

Development and Evaluation of Geometry Optimization Algorithms in Conjunction with ANI Potentials

Dongxiao Hao^{1,2}, Xibing He¹, Adrian E. Roitberg,³ Shengli Zhang², Junmei Wang^{1*}

*¹Department of Pharmaceutical Sciences and Computational Chemical Genomics
Screening Center, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA 15261,*

USA

²School of Physics, Xi'an Jiaotong University, Xi'an 710049, China

³Department of Chemistry, University of Florida, Gainesville 117200, USA

*To whom correspondence should be addressed, Email: junmei.wang@pitt.edu

ABSTRACT

An efficient yet accurate method for producing large amount of energy data for molecular mechanical force field (MMFF) parameterization is on demanding, especially for torsional angle parameters which are typically derived to reproduce *ab initio* rotational profiles or torsional PESs. Recently, an active learning potential (ANI-1x) for organic molecules which can produce smooth and physically meaningful potential energy surfaces (PESs) has been developed. The high efficiency and accuracy make ANI-1x especially attractive for geometry optimization at low cost. To apply ANI-1x potential in MMFF parameterization, one needs to perform constrained geometry optimization. In this work, we first developed a computational protocol to constrain rotatable torsional angles and other geometric parameters for a molecule whose geometry is described by Cartesian coordinates. The constraint is successfully achieved by force projection for the two conjugated gradient (CG) algorithms. We then conducted large-scale assessments on how ANI-1x along with four different optimization algorithms in reproducing DFT energies and geometries for two CG algorithms, CG backtracking line search (CG-BS) and CG Wolfe line search (CG-WS), and two quasi-Newton algorithms, Broyden–Fletcher–Goldfarb–Shanno (BFGS) and low-memory BFGS (L-BFGS). Note that CG-BS is a new algorithm we developed in this work. All four algorithms take the ANI energies and forces to optimize a molecule geometry. Last, we conducted a large-scale assessment of applying ANI-1x in MMFF development in three aspects. First, we performed full optimizations for 100 drug molecules each consisting of five distinct conformations. The average root-mean-square error (RMSE) between ANI-1x and DFT is about 1.3 kcal/mol and the root-mean-square displacement (RMSD) of heavy atoms is about 0.35 Å. Second, we generated torsional

PESs for 160 organic molecules, and constrained optimizations were performed for up to 18 conformations for each PES. We found that RMSE of all the conformers is 1.23 kcal/mol. Last, we carried out constrained optimizations for alanine dipeptide with both ϕ and ψ angles being frozen. The Ramachandran plots indicates that the two CG algorithms in conjunction with ANI-1x potential could well reproduce the DFT optimized geometries and torsional PESs. We concluded that CG-BS and CG-WS are good choices for generating PESs, while CG-WS or BFGS is ideal for performing full geometry optimization. With the continuously increased quality of ANI, it is expected that the computational algorithms and protocols presented in this work will have great applications in improving the quality of an existing small molecule MMFF.

Keywords:

ANI-1x; Molecular Mechanics Force Field Parameterization; Geometry Optimization Algorithms; Constrained Geometry Optimization; Conjugated-Gradient; Quasi-Newton algorithms.

1 INTRODUCTION

The physical and chemical properties of a molecule is closely related to its geometry.¹⁻
³ It is of great importance to obtain accurate molecular geometries. The potential energy surface (PES) has been widely used to depict the relationship between different molecular geometries and their corresponding single-point energies. In computational chemistry, the most important geometric PES types include bond length PES, bond angle PES and dihedral angle PES, besides the usual electrostatics and dispersion terms. To generate a PES, first, we should generate a series of different geometries varying one or multiple geometric parameters for the molecule. However, the ideal low-energy geometries of a molecule with geometric parameters fixed at certain values are usually unknown, thus, constrained geometry optimization is needed to obtain reasonable molecular geometries during a PES generation.⁴

Accurate calculation of potential energies is the foundation of geometry optimization. Quantum-mechanical (QM) methods are very popular for molecule energy calculations and geometry optimization.^{2, 5} Considering the large computational cost of *ab-initio* methods, many approximate methods have been designed to speed up the energy calculations and geometry optimization. Approximate methods replace some of the computationally expensive integrals with empirically determined parameters, to achieve a large speedup.^{2, 6-8} However, the accuracy is also substantially degraded in comparison to the high-level *ab initio* methods due to the imposed approximations.⁹ Additionally, the computational cost of approximate methods is still quite high, potentially limiting the system size that can be treated. Empirical interatomic potentials can also be used to calculate the potential energy, which were developed through fitting to reproduce *ab initio* or experimental data.¹⁰⁻¹¹

However, molecular mechanics force field (MMFF)-based potentials may perform poorly when studying new systems as it is hard to guarantee the satisfactory transferability of force field parameters.¹²⁻¹⁶

Machine learning (ML) is revolutionizing many areas of science and technology,¹⁷ which has also been used to construct various forms of atomistic potentials and has been successfully applied in energetics study of chemical and biological systems.¹⁸⁻²² Encouragingly, ML potentials have achieved a speed-up of as much as 5 orders of magnitude compared to QM calculations without significantly loss of accuracy in predicting molecular energies. One such approach, Accurate Neural network engine for Molecular Energies (ANAKIN-ME or ANI) method successfully constructed a transferable neural network potential (ANI-1) that utilizes a modified version of the symmetry functions to construct single-atom atomic environment vectors.²³⁻²⁴ The ANI-1 potential was trained on 22 million randomly selected molecular conformations from 57 thousand distinct small molecules utilizing density functional theory (DFT) energies.²³ Furthermore, active learning-based ANI potential (ANI-1x) constructed for only 5 million conformations of molecules outperforms the original ANI-1 potential.²⁵ The active learning and transfer learning-based ANI potentials (ANI-1ccx) exceeds the accuracy of DFT, which was retrained to a much smaller data set (about 500 thousand intelligently selected conformations from ANI-1x) at the CCSD(T)/CBS level of accuracy.^{22, 26} ANI-1/1x/1ccx can do only for molecules with C, H, N, and O elements, while ANI-1/2x can also deal with S, F and Cl.

Considering the outstanding advantages of efficiency and accuracy of ANI-1x potential, it is very attractive to apply it in geometry minimization, which is an important

task in computer-aided drug design.^{4, 27-29} However, an efficient minimization engine, especially one can conduct constrained geometry minimization utilizing an ANI potential is currently not available. Although we can borrow the optimization engine implemented in the Gaussian software by taking ANI-1x as an external potential, we found that many jobs were not able to converge successfully and the usages of Central Processing Unit (CPU) are not efficient, especially for optimizations with frozen coordinates. The reason lies that the PESs generated by ML may be less smoothly in comparison with those of *ab initio* potentials. The following are the motivations of this work: (1) to develop/implement different optimization algorithms in combination with ANI-1x potential for performing full geometry optimization and generating torsional PESs with one or more torsional angles being frozen; (2) to evaluate the performance of the optimization algorithms using three molecular datasets; and (3) to work out a set of practical guidance to facilitate us to generate high-quality PESs for MMFF development. The developed optimization algorithm as well as the established guidance can also be applied with other ML-based atomistic potentials, such as ANI-1ccx. The python scripts which implement the two CG and two quasi-Newton optimization algorithms are available in GitHub (https://github.com/junmwang/pyani_mmff). Note that the ANI-1x energy and force are calculated using TorchANI²⁶ implemented as a python module.

2 METHODOLOGIES

2.1 Conjugate gradient methods

Geometry optimization is still a common and complex problem to be solved efficiently.³⁰ Popular atomistic energy optimization strategies include gradient descent

methods, CG methods and Newton's methods.³¹ The "full" Newton's method requires the second derivative matrix (Hessian matrix) of a system in order to find its extrema. The most commonly-used quasi-Newton algorithms include Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm and its low-memory extension L-BFGS,³²⁻³³ which are found to be superior to CG methods in some circumstances. However, for most systems of practical interest, it may be prohibitively expensive to compute the second derivative matrix, which is estimated from successive values of gradient. The CG method is a general method to find a local minimum using the derivatives of the potential energy with respect to the position of the atoms.

For N atoms, the optimization objective function is to minimize the potential energy:

$$V(\mathbf{r}) = V(r_{1x}, r_{1y}, r_{1z}, r_{2x}, r_{2y}, r_{2z}, \dots) \quad (1)$$

An obvious, but not very efficient strategy is to move an atom in the direction of the negative gradient. To improve optimization efficiency, CG uses a new “conjugate” direction of movement that depends on the previous direction of movement, and adopts an iterative procedure taking the form of:

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \alpha \mathbf{h}_i \quad (2)$$

in which $\alpha > 0$ is a step length scalar and $\mathbf{h}_i$ is a search direction. The search direction at the very first iteration is the direction of the negative gradient: $\mathbf{h}_0 = \mathbf{g}_0 = \mathbf{f}_0 \equiv -\nabla V(\mathbf{r}_0)$.

Then the search directions are calculated iteratively according to Eq. 3:

$$\mathbf{f}_i \equiv -\nabla V(\mathbf{r}_i), \quad \mathbf{g}_{i+1} = \mathbf{f}_i, \quad \mathbf{h}_{i+1} = \mathbf{g}_{i+1} + \gamma \mathbf{h}_i \quad (3)$$

Where γ is a scalar factor, computed according to the Polak-Ribiere algorithm as it is more efficient than Fletcher-Reeves method under some circumstances³⁴:

$$\gamma = \max[(\mathbf{f}_i + \mathbf{g}_i) \cdot \frac{\mathbf{f}_i}{\mathbf{g}_i \cdot \mathbf{g}_i}, 0] \quad (4)$$

To ensure the convergence of Eq. 2, α is supposed to satisfy the Wolfe line search conditions³⁵:

$$V(\mathbf{r}_i + \alpha \mathbf{h}_i) - V(\mathbf{r}_i) \leq -\sigma_1 \alpha \mathbf{f}_i^T \mathbf{h}_i \quad (5)$$

$$\nabla V(\mathbf{r}_i + \alpha \mathbf{h}_i)^T \mathbf{h}_i \geq -\sigma_2 \mathbf{f}_i^T \mathbf{h}_i \quad (6)$$

Where $0 < \sigma_1 \leq \sigma_2 < 1$.

Wolfe condition (Eq. 5) can ensure the step size is not excessively large, while wolfe condition (Eq. 6) can ensure the step size is not too small. In this work, we provide two methods to obtain the α value. First, we set $\alpha = \alpha_{min}$, and then increase α from α_{min} to α_{max} using the following formula:

$$\alpha_{i+1} = 1.01\alpha_i \quad (7)$$

To ensure the step length α_i decreases $V(\mathbf{r}_i)$ 'sufficiently', the following Wolfe line search conditions is used to scale down α :

$$V(\mathbf{r}_i + \alpha \beta \mathbf{h}_i) - V(\mathbf{r}_i) \leq -\sigma_1 \alpha \beta \mathbf{f}_i^T \mathbf{h}_i \quad (8)$$

and the iterative form is changed with:

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \alpha \beta \mathbf{h}_i \quad (9)$$

Of note, we reset β to 1.0 after every iteration, and then repeatedly scale it to half until it satisfies the aforementioned Wolfe line search conditions (Eq. 8). We called this backtracking line search algorithm CG-BS. The second method was using the traditional Wolfe line search method,³⁶ thus it is named as CG-WS. In a short, CG-BS employs the first Wolfe condition (Eq. 8) to ensure the step size is not excessively large, while CG-WS employs both conditions (Eqs. 5 and 6). Of note, it is shown that the PRP (Polak-Ribière-Polyak) method is globally convergent when the line search employs a constant step size $\alpha \leq 1/4L$, where L is a Lipschitz constant for force.³⁶ Thus, a small step size can also be employed for CG method.

2.2 BFGS and L-BFGS methods

BFGS and L-BFGS are two quasi-Newton methods which utilizing Hessian matrix in the optimization procedure. The search direction of BFGS and L-BFGS was determined using following equations.

$$\mathbf{p}_i = \mathbf{B}_i^{-1} \mathbf{f}_i \quad (10)$$

Where $\mathbf{B}_i$, an approximation to the Hessian matrix was calculated using Eq. 11 for the BFGS method.

$$\mathbf{B}_{i+1} = (\mathbf{B}_0 - \frac{\mathbf{s}_i \mathbf{y}_i^T}{\mathbf{s}_i^T \mathbf{y}_i}) \mathbf{B}_i^{-1} (\mathbf{B}_0 - \frac{\mathbf{y}_i \mathbf{s}_i^T}{\mathbf{s}_i^T \mathbf{y}_i}) + \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{s}_i^T \mathbf{y}_i} \quad (11)$$

Where $\mathbf{s}_i = \mathbf{r}_{i+1} - \mathbf{r}_i$, $\mathbf{y}_i = \mathbf{f}_i - \mathbf{f}_{i+1}$. While $\mathbf{B}_i$ for L-BFGS was calculated by Eq. 12.

$$\mathbf{B}_{i+1}^{-1} = (\prod_{t=i}^0 \mathbf{U}_t^T) \mathbf{B}_0^{-1} (\prod_{t=0}^i \mathbf{U}_t) + \sum_{j=0}^i (\prod_{t=i}^{j+1} \mathbf{U}_t^T) (\rho_j \mathbf{s}_j \Delta \mathbf{s}_j^T) (\prod_{t=j+1}^i \mathbf{U}_t) \quad (12)$$

Where $\rho_j = \frac{1}{\Delta \mathbf{s}_j^T \Delta \mathbf{y}_j}$, $\mathbf{U}_j = \mathbf{B}_0 - \rho_j \mathbf{y}_j \Delta \mathbf{s}_j^T$. However, according to actual needs, we can take the nearest m item as shown in Eq. 13.

$$\mathbf{B}_{i+1}^{-1} = (\prod_{t=i-1}^{i-m} \mathbf{U}_t^T) \mathbf{B}_0^{-1} (\prod_{t=i-m}^{i-1} \mathbf{U}_t) + \sum_{j=0}^{i-m} (\prod_{t=i}^{j+1} \mathbf{U}_t^T) (\rho_j \mathbf{s}_j \Delta \mathbf{s}_j^T) (\prod_{t=j+1}^i \mathbf{U}_t) \quad (13)$$

Although the step size obtained from Wolfe line search could ensure the BFGS updating is stable, both BFGS and L-BFGS methods may fail for non-convex functions with line searches satisfying the Wolfe conditions.³⁷ The step size $\alpha=1$ is widely applied for both the BFGS and L-BFGS algorithms. Thus, in this study, we applied the constant step size $\alpha=1$ for both methods in geometry optimizations.

2.3 Dihedral angle constraint

To generate PESs, one needs to perform geometry optimization with one or more geometric parameters (such as torsional angles) being frozen. It is straightforward to add

constraints to internal coordinates, however, ANI-1x only takes Cartesian coordinates as an input. Some degrees of freedom can be eliminated through adding a dihedral angle bias potential (restrained optimization) or through revising the force during optimization process (constrained optimization). Adding a bias potential could not completely restrict the change of a dihedral angle. While revising force could eliminate any changes for the dihedral angle in question. For the purpose of generating PESs, we adopt the second approach, i.e., to fix one or multiple geometric parameters by revising the forces. Of note, this approach is in line with the partial optimization in many *ab initio* packages including Gaussian 16.³⁸ For an angle or a torsional angle, the force is revised through a series of rotations along with coordinate axes. The process of revising forces for a fixed torsional angle was illustrated in **Figure 1**. Specifically, for a torsional angle A-B-C-D consisting of atoms A, B, C, and D, we first translate $\mathbf{r}_{ABCD}$ so that B is moved to the coordinate origin, and rotate $\mathbf{r}_{ABCD} = [r_A, r_B, r_C, r_D]$ to make C locate on the x-axis along with force $\mathbf{f}_{ABCD} = [f, f_B, f_C, f_D]$. Note that we use $\mathbf{r}_{ABCD}$ and $\mathbf{r}'_{ABCD}$ to represent coordinates before and after a translational or rotational operation, respectively. This notation rule also applies to forces. In the following equations (Eqs. 14 to 19), T_B and R_{Cx} , etc. are spatial transformation matrices.

$$\mathbf{r}'_{ABCD} = T_B \mathbf{r}_{ABCD} \quad (14a)$$

$$\mathbf{r}'_{ABCD} = R_{Cx} \mathbf{r}_{ABCD} \quad (14b)$$

$$\mathbf{f}'_{ABCD} = R_{Cx} \mathbf{f}_{ABCD} \quad (14c)$$

We then rotate $\mathbf{r}'_{ABCD}$ to make A locate on xy-plane along with force $\mathbf{f}_{ABCD}$:

$$\mathbf{r}'_{ABCD} = R_{Axy} \mathbf{r}_{ABCD} \quad (15a)$$

$$\mathbf{f}'_{ABCD} = R_{Axy} \mathbf{f}_{ABCD} \quad (15b)$$

ABC is constrained in the xy-plane by setting the z-direction forces of A, B and C atoms to zero, and setting the y-direction forces of B and C atoms to zero, i.e., $f_{By} = f_{Bz} = 0$, $f_{Cy} = f_{Cz} = 0$, $f_{Az} = 0$. Then we rotate $\mathbf{r}'_{ABCD}$ so that A is back to its previous plane along with its force:

$$\mathbf{r}'_{ABCD} = R_{Axy}^{-1} \mathbf{r}_{ABCD} \quad (16a)$$

$$\mathbf{f}'_{ABCD} = R_{Axy}^{-1} \mathbf{f}_{ABCD} \quad (16b)$$

Similarly, we rotate $\mathbf{r}_{ABCD}$ so that D is in the xy-plane:

$$\mathbf{r}'_{ABCD} = R_{Dxy} \mathbf{r}_{ABCD} \quad (17a)$$

$$\mathbf{f}'_{ABCD} = R_{Dxy} \mathbf{f}_{ABCD} \quad (17b)$$

Then BCD is constrained in this plane by resetting the z-direction force of D to zero: $f_{Dz} = 0$. Next, we rotate $\mathbf{r}_{ABCD}$ so that the atoms back to their initial translated positions along with the updated forces.

$$\mathbf{r}'_{ABCD} = R_{Cx}^{-1} R_{Dxy}^{-1} \mathbf{r}_{ABCD} \quad (18a)$$

$$\mathbf{f}'_{ABCD} = R_{Cx}^{-1} R_{Dxy}^{-1} \mathbf{f}_{ABCD} \quad (18b)$$

Last, we translate $\mathbf{r}_{ABCD}$ back to its previous position:

$$\mathbf{r}'_{ABCD} = -T_B \mathbf{r}_{ABCD} \quad (19)$$

After all the above translational-rotational operations, $\mathbf{r}'_{ABCD}$ in Eq. 19 equals to $\mathbf{r}_{ABCD}$ in Eq. 14. We also developed algorithms for constraining bond angles and bond lengths. We applied a similar procedure to constrain multiple geometric parameters ($N_{tor} + N_{angle} + N_{bond} \geq 2$, where N_{tor} , N_{angle} , N_{bond} , are numbers of constrained torsional angles, angles and bonds, respectively). For example, to generate a Ramachandran plot, we need to constrain both the ϕ and ψ angles of an amino acid residue. It is pointed out that one may

project forces to a plane or line using alternative approaches. Our approach is convenient and works well with the two CG algorithms described above when a molecule's geometry is described by Cartesian coordinates.

