

Chinese Chemical Society | Xiamen University

**Journal of Electrochemistry**

---

Online First

---

11-6-2023

## **Automated Workflow for Redox Potentials and Acidity Constants Calculation from Machine Learning Molecular Dynamics**

Feng Wang

Jun Cheng

---

# 机器学习加速氧化还原电位和酸度常数计算

王锋<sup>a</sup>, 程俊<sup>a,b,c\*</sup>

(a. 厦门大学化学化工学院, 固体表面物理化学国家重点实验室, 厦门 361005; b. 厦门大学能源材料化学协同创新中心, 厦门 361005; c. 嘉庚创新实验室, 厦门 361005)

**摘要:** 氧化还原电位和酸度常数作为重要的物理化学性质被应用于分析能源材料重要指标值。为了实现能源材料的计算设计, 发展计算电化学的方法, 在复杂电化学环境下计算这些性质至关重要。近年来, 利用计算化学方法计算氧化还原电位和酸度常数已经受到了广泛的关注。然而, 常用的计算方法如基于隐式溶剂化模型的小分子自由能计算, 对于复杂溶剂化环境的处理非常有限。因此, 基于第一性原理分子动力学 (AIMD) 的自由能计算被引入来描述复杂溶剂化环境中的溶质-溶剂相互作用。同时, 基于 AIMD 的自由能计算方法已经被证实可以准确预测这些物理化学性质。然而, 由于 AIMD 计算效率低且计算资源需求大, 需要引入机器学习分子动力学 (MLMD) 加速计算。MLMD 通过机器学习方法, 构建模拟体系结构到第一性原理计算结果的一对一映射, 可以在低成本下实现长时间尺度的 AIMD。对于氧化还原电位和酸度常数计算, 如何构建训练机器学习势函数模型所需的数据集至关重要。本文介绍了如何通过自动化 workflows 实现自由能计算势函数的自动化构建, 通过机器学习分子动力学计算自由能并转化为对应的物理化学性质。

**关键词:** 机器学习分子动力学; 自动化 workflows; 复杂体系; 自由能计算

## 1 引言

氧化还原电位和酸度常数是材料和分子重要的物理化学性质。这些性质对物理、材料、生命科学、地质科学都有非常重要的应用<sup>[1-3]</sup>。这些性质决定了质子和电子的热力学驱动力, 而这些驱动力在许多人造和自然的过程中都扮演非常重要的角色。例如, 电池体系中的电子转移过程、光催化中的电荷转移和地质过程中金属的形态形成和迁移等。几十年来, 开发计算这些性质的第一性原理方法一直是计算化学的中心话题, 并且已经取得了很大进展。配备这些计算工具, 计算科学家能够筛选和设计具有所需功能的新材料。例如, 氧化还原电位的计算已被用于设计用于氧化还原液流电池的具有宽电化学窗口的新型有机电解质。<sup>[4]</sup>

对于一些理想和均相的溶液, 这些性质在计算时会使用连续质模型描述其溶剂化环境。在更复杂或非均相的溶液中 (如离子液体或高浓度电解液), 由于体系的溶剂化环境过于复杂, 连续质模型就失效了。<sup>[5]</sup> 为了解决这个限制, 第一性原理分子动力学 (ab initio molecular dynamics, AIMD) 被引入来描述这类更为复杂的溶剂化结构。区别于连续质模型, AIMD 将溶剂作为分子引入模拟体系, 同时使用和溶

1 质相同的电子结构理论描述溶剂分子，这样有效地解决了复杂溶剂化结构不好描述的问题。<sup>[6-9]</sup>通过结  
 2 合自由能微扰（Free Energy Perturbation, FEP）<sup>[10]</sup>和热力学积分(Thermodynamic Integration, TI)<sup>[11-13]</sup>的方  
 3 法，氧化还原电位和酸度常数可以被准确的计算出来<sup>[14-16]</sup>。当然，计算结果的准确性受限于 AIMD 计算  
 4 的统计收敛和所使用 DFT 泛函的精度<sup>[17-20]</sup>。同时，基于 AIMD 的方法最明显的缺陷在于其对计算资源的  
 5 要求较高，其原因在于体系中包含的电子数比一般的小分子体系多很多同时每一步分子动力学都需要收  
 6 敛体系的波函数已达到精确的能量和原子受力。

7 为了加速 AIMD 模拟，机器学习方法被应用于创建从体系的原子结构到其对应的能量和原子受力的  
 8 一对一映射，即机器学习势函数模型。通过使用代表性的数据集（原子结构与其对应的势能和原子受  
 9 力）训练机器学习势函数（Machine Learning Potential, MLP）<sup>[21-30]</sup>，势能和原子受力可以很准确地被预  
 10 测出来，同时所需要的计算资源相对于 DFT 少非常多。然而，使用机器学习势函数计算自由能，并将自  
 11 由能转化为氧化还原电位和酸度常数并不简单。首先，由于 FEP 和 TI 方法需要引入两个势能面，即对于  
 12 同一个结构有两套势能和原子受力，这个和机器学习势函数的理念有所区别。同时，在势能面上演化结  
 13 构的过程中会引入一些中间态的结构，预测准这类结构的能量和原子受力需要对数据集进行细致的挑选。

14 AI<sup>2</sup>-kit（artificial intelligence × ab initio）<sup>[31]</sup>基于这一问题，通过结合 FEP 采样和迭代更新结构数据  
 15 集<sup>[32, 33]</sup>的方法来构建机器学习势函数并计算这类电化学性质。对于不同体系的势函数，其训练误差非常  
 16 小。在计算酸度常数、氧化还原电位和溶剂化自由能时，机器学习势函数分子动力学（Machine Learning  
 17 Molecular Dynamics, MLMD）计算和 AIMD 吻合地非常好。并且在计算过程中，自由能性质计算速度  
 18 提升了 1000 倍以上。

19

## 20 2 计算方法和计算设置

### 21 2.1 自由能计算方法

22 在正则系综（NVT，恒定粒子数、体积和温度）的理论框架下，模拟体系的亥姆霍兹自由  
 23 能（A）可以表述为

$$A = -k_B T \ln Q_{NVT}, \quad 1$$

24 其中  $Q_{NVT}$  为体系的配分函数。在相空间中，配分函数可以表示成

$$Q_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \int d\mathbf{p}^N \exp\left[-\frac{|\mathbf{p}|^2}{2mk_B T}\right] \int d\mathbf{r}^N \exp\left[-\frac{\mathcal{V}(\mathbf{r}^N)}{k_B T}\right], \quad 2$$

25 其中第一个积分只与体系的动量相关，而第二个积分只和体系中原子的位置相关。与动量相关  
 26 的积分和常数项可以解析得到，为理想气体的配分函数  $Q_{NVT}^{ideal}$ 。而由原子位置决定的积分则是体  
 27 系偏离理想气体（存在相互作用）产生的配分函数  $Q_{NVT}^{excess}$ 。所以正则系综的配分函数  $Q_{NVT}$  可以

1 简化为 $Q_{NVT}^{ideal}Q_{NVT}^{excess}$ 。在此基础上，计算一个化学反应的自由能 $\Delta A$ ，可以通过反应物和产物的配  
2 分函数得到，即

$$\Delta A = A_P - A_R = -k_B T \ln \left[ \frac{Q_P^{excess}}{Q_R^{excess}} \right], \quad 3$$

3 其中 $Q_P^{excess}$ 为产物状态的配分函数 $Q_R^{excess}$ 为反应物状态的配分函数。 $\Delta A$ 的表达式即为自由能计  
4 算的基础公式。将过剩配分函数的表达式带入后可以得到

$$\Delta A = -k_B T \ln \frac{\int dr^N \exp \left[ -\frac{v_P(\mathbf{r}^N) - v_R(\mathbf{r}^N)}{k_B T} \right] \exp \left[ -\frac{v_R(\mathbf{r}^N)}{k_B T} \right]}{\int dr^N \exp \left[ -\frac{v_R(\mathbf{r}^N)}{k_B T} \right]} = -k_B T \ln \left\langle \exp \left[ -\frac{\Delta E}{k_B T} \right] \right\rangle_R, \quad 4$$

