ChemAlive

MOLECULAR DATA FOR MATERIALS AI

Quantum chemistry as a data engine

ConstruQt, InteraQt and Qontext

Industrial use cases in molecular-property datasets and metal–ligand interactions

Peter D. Jarowski, Ph.D.
ChemAlive SA, Lausanne | October 2026



Materials AI needs reliable molecular data at scale

Define the learning task

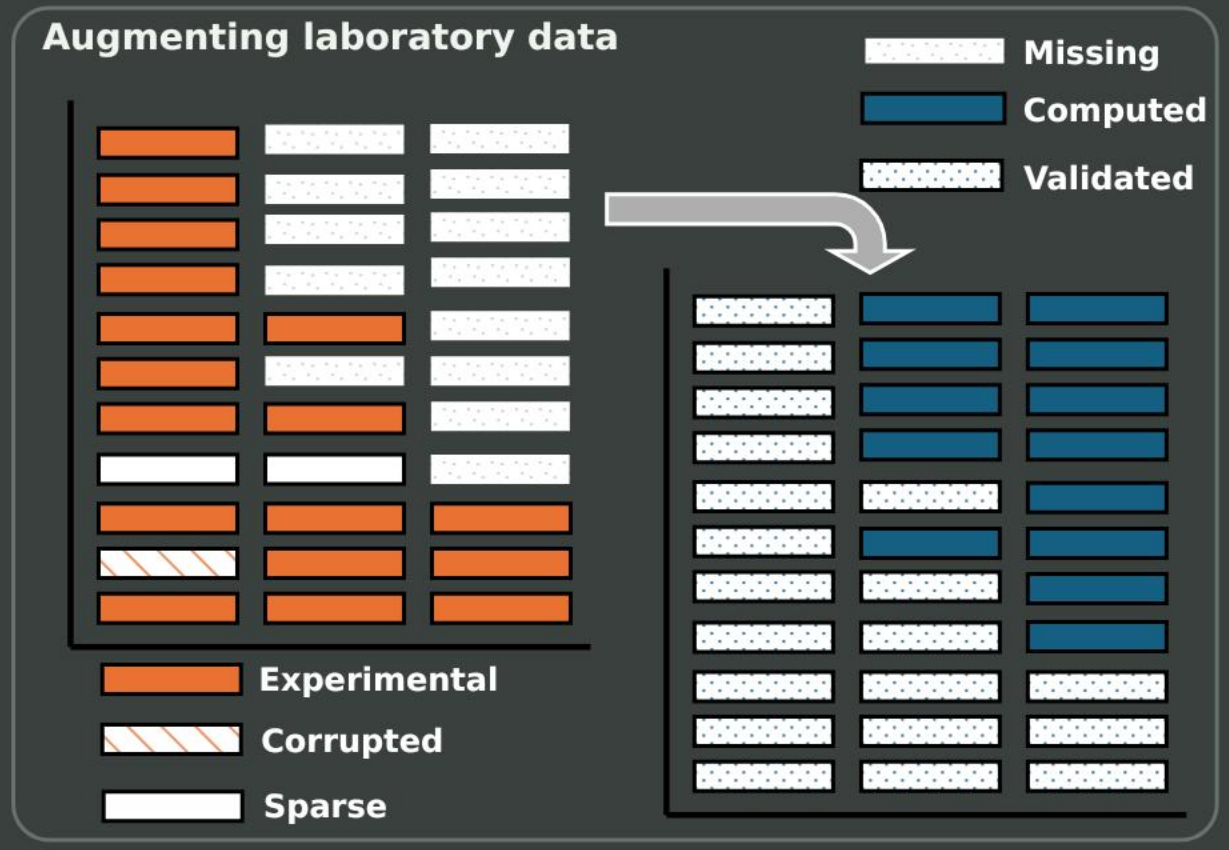
Choose the target property, chemical families and available measurements.

Compute missing physics

Generate molecular states, structures and interaction labels with a consistent protocol.

Build a reusable dataset

Retain the molecular context, quality status and method behind each label.



