Key Points:

* With the recent success of deep-learning, some say that protein structure prediction has reached the single-structure frontier. Conformational ensembles are said to be the future of structural biology. Datasets spanning protein conformational landscapes are becoming increasingly available. Experimental methods and generative learning models bring about static datasets, while molecular dynamics simulations provide dynamic datasets.
* EnGens (**En**semble **Gen**erator) pipeline can process both static and dynamic datasets of protein structure to extract sets of representative conformations (representative conformational ensembles) thus summarizing the complex information contained within the data. To this end, EnGens collects a long list of tools and methods previously proposed for unsupervised learning from protein structure data.
* In this work, EnGens pipeline is showcased on a set of examples from the literature including a large protein complex (PI3K), a peptide drug (Compstatin) and a small molecule (Nelfinavir). EnGens shows the ability to process these datasets faster and provide better insights than previous manual analysis.
* Representative ensembles produced by EnGens will aid downstream tasks related to drug design such as ensemble docking and drug-target interaction prediction.
* EnGens pipeline is available at https://github.com/anon528/supreme-couscous as a python package, wrapped inside a Docker image and accompanied with a set of interactive Jupyter Notebooks. EnGens is accessible to researchers with little to no coding experience. For a computationally experienced audience, EnGens can serve as platform for further algorithmic development.