BeoFold: Simulating How Proteins Fold Using Distributed Computing and State Networks

Author: Bryan P. White (ORCID: 0000-0003-4191-3511) • **Affiliation:** CRV Science Labs • **Preprint ID:** APP-PREPRINT-02 (v1.4.0 Empirical Preprint) • **Categories:** q-bio.BM • physics.chem-ph • cs.DC
Canonical DOI: 10.5281/zenodo.24222370 • **Apps Hub:** https://crvscience.com/apps • **Repository:** https://github.com/bpwhite/crvscience-beofold

Abstract — BeoFold (v1.4.0) is an open-source molecular simulation and cluster-networking program built to investigate how short amino-acid chains fold into stable, functional three-dimensional protein structures. Simulating protein folding inside a single conventional computer at normal biological temperatures is remarkably slow because flexible molecules spend almost all of their time bouncing inside shallow, partially folded pockets separated by tall energy walls. BeoFold solves that sampling trap by orchestrating parallel replicas across multiple temperatures—allowing warm copies to cross energy barriers and trade shapes cleanly with cooler copies—while optionally linking ordinary lab desktops over encrypted QUIC and Raft consensus networking. Once snapshots accumulate, BeoFold converts thousands of noisy atomic coordinates into a clear network of stable states and folding speeds. On the classic 20-amino-acid Trp-cage model protein, BAOAB thermal integration held kinetic motion steady near 282.3828 K, four temperature tiers from 285 K to 405 K achieved a 47.89% exchange rate down to 156.6821 kcal/mol, and spectral state clustering mapped three distinct structural basins with near-exact time-reversibility.

1. Real-World Context & Why This Problem Matters

The Sampling Dilemma at the Heart of Protein Folding

Every living cell depends on proteins to execute mechanical work, transmit signals across cell membranes, ferry oxygen and nutrients, neutralize invading microbes, and speed up essential chemical reactions thousands of times beyond uncatalyzed chemistry. A protein emerges inside the cell as a flexible linear ribbon of amino acids rolled out of a ribosome, yet it achieves biological potency only after folding and packing into a precise three-dimensional native shape. Tracing exactly how floppy chains locate that target structure—and explaining why single inherited mutations or heat stress sometimes push molecules into misfolded tangles—sits right at the core of molecular biophysics, structural medicine, and enzyme engineering.

Watching a protein fold step by step inside a computer confronts researchers with a huge timescale mismatch, much like filming continental drift with a high-speed camera built for hummingbird wings. Individual covalent bonds vibrate trillions of times each second, forcing any physically realistic simulator to advance atomic positions in tiny femtosecond increments (quadrillionths of a second). By contrast, whole loops twisting around a protein core and hydrophobic pockets zipping shut happen thousands or millions of times more slowly. Launching one straightforward trajectory on a desktop computer at physiological temperature near 300 K almost guarantees the chain falls straight into the nearest cozy half-folded valley, bouncing against energy walls for days without ever crossing into the true native fold.

How Temperature Exchanges and State Networks Unlock Folding Pathways

Computational chemists resolved that stubborn trapping obstacle through two elegant statistical ideas that require no artificial shortcuts. First, modern symmetric Langevin thermostats—specifically the BAOAB operator splitting integrator—preserve subtle conformational distributions and steady average temperatures smoothly even when taking practical two-femtosecond steps without injecting runaway energy drift[1]. Second, Temperature Replica-Exchange Molecular Dynamics (T-REMD) launches several parallel copies of the identical protein across a clean staircase of temperatures spanning mild biological warmth up to hot, energetic rungs[2].

On the warm rungs, vigorous thermal jiggling melts stuck partial shapes and lets chains explore new folds effortlessly. Periodically, adjacent replicas compare potential energies via an exact statistical rule: whenever a warmer copy explores a promising compact structure, the ladder swaps temperatures smoothly so cold replicas inherit freshly discovered conformations without corrupting thermodynamic balance. Yet even after gathering thousands of atomic frames, staring at jittering 3D balls and sticks reveals little unless organized cleanly. Markov State Models simplify those raw trajectories by bundling similar molecular shapes into discrete basins and running PCCA+ fuzzy spectral clustering to reveal dominant structural states alongside measurable folding rates[3]. BeoFold brings atomic mechanics, peer-to-peer cluster sharing, and kinetic network mapping cleanly inside one standalone Rust application.

