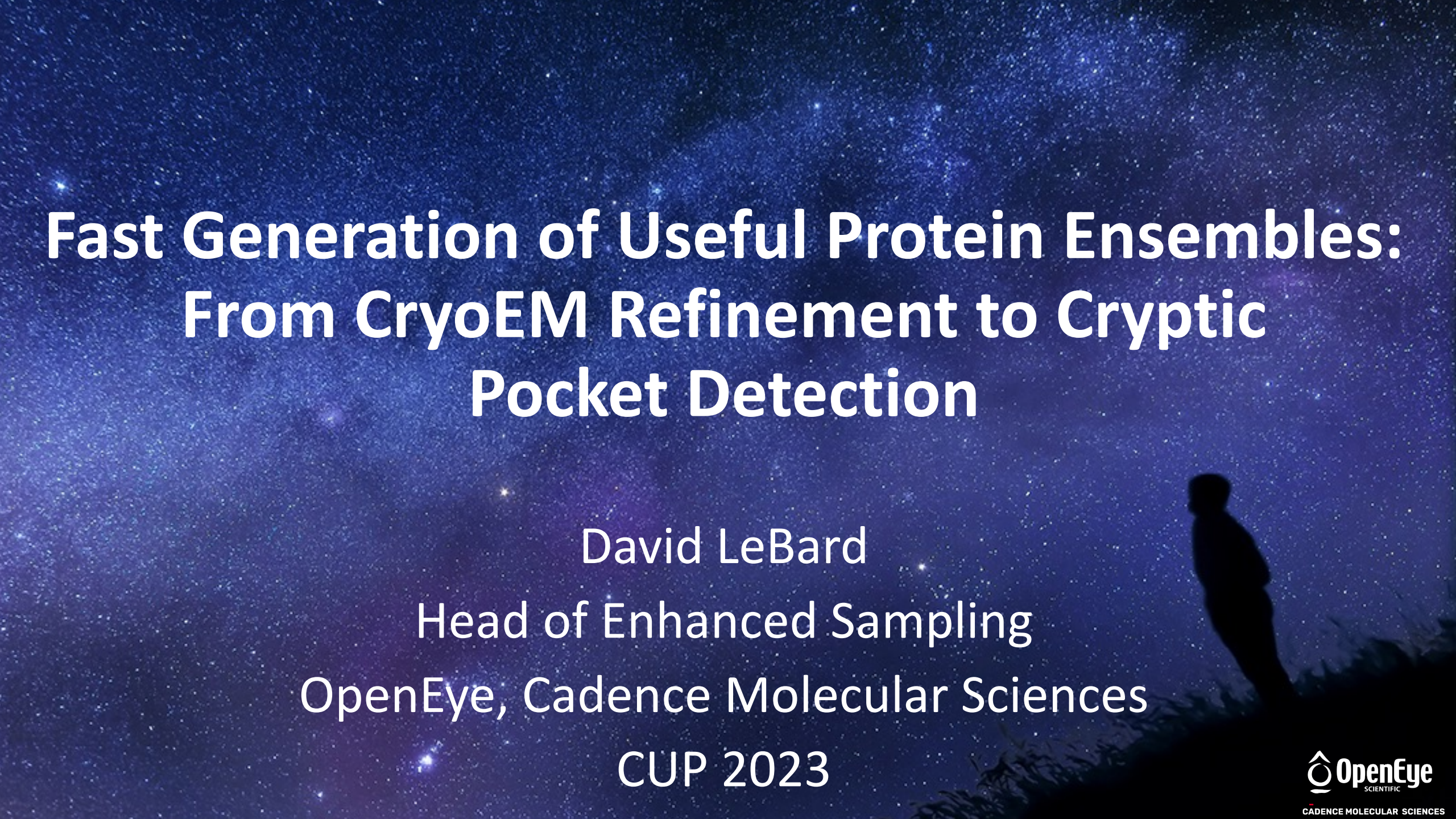
Fast Generation of Useful Protein Ensembles: From CryoEM Refinement to Cryptic Pocket Detection

