Scaling Cheminformatics Computational Simulations on HPC Clusters

Mike Mikailov[1], Nadya Tarasova[2], Yulia Borodina[1],
Marc C. Nicklaus[2], Fu-Jyh Luo[1], Stuart Barkley[1], Kenny Cha[1]

[1]U.S. Food and Drug Administration, Silver Spring, 20993, MD, USA
[2]National Cancer Institute, Bethesda, MD, 20814, USA

Disclaimer

- The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services. This is a contribution of the U.S. Food and Drug Administration and is not subject to copyright.

Introduction/Hypothesis

- The recent explosion of chemical libraries of National Cancer Institute (NCI) beyond a **billion molecules** led to large scale simulations for **Virtual Screening (VS)**.
- **VS** is a simulation technique used in drug discovery to search libraries of molecules to identify structures, likely to bind to a drug target.
- Over **950 years** is needed for processing the billion-sized libraries.
- The FDA CDRH High-Performance Computing team is working with NCI to apply scaling techniques and use more powerful hardware resources (i. e., GPUs) to make this mission critical task feasible and accomplishable in timely manner.
- Theoretically, using the scaling techniques would reduce the **950 years of VS time on one CPU core to 70 days on 5,000 CPU cores** on the FDA CDRH HPC clusters.

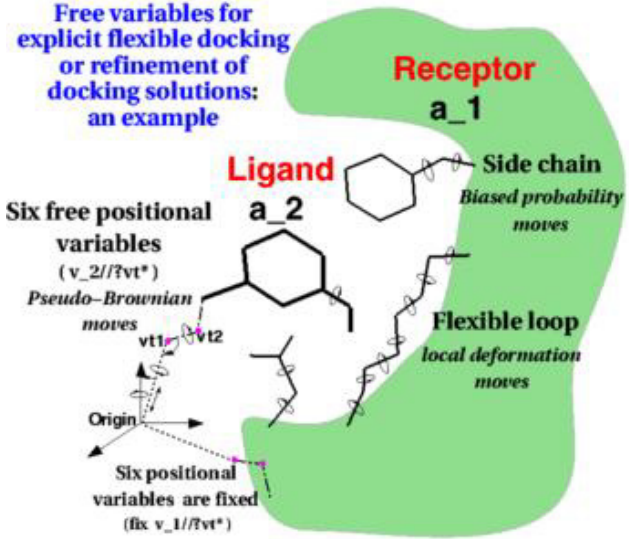
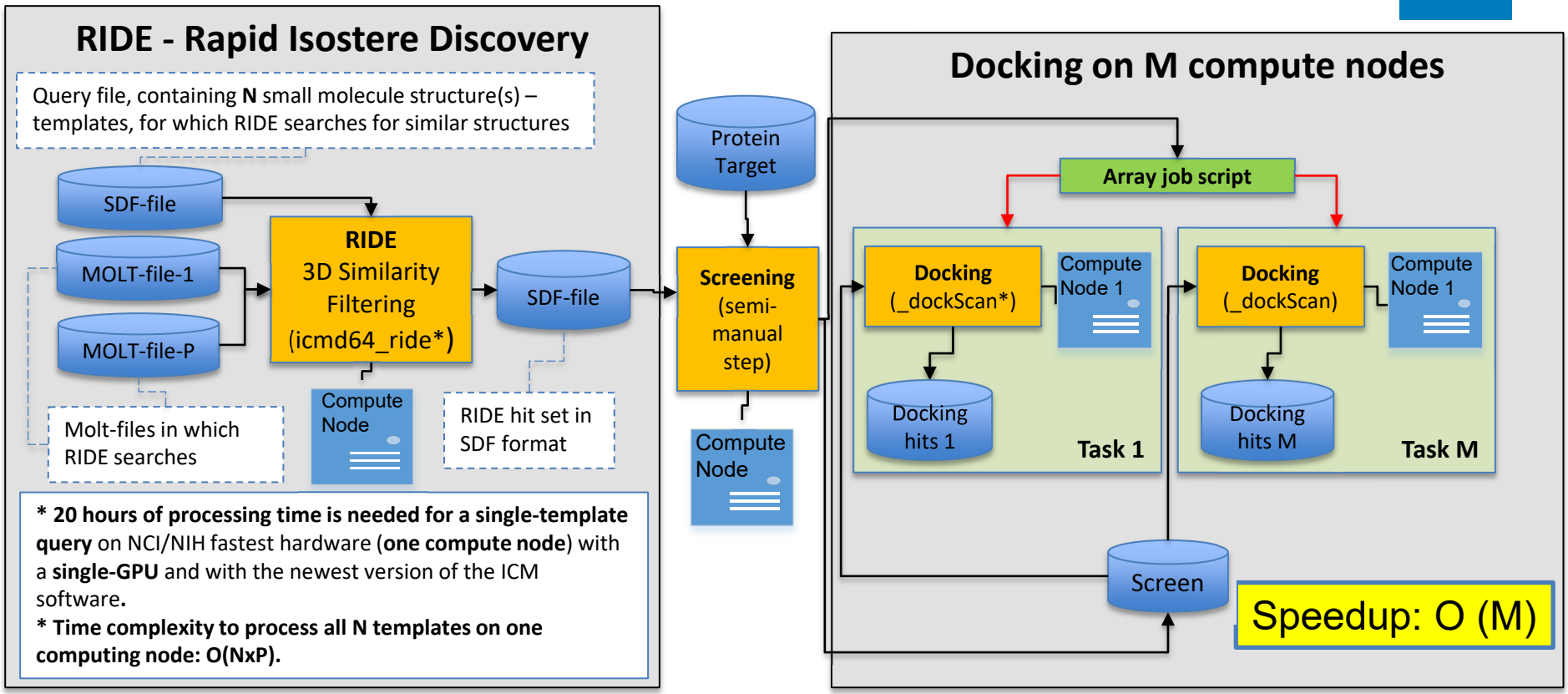


