cation etc.) and inorganic (triplet and singlet OH^+, NO^+, NH_4^+ etc.) species. Tested cations are, however, better described by DFT (Fig. 4) than by AIQM1.

All in all, there is clearly a room for improvement of AIQM1 method for ionic species. Nevertheless, all the tests were performed here for rather small molecules, for which reliable reference data is available, while in case of larger systems, where the charge is more delocalized, AIQM1 is expected to perform better as the electronic density will be more similar to the corresponding neutral species.

Beyond ground-state properties. Finally, AIQM1 method is also transferable to electronically excited states and, e.g., it can be used for multi-reference configuration interaction (MRCI) calculations to predict excitation energies, oscillator strengths and non-adiabatic couplings for simulating spectra and performing non-adiabatic excited-state dynamics. Here we use the graphical unitary-group approach (GUGA) and the same settings (active space, excitation levels, etc.) for MRCI calculations as previously used for benchmarking SQM methods (see Methods for details)[26,66,67]. AIQM1/MRCI is three orders of magnitude faster than popular linear-response time-dependent (TD) DFT approaches such as TD-B3LYP, TD-ωB97X, and TD-ωB97X-D, while the accuracy in vertical excitation energies is similar for

these methods (MAD of AIQM1/MRCI is 0.35 eV, which is close to TD-DFT methods with MAD of 0.32–0.45 eV for the Thiel's data set[66,67], Fig. 4 and Fig. 9a). Thiel's set is often used for benchmarking QM methods and consists of 167 singlet and triplet vertical excitation energies for several states of 28 middle-sized organic compounds represented by unsaturated linear and cyclic hydrocarbons as well as heterocycles calculated with multistate multiconfigurational second-order perturbation theory (MS-CASPT2/aug-cc-pVTZ) for most compounds and with equation-of-motion (EOM)-CCSD(T)/aug-cc-pVTZ for nucleobases cytosine, thymine, and adenine.

MRCI calculations are performed using the SQM (ODM2*) Hamiltonian of AIQM1 and thus, excitation energies are trivially the same as in ODM2* and ODM2. However, when calculating forces for molecules in excited states, NN corrections to forces are added and their effect is not clear as they were trained on ground-state reference data. Thus, we test AIQM1/MRCI forces, by performing geometry optimizations of molecules in excited-states. Such optimizations are of large importance and required, e.g., for simulating fluorescence spectra, but they are very computationally expensive with QM methods and thus the low-cost of AIQM1 makes it potentially attractive for this task. We tested AIQM1/MRCI performance on the ExGeom set[26,66] with excited-state geometries and AIQM1/MRCI MAD for bond lengths is 0.018 Å vs. the approximate coupled cluster singles-and-doubles method (CC2) reference (with TZVP basis set) and 0.019 Å vs. TDDFT reference (specifically, TD-B3LYP/TZVP). This is rather good result given that uncertainties of the reference calculations are in the same order of magnitude (MAD of TDDFT reference vs. CC2 reference is 0.014 Å, Fig. 9b)[66]. The ExGeom set consists of more than 500 reference C–C, C–H, C–O, C–N, and N–H bond lengths of 32 molecules of different classes (e.g., aldehydes, ketones, nucleobases, heterocycles) in different excited states with altogether 100 excited-state equilibrium geometries. Accurate experimental values are very hard to obtain. However, for the available experimental bond lengths in the ExGeom data set, AIQM1/MRCI gives better or similar predictions compared to TDDFT and CC2 for C–O bond in $^1n\pi^*$ and $^3n\pi^*$ excited states, while its error is much bigger for the $^3\pi\pi^*$ excited state of formaldehyde (Fig. 9c).

Overall, AIQM1 seems to be a better choice than current routinely used QM methods in terms of performance/cost ratio at least for some types of excitations, which holds a great promise for using this method for exploration of dynamical properties arising from the manifold of electronic states, e.g., by performing nonadiabatic excited-state dynamics, which should be an interesting topic for future explorations. In any case, the AIQM1 method is only the first step in the direction of creating a general-purpose AI-based method for excited-state simulations—an important, but open topic in chemistry[68]—as obviously training models on excited-state properties will be crucial for future improvements.