David LeBard
Head of Enhanced Sampling
OpenEye, Cadence Molecular Sciences

CUP 2023



OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Outline

Dead person quote

Motivation for generating protein ensembles

Our Normal Mode Analysis (NMA) based ensemble generation method

Generating a simple ensemble: alanine dipeptide

Conclusions

Dead person quote

Motivation for generating protein ensembles

Our Normal Mode Analysis (NMA) based ensemble generation method

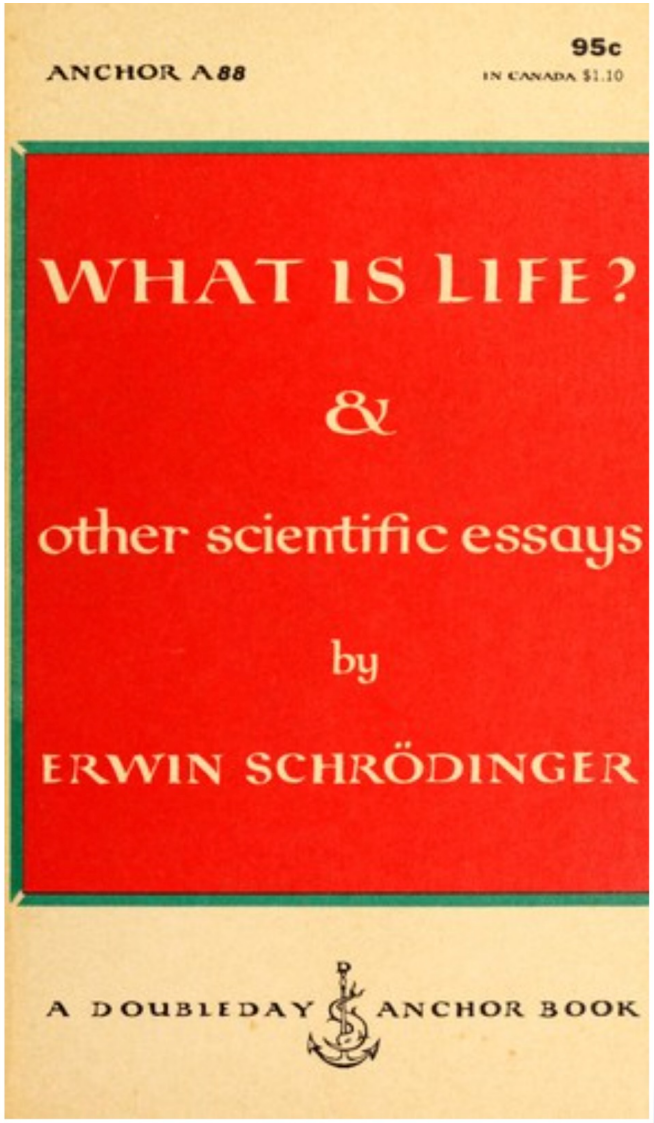
Generating a simple ensemble: alanine dipeptide

Conclusions

E. Schrödinger thought about (protein) ensembles

***What is life?*, 1944**

"I propose to develop first what you might call *'a naive physicist's ideas about organisms*', that is, the ideas which might arise in the mind of a physicist who, after having learnt his physics and, more especially, the statistical foundation of his science, begins to think about organisms and about the way they behave and function…"