5 其中尖括号 $\langle \dots \rangle$ 表示在反应物势能面上的系综平均，而 $\Delta E$ 表示垂直能量差（同一结构在产物和  
6 反应物势能面上对应势能的差值。所以只要在势能面上充分的采样，就可以通过计算 $\Delta E$ 的系综  
7 平均得到准确的自由能。然而，由于分子动力学在势能面上的运动是比较局域的，无偏差的分  
8 子动力学模拟只能在反应物或者产物的区间上采样，无法得到准确的自由能。所以，在此基础  
9 上通过引入自由能微扰和热力学积分的方法来解决采样和自由能计算的问题。

10 在 $\Delta A$ 的统计热力学表达式的基础上，可以通过引入FEP和TI的方法计算体系的自由能变。  
11 FEP的主要原理是在始态和终态之间添加多个中间态，构建自由能计算的路径，增强分子动力  
12 学的采样，从而可以通过热力学积分计算出从始态到终态的自由能。所以通过引入N个中间态，  
13  $\Delta A$ 可以表示为

$$\Delta A = -k_B T \ln \left[ \frac{Q(P)}{Q(N)} \frac{Q(N)}{Q(N-1)} \dots \frac{Q(2)}{Q(1)} \frac{Q(1)}{Q(R)} \right]. \quad 5$$

14 对于中间态势能面的构建，本论文使用的方法是反应物和产物势能面的线性组合，即

$$E_\eta(\mathbf{r}^N) = \eta E_P(\mathbf{r}^N) + (1 - \eta) E_R(\mathbf{r}^N), \quad 6$$

15 其中， $\eta$ 为Kirkwood耦合参数<sup>[34]</sup>，其取值范围从0到1。 $\eta = 0$ 对应的是反应物势能面， $\eta = 1$   
16 对应的是产物势能面。通过改变 $\eta$ 的数值，可以将势能面逐渐从反应物势能面变化到产物的势能  
17 面。当 $\eta$ 的取值足够密时，自由能变即为

$$\Delta A = A(\eta = 1) - A(\eta = 0) = \int_0^1 \frac{\partial A(\eta)}{\partial \eta} d\eta. \quad 7$$

18  
19 将公式1代入后，得到

$$\Delta A = \int_0^1 \frac{\int d\mathbf{r}^N \frac{\partial E_\eta(\mathbf{r}^N)}{\partial \eta} \exp\left[-\frac{E_\eta(\mathbf{r}^N)}{k_B T}\right]}{\int d\mathbf{r}^N \exp\left[-\frac{E_\eta(\mathbf{r}^N)}{k_B T}\right]} d\eta = \int_0^1 \left\langle \frac{\partial E_\eta}{\partial \eta} \right\rangle_\eta d\eta = \int_0^1 \langle \Delta E \rangle_\eta d\eta \quad [12] \quad 8$$

1 此式就是通过热力学积分计算自由能的主要表达式。对于不同的反应过程，氧化还原过程、脱  
 2 质子过程或者溶剂化过程，需要根据势能面的状态设置不同的采样势能面 $E_\eta$ 。同时，不同体系  
 3 对应不同的自由能校正项，需要根据自由能计算过程中做的近似进行校正。因此，通过自由能  
 4 微扰和热力学积分，可以通过分子动力学模拟得到比较准确的自由能计算。

5 对于不同的计算目的设置不同的积分路径，计算其自由能是模拟电化学体系的关键。对于  
 6 计算酸度常数，所需要计算的是酸的脱质子自由能，以酸 XH 为例。其始态势能面为溶剂化的  
 7 XH，终态势能面为溶剂化的 $X^-$ 。其中 XH 上的质子在计算过程中通过 FEP 构建采样势能面的方  
 8 法逐渐变换成一个虚原子。由于计算自由能过程中计算的垂直能量差为该虚原子可逆插入该模  
 9 拟体系的能量差，所以对于虚原子插入的位置有两个要求：其一，虚原子插入的位置应在 XH  
 10 中质子所在平衡位置附近，这样才能表示酸 XH 对应的真实结构；其二，虚原子不能和其他原  
 11 子发生重叠，重叠会导致将其变成真实原子后存在巨大能量，而这个能量在化学过程中非常不  
 12 合理。所以为了解决这个问题，我们需要将虚原子控制在其平衡位置附近，具体的操作方式为  
 13 添加一个谐振子势能<sup>[35]</sup>，即

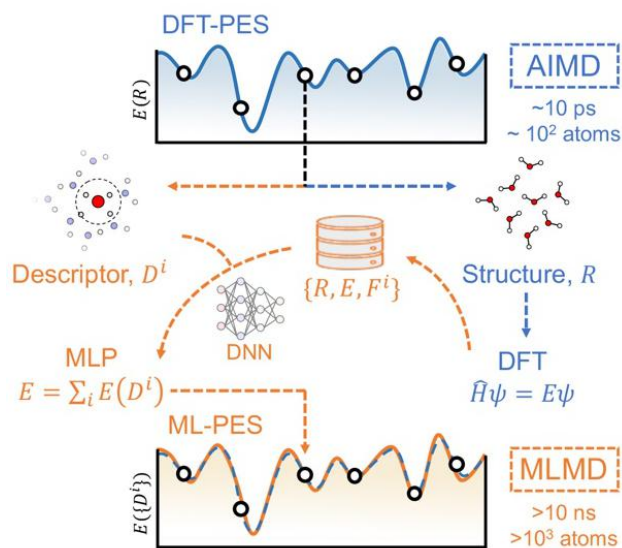
$$V_r(\mathbf{R}^N) = \sum_i^{N_d} \frac{k_{di}}{2} (d_i - d_{i0})^2 + \sum_i^{N_\alpha} \frac{k_{\alpha i}}{2} (\alpha_i - \alpha_{i0})^2, \quad 9$$

14 其中，等式右边的前半部分为控制键长的谐振子势能，其中 $d_{i0}$ 为该键的平衡键长；后半部分为  
 15 控制键角的谐振子势能， $\alpha_{i0}$ 为该键角的平衡键角大小。

16 对于氧化还原电位的计算，由于在插入一个电子后会导致模拟体系中的分子或离子发生自  
 17 发的断键过程。断键的碎片分子/离子会在分子动力学模拟的过程中逐渐被溶剂化分离开来。当  
 18 把这个插入的电子移走，即将这个结构对应到始态势能面，该结构在始态势能面上的电子态就  
 19 会非常奇怪。这个结构对应的电子态已经不是我们想要计算的始态结构的真实电子态，而是碎  
 20 片分子/离子对移除一个电子后对应的电子态。所以，为了保证计算的准确性，在计算氧化还原  
 21 电位的时候，对于发生自发解离的产物（终态势能面），同样需要添加一个谐振子势能将其键  
 22 长固定在产物结构的平衡位置附近。而这一计算方案的合理性在于，对于电极表面的氧化还原  
 23 反应，其与电位相关的部分为电子转移的自由能。而后续发生自发解离的过程对氧化还原电位  
 24 并没有直接贡献。

1  
2

## 2.2 机器学习势函数和机器学习分子动力学



3  
4

图 1 机器学习势函数示意图。 [36]

Figure 1 Systematic illustration of machine learning potential. [36]

6

7 由于 AIMD 所需要的计算资源比较高，同时对于计算复杂体系，经典分子动力学模拟  
8 (Classical MD) 无法满足要求。所以，MLMD 成为代替 AIMD 的一种主要手段。MLP 是通过  
9 构建原子局域环境的描述符，通过机器学习的方法，构建原子局域环境（图 1）到原子受力和原  
10 子能量一对一映射。原子局域环境描述符是一种将原子结构转化为高维数组的一种数学操作。  
11 机器学习势函数的构建基于以下两个假设：1) 体系的势能可以分解成每个原子能量  $E^i(R^i)$  的加  
12 和 [21-30]，即

$$E = \sum_i^N E^i(R^i)。 \quad 10$$

13 2) 每个原子能量和原子受力只与该原子周围很小范围内的原子相关，即只与该原子的局域原子  
14 环境相关。所以，只需要构建原子的局域环境到原子受力和原子能量的一对一映射，就可以通  
15 过机器学习的方法预测出准确的原子受力和体系的总能量。机器学习势函数则是通过机器学习  
16 的方法将原子局域环境描述符映射到体系总能量和原子受力等计算性质的方法。目前常用的几  
17 种机器学习势函数有高维神经网络势能 (High Dimensional Neural Network Potential, HDNNP)  
18 [23-24]、高斯近似势能 (Gaussian Approximation Potential, GAP) [25] 和深度势能 (Deep Potential)  
19 [27] 等。机器学习势函数只需要少量代表性的 DFT 计算数据，就可以很好地重复出 DFT 的势能面，  
20 从而加速分子动力学的模拟。通过机器学习势函数加速，已经可以在亿原子尺度实现第一性原