MATERIALS AI

Molecular library

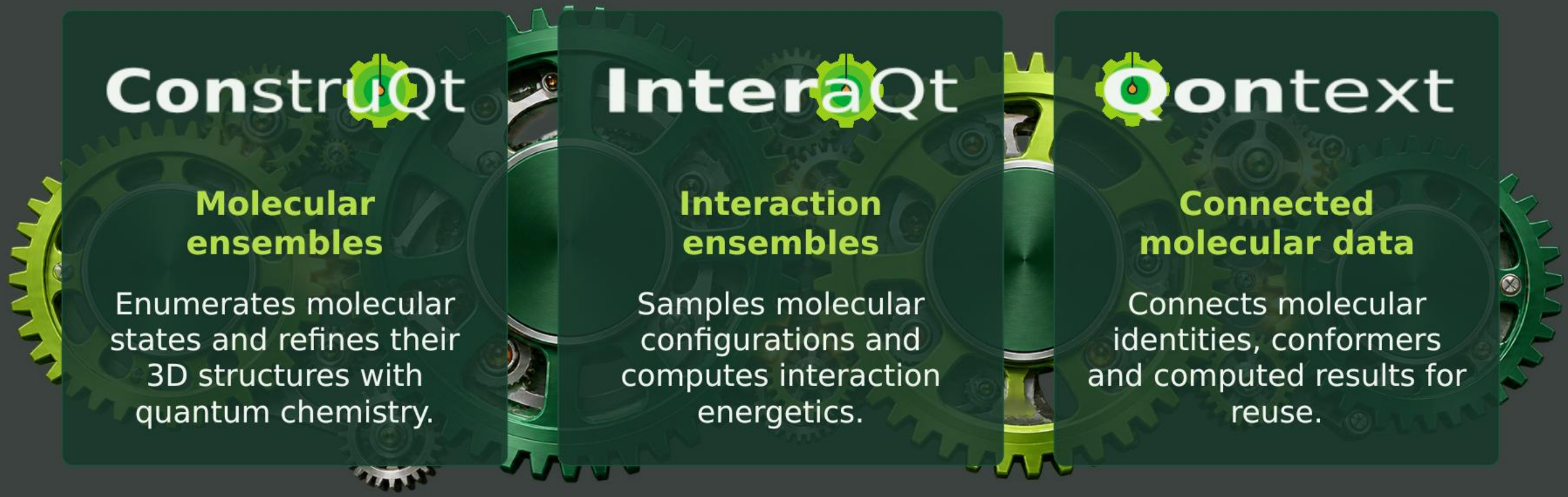
↓

Physics-based reference data

↓

Property model

A connected software stack for molecular data



ConstruQt

Molecular ensembles

Enumerates molecular states and refines their 3D structures with quantum chemistry.

InteraQt

Interaction ensembles

Samples molecular configurations and computes interaction energetics.

Qontext

Connected molecular data

Connects molecular identities, conformers and computed results for reuse.