Credit: The Open Library

Dead person quote

Motivation for generating protein ensembles

Our Normal Mode Analysis (NMA) based ensemble generation method

Generating a simple ensemble: alanine dipeptide

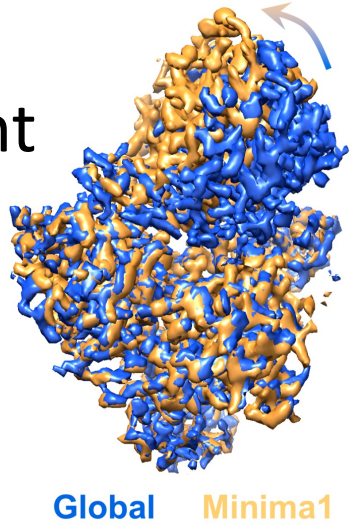
Conclusions

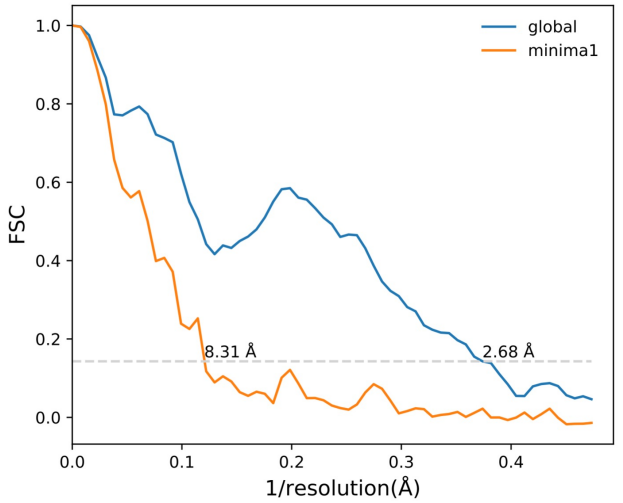
Drug Discovery Realms

Protein Structure, Modeling & Simulation

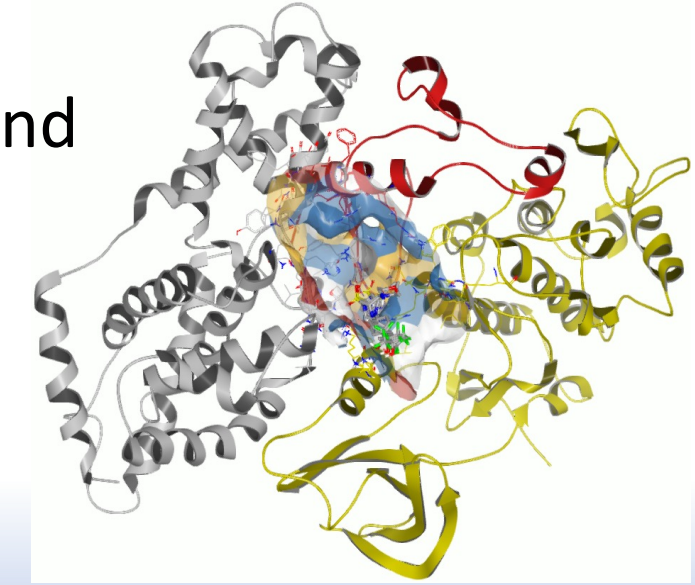
Our uses for protein conformational ensembles

1. Cryo-EM refinement





2. Cryptic pocket identification and classification



OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Dead person quote

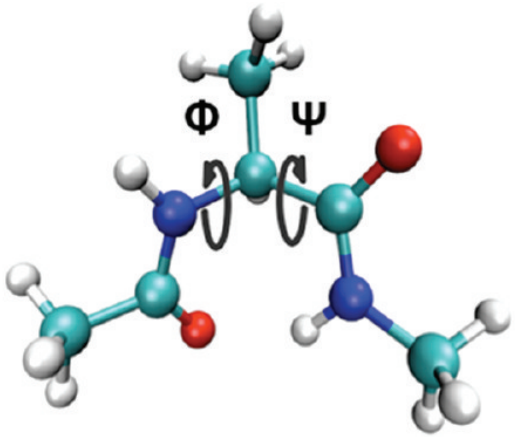
Motivation for generating protein ensembles

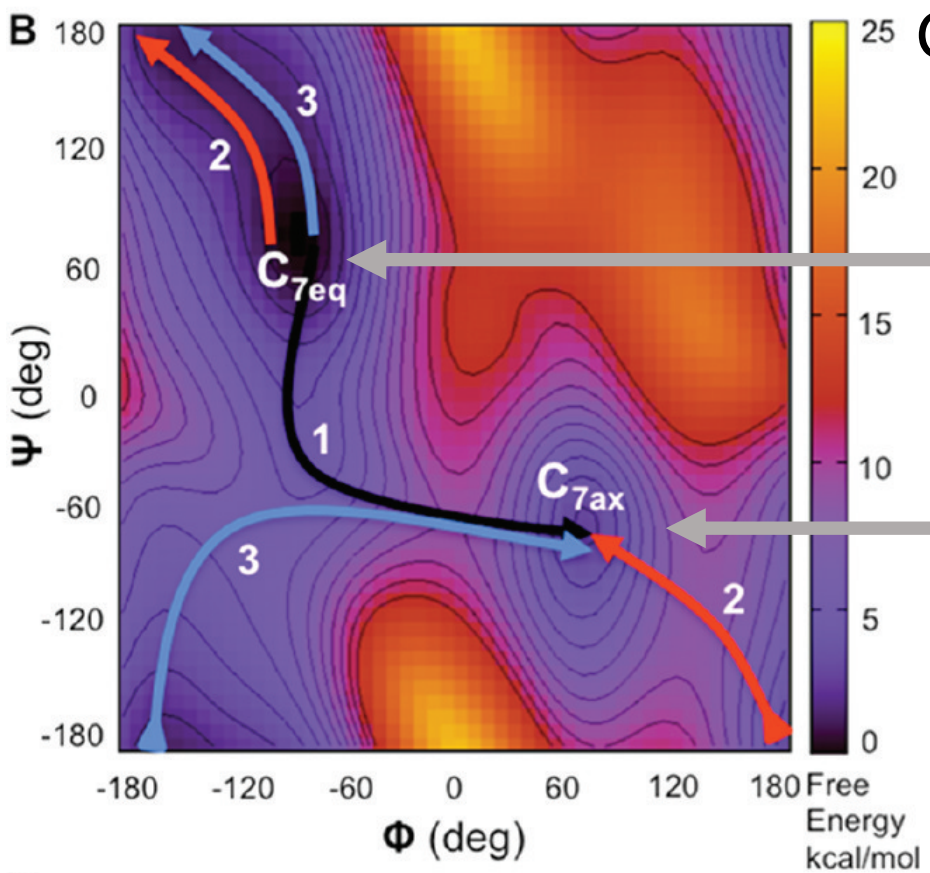
Our Normal Mode Analysis (NMA) based ensemble generation method