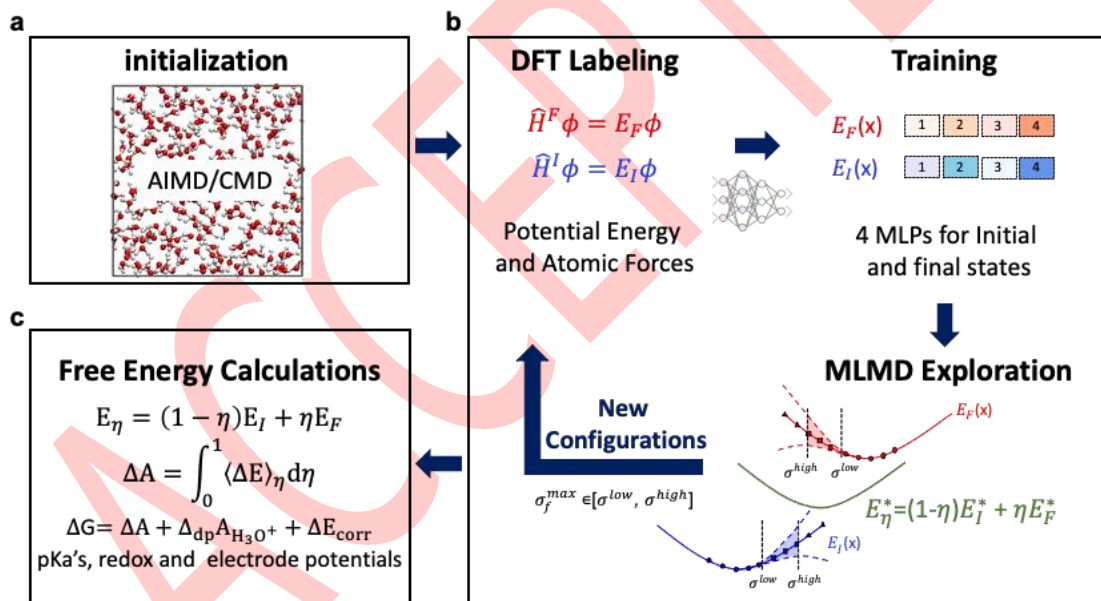
1 理精度的分子动力学模拟。收集具有代表性的数据集是保证机器学习势函数稳定的关键。然而，  
 2 构建数据集则是训练势函数过程中最具挑战的一步。

3

### 4 2.3 自动化 workflow

5

6 由于机器学习势函数是从原子结构到对应的势能和原子受力的一对一映射<sup>[21-30]</sup>，所以我们在  
 7 在训练势函数的过程中，将体系的始态和终态势能面分开来训练，这样两个单一映射就可以用于  
 8 于计算一个结构对应的垂直能量差。同时，为了提升数据集的采样效率和保证势函数的准确性，  
 9 我们引入了一个并行学习（concurrent learning）自动化 workflow<sup>[31-33]</sup>来生成在势能面上均一分布的  
 10 数据集（图 2）。workflow 分成三个部分：（a）初始化，（b）势函数训练和（c）自由能计算。  
 11 初始化 workflow，确定研究体系并通过经典力场分子动力学或者是一小段 AIMD 获得一部分初始  
 12 结构。



13

14 图 2 自动化 workflow 的图示。(a) 初始化；(b) 势函数训练；(c) 自由能计算。<sup>[31]</sup>

15 **Figure 2** Schematic diagram of the automated workflow. (a) Initialization. (b) Training machine learning potential. (c)  
 16 Free energy calculations.

17

18 势函数训练，使用初始化获得的初始结构，分别计算这些结构在始态势能面和终态势能面  
 19 上对应的势能和原子受力，分别构成始态势函数和终态势函数的数据集。使用这两个数据集分  
 20 别使用不同的神经网络初始化参数训练 4 个势函数，这个分别得到 4 个始态势函数和 4 个终态  
 21 势函数。由于使用不同的神经网络初始化参数，对于和数据集中结构差异较大的新结构，4 个势

1 函数预测出来的能量和原子受力会有较大差别，这将用来挑选新结构更新数据集。分别使用一  
 2 个始态势函数和一个终态势函数构建采样势能面 $E_\eta$ ，并同时使用多个 $\eta$ 保证采样势能面能够覆盖  
 3 始态、终态和中间态的所有区域。然后，在采样势能面上进行 MLMD 模拟，生成一系列轨迹。  
 4 然后，分别使用 4 个始态势函数和 4 个终态势函数预测 MLMD 轨迹上结构的原子受力，并计算  
 5 该结构中最大的原子受力的标准差 $\sigma_f^{max} = \max_i \sqrt{\langle ||f_i - \bar{f}_i||^2 \rangle}$ , ( $\bar{f}_i = \langle f_i \rangle$ )<sup>[33]</sup>，其中 i 是一个结构  
 6 中的某个原子的索引。然后，一部分 $\sigma_f^{max}$ 在预设范围( $\sigma_f^{low}$ ,  $\sigma_f^{high}$ ]-之间的结构被随机挑选出来，  
 7 并使用 DFT 计算其对应的势能和原子受力，构成这一轮训练得到的数据集。将这部分新的数据  
 8 集和初始数据集合并，分别再训练 4 个始态和 4 个终态势函数，并开启下一轮 MLMD 采样。随  
 9 着添加的数据集逐渐增多，其对应的势函数可以越来越好得表示始态和终态势能面，  
 10 ( $\sigma_f^{low}$ ,  $\sigma_f^{high}$ ]-区间内的结构会越来越少。当几乎没有新的结构时，表明这时的数据集已经能完美  
 11 描述这个反应的势能面，迭代采样的过程就结束了。然后，使用最终获得的数据集分别训练出  
 12 始态和终态对应的势函数，并计算对应反应的自由能变，将其转化为对应的物理化学性质。

## 2.4 计算设置

16 DFT 计算使用开源的 CP2K 软件<sup>[37]</sup>。计算模拟体系为 63 个水分子和 1 个溶质分子的三维周  
 17 期性盒子。所有的 DFT 计算都基于 BLYP 泛函<sup>[38-39]</sup>和 Grimme D3 色散校正<sup>[40]</sup>。核电子通过  
 18 GTH 赝势描述<sup>[41]</sup>，价电子通过 TZV2P-GTH 基组描述<sup>[42-43]</sup>。深度势能模型 (DeepPot-SE, deep  
 19 potential-smooth edition) <sup>[44]</sup>用于训练机器学习势函数。在分子动力学采样的过程中，Nose-  
 20 Hoover 温标用于执行 NVT 系综，其温度设置为 330, 430, 530 和 630K，步长设置为 0.5 fs。为

21 挑选新的结构,	$\sigma_f^{low}$	0.12 eV/Å	$\sigma_f^{high}$	0.30 eV/Å	MLMD
22 LAMMPS			1 ns		500 ps
23 Nose-Hoover	NVT	330 K	MLP	0.5 ps	DFT

24 表 1 为使用自动化 workflow 得到的势函数的数据集和训练误差的基本信息。对于大部分体系，  
 25 只需要少量的结构就可以构建准确的势函数。对于所有体系，其能量的误差都小于 1meV/atom，  
 26 原子受力的误差小于 60 meV/Å，此误差小于 DFT 计算误差，即势函数可以准确预测 DFT 的计算  
 27 结果。



1 表 1 通过自动化工作流获得的数据集大小和势函数对训练集的预测误差。<sup>[31]</sup>

2 **Table 1.** Datasets for MLP obtained through the automated workflow. The size of datasets training errors of MLPs are  
3 given for demonstration.

