

SYNTHACE ■■

---

## Computer Aided Biology

Delivering biotechnology in the  
21st century

## About the authors

Synthace was founded in 2011. Initially a synthetic biology company, in 2017, that focus shifted fully towards software development. The inspiration for doing so came from the in-house challenges that Synthace scientists faced when trying to undertake complex, multifactorial and time-sensitive experiments with the hardware and software that was available at the time. An in-house solution was designed, which has subsequently been turned into a standalone software product for Computer Aided Biology.

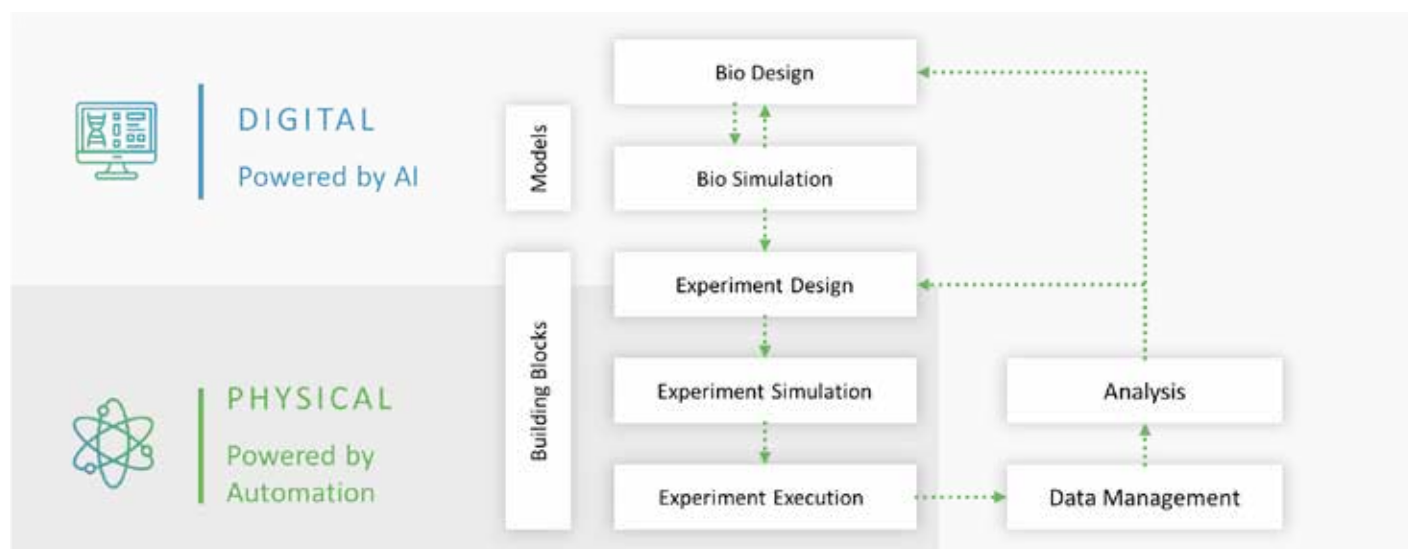
Recognised in 2016 as a World Economic Forum Technology Pioneer, and in 2018 as a Gartner Cool Vendor in Life Sciences, Synthace works with leading biotechnology and pharmaceutical companies across the world, helping them to realise their Lab of the Future/Industry 4.0 ambitions through licensing of its cloud-based automation software.

# Contents

p4	Executive summary
p6	What can biology learn from other industries?
p7	The automotive industry
p8	Moore's Law & EDA
p9	The transition to general purpose solutions
p11	<b>Computer Aided Biology (CAB)</b>
p12	Why CAB? Eroom's Law
p13	Reductionism & complexity
p14	Challenges to CAB adoption
p15	Realising value with CAB
p17	<b>The CAB ecosystem</b>
p18	Bio-Design and Simulation
p19	Experimental Design, Simulation & Execution
p20	Data Management and Analysis
p21	The wider ecosystem
p22	CAB funding trends
p23	Conclusions

# Executive summary

Computer Aided Biology is the emerging ecosystem of tools that augment human capabilities in biological research. It is comprised of two domains: the Digital and the Physical. The Digital, powered by artificial intelligence, includes software for designing and simulating biological systems, as well as methods of collating, connecting, structuring and analysing experimental data from wet lab experiments. The Physical, enabled by automation, includes systems that allow for the seamless transfer of simulated biological designs into real 'wet lab' experiments via abstracted experimental design, logistics simulation and execution. This digital-to-physical workflow is similar to what previously occurred in electronic design automation, the aerospace industry, and the automobile industry. Only by connecting all aspects of the ecosystem together, creating an integrated loop, will scientists be able to address the challenges of 21st century biology.