Generating a simple ensemble: alanine dipeptide

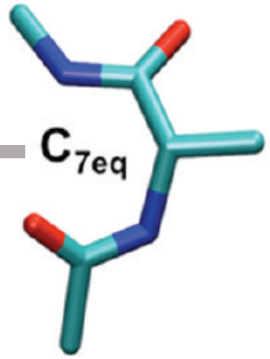
Conclusions

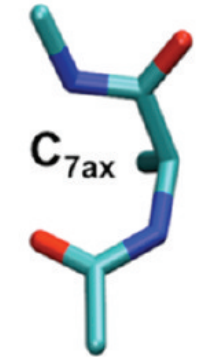
Illustration of protein sampling: alanine dipeptide





Global minimum





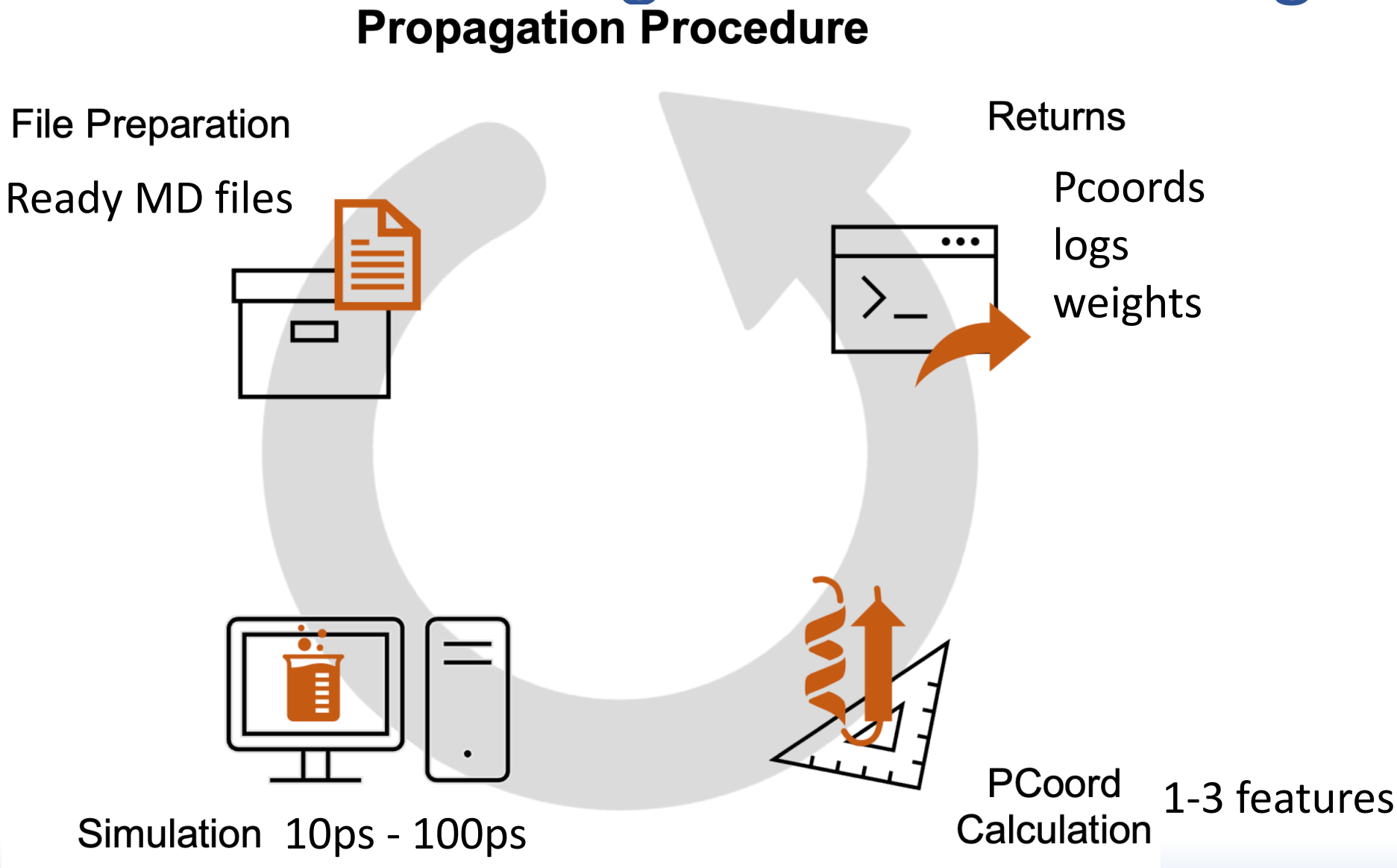
Metastable state

Free energy surface from multiple replica, many-μs of biased MD sampling

Adapted from Patel et al, JCTC, 2011

OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Brief outline of the Weighted Ensemble algorithm

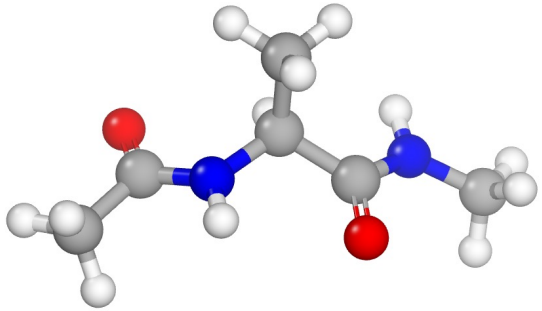


Adapted from: Fig. 5 of Bogetti, et al, A Suite of Advanced Tutorials for WESTPA 2.0, accepted in LiveCOMs 2023

OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Sampling alanine dipeptide using WE simulations

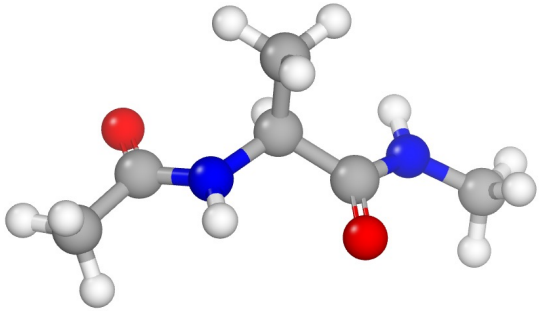
- Molecular system:
  `openmmtools.testsystems.AlanineDipeptideVacuum`
  - Force field: Amber ff96
  - Initial conformation obtained by local energy minimization
- MD integrator: Langevin dynamics at 298 K with a timestep of 1 fs and collision frequency of 1 ps^{-1}.
- WE resampling time: 10 ps
- WE allocation: 4 walkers per bin



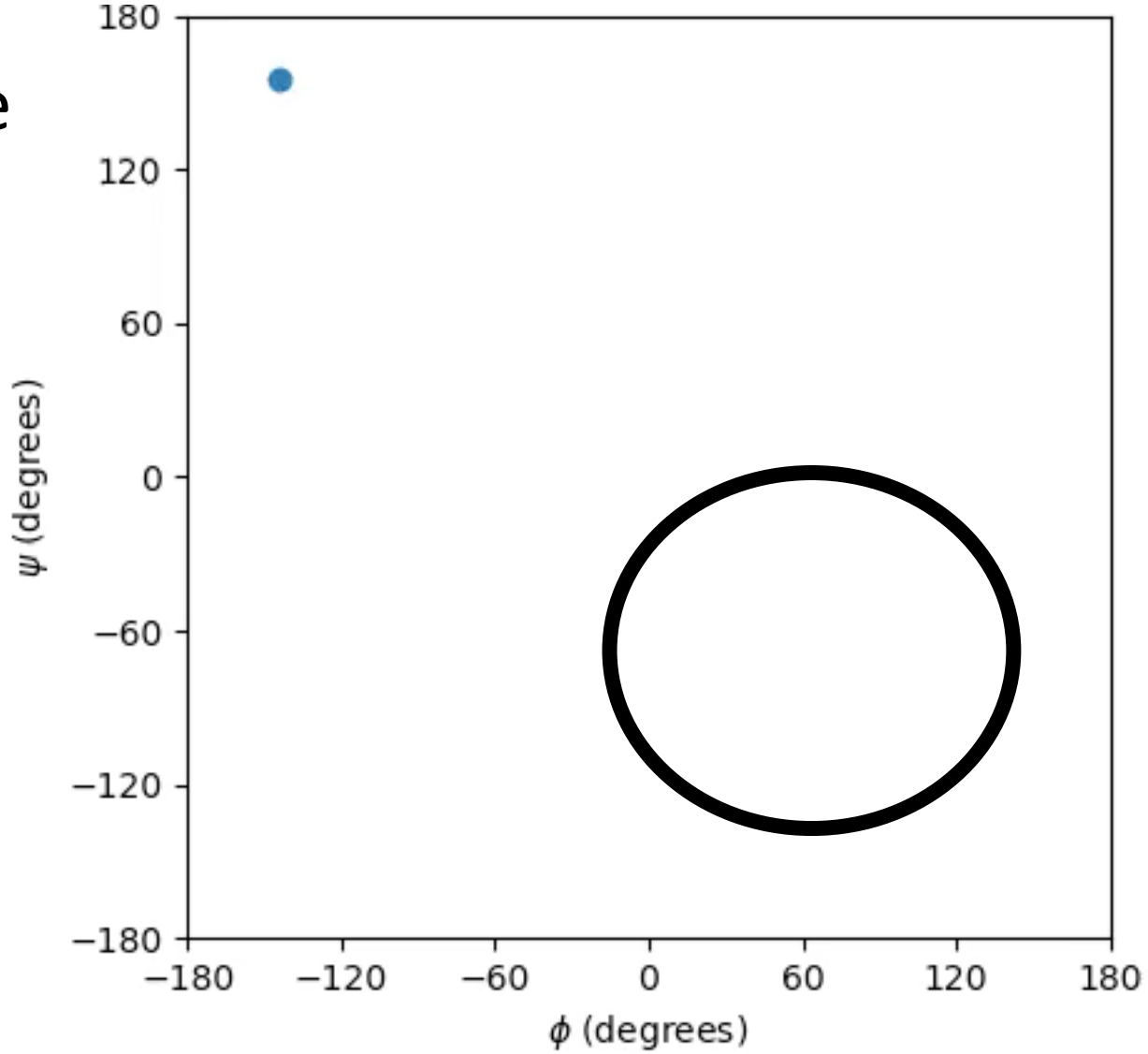
Initial conformation (C5)