Discussion
After initial excitement about great promises AI holds for sub-stituting QM methods, the focus is shifting towards tighter integration of AI with QM instead of substituting QM altogether. This shift is motivated by the need to incorporate correct physical behavior of QM methods, while at the same time exploiting great ability of AI to improve low-level QM methods' accuracy without compromising their speed.

In this work, we have made a step towards creating general-purpose AI-improved QM methods useful for a variety of applications out-of-the-box. Our approach AIQM1 synergistically combines the best of two worlds—transferability of QM and high

accuracy of AI approaches. The success of this approach only became possible with great advances over recent years in methodology development of both QM and AI components as well as generation of numerous carefully curated, high quality reference data. Thus, AIQM1 allows very accurate prediction of ground-state properties such as energies and geometries of closed-shell, neutral organic compounds approaching the gold-standard CCSD(T)/CBS at the speed of semiempirical QM methods. Remarkably, it has accuracy improved in comparison to the parent SQM method (ODM2) also for other cases, not explicitly considered during training of its NN part, e.g., for charged species, showcasing the benefits of using physically-motivated AI. Thus, AIQM1 method has the potential to become a very useful tool for routine simulations with high accuracy.

It is only the beginning of the exciting road for AI-improved QM methods for general-purpose applications. In the near future we expect tighter integration of AI with QM, further optimizing both AI and QM parts, training on more and higher quality reference data, and further extending transferability and accuracy for all properties of interest to chemists and physicists.

Methods
Neural network training. The neural network training and evaluation was performed with the TorchANI software[69]. Each NN-part of AIQM1@DFT* consists of an ensemble of eight ANI-type NNs, which provides better accuracy according to our tests. The ensemble was trained similarly to the previous procedure[12], i.e., the data set was split into nine equal parts, with one part held out for testing and the remaining eight parts were used as cross-validation splits for training eight networks. Each network was trained on seven cross-validation splits and validated on one split using standard rotation of splits. During the training of AIQM1@DFT*, we stopped training NN after 1000 epochs, because we found that longer training does not improve much the performance for the validation set, but deteriorates performance for some of the external data sets. When we analyzed the error between AIQM1@DFT* predicted values and reference DFT values, we found several outliers with error >0.01 a.u. By recalculating the DFT values for these outliers, we found their reference values in ANI-1x data set were wrong, so we used the updated values to train our models. Transfer learning was then used to refit above eight ANI-type networks to 80% of the entire set with CCSD(T)*/CBS values to obtain the final NN part of AIQM1 consisting of ensemble of eight NNs; other 10% were used as the validation set and remaining 10% as the hold-out test set.

Calculation of enthalpies. The enthalpies at 298 K were calculated within harmonic oscillator and rigid rotor approximation in our locally modified version of the MNDO program[70]. Calculating heats of formation requires the evaluation of the atomization energies, which depend on the choice of the atomic energies. Atomic energies calculated with CCSD(T)*/CBS used for fitting NN-part of AIQM1 lead to large errors in atomization energies even for moderate-sized molecules such as naphthalene (error of 25.4 kcal/mol with respect to CCSD(T)/CBS, where the two-point extrapolation scheme was used with cc-pVDZ and cc-pVTZ basis sets); thus we fitted atomic energies of H, C, N, and O elements to reduce the error in heats of formation in the CHNO set. Heats of formation calculated at AIQM1@DFT and AIQM1@DFT* use atomic energies calculated with ωB97X/def2-TZVPP. All values of atomic energies are reported in the Supplementary Data 1 sheet S2.

Heats of formation at other levels (G4, G4MP2, DFT) were calculated using a standard procedure[71]. The procedure for calculating heats of formation with the MNDO program for ODM2, AIQM1, AIQM1@DFT*, and AIQM1@DFT is equivalent, but directly uses experimental reference values for heats of formation of atomic species at 298 K[5], which are slightly different than those used in G4, G4MP2, and DFT.

