Charge Transfer Landscape Manifesting Structure-Rate Relationship in the Condensed Phase via Machine Learning

Dominikus Brian

Division of Arts and Sciences, NYU Shanghai
NYU-ECNU Center for Computational Chemistry at NYU Shanghai
Department of Chemistry, New York University

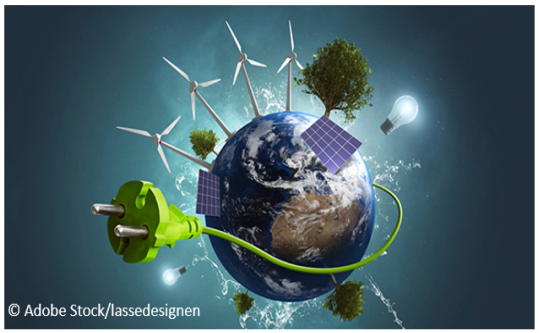
VISTA Seminar 43, October 19, 2022

Outline

Outline

Global Energy Transition and Sustainable Development

Global Energy Transition & Sustainable Development

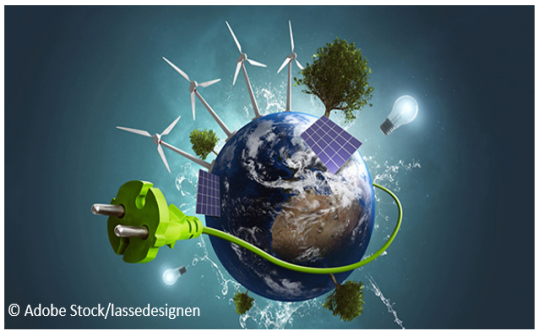


Energy

- Harvesting
- Storage
- Distribution

Global Energy Transition and Sustainable Development

Global Energy Transition & Sustainable Development



Energy

- **Harvesting**
- **Storage**
- **Distribution**

Solution-Processed OPV Device for Large-Scale Energy Harvesting

Organic Photovoltaics (OPV) : From lab to the city

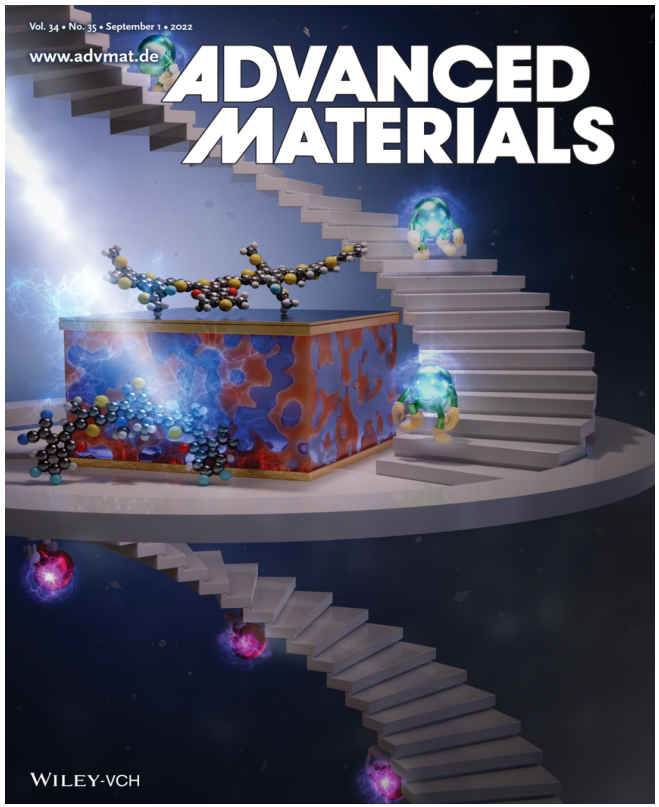


Figure 1: Solution Processed Bulk-Heterojunction OPV. Image: Adv. Mater. 2022, 34, 2202575.

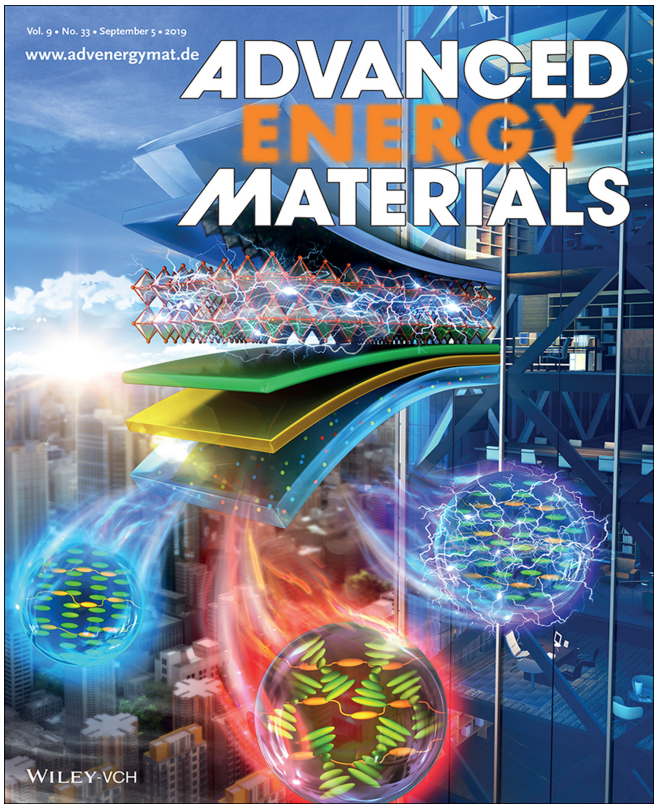
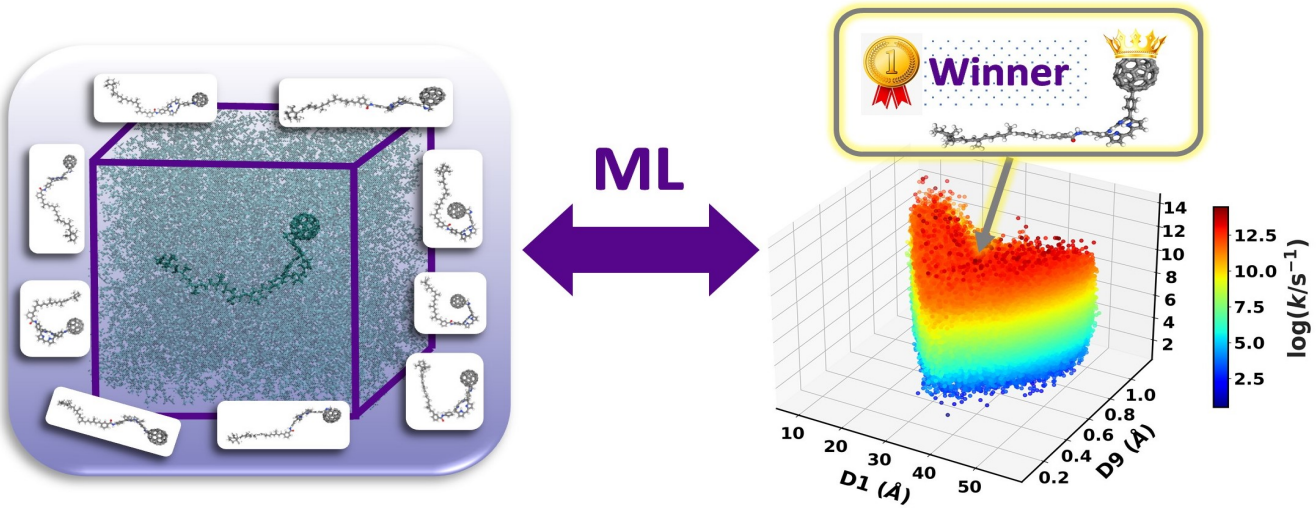


Figure 2: Solar Cells for Energy Harvesting Window. Image: Adv. Energy Mater. 2019, 9, 1900720.

Objectives: Charge Transfer (CT) Landscape Construction and Exploration.

