

Target 2035 – Pharmacological modulators for all human proteins

To paraphrase Sydney Brenner: “Progress in science and drug discovery depends on new techniques and reagents, new discoveries and new ideas, probably in that order.”

And among all the possible techniques and reagents, the bibliometric evidence clearly shows that pharmacological modulators (such as chemical probes) have the greatest scientific impact (Edwards et al, Nature 470:163 2011). *For nearly every protein for which there is a high quality and openly available chemical probe, the original paper describing the probe is the most highly cited paper on that protein.* Pharmacological modulators are also key parts of the pharmaceutical industry’s target validation arsenal as they have a variety of features that are complementary to genetic approaches: they are often reversible, have immediate action, can target specific functions, provide dose-dependent response, more closely mimic therapeutics, can be used on many cell types, including primary cells, may be useful in vivo, etc.

In the mid 1990’s, many bandied about the concept of creating a pharmacological modulator for every human protein. At the time, this was both popular and simplistic to state, and also sounded “forward thinking” and innovative, but there was no practical plan. The situation has changed. In the intervening years, largely due to the efforts of the pharmaceutical sector and its support for public-private partnerships like the SGC and the IMI, progress has been made and an actionable plan has emerged. Importantly, the notion that probe discovery should be ‘pre-competitive’ has also been embraced.

We believe that with the correct organization that brings in, or structures, complementary efforts around the world (for example, as a group of federated projects), and the active participation of the pharmaceutical industry, as well as with some technology development, we can aim to generate potent and well-characterized functional modulators for nearly all protein in the human genome by 2035, with an intermediate goal of developing potent, well-characterized and reversible functional modulators for a targeted set of proteins by 2025.

There are many scientific, organizational and financial questions that must be considered, likely best within a smaller group of experts.

Potential five and ten-year plans

The path to 2035 could be structured as two phases.

The first phase, from 2020-2025, would be structured to build the foundation for a concerted global effort, and would aim to collect, characterize and make available existing pharmacological modulators for key representatives from all proteins families in the current druggable genome (~4,000 proteins), as well as to develop critical and centralized infrastructure to facilitate data collection, curation, dissemination, and mining that will power the scientific community worldwide. This phase might also create centralized facilities to provide quantitative genome-scale biochemical and cell-based profiling assays to the federated community, as well as to coordinate the development of new technologies to extend the definition of druggability. This first phase will complement and extend ongoing efforts to create chemical tools and chemogenomics libraries to blanket priority gene families, such as kinases and epigenetics families.

The second phase, from 2025-2035, will be to apply the new technologies and infrastructure to generate a complete set of pharmacological modulators for > 90% of the ~20,000 proteins encoded by the genome.

“Target 2035” sounds ambitious, but its concept and practicality is on firm ground based on a number of pilot studies, which revealed the following success parameters:

1. Collaborate with the pharmaceutical sector to access unparalleled expertise, experience, materials, and logistics
2. Establish clear and quantitative quality criteria for the output (target chemical tool profiles) to provide focus
3. Organize the project around protein families – it is the most efficient, practical and scientifically sound way to divide this large project into teams
4. Establish clear open science principles to eliminate or reduce conflicts of interest, to reduce legal encumbrances, and to encourage participation by the community.

The first five years.

By 2025, the aim will be to have collected and characterized a modulator(s) for at least one protein within each druggable protein family as a key deliverable (many of these modulators would come from existing or completed pharma programs), while in parallel creating the computational and experimental infrastructure to support the project and its expansion to the entire human proteome, as well as developing the technologies to tackle the more challenging proteins. The foundational, and ideally centralized, infrastructure should comprise:

1. Data collection, curation, dissemination, and mining
2. Compound and/or antibody accrual, storage and distribution
3. Characterization capabilities, including screening panels for specificity, toxicity profiling, cell-based “omics” read-outs

The project would welcome the involvement of SMEs, which might be instrumental in creating the infrastructure and in developing assays. The open data from the project should in turn provide opportunities for data-driven companies, such as those developing AI methods.

2025-2035 – completing the task:

Having established the organizational structure, the management team, created the scientific units, established a scalable infrastructure and developed new technologies, the project would be poised to tackle the wider proteome.

Target 2035 Strategy

I. Establish footholds in all target families

- Nominate team leaders for all target families
- Establish international partnerships, as appropriate
 - Bring in appropriate technologies (DNA encoded libraries, peptide technology: phage and mRNA display, new protein production methods, new assay formats, etc)
- Establish open access principles
- Establish quantifiable metrics relevant for each protein family
 - affinity, selectivity etc
- Collect available compounds and antibodies from different sources (e.g. pharma)
- Undertake limited mandatory characterization of all reagents
 - Biochemical/biophysical assay for primary target
 - Cell permeability
 - in silico properties
 - Generic & published selectivity profile (see also part 3)
- Develop plan to increase selectivity panel as more target families will be included
 - Broaden functional read-outs
 - Move to in-cell profiling as the ultimate measure of selectivity, and develop unbiased “omics” quantitative readouts –
 - CETSA
 - BRET
 - Technologies that are developed during course of project

II: Dive deeper into three “pilot” families

- Select three representative protein families to establish SOPs and a framework for scaling up, *for example*:
 - E3's (challenging enzymes)
 - SLC's (membrane proteins)
 - Cytokines (secreted proteins)
- These families should be selected to complement ongoing SGC epigenetics and kinase programs and the NIH-funded programs on GPCRs and Ion Channels
- Aggregate existing compounds and generate small set of new ones
 - Use chemogenomics, or probe strategy, or a new approach - let science dictate?
- Specific outputs (all meeting pre-defined QA/QC criteria) *might* include:
 - Proteins
 - Structures
 - Biochemical and cell-based assays incl. selectivity assays
 - Cell lines with targeted mutations
 - Co-crystal structures with ligands or fragments
 - Chemistry
 - Antibodies
- Characterize all modulators (also from part A)

III: Establish infrastructure for execution phase

- Database – searchable, curated and updated – and FAIR
- Automated compound store, state-of-the art compound logistics
- Communication tools to disseminate reagents and knowledge and to ensure proper use
- Infrastructure for general assays (cytotox, solubility, cell permeability; in vitro & in silico) and assays to be developed
 - This central resource might be structured to “service” the global chemical biology community

IV: Develop strategy to invent technologies to drug unprecedented proteins (and potentially other biomolecules, such as RNA)

- Open challenges like Innocentive
- Grant funding
- Open “hackathons”