

## NUMERATE CASE STUDY: DESIGNING NEW DRUGS FOR A MULTI-TARGETED TYPE 2 DIABETES THERAPY

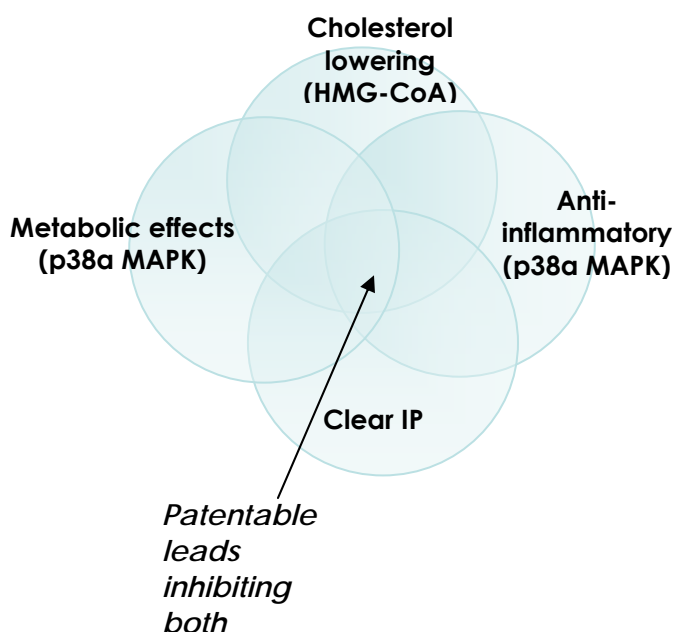
### Introduction

For many years the pharmaceutical and biotechnology industries have sought to reduce the expense of drug discovery by developing computational approaches to aid in the design of new drug candidates. Numerate, a company whose mission is to transform drug discovery into a reliable and predictable engineering process, has applied new developments in machine learning to fulfill the promise of *in silico* design. Numerate's scientists and engineers have developed a new process that delivers the same output as a successful medicinal chemistry effort: proprietary leads that regularly demonstrate a high probability of success.

The following case study demonstrates the effectiveness of Numerate's approach in the only truly relevant way—by designing novel, patentable molecules *in silico* that are then synthesized by chemists and validated *in vitro* and *in vivo* for their biological activity. This case is particularly powerful because the requirement for the newly designed molecules was activity against not one but two distinct drug targets, a challenge that is virtually impossible for traditional methods of drug discovery.

### The Drug Design Problem

There exists an unmet medical need and a market opportunity to develop a single therapy capable of addressing multiple aspects of cardiovascular risk in Type 2 diabetics. Type 2 diabetics are at increased cardiovascular risk because they suffer from hyperglycemia, combined hyperlipidemia and heightened inflammatory status. These derangements provoke damage to blood vessels through inflammation-mediated processes, resulting in atherosclerosis. Combining existing therapies that target different aspects of diabetes has proven sub-optimal from an efficacy standpoint and problematic from a safety perspective.



To create a molecule that would address these multiple medical requirements, Numerate scientists took on the challenge of designing a single molecule that could inhibit both 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA R, the well-studied target of the statins) and p38 $\alpha$  mitogen-activated protein kinase (p38 $\alpha$  MAPK). Signaling through p38 $\alpha$  MAPK controls the release of pro-inflammatory cytokines, which contribute to atherogenesis and elevations in very low density lipoprotein (VLDL) and triglycerides. In the liver, p38 $\alpha$  MAPK-signaling controls the release of C-reactive protein (CRP) and stimulates gluconeogenesis in response to free fatty acids in the blood, thereby providing a link between obesity and diabetes. Because these two targets are not structurally or functionally homologous, it would be extremely difficult to address this drug design problem by traditional techniques. Now this challenge can be efficiently

addressed with Numerate's computational approach.

## Methods

The goal of this case study is to identify a small set of novel and structurally diverse compounds that are highly likely to meet the preclinical candidate specification. Numerate does not attempt to predict precise activities (e.g., IC<sub>50</sub> values) from small amounts of noisy training data that is subject to statistical variation from different experimental conditions. Instead, Numerate ranks compounds by their likelihood of success in the relevant assays, in this case, HMG Co-A and p38 $\alpha$  activity. To start, Numerate collects SAR data for each target, in this example over 100 compounds per target. Numerate represents these compounds based on their shape and electrostatic features, a representation that is scaffold independent and captures all potential interactions responsible for activity. Determining which sets of features (out of tens of millions) are actually relevant to binding (given the noise and bias in the data) is where traditional methods have been inadequate and where Numerate has made significant and enabling breakthroughs. Finally, careful cross-validation provides a reliable estimate of expected laboratory performance.

This precise set of features related to shape and electrostatics is the equivalent of an *in silico* assay. Once Numerate has designed *in silico* assays with accuracy similar to an *in vitro* assay, these assays are used to search through very large (>100B compounds) and diverse virtual chemical spaces. In this case, the space was designed to encode the key structural features and functional groups expected by a medicinal chemist to carry activity against p38 $\alpha$  or HMG CoA. The throughput and speed of Numerate's virtual assays make it possible to look at ~10M molecules per day and are crucial because the more molecules that are analyzed, the higher the probability of success.