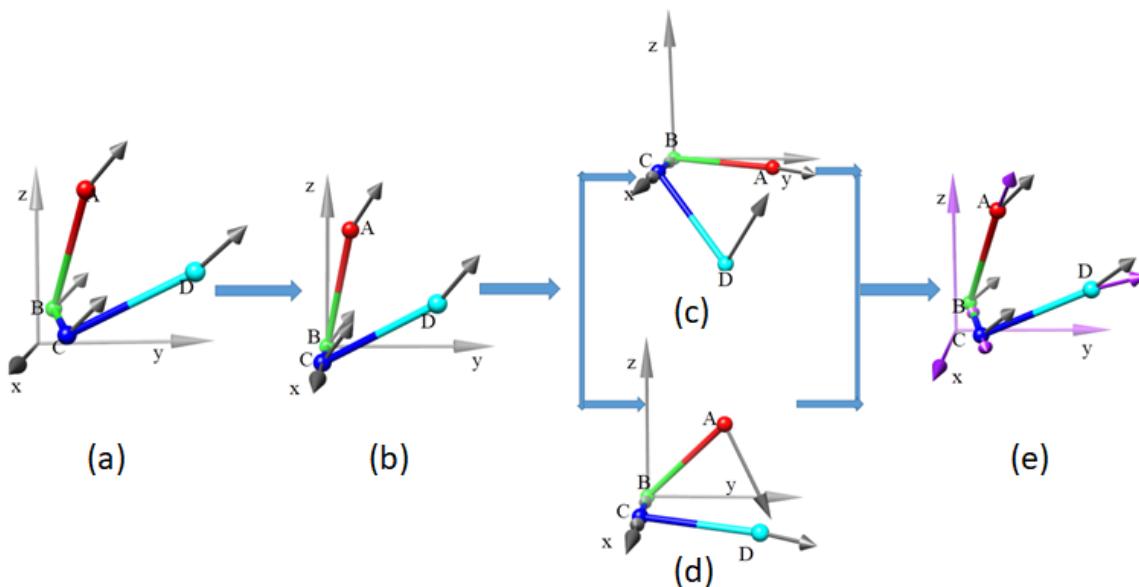


Figure 1. Schematic illustration of revising forces for a torsional angle which is frozen during the CG-WS optimization. (a). Initial coordinates and forces of Atoms A, B, C and D; (b). Translating $\mathbf{r}_{ABCD}$ so that B is at the coordinate origin and aligning $\mathbf{r}_{ABCD}$ so that Atom C is in x -axis through a rotational operation; (c). Rotating $\mathbf{r}_{ABCD}$ so that D is in the xy -plane, and resetting the z -direction force of D to zero; (d) Rotating $\mathbf{r}_{ABCD}$ so that A is in the xy -plane, and resetting the z -direction force of A to zero; (e). Rotating $\mathbf{r}_{ABCD}$ so that the atoms back to their initial positions along with the updated forces.

2.3 PES calculation with DFT

We prepared three datasets for the evaluation of four optimization algorithms in

geometry optimizations using the ANI-1x energies and forces. **Dataset 1** consists of 100 drug molecules (**Figure S1**) with varied numbers of atoms NATOM from 13 to 50. **Dataset 2** has 8 large molecules identified as potential inhibitors of SARS CoV-2 3CL protease.³⁹ The eight molecules have NATOM ranged from 65 to 120, as shown in **Figure S2**. The first two molecule sets were applied to assess the four algorithms in full geometry optimization. Each molecule in both sets has five distinct conformations generated using Omega2 module of the OpenEye software package (www.eyesopen.com). Full DFT optimization at ω B97X/6-31G(d) level were performed for every conformer using the Gaussian 16 software.³⁸ Note that in the rest of the manuscript, DFT refers to ω B97X/6-31G(d) unless explicitly stated otherwise.

The last dataset, **Dataset 3**, consists 160 molecules which were used to develop GAFF torsional angle parameters.¹⁵ Each molecule in this dataset has 2 to 23 heavy atoms (C, N and O) and only one rotatable torsional angle was applied in torsional PES generation. The 2D structures of those molecules and their rotatable bonds which are colored in red are shown in **Figure S3**. To generate a torsional PES for a molecule, the torsional angle was rotated from 0 to 360 degrees at a step of 20 degrees. In total 18 conformations were generated.

Next, the AMBER topology for a molecule was generated utilizing GAFF force field.¹⁵ For each conformation, we performed minimization using the Pmemd module of AMBER18⁴⁰ to generate a Z-matrix input with the torsional angle in question being frozen using the Antechamber module,⁴¹ and conducted constrained optimization at the B3LYP/6-

31G* level using the Jaguar software.⁴² Solvent effect was taken into consideration using polarizable continuum model (PCM) with the exterior dielectric constant being set to a value according to the calculated hydration free energy: $\epsilon_{\text{out}} = 2$ if $\Delta G_{\text{solv}} \geq -1.0$; $\epsilon_{\text{out}} = 4$ if $-9.0 \leq \Delta G_{\text{solv}} < -1.0$; $\epsilon_{\text{out}} = 10$ if $-18.0 \leq \Delta G_{\text{solv}} < -9.0$; and $\epsilon_{\text{out}} = 80$ if $\Delta G_{\text{solv}} < -18.0$ kcal/mol. Interior dielectric constant ϵ_{in} was set to 1. Note a hydration free energy ΔG_{solv} was calculated with the inner dielectric constant ϵ_{in} being set to 1 and exterior dielectric constant ϵ_{out} being set to 78.3. In above, we provided details on how we generated molecular geometries for Dataset 3 molecules. However, for the following DFT and ANI-1x calculations, only the molecular geometry information was applied.

The optimized geometries by Jaguar were then applied to calculate single-point energies at $\omega\text{B97X/6-31G(d)}$ level,⁴³ as well as constrained optimizations at the same level using the Gaussian 16 package.³⁸ It is pointed out that no solvent effect was considered in DFT calculations. A single-point energy of the ANI-1x²⁵ potential was calculated using the DFT optimized geometry, while a full or constrained optimization was performed using the same starting geometry as used in DFT optimization. Moreover, the same convergence criteria were utilized in geometry optimizations by both $\omega\text{B97X/6-31G(d)}$ and ANI-1x/CG, i.e., the maximum force and displacement are 0.00045 and 0.0018 in atomic units, respectively; and the RMS force and displacements are 0.0030 and 0.0012 in atomic units, respectively. An optimization job fails if it cannot converge within 10,000 steps, the predefined maximum number of iterations. This is the worst scenario in a geometry optimization in this study. However, a failed job can be restarted using the last geometry

of the last optimization.

To establish computational protocols of applying CG/ANI-1x in torsional force field parameterization, we also selected a subset of molecules which have at least 14 heavy atoms. The geometries of those molecules were optimized by GAFF force field prior to ANI-1x/CG-WS and DFT optimization.

To generate Ramachandran plots for alanine dipeptide, we rotated the ϕ and φ angles from 0 to 360 degrees at a step of 10 degrees. Constrained optimization was then performed for each conformation using the DFT and ANI-1x/CG-WS method. The PESs and Ramachandran plots were then produced for comparison using the Python Matplotlib module⁴⁴ and Matlab (<https://www.mathworks.com>) software.

3. RESULTS

3.1 Efficiency and accuracy of four optimization methods in full geometry optimization

We evaluated the performance of full geometry optimization using two datasets. The performance, measured by the Central Processing Unit (CPU) time, number of optimization steps, AUE (average unsigned error), RMSE (root-mean-square error) and correlation coefficient R, is listed in Table 1 for 100 drug molecules and Table S1 for 8 large molecules. Each molecule has five distinct conformations. The AUE, RMSE and R measure how well an ANI/CG model reproduces the relative conformational energies by DFT. Note that we used the mean value of all five conformations as the reference to

calculate the relative conformational energies. As shown in **Table 1**, CG-WS is most efficient method while L-BFGS is the least efficient. The computer time of CG-BS and BFGS are slightly longer than that of CG-WS. As the accuracy is concerned, the BFGS and L-BFGS methods have smaller AUE/RMSE than CG-WS and CG-BS. However, the accuracy of all the four methods is comparable and the differences of AUE and RMSE are less than 0.1 kcal/mol. As far as the absolute energy is concerned, most ANI-1x energies are more positive than the *ab initio* ones and the average $\Delta E = E_{ANI} - E_{QM}$ are 2.58, 2.57, 2.32 and 2.35 kcal/mol, for the CG-BS, CG-WS, L-BFGS and BFGS respectively. The number of conformations (the total is $100 \times 5 = 500$) that have the closest energies to ω B97X/6-31(d) are 5, 19, 368 and 108, for the four optimization methods correspondingly.

As for large molecules shown in **Figure S2**, the rank orders of computer time and optimization steps are the same as Dataset 1. Interestingly, the two CG methods achieved smaller AUE and RMSE than the two quasi-Newton methods.

We also compared the conformational energies calculated by another DFT method and GAFF. As shown in **Table 1**, the AUE and RMSE of B3LYP/6-31G*, 0.60 and 0.73 kcal/mol, are much smaller than those of ANI-1x methods. The AUE and RMSE of GAFF are about 0.3 kcal/mol larger, partially due to the fact that GAFF¹⁵ was developed to reproduce MP2/aug-cc-pVTZ instead of ω B97X/6-31G(d).

Table 1. Efficiency and accuracy of various optimization method combined with constraint method

	B3LYP/	ANI/	ANI/	ANI/	ANI/	GAFF
--	--------	------	------	------	------	------

	6-31G*	CG-BS	CG-WS	L-BFGS	BFGS	
CPU time (second)	11183.0	98.6	60.0	173.4	76.1	0.14
Optimization Steps	25.5	410.1	152.7	1007.3	443.0	1146.7
AUE (kcal/mol)	0.60	1.11	1.13	1.05	1.07	1.34
RMSE (kcal/mol)	0.73	1.31	1.34	1.25	1.26	1.56
R	0.88	0.70	0.70	0.75	0.74	0.70
Ave. RMSD (Å)	0.16	0.36	0.35	0.31	0.32	0.39

3.2 Efficiency and accuracy of two CG optimization methods in producing one-dimensional torsional PES

We evaluated the two CG optimization methods in generating one-dimensional PES for 160 molecules in **Dataset 3** and two-dimensional PES for alanine dipeptide. Here, we only reported the results obtained by CG-WS, as the results of CG-BS are essentially similar to those of CG-WS.

Each molecule in **Dataset 3** belongs to one of five groups according to the number of heavy atoms it has (**Figure S3**). The aim is to study how the predictions depend on the sizes of molecules. For a given molecule, *ab initio* optimizations of some conformations may fail due to the occurrence of steric clashes between atoms when the rotatable torsional angle takes some specific values. In total, there are 2868 conformations left for 160 molecules (on average 17.92 conformations per molecule).

To generate a PES of a molecule, we applied the mean value of all the conformations as the reference. By this way, two PESs of the same molecule can be compared. To evaluate the reliability of ANI-1x energy, Smith et al. developed a ρ parameter, which is the standard deviation of ANI-1x potentials predicted by an ensemble of models, weighed by square

root of the number of atoms in a molecule.²⁵ They found that with ρ smaller than 0.23 kcal/mol, about 98% of molecules in their molecular set have the predicted errors smaller than 1.5 kcal/mol.

Single Point Energy Calculations. Among the 2868 conformations, 14 have the relative conformational energies larger than 15.0 kcal/mol. The $\Delta\Delta G$ between the ANI-1x and ω B97x/6-31G(d) for those conformations are very big, with an RMSE of ~ 19.5 kcal/mol. As expected, the mean ρ values are also very big, 1.6 kcal/mol. Therefore, the following performance evaluation was conducted without those outliers. Of note, those outliers come from three molecules (#25 in Figure S1A, #21 and #22 in Figure S1E), for which the inner bond of the torsional angle is a double bond. The PESs generated using the single-point ω B97X/6-31G(d) and ANI-1x energies are demonstrated in **Figure S4** of the Supporting Information. It is demonstrated most ANI-1x PESs are fitted very well to those predicted by DFT. **Figure 2A** shows the distribution of the mean absolute energy deviations (MAD) between the ANI-1x and DFT single-point energies for the 160 molecules. The average MAD for ANI-1x single-point energy vs. reference DFT single-point energy is 0.63 kcal/mol, while the significantly larger RMSE, 1.03 kcal/mol, is caused by about 9.7% conformations whose $\Delta\Delta G$ are larger than 1.5 kcal/mol. The scatter plot of the two types of single-point energies is shown in **Figure 2C**. Exploration of the PESs shown in **Figure S4** can identify those outliers shown in **Figure 2A** and **2C**.

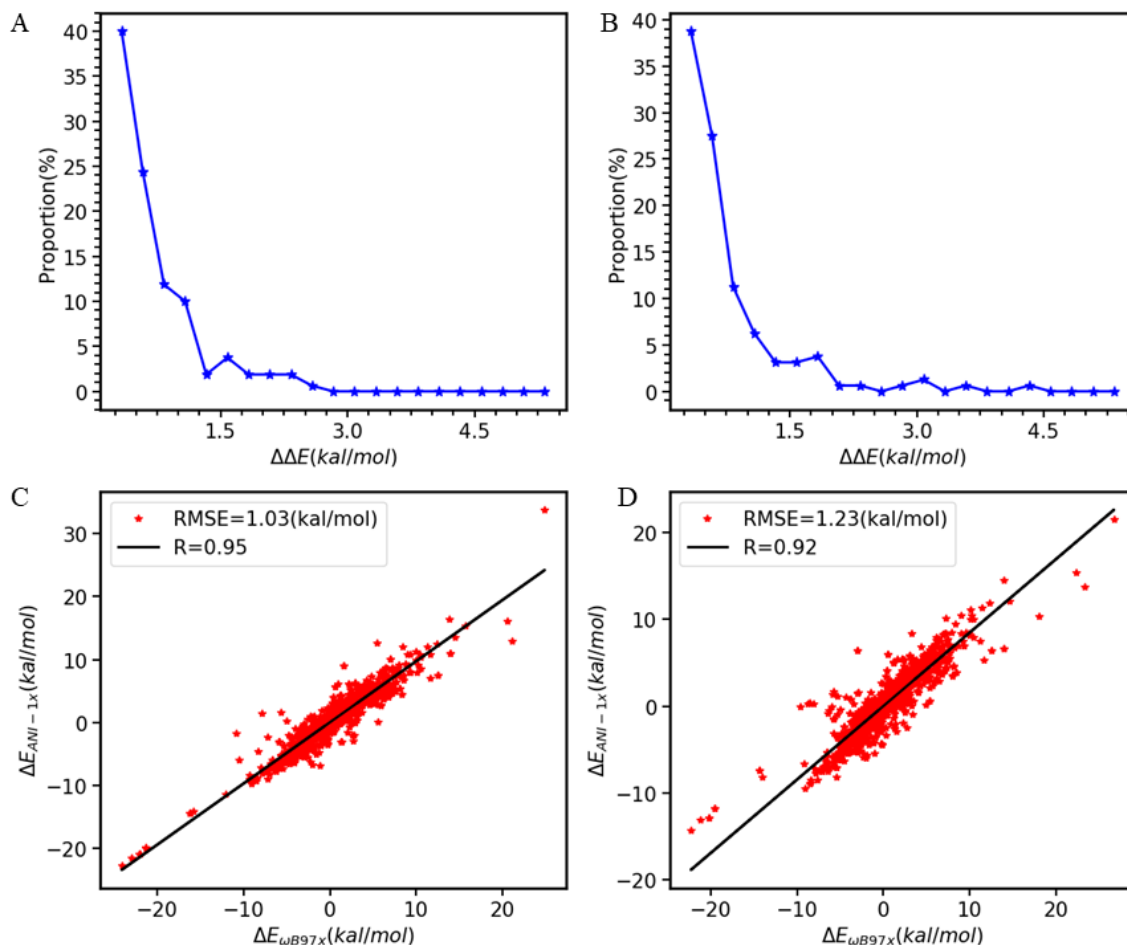


Figure 2. Performance of torsional PESs generated using ANI-1x and CG-WS for 160 molecules. A: mean absolute energy deviations (MAD) between single-point ANI-1x potential energies and single-point $\omega B97X/6-31G(d)$ energies on the Jaguar optimized geometries. B: MAD of the conformational energies by ANI-1x/CG-WS and $\omega B97X/6-31G(d)$ after geometric optimization. C: Scatter plot between single-point ANI-1x potential energies and single-point $\omega B97X/6-31G(d)$ energies. D: Scatter plot between ANI-1x potential energies and $\omega B97X/6-31G(d)$ energies after geometric optimization.

As pointed out above, the accuracy of ANI-1x potential energy can be evaluated by ρ parameter. In **Table 2**, we demonstrated how the performance of ANI-1x varies as a

function of ρ . Interestingly, we also observed that using ρ cutoff of 0.23 kcal/mol, 98% of conformations have $\Delta\Delta G$ between $\omega B97X/6-31G(d)$ and ANI-1x smaller than 1.5 kcal/mol. Unfortunately, with such a stringent cutoff, the data coverage, the percent of data with ρ smaller than the cutoff, is only 57%. As shown in **Table 2**, the ρ cutoff of 0.6 achieves a good balance between data coverage (93%) and ANI-1x accuracy (MAD and RMSE of 0.55 and 0.89 kcal/mol, respectively, and 92% of conformations satisfy the condition of $\Delta\Delta E < 1.5$ kcal/mol).

Table 2. Summary of ANI-1x performance using ρ as the selection parameter, all energies are in kcal/mol. MAD and RMSE are mean absolute energy deviation and root-mean-square energy deviation, respectively.

