Physical Model Induction with QuanSA™

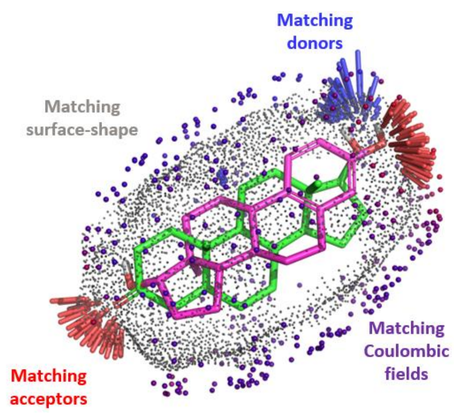
Affinity Prediction that is Synergistic with Simulation-Based Methods

Ajay N. Jain, PhD, VP of Research, Ann E. Cleves, PhD, VP of Application Science
BioPharmics Division, Optibrium Limited, Cambridge, UK info@optibrium.com

Introduction to QuanSA: *Quan*titative *S*urface-field *A*nalysis

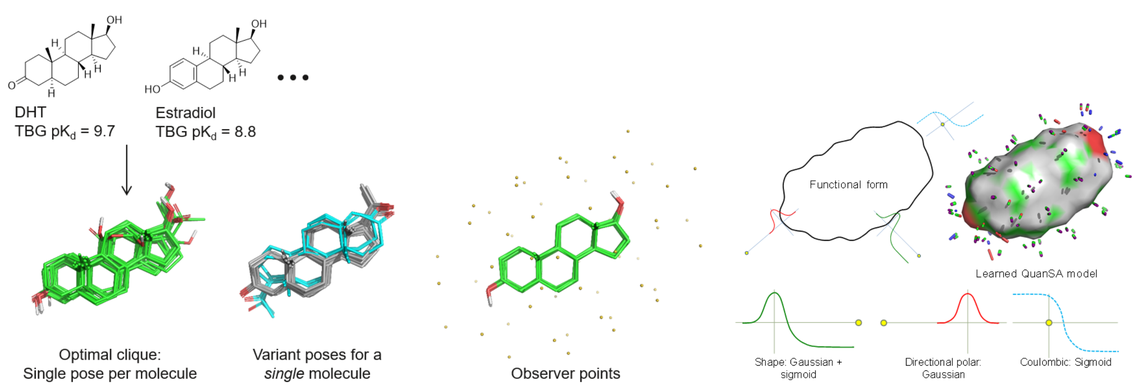
Affinity prediction challenges:
The things we want to predict are in the future. They do not come from the same statistical population as the molecules/activity-data from which we can induce models. This violates the central assumption of machine-learning: predict on things that come from the same population as things used to train.

QuanSA uses a surface representation:
To address these challenges, it is necessary to use a physics-driven domain knowledge in the model induction process. The actual molecular surfaces and their properties are not well represented by the atom/bond depictions used to symbolize molecules. Surfaces are congruent even when they don't look like they should be.



The QuanSA method

To define a 'pocket field', an initial alignment of all training molecules is constructed and function parameters at the observer points are learned based on activity data [1].



The QuanSA pocket field is iteratively refined using multiple instance machine learning; considering multiple poses for each compound means that no assumptions are made about the 'right' pose.

Building/applying a model is tractable, **taking just hours to build or refine**.

QuanSA models require no known target; models can be informed by protein structure or applied on purely phenotypic data.

A new molecule can typically be run in seconds; thus, **very large-scale applications are possible.**

Predictions are supported with a **score, a pose and quality metrics.**

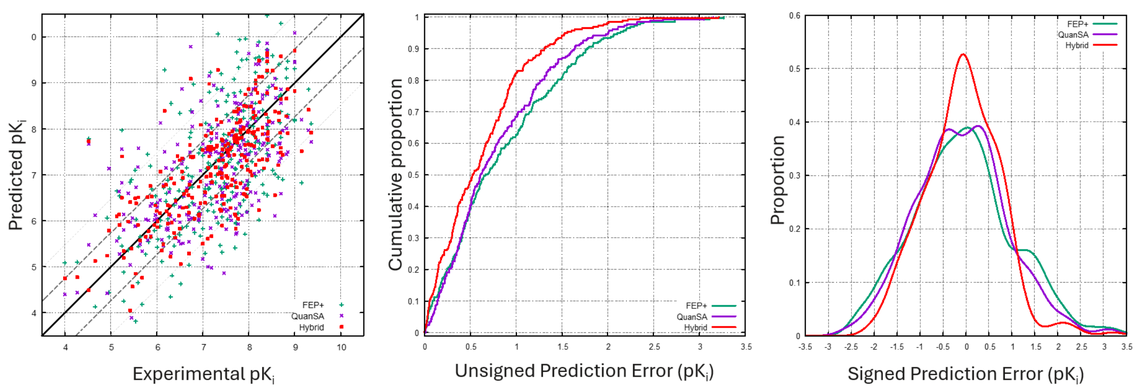
Structurally novel molecules are often well within the domain of applicability, **accurately supporting scaffold-hopping**.

