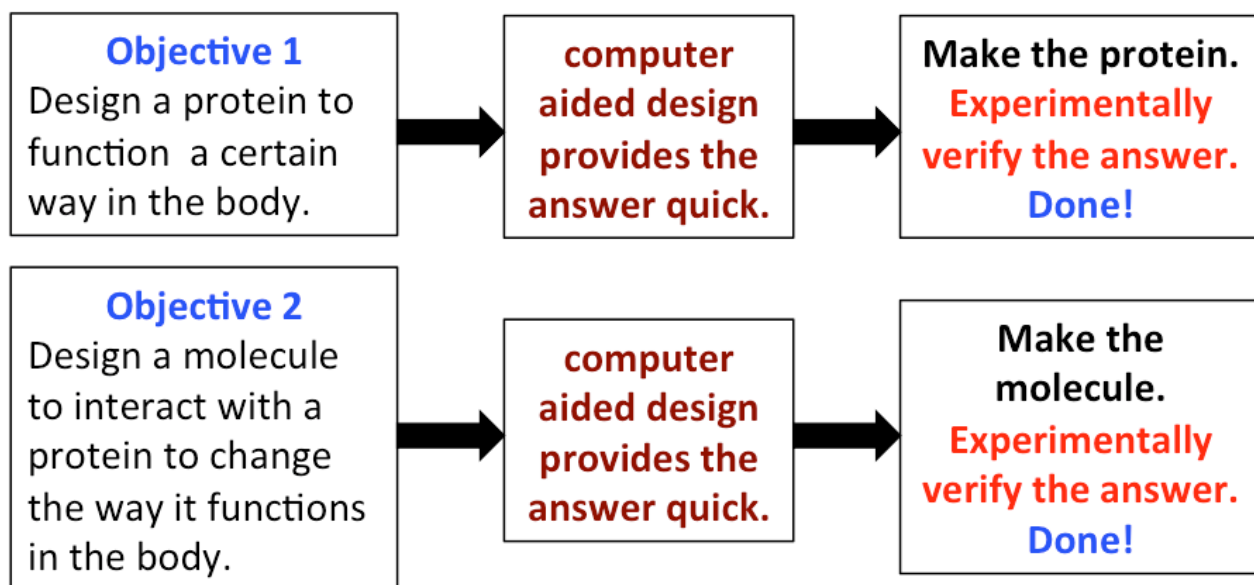


Drug Discovery is a Grand Challenge for the 21st Century

Proteins are macromolecular machines that carry out biological functions in living organisms, making them an important type of biomolecule to study for a myriad number of applications that range from agriculture to human health. The proteins produced in living organisms are very special because they support life on Earth. From a mechanistic point of view, understanding how proteins function is a step toward the ability to prevent and cure diseases through molecular engineering. In order to understand how a protein functions, its detailed mechanical and thermodynamic properties must be understood. Ultimately it is the atomic motions of a protein that control its biological function as a physical process. Imagine the following scenario that *inaccurately* depicts the grand challenge:



The grand challenge presents difficult problems that must be solved before such a computer-aided approach could ever be possible. Below is a short list of problems (very incomplete) that immediately present themselves, which makes the above scenario too simplified of a workflow. However, note that the phrase “**Experimentally verify**” includes every single experimental method at every stage of research and development, including all aspects of clinical trials involving short and long-term studies.