CT Landscape $\approx$ Conformations $\Leftrightarrow$ CT Properties



Dominikus Brian and Xiang Sun, JPCB, 125, 13267-13278, (2021).

1. **Construct** : Construct Conformations-CT properties ML models to bypass exhaustive and expensive calculations. Exploit the ML models to interpolate for a larger conformations-CT properties database. [5 Steps]
2. **Explore** : Perform exploration and analysis to obtain molecular insights. [3 Examples]

Subject of the Study

Molecular triad (Fullerene - Porphyrin - Carotenoid) in Polar Organic Solvent (THF). Chosen for the following reason:

1. Preliminary knowledge of the influence of conformation on the calculated CT properties.
2. Accessible literature for the multiscale simulation (quantum chemistry and molecular dynamics) aspect of the system.
3. Availability of experimental kinetics data in condensed phase.

Outline

Charge Transfer Landscape

Formal Definition

Charge Transfer (CT) Landscape : Describe the full conformation space of a molecule or interacting molecular systems that manifest mapping of the possible states to their charge transfer properties, such as electron transfer kinetic rate, activation energies, diabatic coupling, reorganization energies, and other relevant intermediate molecular properties, a collection of which can be called charge transfer fingerprints (CTFP).

Informal Definition

Charge Transfer (CT) Landscape : A kind of energy landscape for charge transfer properties.

CT Landscape Strategy

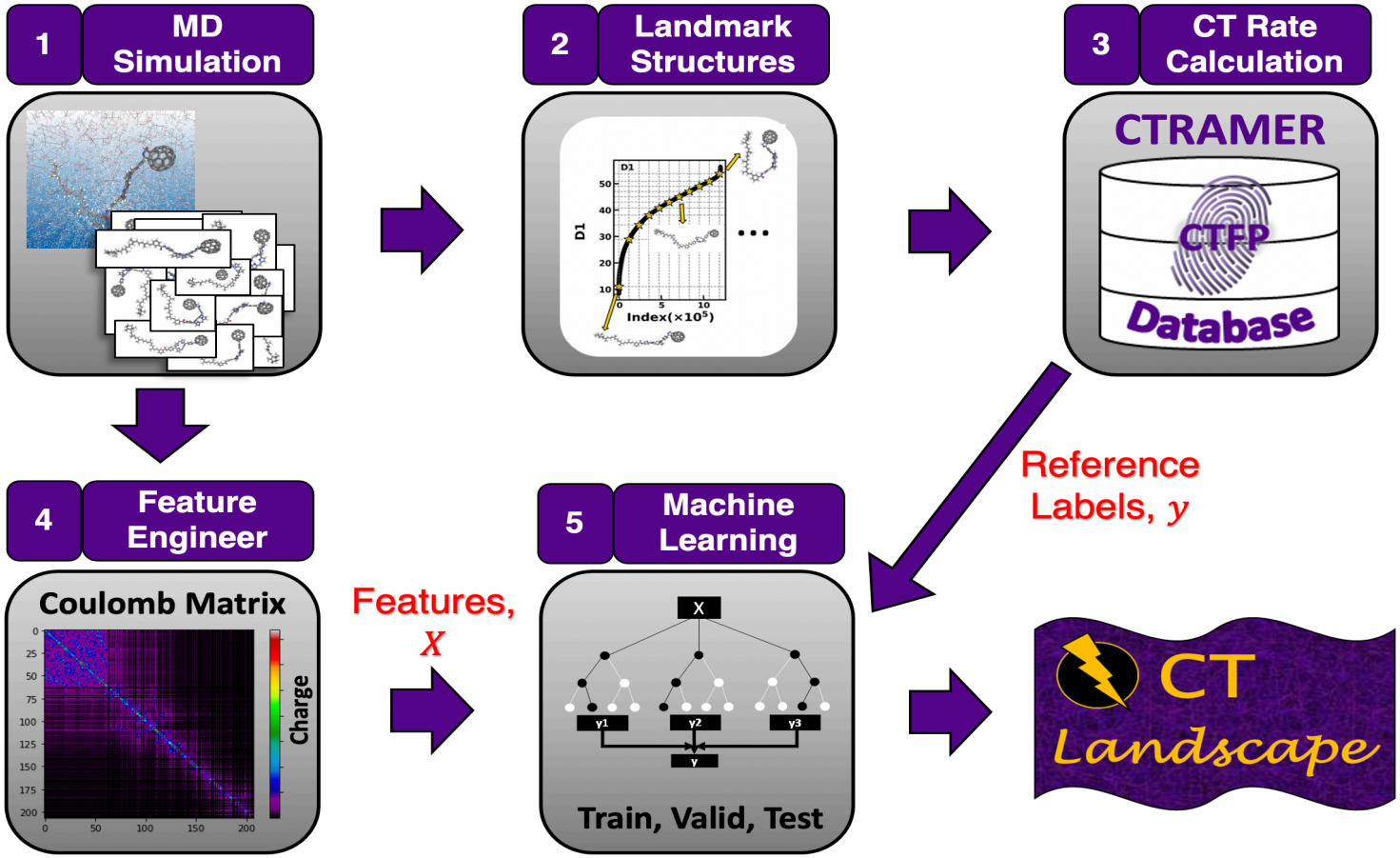


Figure 3: Schematic illustration of the machine learning model construction for charge transfer (CT) landscape of organic photovoltaic molecule in condensed phase.

CT Landscape Strategy

The conundrum for finding optimal combinations of methodology and protocols.

1. MD Simulations : Force Field, Enhanced Sampling, Simulations Parameters and Steps.
2. Landmark Structures : Number of samples, Importance Sampling, Selection Criteria.
3. CT Rate Calculation : Quantum Chemistry, Rate Theory, Intermediate Properties.
4. Feature Engineer : Feature/spatial Representation and Dimension Reduction.
5. Machine Learning : ML algorithm, Training Strategy/Protocol, Hyperparameters Optimization.