2. Conceptual Design & How BeoFold Works

From Bond Forces and BAOAB Integration to Replica Swapping

BeoFold (v`1.4.0`) divides molecular forces, multi-computer networking, and trajectory post-processing into clean, modular Rust crates so both single-laptop demonstrations and multi-workstation lab clusters run identical physics. Entering through `fold-`

`physics`, `forcefield.rs` computes how every atom feels stretched chemical bonds, bent bond angles, preferred backbone twists, and non-bonded spatial attractions or steric clashes. Advancing frame by frame, `integrator.rs` updates positions and velocities using the symmetric BAOAB splitting pattern—interleaving deterministic force kicks, inertial drifts, and gentle thermal friction—specifically prized for keeping conformational probabilities sharp across practical time steps[1].

To prevent simulations from getting stranded inside local valleys, `mremd.rs`—accelerated optionally by compute shaders inside `gpu.rs`—runs synchronized temperature tiers in parallel and orchestrates periodic coordinate exchanges between neighboring heat levels while logging every attempted and accepted swap[2].

Ad-Hoc Beowulf Clusters and PCCA+ Kinetic Landscapes

Traditionally, distributing molecular dynamics across multiple computers demanded specialized university supercomputer schedulers and cumbersome job scripts. Inside BeoFold, `fold-node` replaces complex cluster administration with lightweight ad-hoc networking across ordinary lab or classroom laptops. Peer discovery and node health run cleanly over `mesh.rs`, workunit dispatch guarantees consistency across crashes or dropped Wi-Fi connections via Raft distributed consensus (`raft.rs`), and compressed coordinate batches travel over fast encrypted QUIC streams (`quic_net.rs`) structured with versioned schemas in `fold-proto` and written safely to disk inside `fold-storage`. If one student's laptop closes mid-lecture, Raft log replication ensures no folding trajectories vanish mid-flight.

As workunits return to the coordinator, `fold-msm` converts massive frame piles into an intuitive map of how the protein moves. First, `clustering.rs` groups geometrically similar snapshots into fine-grained conformational microstates. Next, `transition.rs` counts lag-time jumps moving from one shape to another and enforces equilibrium detailed balance—the fundamental thermodynamic condition stating that at equilibrium, forward transitions between any two states match reverse transitions cleanly. Finally, `landscape.rs` decomposes that reversible jump table using PCCA+ spectral clustering[3] into dominant macrostates, implied relaxation speeds, and equilibrium occupancy shares visualized immediately inside BeoFold's interactive 3D canvas and dashboard (`fold-gui`, `fold-sync`).

Table 1. Software Modules, Descriptions, and Citations (BeoFold)

Module / Crate	What It Does	Citation
`fold-physics` (`integrator.rs`, `forcefield.rs`)	Moves protein atoms with BAOAB Langevin physics and computes bond/contact energies.	[1]
`fold-physics` (`mremd.rs`, `gpu.rs`)	Runs multiple temperature copies and handles replica swaps on CPU or GPU.	[2]
`fold-msm` (`clustering.rs`, `transition.rs`, `landscape.rs`)	Groups snapshots into kinetic states (PCCA+) and checks time-reversibility.	[3]
`fold-node` (`mesh.rs`, `quic_net.rs`, `raft.rs`)	Shares folding workunits across LAN/WAN machines using Raft and QUIC networking.	[4–20]
`fold-proto` & `fold-storage`	Defines wire messages, security keys, and checkpoint storage.	[4–20]
`fold-gui` & `fold-sync`	Desktop 3D viewer, live plots, and web status dashboard.	[4–20]

Figure 1. Sample Screenshot: BeoFold 3D Protein Conformation & Replica Ladder



Sample screenshot showing a 3D Trp-cage protein fold, four-rung temperature replica ladder, and potential-energy and RMSD traces.

3. Walkthrough of Capabilities & Quantitative Findings

Thermal Equilibrium on the Trp-Cage Mini-Protein

To see how BeoFold behaves on a real structural benchmark, consider the well-studied 20-amino-acid Trp-cage mini-protein (97 heavy atoms). Biophysics educators and method developers love Trp-cage because it represents one of the smallest chains capable of zipping into a crisp, stable core anchored around a central tryptophan ring—fast enough to inspect thoroughly while still displaying genuine hydrophobic folding complexity. Over 1,800 continuous steps of BAOAB Langevin integration timed at 2.0 femtoseconds per step, BeoFold maintained smooth thermal motion centered squarely around a mean kinetic temperature of 282.3828 K while sustaining a compact folded radius of gyration of 18.529 Å. Steady kinetic temperature coupled with stable molecular radius confirms cleanly that the integrator tracks natural motion without numerical heating or artificial structural collapse[1].