Electronic structure and benchmark calculations. All ODM2 and ODM2* calculations were carried out with the MNDO program[70]. CCSD(T)*/CBS calculations were performed with the ORCA 4.2.0 software package[72,73] following the procedure defined in literature[12]. The ωB97X-D4 calculations were performed with ORCA 4.2.0, and ωB97X-D calculations were performed with Gaussian 16[74]. The ωB97X/6-31G* calculations were performed with Gaussian 16, while ωB97X/def2-TZVPP calculations were performed with ORCA 4.2.0. D4-dispersion corrections were calculated with the dftd4 program[75]. We performed benchmarks of AIQM1, AIQM1@DFT*, and AIQM1@DFT with the locally modified version of the MNDO program[70] interfaced to TorchANI[69] and dftd4[75]. For benchmarking excited-state properties with AIQM1/MRCI, we used the same settings (active spaces, excitation levels, etc.) as in MRCI calculations with ODM2/MRCI[26,66,67]. Specifically, for benchmarking vertical excitations in the Thiel's set, we used the single-reference CISDTQ, which closely approximates full CI and is more accurate than, e.g., MR-CISD approximation to full CI; the active spaces include all π molecular orbitals for π → π* excitations and also

include lone-pair molecular orbitals for $n \rightarrow \pi^*$ excitations; the starting reference electronic configuration is the ground-state SCF determinant. For benchmarking excited-state geometry optimizations with MRCI, we used MR-CISD level in most cases as well as MR-CISDT and MR-CISDTQ in a few cases; the active spaces and reference electronic configurations are the same as provided in the Supporting Information of ref. [66]. All the data for benchmarks can be found in the Supplementary Data 1. The Supplementary Data 2 with Cartesian coordinates for the CHNO, CATIONS41, and ExGeom data sets is also provided.

Geometry optimizations. The ωB97X/def2-TZVPP, ωB97X-D4/def2-TZVPP geometry optimizations were performed with the ORCA program using the default BFGS algorithm, while B3LYP/6-31G*, ωB97X/6-31G*, ωB97X-D/6-31G* geometry optimizations were performed with Gaussian 16 using the default Berny algorithm GEDIIS. For ANI-1ccx and AIQM1, the geometry optimizations are performed by interfacing to the MNDO program using the default BFGS algorithm for most data sets except for optimizations of C_{60}, C_{18}, **1b** (optimized by interfacing to Gaussian 16) and the torsion benchmark (optimized by interfacing to ASE[76] using the LBFGS algorithm).

Data availability
The data (calculated energies and optimized geometries) generated in this study are provided in the Supplementary Information. Any other relevant data are available from the authors upon reasonable request.

Code availability
The AIQM1, AIQM1@DFT, and AIQM1@DFT* methods are available as the open-source code free of charge for non-commercial and non-profit uses, such as academic research and education, as described in http://MLatom.com/AIQM1. Any other relevant code is available from the authors upon reasonable request.

