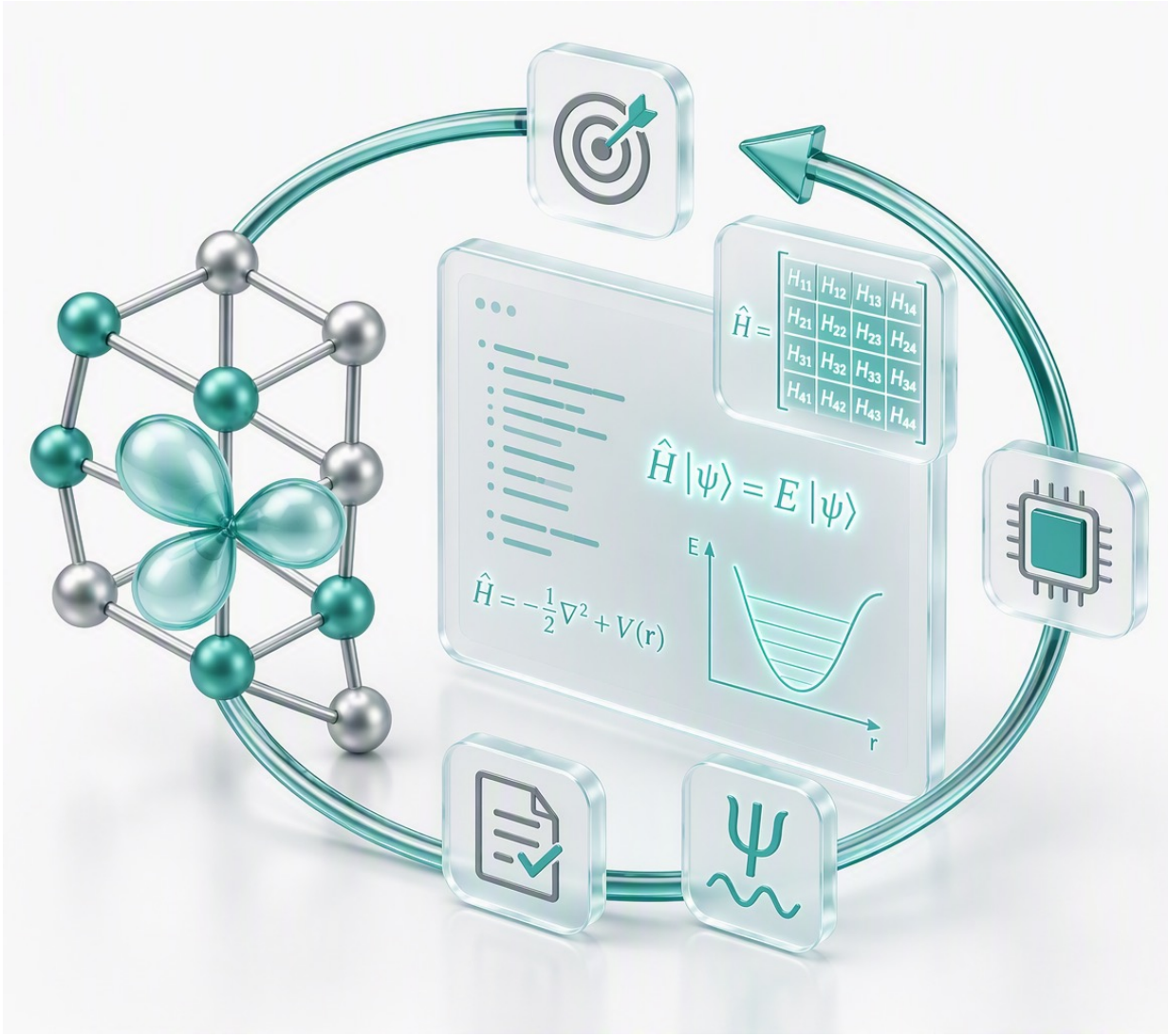
中国科学院大学 University of Chinese Academy of Sciences
前沿交叉科学学院 School of Advanced Interdisciplinary Sciences

Environment-Adaptive Atomic Orbital Basis Function, Agent Auto-Research for Molecular Designing



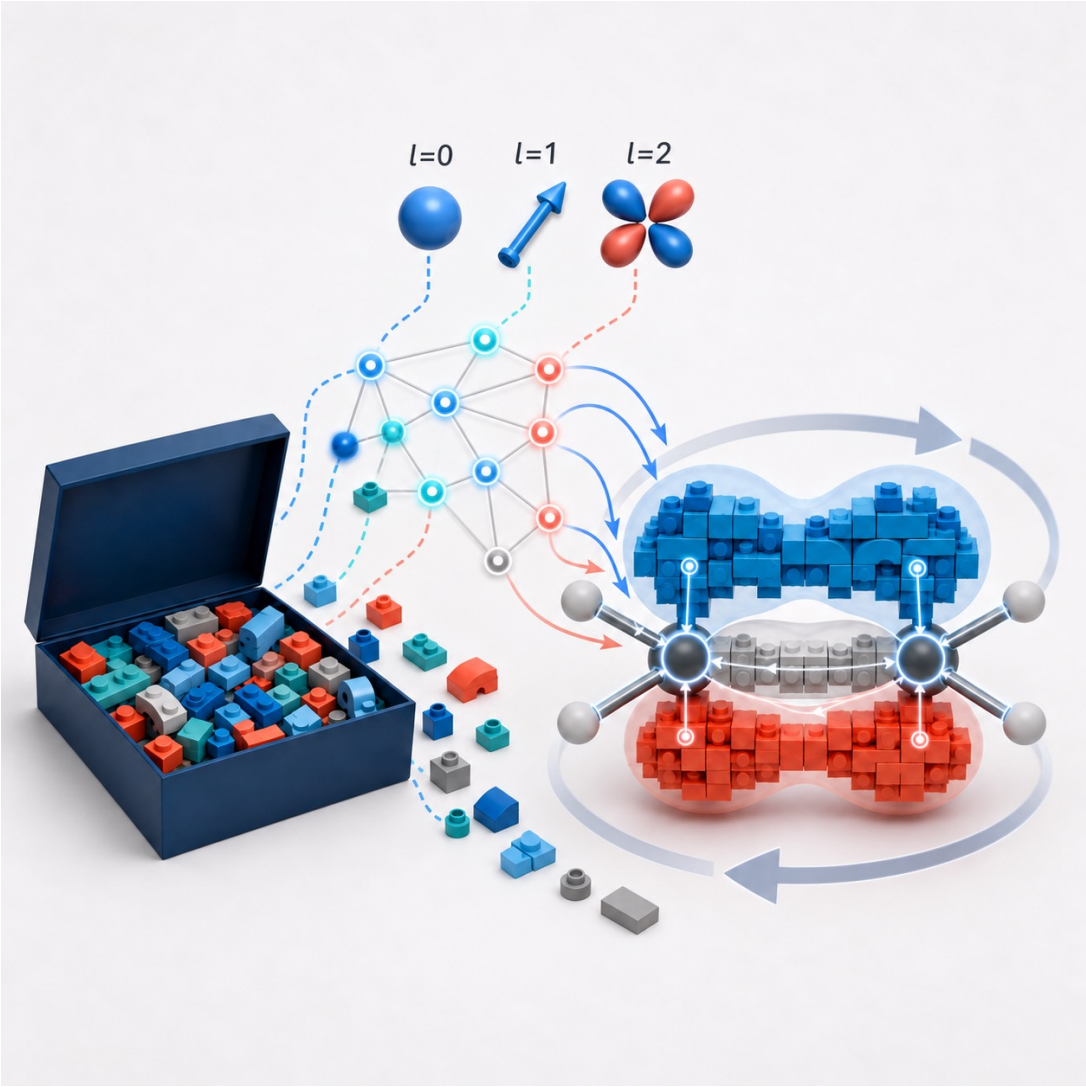
Yuan Jiao

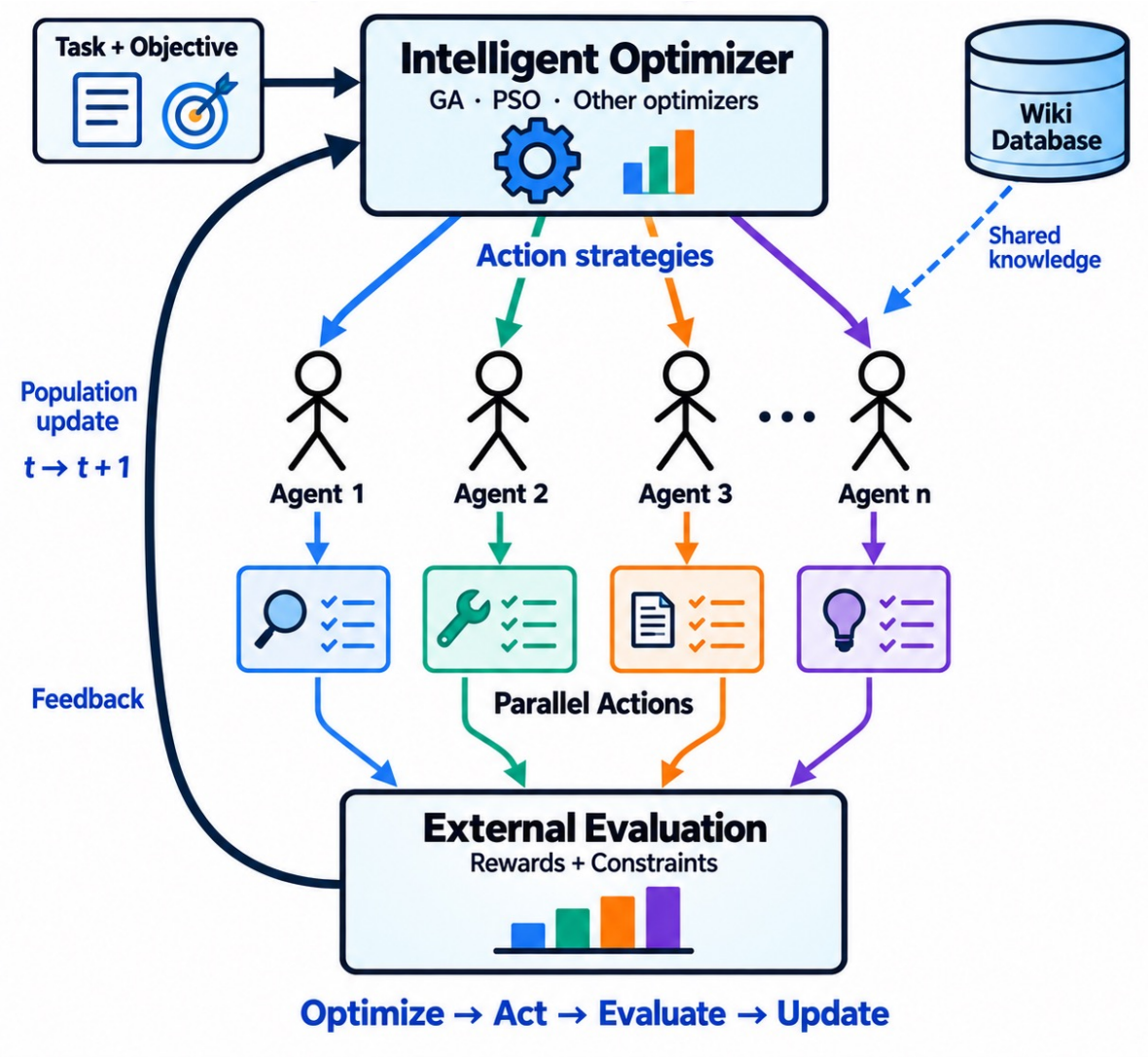
School of Advanced Interdisciplinary Sciences, SAIS

University of Chinese Academy of Sciences

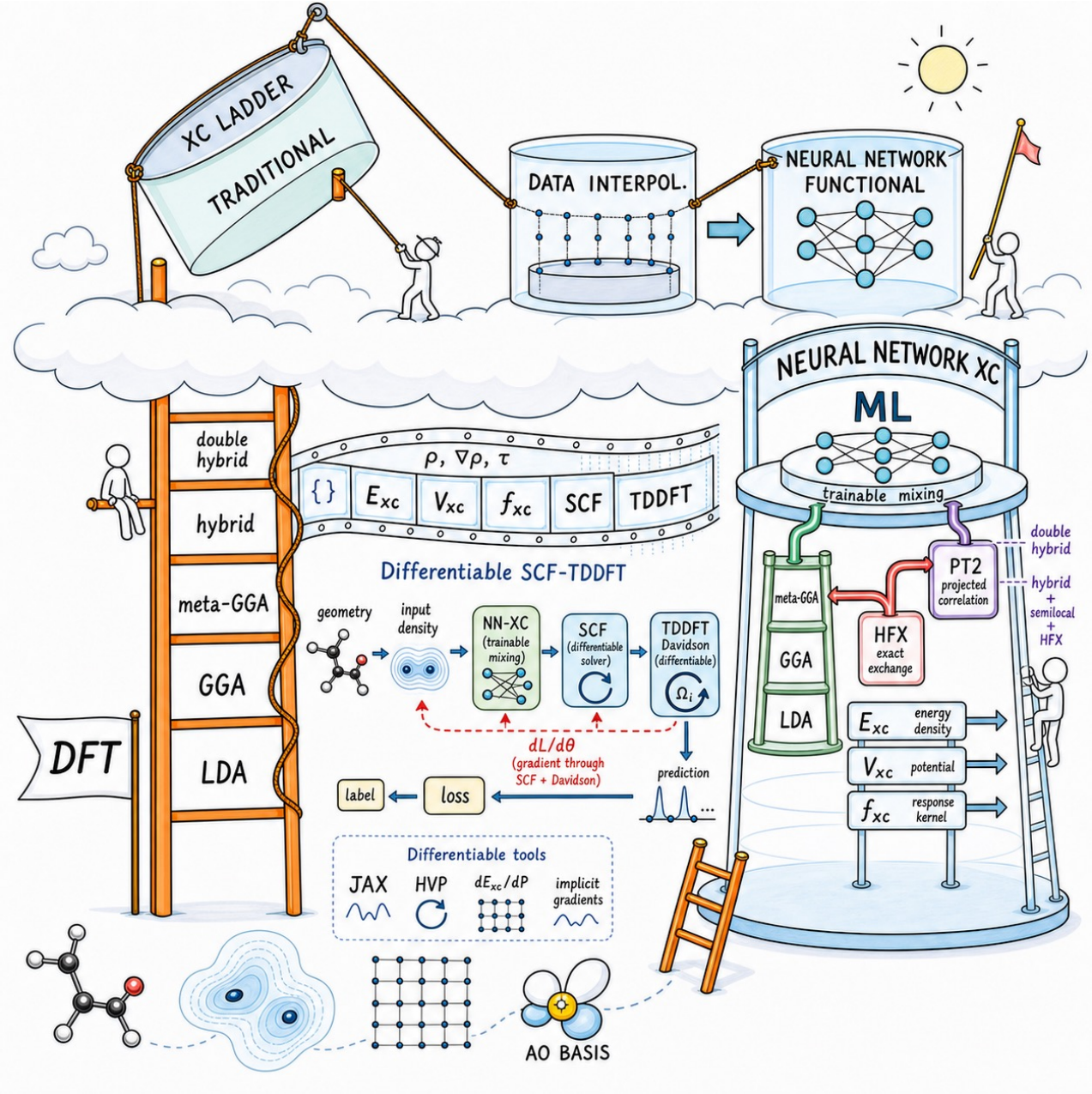
■ Environment-Adaptive Atomic Orbital Basis Function