	$N_{\text{ini}}, N_{\text{fin}}$	RMSE <sub>E</sub> (meV/atom)	RMSE <sub>F</sub> (meV/Å)
HCl → Cl <sup>-</sup>	800, 800	0.553, 0.615	42.8, 44.2
H <sub>3</sub> O <sup>+</sup> → H <sub>2</sub> O	6016, 5234	0.653, 0.809	47.5, 50.8
H <sub>2</sub> S → HS <sup>-</sup>	835, 835	0.580, 0.700	44.0, 45.7
H <sub>2</sub> O → OH <sup>-</sup>	11048, 10648	0.556, 0.579	44.0, 43.4
HS <sup>-</sup> → S <sup>2-</sup>	850, 827	0.544, 0.551	43.4, 45.5
Cl <sup>-</sup> → Cl <sup>·</sup>	875, 873	0.532, 0.586	44.7, 48.1
OH <sup>-</sup> → OH <sup>·</sup>	1996, 1996	0.797, 0.706	51.6, 49.5
HS <sup>-</sup> → HS <sup>·</sup>	887, 887	0.722, 0.604	47.7, 47.8
O <sub>2</sub> <sup>-</sup> → O <sub>2</sub>	5499, 5499	0.819, 0.753	53.5, 52.1
CO <sub>2</sub> <sup>-</sup> → CO <sub>2</sub>	5791, 5791	0.733, 0.603	57.1, 47.7

4

5

## 6 3 结果与讨论

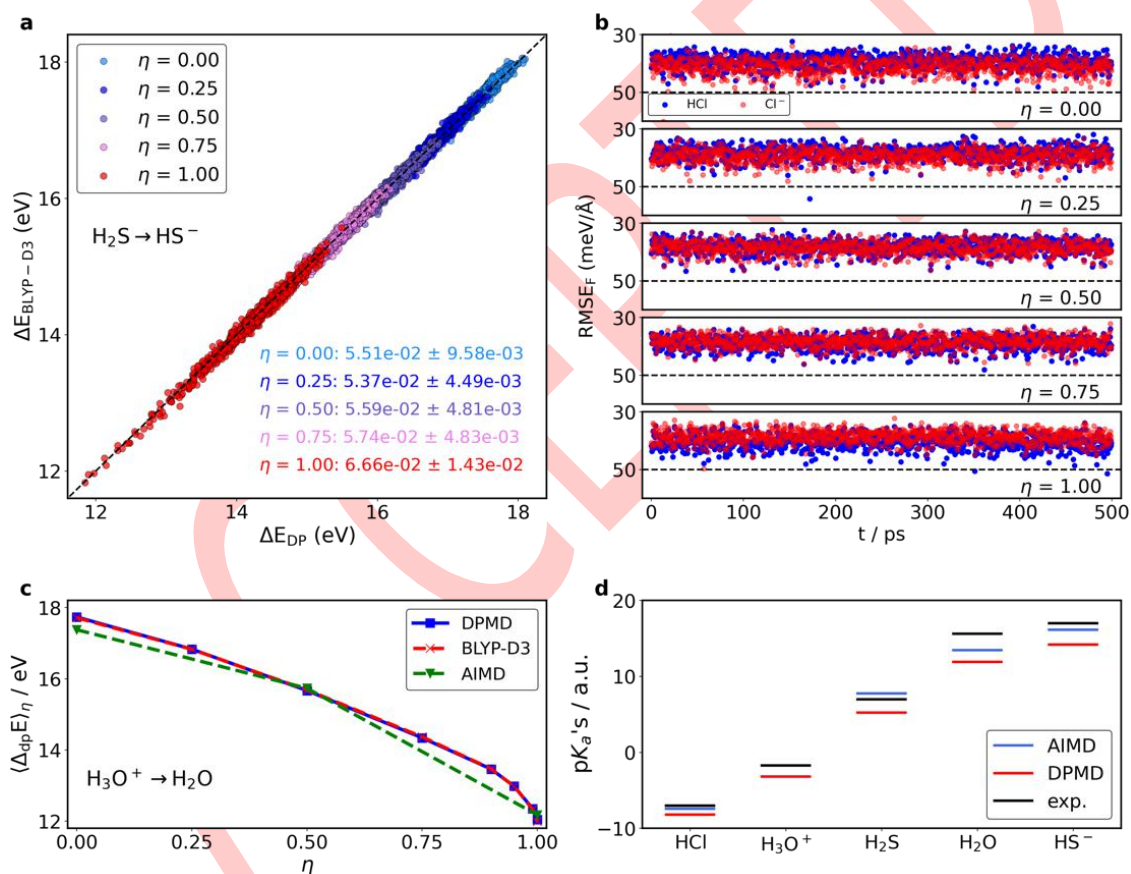
### 7 3.1 酸度常数

8 酸度常数的反应方程式为:  $\text{XH}(\text{aq.}) \rightarrow \text{X}^-(\text{aq.}) + \text{H}^+(\text{aq.})$ 。这个反应可以分解为两个半反  
9 应: (a)  $\text{XH}(\text{aq.}) \rightarrow \text{X}^-(\text{aq.}) + \text{H}^+(\text{g.})$ 和(b)  $\text{H}^+(\text{g.}) \rightarrow \text{H}^+(\text{aq.})$ 。所以计算 XH 的酸度常数需要计算  
10 XH 的脱质子自由能  $\Delta_{\text{dp}}A_{\text{XH}}$ 和质子的溶剂化自由能。根据之前发展的方法, 将质子的溶剂化自由  
11 能近似为负的  $\text{H}_3\text{O}^+$ 的脱氢自由能  $-\Delta_{\text{dp}}A_{\text{H}_3\text{O}^+}$ 。同时, 由于 XH 脱质子后体系缺少了一个质子,  
12 所以需要体系的平动熵进行校正, 即  $k_B T \ln [c^\circ \Lambda^3]$ , 其数值为 -0.19 eV。所以酸度常数  $\text{p}K_a$ 对  
13 应的自由能为:

$$2.303 k_B T \text{p}K_a = \Delta_{\text{dp}}A_{\text{XH}} - \Delta_{\text{dp}}A_{\text{H}_3\text{O}^+} + k_B T \ln [c^\circ \Lambda^3] \quad [14] \quad 11$$

14 基于这个计算公式, 我们使用 MLMD 模拟计算了  $\text{H}_2\text{S}$ 、 $\text{HS}^-$ 、 $\text{HCl}$ 、 $\text{H}_2\text{O}$  和  $\text{H}_3\text{O}^+$ 的酸度常  
15 数。如图 3 (a) 所示, 对于脱质子自由能的计算, 每个  $\eta$ 对应的垂直能量差的 RMSE 在 50 到 70  
16 meV 之间。原子受力的 RMSE 随模拟时间无太大变化, 如图 3(b)所示。以  $\Delta_{\text{dp}}A_{\text{H}_3\text{O}^+}$ 为例, 图 3(c)  
17 将 MLMD、AIMD 和 DFT 验证的结果放在一起比较, 我们可以发现 DFT 验证的结果和 MLMD  
18 计算的结果相同, 即 MLMD 可以准确预测 DFT 的计算结果。同时和 AIMD 比较可以发现,

1 MLMD 可以使用更多的垂直能量差进行热力学积分计算，最终结果更加准确。如图 3(d)所示，  
 2 通过 MLMD 计算的  $pK_a$ 's 和实验值以及 AIMD 计算的结果放在一起做比较，可以看出 MLMD  
 3 计算结果的平均误差 (mean signed error, MSE) 为-2.2 (-0.14 eV)，这比 AIMD 计算结果的误  
 4 差大了 0.1 eV (-0.8  $pK_a$ , -0.05 eV)。同时 MLMD 计算结果存在明显的负的系统误差，即计算  
 5 结果全部出现大小相近的负偏差。这个负的偏差来源于  $pK_a$ 's 的计算公式，且具体表现在  $H_3O^+$   
 6 的  $pK_a$  上。因为我们使用  $\Delta_{dp}A_{H_3O^+}$  作为参考积分，所以  $H_3O^+$  的  $pK_a$  的误差只取决于校正项  
 7  $k_B T \ln [c^0 \Lambda^3]$ ，这个误差为-1.5  $pK_a$  (-0.09 eV)。如果将这个误差从 MLMD 的计算结果中扣除  
 8 (如表 2, MLMD\*)，发现其 MSE 为 0.7  $pK_a$  (0.04 eV)，而这个误差小于势函数训练的误差。  
 9