References
1. Raghavachari, K., Trucks, G. W., Pople, J. A. & Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **157**, 479–483 (1989).
2. Thomas, J. R. et al. The balance between theoretical method and basis set quality: a systematic study of equilibrium geometries, dipole moments, harmonic vibrational frequencies, and infrared intensities. *J. Chem. Phys.* **99**, 403–416 (1993).
3. Helgaker, T., Gauss, J., Jørgensen, P. & Olsen, J. The prediction of molecular equilibrium structures by the standard electronic wave functions. *J. Chem. Phys.* **106**, 6430–6440 (1997).
4. Bak, K. L. et al. The accurate determination of molecular equilibrium structures. *J. Chem. Phys.* **114**, 6548–6556 (2001).
5. Husch, T., Vaucher, A. C. & Reiher, M. Semiempirical molecular orbital models based on the neglect of diatomic differential overlap approximation. *Int. J. Quantum Chem.* **118**, e25799 (2018).
6. Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.* **87**, 897–923 (2015).
7. Dral, P. O. Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.* **11**, 2336–2347 (2020).
8. von Lilienfeld, O. A., Müller, K.-R. & Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **4**, 347–358 (2020).
9. Devereux, C. et al. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *J. Chem. Theory Comput.* **16**, 4192–4202 (2020).
10. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
11. Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less is more: sampling chemical space with active learning. *J. Chem. Phys.* **148**, 241733 (2018).
12. Smith, J. S. et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **10**, 2903 (2019).
13. Riplinger, C., Pinski, P., Becker, U., Valeev, E. F. & Neese, F. Sparse maps—a systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **144**, 024109 (2016).
14. Smith, J. S. et al. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Sci. Data* **7**, 134 (2020).
15. Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **121**, 10037–10072 (2021).
16. Yao, K., Herr, J. E., Toth, D. W., McKintyre, R. & Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* **9**, 2261–2269 (2018).
17. Unke, O. T. & Meuwly, M. PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
18. Ko, T. W., Finkler, J. A., Goedecker, S. & Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **12**, 398 (2021).
19. Muhli, H. et al. Machine learning force fields based on local parametrization of dispersion interactions: Application to the phase diagram of C_{60}. *Phys. Rev. B* **104**, 054106 (2021).
20. Manzhos, S. Machine learning for the solution of the Schrödinger equation. *Mach. Learn.: Sci. Technol.* **1**, 013002 (2020).
21. Westermayr, J., Gastegger, M., Schütt, K. T. & Maurer, R. J. Perspective on integrating machine learning into computational chemistry and materials science. *J. Chem. Phys.* **154**, 230903 (2021).
22. Zubatiuk, T. & Isayev, O. Development of multimodal machine learning potentials: toward a physics-aware artificial intelligence. *Acc. Chem. Res.* **54**, 1575–1585 (2021).
23. Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Big data meets quantum chemistry approximations: the Δ-machine learning approach. *J. Chem. Theory Comput.* **11**, 2087–2096 (2015).
24. Caldeweyher, E., Bannwarth, C. & Grimme, S. Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **147**, 034112 (2017).
25. Caldeweyher, E. et al. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **150**, 154122 (2019).
26. Dral, P. O., Wu, X. & Thiel, W. Semiempirical quantum-chemical methods with orthogonalization and dispersion corrections. *J. Chem. Theory Comput.* **15**, 1743–1760 (2019).
27. Axilrod, B. M. & Teller, E. Interaction of the van der Waals type Between three atoms. *J. Chem. Phys.* **11**, 299–300 (1943).
28. Muto, Y. Force between nonpolar molecules. *Proc. Phys. Math. Soc. Jpn.* **17**, 629–631 (1943).
29. Folmsbee, D. & Hutchison, G. Assessing conformer energies using electronic structure and machine learning methods. *Int. J. Quantum Chem.* **121**, e26381 (2020).
30. Zubatyuk, R., Smith, J. S., Nebgen, B. T., Tretiak, S. & Isayev, O. Teaching a neural network to attach and detach electrons from molecules. *Nat. Commun.* **12**, 4870 (2021).
31. Zubatyuk, R., Smith, J. S., Leszczynski, J. & Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Sci. Adv.* **5**, eaav6490 (2019).
32. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**, 1345–1359 (2010).
33. Hu, L. H., Wang, X. J., Wong, L. H. & Chen, G. H. Combined first-principles calculation and neural-network correction approach for heat of formation. *J. Chem. Phys.* **119**, 11501–11507 (2003).
34. Wu, J. & Xu, X. The X1 method for accurate and efficient prediction of heats of formation. *J. Chem. Phys.* **127**, 214105 (2007).
35. Dandu, N. et al. Quantum-chemically informed machine learning: prediction of energies of organic molecules with 10 to 14 non-hydrogen atoms. *J. Phys. Chem. A* **124**, 5804–5811 (2020).
36. Wan, Z., Wang, Q. D. & Liang, J. Accurate prediction of standard enthalpy of formation based on semiempirical quantum chemistry methods with artificial neural network and molecular descriptors. *Int. J. Quantum Chem.* **121**, e26441 (2021).
37. Dral, P. O. et al. Semiempirical quantum-chemical orthogonalization-corrected methods: theory, implementation, and parameters. *J. Chem. Theory Comput.* **12**, 1082–1096 (2016).
38. Dral, P. O., Wu, X., Spörkel, L., Koslowski, A. & Thiel, W. Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks for ground-state properties. *J. Chem. Theory Comput.* **12**, 1097–1120 (2016).
39. Curtiss, L. A., Raghavachari, K., Redfern, P. C. & Pople, J. A. Assessment of Gaussian-3 and density functional theories for a larger experimental test set. *J. Chem. Phys.* **112**, 7374–7383 (2000).
40. Curtiss, L. A., Redfern, P. C. & Raghavachari, K. Gaussian-4 theory. *J. Chem. Phys.* **126**, 084108 (2007).
41. Curtiss, L. A., Redfern, P. C. & Raghavachari, K. Gaussian-4 theory using reduced order perturbation theory. *J. Chem. Phys.* **127**, 124105 (2007).
42. Weber, W. *Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Implementierung, Parametrisierung und Anwendung* DOI, (Universität Zürich, 1996).