■ Agent Auto-Research for Molecular Designing





■ Background-From GradTDDFT to GradSCF



- **One-Electron Integral**

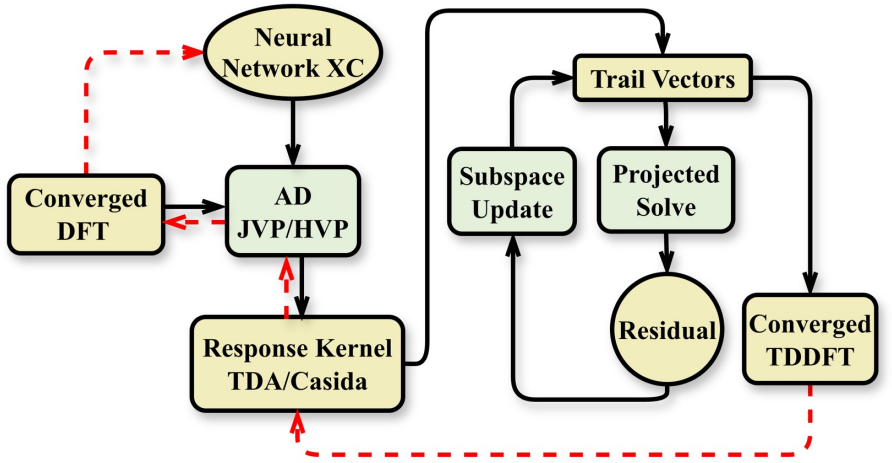
$$\boldsymbol{S_{\mu\nu}} = \int \chi_\mu(r)\chi_\nu(r)dr \qquad \mathrm{T}_{\boldsymbol{\mu\nu}} = \int \chi_\mu(r)\nabla^2\ \chi_\nu(r)dr$$

$$\boldsymbol{h_{\mu\nu}} = \int \chi_\mu(r)\hat{\mathrm{h}}\ \chi_\nu(r)dr \quad \mathrm{v}_{\boldsymbol{\mu\nu}} = \int \chi_\mu(r)\frac{1}{\mathrm{r}}\ \chi_\nu(r)dr$$

- **Two-Electron Integral**

$$(\boldsymbol{\mu\nu}|\boldsymbol{\lambda\sigma}) = \int \frac{\chi_\mu^*(r_1)\chi_\nu(r_1)\chi_\lambda^*(r_2)\chi_\sigma(r_2)}{r_{12}} dr_1 dr_2.$$

- **HF/DFT and After Single Particle Orbital**



■ Atomic Orbital Basis Function

➢ Slater-Type Orbitals(STOs)

$$\phi^{\mathrm{STO}}_{\zeta n\ell m}(\boldsymbol{r}) \propto r^{n-1}e^{-\zeta r}Y_{\ell m}(\hat{\mathrm{r}})$$

$$\phi^{\mathrm{STO}}_{1\mathrm{s}}(r) \propto e^{-\zeta r} \qquad \frac{d\phi^{\mathrm{STO}}_{1\mathrm{s}}(r)}{d\mathrm{r}}|_{r=0} = -\zeta$$

- atom-centered
- Natural nuclear cusp and exponential tail

➢ Gaussian-Type Orbitals(GTOs)

$$\phi^{\mathrm{GTO}}_{\alpha n\ell m}(r) \propto r^{2(n-1)-\ell}e^{-\alpha r^2}Y_{\ell m}(\hat{r})$$

$$\phi^{\mathrm{GTO}}_{1\mathrm{s}}(r) \propto e^{-\alpha r^2} \qquad \frac{d\phi^{\mathrm{GTO}}_{1\mathrm{s}}(r)}{d\mathrm{r}}|_{r=0} = 0$$

- atom-centered
- Standard choice for molecular calculations
- Efficient evaluation of multicenter integrals

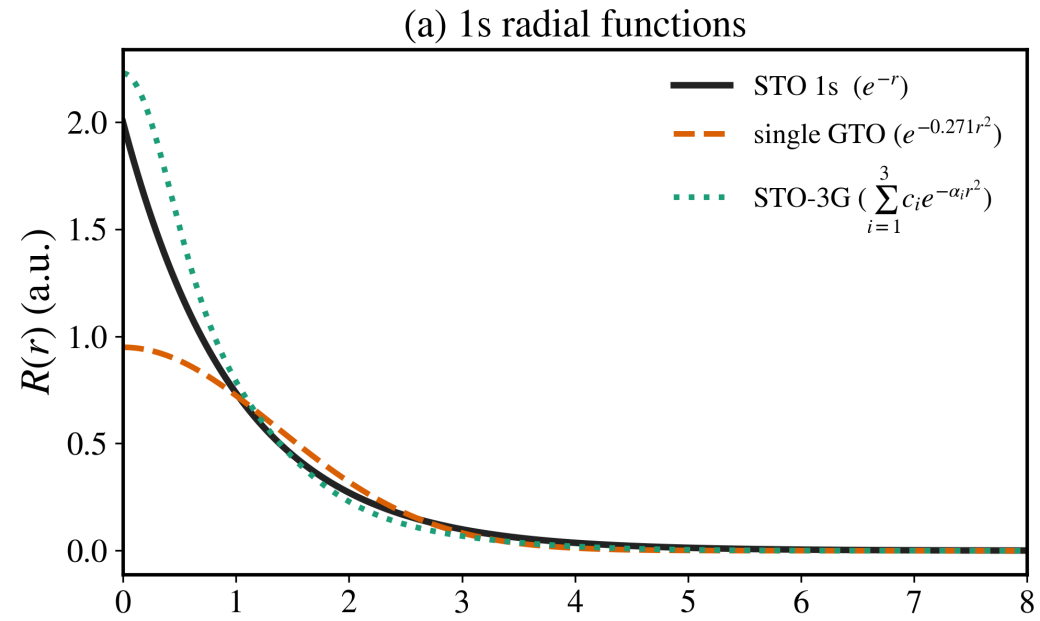
➢ Numerical Atomic Orbitals(NAOs)

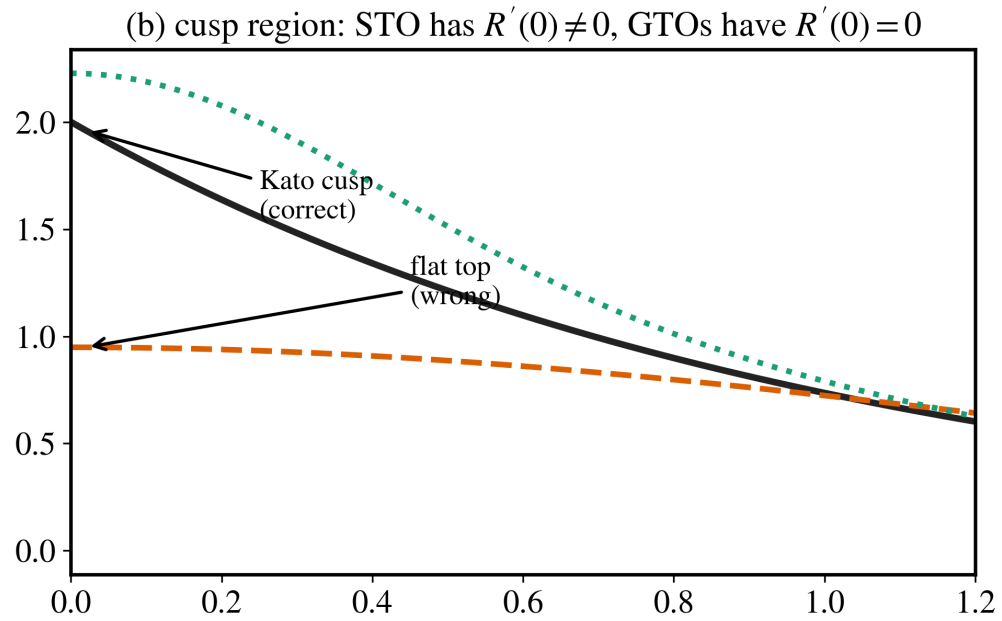
- Radial functions tabulated numerically
- Compact and flexible atom-centered basis

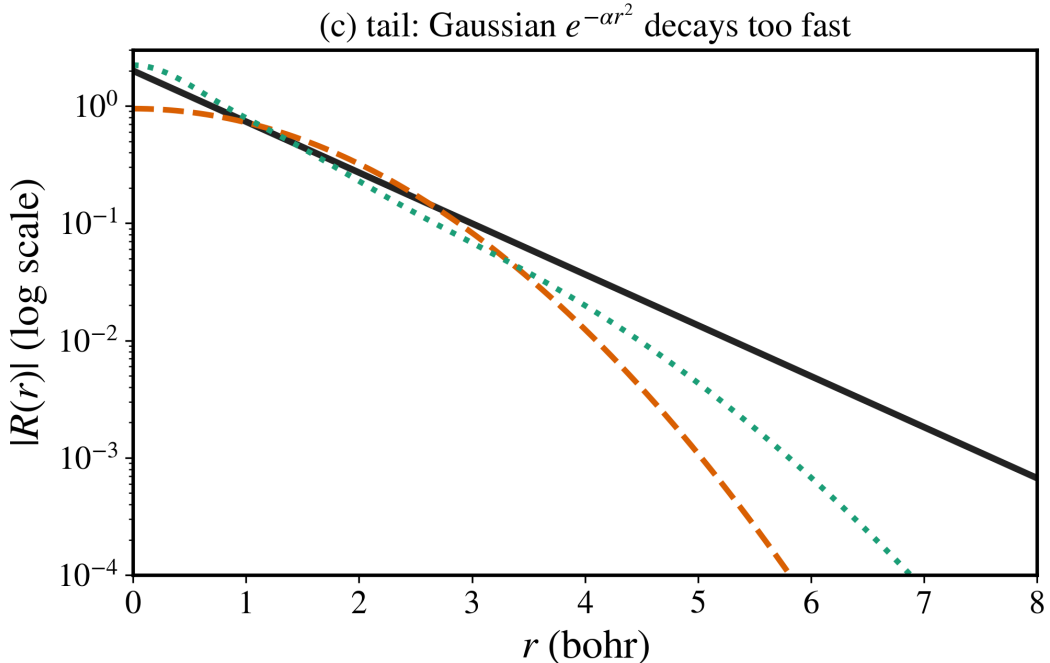
➢ Plane Waves(PWs)

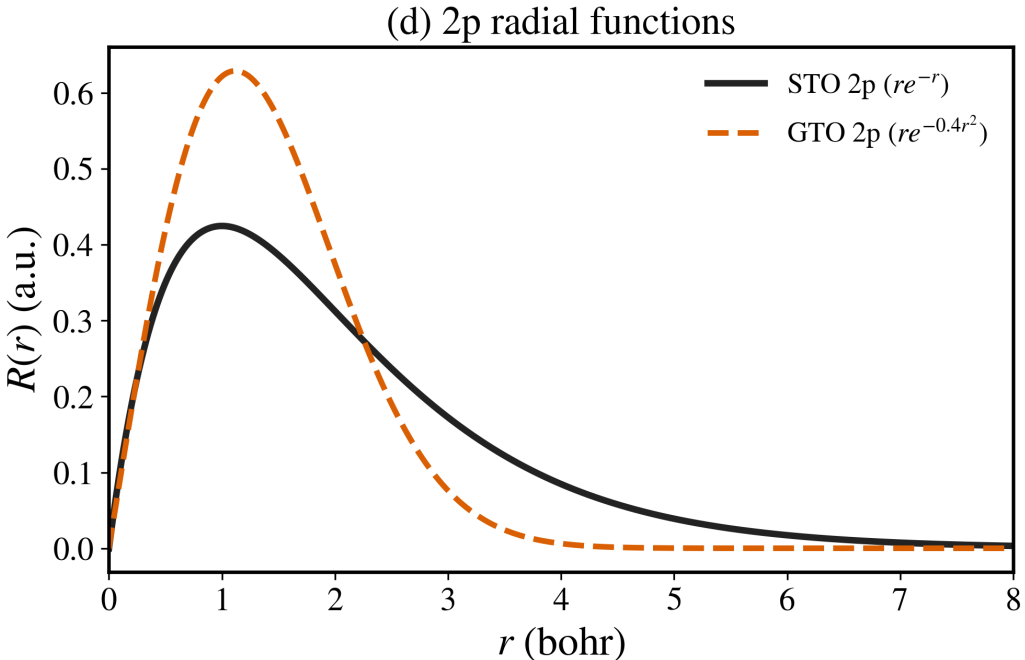
- Delocalized and orthonormal
- Naturally adapted to periodic systems
- Systematically converged with an energy cutoff

■ Atomic Orbital Basis Function









■ Atomic Orbital Basis Function

➢ Basis Set Family

- **Valence flexibility:** use double-, triple-, or higher-ζ descriptions to give each valence orbital several radial shapes
- **Polarization functions:** add higher angular momenta to describe anisotropic density, bonding, and angular correlation
- **Diffuse functions:** add small exponents to describe long-range density (anions, Rydberg states, noncovalent interactions, multipoles, (hyper)polarizabilities, etc)
- **Tight functions:** add large exponents to descrive core region (core correlation, nuclear magnetic shielding, spin-spin coupling constants, etc)