Molecular ensembles and interaction labels retain their computational context

One chemical input expands into a molecular ensemble

Chemical input

Molecular states and conformers



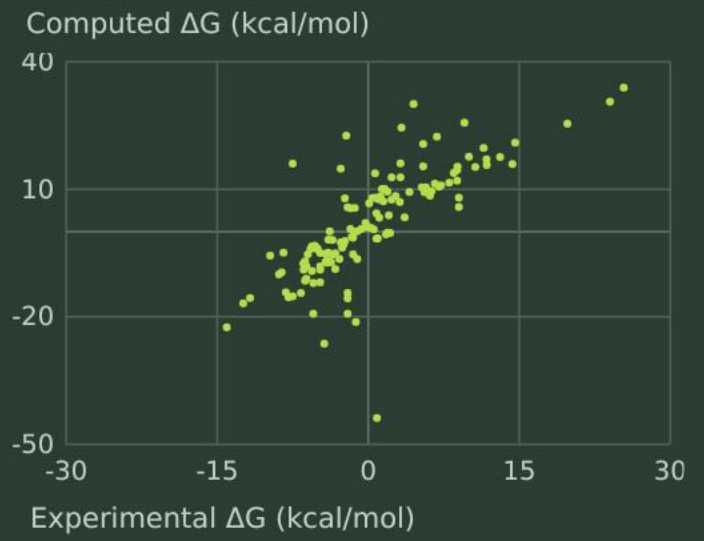
Enumeration

Quantum refinement and weighting

Qtautomers validation

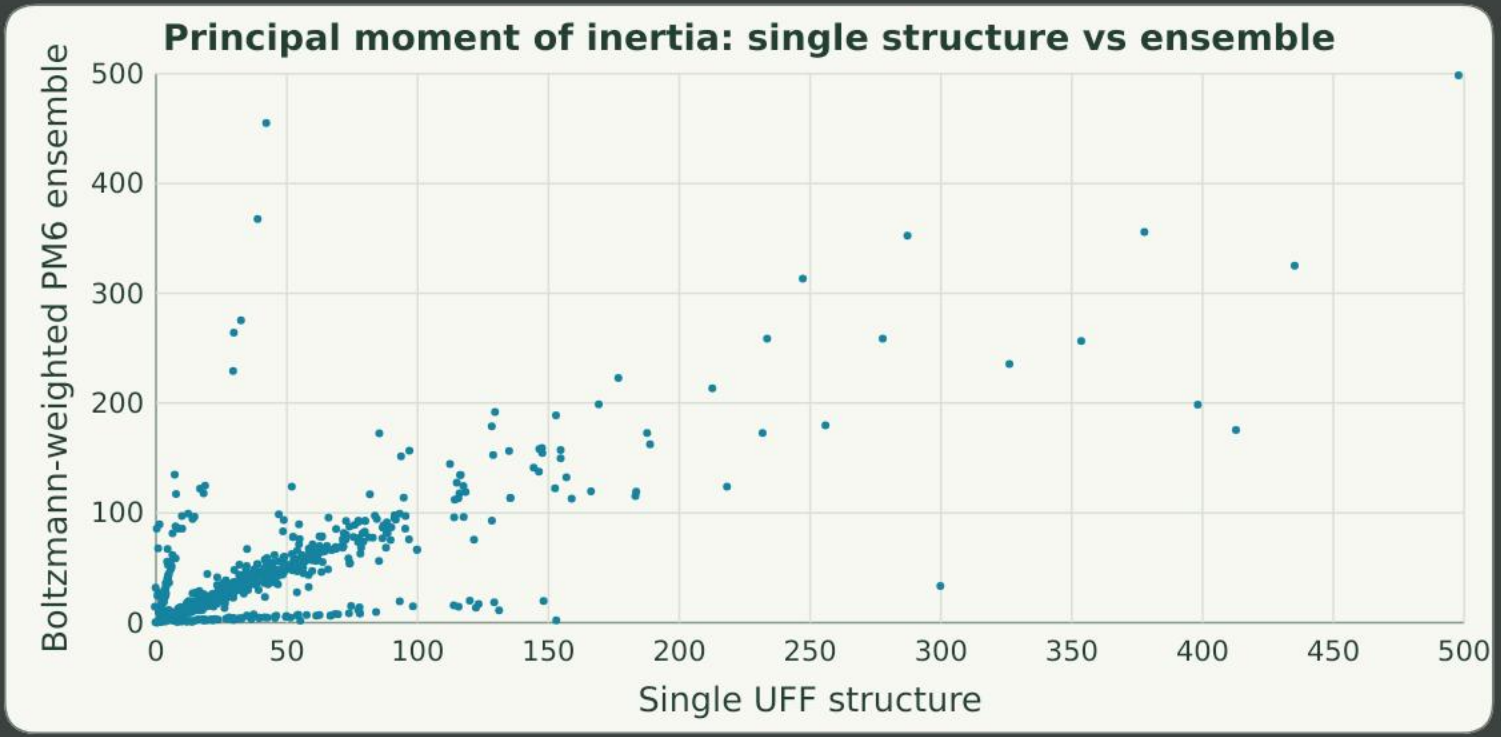
136 tautomer pairs in water

B3LYP/6-31+G(d)



State-specific structures and energetics provide the basis for ensemble descriptors