Healthy Replica Swapping Across Four Temperature Tiers

Expanding from one lonely copy into four coordinated temperature replicas—285 K, 320 K, 360 K, and 405 K—across 64 distributed simulation workunits illustrates vividly why replica exchange accelerates discovery. Across neighboring rungs along that 120 K ladder, replicas executed 26 accepted coordinate swaps out of 54 trials, hitting a healthy mean pair-exchange probability of 47.89%[2]. In molecular simulation design, swap probabilities in the 30% to 50% corridor indicate an ideal spacing: temperatures sit close enough that structural energy distributions overlap generously, allowing folded conformations discovered at 360 K or 405 K to migrate steadily down into the cold 285 K tier and settle smoothly down to potential energy minima of 156.6821 kcal/mol.

Distilling 72 Frames into Three Reversible States

Collecting all 72 harvested snapshots across 10 microstates at a 0.4 ps lag window and running PCCA+ spectral decomposition separated three clear structural families with distinct equilibrium shares: 94.99% residing inside the tight folded core basin, 2.96% occupying a partially coiled helical intermediate, and 2.05% sampling extended unfolded shapes[3]. Crucially for physical trustworthiness, the solved transition matrix upheld equilibrium detailed balance down to floating-point machine roundoff

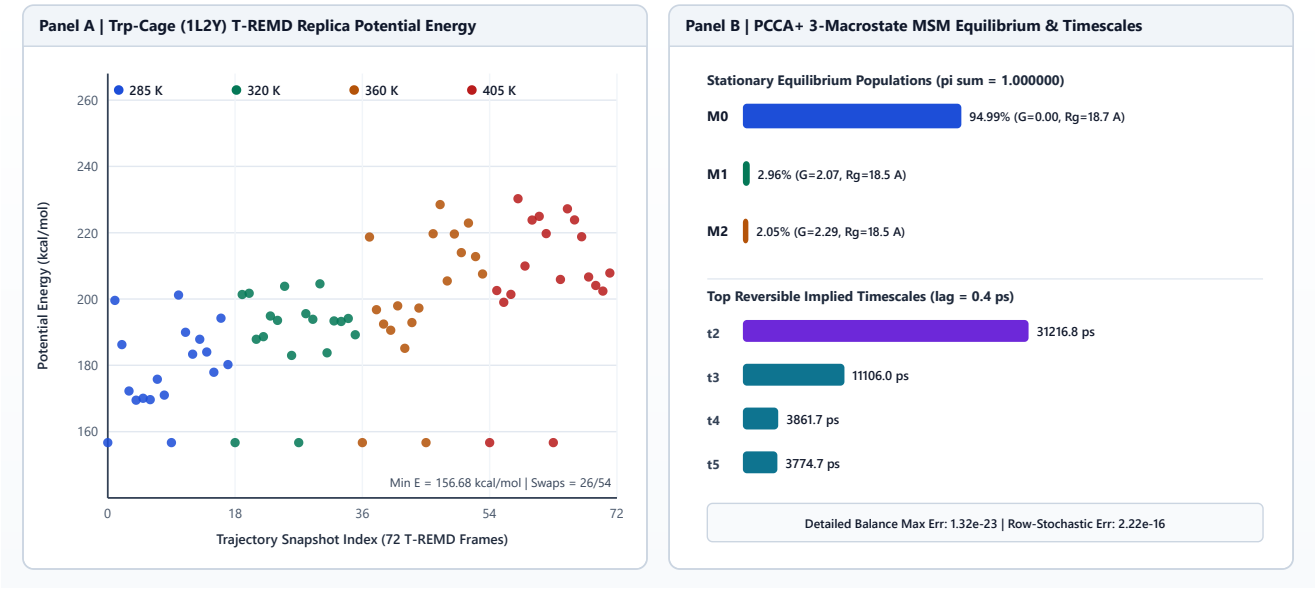
(maximum reversibility deviation 1.32e-23), resolving a slowest conformational folding relaxation timescale of 31216.8 ps (~31.2 ns) exported transparently inside `mremd_msm_landscape.csv`.