6-311++G**

6-> Six primitive functions for core orbital

3-> three primitive functions for inner orbital

1-> one primitive functions for valence orbital

++->diffusion primitive function

****->polarized primitive function**

■ Atomic Orbital Basis Function

➢ Basis Set Family

cc-pVDZ: 3s2p1d

s	s	s
p	p	
d		

aug-cc-pVDZ: 4s3p2d

s	s	s	s
p	p	p	
d	d		

d-aug-cc-pVDZ: 5s4p3d

s	s	s	s	s
p	p	p	p	
d	d	d		

cc-pVTZ: 4s3p2d1f

s	s	s	s
p	p	p	
d	d		
f			

aug-cc-pVTZ: 5s4p3d2f

s	s	s	s	s
p	p	p	p	
d	d	d		
f	f			

d-aug-cc-pVTZ: 6s5p4d3f

s	s	s	s	s	s
p	p	p	p	p	
d	d	d	d		
f	f	f			

cc-pVQZ: 5s4p3d2f1g

s	s	s	s	s
p	p	p	p	
d	d	d		
f	f			
g				

aug-cc-pVQZ: 6s5p4d3f2g

s	s	s	s	s	s
p	p	p	p	p	
d	d	d	d		
f	f	f			
g	g				

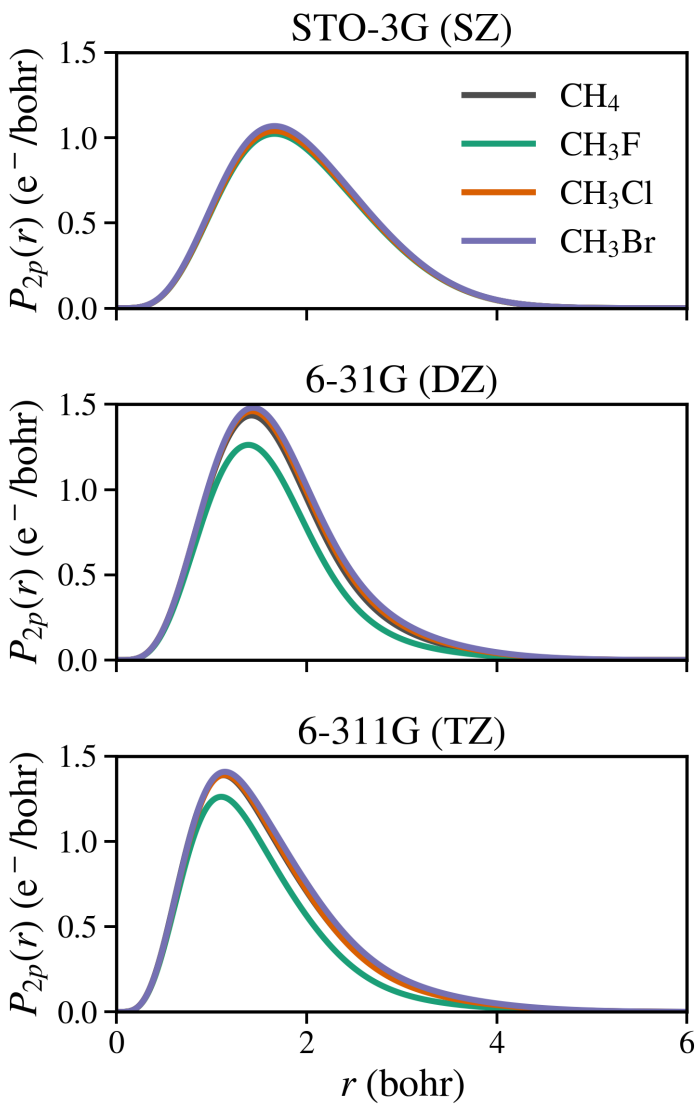
d-aug-cc-pVQZ: 7s6p5d4f3g

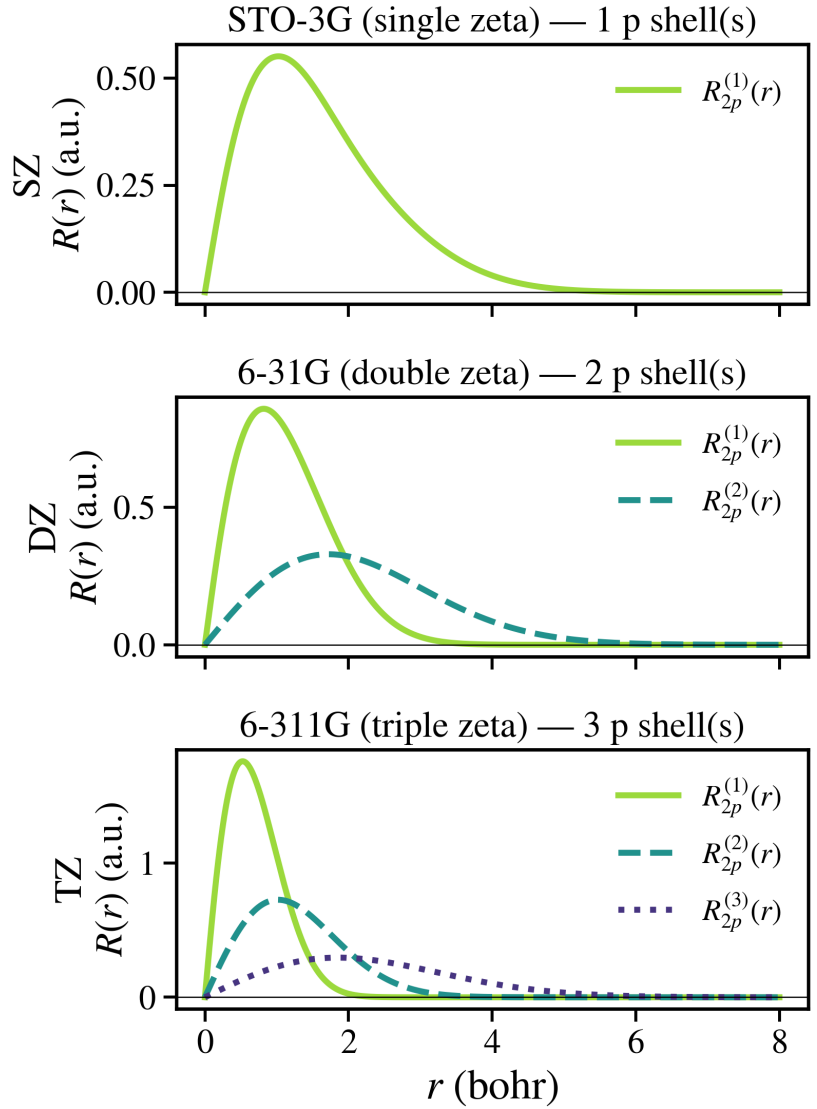
s	s	s	s	s	s	s
p	p	p	p	p	p	
d	d	d	d	d		
f	f	f	f			
g	g	g				

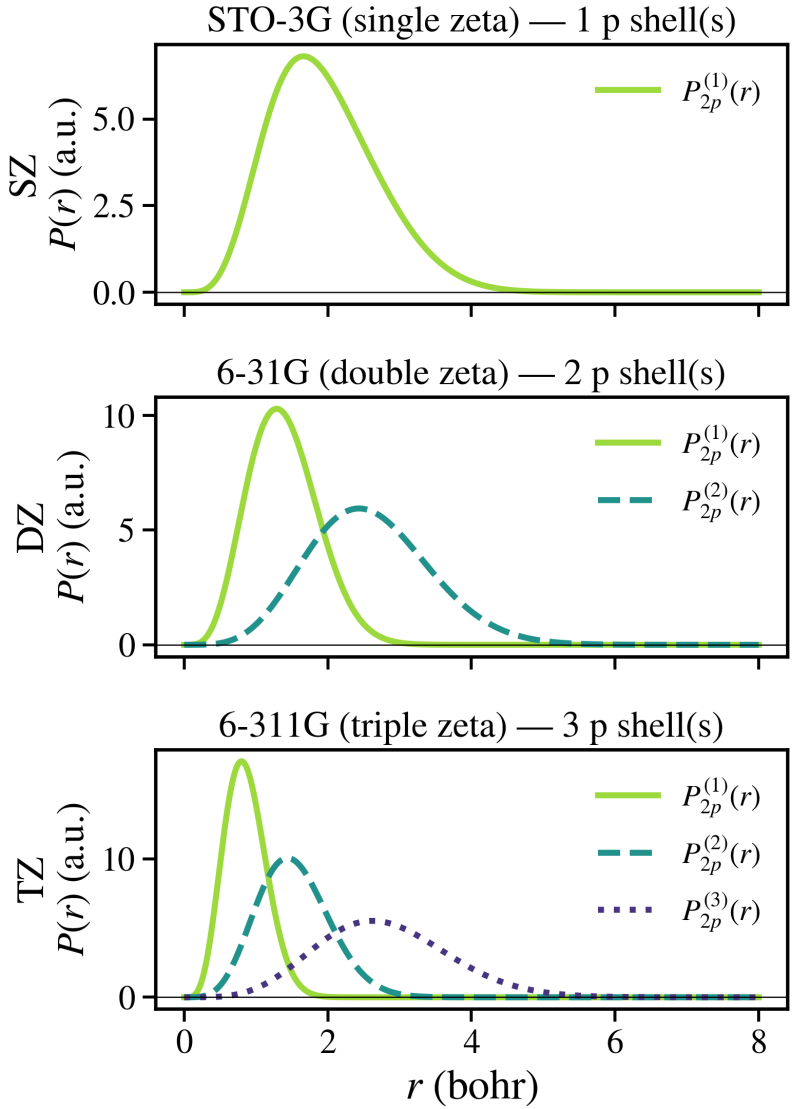
Github.com/pfloos/ESQC_2026

■ Atomic Orbital Basis Function

➢ Contracted Multi-Zeta Basis Set

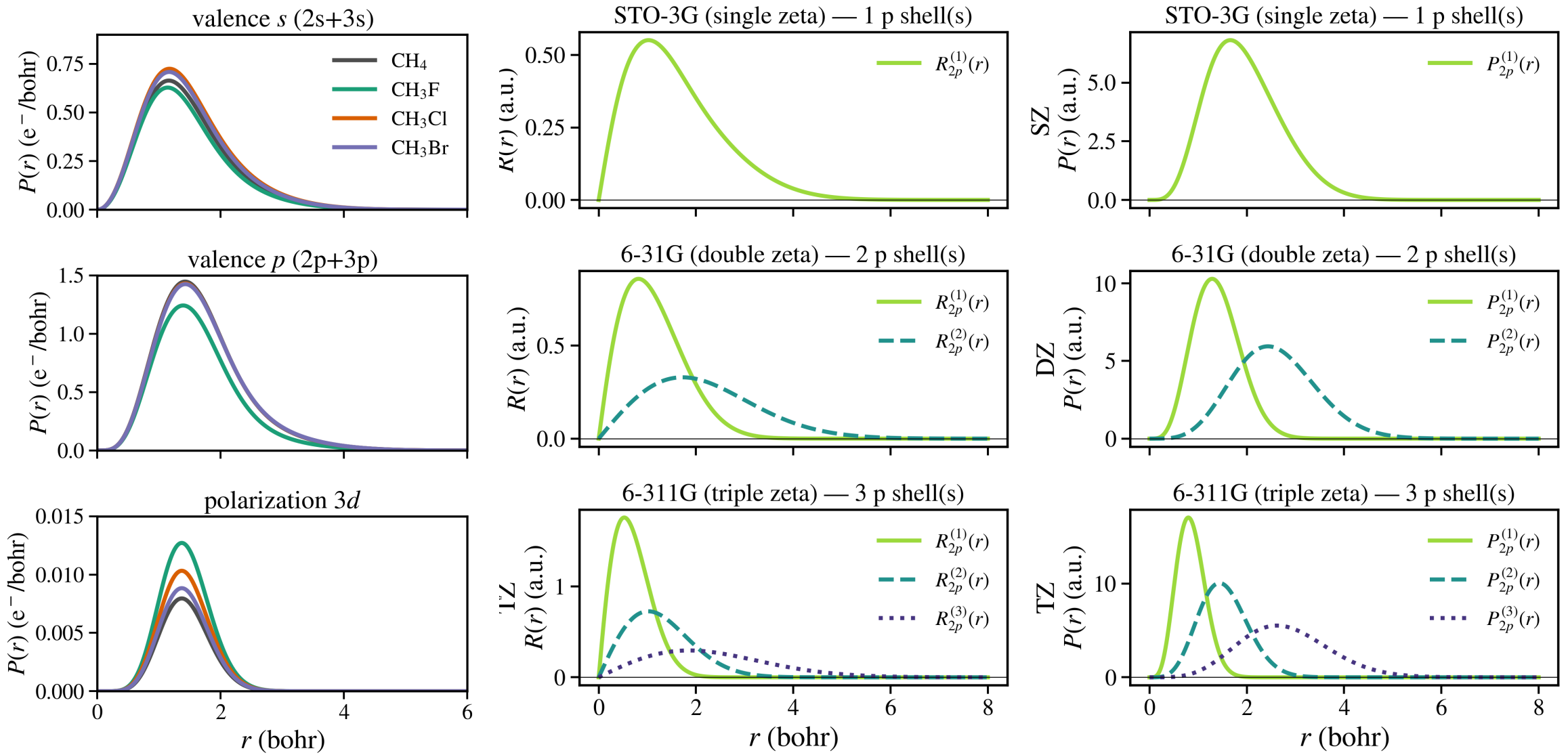






■ Atomic Orbital Basis Function

➢ Contracted Multi-Zeta Basis Set



■ Chemical Environment-Adaptive Basis

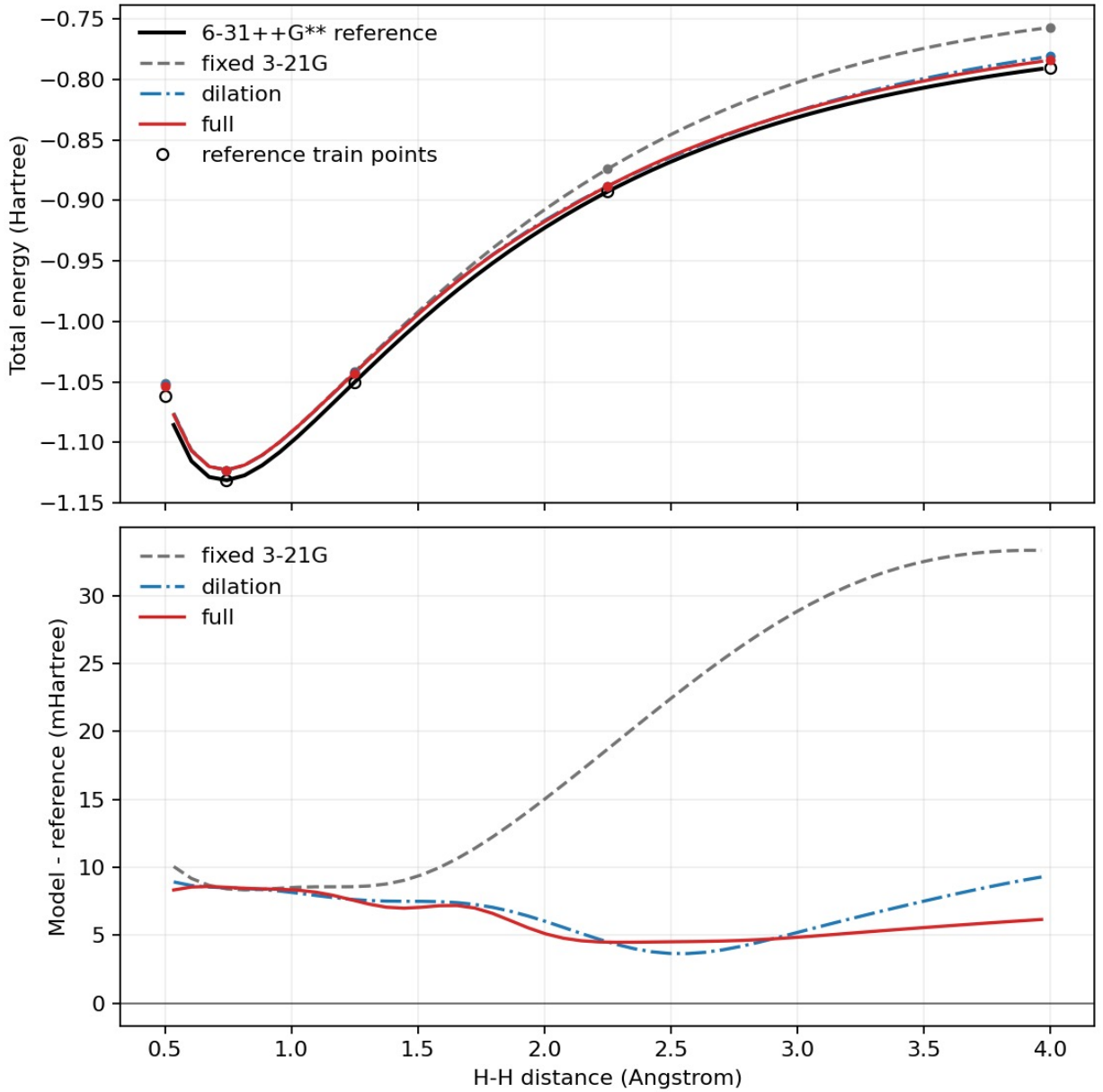
➢ 1: Neural Network Atom Basis Designing and Training

$$\phi^0_{c,\rho}(r) = \sum_p d_{cp} N_s\left(\rho_c^2 \alpha_{cp}\right) \exp\left(-\rho_c^2 \alpha_{cp} r^2\right)$$

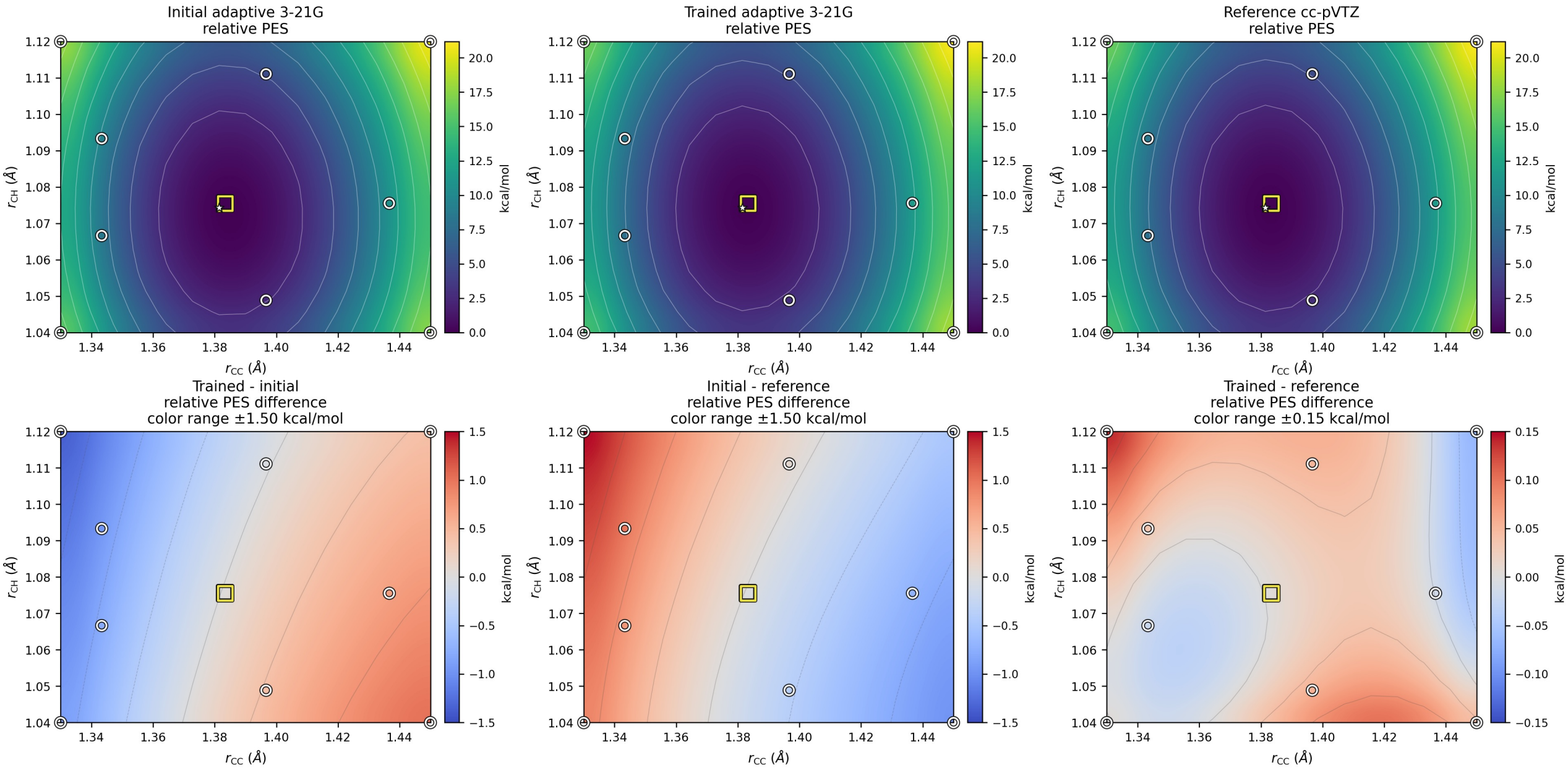
$$\rho_{Ac} = \exp[(\ln 2)\tanh(\eta_{Ac})], \qquad \rho_{Ac} \in (0.5,2)$$

$$\eta_{Ac} \rightarrow \{\phi^0_{c,\rho}(r)\} \rightarrow \{S, h, (\mu\nu|\lambda\sigma)\} \rightarrow SCF \rightarrow observables$$

atom	shell	primitive	basis cofficient / $bohr^{-2}$	model
C	core S(3)	3	172.256, 25.9109, 5.53335	fixed
C	valence S(2)	2	3.66498, 0.770545	$\times(\lambda_C^2)$
C	valence S(1)	1	0.195857	$\times(\lambda_C^2)$
C	valence P(2)	2	3.66498, 0.770545	$\times(\lambda_C^2)$
C	valence P(1)	1	0.195857	$\times(\lambda_C^2)$
H	S(2)	2	5.447178, 0.824547	$\times(\lambda_C^2)$
H	S(1)	1	0.183192	$\times(\lambda_C^2)$