WE sampling when the best coordinate is known

- 2ns of total molecular time
- WESTPA in Orion using the Garden floe package
- Sampling φ/ϕ backbone angles
- Minimal Adaptive Binning
- Starting in C5/$C7_{eq}$ basin

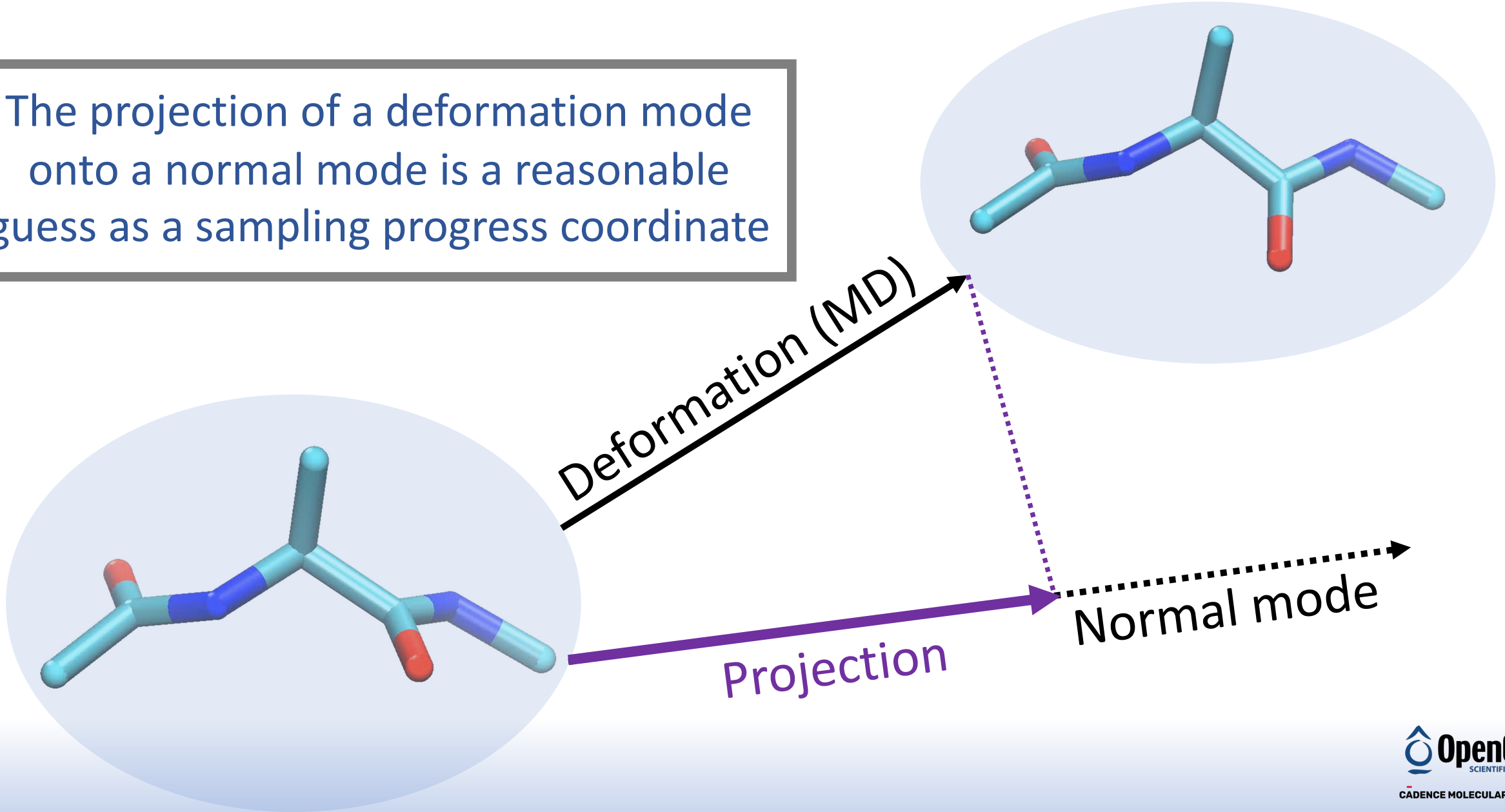


Initial conformation (C5)



What if the best progress coordinate is unknown?