Outline

CT Landscape Construction

5 Steps of Charge Transfer Landscape Construction

1. MD Simulations for Conformational Sampling

Enhanced Sampling vs Long Duration MD simulations

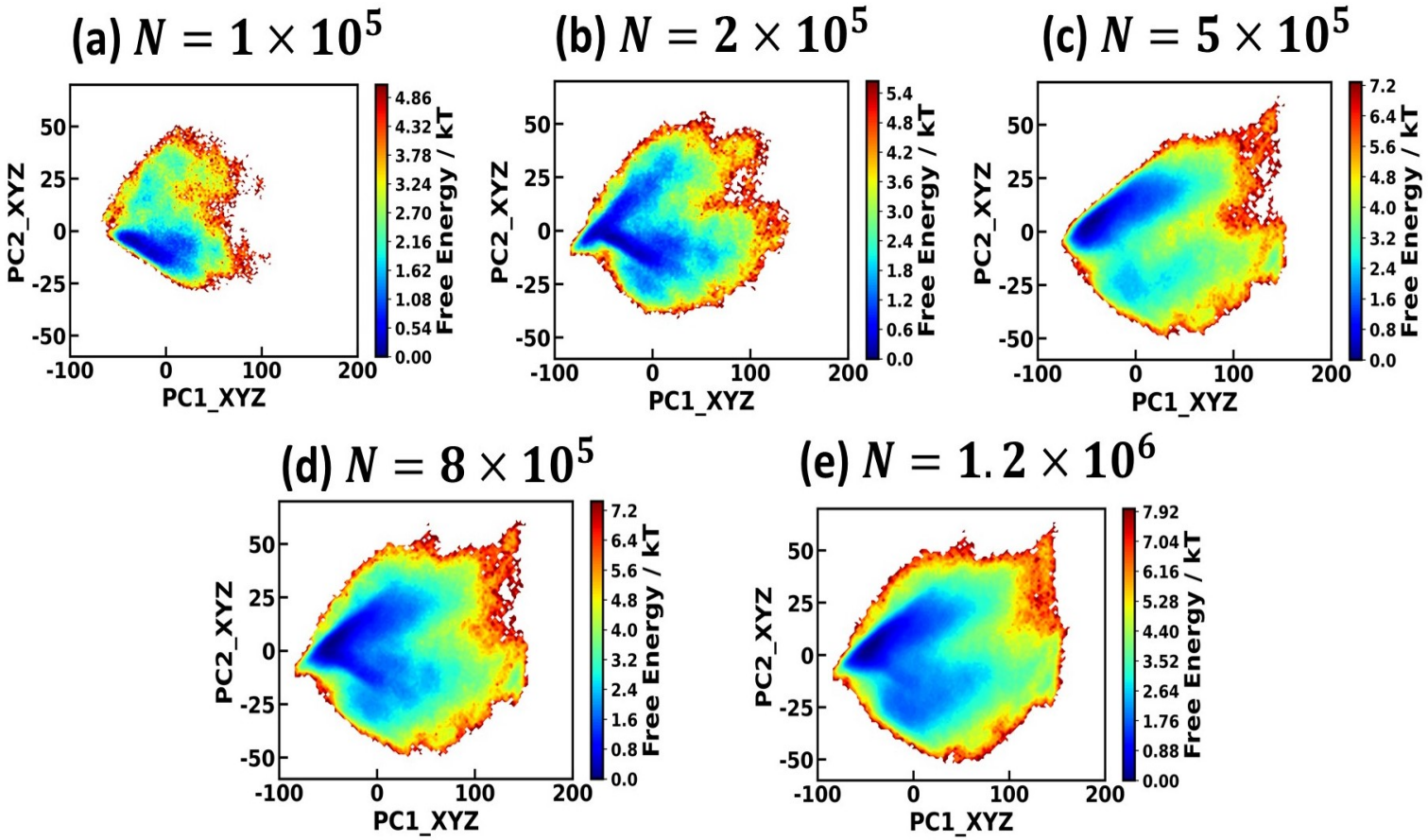
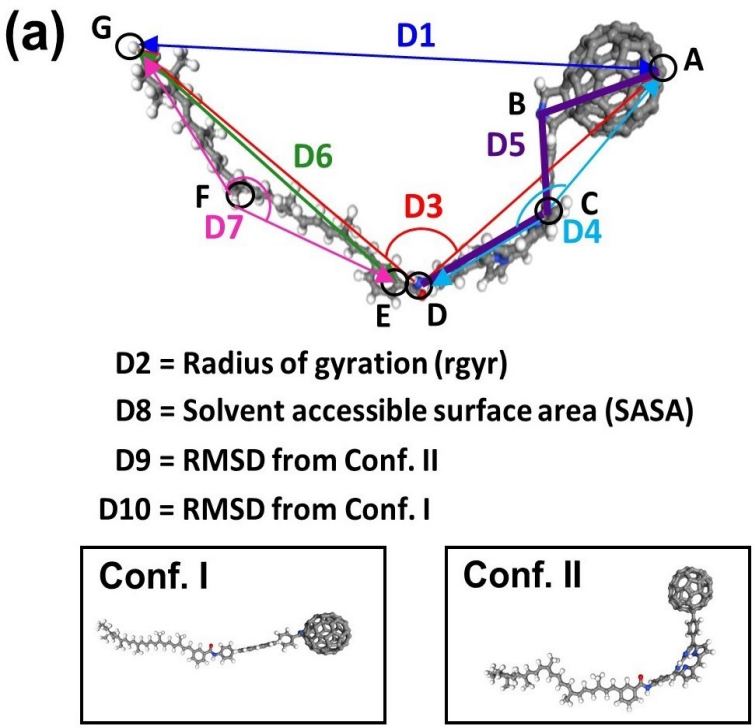
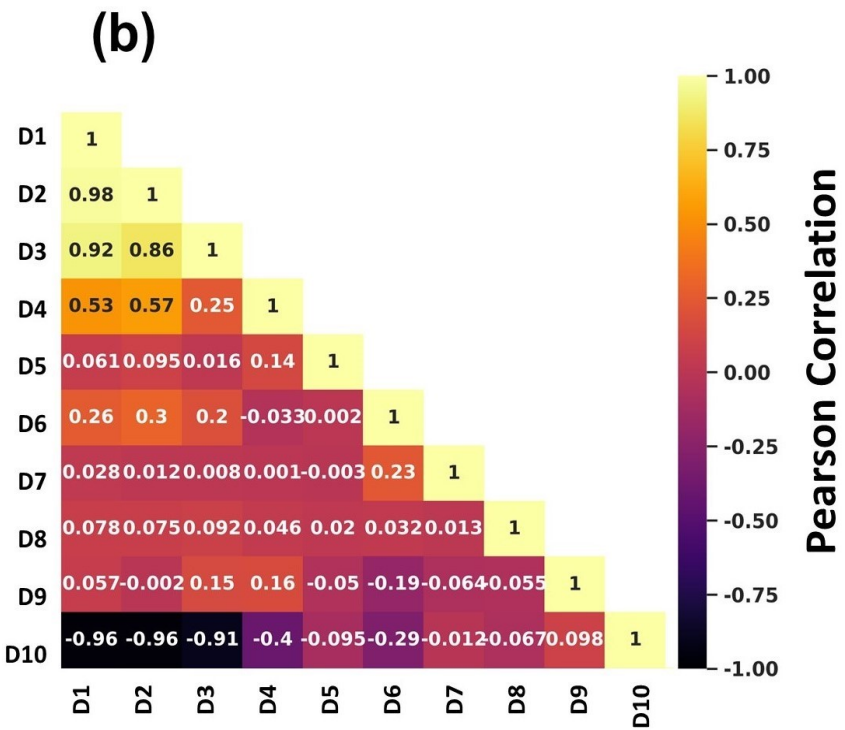


Figure 4: Free energy landscape obtained with data sets of increasing size.

2. Extracting Landmark Structures

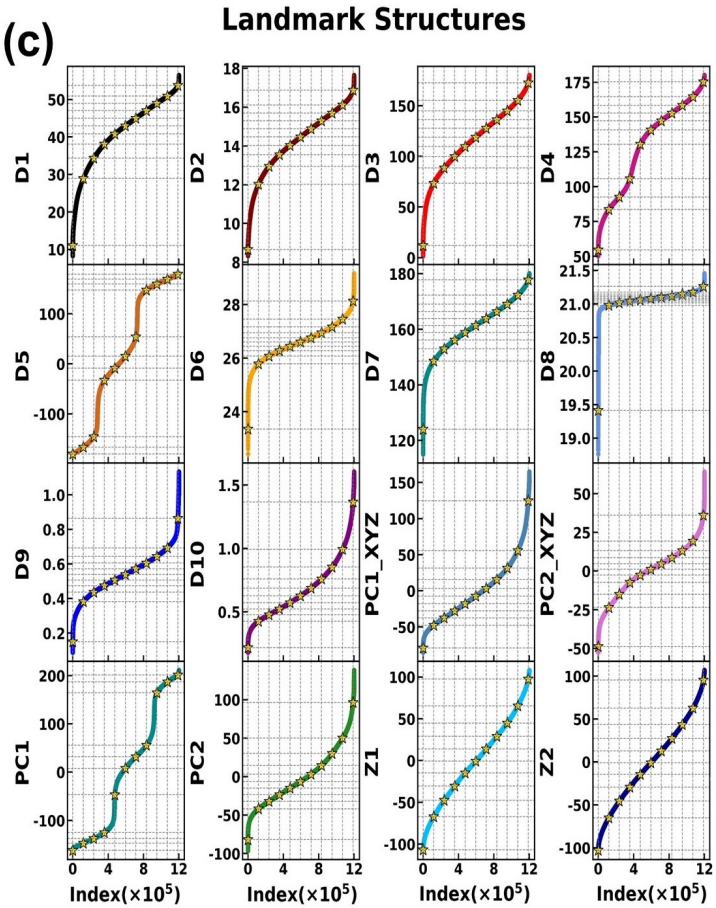
(a) Molecular descriptors ;
(b) Correlation between descriptors