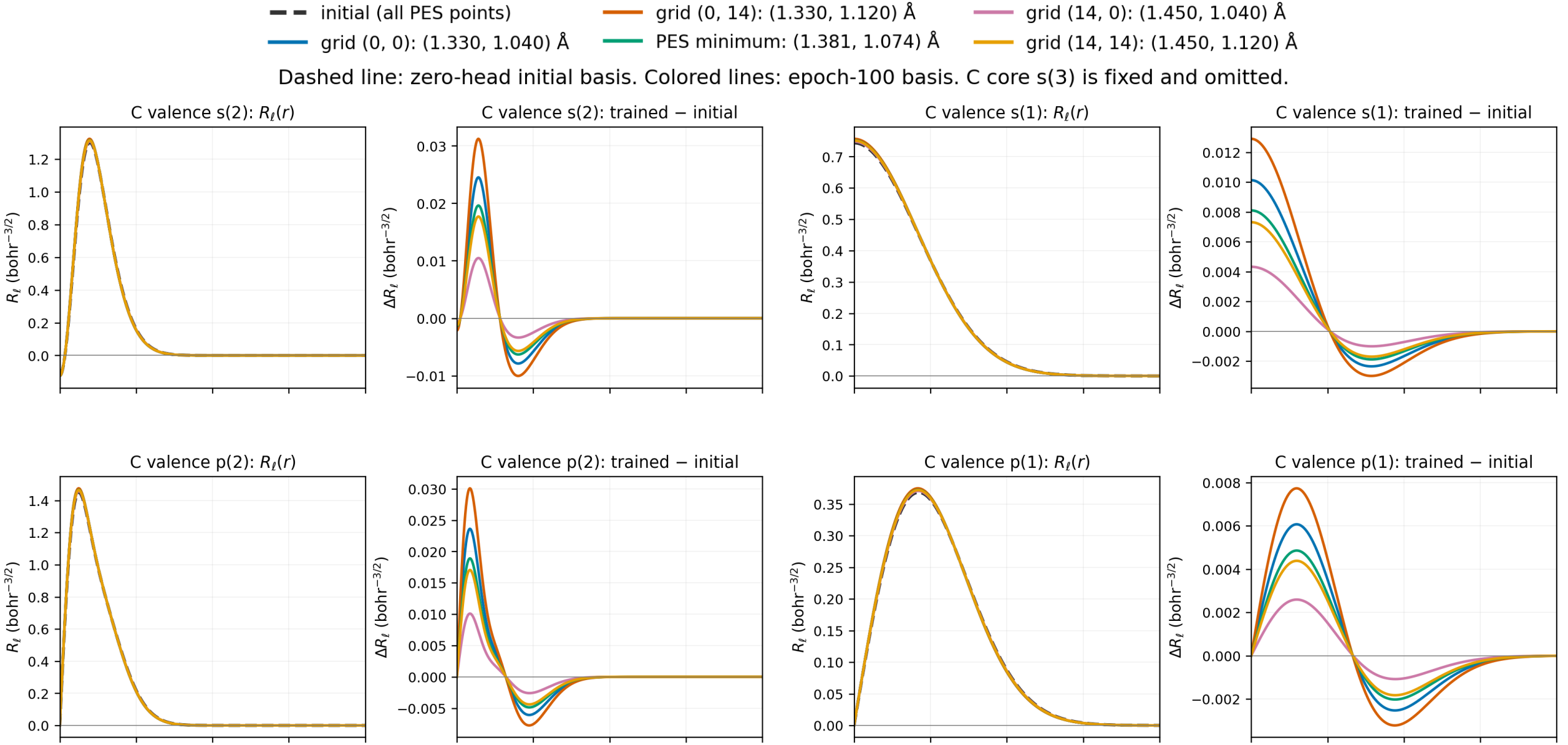


■ Chemical Environment-Adaptive Basis



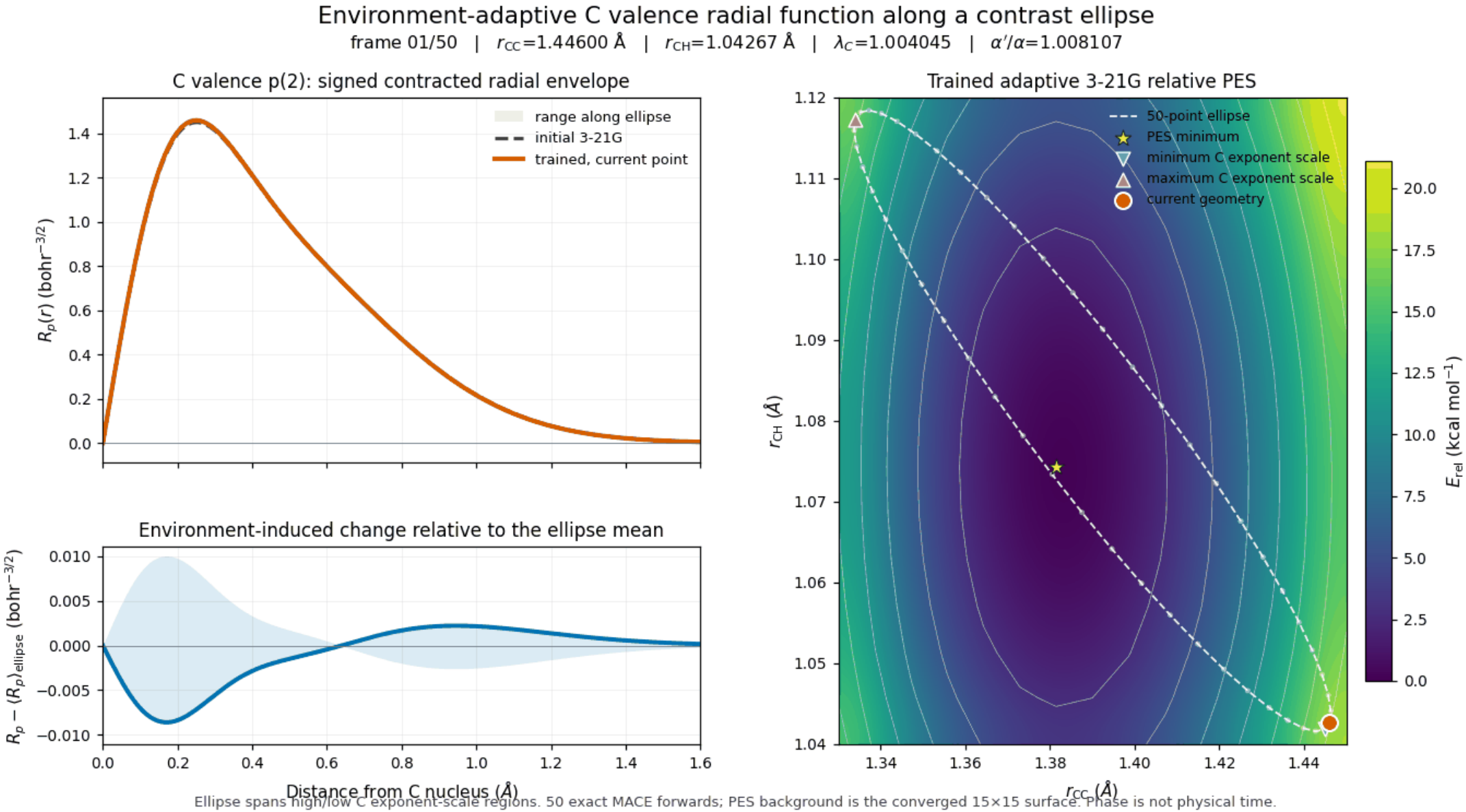
■ Chemical Environment-Adaptive Basis

➢ 1: Neural Network Atom Basis Designing and Training

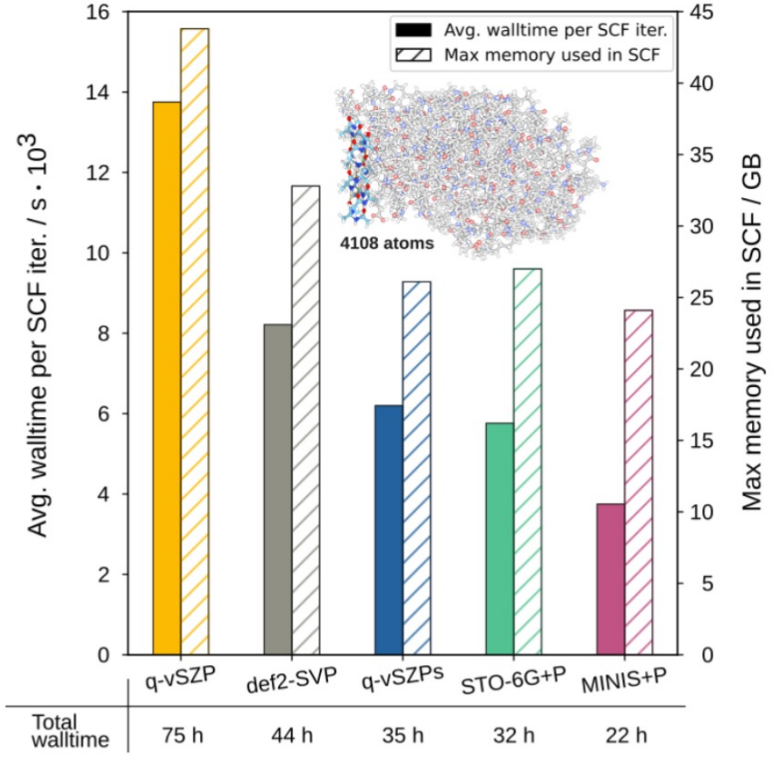


■ Chemical Environment-Adaptive Basis

➢ 1: Neural Network Atom Basis Designing and Training

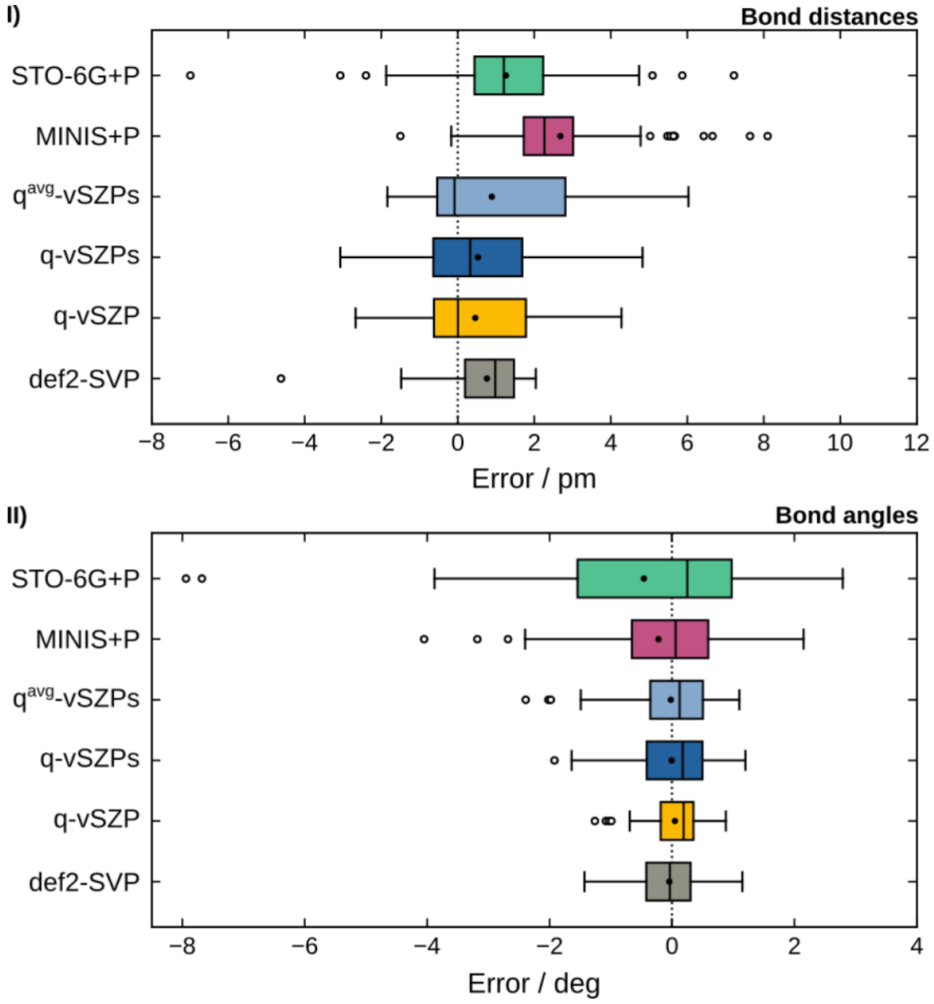


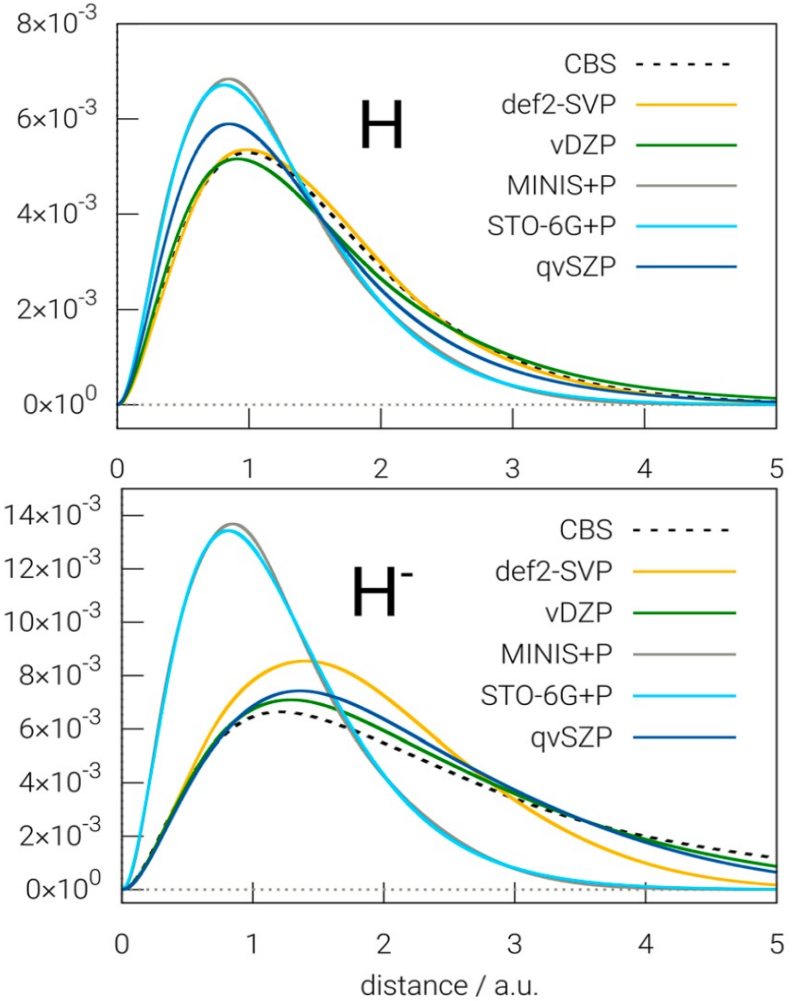
■ Chemical Environment-Adaptive Basis q-vSZP/q-vSZPs



$$\boldsymbol{R}_{A\ell}(\boldsymbol{r}; q_A^{\mathrm{eff}}) = \sum_p{}' \, \mathrm{C}_{\mathrm{p,A}}(q_A^{\mathrm{eff}}) \, N_{\ell p} r^{\ell} e^{-\alpha_{\ell p} r^2}$$

$$\mathrm{C}_{\mathrm{p,A}}(q_A^{\mathrm{eff}}) = \mathrm{C}_{\mathrm{p,A}}(q_A^{\mathrm{eff}}) + C_{\mathrm{p},0}$$