The projection of a deformation mode onto a normal mode is a reasonable guess as a sampling progress coordinate



Dead person quote

Motivation for generating protein ensembles

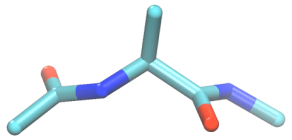
Our Normal Mode Analysis (NMA) based ensemble generation method

Generating a simple ensemble: alanine dipeptide

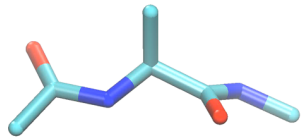
Conclusions

Automated rare event sampling of alanine dipeptide

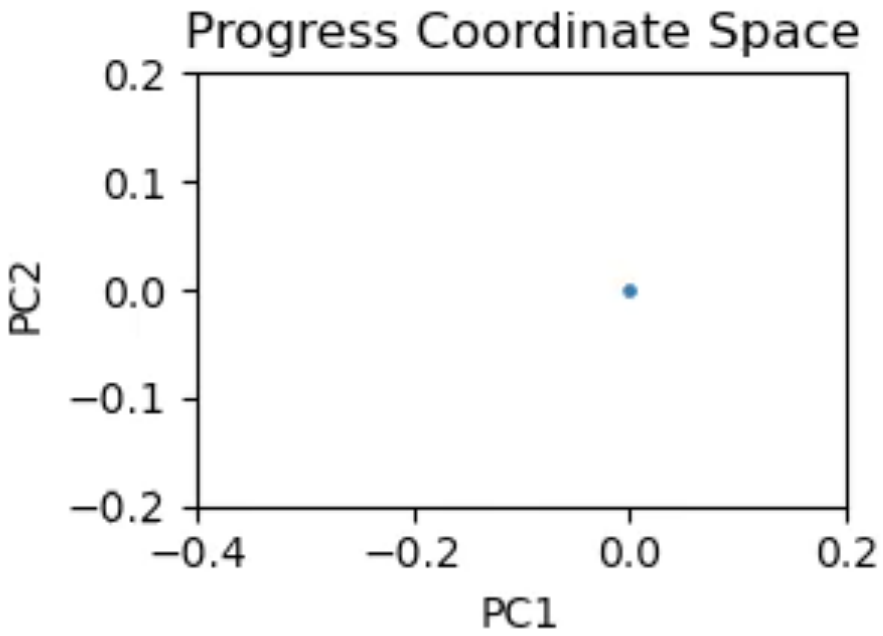
Automated progress coordinates: Projection onto the first two **principal components** (**PC**) of heavy-atom displacements from the initial conformation, fit to a 100 ns MD trajectory that remained trapped in the $C5/C7_{eq}$ basin.