43. Luo, S., Zhao, Y. & Truhlar, D. G. Validation of electronic structure methods for isomerization reactions of large organic molecules. *Phys. Chem. Chem. Phys.* **13**, 13683–13689 (2011).

44. Peverati, R., Zhao, Y. & Truhlar, D. G. Generalized gradient approximation that recovers the second-order density-gradient expansion with optimized across-the-board performance. *J. Phys. Chem. Lett.* **2**, 1991–1997 (2011).

45. Sellers, B. D., James, N. C. & Gobbi, A. A comparison of quantum and molecular mechanical methods to estimate strain energy in druglike fragments. *J. Chem. Inf. Model.* **57**, 1265–1275 (2017).

46. Hawkins, J. M., Meyer, A., Lewis, T. A., Loren, S. & Hollander, F. J. Crystal structure of osmylated C_{60}: confirmation of the soccer ball framework. *Science* **252**, 312–313 (1991).

47. Hedberg, K. et al. Bond lengths in free molecules of buckminsterfullerene, C_{60}, from gas-phase electron diffraction. *Science* **254**, 410–412 (1991).

48. Liu, S., Lu, Y. J., Kappes, M. M. & Ibers, J. A. The structure of the C_{60} molecule: X-ray crystal structure determination of a twin at 110 k. *Science* **254**, 408–410 (1991).

49. Yannoni, C. S., Bernier, P. P., Bethune, D. S., Meijer, G. & Salem, J. R. NMR determination of the bond lengths in C_{60}. *J. Am. Chem. Soc.* **113**, 3190–3192 (2002).

50. Zhao, Y. & Truhlar, D. G. Construction of a generalized gradient approximation by restoring the density-gradient expansion and enforcing a tight Lieb-Oxford bound. *J. Chem. Phys.* **128**, 184109 (2008).

51. Peverati, R. & Truhlar, D. G. Exchange-correlation functional with good accuracy for both structural and energetic properties while depending only on the density and its gradient. *J. Chem. Theory Comput.* **8**, 2310–2319 (2012).

52. Hoffmann, R. Extended hückel theory—v: cumulenes, polyenes, polyacetylenes and C_n. *Tetrahedron* **22**, 521–538 (1966).

53. Arulmozhiraja, S. & Ohno, T. CCSD calculations on C_{14}, C_{18}, and C_{22} carbon clusters. *J. Chem. Phys.* **128**, 114301 (2008).

54. Diederich, F. et al. All-carbon molecules: evidence for the generation of cyclo[18] carbon from a stable organic precursor. *Science* **245**, 1088–1090 (1989).