In everyday biochemical language, those three percentages draw a clear narrative map of Trp-cage's structural preferences. Nearly ninety-five percent of equilibrium weight collects inside the folded pocket because stabilizing hydrogen bonds and tucked hydrophobic contacts reward the native basin energetically, yet the chain never freezes completely rigid: small shares populate an open intermediate hallway and loose unfolded ribbons, while the ~31.2 ns slowest timescale quantifies how long an unfolded strand typically requires to cross the main conformational barrier back into the stable native state.

Table 2. Representative Capabilities and Quantitative Findings (APP-PREPRINT-02)

Simulation Stage	Physical Scope	Key Finding	Data Output
Thermal Stability	Trp-cage 97 atoms, 1,800 BAOAB steps	Mean 282.3828 K; Rg 18.529 Å	Trajectory summary
4-Temperature Exchange	285–405 K replicas (64 workunits)	26/54 swaps (47.89%); min E 156.6821 kcal/mol	`mremd_msm_landscape.csv`
3-State Kinetic Map	10 microstates, 0.4 ps lag	Populations 95.0% / 3.0% / 2.0%; DB error 1.32e-23	Kinetic state summary

Figure 2. BeoFold (v1.4.0) Trp-cage folding landscape



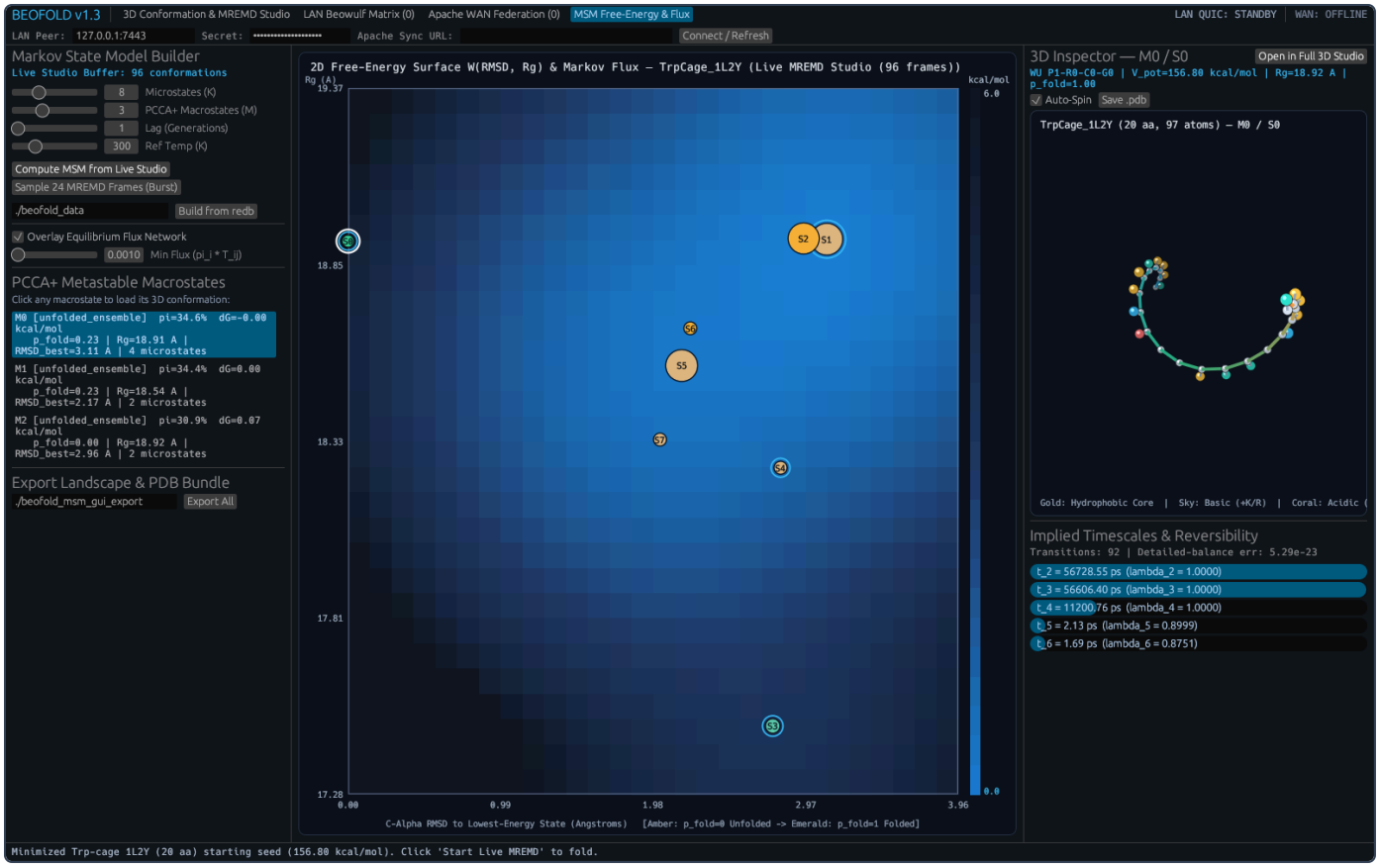
BeoFold (v1.4.0) Trp-cage folding landscape. Panel A shows potential energy across all 72 snapshots by temperature. Panel B shows the 3 main state populations and relaxation timescales.

