Thomas Schiex

Automated reasoning and learning for Computational Protein Design

Computational protein design aims at designing new useful proteins from first principles, by reasoning on 3D atomic forces. It defines a complex decisionproblem that has been ultimately reduced to more manageable formulations, as NP-hard discrete energy optimization problems over large discrete MarkovRandom Field. Traditionally solved using Monte Carlo approaches, recent Artificial Intelligence algorithms are able to provably solve these problems onsignificant cases. Because these formulations remain approximate, information extracted by Machine Learning (ML) from protein sequences, structures orother experimental data can be used to refine or even replace physics-based energy functions. However, design often requires to produce proteinsequences which satisfy radically new properties, something that remains challenging for most ML approaches.