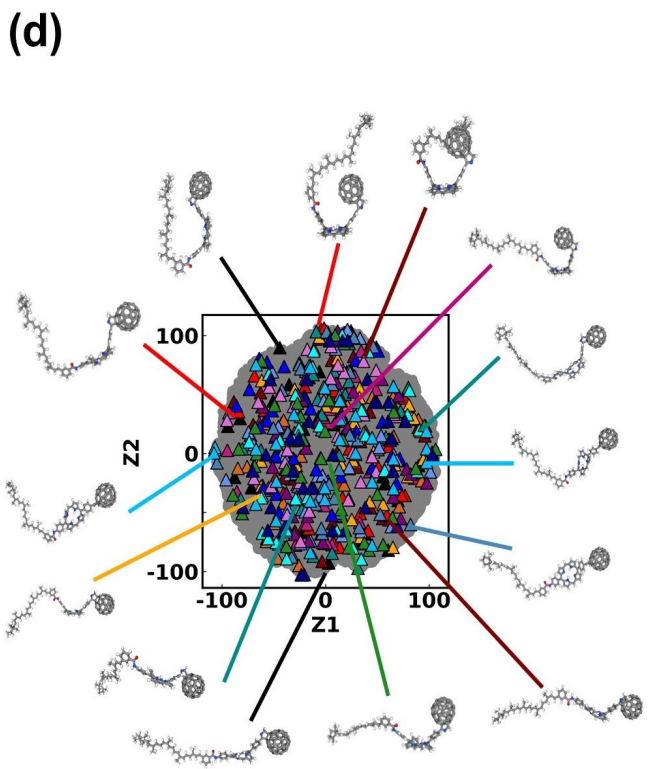




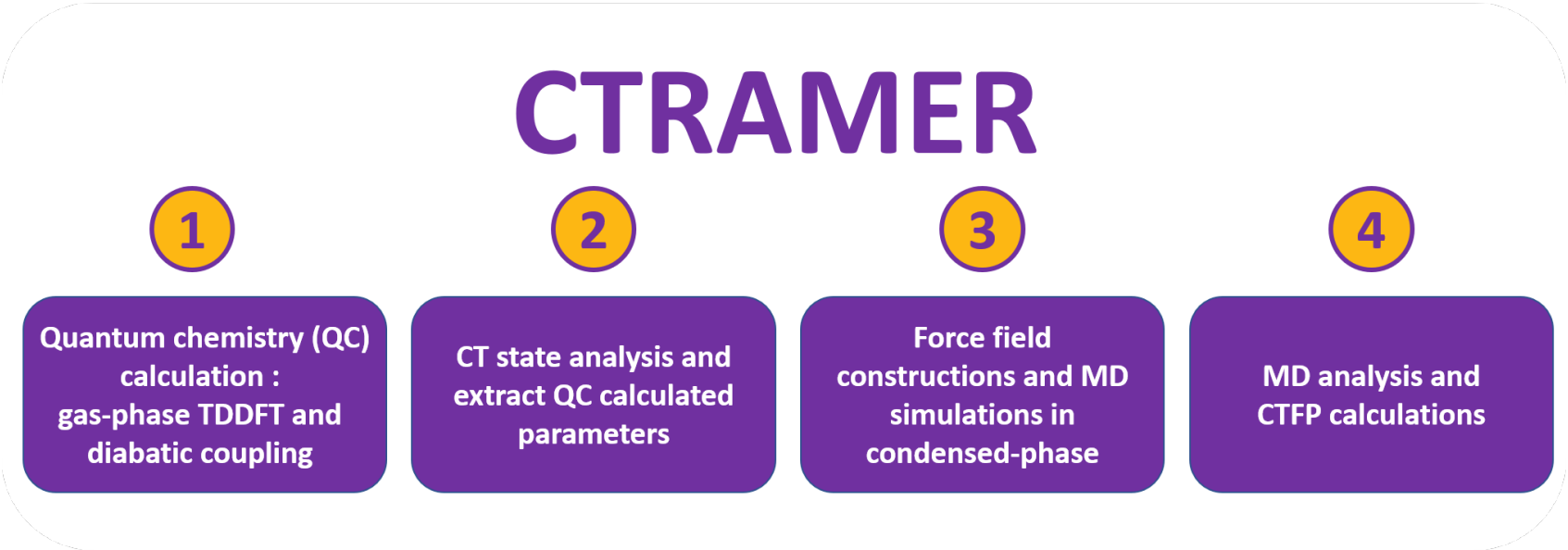
2. Extracting Landmark Structures

(c) Stochastic sampling around equipartition grid points;
(d) Selected landmark structures.





3. CT Rate Constant Calculation



CTRAMER (Charge-Transfer RAtes from Molecular dynamics, Electronic structure, and Rate theory).
For more detail see: JPCB, 125, 13267-13278, (2021) and the CTRAMER paper JCP, 154, 214108 (2021).

3. CT Rate Constant Calculation

The CT rate constant was calculated based on linearized semiclassical Fermi's golden rule (LSC FGR), with Marcus level of approximation

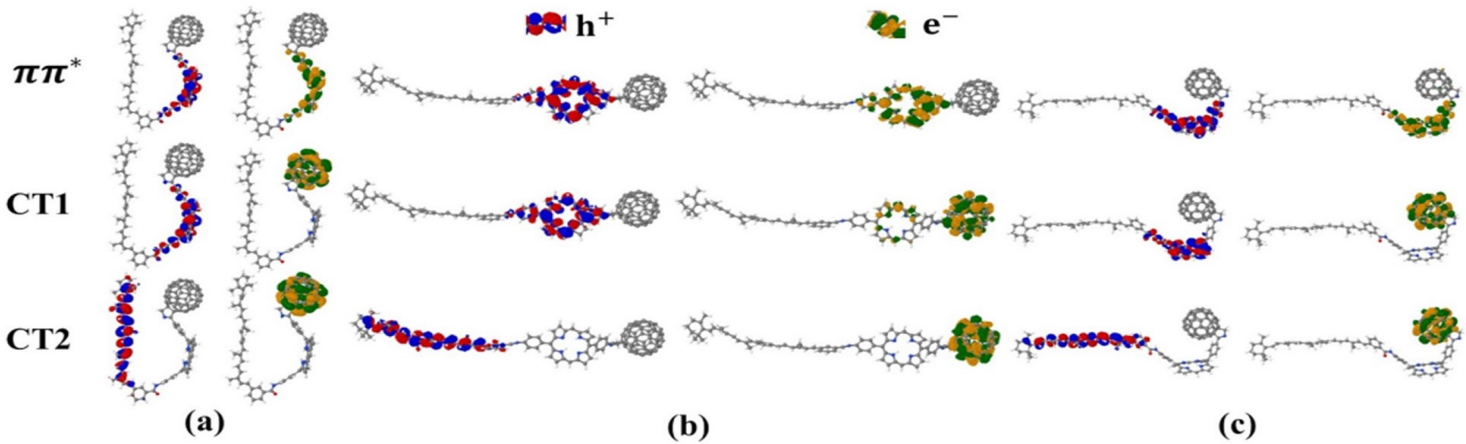
$$k_{D\to A} = \frac{\Gamma^2}{\hbar}\sqrt{\frac{2\pi}{\sigma_U^2}}\exp\left(-\frac{\langle U\rangle^2}{2\sigma_U^2}\right), \tag{1}$$

where Γ is the electronic coupling, $\langle U\rangle$ and σ_U^2 are the ensemble average and corresponding variance of the donor-acceptor energy gap, respectively. These properties are obtained from quantum chemistry calculation and molecular dynamics simulation.
Alternatively, expressing in term of the Marcus parameters

$$k_{D\to A}^{\mathrm{M}} = \frac{\Gamma^2}{\hbar}\sqrt{\frac{\pi}{k_B T E_r}}\exp\left(-\frac{(\Delta E + E_r)^2}{4k_B T E_r}\right), \tag{2}$$

where the reorganization energy $E_r = \sigma_U^2/(2k_B T) = -\Delta E - \langle U\rangle$, and ΔE is the donor-to-acceptor reaction free energy.