QuanSA benchmarking vs FEP+

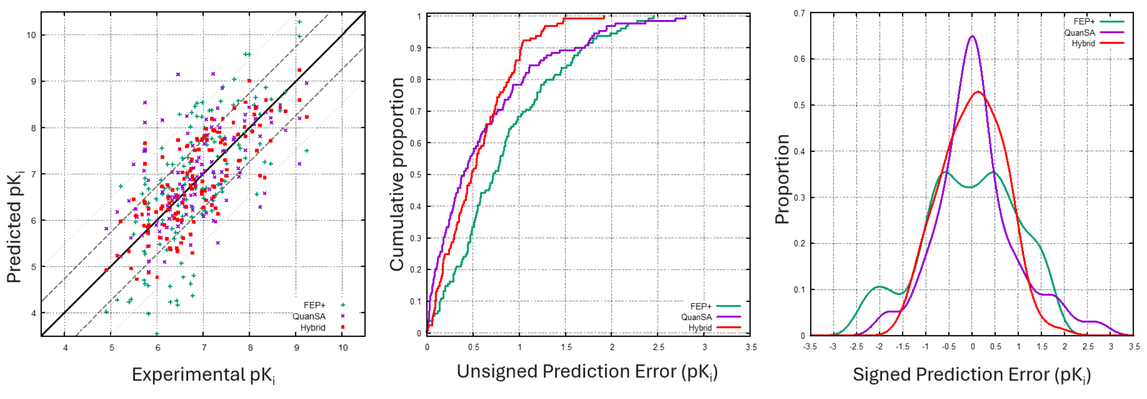
Schindler 2020 and Abel 2015 FEP+ comparison

A critical application is to accurately predict affinities for future molecules. QuanSA and FEP+ models were built and evaluated [2] for sixteen targets from two published datasets using temporal segregation. Training set compounds were selected based upon similarity to the FEP+ reference ligand, forcing the QuanSA models to extrapolate. The study compared the accuracy across the targets, as summarized in the plots below.

Schindler 2020 results (8 targets)



Abel 2015 results (8 targets)



QuanSA and FEP+ have similar accuracy.

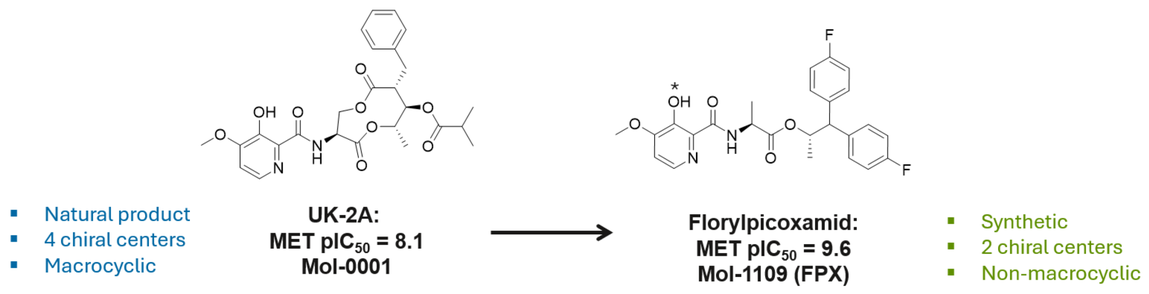
Both methods are highly synergistic; a hybrid (mean) score increases accuracy compared to either method.