ρ cutoff	MAD	RMSE	No. Outliers ($\Delta\Delta E \geq 1.5$)	No. Data $\leq \rho$	Percent of data with $\Delta\Delta E < 1.5$ (%)	Percent of data with $\rho < \text{cutoff}$ (%)
<i>Single point calculations with ANI-1x and $\omega B97x/6-31G(d)$</i>						
0.23	0.35	0.59	36	1622	98	57.1
0.25	0.37	0.6	48	1775	97	62.5
0.3	0.42	0.66	82	2092	96	73.7
0.35	0.46	0.73	113	2271	95	80.0
0.4	0.48	0.77	139	2396	94	84.4
0.45	0.51	0.81	172	2498	93	88.0
0.5	0.53	0.84	185	2564	93	90.3
0.6	0.55	0.89	208	2641	92	93.0
0.7	0.56	0.91	222	2678	92	94.3
0.8	0.56	0.91	228	2704	92	95.2
0.9	0.57	0.92	239	2738	91	96.4
1	0.58	0.95	248	2758	91	97.1
unlimited	0.61	1.03	276	2840	90	100.0
<i>After optimizations with ANI-1x/CG-WS and $\omega B97x/6-31G(d)$</i>						
0.23	0.37	0.76	36	1596	98	56.2
0.25	0.38	0.76	43	1728	98	60.8
0.3	0.44	0.84	83	2064	96	72.7
0.35	0.49	0.95	128	2264	94	79.7
0.4	0.53	1.01	162	2415	93	85.0
0.45	0.55	1.03	182	2501	93	88.1
0.5	0.57	1.07	201	2577	92	90.7
0.6	0.59	1.09	230	2670	91	94.0
0.7	0.61	1.14	248	2701	91	95.1

0.8	0.63	1.17	266	2742	90	96.5
0.9	0.63	1.18	268	2755	90	97.0
1	0.64	1.2	273	2775	90	97.7
unlimited	0.65	1.23	288	2840	90	100.0

Conformational Energies after Constrained ANI-1x/CG-WS optimization.

Torsional angle PESs calculated using the optimized geometries play an important role in developing MMFFs. The ANI-1x/CG-WS optimized PESs and ω B97X/6-31G(d) optimized PESs for the 160 molecules are demonstrated in **Figure S5** of the Supporting Information. Overall, the ANI-1x/CG-WS PESs reproduce the DFT PESs very well as demonstrated in **Figure 2**. The distribution of MAD between the ANI-1x and DFT energies after geometric optimization for the 160 molecules is shown in **Figure 2B**, while the scatter plot of the two types of conformational energies after optimization is shown in **Figure 2D**. The average MAD for ANI-1x versus DFT energies is 0.65 kcal/mol, slightly larger than that for the single-point energies which is 0.61 kcal/mol. As expected, the RMSE value for ANI-1x versus DFT energies after optimization, 1.23 kcal/mol, is also slightly larger than the corresponding single-point calculation counterpart. Again, exploration of the PESs shown in **Figure S5** can identify those outliers shown in **Figure 2B** and **2D**. The above result is reasonable as ANI-1x/CG-WS optimized geometries may be slightly different from those optimized by DFT as shown in **Figure 3A**. We also plotted the MADs before and after geometric optimization in **Figure 3B**. Apparently, for most conformations the difference of MAD values is small, however, for about 10% of conformations, the MADs between ANI-1x and ω B97/6-31G(d) are larger than 1.5 kcal/mol. As shown in **Table 2**, for using a ρ cutoff of 0.6 kcal/mol, the data coverage (94%) is slightly larger than that of single point calculations, while the percent of conformations satisfying $\Delta\Delta E < 1.5$ kcal/mol, 91%, is slightly smaller.

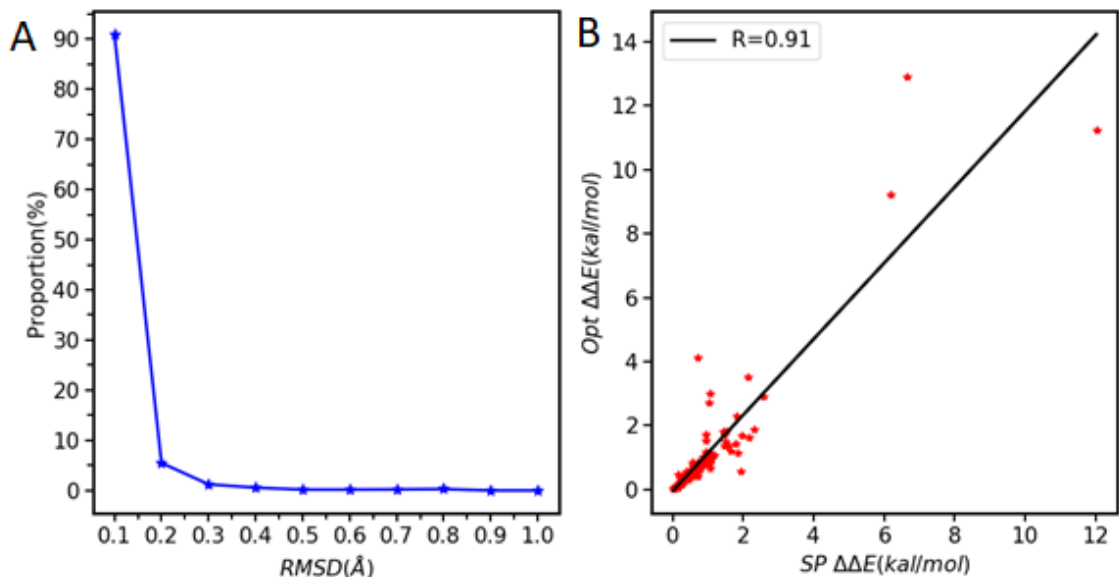


Figure 3. Impact of different optimization methods on structures and mean absolute energy deviations (MAD) of the two potentials ($MAD = [\Delta E_{ANI-1x} - \Delta E_{\omega B97X/6-31G(d)}]$). A: Distribution of RMSD between ANI-1 optimized structures and $\omega B97/6-31G(d)$ optimized structures for 160 molecules. B: Scatter plot MADs before (x-axis) and after geometric optimization (y-axis).

Structure similarity between ANI-1 optimization and DFT optimization. To critically assess the performance of CG-WS algorithm in conjunction with ANI-1x potential in geometry optimization, we performed geometric optimizations for the 160 molecules consisting of C, H, N and O and each has up to 18 conformations. $\omega B97X/6-31G(d)$ optimization was also performed for the same set of conformations of 160 molecules. The distribution of RMSDs is illustrated in **Figure 3A**. The distribution has a mean value of 0.05 Å and 90.9% of RMSDs are lower than 0.1 Å. The detailed RMSDs for each molecule was presented in **Figure S5** of the Supporting Information. In brief, our CG-WS algorithm in conjunction with ANI-1x can produce very similar geometries as those optimized by DFT.

3.3 Efficiency and accuracy of two CG optimization methods in producing two-dimensional torsional PES

We generated 1296 alanine dipeptide (alanine with the N-terminal linking an acetyl group and the C-terminal being a N-methyl amide group) conformations with the peptide ϕ and φ angles being rotated from 0 to 360 degrees at a step of 10 degrees. Without further optimization using a MMFF, we then performed constrained geometric optimization using calculated ω B97x/6-31G(d) and ANI-1x/CG-WS. The structure as well as ϕ (blue and red bonds) and φ (red and green) angles are shown in **Figure 4A**. About 40% of conformations were not successfully optimized by ω B97x/6-31G(d) using the default setting of Gaussian 16.³⁸ With several rounds of restarting, still 9 conformations cannot meet the default convergence condition. On the other hand, all jobs finished with the first try by ANI-1x/CG-BS; and only 9 failed by ANI-1x/CG-WS. For the 9 failed jobs, all optimization jobs succeeded after restarting. Therefore, ANI-1x/CG-BS is more robust than ANI-1x/CG-WS. As for the computational time, the average number of optimization steps and computation time for ANI-1x/CG-WS are 116.8 and 36.9 seconds, while the corresponding values are 24.4 and 6574.9 seconds for DFT. On average, ANI-1x/CG-WS optimization only uses 0.6% of the CPU time required for ω B97x/6-31G(d) optimization.

The average value of RMSDs of the heavy atoms are 0.10 ± 0.06 Å between the ANI-1x/CG-WS and DFT optimized geometries. As expected, DFT has lower absolute energies than ANI-1x/CG-WS for most conformations and the mean difference ($E_{ANI} - E_{DFT}$) is 1.36 kcal/mol. However, the MAE and RMSE values of the relative energies are much smaller, which are 0.32 and 0.49 kcal/mol, respectively. The correlation between the two sets of relative energy data is 0.99. The performance of ANI-1x/CG-BS is same as that of ANI-1x/CG-WS except that $E_{ANI} - E_{DFT} = 1.39$ kcal/mol is slightly larger. However, CG-BS is more robust than CG-WS, as the former has no failed jobs while the latter has 10 out of 1296 optimization jobs failed. It is worth mentioning that after one-round

restarting, all the 10 jobs achieved the convergence. We generated the Ramachandran plots of the two-torsional PESs calculated using DFT and ANI-1x/CG-WS (**Figure 4C and 4D**). Encouragingly, the overall shapes of the two Ramachandran plots resemble to each other very well. The RMSD values are also shown in a contour plot with 20 filled intervals using the 2D-contour function in Matlab (**Figure 4B**). Note that the RMSDs between the DFT and ANI-1x/CG-WS geometries were calculated using least-square fitting for the heavy atoms only.

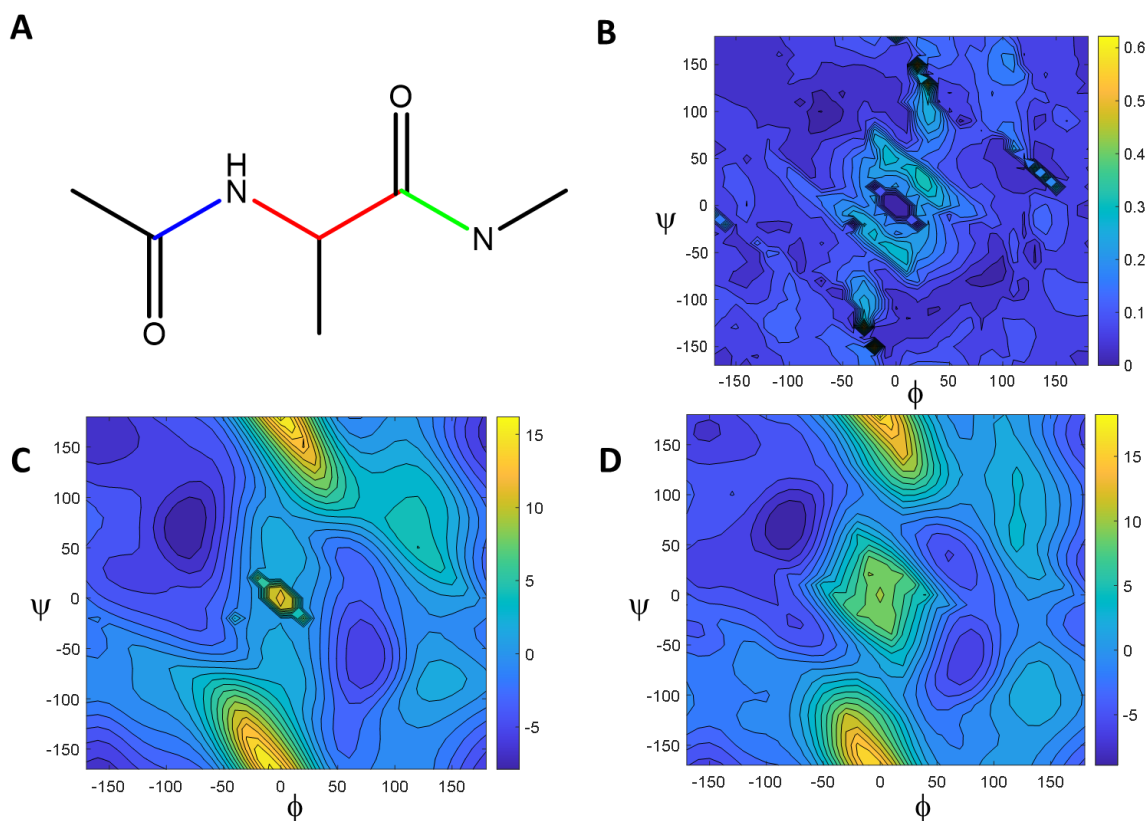


Figure 4. Ramachandran plots for the alanine dipeptide. A: Definition of ϕ (blue and red bonds) and ψ (red and green bonds) torsional angles. B: a contour plot of RMSD values between the ANI-1x/CG-WS and DFT optimized geometries. C: A Ramachandran plot using the DFT potential energies. D: A Ramachandran plot using the ANI-1x/CG-WS potential energies.

4. DISCUSSION

4.1 Optimization algorithms

CG, BFGS and L-BFGS all can use the Wolfe line search. However, we found that both BFGS and L-BFGS failed under some circumstances when applying Wolfe line searches, which is consistent with previous study. Thus, we utilized a constant step size of $\alpha=1$ for both the quasi-Newton algorithms.

For two CG algorithms, we modified the forces of relevant atoms through a series of rotational operations. The successful implementation of the algorithm is demonstrated by very small RMSDs in generating one-dimensional PESs (~ 0.05 Å) and two-dimensional PESs (~ 0.1 Å). However, the algorithm of modifying forces for the restrained atoms is not applicable to the LBFGS and BFGS methods, as both algorithms modify the search directions determined by Eq. 3. Thus, both the search direction and force to constrain the dihedral angle should be modified for the two quasi-Newton methods. We will improve the two quasi-Newton methods so that they can be applied to perform constrained geometry optimization in future.

4.2 Can pre-geometry optimization using ANI-x speed up DFT optimization?

We also investigated if using an optimized geometry with ANI-1x can shorten the DFT optimization time. For 100 molecules in Dataset 1, the DFT optimization time per molecule is shorten about 23%, while the optimization steps reduced 18% using geometries pre-optimized by ANI-1x/CG-WS. In contrast, after the ANI/L-BFGS optimization, the DFT

optimization time and steps reduced 16.0% and 18.0%, respectively. Therefore, the two optimization methods can reduce *ab initio* compute time almost equally efficient. However, for large molecules in **Dataset 2**, the optimization time reduced only about 9% and 12% after pre-optimization using ANI/CG-WS and ANI/L-BFGS, respectively. It is obvious that beneficial effect wanes for large molecules. In conclusion, one can save considerable CPU time by conducting ANI-1x/CG-WS optimization prior to DTF optimization, which is essential in large-scale *ab initio* calculations.

4.3 Dependence of ANI-1x performance on molecular size

To evaluate the dependence of ANI-1x performance on molecular size, we divide the 160 molecules in Dataset 3 into 5 groups and each group has 32 molecules (**Figure S3**). Group 1 has smallest numbers of heavy atoms, while Group 5 has largest numbers of heavy atoms. We investigated the dependence of molecular size for the following properties, RMSD of the heavy atoms between the ANI-1x/CG-WS and DFT optimized geometries, MAD and RMSE of the conformational energies, and the correlation coefficient between the two set of energies.

As shown in **Figure 5A**, the smaller the molecule size is, the smaller RMSD of the ANI-1x/CG-WS optimized structure relative to the DFT optimized structure is. However, even for Group 5, the median RMSD, 0.055 Å, is still acceptable. Not as obvious as RMSD, MAD becomes larger from Group 1 to Group 5 when molecular size increases (**Figure 5B**). We noticed that the Group 3 and Group 4 have the similar MAD values for both single-

point and optimized energies. Unlike RMSD and MAD, the molecular size-dependence is not obvious for RMSE as illustrated by **Figure 5C**. The RMSE of Group 1 was higher than Groups 2, 3 and 4, mainly due to ANI-1x performs very badly for two outliers. As to correlation between the ANI-1x/CG-WS and DFT energies, the difference is quite small, all five molecule groups achieved good correlations. .

Apparently, geometric optimization has a large impact on MAD, RMSE and R. The decrease of R value for Group 5 after optimization is understandable (Figure 5D), as the optimized geometries by ANI-1x/CG-WS and DFT may be different (**Figure S5**). If the RMSD values are small, as for the first four groups, the MAD and RMSE are comparable between single-point and optimized energies.

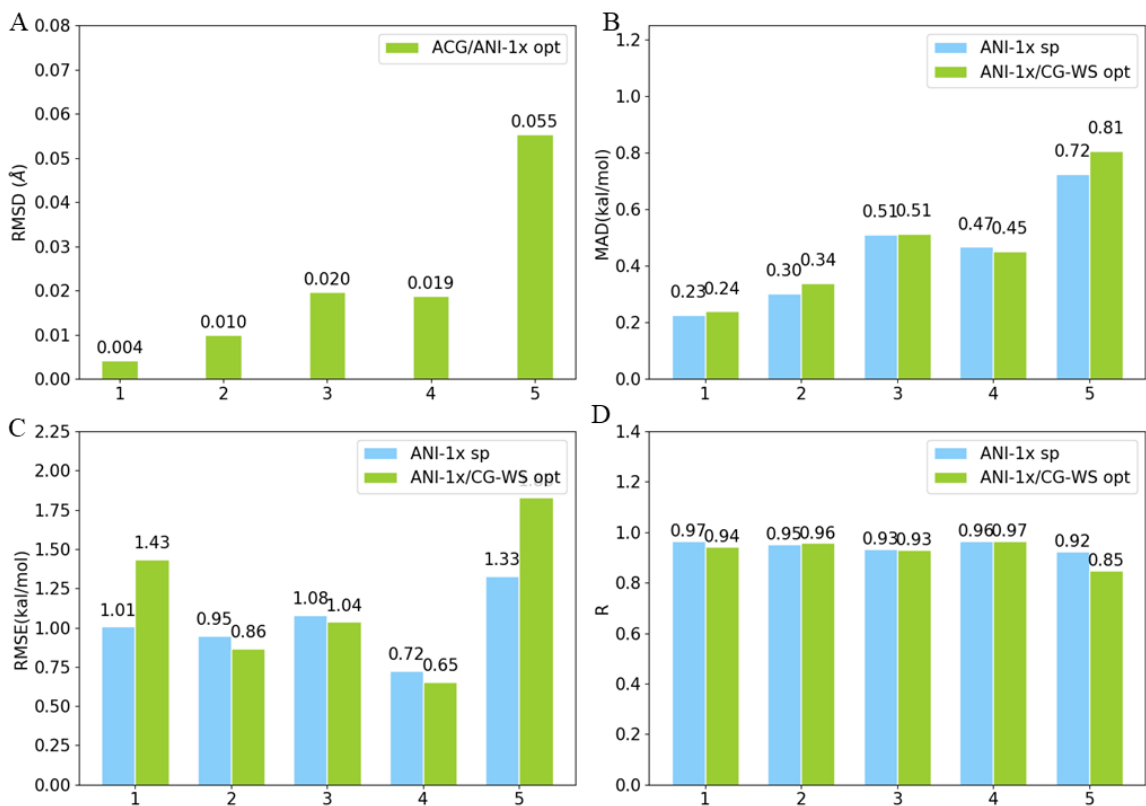


Figure 5. The dependence of ANI-1x/CG-WS performance on molecular size. A: Median

RMSD between the structures optimized by ANI-1x/CG-WS and ω B97x/6-31G(d). B and C: Median MAD (Panel B) and RMSE (Panel C) for single-point and optimized energies by ANI-1x/CG-WS and DFT. D: Correlation coefficient between single-point and optimized energies by ANI-1x/CG-WS and DFT.