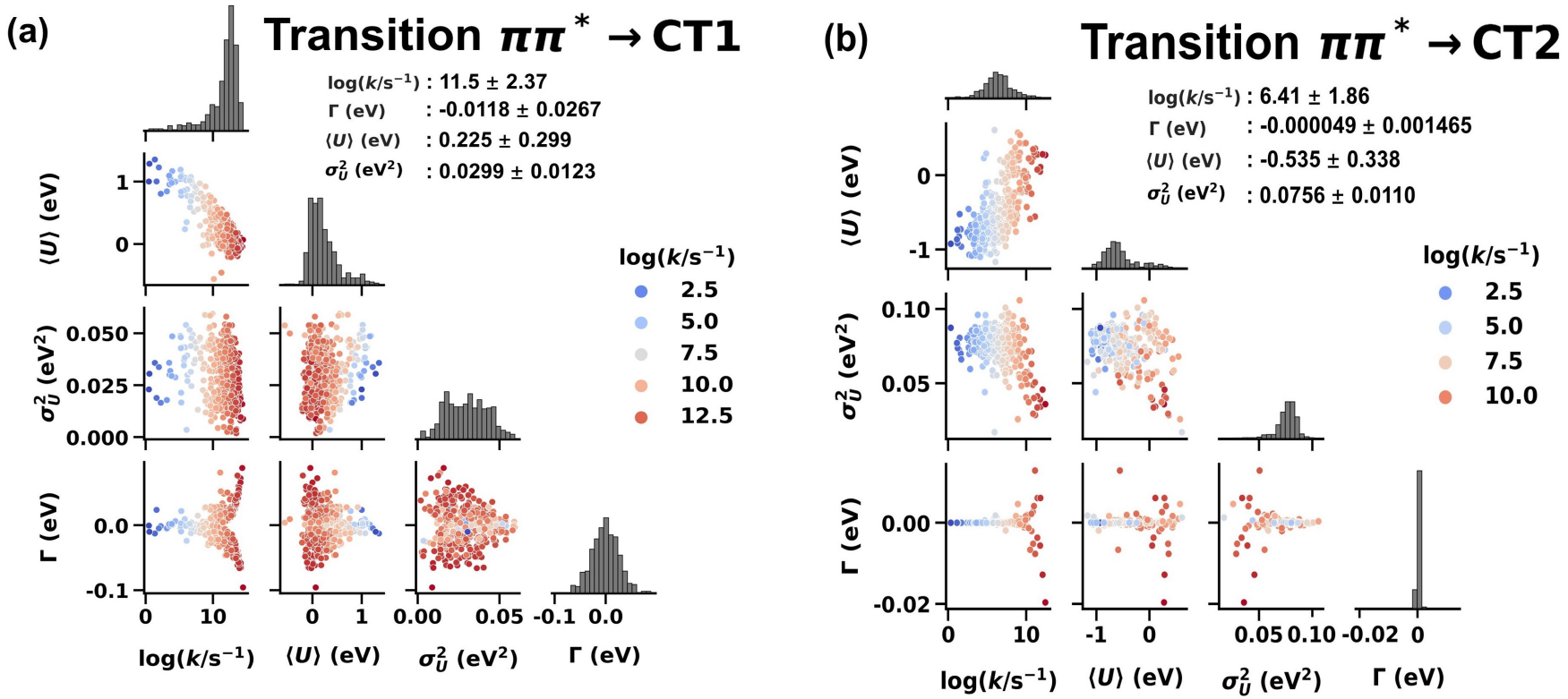
3. CT Rate Constant Calculation



Excitation	(a) Conf #0		(b) Conf #32		(c) Conf #364	
$E_{\pi\pi^*}$	2.58		1.74		2.67	
E_{CT1}	1.87		2.04		1.49	
E_{CT2}	2.01		2.65		2.53	
Transition	$\pi\pi^* \rightarrow$ CT1	$\pi\pi^* \rightarrow$ CT2	$\pi\pi^* \rightarrow$ CT1	$\pi\pi^* \rightarrow$ CT2	$\pi\pi^* \rightarrow$ CT1	$\pi\pi^* \rightarrow$ CT2
k (s^{-1})	8.12×10^{9}	4.41×10^{6}	5.56×10^{13}	8.16×10^{5}	1.21×10^{14}	8.03×10^{8}
Γ (eV)	-2.68×10^{-3}	2.01×10^{-5}	5.47×10^{-2}	-7.20×10^{-5}	6.58×10^{-2}	-4.64×10^{-4}
$\langle U \rangle$ (eV)	0.559	0.2064	−0.124	−0.835	0.000150	−0.440
σ_U^2 (eV^2)	0.0594	0.0726	0.0177	0.0785	0.0186	0.0723
E_r (eV)	1.15	1.40	0.341	1.52	0.360	1.40
ΔE (eV)	−1.71	−1.61	−0.217	−0.681	−0.360	−0.958

Figure 5: Charge transfer properties for prototypical triad conformations.

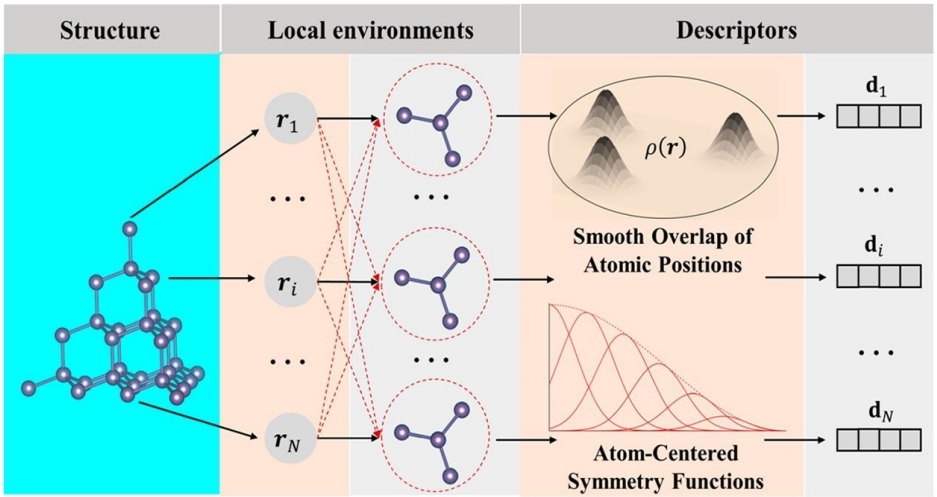
3. CTRAMER for CTFP Database

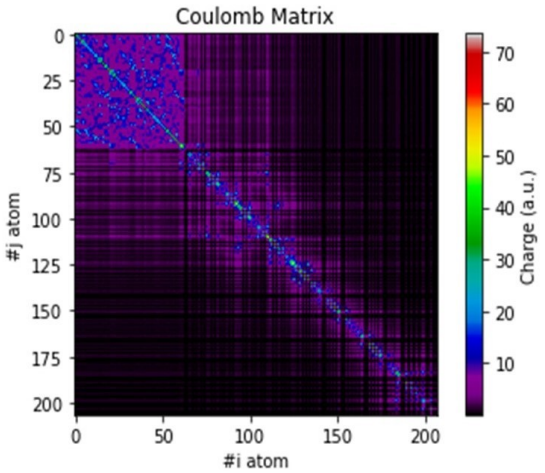


(a) (b)

Figure 6: Data distribution in the CTFP database

4. Feature Engineering





Descriptors for molecular structure
JPCL, 11, 8710, (2020)

Coulomb Matrix

The diagonal CM elements are defined as $C_{ii} = 0.5 Z_i^{(2.4)}$ and the off-diagonal elements are $C_{ij} = Z_i Z_j / |\mathrm{r}_i - \mathrm{r}_j|$ $(i \neq j)$, where Z_i and r_i are the nuclear charge and position of the i-th atom.