10  
 11 **图 3** (a)  $H_2S \rightarrow HS^-$ 体系， $\Delta_{dp}E$  的准确性分析。(b)  $HCl \rightarrow Cl^-$ 体系中每个 $\eta$ 预测原子受力的 RMSE 随模拟时间  
 12 演变情况。(c)  $H_3O^+ \rightarrow H_2O$  体系 AIMD、MLMD 和基于 MLMD 轨迹的 DFT 计算得到的热力学积分曲线的对  
 13 比。(d) 酸度常数的 AIMD 和 MLMD 的计算结果以及实验值。

14 **Figure 3** (a) Accuracy analysis of  $\Delta_{dp}E$  for  $H_2S \rightarrow HS^-$ . (b) Evolution of the RMSE of each  $\eta$  predicted atomic force  
 15 in the  $HCl \rightarrow Cl^-$  system with simulation time. (c) Comparison of the TI-path from AIMD, MLMD, and DFT  
 16 validation for  $H_3O^+ \rightarrow H_2O$ . (d) Acidity constants from AIMD, MLMD and experimental values.

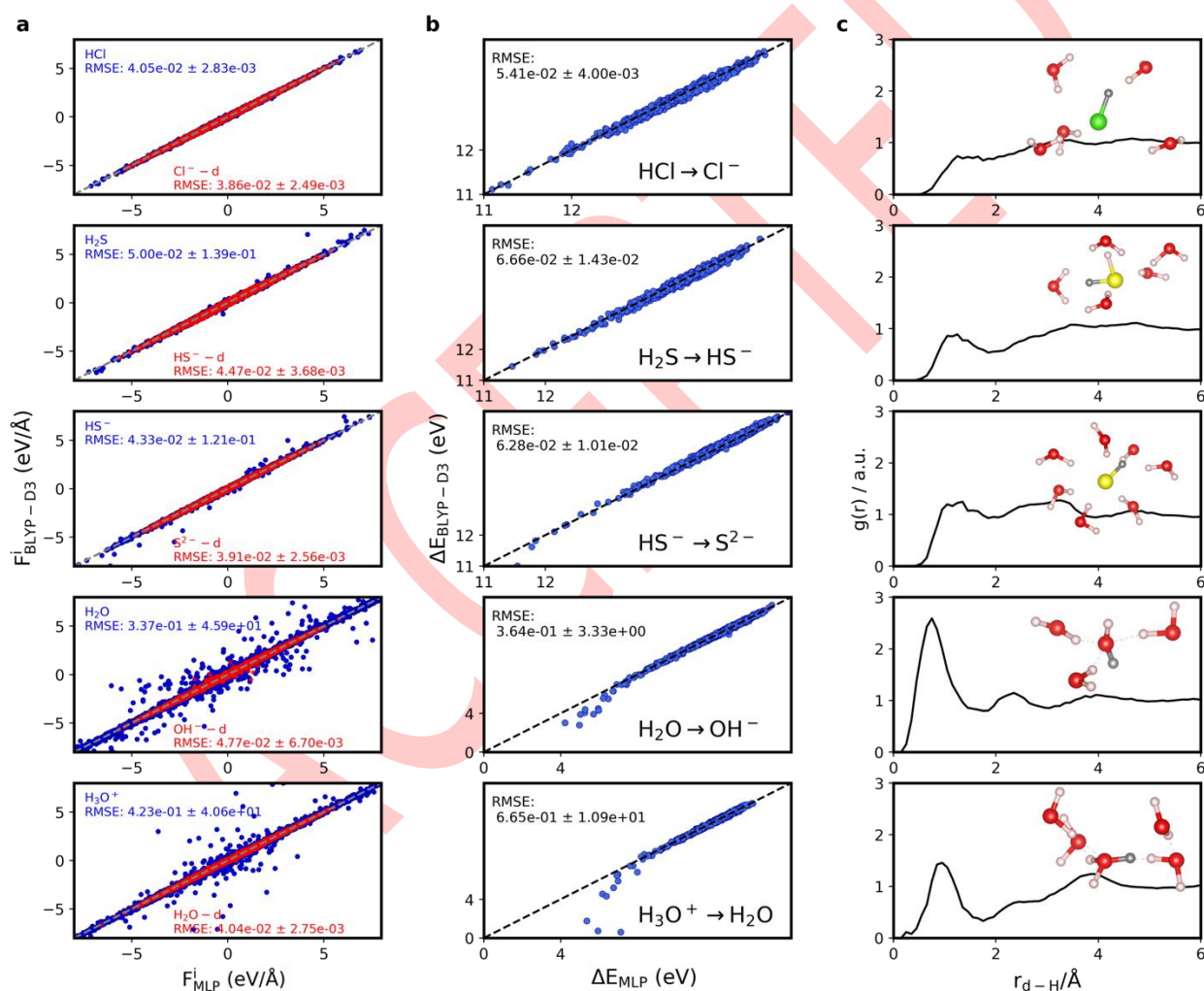
17  
 18

1 表 2 通过基于 MLMD 模拟得到的酸度常数。[31]

2 **Table 2.** Acidity constants obtained through MLMD simulations.

	AIMD	MLMD	MLMD *	exp.
HCl	-7.0	-8.2 (-8.8)	-6.7 (-7.3)	-7.0
H <sub>3</sub> O <sup>+</sup>	-3.2	-3.2 (-3.2)	-1.7 (-1.7)	-1.7
H <sub>2</sub> S	7.8	5.2 (5.0)	6.7 (6.5)	7.0
H <sub>2</sub> O	13.5	11.9 (11.6)	13.4 (13.1)	15.6
H <sub>2</sub> S	16.2	14.2 (14.5)	15.7 (16.0)	17.0

3



4

5 图 4 不同体系脱质子自由能计算中  $\eta = 1.00$  的误差分析。(a) 原子受力误差分析。(b) 垂直能量差  $\Delta E_{\eta=1.0}$  的误  
6 差分析。(c) 虚原子和水分子上氢原子的径向分布函数。[31]

7 **Figure 4.** Errors in  $\eta = 1.00$ . (a) Errors in atomic force. (b) Error in  $\Delta E_{\eta=1.0}$ . (c) Radial distribution functions of the  
8 dummy atom and hydrogen atoms in water molecules.

9

1 对于所有的 MLMD 模拟，MLP 展现相对于 DFT 验证的高准确性。虽然对于几条  $\eta = 1.00$   
 2 的 MD 模拟（如图 4 (a),  $\text{H}_2\text{O}$  和  $\text{H}_3\text{O}^+$ ），始态势函数对原子受力和能量的 RMSE 比其他的 MD  
 3 大了约 10 倍，但是这些误差对热力学积分的计算结果影响不大。这也导致了数据集大小明显大  
 4 于其他数据集。其原因在于，首先，对于  $\eta = 1.00$  由于  $E_\eta = E_F$ ，MD 是由终态势函数驱动的，  
 5 始态势函数在这个点的误差对 MD 模拟没有影响。同时，如图 4 (b),  $\Delta E$  的离群点少于 1%且对  
 6 整体的统计平均影响较小。这个误差的主要来源是，对于  $\eta = 1.00$  时，虚原子和其他原子没有  
 7 相互作用，由于 O 原子和水分子上的 H 原子形成氢键，所以导致了虚原子和水分子上 H 的距离  
 8 非常短（图 4 (c)）。当虚原子被重新打开成为质子（在始态势能面），这时候就出现了非常大  
 9 的力，而势函数对这种巨大的变化拟合存在较大的误差。[45]