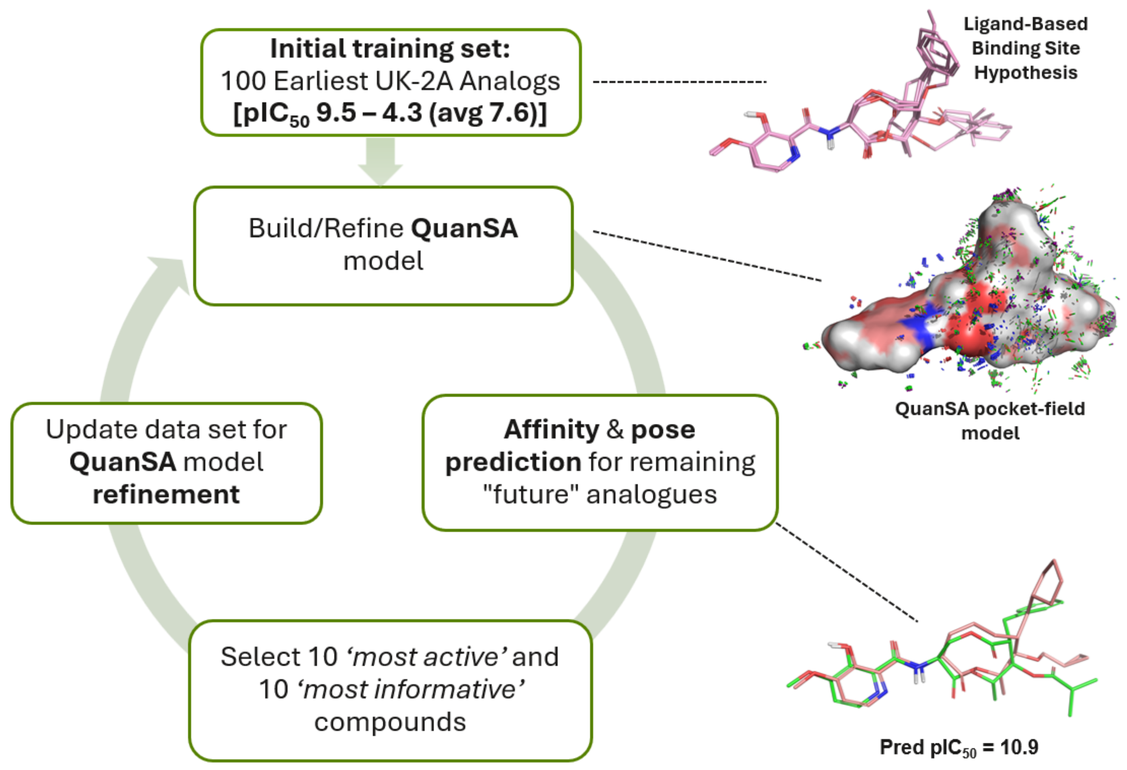
QuanSA is ~1000x faster than FEP+, alleviating screening bottlenecks.

QuanSA project application

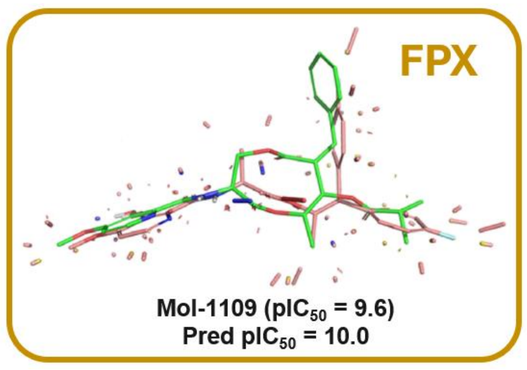
Active learning to identify a mimic of a macrocyclic natural product

Scaffold replacement as part of an optimization process is a complex challenge. Using a data set of ~1,100 time-stamped compounds, we applied an iterative procedure to refine a QuanSA model, starting with a macrocyclic natural product lead (UK-2A), and rapidly identify a non-macrocyclic fully synthetic broad-spectrum crop anti-fungal (FPX) [3].





Iterative model refinement efficiently guided candidate selection to the desired product.
FPX was **identified in round 5** as one of the most active predicted molecules
The model effectively **learned the non-macrocyclic scaffold**.
Only 100 molecules were selected vs over 1,000 in the project, representing a **10x improvement in efficiency**.



Conclusions

- **QuanSA builds physically realistic causal models based on ligand structures alone.**
- **QuanSA and FEP+ are equivalent in accuracy and synergistic, but QuanSA is ~1000x faster and has a broader domain of applicability.**
- **Active learning with QuanSA enables more efficient lead-to-candidate design - 10x in this case study.**

References

[1] Cleves, A.E. and Jain, A.N. (2018). JCAMD, 32, 731-757
[2] Cleves, A.E., Johnson, S.R., and Jain, A.N. (2021). JCIM, 61, 5948-5966
[3] Cleves, A.E., Jain, A.N., Demeter, D.A., et al. (2024). JCAMD 38, 19

Acknowledgements

BMS Pharmaceuticals: Stephen Johnson;
Corteva Agrisciences: Erin Hancock, Zachary Buchan, Jeremy Wilmot, Dave Demeter;
Optibrium: Himani Tandon

Learn more on our webpage