4.4 The quality of ANI-1x directly determines the performance of ANI-1x/CG-WS

As mentioned above, large structural RMSD after geometric optimization contributes to large RMSE between ANI-1x/CG-WS and DFT energies. As our initial structures were already optimized even though using a different *ab initio* model (B3LYP/6-31G* with solvent effect being taken into consideration), we expect that the RMSD of the geometries, MAD and RMSE values between the ANI-1x/CG-WS and DFT optimized energies should be very small. Indeed, we observed the two types PESs agree with each other very well for most molecules in our dataset. We selected a subset of 9 molecules, whose molecular sizes and RMSD values are relatively large, for further investigation. The 2D-structures of the 9 molecules are shown in **Figure 6**.

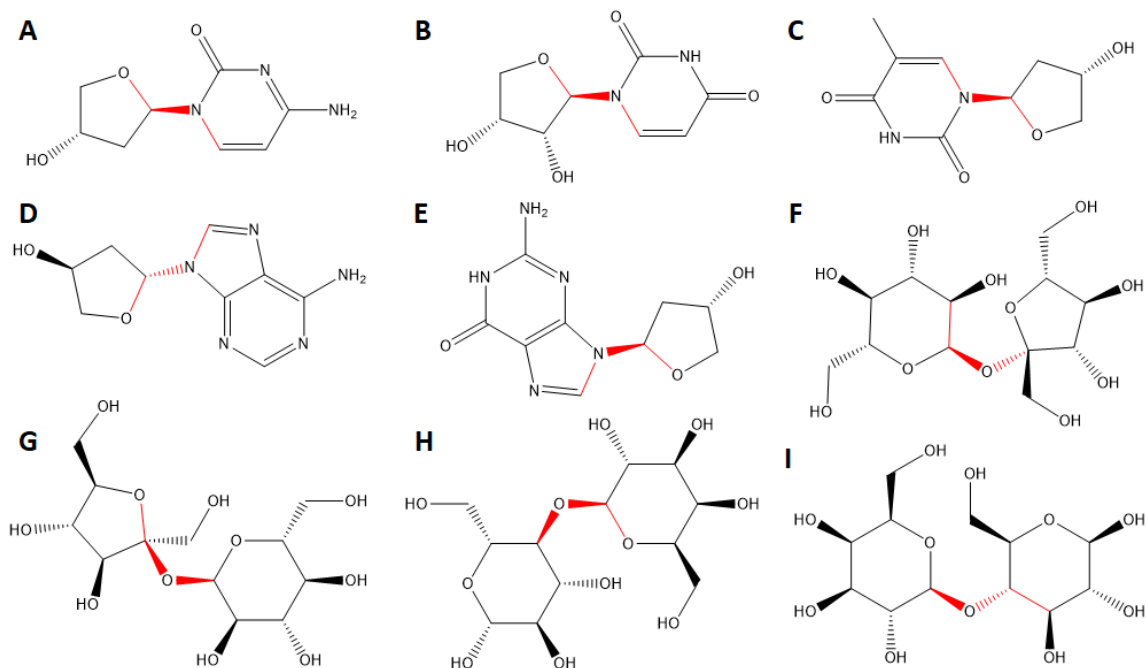


Figure 6. 2D-structures of the nine-molecule subset with the torsional angles for generating PESs colored in red. A: dC B: rU C: dT D: dA E: dG F: Sucrose1 G: Sucrose2, H: Lactose1, I: Lactose2.

We calculated four types of PESs, which were generated using (1) ANI-1x energies based on the ω B97x/6-31G(d) optimized geometries using Jaguar (denoted by ANI-1x//DFT), (2) ω B97x/6-31G(d) energies based on the ω B97x/6-31G(d) optimized geometries (denoted by ω B97x/6-31G(d)//DFT), (3) ANI-1x energies after ANI-1x/CG-WS optimization (denoted by ANI-1x//CG-WS), and (4) ω B97x/6-31G(d) energies calculated using the CG-WS optimized geometries (denoted by ω B97x/6-31G(d)//CG-WS). **Figure 7** illustrate the four types of PESs for the nine molecules. A general observation is that if a single-point ANI-1x//DFT PES fitted well to the corresponding single-point ω B97x/6-31G(d)//DFT, the two PESs based on the ANI-1x/CG-WS optimized

geometries, ANI-1x//CG-WS and ω B97x/6-31G(d)//CG-WS also fitted well to each other, as the cases for first five molecules (**Figure 7A-7E**). More importantly, ANI-1x/CG-WS optimization can improve the overlay of the PES curves (blue and brown curves). However, when the ANI-1x//DFT PES of a molecule cannot reproduce its ω B97x/6-31G(d)//DFT PES well, as the cases for the last four molecules (**Figure 7F-7I**), ANI-1x/CG-WS optimization cannot improve the overlay of the PES curves. This result suggest that a high-quality ANI potential is the key to achieve high accuracy on prediction the ANI energies after CG-WS geometric optimization.

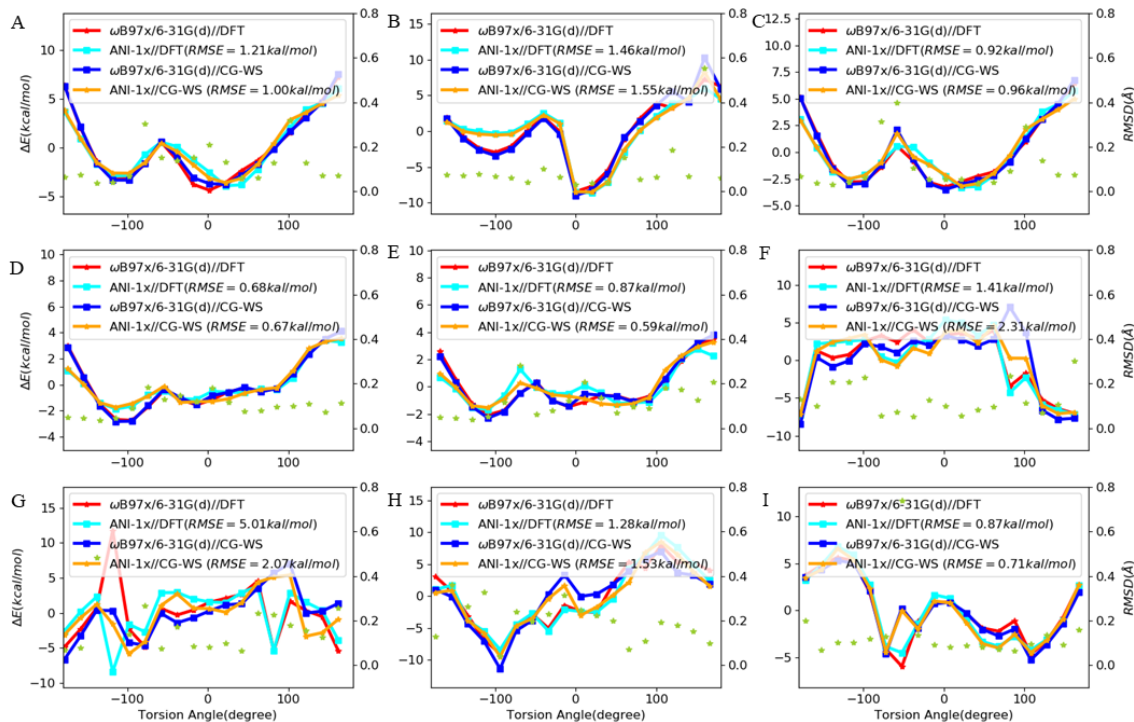


Figure 7. Potential energy surfaces of nine large molecules calculated using four different protocols. The starting geometries were optimized by B3LYP/6-31G* using the Jaguar software, and the RMSDs between the initial and ANI-1x/CG-WS are shown as green dots.

4.5. Application of ANI-1x/CG-WS in torsional parameter development

So far, our evaluation on ANI-1x/CG-WS is based on molecular geometries optimized by using B3LYP/6-31G* with the solvent effect taken into consideration. However, in practice, we do not have structures optimized by an *ab initio* model. To further test the applicability of the CG-WS algorithm in MMFF development, we followed the standard procedure to prepare structures prior to ANI-1x/CG-WS. First, we draw molecules shown in **Figure 6** using the Maestro molecular graphics software (www.schrodinger.com), and then generated residue topologies using the Antechamber module,⁴¹ next we conducted constrained minimizations for molecules described by GAFF.¹⁵ The GAFF-optimized geometries were then subjected to ANI-1x/CG-WS optimization. We found that the median RMSD of the nine molecules (in total 162 conformations) starting from GAFF-optimized structures was 0.091 Å, which is slight lower than that using the B3LYP/6-31G* optimized structures, which is 0.095 Å.

We then generated four types of PESs, which are ANI-1x//DFT, ω B97x/6-31G(d)//DFT, ANI-1x//CG-WS, and ω B97x/6-31G(d)//CG-WS. Note that DFT here refers to ω B97x/6-31G(d). The initial structures are GAFF-optimized geometries. The RMSD values between DFT and ANI-1x/CG-WS optimized geometries were also calculated and shown in **Figure 8** (dotted lines). Comparisons on the four types of PESs suggest that the four types of PESs overlaid very well. It is encouraging to observe that the accuracy of fitting the ANI-1x/CG-WS PESs reproducing the DFT ones is significantly improved using GAFF-optimized geometries, due to the much smaller RMSD values.

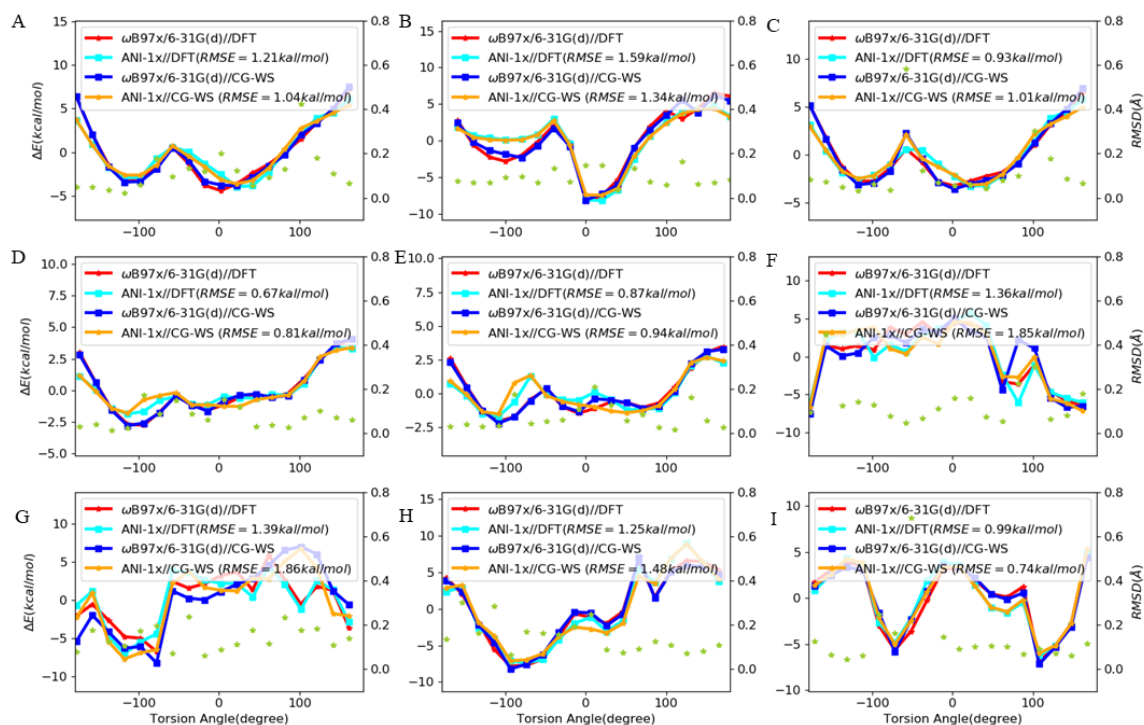


Figure 8. Potential energy surfaces of nine large molecules calculated using four different protocols. The starting geometries were generated by performing constrained optimizations using GAFF.

In MMFF development for arbitrary organic molecules, an important task is to develop high-quality torsional angle parameters to depict the torsional PESs. It will become a burden to generate PESs with high-quality *ab initio* models for a large set of organic molecules. On another hand, ANI-1x/CG-WS could significantly speed up the geometric optimization and potential energy calculation, and its accuracy is still acceptable compared to those obtained by *ab initio* calculations. For a typical molecule with a size of regular small-molecule drugs, such as the Group 5 molecules, the RMSE is about 1.5 kcal/mol. Interesting, for the 9 molecules in **Figure 6**, only one conformation has its ρ value, 0.61, slightly larger than the suggested cutoff, 0.6. No matter what, it is a good practice to

eliminate low quality data using ρ parameter in generating PES with ANI-1x/CG-WS. Thus, we may not apply ANI-1x/CG-WS potential energies directly to derive force field parameters, however, the accuracy is sufficient to detect bad torsional angle parameters. Once those low-quality force field parameters being discovered, we can then generate torsional PES with a high-quality *ab initio* model, and subsequently conduct force field parameterization. Instead of developing ANI potentials using training data produced by high-quality and computer time-demanding *ab initio* models, in our opinion, it is more important to expand the chemical space coverage and to improve the single-point accuracy of an ANI model, since according to **Figure 3B**, there is a strong correlation between the single-point MADs and those after geometric optimization. With the continually improved ANI potentials, we believe that our CG-WS and CG-BS algorithms can be a useful tool in general purpose force field development.

Conclusions

In this work, we developed and assessed four geometry optimization algorithms in conjunction with ANI-1x potential. The four geometry optimization algorithms consist of two conjugated gradient (CG) and two quasi-Newton algorithms, namely CG-BS (CG with backtracking line search), CG-WS (CG with Wolfe line search), BFGS (Broyden–Fletcher–Goldfarb–Shanno) and L-BFGS (low memory BFGS). CG-WS was a new algorithm developed in this work. All four algorithms can be applied to conduct full geometry optimization. Among the four algorithms, CG-WS is the most efficient, CG-BS is the most robust, and L-BFGS and BFGS achieve the smallest $\Delta E = E_{ANI} - E_{QM}$ values.

Both CG-WS and BFGS methods are recommended for this type of tasks. We developed an algorithm to project forces to a plane or line so that the two CG methods can perform constrained geometry optimization for a molecule described by Cartesian coordinates. The performance of CG-BS and CG-WS is quite similar except that the former is relatively more robust, and the latter is slightly more efficient and accurate. Therefore, both CG-WS and CG-BS are recommended for performing constrained geometry optimization.

We have explored the potential of applying ANI in molecular mechanics force field (MMFF) development. Although the accuracy of CG-WS or BFGS in conjunction with ANI-1x potential energies is less accurate in comparison with those obtained through *ab initio* optimization, they are much more efficient and the computer time is less than 1% of the DFT for molecules in our datasets. Thus, we concluded that CG-WS, CG-BS or BFGS in conjunction with ANI-1x potential although is not accurate enough to produce reference data for force field parameterization, they are good enough to detect problematic molecules or torsional parameters in MMFF development. Moreover, we suggested using the ρ cutoff of 0.6 kcal/mol, to eliminate those datapoints which may have poor accuracy of ANI-1x potential. With the continuously-increased quality of ANI, it is expected that CG-WS, CG-BS and BFGS algorithms, which in principle can be combined with any ANI potentials, will have a great application in improving the quality of an existing small molecule MMFF.

Supporting Information

Table S1 demonstrates the efficiency and accuracy of four optimization methods for 8 large molecules. Figures S1 and Figure S2 shows 100 and 8 molecules in Dataset 1 and Dataset 2, respectively. Figure S3 shows the 2D-structures of the 160 molecules in Dataset 3 and the torsional angles for which torsional scanning were performed (colored in red);

Figure S4 shows the torsional PESs using the ω B97x/6-31G(d) and ANI-1x based on the B3LYP/6-31G* optimized geometries; and Figure S5 shows the torsional PESs using the ω B97x/6-31G(d) and ANI-1x after geometric optimization using the same *ab initio* model and ANI-1x/CG-WS, respectively.

Competing Interests

There is no competing interest to declare.

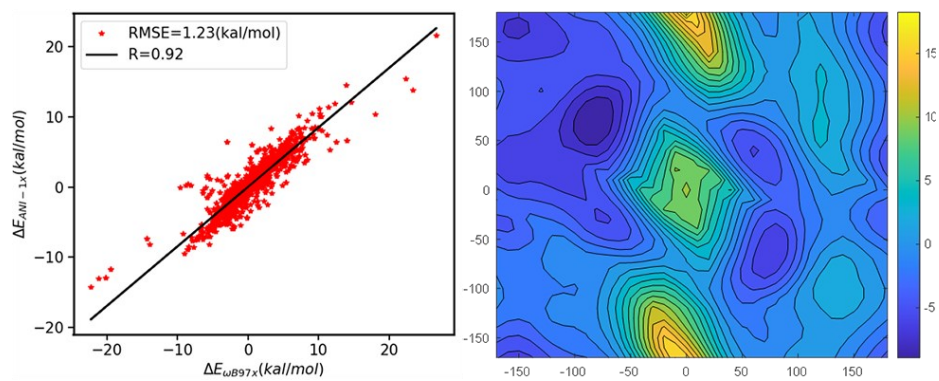
Acknowledgements

The authors gratefully acknowledge the funding support from the National Science Foundation (NSF) to J.W. (No. 1955260) and A.E.R. (No. 11774279), National Institutes of Health (NIH) to J.W. (R01GM079383), National Natural Science Foundation of China to S.Z (No. 11774279). The authors also thank the Center for Research Computing (CRC) at University of Pittsburgh for the computing resources, and OpenEye Scientific Inc. for providing us software tools.