10

### 11 3.2 氧化还原电位

12 氧化还原电位的计算需要参比到计算氢标电极上（computational standard hydrogen electrode,  
 13 cSHE）[14]。这存在两个原因：其一是计算得到的  $\Delta_{ox}A$  由于周期性边界条件下静电势能的参考零  
 14 点存在不确定性，所以带静电荷体系的能量存在不确定性，即 Hartree potential shift (HPS)。  
 15 所以计算求得的  $\Delta_{ox}A$  没有实际的物理意义，只有通过将这个自由能参考到一个具有相同 HPS 的  
 16 自由能才可以将这个误差抹去。其二在于氢标电极 (SHE) 是实验中常用的电极，将计算的氧  
 17 化还原电位参考到 SHE 有助于计算数值和实验值的直接比较。首先，氧化还原半反应的表达式  
 18 为  $X^-(\text{aq.}) + e^-(\text{vac.}) \rightarrow X(\text{aq.})$ ，其对应的自由能为  $-\Delta_{ox}A_{X^-}$ 。其次，SHE 半反应的表达式为  
 19  $\text{H}^+(\text{aq.}) + e^-(\text{vac.}) \rightarrow \frac{1}{2}\text{H}_2(\text{g.})$ ，可以将其分解为两个半反应：(1)  $\text{H}^+(\text{aq.}) \rightarrow \text{H}^+(\text{g.})$ ，即质子的脱  
 20 溶剂化反应，其自由能根据  $\text{p}K_a$ 's 计算的推导可以使用  $\Delta_{dp}A_{\text{H}_3\text{O}^+}$  近似；(2)  $\text{H}^+(\text{g.}) + e^-(\text{vac.}) \rightarrow$   
 21  $\frac{1}{2}\text{H}_2(\text{g.})$ ，即从气态质子生成氢气，其自由能为负的  $\text{H}^+(\text{g.})$  标准吉布斯生成自由能  $-\mu_{\text{H}^+}^{g,0}$  (-15.81  
 22 eV)。同时，需要减去质子对应的零点能  $\Delta_{ZPE}$  (0.35 eV)。所以总反应的自由能为两个半反应  
 23 的自由能之差：

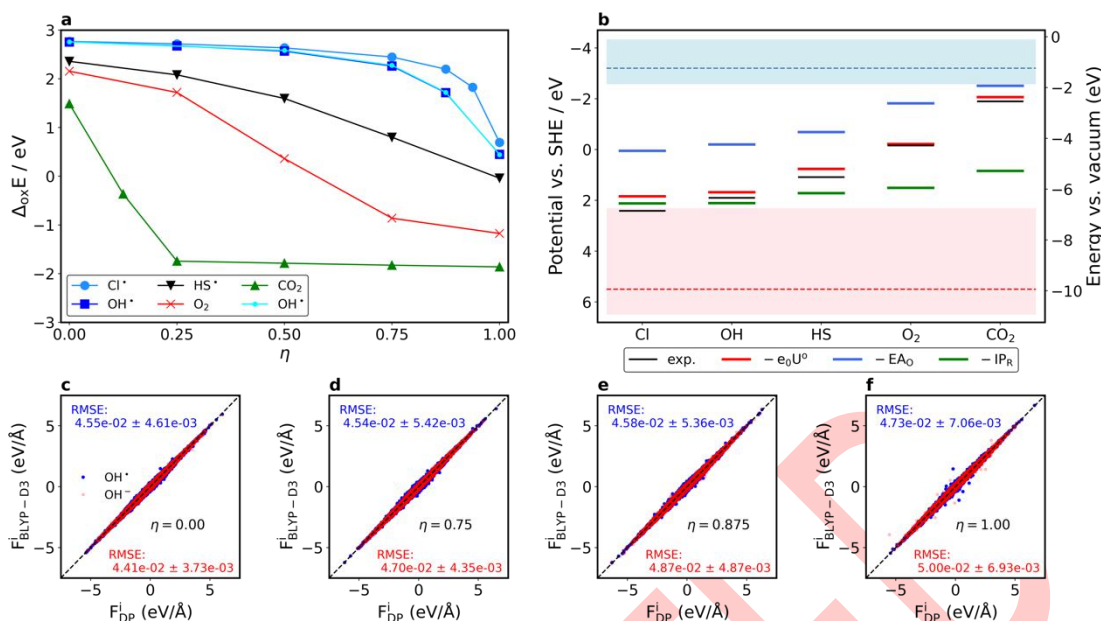
$$\Delta_{ox}G = -\Delta_{ox}A_{X^-} - \Delta_{dp}A_{\text{H}_3\text{O}^+} + \mu_{\text{H}^+}^{g,0} + \Delta_{ZPE} \quad 12$$

24 根据  $\Delta G = -e_0U$ ，可以得到氧化还原电位的具体表达式为：

$$U^0(\text{SHE}) = (\Delta_{ox}A + \Delta_{dp}A_{\text{H}_3\text{O}^+} - \mu_{\text{H}^+}^{g,0} - \Delta_{ZPE})/e_0 \quad [11] \quad 13$$

25 根据这个计算公式我们计算的 5 个电对的氧化还原电位，其计算结果如表 3 所示。

26



1  
2 **图 5** (a) 各个氧化还原自由能计算的热力学积分曲线。(b) MLMD 计算得到的氧化还原电位。(c-f) OH<sup>•</sup>/OH<sup>-</sup> 热  
3 力学积分上  $\eta = 0.00, 0.75, 0.875$  和  $1.00$  上 MLP 预测的原子受力和 DFT 验证计算的对比。<sup>[31]</sup>

4 **Figure 5.** (a) TI-path for redox free energy calculations. (b) Redox potentials obtained from MLMD calculations. (c-f)  
5 Comparison of MLP-predicted atomic forces and DFT validated calculations on the TI-path of OH<sup>•</sup>/OH<sup>-</sup> with  $\eta = 0.00,$   
6  $0.75, 0.875$  and  $1.00$ .

7  
8 为了获取准确的氧化还原电位，我们对于热力学积分线性较差体系添加额外的 $\eta$ 以确保计算  
9 值的准确性。同时，如图 5(a)，MLMD 计算得到的 OH<sup>-</sup>热力学积分和通过 DFT 验证的结果相差  
10 基本为 0，即对于氧化还原体系，MLMD 可以得到 DFT 精度的计算结果。同时从表 3 展示的积  
11 分值来看，AIMD 三个点的积分值和 MLMD 三个点的积分值结果相差较小，这部分的误差主要是  
12 是由于 AIMD 对于氧化还原体系的统计采样不够造成的。如图 5(a)，Cl<sup>•</sup>/Cl<sup>-</sup>和 OH<sup>•</sup>/OH<sup>-</sup>的热力  
13 学积分线性程度较差，所以在 0.75 和 1.00 之间插入了额外的点来收敛热力学积分。如图 5(c-f)  
14 所示，通过对 OH<sup>-</sup>积分路径上 $\eta = 0.00, 0.75, 0.875$  和  $1.00$  上原子受力的误差进行分析，我们发  
15 现 MLP 对 0.875 的结构预测的原子受力非常准确。由于在势函数训练的过程中没有直接包含  
16 0.875 这个中间态对应的结构，所以可以证明 MLP 对自由能计算体系展现出非常好的内插性。  
17 只使用 3 个点积分数值和收敛之后的积分数值之间的差距约为 0.1 eV，这个误差大于势函数训  
18 练的误差且对氧化还原电位的计算起到提升精度的作用。通过表 3 和图 5(b) 的对比不难发现，  
19 和 AIMD 相比，MLMD 计算的氧化还原电位误差下降了约 0.3 V。计算提升的原因主要是  
20 MLMD 对 $\Delta_{ox}A$  和 $\Delta_{dp}A_{H_3O^+}$ 数值积分收敛的提升。然而，对于 HS<sup>•</sup>/HS<sup>-</sup>、OH<sup>•</sup>/OH<sup>-</sup>和 Cl<sup>•</sup>/Cl<sup>-</sup>，氧

1 化还原电位的计算误差仍有 0.3-0.6 V。这个误差的主要来源是 BLYP 泛函的误差。这部分的误  
 2 差可以通过更高精度的泛函或者量子化学方法来解决<sup>[17]</sup>。

3

4 **表 3** 通过基于 MLMD 模拟得到的氧化还原电位。MLMD 计算使用的 $\Delta_{dp}A_{H_3O^+}$ 为 MLMD 计算出的 15.512 eV。  
 5 <sup>[31]</sup>

6 **Table 3** Redox potentials obtained through MLMD simulations. The  $\Delta_{dp}A_{H_3O^+}$  used in the MLMD calculations is  
 7 15.512 eV from MLMD.