Boltzmann weighting can change molecular shape descriptors



12,390 substances

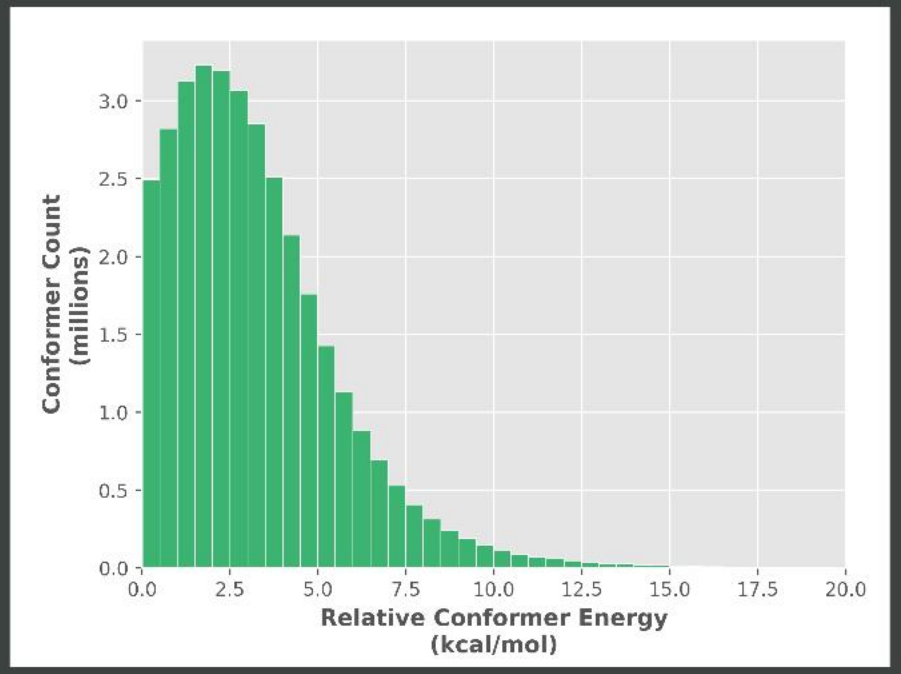
157,430 PM6 conformers across two property datasets

Class-dependent departures

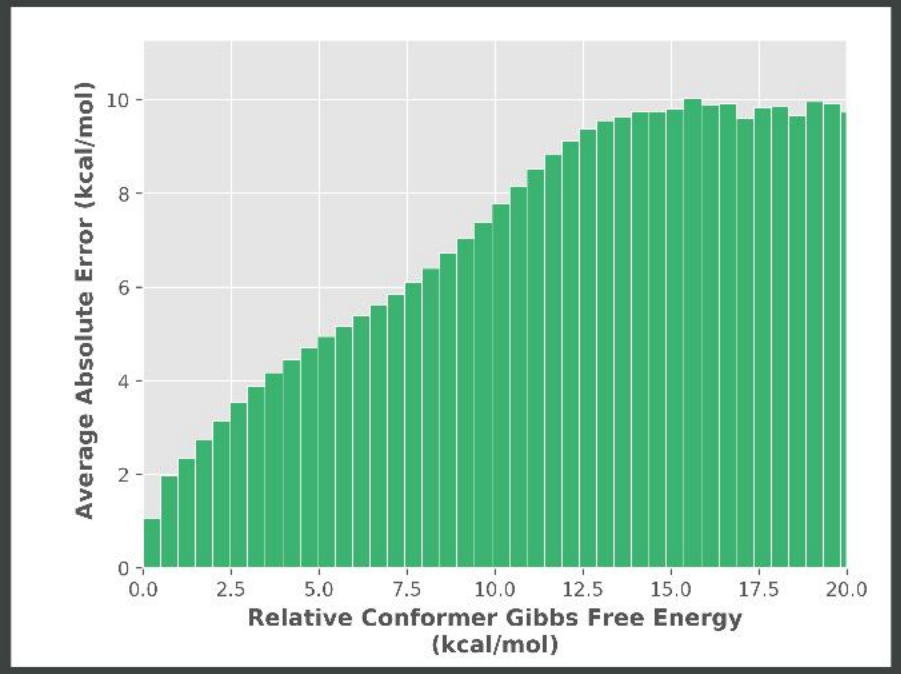
Certain molecular classes depart from the overall correlation when descriptors reflect a Boltzmann-weighted ensemble.