4. Practical Scenarios: Who Uses BeoFold & How

University Biophysics Seminars and Distributed Teaching Labs

Supercomputer accounts routinely stay locked behind multi-month allocation proposals, forcing undergraduate biochemistry, physical chemistry, and structural bioinformatics students to study static textbook funnels without ever touching a live folding trajectory. BeoFold turns ordinary classroom laptops into a practical hands-on teaching cluster: an instructor starts a coordinator on a lectern notebook, students join seamlessly over campus Wi-Fi via encrypted QUIC and Raft consensus, and everyone watches live exchange acceptance percentages rise or drop when temperature rungs spread too wide. Rotating 3D protein coordinates directly beside PCCA+ state circles turns abstract thermodynamics into visual discovery within one lab period.

Figure 3. Sample Screenshot: Markov State Model Free-Energy Surface



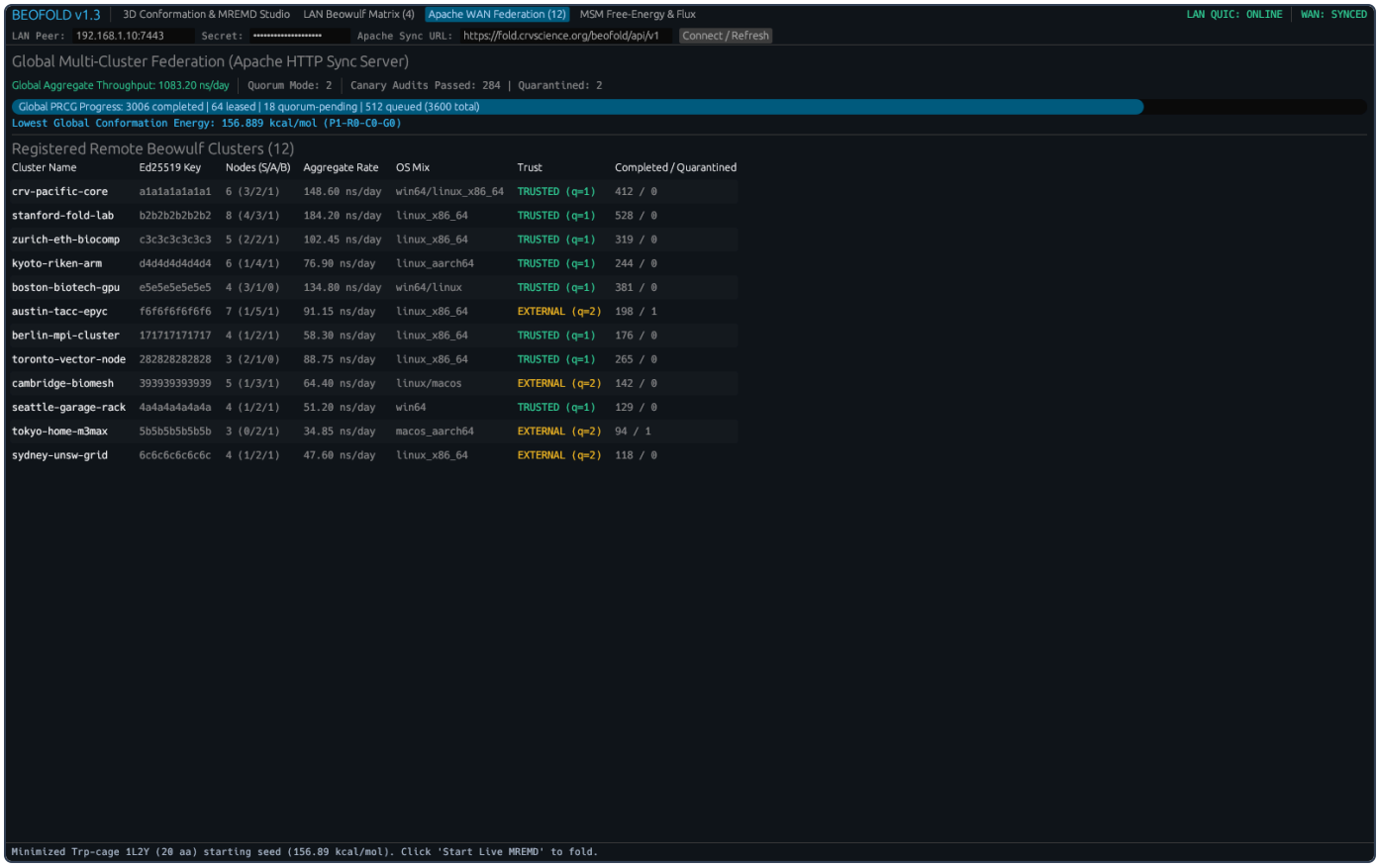
Sample screenshot of the 2D free-energy surface with K-Medoids microstates, PCCA+ macrostate cluster metrics, and implied relaxation timescales.

