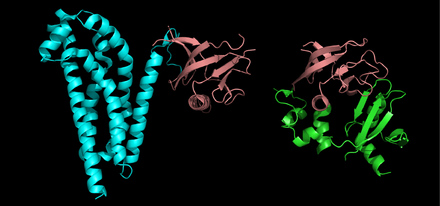


Next-generation sequencing technologies

Source: [Biochemical Society Transactions | Portland Press](https://portlandpress.com/biochemsoctrans/issue/46/5)

**Business use case:** non-redundant representative subsets of protein sequence using UD-MIS

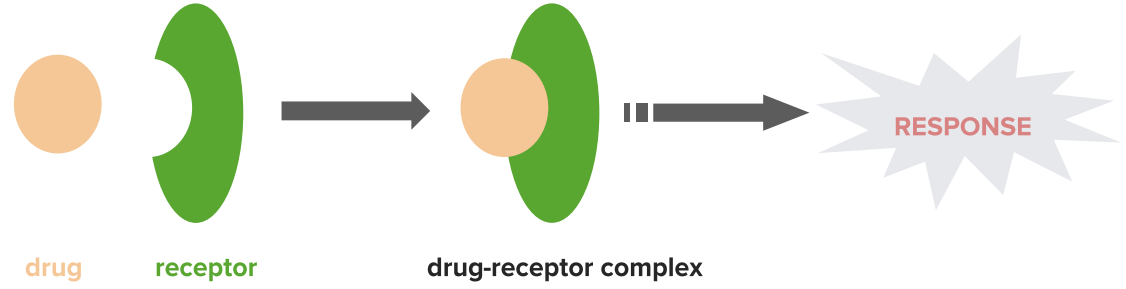
The application of UD-MIS for choosing non redundant representative sets of protein from sequence data sets can be served in a variety of applications. For instance, selecting representative sets of protein structure which is one of the biology problems in the world.



Two protein structure

Source: https://www.wikiwand.com/en/Protein\_structure

Protein structure is the three-dimensional arrangement of atoms in an amino acid-chain molecule. It enables us to understand ligand protein binding and develop binding affinity techniques [1]. Binding affinity is the strength of the binding interaction between a single biomolecule (e.g. protein or DNA) to its ligand/binding partner (e.g. drug or inhibitor) [2].



Drug receptor

Source: <https://www.lecturio.com/magazine/biological-interaction/>

**Business pitch.**

**Are you thinking how to increase your market share? Do you want to make more effective medicine with less side effects to the human body? Here is our reliable solution:**

It becomes possible to firstly predict how and where a particular small molecule might interact with a protein and then to identify putative ligands for a specific protein site [1].

Targeting specific chosen protein to evaluate ligand affinity can be done by selecting and analyzing representative sets of protein structure. Target-oriented drug design aims to find a high affinity ligand that would bind the target protein in order to block its disease-associated function catalytic activity or interaction with other molecules.

Structural knowledge on the exact interactions of drugs with their target protein has been applied mainly to predict potency changes of chemically modified lead compounds in protein structure-based drug design industry [3][4].

**Drug design industry overview**:

In recent years, the drug discovery technologies industry has seen remarkable developments and rapid progress. Biology, chemistry and computation provide an amalgamation that is reflective of the industry’s future.

The 10 largest pharmaceutical companies now pay nearly $80 billion a year to come up with fewer and fewer successful drugs. The reason is that drugs that are easiest to find and that safely and effectively treat common disorders have all been found; what is left is hunting for drugs that address problems with complex and elusive solutions and that would treat disorders affecting only tiny portions of the population and thus could return far less in revenue. Because finding new, successful drugs has become so much harder, the average cost of bringing one to market nearly doubled between 2003 and 2013 to $2.6 billion. These same challenges have increased the lab-to-market time line to 12 years, with 90 percent of drugs washing out in one of the phases of human trials. Many scientists in the field think that computer-aided techniques will ultimately improve drug development in several ways such as identifying more promising drug candidates and by speeding up the overall process. Where the next generation sequencing techniques can make huge impact is having drugs that fail early on, before we make all that investment in them. That’s why Every one of the major pharmaceutical companies has announced a partnership with at least one of the AI-based drug-discovery start-ups. They raised more than $1 billion in funding in 2018, and as of last September, they were on track to raise $1.5 billion in 2019.

According to the market research reports, the global market for drug discovery should grow from $69.8 billion in 2020 to $110.4 billion by 2025 with a compound annual growth rate (CAGR) of 9.6% for the period of 2020-2025[5].

**Benchmarking:**

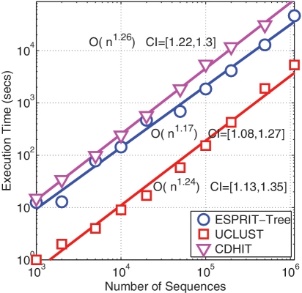
There are several conventional methods to select non-redundant representative subsets of protein sequence. Our technology has dramatically promoted sequencing technologies by offering low-cost and ultra-high-throughput sequencing. In addition to accuracy, computational complexity is an important issue.

Previous methods for this task, such as CD-HIT, PISCES and UCLUST, apply a heuristic threshold-based algorithm.

Next-generation sequencing technologies are able to demonstrate a good speedup derived from solving UD-MIS problem quantumly compared with conventional famous methods based on the Threshold algorithm.

Table: Computational Complexity (time)

|Conventional Method|US-MIS Quantum|



Source: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3251834/>

Computational Costs: Our technologies to provide non-redundant representative subsets of protein sequence reduces computational complexity which result in a lower computational cost.

The next generation sequencing techniques can cut the time spent in discovery from 4.5 years to as little as one year, reduces discovery costs by 80 percent and results in one-fifth the number of synthesized compounds as is normally needed to produce a single winning drug.

The “production” costs are accounted for:

* Labor, administration, management, utilities, reagents, and consumables
* Sequencing instruments and other large equipment (amortized over three years)
* Informatics activities directly related to sequence production (e.g., laboratory information management systems and initial data processing)
* Indirect Costs as they relate to the above items

**Potential customer**:

Company name: Pfizer

Country: USA

Profit: $ 41.9 billion – 2020

Company name: AstraZeneca

Country: United Kingdom

Profit: $ 26.6 billion – 2020

Company name: Regeneron pharmaceuticals:

Country: USA

Profit: $ 8.497 billion – 2020

Company name: Sygnature Discovery

Country: United Kingdom

Profit: $ 68.9 million – 2020

**Competetors:**

[1] <https://portlandpress.com/biochemsoctrans/article-abstract/46/5/1367/67796/Protein-structure-and-computational-drug-discovery?redirectedFrom=fulltext>

[2] [Binding Affinity | Dissociation Constant | Malvern Panalytical](https://www.malvernpanalytical.com/en/products/measurement-type/binding-affinity)

[3] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4236214/>

[4] [Frontiers | Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview | Pharmacology (frontiersin.org)](https://www.frontiersin.org/articles/10.3389/fphar.2018.00923/full)