Author contributions:

J. Wang designed the project and conducted *ab initio* calculations. D. Hao implemented the algorithm and performed the ANI-1x optimization. All authors discussed the project and wrote the manuscript.

TOC



References

1. Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A., Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *Chem. Rev.* **2010**, *110*, 5714-5789.
2. Sun, H., Ab initio calculations and force field development for computer simulation of polysilanes. *Macromolecules* **1995**, *28*, 701-712.
3. Wang, J.; Tingjun, H., Application of Molecular Dynamics Simulations in Molecular Property Prediction I: Density and Heat of Vaporization. *J. Chem Theory. Comput.* **2011**, *7*, 2151-2165.
4. Anderson, A. C., The process of structure-based drug design. *Chem. Biol.* **2003**, *10*, 787-797.
5. Ferguson, D. M.; Gould, I. R.; Glauser, W. A.; Schroeder, S.; Kollman, P. A., Comparison of ab initio, semiempirical, and molecular mechanics calculations for the conformational analysis of ring systems. *J. Comput. Chem.* **1992**, *13*, 525-532.
6. Elstner, M., The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.* **2006**, *116*, 316-325.
7. Stewart, J. J. P., Application of the PM6 method to modeling proteins. *J. Mol. Model.* **2009**, *15*, 765-805.
8. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
9. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-42.
10. Sagui, C.; Pedersen, L. G.; Darden, T. A., Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J. Chem. Phys.* **2004**, *120*, 73-87.
11. van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396-9409.
12. Clark, M. A.; Cramer, R. D.; van Opdenbosch, N., Validation of the general purpose tripos 5.2 force field. *J. Comp. Chem.* **1989**, *10*, 982-1012.
13. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281-296.
14. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*,

671-690.

15. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
16. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; Head-Gordon, M.; Clark, G. N.; Johnson, M. E.; Head-Gordon, T., Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549-2564.
17. Schmidhuber, J., Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85-117.
18. Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A., Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
19. Manzhos, S.; Jr., T. C., A random-sampling high dimensional model representation neural network for building potential energy surfaces. *J. Chem. Phys.* **2006**, *125*, 084109.
20. Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A., Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*, 13890.
21. Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E., Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, 241733.
22. Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E., Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10*, 2903.
23. 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.* **2017**, *8*, 3192-3203.
24. Behler, J.; Parrinello, M., Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
25. Smith, J.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A., Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*.
26. 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. Info. Model.* **2020**, *60*, 3408-3415.
27. Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T., End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, *119*, 9478-9508.
28. Brown, N.; Ertl, P.; Lewis, R.; Luksch, T.; Reker, D.; Schneider, N., Artificial intelligence in chemistry and drug design. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 709-715.
29. Schneider, P.; Walters, W. P.; Plowright, A. T.; Sieroka, N.; Listgarten, J.; Goodnow, R. A., Jr.; Fisher, J.; Jansen, J. M.; Duca, J. S.; Rush, T. S.; Zentgraf, M.; Hill, J. E.;

- Krutoholow, E.; Kohler, M.; Blaney, J.; Funatsu, K.; Luebke, C.; Schneider, G., Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **2020**, *19*, 353-364.
30. Watowich, S. J.; Meyer, E. S.; Hagstrom, R.; Josephs, R., A stable, rapidly converging conjugate gradient method for energy minimization. *J. Comput. Chem.* **1988**, *9*, 650-661.
 31. Gilbert, J. C.; Nocedal, J., Global Convergence Properties of Conjugate Gradient Methods for Optimization. *SIAM J. Optim.* **1992**, *2*, 21-42.
 32. Byrd, R.; Lu, P.; Nocedal, J.; Zhu, C., A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.* **2003**, *16*.
 33. Curtis, F. E.; Que, X., A quasi-Newton algorithm for nonconvex, nonsmooth optimization with global convergence guarantees. *Math. Program. Comput.* **2015**, *7*, 399-428.
 34. Grippo, L.; Lucidi, S., A globally convergent version of the Polak-Ribière conjugate gradient method. *Math. Program.* **1997**, *78*, 375-391.
 35. Wolfe, P., Convergence Conditions for Ascent Methods. *SIAM Rev.* **1969**, *11*, 226-235.
 36. Hager, W. W.; Zhang, H., A survey of nonlinear conjugate gradient methods. *Pacific J. Optim.* **2006**, *2*, 35-58.
 37. Mascarenhas, W. F., The BFGS method with exact line searches fails for non-convex objective functions. *Math. Program.* **2004**, *99*, 49-61.
 38. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J., Gaussian 16, Revision C.01. *Gaussian, Inc., Wallingford CT* **2016**.
 39. Wang, J., Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) through Computational Drug Repurposing Study. *J. Chem. Info. Model.* **2020**, *60*, 3277-3286.
 40. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, I., T.E. ; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.;

Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; D.M., Y.; Kollman, P. A., AMBER 2018. *University of California, San Francisco* **2018**.

41. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247-260.
42. Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A., Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110-2142.
43. Chai, J. D.; Head-Gordon, M., Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106.
44. Hunter, J. D., Matplotlib: A 2D Graphics Environment. *Comput. Sci. & Eng.* **2007**, *9*, 90-95.

Development and Evaluation of Geometry Optimization Algorithms in Conjunction with ANI Potentials

[Supporting Information]

Dongxiao Hao^{1,2}, Xibing He¹, Adrian E. Roitberg,³ Shengli Zhang², Junmei Wang^{1*}

*¹Department of Pharmaceutical Sciences and Computational Chemical Genomics
Screening Center, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA 15261, USA*

²School of Science, Xi'an Jiaotong University, Xi'an 710049, China

³Department of Chemistry, University of Florida, Gainesville 117200, USA

Table S1. Efficiency and accuracy of various optimization method combined with constraint method for 8 large molecules

	B3LYP/ 6-31G*	ANI/ CG-BS	ANI/ CG-WS	ANI/ L-BFGS-WS	ANI/ BFGS	GAFF
CPU time (second)	335438.0	1281.1	462.8	2004.1	959.0	0.56
Optimization Steps	57.7	1093.3	340.8	3161.7	1161.6	1580.4
AUE (kcal/mol)	2.28	2.99	2.85	2.92	2.94	3.74
RMSE (kcal/mol)	2.80	3.32	3.17	3.51	3.55	4.20
R	0.71	0.69	0.72	0.78	0.72	0.50
Ave. RMSD (Å)	1.27	1.39	1.36	1.19	1.24	1.41

*For minimization using GAFF, drms was set to 0.05.

Figure S1A

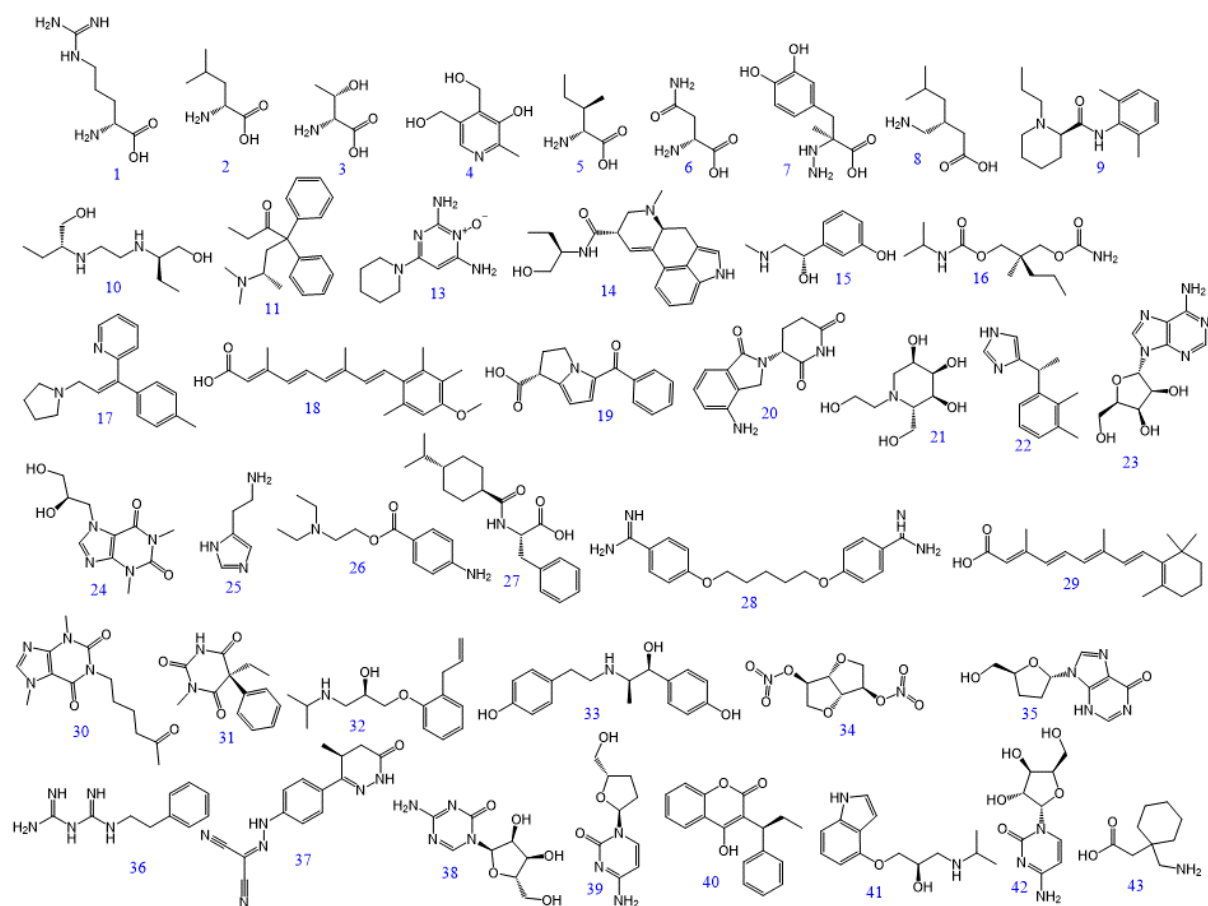


Figure S1B

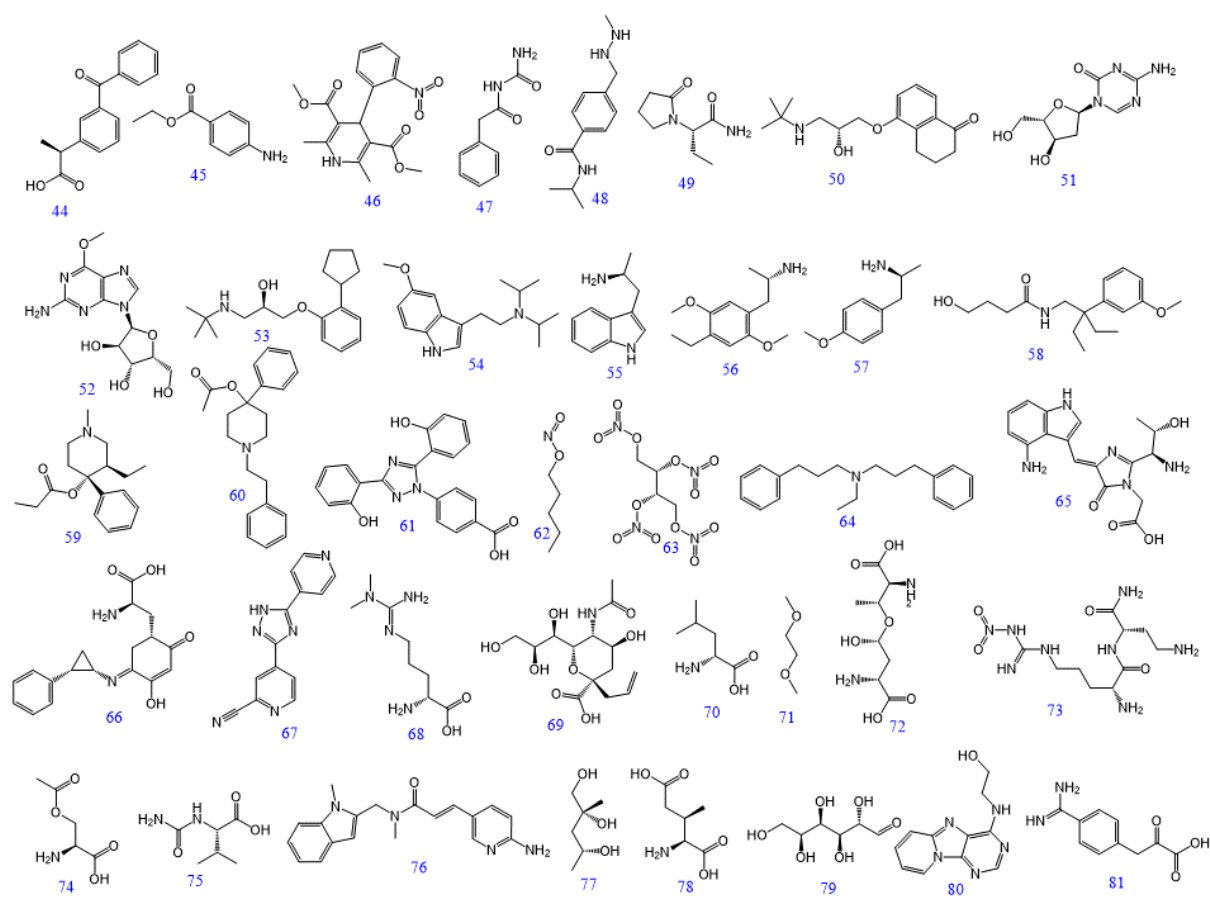


Figure S1C

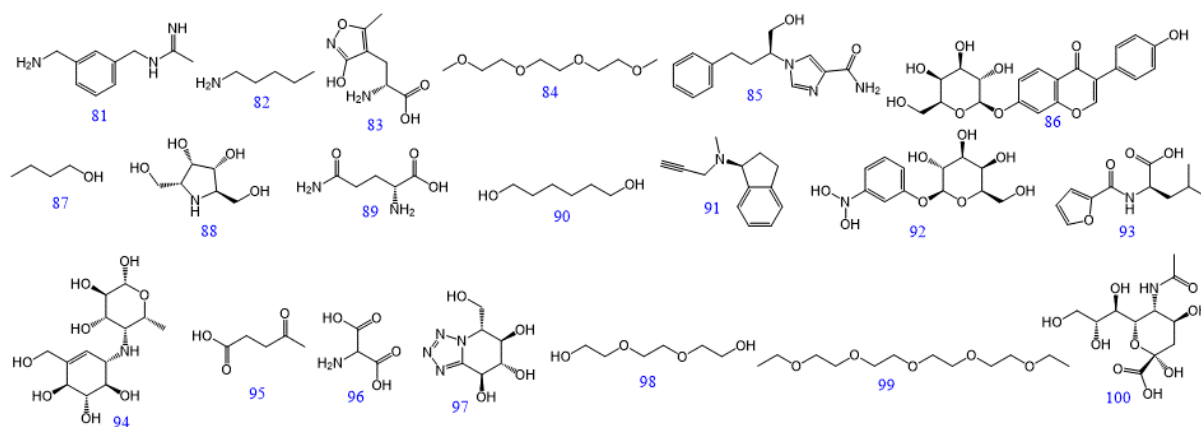


Figure S1. 2D-structures of 100 drug molecules with atom numbers ranged from 13 to 50.

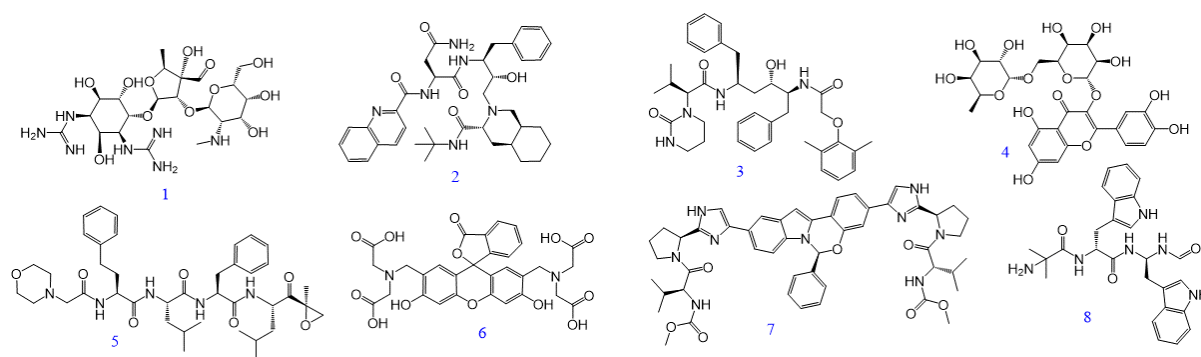


Figure S2. 2D-structures of 8 large molecules with atom numbers ranged from 65 to 120.

