Short Proposal

Gravitational physics meets Computational biology

Key question: Somewhat surprisingly, it remains unknown if the Molecular **Dynamics** simulations, popular tool of structural biology, gives correct **dynamics** of biomolecules; this is because the extended ensemble methods used are aimed at calculating **averages**, not dynamics.
Key result: Mathematical proof if the dynamics are correct or not.
Key impact: A 'YES' would set the approach on the firm footing it deserves; a 'NO' would initiate a new field of study to look into the consequences of and remedies to this.

---

1. Research idea and context

Starting point. Molecular Dynamics (MD) simulations has become a hugely prevalent tool for studying the structure and **dynamics** of biomolecules. It is especially useful when molecules are intrinsically unstructured in their biologically relevant state—such as lipids forming the core of the cell membrane or the intrinsically disordered proteins—so a single snapshot provided by a crystal structure is not fully informative, but rather an ensemble of typical structural states, with the rates of movement between them, is preferred. MD can deliver just this dynamic view. Or so it is believed.

Open question. To be precise, nobody knows if MD really can reach the above proclaimed interrogative power, see 'On extended ensembles'. That is, we do not know if, mathematically speaking, our choice of equations of motion leads to correct **dynamics** of the biomolecules under study. While it is reasonable to assume that this is the case—as the equations do give correctly weighted averages—the real proof is lacking. Here we wish to pursue this matter.

Relevance. Thousands of scientific articles describing MD simulations of biomolecules are published yearly, and most of these discuss **dynamics**: protein folding pathways, mechanisms of viral assembly, enzyme–ligand dynamics, phospholipid diffusion mechanisms, untwisting of DNA, *et cetera*. Scrutinizing the underlying assumption of this huge body of work is a highly relevant task.

Perspective. The project has two possible successful outcomes: 1) NO, the currently used extended ensemble methods do not produce the correct dynamics; this would initiate a new research field to study the consequences of the problem and its possible remedies. 2) YES, the

extended ensemble MD does indeed produce the correct dynamics. Proving this would give this widely used research tool a firm mathematical foundation—a very relevant fundamental result.

On extended ensembles. Classical MD solves Newton's equations of motion for N atoms within volume V; solutions conserve the total energy E. With N, V, and E fixed, any snapshot along the time-integrated trajectories appears as if coming from the *microcanonical* thermodynamic ensemble. This motivates the *ergodic hypothesis* (EH): Time averages over these *NVE* trajectories are equal to the microcanonical ensemble averages. If EH holds, then by accurately solving Newton's equations one can use MD to sample the microcanonical ensemble.

As the N-particle equations of motion are not analytically solvable, one must resort to numerical time-integration. But it turns out that no numerical integration scheme can provide the exact solution needed to invoke EH. Interestingly, however, one class of numerical integrators, the *symplectic* integrators, has the property of exactly solving not the original system, but a system (in a mathematical sense) close to it. Thus symplectic integrators indeed produce exact solutions (of the close-by *NVE* system), and one can argue by EH that the time averages are indeed equal to the microcanonical ensemble averages. Bottom line: In its basic form, the MD trajectories describe *NVE* dynamics correctly.

However, current biomolecular simulations do not use MD in its basic form. The equations of motion are modified to impose fluctuations on E such that the system temperature T has a distribution characteristic of the *canonical* ensemble (average(T)=T_b, variance(T)=$2T_b^2/3N$, with T_b the fixed temperature of the surroundings); the *isothermal–isobaric* ensemble is reached when also V is made to fluctuate to control the pressure p. Importantly, while these modifications, also known as *extended ensemble methods*, do allow sampling from thermodynamic ensembles closer to the experimental setting (one typically does not control E of a biomolecular experiment, but rather the T and p of the surroundings), the **modifications do not need to maintain realistic dynamics**.

2. Proposed solution or concept

The extended ensemble MD, unlike classical MD in *NVE*, does not preserve the volume of the *phase space* ($6N$-dimensional space of the positions and momenta of the N atoms). That is, whereas the trajectories produced by MD in *NVE* exist in flat space, to the extended ensemble trajectories the phase space appears curved. In fact, here things might be even weirder, as there are indications that in the extended ensemble simulations the *metric* of the phase space might be time dependent. In other words, the distance between structures A and B (of say, a protein) might

not depend only on **what** are the positions and velocities of the atoms in A, but also **when** they are. This deep interconnectedness between the time and space coordinates aroused the interest of a (rather brilliant) post doc candidate, who in his PhD (to be handed in this autumn) specialized in the geometric analysis in Gravitational physics. He proposed that we try out the following.

Unique approach: Import mathematical tools from Gravitational Physics to understand MD

We believe that as the extended ensemble MD approaches can be interpreted as evolution of the phase space geometry, the tools used by the mathematical relativity community to tackle the black hole stability problem (such as the Energy and Morawetz estimates on curved backgrounds) and the tools developed by the numerical relativity community to deal with evolving metrics would find interesting use also in the MD setting.

---

3. Objective of the exploratory phase

We apply the tools developed in the context of Gravitational physics to analyze the phase space structure and evolution caused by different deterministic MD extended ensemble techniques, such as the widely applied Nose–Hoover thermostat and Parrinello–Rahman barostat. In particular, we will characterize if solutions for these systems exist, are unique, and are stable under small perturbations.

Tasks: 1) Formulate the question in precise mathematical language: Which equations, under which conditions, have to satisfy which properties. Note that formulating the problem in the language of the mathematical relativity community will already increase its visibility and make it accessible for many more skilled people to work on. 2) Estimate how difficult the mathematical problem itself is. 3) Work out the mathematical problem. Time permitting: 4) utilize thus gained understanding of the mathematical structure of these systems to develop optimal numerical solvers for them, similar to symplectic integrators understood to be optimal for *NVE*.