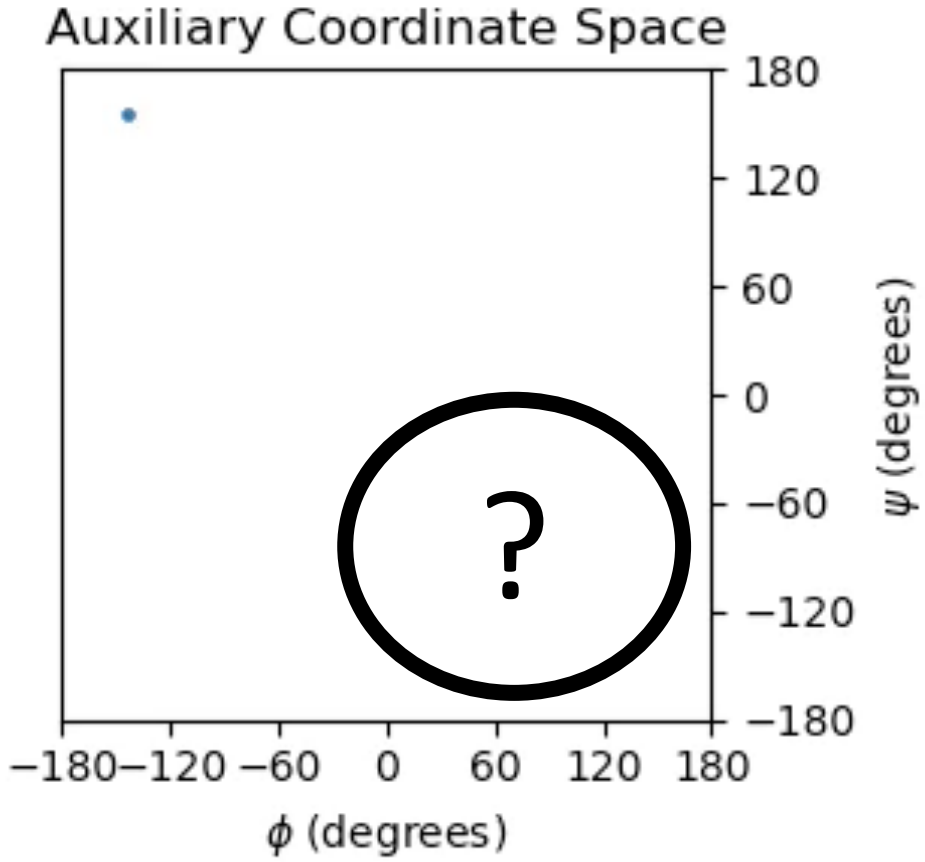


PC1



PC2





$$0 \leq t \leq 2 \text{ ns}$$

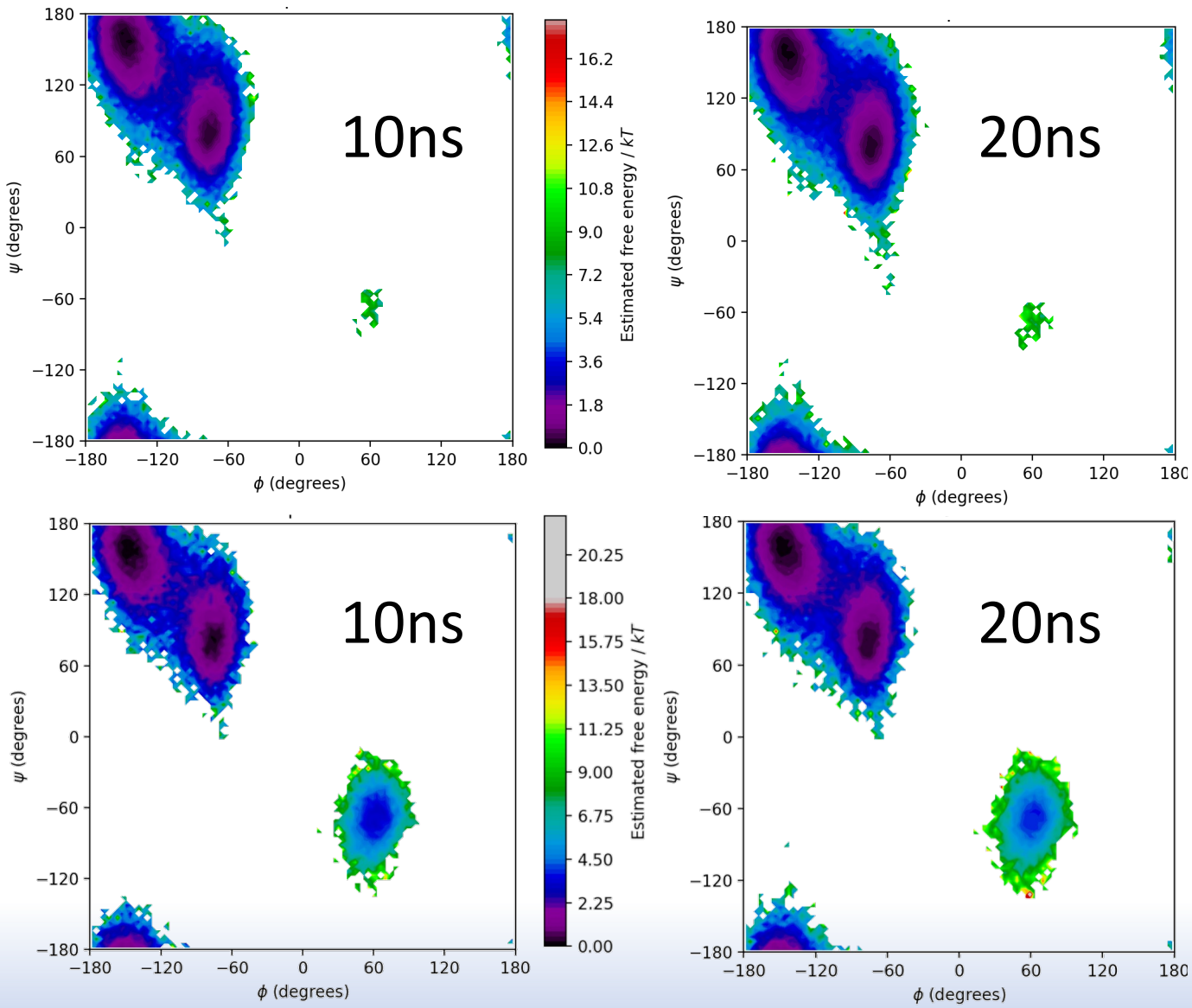
OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Normal mode sampling is good, but not perfect

Progress coordinate

PC1, PC2

ϕ, ψ



OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Dead person quote

Motivation for generating protein ensembles

Our Normal Mode Analysis (NMA) based ensemble generation method

Generating a simple ensemble: alanine dipeptide

Conclusions

Conclusions

Protein ensemble generation is important for many applications in the drug discovery pipeline.

Here, we will show two applications of protein ensemble generation:

1) Cryo-EM structure refinement
2) Cryptic pocket identification and classification

Finally, the protein sampling methods (including cryptic pocket analysis) will be available in our OE Floes release this summer!


Thank You
The End
OpenEye
SCIENTIFIC
CADENCE MOLECULAR SCIENCES