Figure S3A

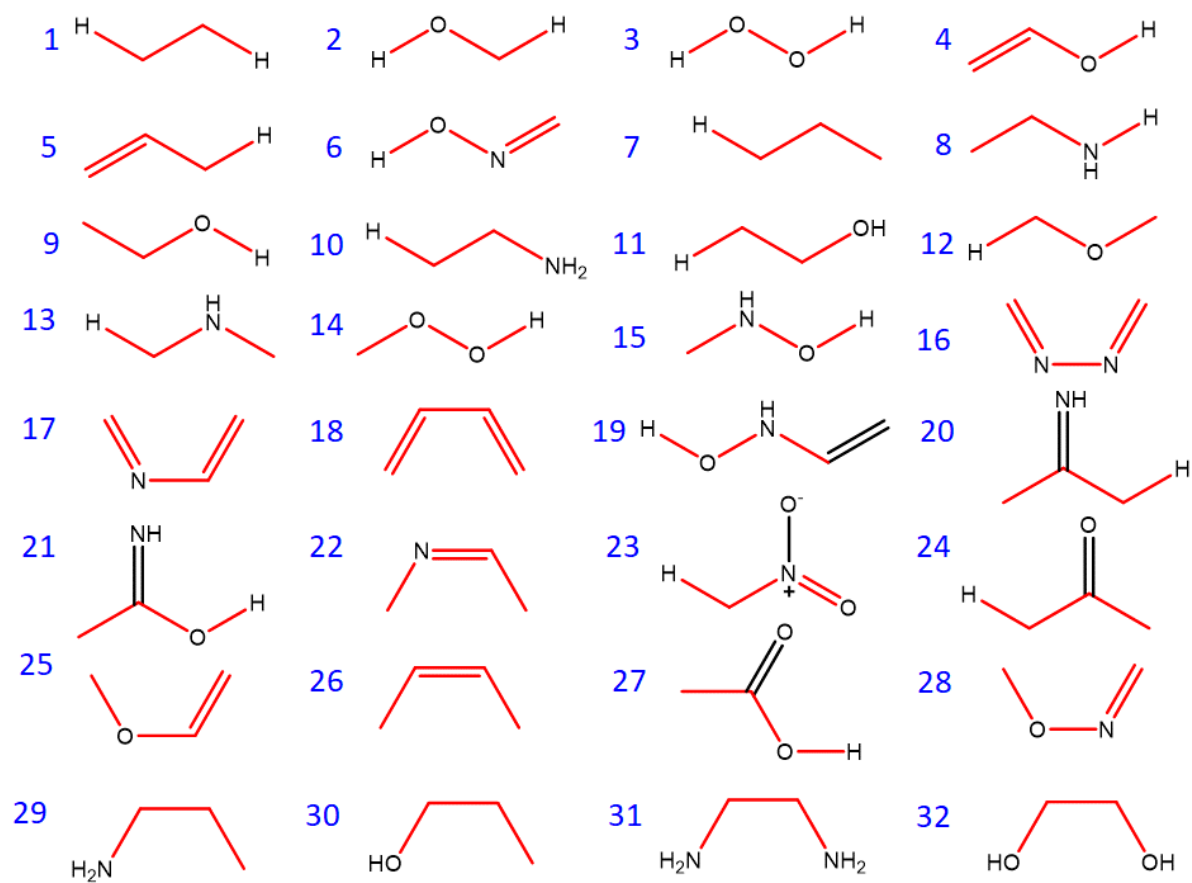


Figure S3B

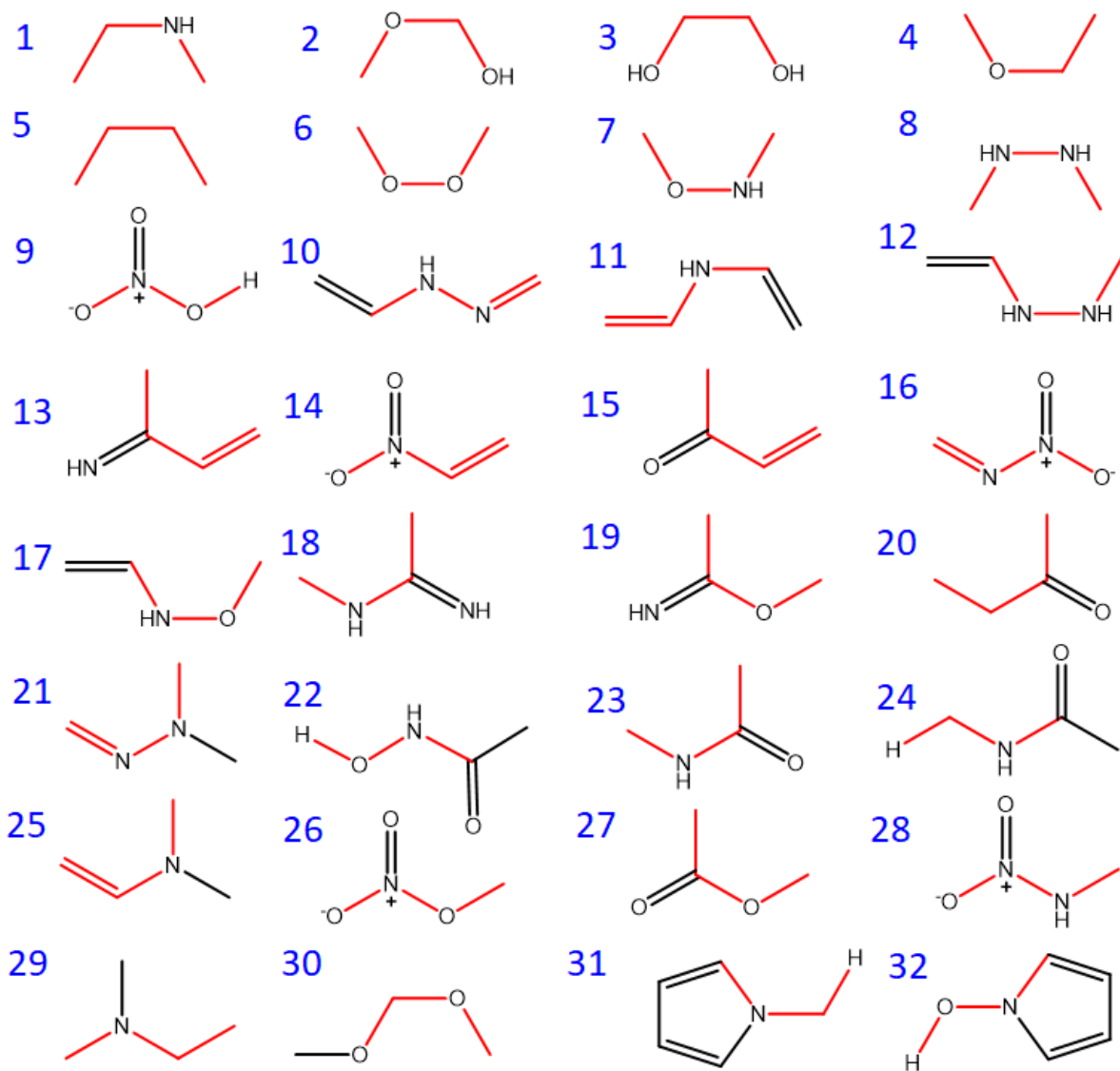


Figure S3C

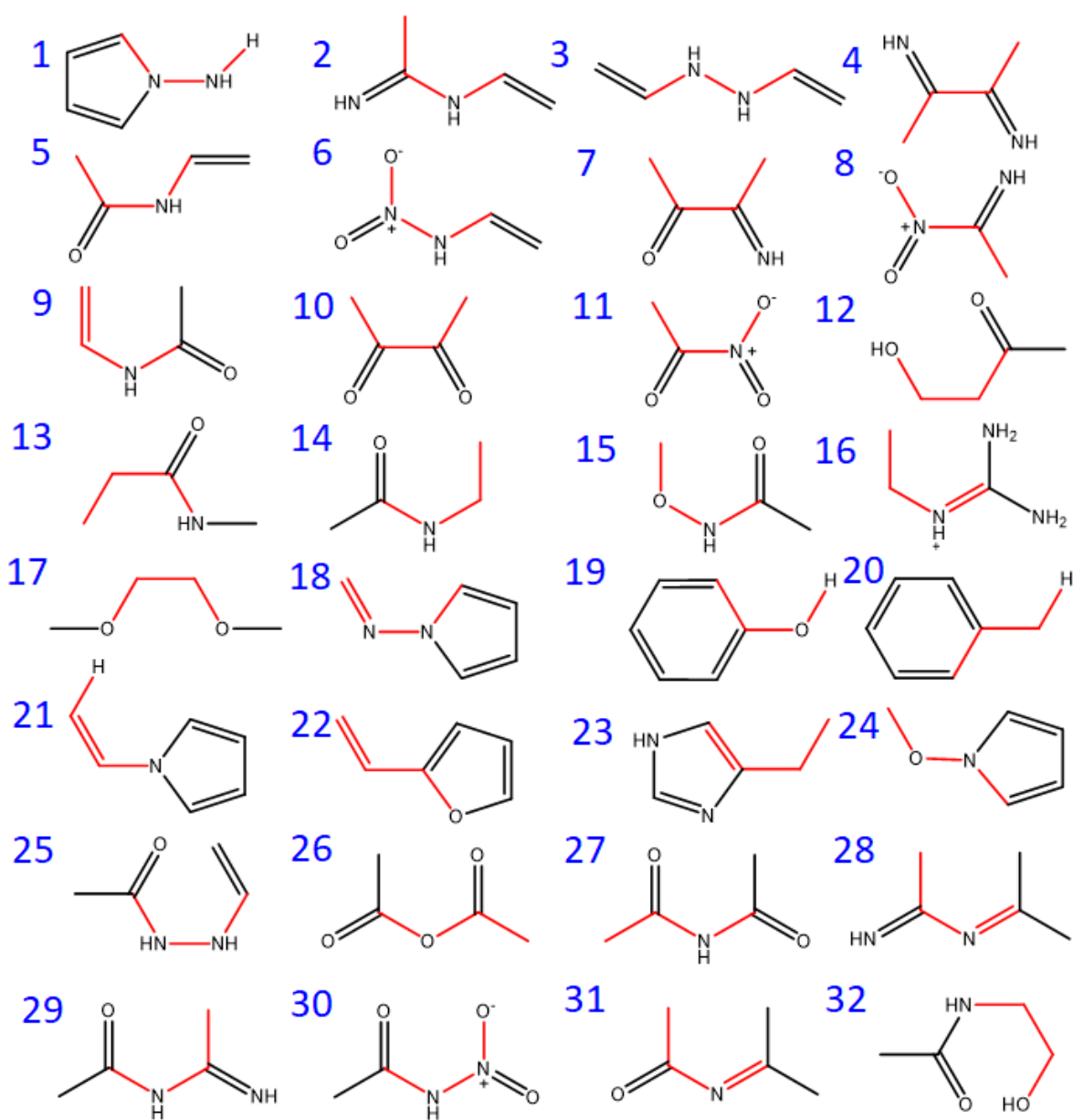


Figure S3D

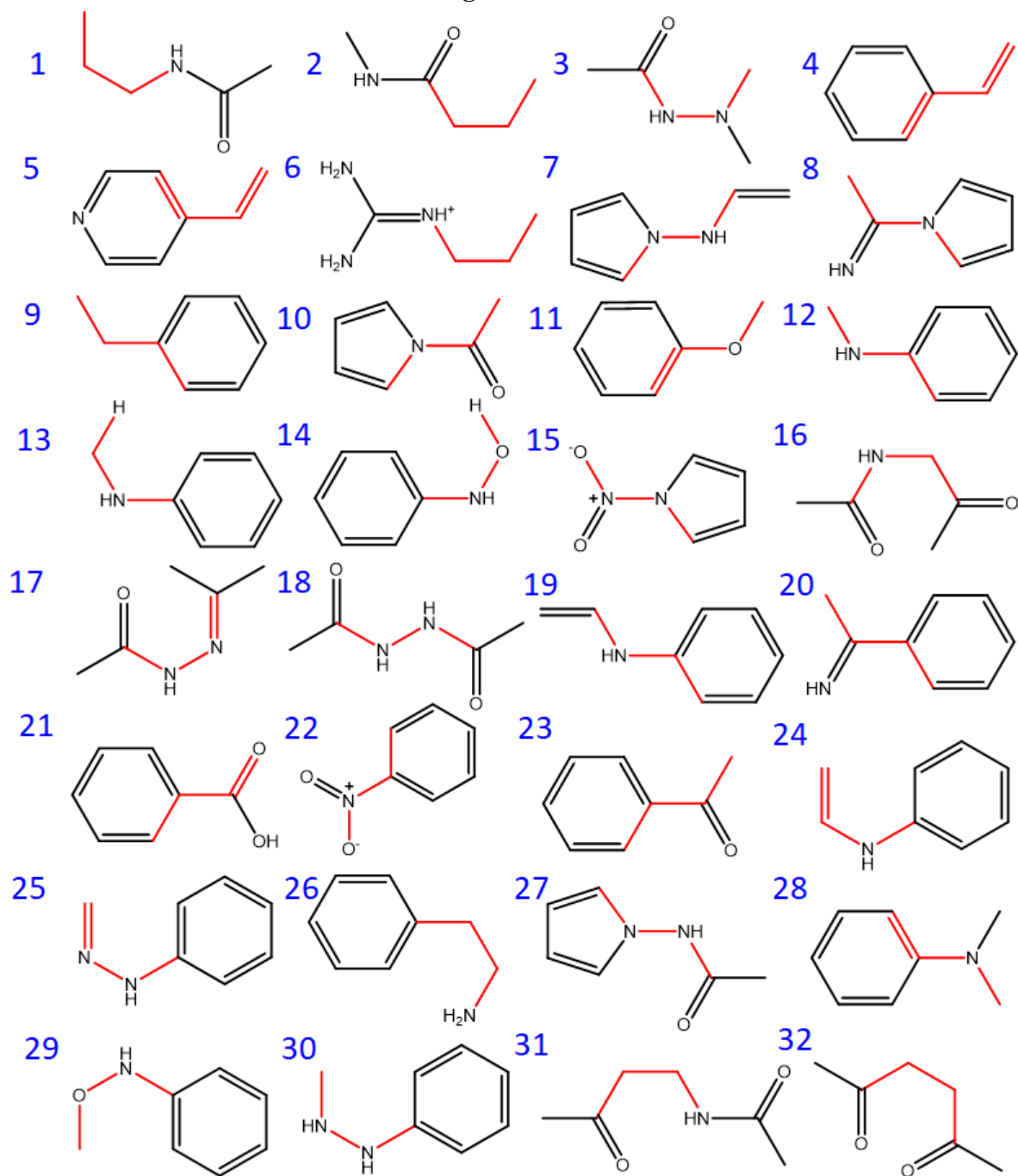


Figure S3E

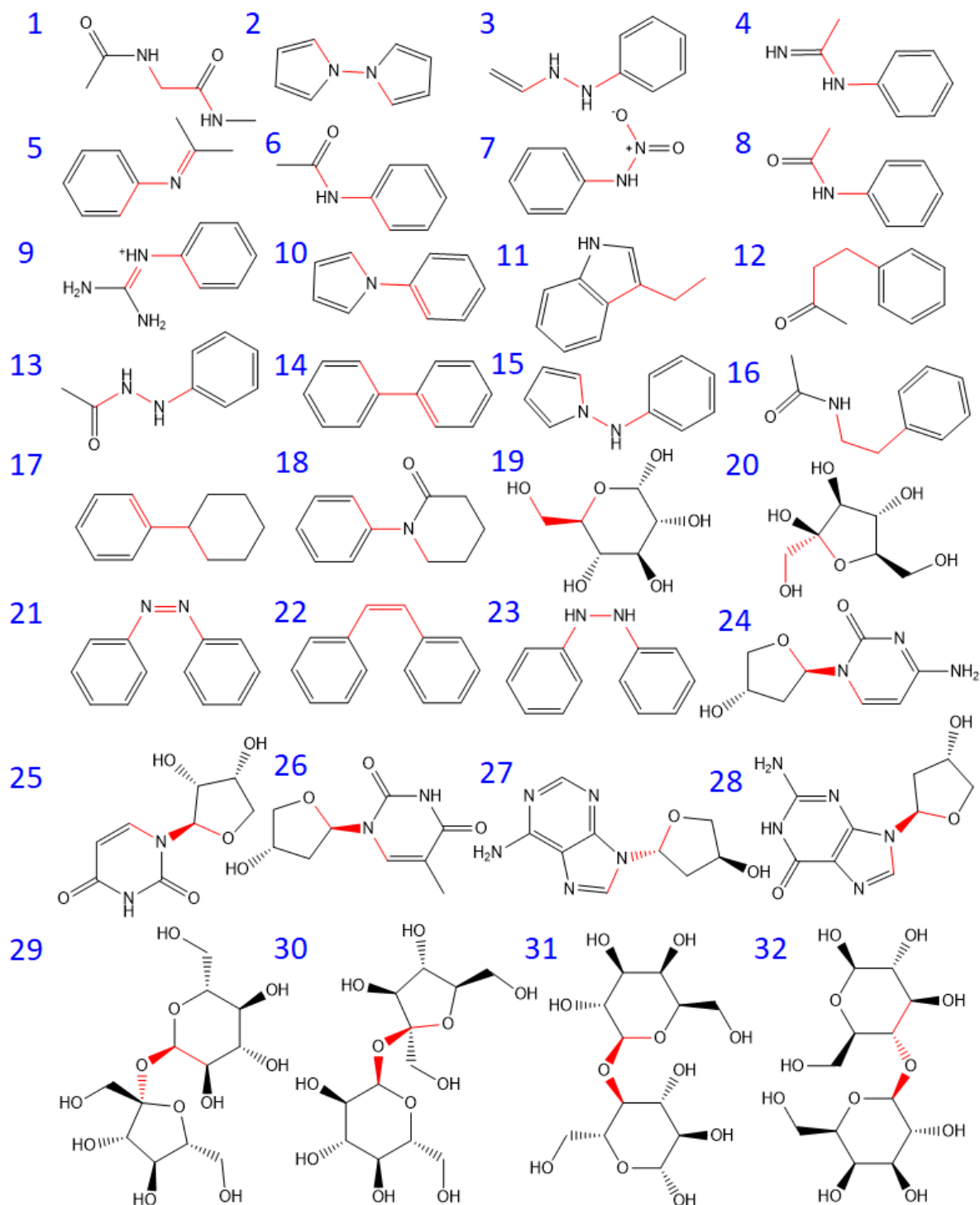


Figure S3. 2D-structures of 160 molecules arranged into five groups according to the numbers of heavy atoms, from the lowest (Group 1, Figure S3A) to highest (Group 5, Figure S3E). The torsional angles for which PESs were generated are colored in red.