J. Chem. Phys. 165, 054109 (2026)

J. Chem. Phys. 159, 164108 (2023)

■ Chemical Environment-Adaptive Basis EGNN for Contraction Coefficients

➢ SO(3) Convolution Filter

$$F_m^{l_{in}l_f}(\boldsymbol{r_{ij}}) = R^{l_{in}l_f}\left(|\boldsymbol{r_{ij}}|\right)Y_m^{l_f}\left(\frac{\boldsymbol{r_{ij}}}{|\boldsymbol{r_{ij}}|}\right)$$

$Y_m^{l_f}\left(\frac{r_{ij}}{|r_{ij}|}\right)$ →**Harmonic Spherical Function**

$R^{l_{in}l_f}\left(|\boldsymbol{r_{ij}}|\right)$ →**Feed Forward Network**

➢ SelfNet Layer

$$f_{ii}^{l_{out}} = \left(\boldsymbol{W^{l_{in_r}l_{in_l}l_{out}}xl_i^{l_{in}} \otimes xr_i^{l_{in}}}\right)^{l_{out}}$$

$$m_{ijc}^{l_{out}} = \sum_{ll}\left(F_c^{l_{in}l_f l_{out}}\left(r_{ij}, \frac{r_{ij}}{|r_{ij}|}\right) \otimes \left(x_{jc}^{l_{in}}\right)\right)^{l_{out}}$$

➢ PairNet Layer

$$f_{ij}^{l_{out}} = \left(\boldsymbol{F^{l_{in_r}l_{in_l}l_{out}}(r_{ij})x_i^{l_{in_i}} \otimes x_j^{l_{in_j}}}\right)^{l_{out}}$$

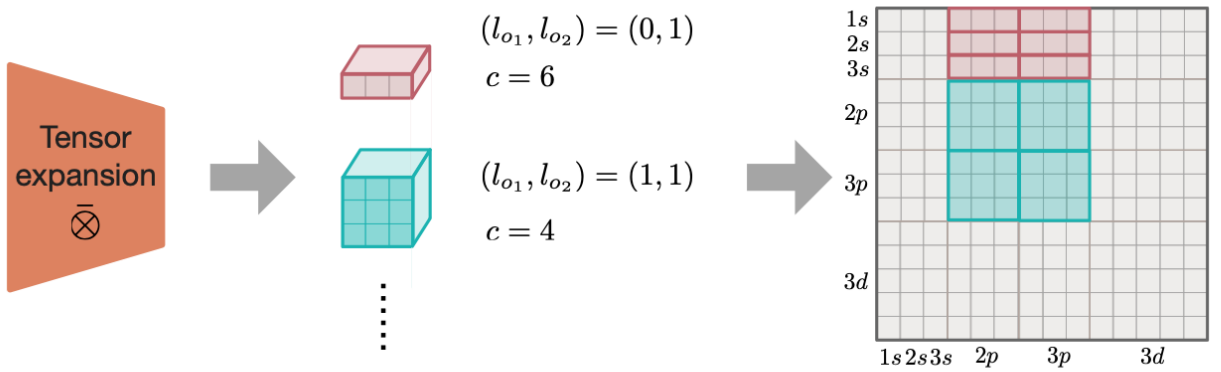
$$\tilde{x_i} = f_{linear}(x_i + \sum_j m_{ij})$$

➢ Expansion Layer

$$\left(\overline{\otimes} f_{ij}^{l_3}\right)_{m_1,m_2}^{l_1,l_2} = \sum_m C_{l_1m_1,l_2m_2}^{l_3m_3} f_{ij}^{l_3}$$

Hamiltonian Learning ✅

Neural Network Basis Set ❌



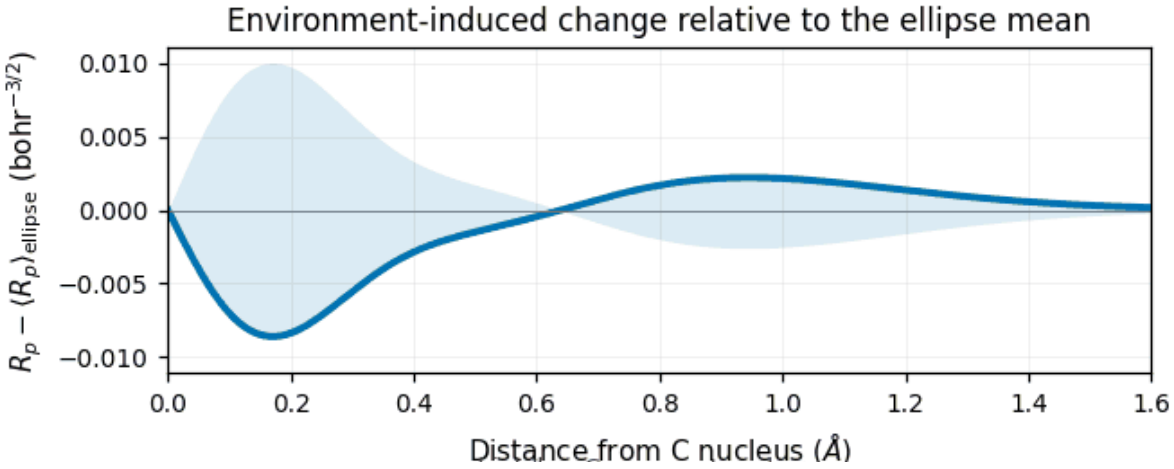
■ Functional

● Functional-GradTDDFT

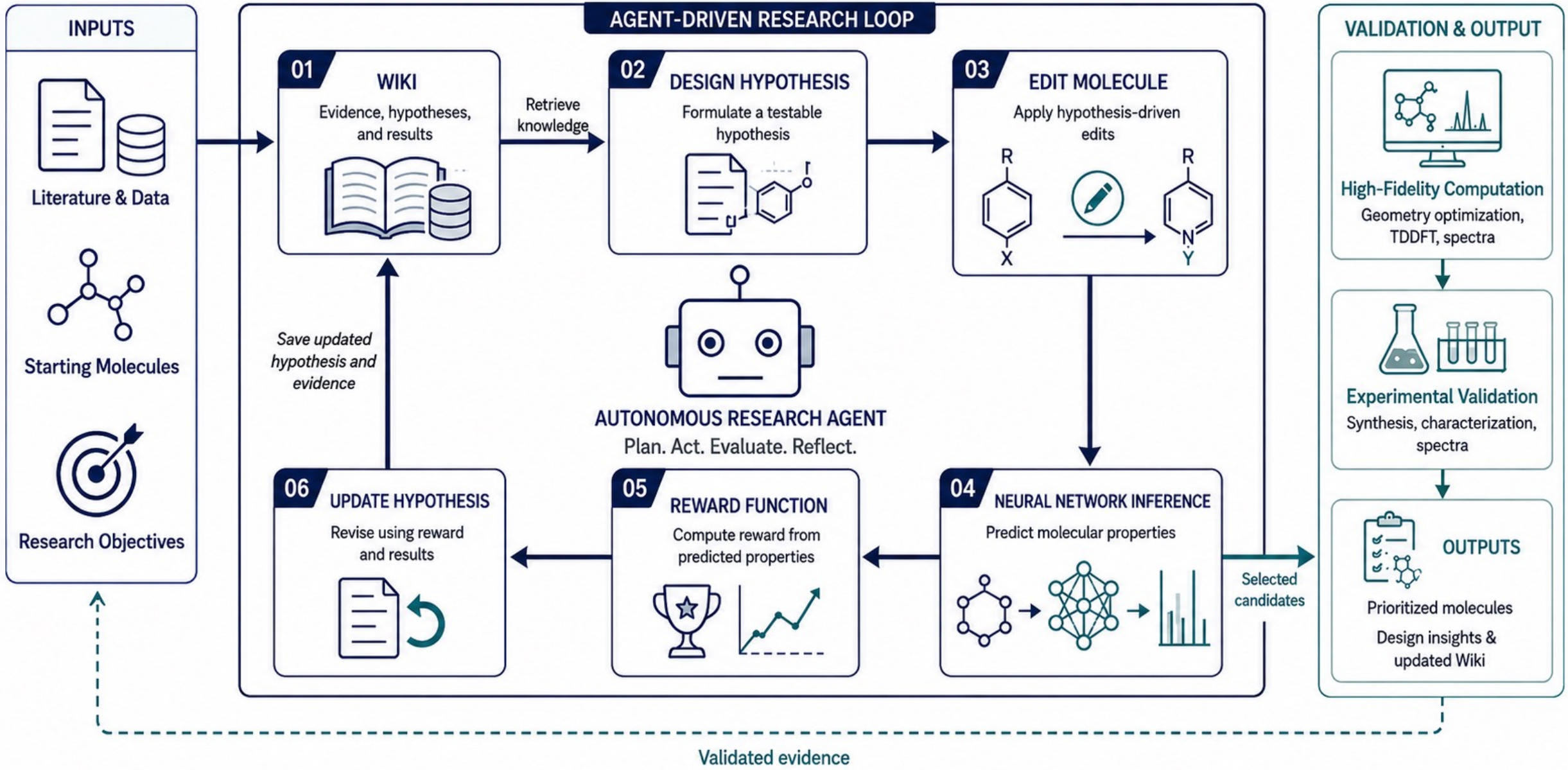
- Differential SCF code for DFT/HF and implicit differential Davidson solver code for TDDFT
- Basis Set Library and Electronic Integral Module With Jax Based Contraction Operator
- Neural Network for Optimization of Gaussian-Type Orbital

● Outlook

- Highly Efficient Integral Module for AD
- SE(3) Equivariant Neural Network Basis Set
- Neural Network Based NAO(More Flexible Form)
- Form GradTDDFT to GradSCF (Electronic Structure Infra for AI)