Some molecular classes lose shape-descriptor correlation after Boltzmann weighting

Force-field errors grow away from equilibrium



Conformer population across relative energy

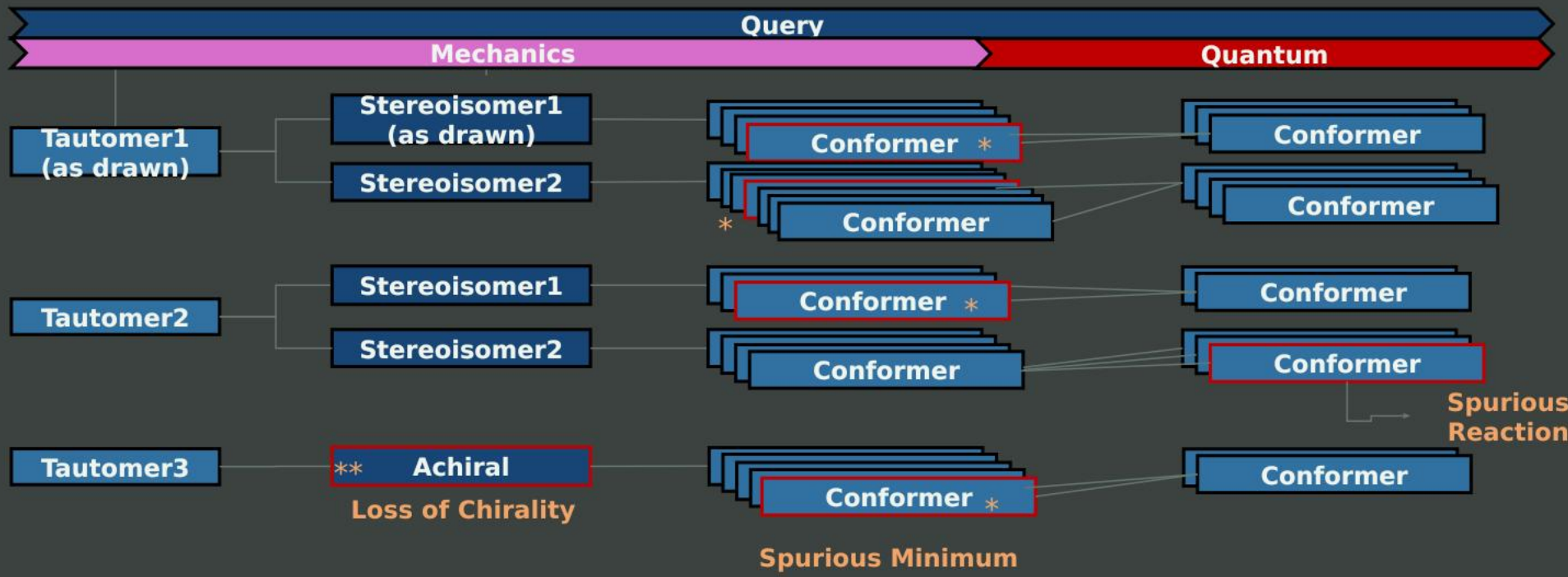


UFF–PM6 error across the energy landscape

Historical Qontext analysis: over 1.5 million molecules and 35 million conformations.