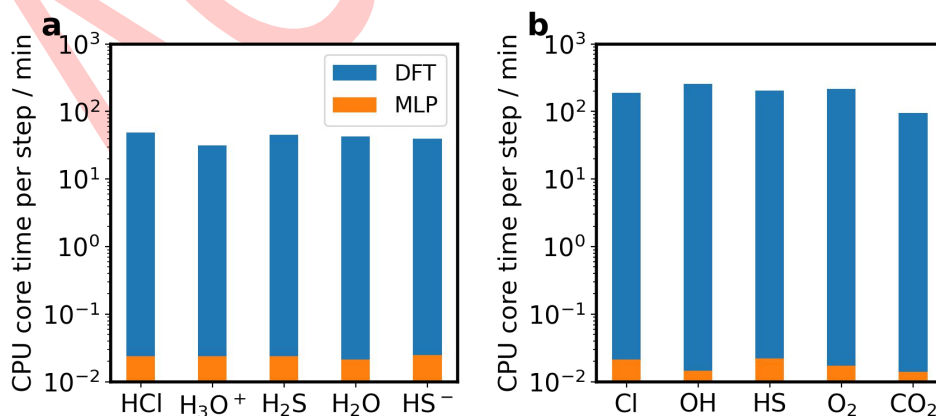
	AIMD	MLMD (3P)	MLMD	exp.
Cl <sup>+</sup> /Cl <sup>-</sup>	1.5	1.58	1.84	2.41
OH <sup>+</sup> /OH <sup>-</sup>	1.3	1.49 (1.51)	1.68 (1.68)	1.90
HS <sup>+</sup> /HS <sup>-</sup>	0.5	0.69	0.76	1.08
O <sub>2</sub> /O <sub>2</sub> <sup>-</sup>	-0.5	-0.35	-0.22	-0.16
CO <sub>2</sub> /CO <sub>2</sub> <sup>-</sup>	-2.07	-1.98	-2.06	-1.90

8

### 9 3.3 计算效率

10 如图 6 所示，通过对比 DFT 和 MLP 计算一个数据点所需的计算时间，不难发现机器学习势  
 11 函数的计算效率相对于 DFT 提升了 1000 以上。对于氧化还原电位这种难以收敛的体系，其计算  
 12 效率可以提升 10000 倍左右。同时需要指出的是，当计算使用的泛函换成更高阶的电子结构计  
 13 算方法，例如杂化泛函、random phase approximation、MP2 和 GW 等，只需要较小的数据集就  
 14 可以得到准确的机器学习势函数，而计算效率相对于这类方法可以提升数万倍。

15



16

17 **图 6** (a) 脱质子自由能计算和(b)氧化还原自由能计算中 DFT 和 MLP 计算效率对比。

18 **Figure 6.** Comparison of DFT and MLP calculation efficiency in (a) deprotonation free energy and  
 19 energy.

## 4 结论

本文介绍了如何通过自动化 workflows 构建机器学习势函数并通过机器学习分子动力学加速自由能计算。为了得到准确的机器学习势函数，需要迭代更新数据集，确保势函数可以准确预测反应路径上所有构象的能量和原子受力。进一步，通过机器学习分子动力学，可以实现纳秒级别的分子动力学采样，同时使用更密集的数据点计算热力学积分值，可以得到更准确的计算结果。同时需要指出的是，由于构建数据集只需要较小的数据量，所以这一自动化 workflow 使通过更准确的泛函计算对应的物理化学性质成为可能。同时，基于 AIMD 的自由能计算早已被应用于界面酸度常数和氧化还原电位的计算。而本文描述的方法基于这一方法实现机器学习加速，因此这一自动化 workflow 可以直接用于计算界面酸度常数和氧化还原电位，加速界面电化学研究。

## 参考文献(References):

- [1]. Blumberger J. Recent advances in the theory and molecular simulation of biological electron transfer reactions[J]. *Chem. Rev.*, 2015, 115(20): 11191-11238.
- [2]. Wardman P. Reduction potentials of one - electron couples involving free radicals in aqueous solution[J]. *J. Phys. Chem. Ref. Data*, 1989, 18(4): 1637-1755.
- [3]. Weinberg D R, Gagliardi C J, Hull J F, Murphy C F, Kent C A, Westlake B C, Meyer, T J. Proton-coupled electron transfer[J]. *Chem. Rev.*, 2012, 112(7): 4016-4093.
- [4]. Yu J, Zhao T S, Pan D. Tuning the Performance of Aqueous Organic Redox Flow Batteries via First-Principles Calculations[J]. *J. Phys. Chem. Lett.*, 2020, 11(24): 10433-10438.
- [5]. Tomasi J, Mennucci B, Cammi R. Quantum mechanical continuum solvation models[J]. *Chem. Rev.*, 2005, 105(8): 2999-3094.
- [6]. Leung K. Predicting the voltage dependence of interfacial electrochemical processes at lithium-intercalated graphite edge planes[J]. *Phys. Chem. Chem. Phys.*, 2015, 17(3): 1637-1643.
- [7]. Le J, Iannuzzi M, Cuesta A, Cheng, J. Determining potentials of zero charge of metal electrodes versus the standard hydrogen electrode from density-functional-theory-based molecular dynamics[J]. *Phys. Rev. Lett.*, 2017, 119(1): 016801.
- [8]. Rossmeisl J, Skúlason E, Björketun M E, Tripkovic V, Nørskov J K. Modeling the electrified solid-liquid interface[J]. *Chem. Phys. Lett.*, 2008, 466(1-3): 68-71.
- [9]. Zhang C, Sprik M. Finite field methods for the supercell modeling of charged insulator/electrolyte interfaces[J]. *Phys. Rev. B*, 2016, 94(24): 245309.
- [10]. King G, Warshel A. Investigation of the free energy functions for electron transfer reactions[J]. *J. Chem. Phys.*, 1990, 93(12): 8682-8692.
- [11]. Cheng J, Sulpizi M, Sprik M. Redox potentials and pKa for benzoquinone from density functional theory based molecular dynamics[J]. *J. Chem. Phys.*, 2009, 131(15): 154504.
- [12]. Blumberger J, Tavernelli I, Klein M L, et al. Diabatic free energy curves and coordination fluctuations for the aqueous  $\text{Ag}^+/\text{Ag}^{2+}$  redox couple: A biased Born-Oppenheimer molecular dynamics investigation[J]. *J. Chem. Phys.*, 2006, 124(6): 064507.
- [13]. Costanzo F, Sulpizi M, Valle R G D, Sprik M. The oxidation of tyrosine and tryptophan studied by a molecular dynamics normal hydrogen electrode[J]. *J. Chem. Phys.*, 2011, 134(24): 06B615.
- [14]. Cheng J, Liu X, VandeVondele J, Sulpizi M, Sprik M. Redox potentials and acidity constants from density functional theory based molecular dynamics[J]. *Acc. Chem. Res.*, 2014, 47(12): 3522-3529.
- [15]. Yang X H, Cuesta A, Cheng J. Computational Ag/AgCl reference electrode from density functional theory-based molecular dynamics[J]. *J. Phys. Chem. B*, 2019, 123(48): 10224-10232.
- [16]. Leung K, Tenney C M. Toward first principles prediction of voltage dependences of electrolyte/electrolyte interfacial processes in lithium ion batteries[J]. *J. Phys. Chem. C*, 2013, 117(46): 24224-24235.