55. Kaiser, K. et al. An sp-hybridized molecular carbon allotrope, cyclo[18] carbon. *Science* **365**, 1299–1301 (2019).

56. Liu, Z., Lu, T. & Chen, Q. An sp-hybridized all-carboatomic ring, cyclo[18] carbon: Bonding character, electron delocalization, and aromaticity. *Carbon* **165**, 468–475 (2020).

57. Chalifoux, W. A. & Tykwinski, R. R. Synthesis of polyynes to model the sp-carbon allotrope carbyne. *Nat. Chem.* **2**, 967–971 (2010).

58. Dral, P. O. & Clark, T. Semiempirical UNO–CAS and UNO–CI: method and applications in nanoelectronics. *J. Phys. Chem. A* **115**, 11303–11312 (2011).

59. Simonetta, M. & Gavezzotti, A. in *The Carbon–Carbon Triple Bond: Part 1* **1** (ed Saul Patai) 1–56 (John Wiley & Sons Ltd., 1978).

60. Müller, P. in *Crystal Structure Refinement: A Crystallographer's Guide to SHELXL* (ed Peter Müller) 152–153 (Oxford University Press, 2006).

61. Hirshfeld, F. L. Hellmann–Feynman constraint on charge densities, an experimental. *Test. Acta Cryst.* **B40**, 613–615 (1984).

62. Rezac, J., Riley, K. E. & Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **7**, 2427–2438 (2011).

63. Goerigk, L. & Grimme, S. Efficient and accurate double-hybrid-meta-GGA density functionals-evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **7**, 291–309 (2011).

64. Anacker, T. & Friedrich, J. New accurate benchmark energies for large water clusters: DFT is better than expected. *J. Comput. Chem.* **35**, 634–643 (2014).

65. Kolb, M. *Ein neues semiempirisches Verfahren auf Grundlage der NDDO-Näherung: Entwicklung der Methode, Parametrisierung und Anwendung* DOI (Belgische Universität-Gesamthochschule Wuppertal, 1991).

66. Tuna, D., Lu, Y., Koslowski, A. & Thiel, W. Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks of electronically excited states. *J. Chem. Theory Comput.* **12**, 4400–4422 (2016).

67. Silva-Junior, M. R. & Thiel, W. Benchmark of electronically excited states for semiempirical methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J. Chem. Theory Comput.* **6**, 1546–1564 (2010).

68. Dral, P. O. & Barbatti, M. Molecular excited states through a machine learning lens. *Nat. Rev. Chem.* **5**, 388–405 (2021).

69. Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J. S. & Roitberg, A. E. TorchANI: a free and open source PyTorch-based deep learning implementation of the ANI neural network potentials. *J. Chem. Inf. Model.* **60**, 3408–3415 (2020).

70. Thiel, W. *MNDO, Development Version* (Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, 2019).

71. Curtiss, L. A., Raghavachari, K., Redfern, P. C. & Pople, J. A. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.* **106**, 1063–1079 (1997).

72. Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1327 (2018).

73. Neese, F. The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 73–78 (2012).

74. Frisch, M. J. et al. *Gaussian 16, Rev. A.01* (Wallingford, CT, 2016).

75. Caldeweyher, E., Ehlert, S. & Grimme, S. *DFT-D4, Version 2.5.0* (Mulliken Center for Theoretical Chemistry, University of Bonn, 2020).

76. Hjorth Larsen, A. et al. The atomic simulation environment-a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).

Acknowledgements

Author contributions

P.O.D. conceived the idea. P.Z. carried out method implementation with the help from R.Z. P.Z. carried out all calculations. P.Z. and P.O.D. performed data analysis and visualization. R.Z. generated the NN training data. P.O.D. wrote the manuscript with assistance from P.Z. All authors provided critical feedback and helped shape the research, analysis, and manuscript. P.O.D., W.W., and O.I. supervised and acquired funding for the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-27340-2.

Correspondence and requests for materials should be addressed to Olexandr Isayev or Pavlo O. Dral.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.