Energy models affect which structures receive weight in molecular datasets

Qontext retains the relationships behind computed data

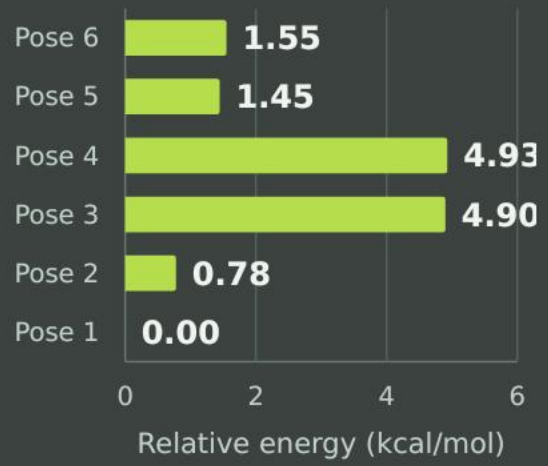


Linked molecular states, conformers and quality-control outcomes

Metal-complex ensembles resolve distinct energy ranges

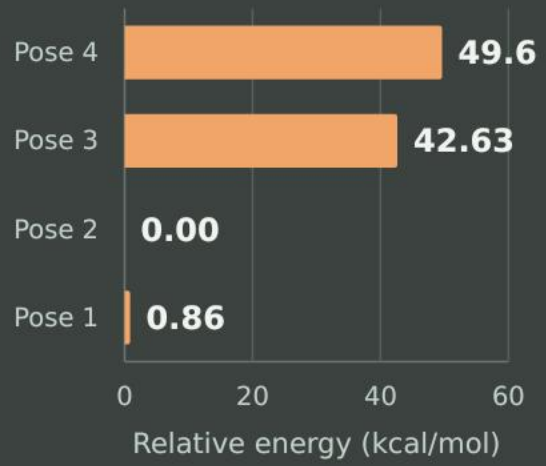
Cu complex

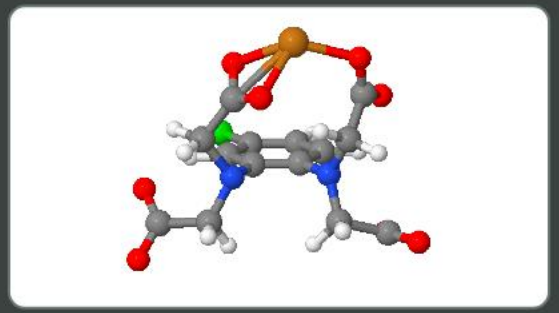
6 poses 0–6 kcal/mol scale

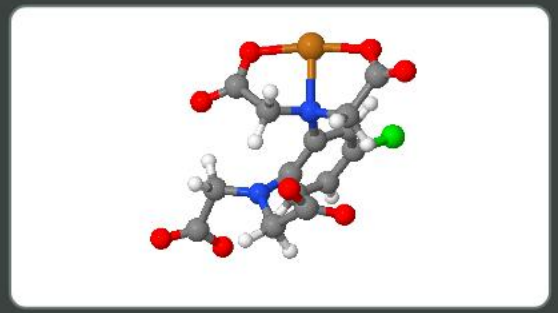


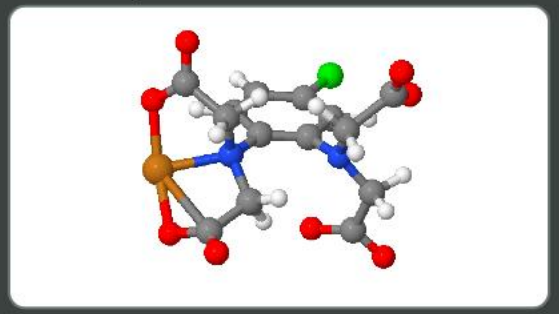
Ca complex

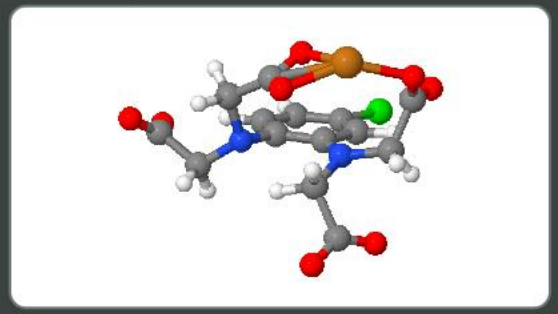
4 poses 0–60 kcal/mol scale




Cu-complex view 1


Cu-complex view 2

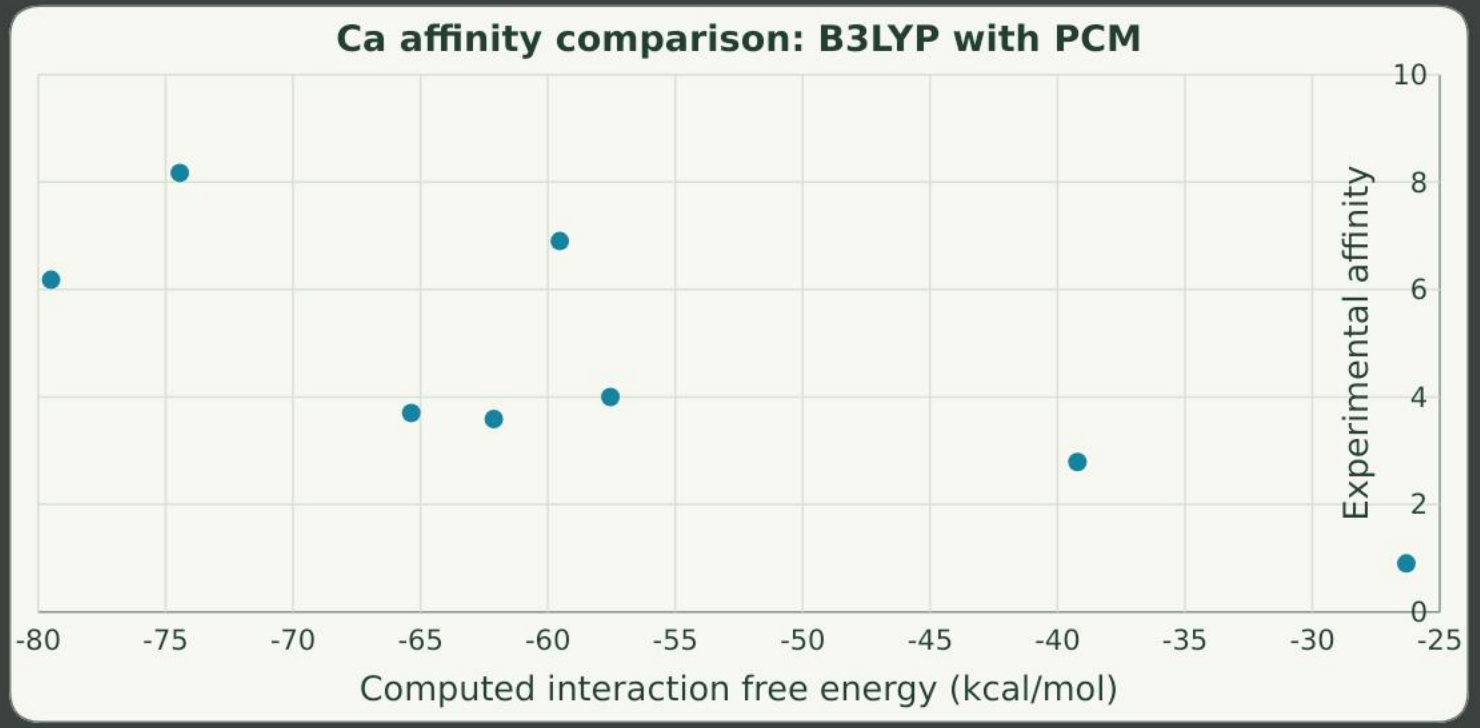

Cu-complex view 3

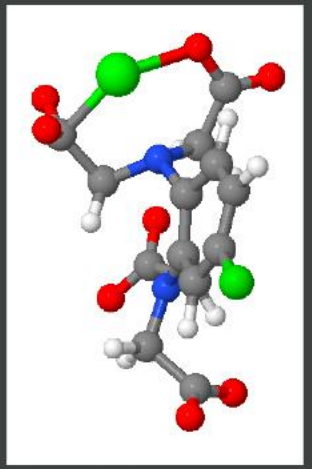

Cu-complex view 4

Relative energies are zeroed within each metal ensemble. Images illustrate each complex.

Cu poses span 4.9 kcal/mol, while the Ca ensemble spans 49.6 kcal/mol

Computed affinity labels can be checked against experiment





$R^2 = 0.655$

Eight compounds
Source experimental-affinity comparison

The calculation protocol must represent the physical quantity needed for learning

Qontext connects molecular data to property prediction

ConstruQt InteraQt Qontext

Train and validate



Physics-based data
Structures and interaction labels

Connected records
States, methods and quality status

Property predictions
New candidates in the target domain

A defined learning task connects reusable molecular records to validated predictions

Paid pilots build reusable molecular-data workflows



The client owns the application; ChemAlive develops the molecular-data workflow

Bring physics-based molecular data to your materials AI program

We define the molecular dataset and validation strategy needed to connect your target property to a scalable AI workflow.

Peter D. Jarowski, Ph.D.
p.jarowski@chemalive.com
+41 78 817 27 36
chemalive.com