5. Machine Learning for CT Landscape

Screening of Features, ML Model, and Learning Strategy

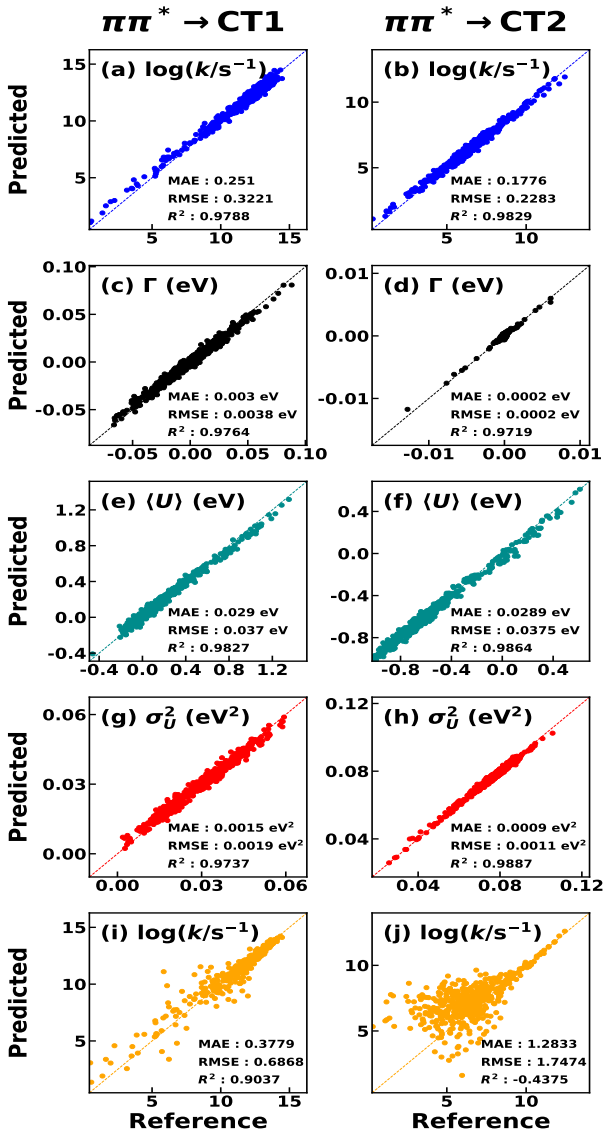
Features : {SOAP, CM, ACSF, PCA, Descriptors, ... }
ML Model : {ANN, KRR(lin, poly, rbf, chi2, Laplacian, Gaussian, custom), RF, Lasso Reg, XGBoost, ... }
Learning Strategy : {LOOCV, 5-Fold CV, Custom Split, ... }

= Coulomb Matrix (CM) + Kernel Ridge Regression (KRR) + 5-fold Cross-Validation

Training Duration : $\sim <$ 1 CPU core hour / parameter set.

$\sim$ 150 parameter set were screened.

5. Machine Learning for CT Landscape Interpolation



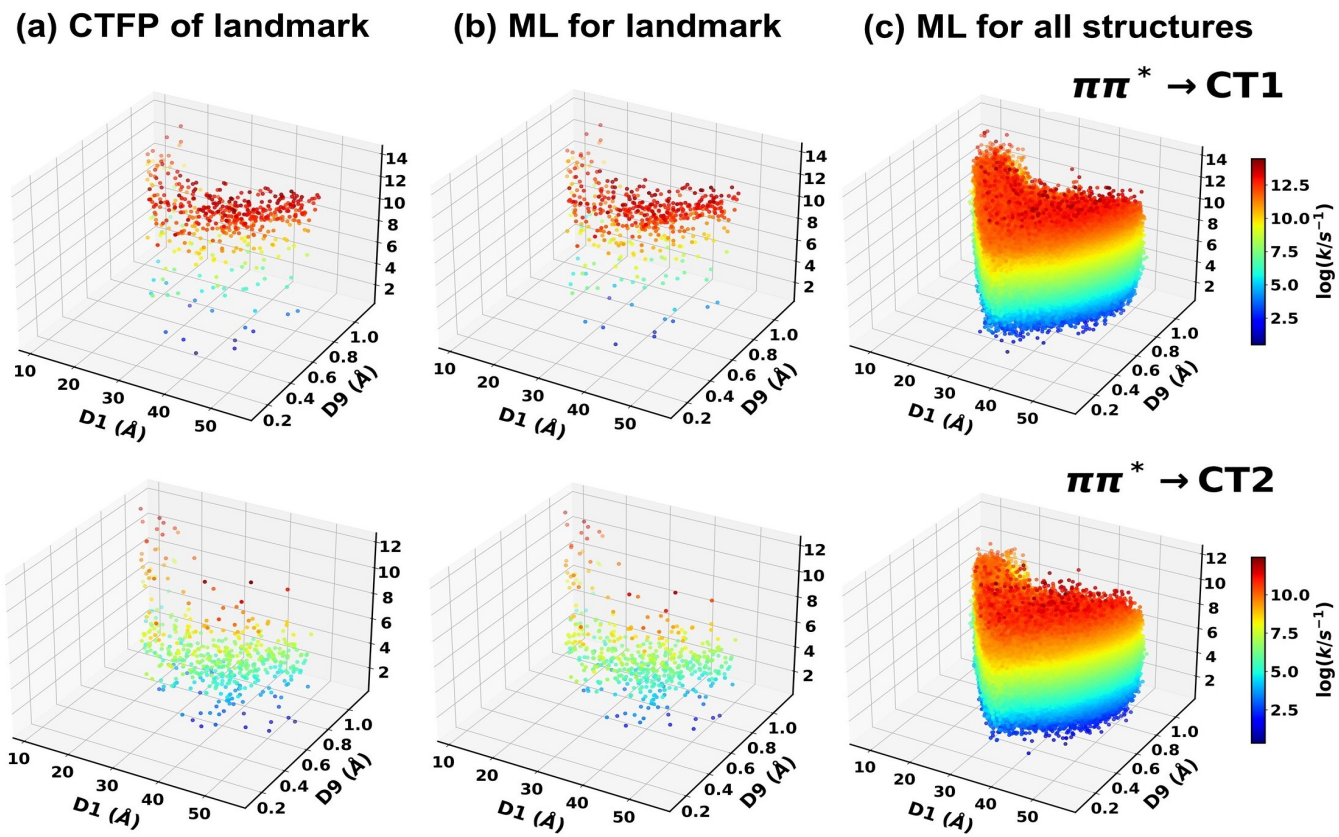
Insights learned:

- KRR works well for direct QSPR mapping using small dataset.
- Calculated $\log(k/\mathrm{s}^{-1})$ (from ML predicted parameters) are prone to error propagation.
- ML model trained to directly predict $\log(k/\mathrm{s}^{-1})$ is superior to that trained on the linear scale k/s^{-1}.