## Results

The output of Numerate's design process is a set of potential lead molecules ranked by their likelihood of success. In the world of drug discovery, the key finite resource is chemistry. Because Numerate's approach creates a relatively small set of compounds that has a high probability of fitting the design criteria, the medicinal chemist spends far less time synthesizing new compounds. In the same way that someone performing a Google search looks only at the top ranked hits and doesn't waste time sifting through pages of results, a medicinal chemist using Numerate's approach need only synthesize the few dozen molecules that are highly ranked.

Compound	HMG CoA R <sup>a</sup>	P38 $\alpha$ MAPK <sup>b</sup>	Cellular Inflammation <sup>c</sup>
Lipitor	7 nM	> 100000 nM	93000 nM
Baycol	1.3	> 100000	82000
MTT2DT-1	33	11000	31000
MTT2DT-2	2	36000	30000
MTT2DT-3	3.7	11000	21000
MTT2DT-4	130	1900	18000
MTT2DT-5	11	6000	230000
MTT2DT-6	1.2	400	8000
MTT2DT-7	4.8	270	9700
MTT2DT-8	98	< 100	9700

a Concentration required to inhibit rat liver HMG-CoA R activity by 50%. Data for atorvastatin from Roth et al. J. Med. Chem. 1991, 34, 357. Data for cerivastatin from Bischoff et al. Atherosclerosis 1997, 135, 119.

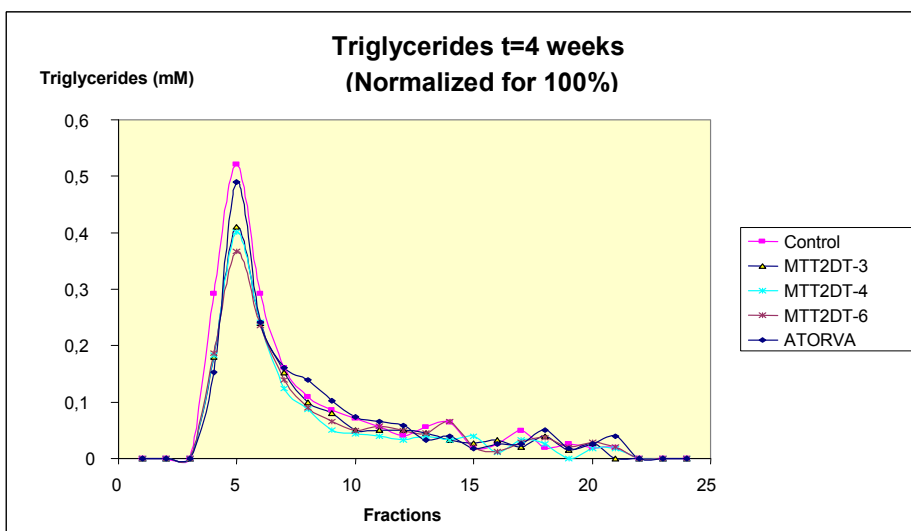
b Concentration required to inhibit recombinant human p38 $\alpha$  MAPK activity by 50%.

c Concentration required to inhibit LPS-stimulated release of TNF from human PBMC by 50%

After the analysis was completed, Numerate's chemists examined the list with the purpose of finding not only the highest ranked compounds but also compounds in several different chemical series. With these criteria in mind, 19 compounds in 4 chemical series were determined to be good candidates. These compounds were synthesized by Sygnature Chemical Services and

then tested for activity by MDS Pharma Services. In the table below, we compare the activity of the new compounds with Lipitor®, the market leader, and a second statin, Baycol, withdrawn from the market because it was associated with rhabdomyolysis. The IC<sub>50</sub> values in the table clearly show nanomolar inhibitory activity of 8 compounds against both HMG CoA R and p38α MAPK, satisfying the design profile and demonstrating the strength of Numerate's engineering approach. The third column in the table indicates that most of the new compounds are 3- to 10-fold more potent than Lipitor in a whole-cell model of inflammation.

.In addition, *in vivo* characterization of the compounds displays good-to-excellent metabolic stability in mouse, dog and human microsomes, and three of the leads that advanced to pharmacodynamic studies in transgenic mice have shown significant reductions in plasma cholesterol concentrations. More importantly, these compounds show significant reductions in plasma triglycerides, an effect that is novel for a statin and consistent with targeting both p38α and HMG CoA R.



## Conclusion

Numerate's scientists have succeeded in designing 8 novel compounds that, when synthesized, have demonstrated activity against two independent drug targets. Not only did they succeed in meeting this difficult challenge, they also accomplished this goal in a short period of time. It took roughly one month to gather the data and one month to perform the computations to determine the best molecules to synthesize. The total time, including synthesis of the high ranked compounds and biological testing, was one year, at a cost of \$950,000. It is clear in this case that it was far less costly and time-consuming to apply Numerate's technology than it would have been to apply traditional discovery methods. Numerate's approach is so efficient that it compares favorably in cost and time to traditional discovery (in-house or outsourced), and even to licensing new compounds. The compounds developed in this case study have issued patents and are now undergoing further characterization and optimization, showing the promise of providing a novel treatment for vascular disease in the large and rapidly growing diabetic population.

## For More Information...

If you'd like more information about partnering with Numerate to advance your therapeutic programs, please call (650) 472-0632, or send an email to [info@numerate.com](mailto:info@numerate.com).