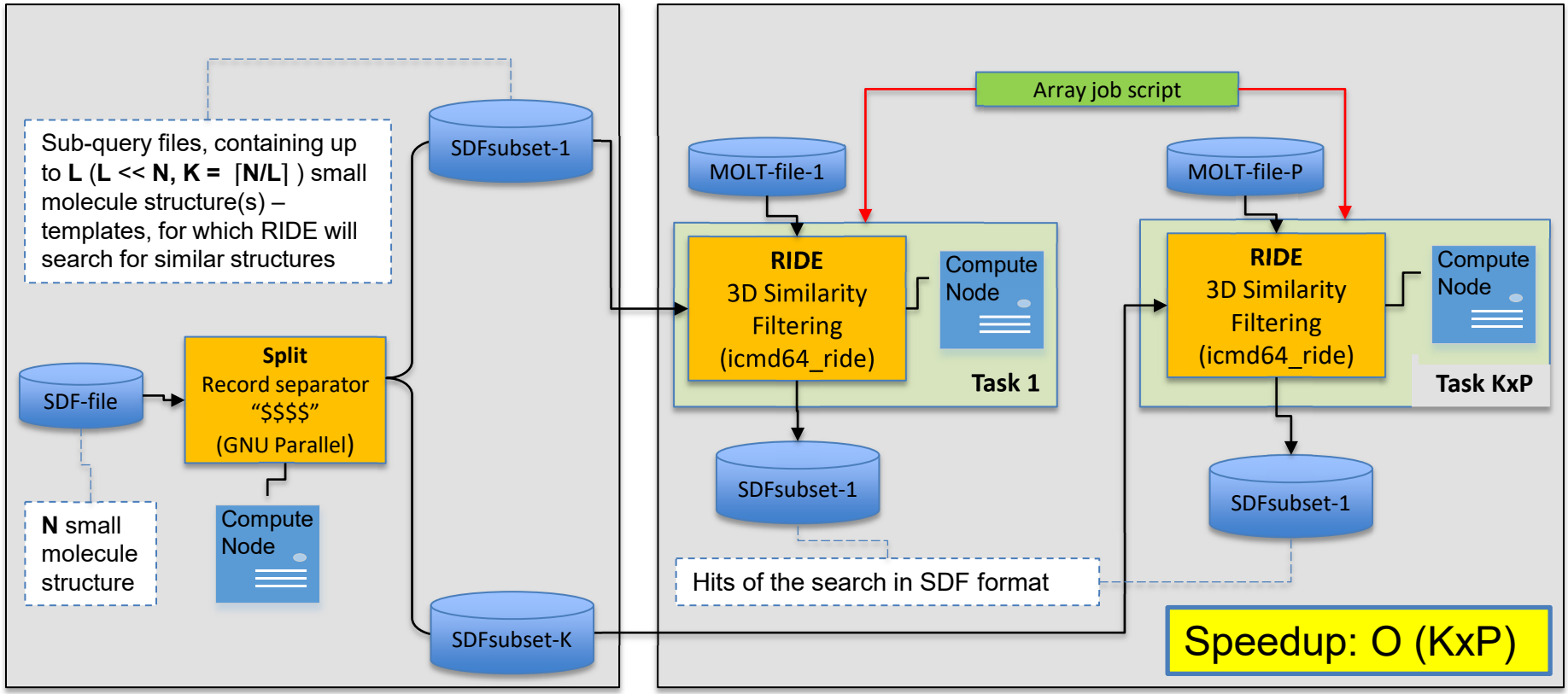
Image source: [Molsoft L.L.C.: Molsoft's Technology]

Current virtual screening workflow



*MolSoft ICM-Pro software

Scaling rapid isostere discovery



Conclusion

- Divide and Conquer
  - Data partitioning.
  - Array job-based technique.
  - Distributed computing across the HPC clusters.

U.S. FOOD & DRUG ADMINISTRATION

Thank you!

Disclaimer: *The information in this presentation represents the opinions of the speaker and does not necessarily represent FDA's position or policy.*


FDA
U.S. FOOD & DRUG
ADMINISTRATION