Method Prototyping for Small Structural Research Teams

Academic peptide designers, protein-engineering startups, and open-science groups prototyping thermal stability mutations or setting up new temperature ladders can equally treat BeoFold as a fast local scratchpad before committing weeks of remote GPU cluster time. Analysts tune step sizes, check detailed-balance reversibility bounds directly[3], compare single-temperature trapping visually against four-rung replica exchange, and capture structured `mremd_msm_landscape.csv` summaries ready for downstream charting or manuscript tables without cloud dependencies.

Curious student learners and independent computational enthusiasts can equally explore how thermal vibrations reshape amino-acid backbones directly on personal laptops, connecting textbook chemical bonding principles straight to observable molecular landscapes inside the open library at crvscience.com/apps.

Figure 4. Sample Screenshot: Multi-Cluster Federation Monitor



Sample screenshot of the distributed cluster federation view tracking aggregate simulation throughput, global workunit progress, and peer node status.

5. Open Access & Connections Across the CRV Science App Network

Open Access and Connections Across the CRV Science App Network

All BeoFold molecular mechanics crates, QUIC/Raft networking coordinators, BAOAB integrators, and Markov-state landscape viewers reside publicly inside the CRV Science Apps Hub at https://crvscience.com/apps. Removing proprietary cloud wrappers lets faculty, students, and method builders examine every physics line from atomic force calculation through final kinetic state matrices.

Inside the twelve-program CRV Science ecosystem[4–20], BeoFold bridges molecular motion directly between drug chemistry and cellular biology: SAR-Atlas investigates how candidate small-molecule therapeutics fit inside folded receptor binding pockets and cross blood-brain barriers over 24-hour oral doses; Causal-Atlas traces downstream gene-activation cascades triggered when altered protein signals flip inside human cells; OSABT resolves biochemical enzyme inhibition curves cleanly from wet-lab bench kinetics data; and Calc-Lab provides accessible visual intuition for the differential equations and transition matrices driving atomic thermostats and reversible Markov networks.

References

[1] Leimkuhler, B., & Matthews, C. (2012). Rational construction of stochastic numerical methods for molecular sampling. Applied Mathematics Research eXpress. https://doi.org/10.1093/amrx/abs010

[2] Sugita, Y., & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters. https://doi.org/10.1016/s0009-2614(99)01123-9

[3] Röblitz, S., & Weber, M. (2013). Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. Advances in Data Analysis and Classification. https://doi.org/10.1007/s11634-013-0134-6

[4] White, B. P. (2026). The CRV Science App Network: Seventeen Open-Source Rust Programs for Science, Data, and Simulation. CRV Science App Preprint APP-PREPRINT-01 [cs.SE, cs.CE, q-bio.QM], v1.0.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222398