■ Background - Molecular Designing



■ Molecular Edition Protocol

➢ Basic Operation

■ Atom operations

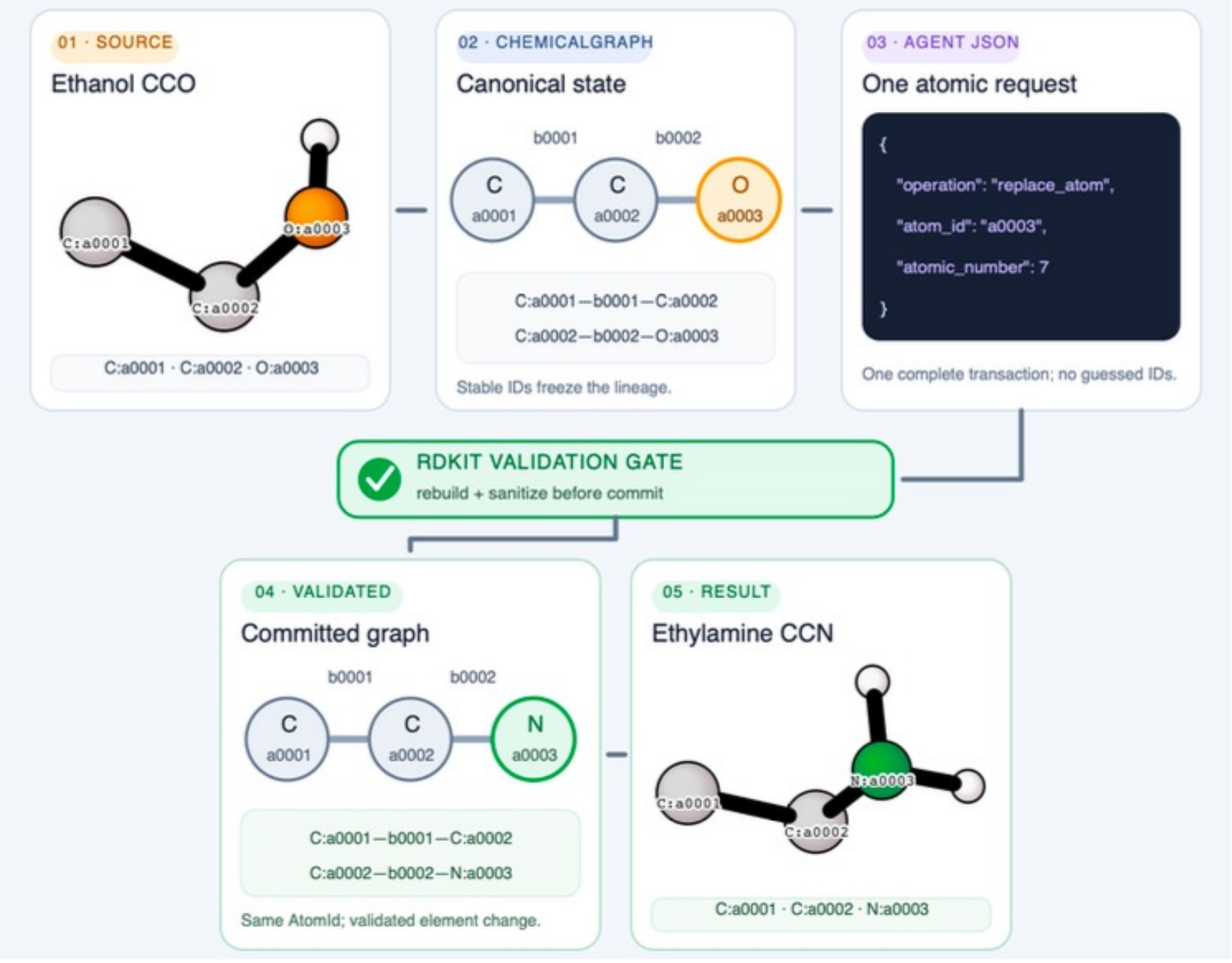
- Add Atom
- Remove Atom
- Replace Atom

■ Bond operations

- Add Bond
- Remove Bond
- Change Bond

■ Fragment operations

- Attach Fragment
- Detach Fragment
- Substitute Fragment



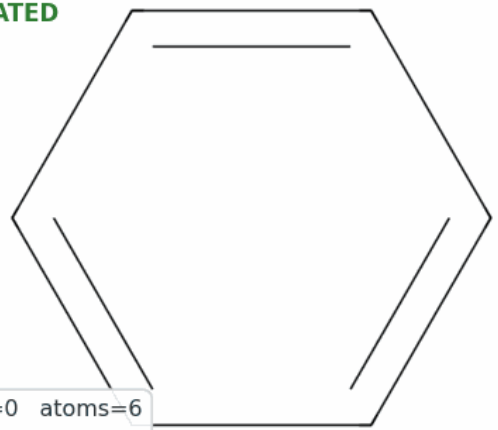
■ Molecular Edition Protocol

Benzene -> all recorded attempts | frame 1/101 | CALCULATED

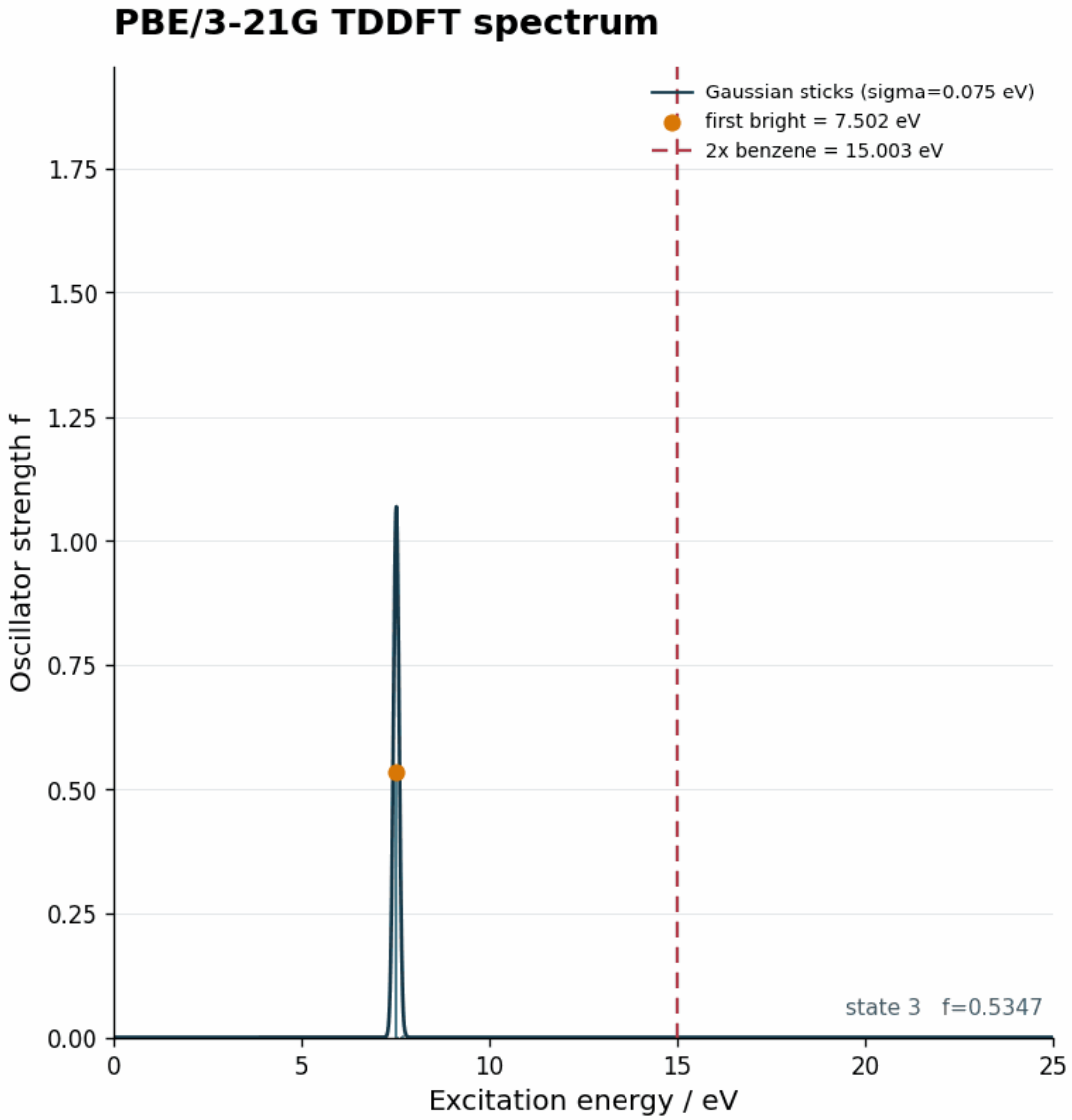
Molecule | MAIN LINE

parent: step-000-benzene

CALCULATED

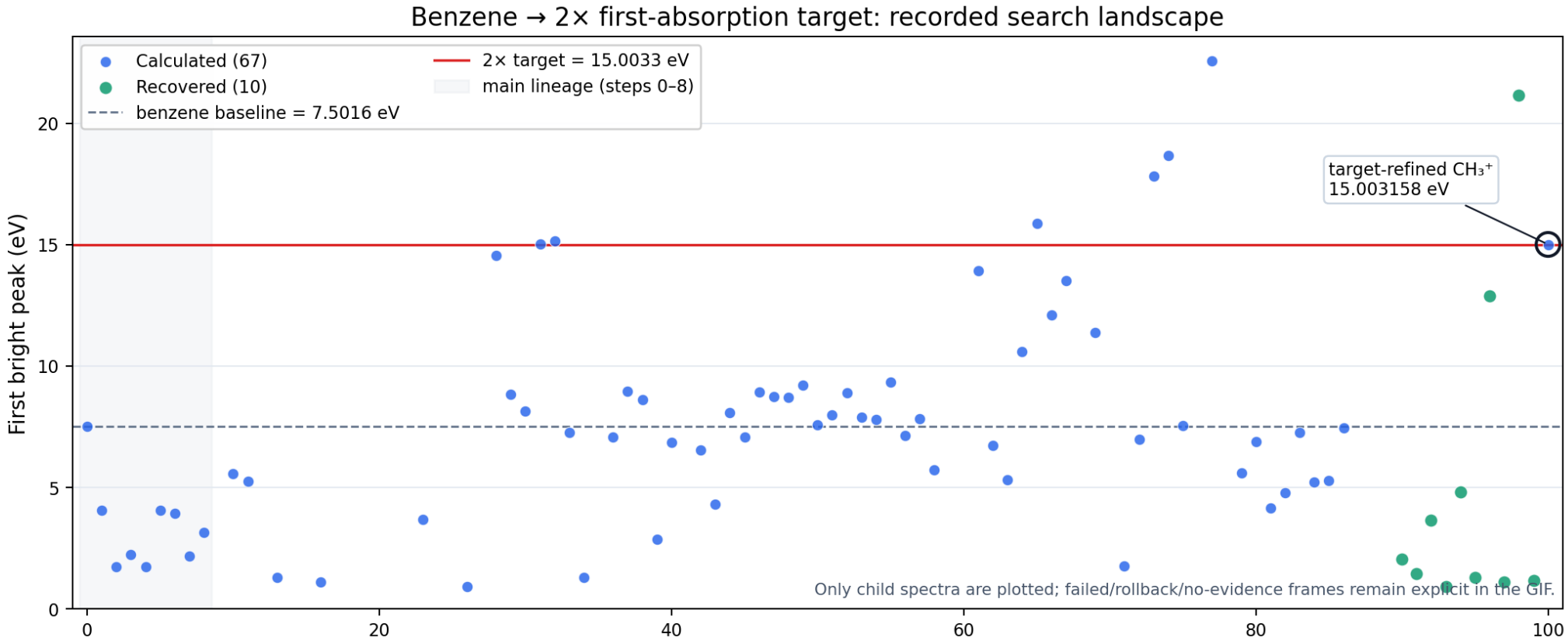


C6H6 q=0 atoms=6
benzene
edit: initial_structure



Every frame is a recorded attempt; branch parent is explicit. Failed or rolled-back children do not receive invented spectra.

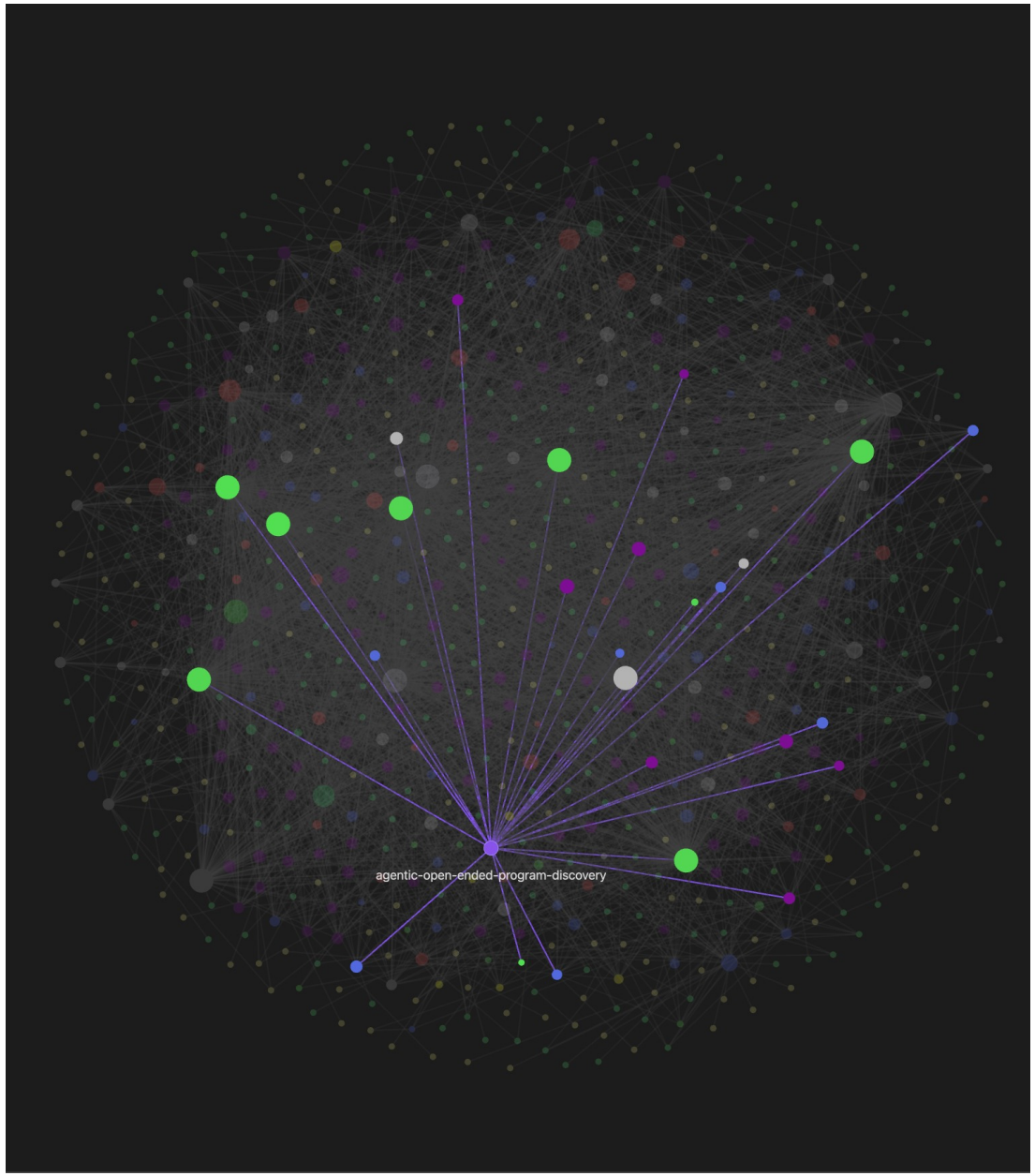
■ Molecular Edition Protocol



Hypothesis from LLM-Wiki (BTW)

```
.wiki/
├── raw/       immutable local sources; never published by this repository
├── wiki/
│   ├── sources/
│   ├── concepts/
│   ├── topics/
│   ├── queries/
│   ├── relations.md
│   └── method-improvements.md
└── output/    derived local artifacts
```

```
PDF reader + existing wiki
        |
        v
Evidence Pack
        |
        +--> Organizer: source draft
        |
        +--> Questioner: 3 questions by default
                         |
                         v
Round 1: Organizer answers
         -> Evaluator scores/feedback
Round 2: Organizer revises
         -> Evaluator scores/feedback
Round 3: Organizer revises
         -> Evaluator final score
                         |
                         v
Main Agent validates and writes:
source page + agent-queries page + relations/topic/index/log
```



Github.com/STOKES-DOT/oh-my-llm-wiki

■ Agent for Hypothesis, Molecular Edition and Evolution

Parent molecule

CC(=O)C=Cc1c(N(C)C)cc2c(=O)c3c(C=CC=O)cccc3n3c4ccccc4c(=O)c1c23

The parent contains two conjugated peripheral arms:

- an acetyl-vinyl arm, CC(=O)C=C–;
- an aldehyde-vinyl arm, –C=CC=O.

The proposed edit retains the first arm and replaces the second one.

Hypothesis

Replace the aldehyde-vinyl arm –C=CC=O with a stronger conjugated dicyanovinyl acceptor, –C=C(C#N)C#N, while preserving the fused polycyclic chromophore and the existing dimethylamino donor.

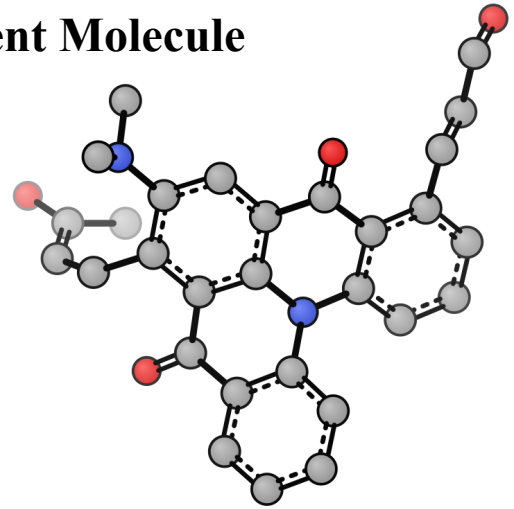
The two cyano groups are expected to strengthen the electron-withdrawing character of the side chain. Coupling this acceptor to the π-conjugated core may increase donor–acceptor polarization, stabilize the relevant excited state, reduce the effective optical gap, and red-shift the first absorption maximum toward the target range of 620–750 nm.

The hypothesis is falsified at two levels:

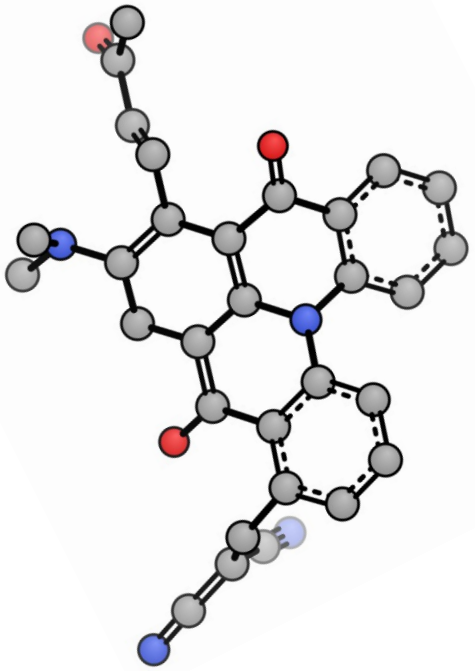
1. **Directional falsification:** the absorption maximum does not shift to a longer wavelength.
2. **Target-level falsification:** the absorption maximum red-shifts but remains substantially below 620 nm.

PLQY and log10(ε) remain weak secondary objectives and are not required to improve simultaneously.