Figure S4A

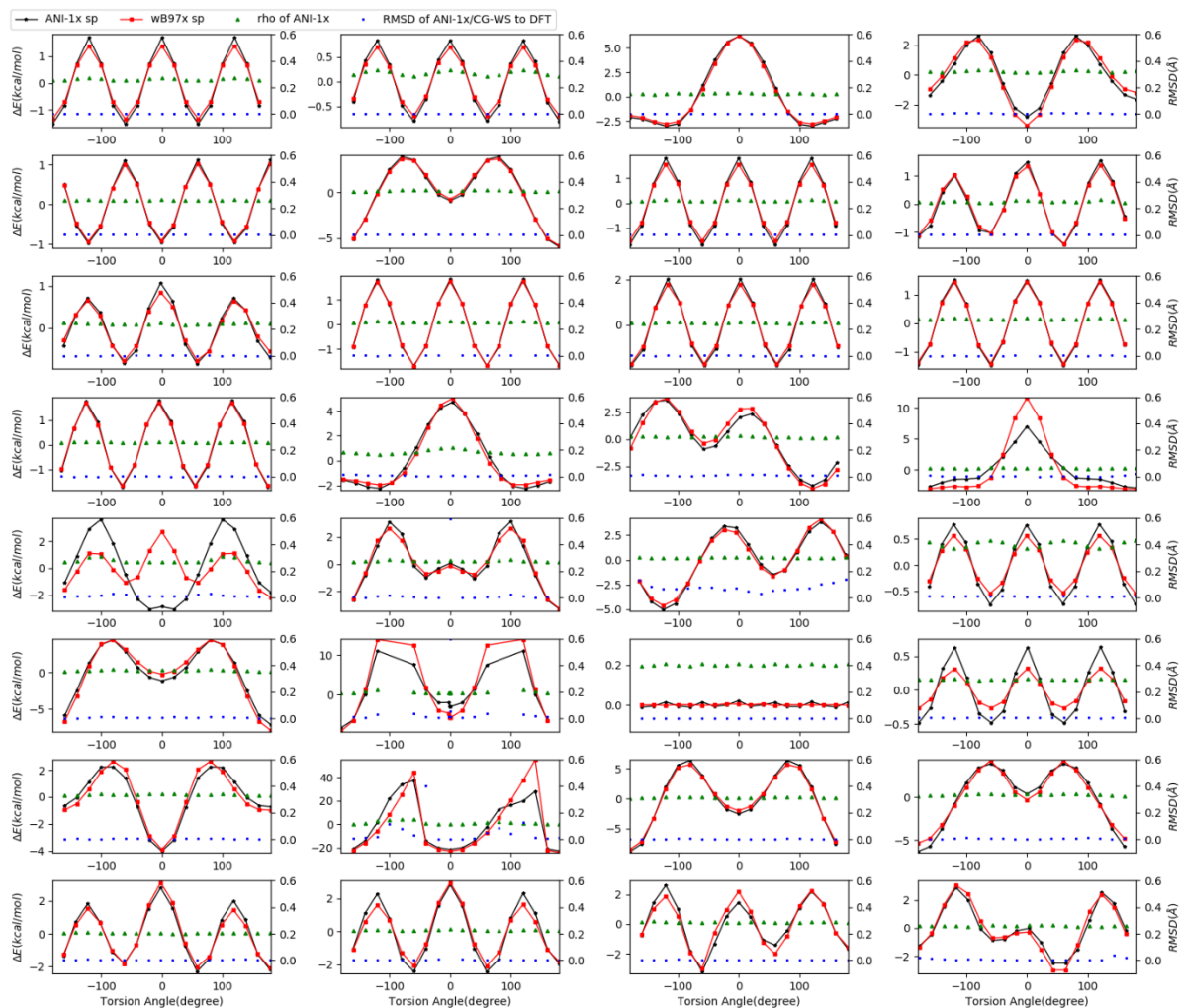


Figure S4B

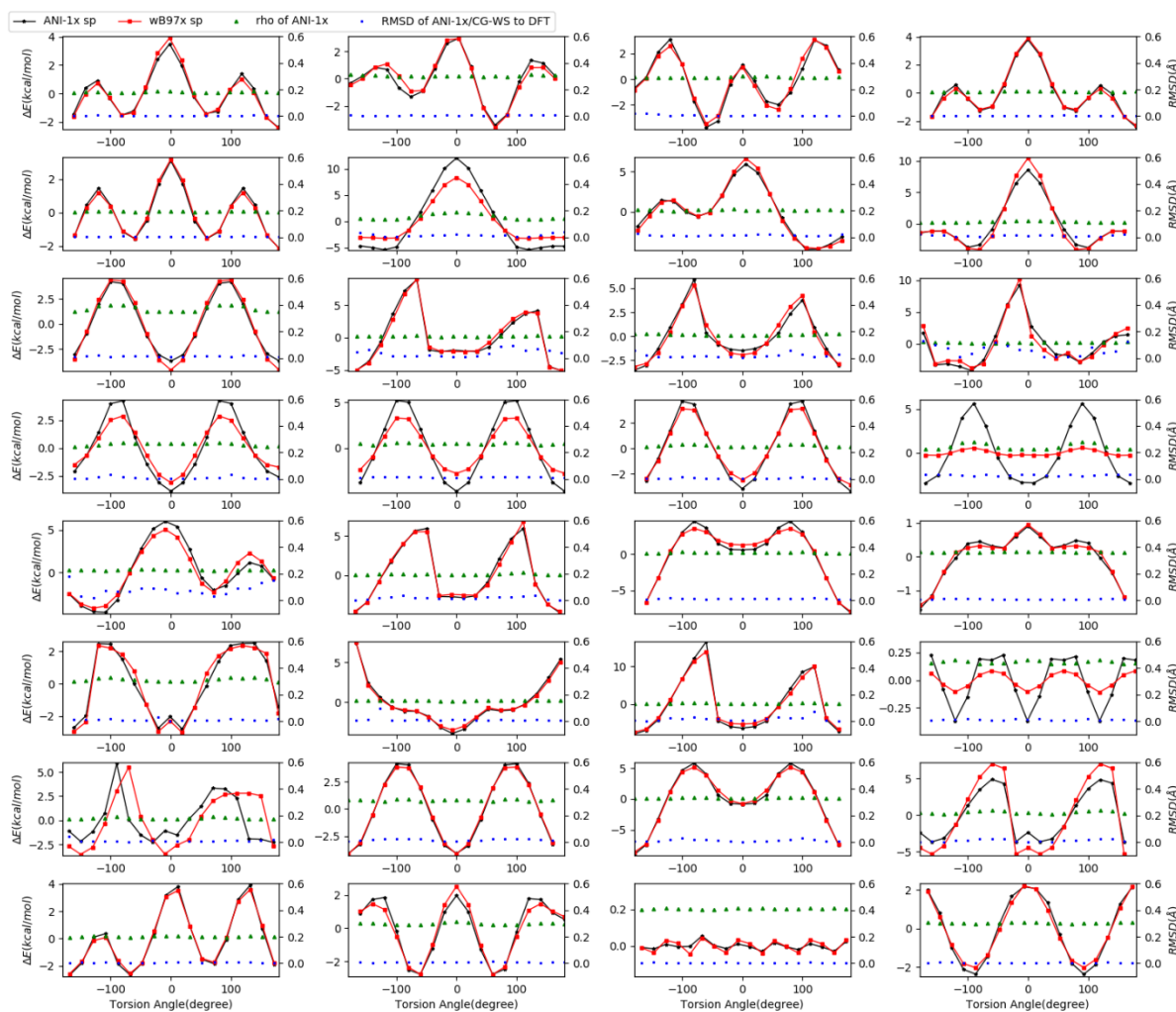


Figure S4C

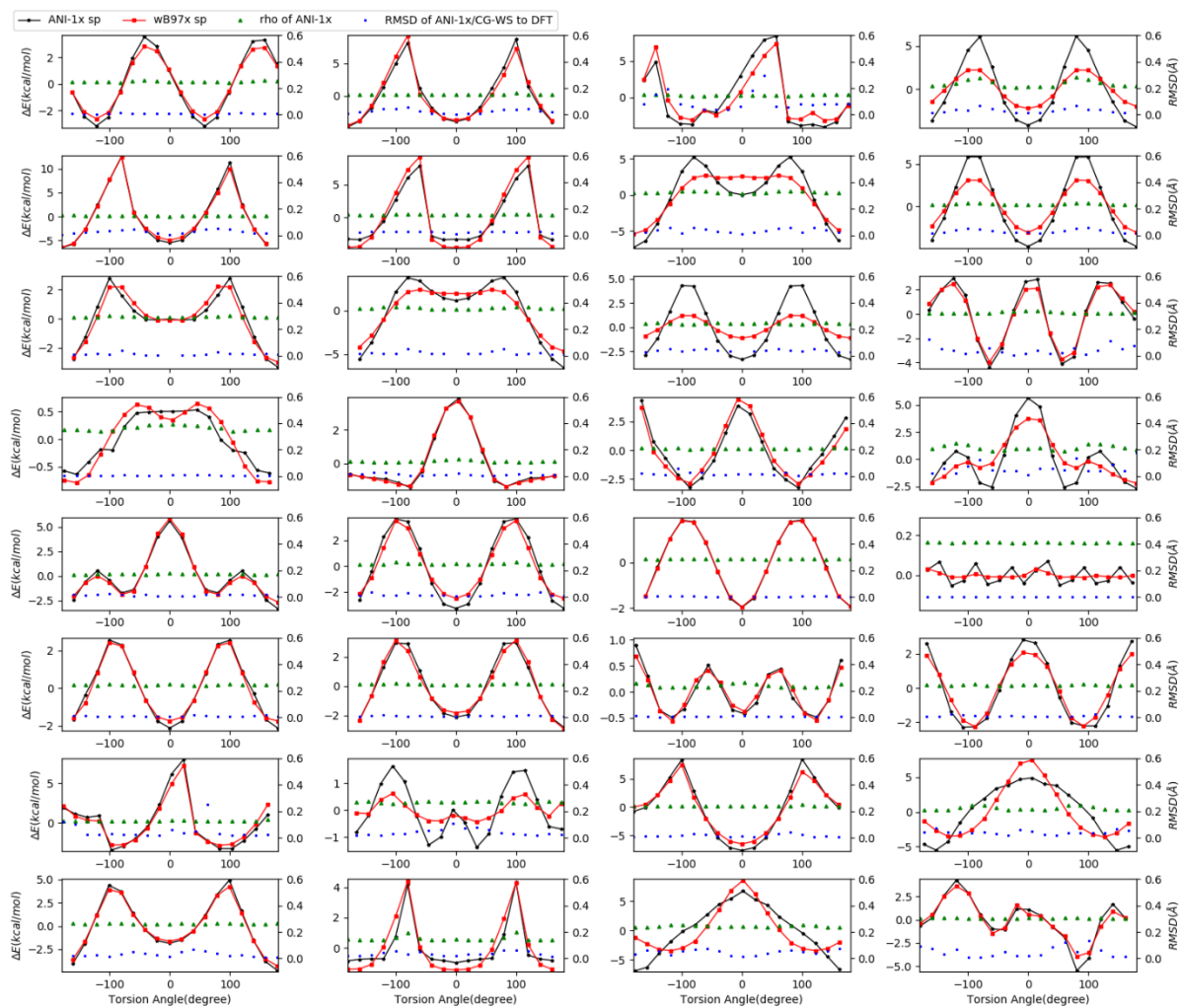


Figure S4D

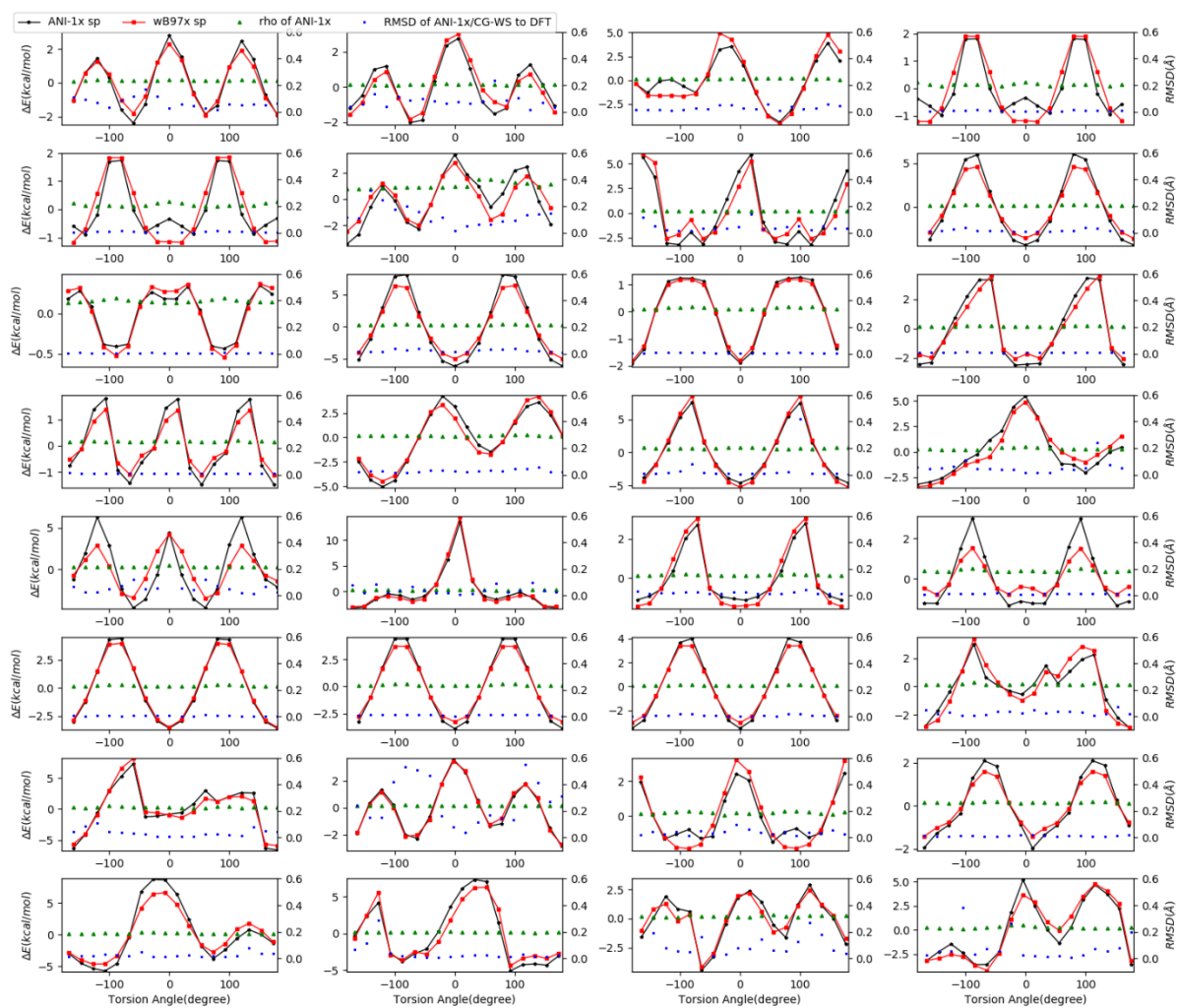


Figure S4E

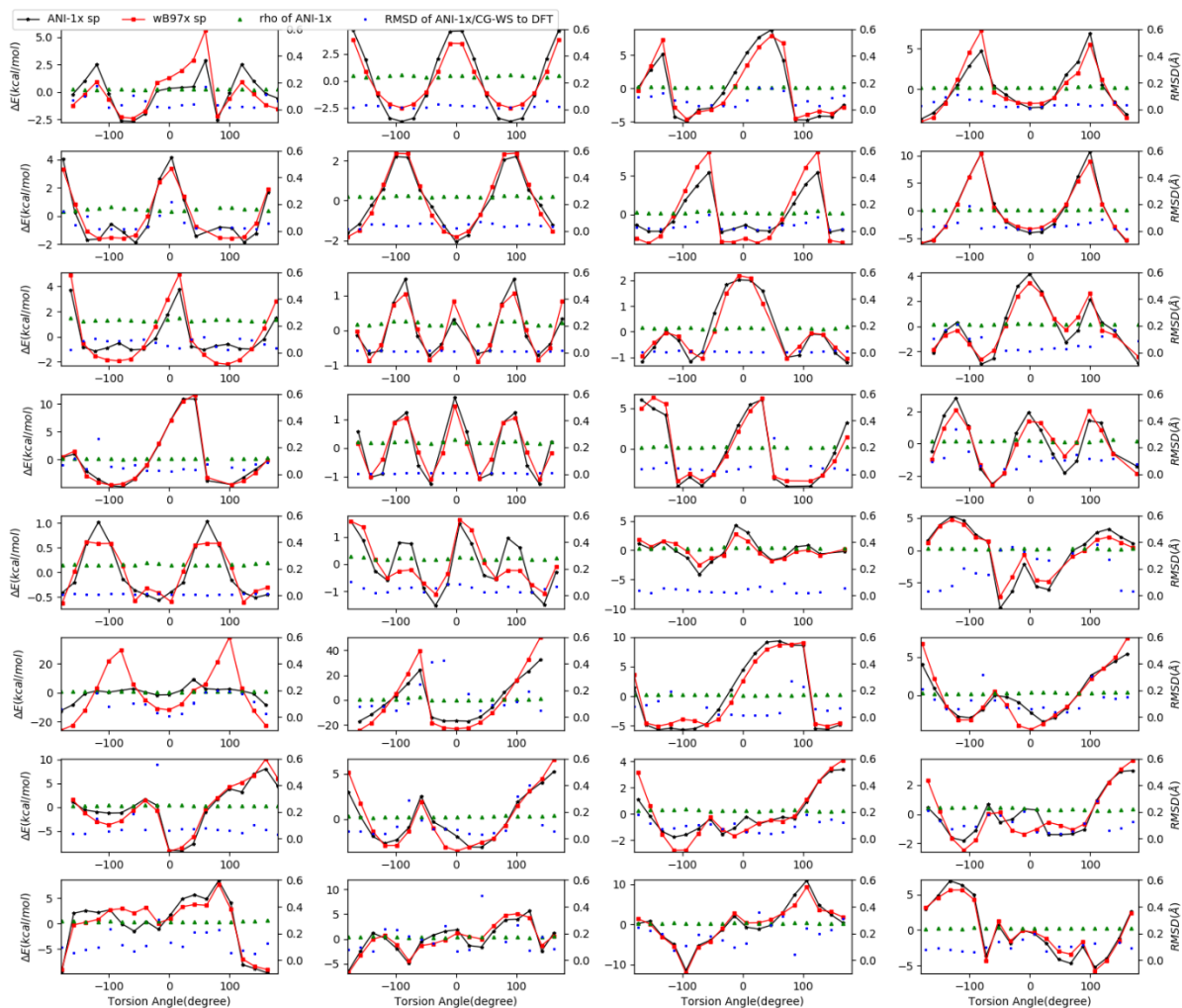


Figure S4. PESs for 160 molecules. A to E correspond to Group 1-5 molecules in Figure S3. Each subplot shows a one-dimensional potential surface scan generated using ω B97/6-31G(d) energies (red curves), and the ANI-1x energies (black curves) based on the B3LYP/6-31G* optimized structures using Jaguar. The rho values for ANI-1x calculations and RMSDs between ANI-1x-optimized and DFT*-optimized geometries are shown as green triangles and blue squares, respectively.

