Arc Science
Technical report

Snöggo

Danil Kotelnikov

Abstract. Benzimidazole drugs lose efficacy in helminth populations that carry beta-tubulin substitutions at F167, E198 and F200, and whether these substitutions change the free energy of drug binding is an open question. Snöggo computes that change by bidirectional non-equilibrium alchemical switching with Bennett acceptance-ratio estimation, over cycles whose bound and free legs share one Amber14SB/GAFF2/TIP3P parameter set. The study design is serialised and hashed before any sampling, and a result that fails a registered diagnostic is reported as indeterminate. A charge-changing substitution is handled by transforming a bulk counter-ion in the opposite direction. The remaining term is reported without a continuum correction, because a Poisson–Boltzmann correction would double-count dielectric burial. Release 1.0.0 reports the design and its verification. Qualification of the engines on the target hardware is the next stage.

In short

Snöggo is a pipeline that computes how a point mutation changes the free energy of ligand binding, with every parameter and threshold fixed and hashed before sampling. Release 1.0.0 is the control and analysis software with its registered study design.

1. Thermodynamic cycle

The quantity of interest is the change in binding free energy caused by a mutation. The cycle obtains it from two mutation steps in place of two binding steps. The side chain is transformed once in the ligand-bound protein and once in the free protein, and the mutation free energy of the bound leg minus that of the free leg gives the change in binding free energy. In the cycle described for E198A, the horizontal arrows are the alchemical transformations and the vertical arrows are the binding steps.

Both legs are built from one shared assignment of force-field parameters, so a difference between two variants cannot come from a difference in force-field treatment. Closure of the cycles over the network of neutral mutations is computed and compared with a registered threshold. The same bootstrap draws for the free-protein leg are reused across ligands, so that the shared contribution of that leg cancels correctly when two drugs are compared.

2. Alchemical paths

A mutation free energy is obtained by non-equilibrium switching. Independent equilibrium simulations run at both physical end states of a leg, and short transitions of fixed length are started from each in both directions. The work values of the forward and reverse transitions are combined with the Bennett acceptance-ratio estimator. The two wider profiles run a published reference protocol calibrated on 134 mutations: five independent 3 ns equilibrium simulations per leg and direction, each starting 30 transitions of 80 ps, which gives 150 transitions per direction. The reference reports a root-mean-square error of 1.2 kcal/mol against experiment. That figure describes the published method.

E198A removes a formal negative charge. In a periodic box a change of net charge introduces finite-size artefacts that a charge-conserving mutation does not have. The pipeline keeps the box neutral: a counter-ion in bulk solvent is transformed in the opposite direction within the same simulation, so the total charge is zero at every value of the coupling parameter. A flat-bottom restraint holds the ion in a shell 1.7 to 2.3 nm from an anchor on the protein. A first geometry implied 90.4% occupancy of that shell against a required 99%. Moving the restraint walls 0.1 nm inward gives a predicted occupancy of 0.999991. Both values are analytic, calculated from the radial Boltzmann distribution.

3. Force fields

The Hamiltonian is Amber14SB for the protein, GAFF2 for the ligand and TIP3P for water. Hybrid topologies, which hold both end states of a mutation in one system, are built with the pmx software. One parameter assignment serves the bound and free legs of every variant. The sources name no learned component. The free-energy estimates described here come from this classical force field and from the work values of the switching transitions.

The neutral construction leaves one term untreated. A discrete-solvent contribution persists at infinite box size and originates in the water model, so a continuum solver cannot supply it. A correction based on Poisson–Boltzmann calculations would double-count dielectric burial already present in the explicit-solvent free energy. Release 1.0.0 therefore carries no such correction, the software refuses to compute one, and the claim that depends on it stays blocked.

4. Workflow

A campaign begins with a proposal that lists the whole study: structural templates, ligand microstates, mutation edges, replica counts, simulation and transition lengths, temperature, ionic strength, every quality-control threshold, the adjudication rule and the wall-clock budget. The proposal is serialised in canonical form and hashed with SHA-256, and the digest accompanies every later artefact. Parameters outside the registered bounds are refused. Observables whose methods are not qualified stay in the proposal with the status blocked, and every case outside the chosen scope is recorded with its reason.

One command installs a hash-verified runtime, qualifies the engines, freezes the proposal, executes the campaign and exports the evidence. Every result is committed under a content digest. An interrupted stage resumes only through a recovery plan that an operator has reviewed, and the worker stops the engine at the registered hard stop with committed results kept. Three profiles can be launched. The default covers the neutral cycle over the wild type, F167Y, F200Y and the double mutant with 880.8 ns of planned simulated time, and the two wider profiles plan 1691.0 ns and 2209.0 ns. Their wall-clock figures are scaling arithmetic.

5. Qualification

Three kinds of evidence are kept apart. Portable software checks show that the control and analysis code behave as specified, and they execute no scientific engine. Engineering smoke runs execute the native cases on general-purpose hardware, which shows that the engines run. Qualification is the execution of registered acceptance cases on the target machine itself, and only a passing record activates a runtime. Scientific results are a fourth category, produced by a completed campaign.

A separate gate asks whether the committed sampling supports a claim at the registered margin of 1 kcal/mol. The familywise confidence interval across registered contexts must fall within the margin. No single replica may control the estimate, and the first half of the replicas must agree with all of them. The drift and overlap diagnostics must hold, and no single transition may move a leg by more than a quarter of the margin. A wide interval leads to a proposal for more sampling, which an operator reviews. An unstable estimate is reported as indeterminate with the failing diagnostic named.

6. Limits

Release 1.0.0 holds the workflow for free-energy, docking and quantum-mechanical calculations. Qualification of the engines on the target hardware is the precondition for any result, and the results section of the preprint follows it. The charge-changing substitution E198A yields numbers with no associated claim until the residual correction is redesigned. A fourth profile, which includes absolute binding free energies, is registered and cannot be launched.

The contribution is an execution discipline around a published estimator. A systematic review cited by the preprint finds insufficient peer-reviewed evidence to establish the association between these polymorphisms and phenotypic resistance across human hookworms. The repository states that the outputs are conditional computational estimates and that no output is a statement about a resistance mechanism in an organism. References to the preprint give its sections.

System design

Stages of a campaign as implemented in release 1.0.0; the residual correction is withheld.

Implemented components

- Locked runtime.
- Qualification cases.
- Frozen proposal.
- Time budget.
- Hybrid topologies.
- Equilibrium replicas.
- Switching transitions.
- Acceptance-ratio estimate.
- Registered diagnostics.
- Familywise intervals.
- Claim margin gate.
- Claim matrix.

Specified components

- Residual correction.

Connections

- Locked runtime → Qualification cases.
- Qualification cases → Frozen proposal.
- Frozen proposal → Time budget.
- Time budget → Hybrid topologies.
- Hybrid topologies → Equilibrium replicas.
- Equilibrium replicas → Switching transitions.
- Switching transitions → Acceptance-ratio estimate.
- Acceptance-ratio estimate → Registered diagnostics.
- Registered diagnostics → Familywise intervals.
- Familywise intervals → Claim matrix.
- Claim matrix → Claim margin gate.
- Claim margin gate → Time budget.
- Residual correction → Claim matrix.

Source

This report is the printed form of the page https://arcscience.app/snoggo/. The page is the source of record for every statement in it.