Parent Molecule



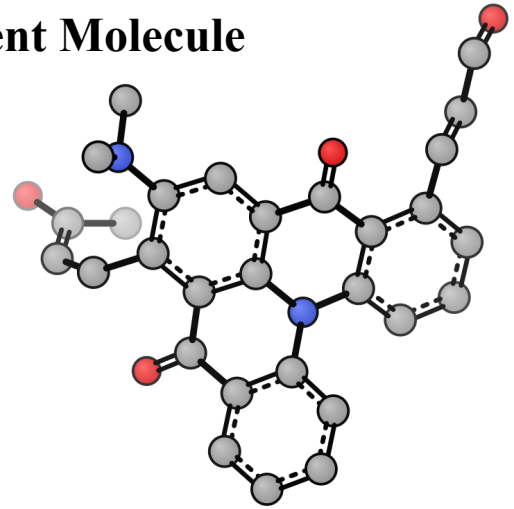
Edited Molecule



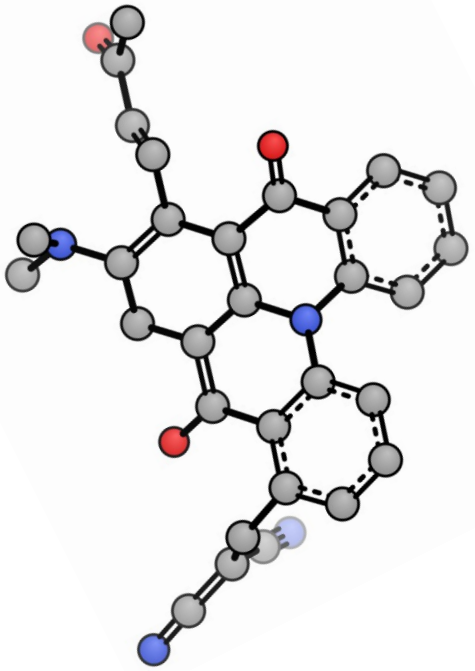
■ Agent for Hypothesis, Molecular Edition and Evolution

```
### Proposed molecular edit

Parent side chain:
core-C=CH-CHO

          substitute_fragment
                   ↓

Designed side chain:
core-C=C(C#N)2

More explicitly:

...c3c(C=CC=O)cccc3...
              ↓
...c3c(C=C(C#N)C#N)cccc3...

This is a substitute_fragment operation rather than a single-atom replacement:

Operation:              substitute_fragment
Cleaved bond:           b0028
Retained parent atom:   a0024
New attachment bond:    DOUBLE
Added fragment:         C3N2
Complete edited arm:    C=C(C#N)C#N

The complete dicyanovinyl arm contains six heavy atoms, one of which is retained from
the parent structure; the newly attached fragment contributes five heavy atoms.
```

Parent Molecule



Edited Molecule

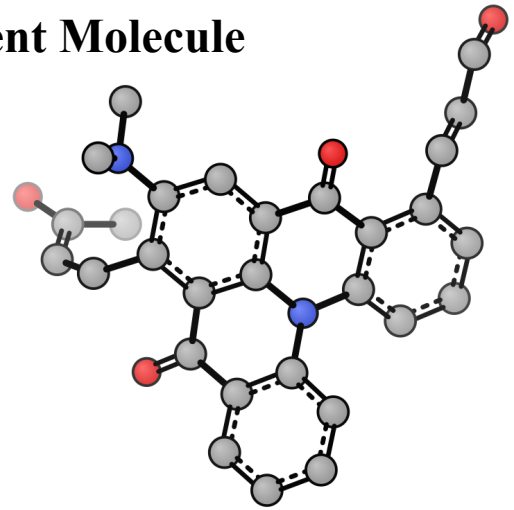


■ Agent for Hypothesis, Molecular Edition and Evolution

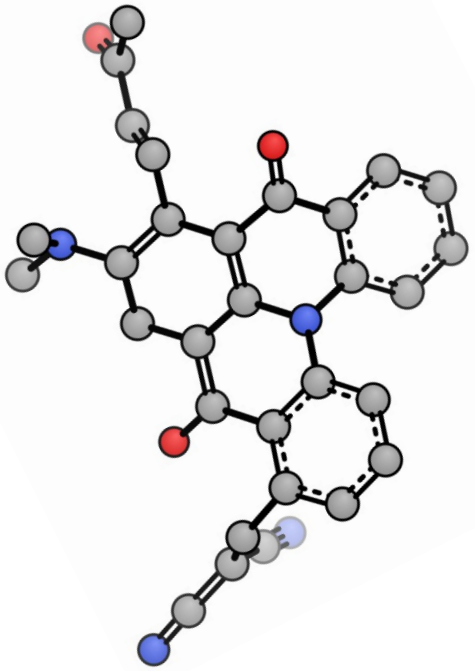
Results

Property	Parent	Edited molecule	Change
Absorption maximum, λabs	501.981 nm	**525.261 nm**	**+23.280 nm**
Distance from 620 nm	118.019 nm	**94.739 nm**	Improved by 23.280 nm
Emission maximum, λemi	483.260 nm	478.009 nm	−5.251 nm
PLQY	0.2088	0.0976	Decreased
Molar extinction coefficient	26,366.4 M^{-1} cm^{-1}	29,303.6 M^{-1} cm^{-1}	Increased
log10(ε)	approximately 4.421	4.467	Slight increase
Fitness	−0.904427	**−0.725828**	Improved
Parent similarity	–	**0.7385**	Main scaffold retained

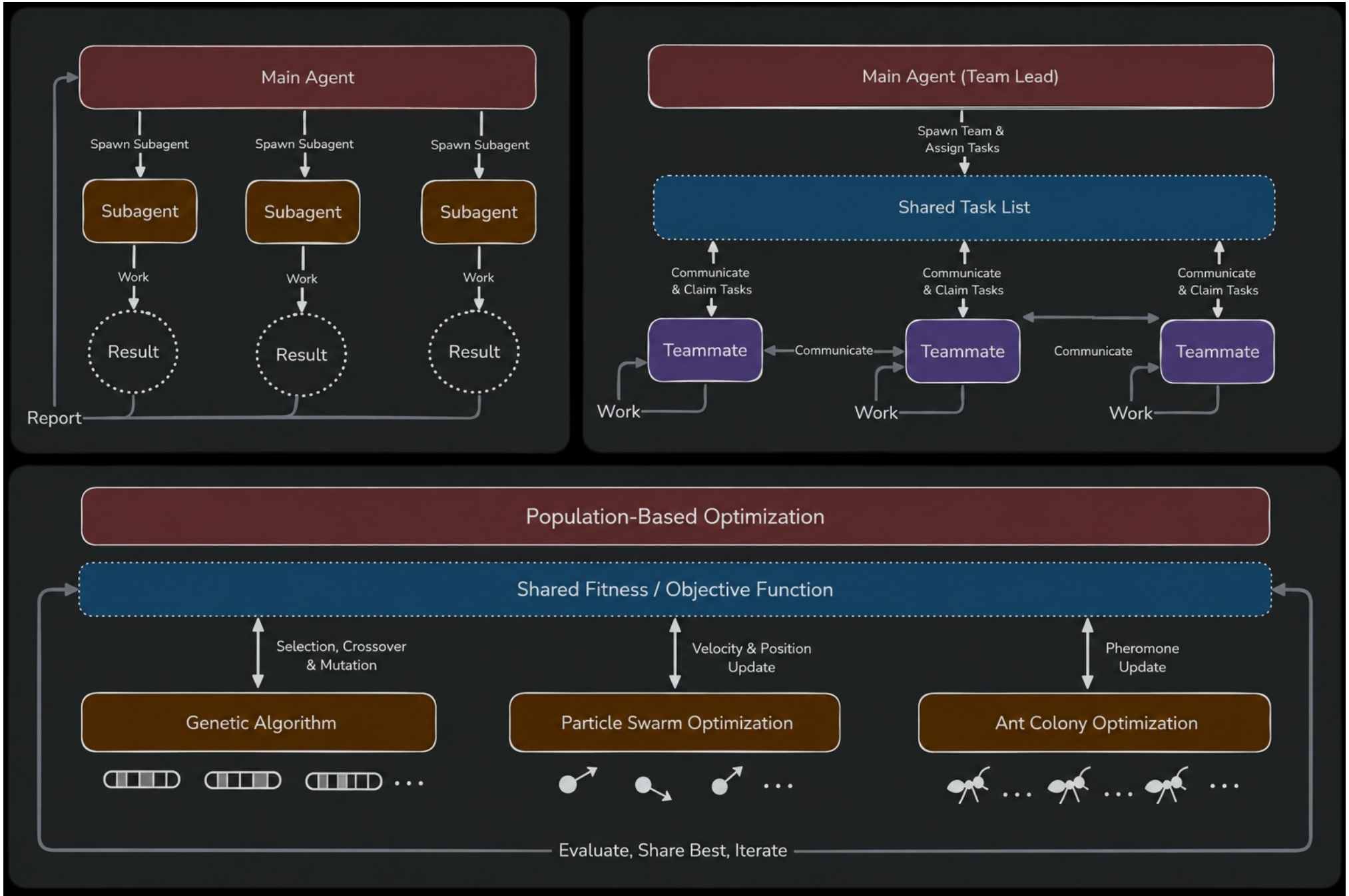
Parent Molecule



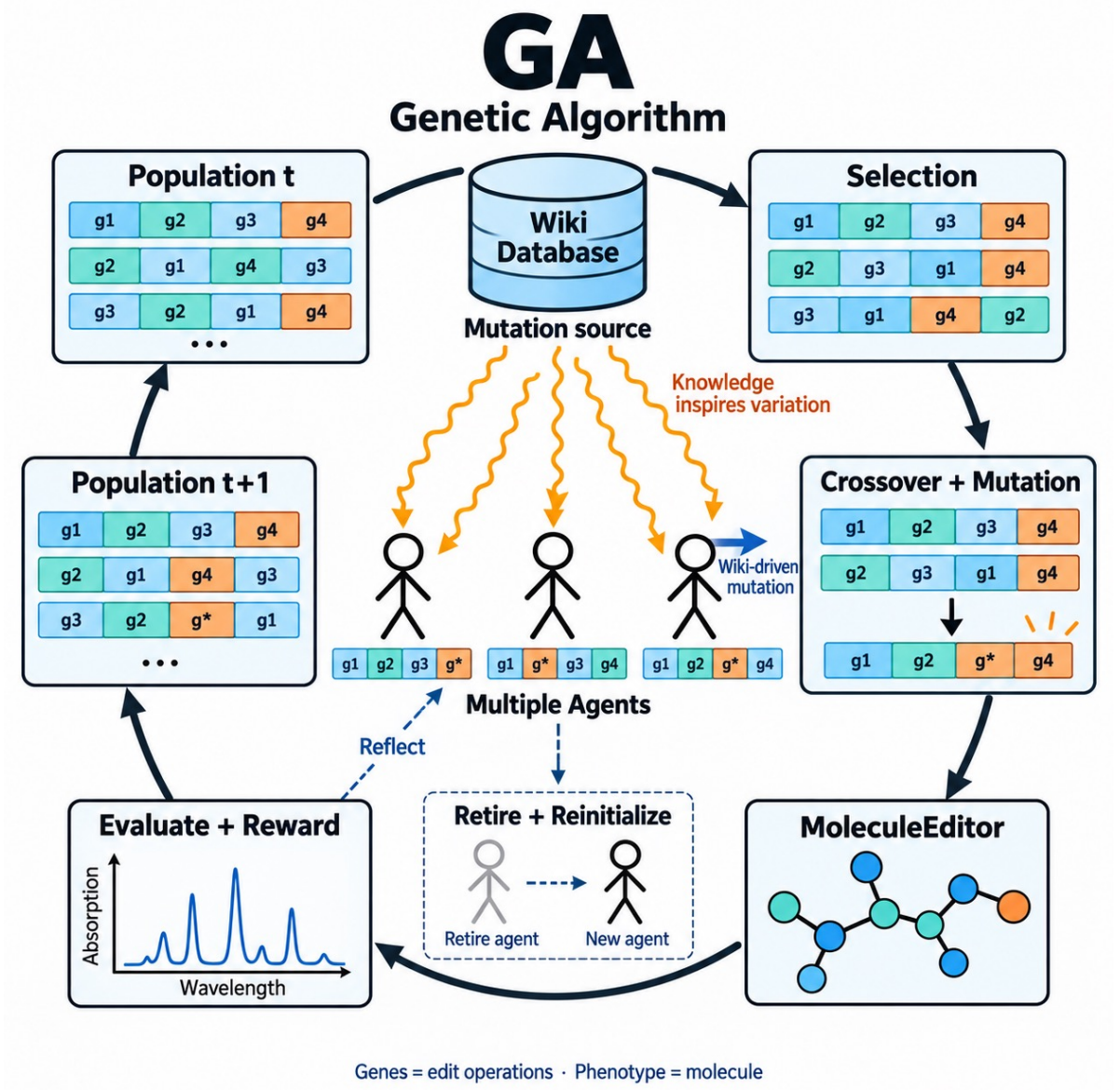
Edited Molecule

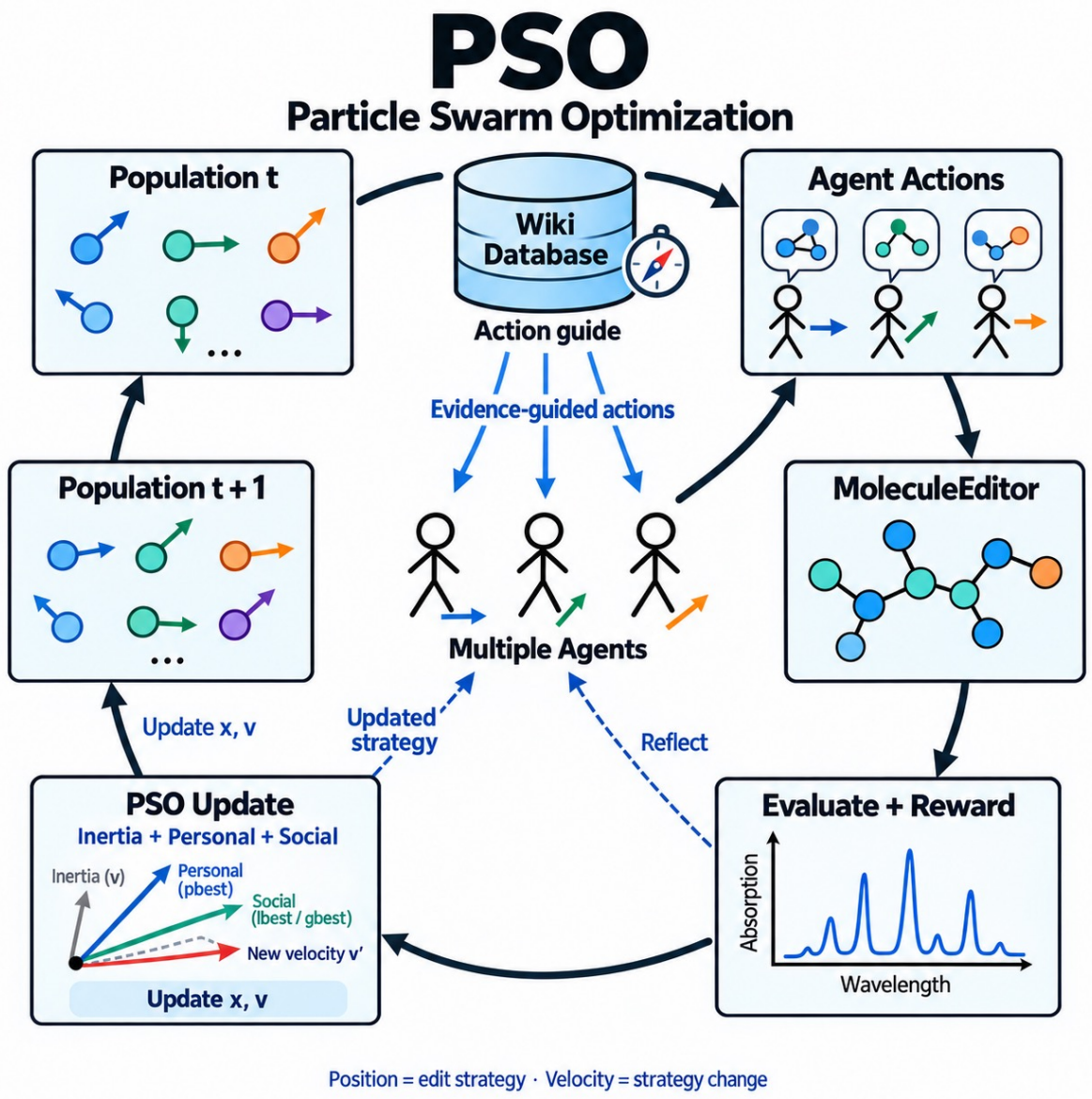


■ Agent for Hypothesis, Molecular Edition and Evolution



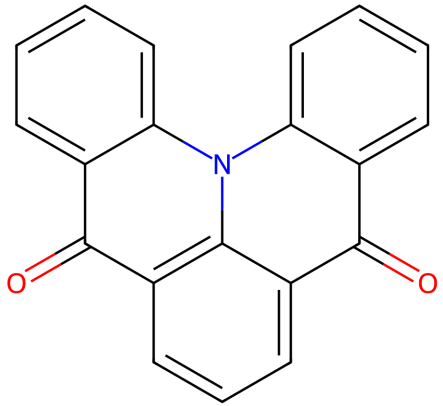
■ Agent for Hypothesis, Molecular Edition and Evolution