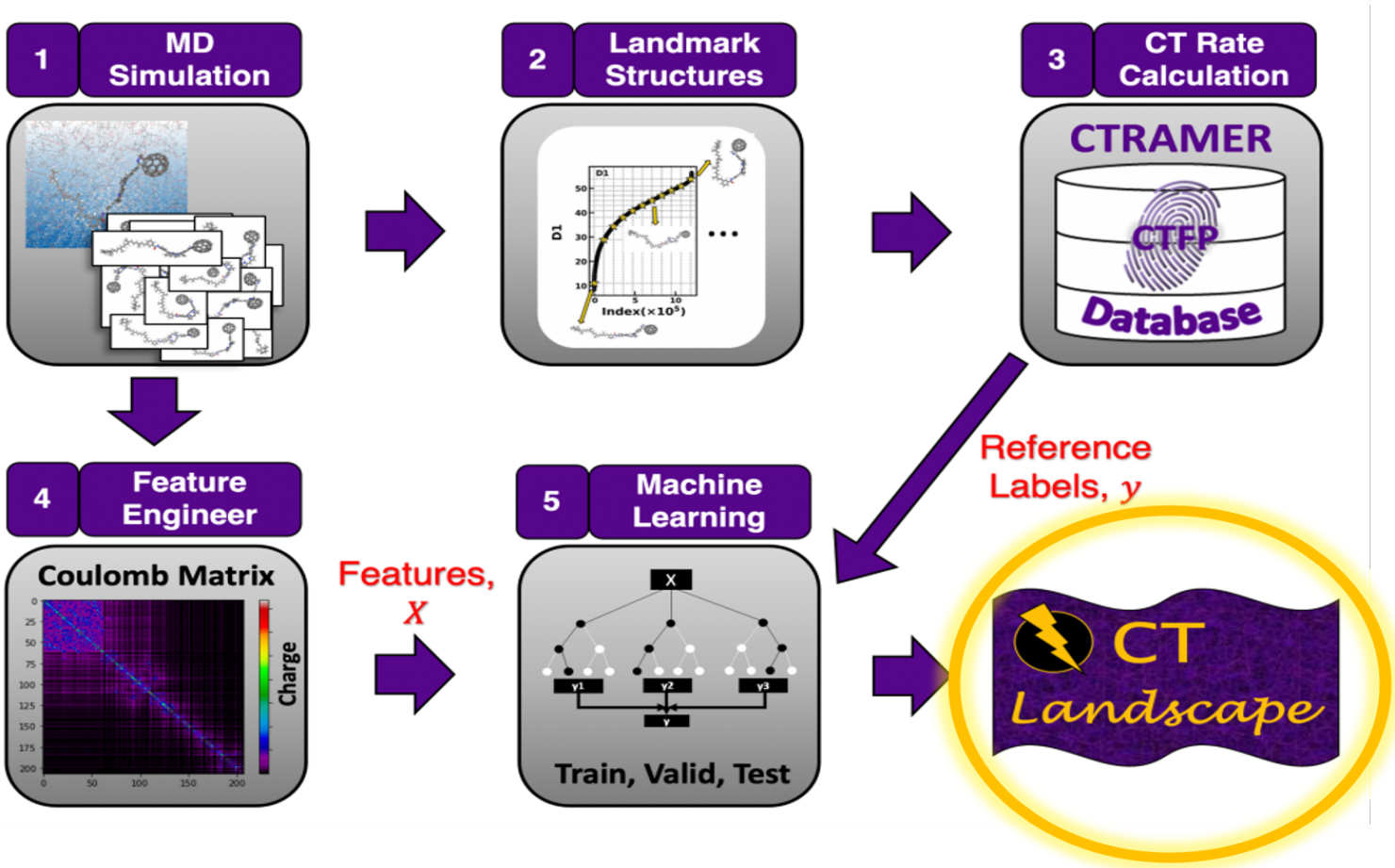
*KRR : Kernel Ridge Regression;
QSPR : Quantitative Structure-Property Relationship.

5. Machine Learning for CT Landscape Interpolation

(a) *Ab-Initio* calculated (500 GPU hours + 97,000 CPU core hours)
(b) ML-Predicted ($<$ 1 s w/ single core CPU)
(c) ML-Constructed ($<$ 10 minutes w/ single core CPU)



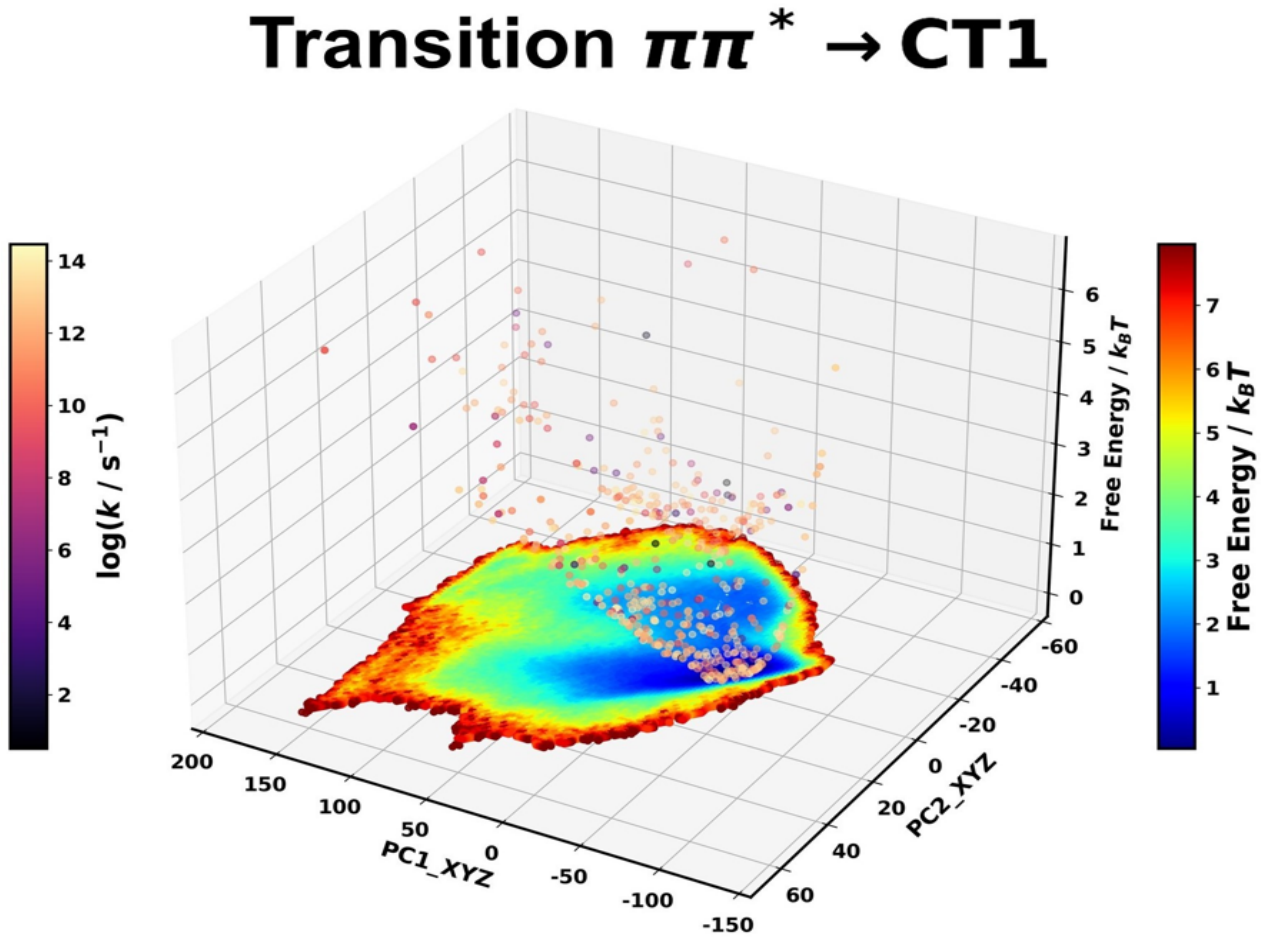
5 Steps of Charge Transfer Landscape Construction



CT Landscape Exploration

3 Examples of Charge Transfer Landscape Exploration

1. Ensemble-averaged CT-Rate Constant



Floor : Free Energy Landscape from 1.2 Million Conformations ;
Points : CT rate of 495 Landmark Structures.

1. Ensemble-averaged CT Rate Constant

$$\langle k_{D\to A}\rangle = \sum_{i=1}^{n} W_i \; k_{D\to A,i}. \tag{3}$$

with $W_i = e^{-\beta F_i} / \sum_{i=1}^{n} e^{-\beta F_i}$, n is the number of sampled structures and $\beta = 1/k_B T$ with $T = 300$K, and $F_i = -k_B T \ln[\mathrm{Prob.}(\mathrm{D}_i)]$, for the i-th landmark structure.