Figure S5A

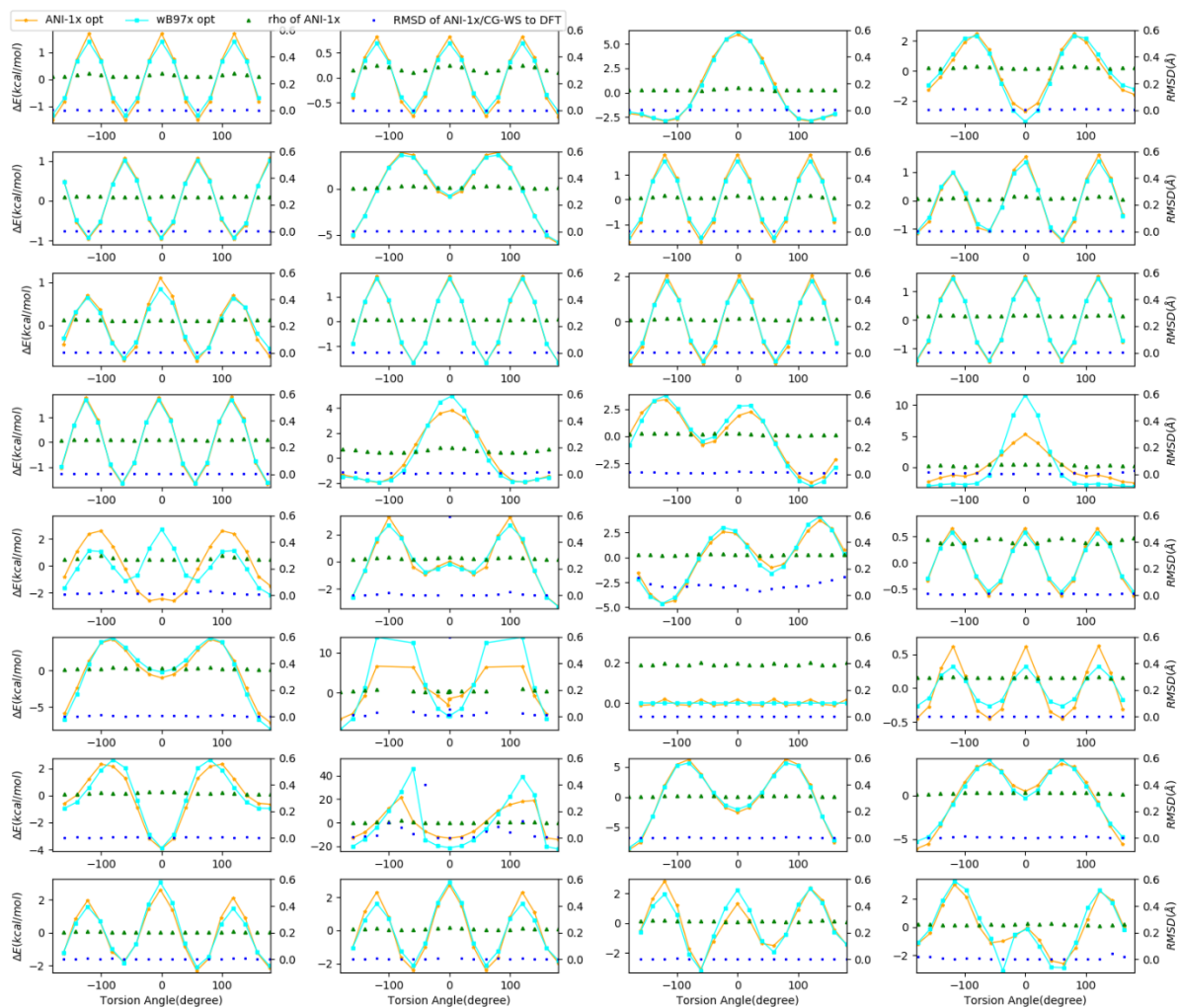


Figure S5B

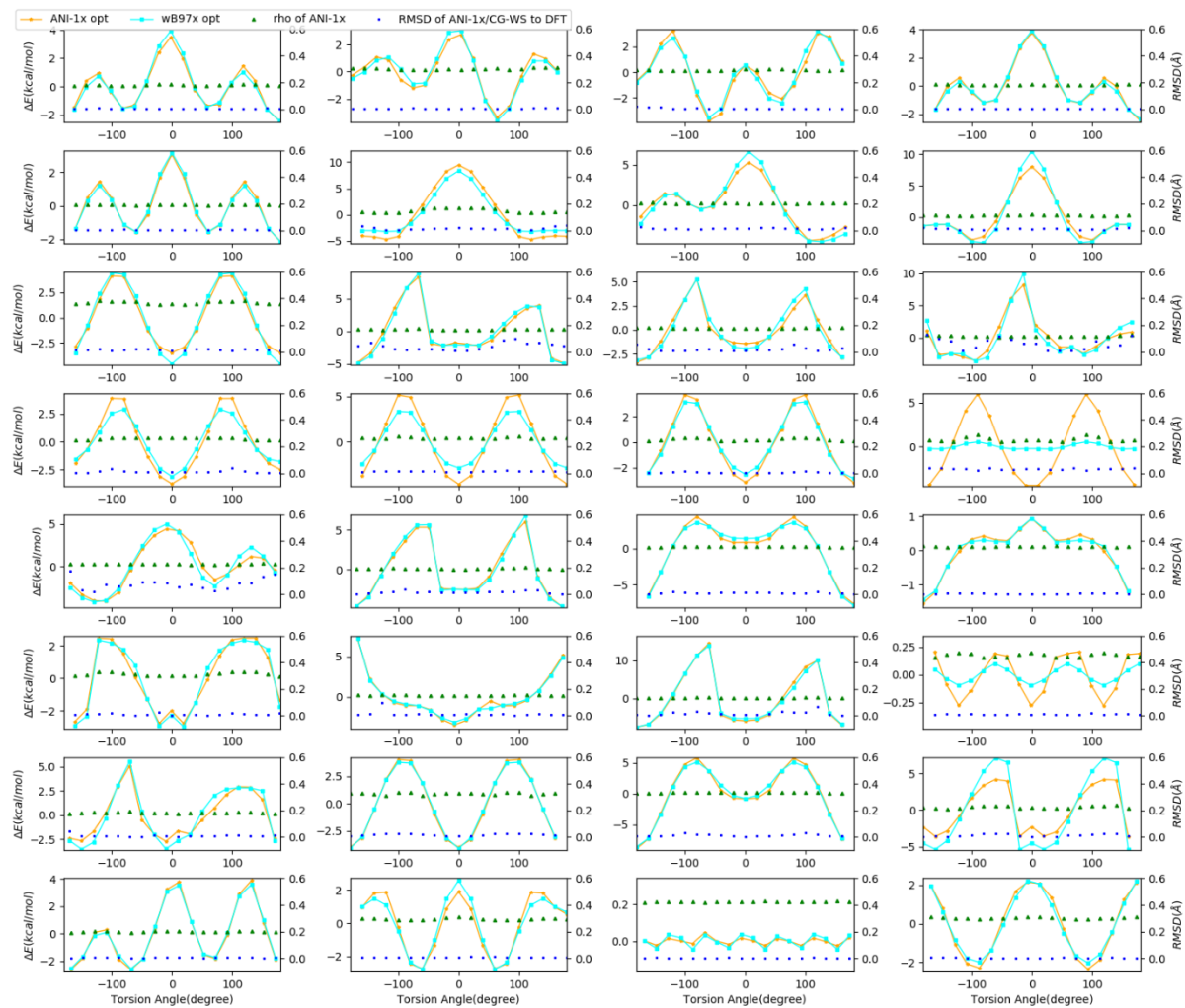


Figure S5C

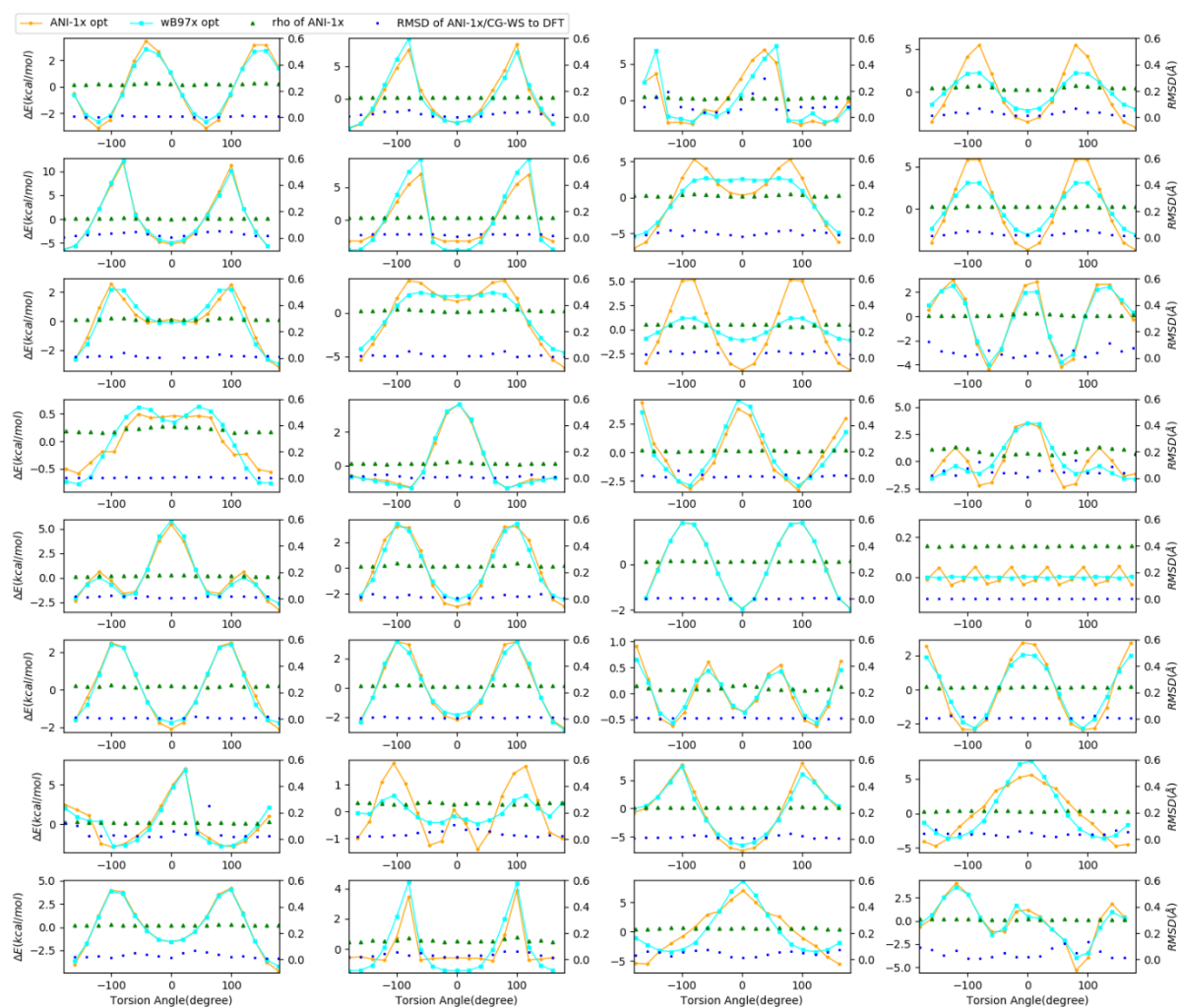


Figure S5D

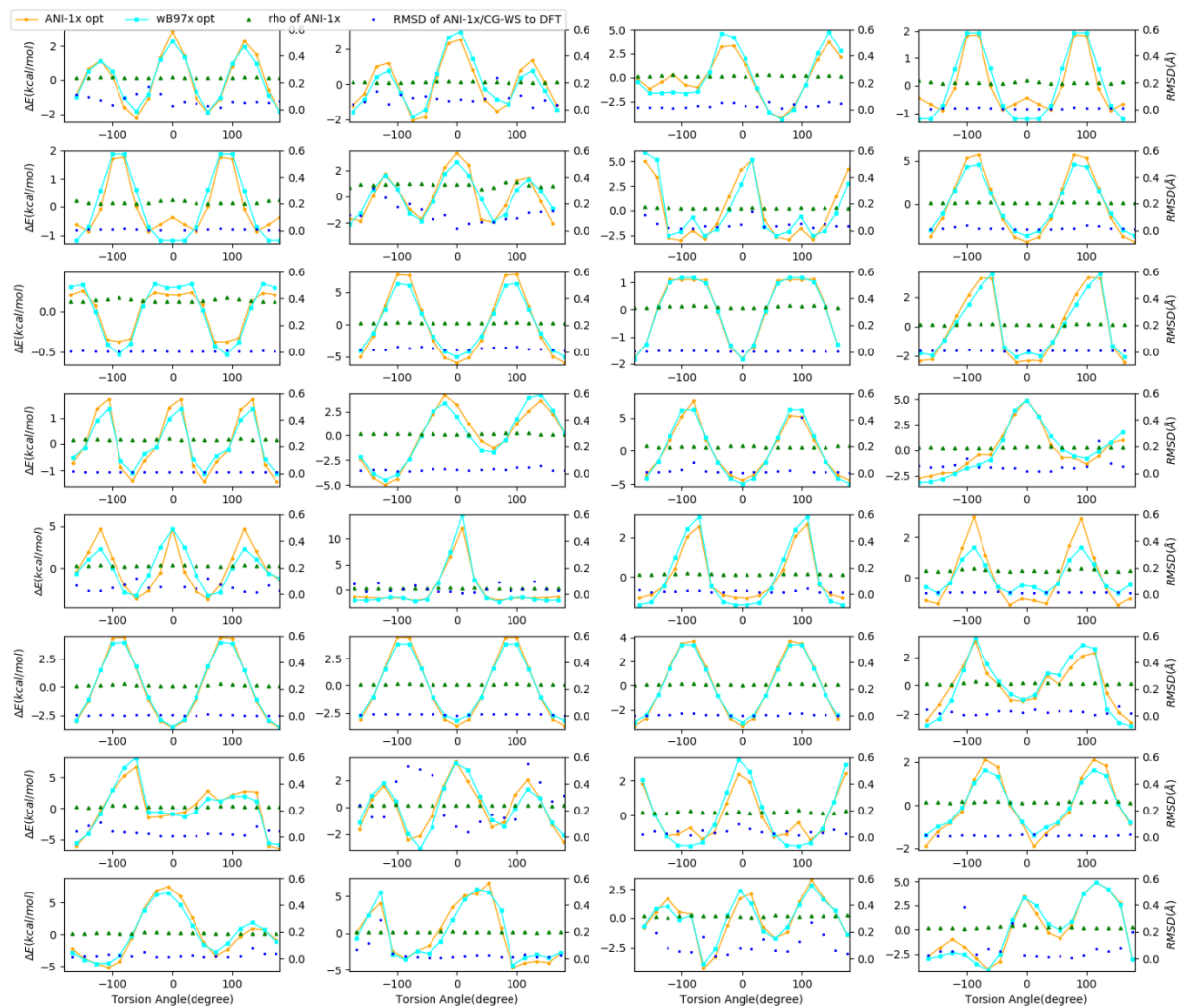


Figure S5E

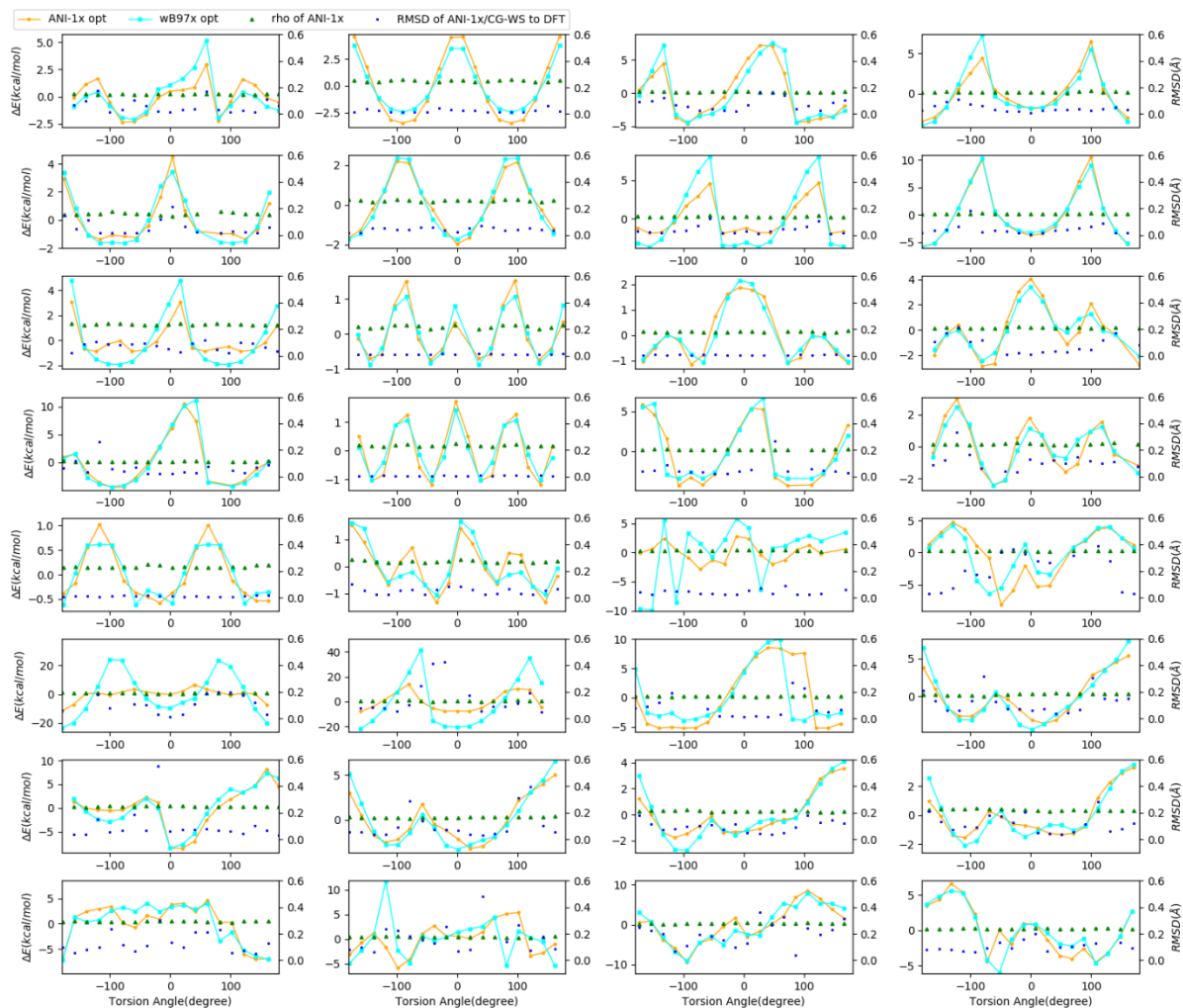


Figure S5. PESs for 160 molecules. A to E correspond to Group 1-5 molecules in Figure S3. Each subplot shows a one-dimensional potential surface scan generated using ω B97/6-31G(d) energies after geometric optimization (cyan curves), and using ANI-1x after optimization by I-ACG/ANI-1x (orange curves). The rho values for ANI-1x calculations and RMSDs between ANI-1x-optimized and wB97x/6-31G(d)-optimized geometries are shown as green triangles and blue squares, respectively.