Github.com/STOKES-DOT/Multi-Agent-Swarm

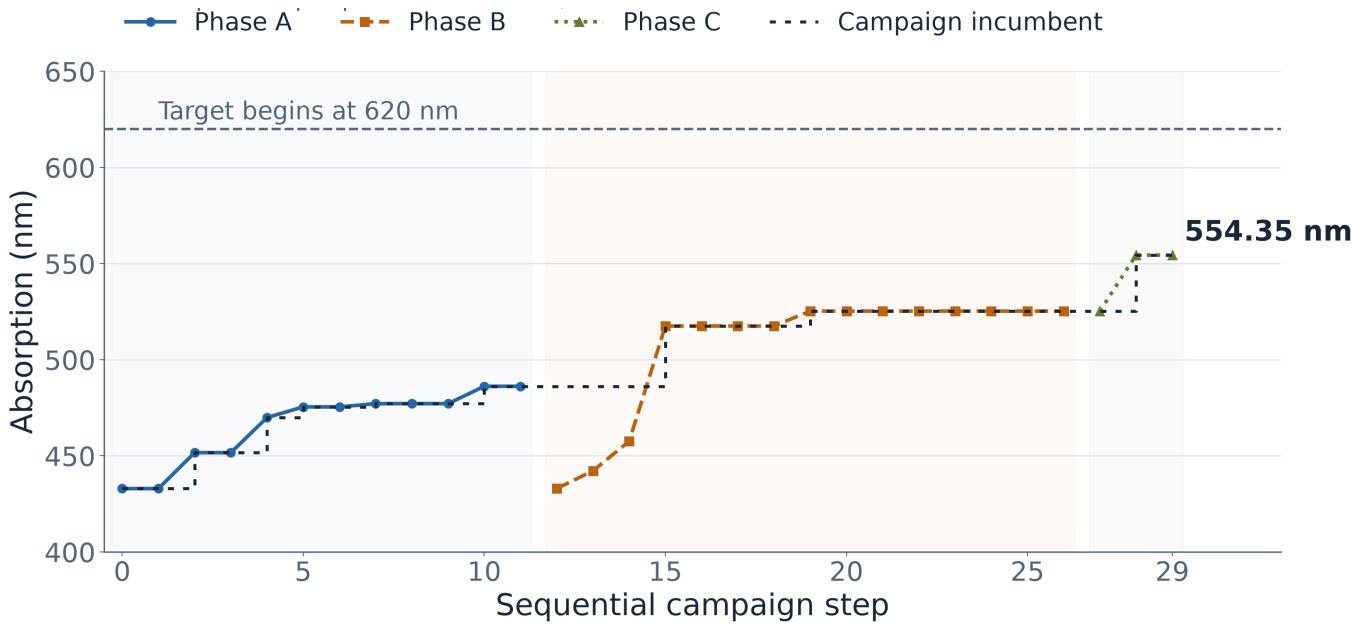
■ Agent for Hypothesis, Molecular Edition and Evolution-PSO

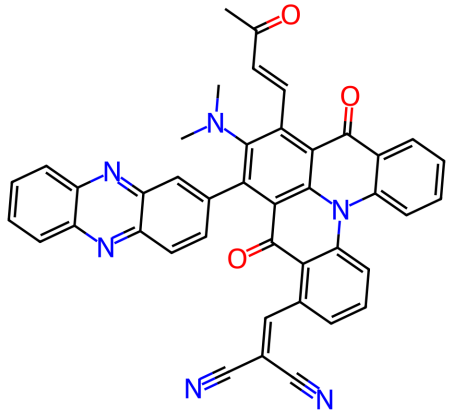


Initial DiKTa

Absorption: 432.94 nm

Reward: -1.434828

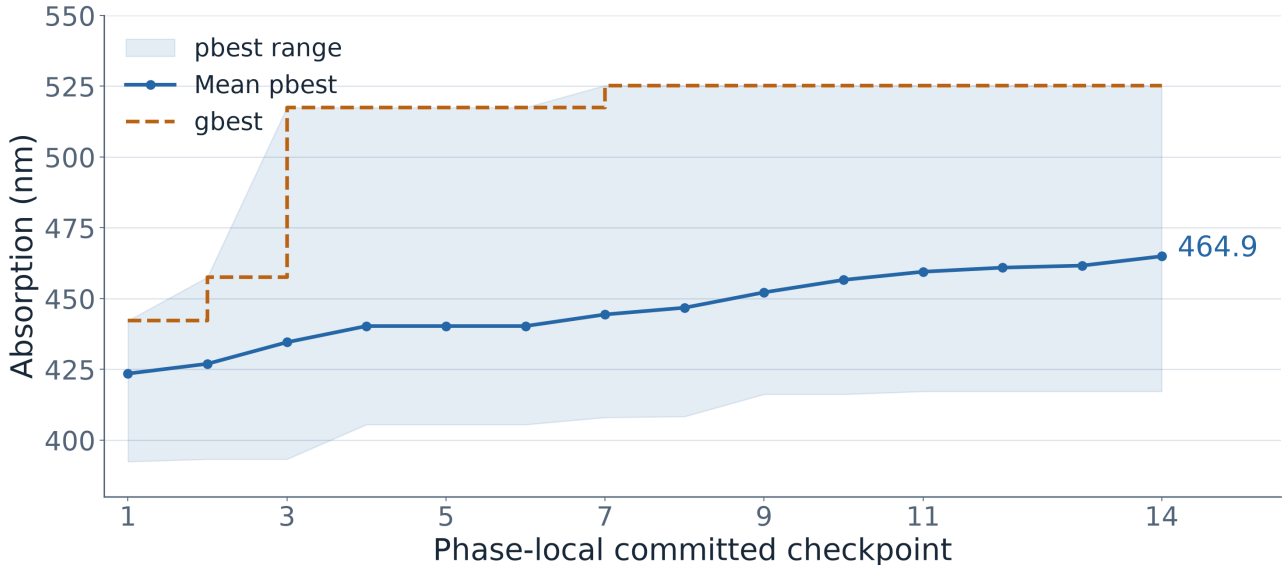




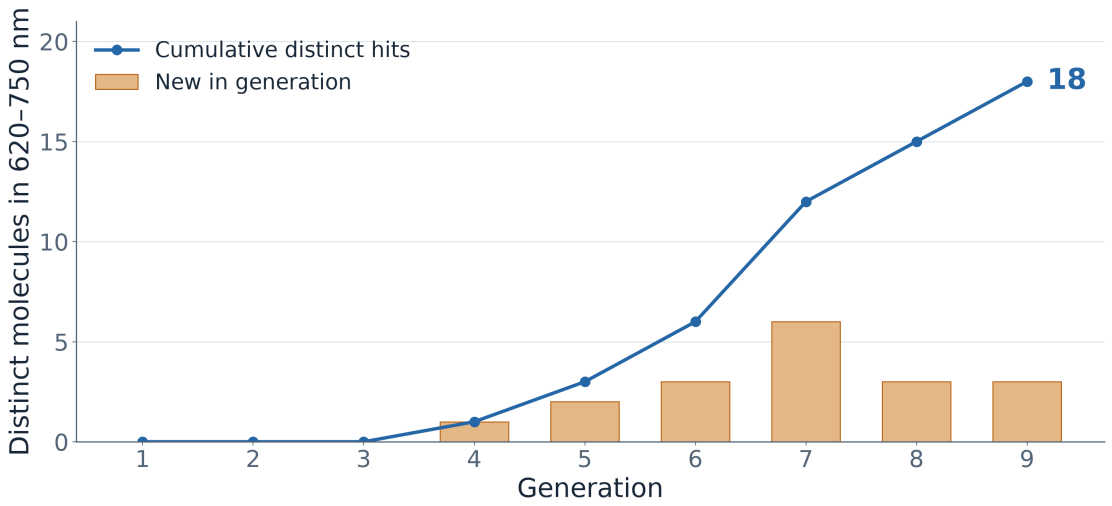
Phase C endpoint

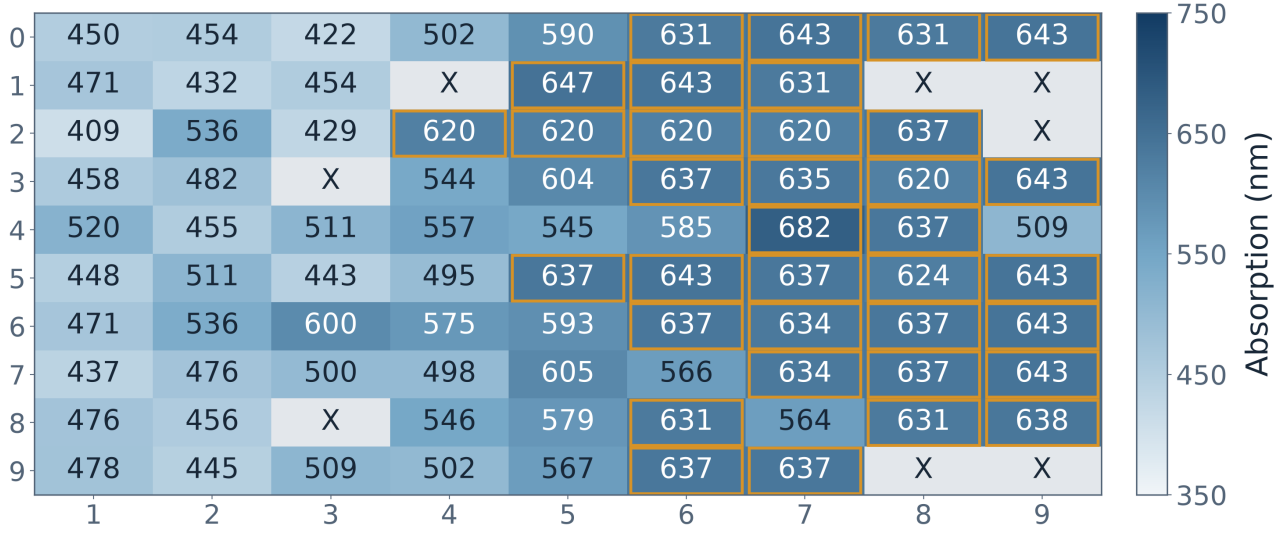
Absorption: 554.35 nm

Reward: -0.502542

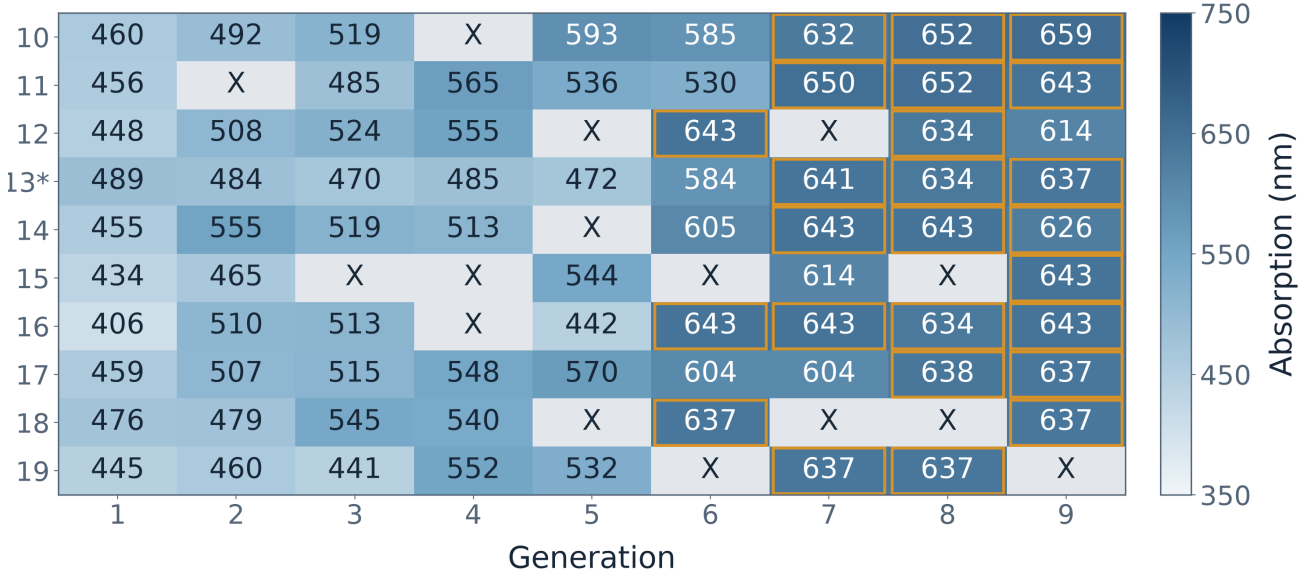


■ Agent for Hypothesis, Molecular Edition and Evolution-GA

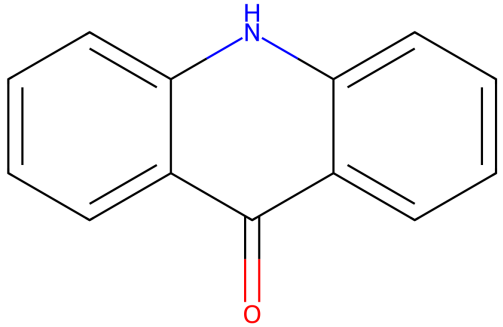




Rank	Abs. (nm)	Emi. (nm)	PLQY	log₁₀ ε	Reward
1	620.32	669.59	0.535	4.296	1.004832
2	635.15	667.76	0.590	4.123	1.004824
3	643.46	658.90	0.542	4.001	1.004378
4	638.47	670.15	0.546	3.950	1.004312
5	650.32	671.72	0.514	4.018	1.004268
6	636.71	645.07	0.431	4.239	1.004222



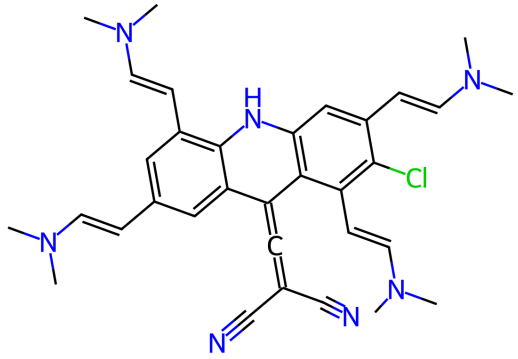
■ Agent for Hypothesis, Molecular Edition and Evolution-GA



Initial acridone

Absorption: 393.30 nm

Reward: -1.741050

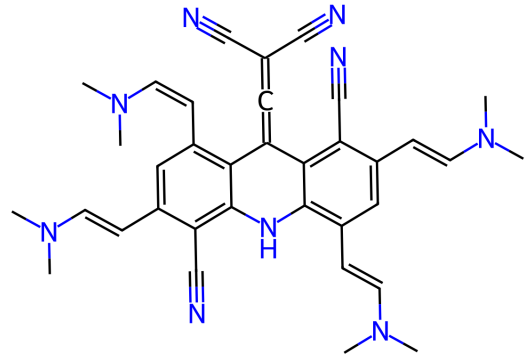


GA-09

Absorption: 638.30 nm

Emission: 678.71 nm | PLQY: 0.535

ε: 8102 M^{-1} cm^{-1} | Reward: 1.004189

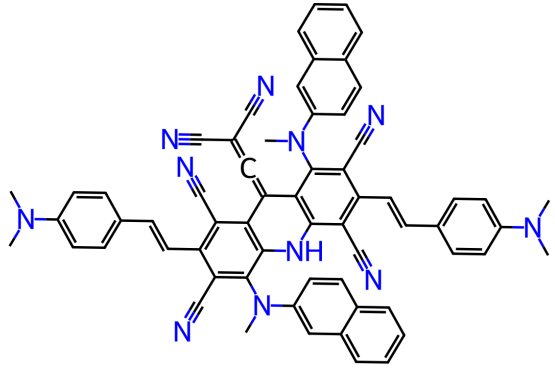


GA-10

Absorption: 651.91 nm

Emission: 693.82 nm | PLQY: 0.486

ε: 11239 M^{-1} cm^{-1} | Reward: 1.004179

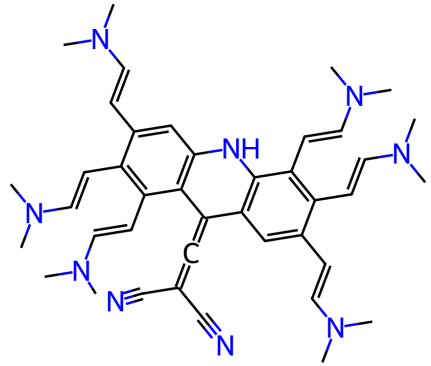


GA-17

Absorption: 646.62 nm

Emission: 772.08 nm | PLQY: 0.194

ε: 40431 M^{-1} cm^{-1} | Reward: 1.003646



GA-18

Absorption: 631.69 nm

Emission: 681.10 nm | PLQY: 0.330

ε: 13600 M^{-1} cm^{-1} | Reward: 1.003537