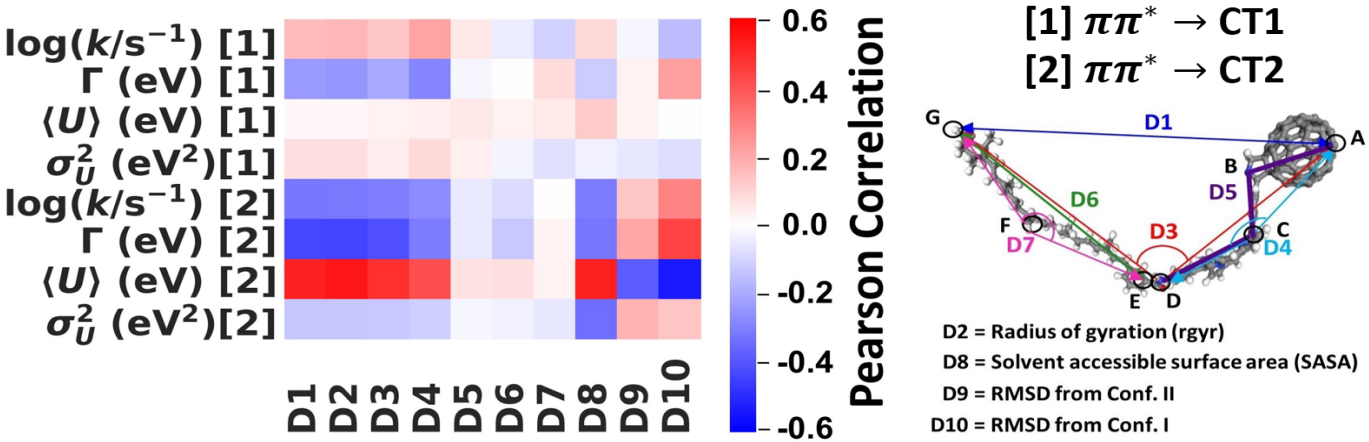
[495 Landmark Structures] [770 Random Structures]

$\langle k_{\pi\pi^*\to \mathrm{CT1}}\rangle = 0.82 \pm 0.23 \times 10^{11}\ \mathrm{s}^{-1} = 5.9 \pm 1.0 \times 10^{11}\ \mathrm{s}^{-1}$

Experimental data for similar triad molecules (T = 292K) in 2-MeTHF solvent, $\langle k_{\pi\pi^*\to \mathrm{CT1}}\rangle = 1.1 \times 10^{11}$ and $3.31 \times 10^{11} s^{-1}$.
JACS, 119, 1400, (1997); JPCB, 104, 18, (2000); JPCA, 107, 38, (2003).

Boltzmann weighting is essential to correctly reflect the probabilities of each conformation in the ensemble.

2. Correlation Studies for Molecular Insights



1. All the descriptors are less correlated with the CT properties of the transition $\pi\pi^* \rightarrow$ CT1 than that of $\pi\pi^* \rightarrow$ CT2.
2. The descriptors D1–D4 and D8–D10 have strong (anti)correlation with CT properties. Relative distance and orientation of Carotenoid and Fullerene is important.
3. The descriptors D5–D7 was found to be almost uncorrelated to CTFP. CTFP properties are less sensitive to the Fullere-Porphyrin orientation (D5), and the Carotenoid local structure (D6-D7).

3. Reverse Design using the CT landscape

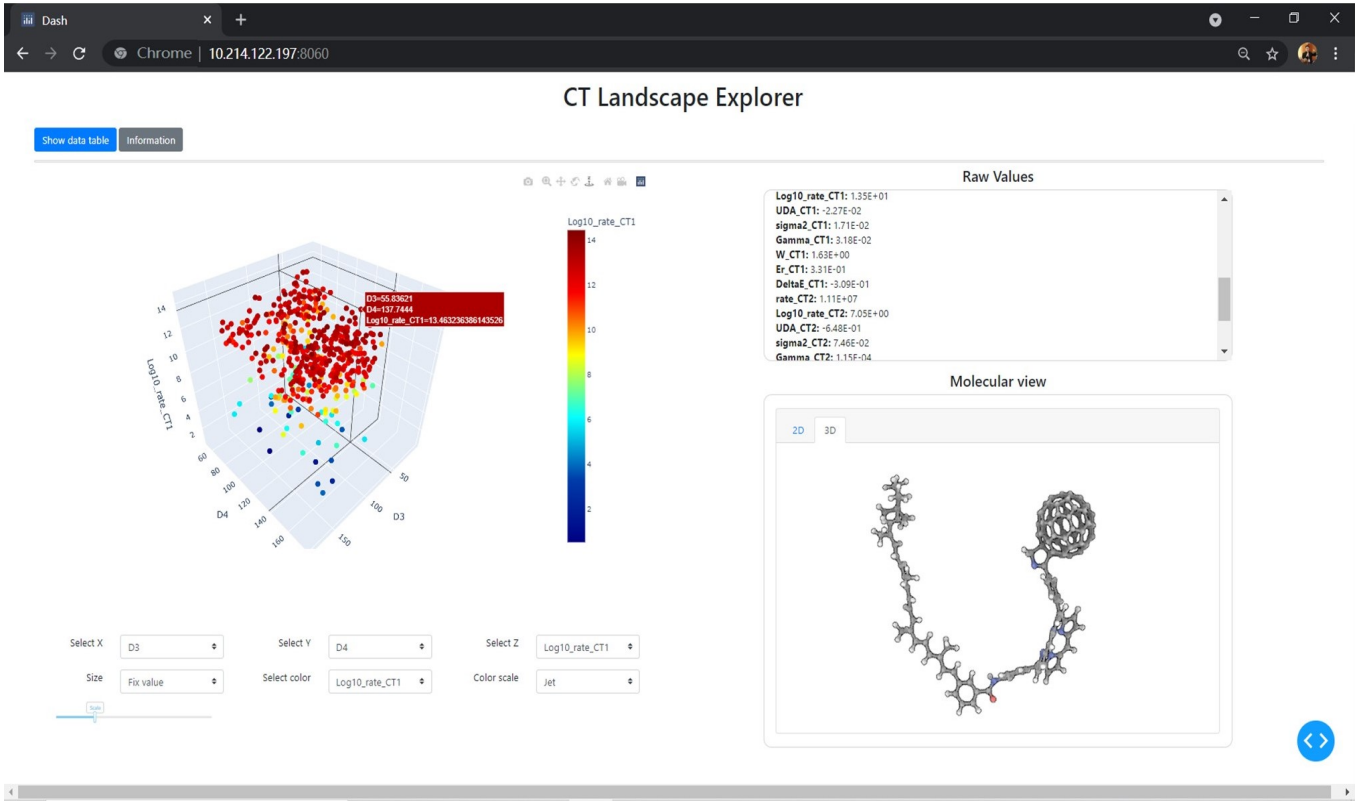


Figure 7: Screenshot of the CT landscape explorer interactive web interface. The CT landscape explorer source code is accessible from `https://github.com/xiangsunlab/ct_landscape`.

Server access

Outline

Summary

Summary

- We have developed ML models for Conformation-CTFP mapping with $R^2 > 0.97$ and MAE and RMSE of $\lessapprox 1 k_B T$.
- We constructed CT landscape for 1.2 million MD-sampled conformations employing ML models with the 5 steps strategy.
- We demonstrated 3 examples of CT landscape exploration to extract chemical insights.

Outlook

Construction and exploration of charge transfer landscape will be a useful tool to investigate the influence of conformation space on charge transfer properties and accelerate molecular/material design for various applications.

Acknowledgement

Sun Group @ NYU 上海 SHANGHAI 纽约大学

Funding :

上海纽约大学 NYU SHANGHAI

NSFC National Natural Science Foundation of China

References

1. JPCB, 125, 13267-13278, (2021)
2. JCP, 154, 214108, (2021).
3. Adv. Mater., 34, 2202575, (2022).
4. Adv. Energy Mater., 9, 1900720, (2019).
5. JPCL, 11, 8710, (2020).
6. JACS, 119, 1400, (1997).
7. JPCB, 104, 18, (2000).
8. JPCA, 107, 38, (2003).

Thank You !