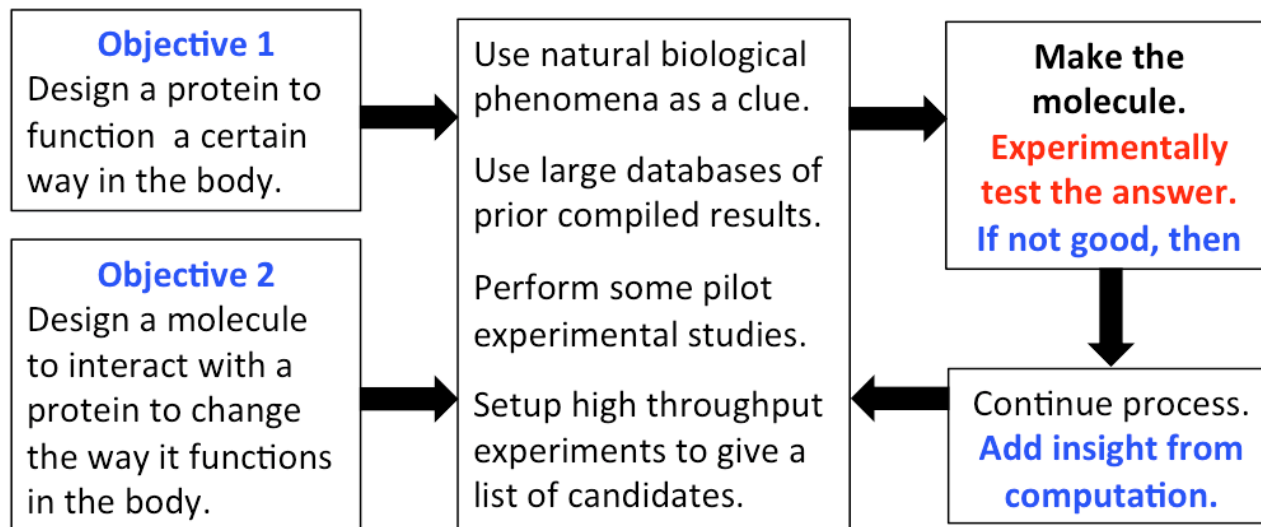
List of problems that need to be overcome to meet the grand challenge:

- 1) The environment of a protein consists of a multitude of other molecules that it interacts with. Therefore, the design of a protein must take into consideration the solvent conditions that the protein will be exposed to. Note that a protein may be transported to different environments.
- 2) To control function, one must track how molecules of all types in the body will interact with the designed protein in different environmental conditions. Therefore, successful design of function depends on what the protein should not interact with, in addition to what it should interact with.
- 3) Finding “**the** answer” is of measure zero not because of the improbability of finding such an answer, but rather, there will invariably be a multitude of different ways to achieve the desired outcome. Perhaps some answers will be better than others, but due to the nature of the complexity found in biological systems as empirically evident from evolutionary biology, almost surely a multitude of different answers will achieve the desired result. (Thank God!)
- 4) Assuming a rank ordered list of predictions are given, instead of one, there is a high likelihood that despite having a very large number of predictions, the designed protein or molecule will not work. Attempting to experimentally verify the computational prediction will most frequently

fail. This verification process will have many shades of grey because most of the predictions will not work at all. The majority of the remaining predictions will only minimally suffice, while just a tiny fraction of designs (if any) will have the superior characteristics that are desired.

- 5) Only after extensive long-term clinical testing, it isn't possible to know a priori what the desired characteristics are! An idea of targets can be surmised initially, but due to the sheer complexity of the biological system (how the body as a whole responds to perturbations) one cannot specify the desired characteristics due to the overwhelming amount of potential molecular interactions. However, we know superior characteristics when we see them based on the effectiveness of the drug and its associated unintended side effects that define risk levels.

The next scenario **accurately** depicts the grand challenge according to the current state of affairs:



The actual added value of insight that is obtained from computation is limited for two reasons. First, accuracy is poor when it comes to calculation of free energies because entropy effects are poorly estimated. The second reason is the rate at which predictions can be made is much slower than how fast high-throughput experimental screening can be performed. Nevertheless, simulation results and other computer-based predictions help the drug discovery process, which is why they are used and are included in the workflow diagram above. However, unlike the first scenario, computational results make exceptional use of the human brain for interpretation of the results. The computational results can help a scientist better understand mechanisms at a detailed level, and from this insight, allow the scientist to tune experiments to more effectively filter out poor candidates. The essential idea of the drug discovery process is to perform trial and error inquires with as many constraints as possible to reduce the time it takes to get “lucky” in finding a solution. In short, computational models are used to increase “luck”, but they are insufficient for leading the drug discovery process, and this deficiency presents many opportunities in the field of protein chemistry that are aligned with my interests.