- 1 [17]. Cheng J, VandeVondele J. Calculation of electrochemical energy levels in water using the random phase  
2 approximation and a double hybrid functional[J]. *Phys. Rev. Lett.*, 2016, 116(8): 086402.
- 3 [18]. Su N Q, Xu X. The XYG3 type of doubly hybrid density functionals[J]. *WIREs. Co. Mol. Sci.*, 2016, 6(6): 721-  
4 747.
- 5 [19]. Su N Q, Zhu Z, Xu X. Doubly hybrid density functionals that correctly describe both density and energy for  
6 atoms[J]. *P. Natl. Acad. Sci.*, 2018, 115(10): 2287-2292.
- 7 [20]. Zhang I Y, Wu J, Xu X. Accurate heats of formation of polycyclic saturated hydrocarbons predicted by using the  
8 XYG3 type of doubly hybrid functionals[J]. *J. Comput. Chem.*, 2019, 40(10): 1113-1122.
- 9 [21]. Thompson A P, Swiler L P, Trott C R, Foiles S M, Tucker G J. Spectral neighbor analysis method for automated  
10 generation of quantum-accurate interatomic potentials[J]. *J. Comput. Phys.*, 2015, 285: 316-330.
- 11 [22]. Huan T D, Batra R, Chapman J, Krishnan S, Chen L, Ramprasad R. A universal strategy for the creation of  
12 machine learning-based atomistic force fields[J]. *NPJ Comput. Mater.*, 2017, 3(1): 1-8.
- 13 [23]. Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy  
14 surfaces[J]. *Phys. Rev. Lett.*, 2007, 98(14): 146401.
- 15 [24]. Behler J. Perspective: Machine learning potentials for atomistic simulations[J]. *J. Chem. Phys.*, 2016, 145(17):  
16 170901.
- 17 [25]. Bartók A P, Payne M C, Kondor R, Csányi G. Gaussian approximation potentials: The accuracy of quantum  
18 mechanics, without the electrons[J]. *Phys. Rev. Lett.*, 2010, 104(13): 136403.
- 19 [26]. Rupp M, Tkatchenko A, Müller K R, Von Lilienfeld O A. Fast and accurate modeling of molecular atomization  
20 energies with machine learning[J]. *Phys. Rev. Lett.*, 2012, 108(5): 058301.
- 21 [27]. Zhang L, Han J, Wang H, Car R, Weinan E. Deep potential molecular dynamics: a scalable model with the  
22 accuracy of quantum mechanics[J]. *Phys. Rev. Lett.*, 2018, 120(14): 143001.
- 23 [28]. Wang H, Zhang L, Han J, Weinan E. DeePMD-kit: A deep learning package for many-body potential energy  
24 representation and molecular dynamics[J]. *Comput. Phys. Commun.*, 2018, 228: 178-184.
- 25 [29]. Chmiela S, Tkatchenko A, Sauceda H E, Poltavsky I, Schütt K T, Müller K R. Machine learning of accurate  
26 energy-conserving molecular force fields[J]. *Sci. Adv.*, 2017, 3(5): e1603015.
- 27 [30]. Schütt K T, Arbabzadah F, Chmiela S, Müller K R, Tkatchenko A. Quantum-chemical insights from deep tensor  
28 neural networks[J]. *Nat. Commun.*, 2017, 8(1): 1-8.
- 29 [31]. Wang F, Cheng J. Automated workflow for computation of redox potentials, acidity constants, and solvation free  
30 energies accelerated by machine learning[J]. *J. Chem. Phys.* 2022; 157 (2): 024103
- 31 [32]. Zhang L, Lin D Y, Wang H, Car R, Weinan E. Active learning of uniformly accurate interatomic potentials for  
32 materials simulation[J]. *Phys. Rev. Mater.*, 2019, 3(2): 023804.
- 33 [33]. Zhang Y, Wang H, Chen W, Zeng J, Zhang L, Wang H, Weinan E. DP-GEN: A concurrent learning platform for  
34 the generation of reliable deep learning based potential energy models[J]. *Comput. Phys. Commun.*, 2020, 253:  
35 107206.
- 36 [34]. Kirkwood J G. Statistical mechanics of fluid mixtures[J]. *The Journal of Chemical Physics*, 1935, 3(5): 300-313.
- 37 [35]. Sulpizi M, Sprik M. Acidity constants from vertical energy gaps: density functional theory based molecular  
38 dynamics implementation[J]. *Phys. Chem. Chem. Phys.*, 2008, 10(34): 5238-5249.
- 39 [36]. Sun Y, Wu C R, Wang F, Bi R H, Zhuang Y B, Liu S, Cheng J. Step-induced double-row pattern of interfacial  
40 water on rutile TiO<sub>2</sub> (110) at electrochemical conditions[J]. 2023.
- 41 [37]. VandeVondele J, Krack M, Mohamed F, Parrinello M, Chassaing T, Hutter J. Quickstep: Fast and accurate  
42 density functional calculations using a mixed Gaussian and plane waves approach[J]. *Comput. Phys. Commun.*,  
43 2005, 167(2): 103-128.
- 44 [38]. Becke A D. Density-functional exchange-energy approximation with correct asymptotic behavior[J]. *Phys. Rev.*  
45 *A*, 1988, 38(6): 3098.
- 46 [39]. Lee C, Yang W, Parr R G. Development of the Colle-Salvetti correlation-energy formula into a functional of the  
47 electron density[J]. *Phys. Rev. B*, 1988, 37(2): 785.
- 48 [40]. Grimme S, Antony J, Ehrlich S & Krieg H. A consistent and accurate ab initio parametrization of density  
49 functional dispersion correction (DFT-D) for the 94 elements H-Pu[J]. *J. Chem. Phys.*, 2010, 132(15): 154104.
- 50 [41]. Hartwigsen C, Goedecker S, Hutter J. Relativistic separable dual-space Gaussian pseudopotentials from H to  
51 Rn[J]. *Phys. Rev. B*, 1998, 58(7): 3641.
- 52 [42]. Goedecker S, Teter M, Hutter J. Separable dual-space Gaussian pseudopotentials[J]. *Phys. Rev. B*, 1996, 54(3):  
53 1703.
- 54 [43]. VandeVondele J, Hutter J. Gaussian basis sets for accurate calculations on molecular systems in gas and  
55 condensed phases[J]. *J. Chem. Phys.*, 2007, 127(11): 114105.
- 56 [44]. Zhang L, Han J, Wang H, Saidi W & Car R. End-to-end symmetry preserving inter-atomic potential energy  
57 model for finite and extended systems[J]. *Adv. Neural Inf. Process. Syst.*, 2018, 31.



1 [45]. Sulpizi M, Sprik M. Acidity constants from vertical energy gaps: density functional theory based molecular  
2 dynamics implementation[J]. Phys. Chem. Chem. Phys., 2008, 10(34): 5238-5249.  
3

ACCEPTED

# Automated Workflow for Redox Potentials and Acidity Constants Calculation from Machine Learning Molecular Dynamics

Feng Wang<sup>a</sup>, Jun Cheng<sup>a,b,c\*</sup>

(a. State Key Laboratory of Physical Chemistry of Solid Surfaces, b. Collaborative Innovation Center of Chemistry for Energy Materials, Department of Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, c. Tan Kah Kee Innovation Laboratory, Xiamen 361005, Fujian, China)

**Abstract:** Redox potentials and acidity constants are key properties for evaluating the performance of energy materials.

To achieve computational design of new generation of energy materials with higher performances, computing redox potentials and acidity constants with computational chemistry have attracted lots of attention. However, lots of works are done by using implicit solvation models, which is difficult to applied to complex solvation environments due to hard parameterization. Recently, ab initio molecular dynamics (AIMD) has been applied to investigate real electrolytes with complex solvation. Furthermore, AIMD based free energy calculation method have been established to calculate these physical chemical properties accurately. However, due to the low efficiency of ab initio calculations and the high computational costs, AIMD based free energy calculation are limited to systems with less than 1000 atoms. To solve the dilemma, machine learning molecular dynamics (MLMD) is introduced to accelerate the calculations. By using machine learning method to construct one-to-one mapping from structures to computed potential energies and atomic forces, molecular dynamics can be carried out with much low costs under ab initio accuracy. In order to achieve the MLMD based free energy calculation, a new scheme for machine learning potential (MLP) should be introduced to collect training datasets. By combining the free energy perturbation sampling method and concurrent learning scheme, the training datasets can be collected along the reaction's pathway (insertion of an electron or a proton) with high efficiency and the free energy calculation based on MLMD show good accuracy in comparison with AIMD simulation. This paper describes how to constructing machine learning potential for free energy calculation through the automated workflow and how to use MLMD to compute accurate free energy differences and corresponding physical chemical properties.

**Key words:** machine learning molecular dynamics, automated workflow, complex systems, free energy calculation

## Entry for the Table of Contents

((Insert graphic for Table of Contents here (300 DPI resolution: up to 1240x1180 pixels with a maximum width 10.5 cm and height 10 cm).))

