17th Global Summit on
Stem Cell & Regenerative Medicine

15th International Conference on
Genetic Disorders and Gene Therapy

conferenceseries.com

November 22-23, 2021

WEBINAR

Eric J Sorin, J Stem Cell Res Ther 2021, Volume 11

Overcoming the heuristic nature of K-means clustering: Identification and characterization of binding modes from simulations of molecular recognition complexes

Eric J Sorin
California State University at Long Beach, USA

States in computational data is a non-trivial problem, particularly when the number of states is unknown a priori and for large, flexible chemical systems and complexes. To this end, we report a novel clustering protocol that combines high-resolution structural representation, brute-force repeat clustering, and optimization of clustering statistics to reproducibly identify the number of clusters present in a data set (k) for simulated ensembles of Butyrylcholinesterase in complex with two previously studied organophosphate inhibitors. Each structure within our simulated ensembles was depicted as a high-dimensionality vector with components defined by specific protein-inhibitor contacts at the chemical group level and the magnitudes of these components defined by their respective extents of pair-wise atomic contact, thus allowing for algorithmic differentiation between varying degrees of interaction. These surface-weighted interaction fingerprints were tabulated for each of over one million structures from more than 100 μs the accurate and reproducible detection and description of thermodynamic of all-atom molecular dynamics simulation per complex and used as input for repetitive k-means clustering. Minimization of cluster population variance and range afforded accurate and reproducible identification of k, thereby allowing for the characterization of discrete binding modes from molecular simulation data in the form of contact tables that concisely encapsulate the observed intermolecular contact motifs. While the protocol presented herein to determine k and achieve non-heuristic clustering is demonstrated on data from massive atomistic simulation, our approach is generalizable to other data types and clustering algorithms, and is tractable with limited computational resource.

Biography

Eric J. Sorin is a physical chemist with expertise in statistical mechanics and the modeling and simulation of bimolecular systems. Primary areas of study in his lab include RNA folding and non-covalently bound molecular recognition complexes, with a focus on the interactions that stabilize binding of inhibitors of the cholinesterase family of enzymes. Sorin was a founding member of the Folding@Home distributed computing project, which employs processors donated by the public from around the globe to collect simulations of bimolecular systems associated with human health and disease. He enjoys pursuing cutting-edge research with a diverse and talented pool of undergraduate and graduate students in his native Southern California.