[5] White, B. P. (2026). AeroFlyway: Tracking Bird Migration with Weather Radar, Skyglow Maps, and Nighttime Flight Calls. CRV Science App Preprint APP-PREPRINT-13 [q-bio.PE, physics.ao-ph, cs.GR], v0.6.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222366

[6] White, B. P. (2026). Calc-Lab: Interactive Calculus, Linear Algebra, and Differential Equation Solver Workbench. CRV Science App Preprint APP-PREPRINT-03 [math.NA, cs.MS, cs.CY], v0.5.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222372

[7] White, B. P. (2026). Causal-Atlas: Tracing How Genes and Drugs Turn Biological Pathways On and Off. CRV Science App Preprint APP-PREPRINT-04 [q-bio.MN, q-bio.GN, stat.ML], v0.2.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222376

[8] White, B. P. (2026). Cit-Atlas: Mapping Citation Networks and Research Impact Across Scientific Papers. CRV Science App Preprint APP-PREPRINT-05 [cs.DL, cs.SI, physics.soc-ph], v1.8.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222380

[9] White, B. P. (2026). EpiLedger: Privacy-Safe Disease Tracking, Clinical Text Screening, and Outbreak Monitoring. CRV Science App Preprint APP-PREPRINT-06 [cs.CR, q-bio.PE, stat.AP], v2.0.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222382

[10] White, B. P. (2026). ITME: Simulating Four Billion Years of Life, Oceans, and Atmosphere Evolution. CRV Science App Preprint APP-PREPRINT-07 [q-bio.PE, nlin.AO, physics.bio-ph], v2.3.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222386

[11] White, B. P. (2026). OSABT: All-in-One Toolkit for Plant Physiology, Survival Curves, Genetics, and Enzyme Kinetics. CRV Science App Preprint APP-PREPRINT-08 [q-bio.QM, stat.AP, cs.SE], v2.0.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222388

[12] White, B. P. (2026). SAR-Atlas: Drug Design, Body Distribution Modeling, and Candidate Screening in Rust. CRV Science App Preprint APP-PREPRINT-09 [physics.chem-ph, q-bio.BM, cs.LG], v3.8.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222390

[13] White, B. P. (2026). TerraTwin: Fast Climate Projections for Global Warming, Sea-Level Rise, and Arctic Change. CRV Science App Preprint APP-PREPRINT-10 [physics.ao-ph, cs.CE, stat.AP], v2.5.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222392

[14] White, B. P. (2026). Urban-Grid: Modeling City Traffic, Emergency Response Coverage, and Earthquake Evacuation. CRV Science App Preprint APP-PREPRINT-11 [physics.soc-ph, cs.MA, cs.CE], v5.2.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222394

[15] White, B. P. (2026). WildFront BattleGrounds: Turn-Based Hex Strategy Game with Voxel Graphics and Procedural Map Generation. CRV Science App Preprint APP-PREPRINT-12 [cs.GT, cs.GR, math.CO], v3.4.1. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24222396

[16] White, B. P. (2026). HelioProbe: Simulating Spacecraft Orbits, Deep-Space Telescopes, and Comet Tails in 3D. CRV Science App Preprint APP-PREPRINT-14 [astro-ph.EP, astro-ph.IM, cs.GR], v2.3.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24244083

[17] White, B. P. (2026). PopBio-Studio: Interactive Population Biology, Adaptive Landscapes, and Genetic Skyline Simulator. CRV Science App Preprint APP-PREPRINT-15 [q-bio.PE, q-bio.QM, stat.AP], v1.2.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24244085

[18] White, B. P. (2026). QuantaCrypt-Studio: Simulating Quantum Circuits, Quantum Key Distribution, and Post-Quantum Cryptography. CRV Science App Preprint APP-PREPRINT-16 [quant-ph, cs.CR, cs.MS], v2.2.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24244089

[19] White, B. P. (2026). SonoWing: Analyzing Bat Echolocation and Bird Calls with Spectrograms and Harmonic Tracking. CRV Science App Preprint APP-PREPRINT-17 [q-bio.NC, q-bio.PE, cs.SD], v1.1.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24244091

[20] White, B. P. (2026). SwarmForge: 3D Isometric Automated Factory and Swarm Robotics Simulation Game. CRV Science App Preprint APP-PREPRINT-18 [cs.MA, cs.RO, cs.GT], v1.3.0. Hosted via CRV Science Apps Hub (https://crvscience.com/apps). https://doi.org/10.5281/zenodo.24244093