**Exhibit 1.** The Computer Aided Biology Landscape, connecting the digital word with the physical; powered by AI and Automation.

Even with rapid advances in biological and medical research, little has changed in the way that this research is conducted. New experimental and computer methods simply fit into or promote reductionist ways of working. In this white paper, we show how the emerging Computer Aided Biology ecosystem is evolving, and how many companies are striving towards the same goal of an integrated suite of tools, albeit via different paths.



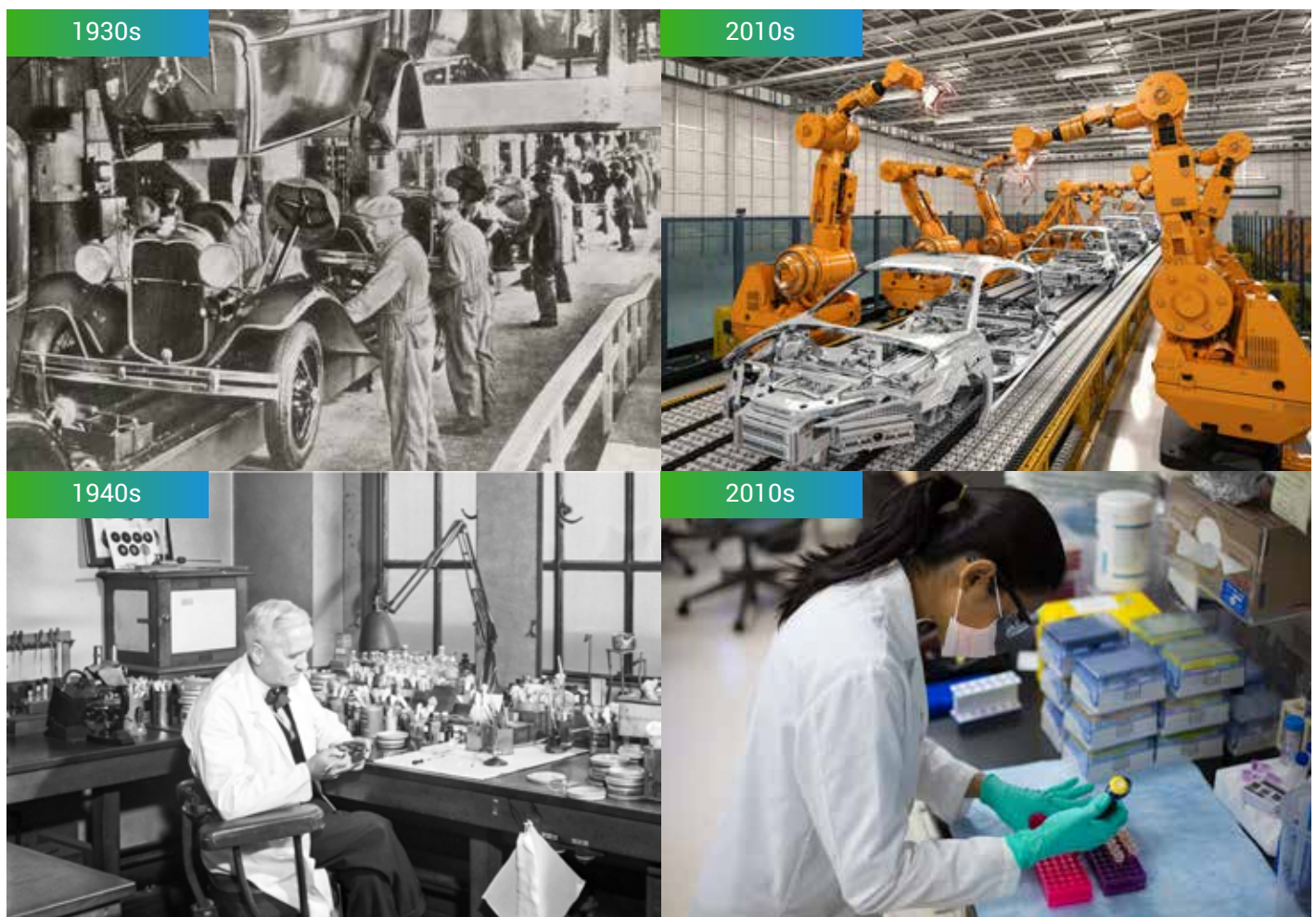


# Biology can learn valuable lessons from other industries

Despite rapid advances in the complexity and power of research methodologies, the majority of biological laboratory research is still conducted in a similar manner as it has always been: manually, and in a reductionist ‘one factor at a time’ way. As such, despite huge advances, biological research still translates poorly from controlled environments into complex dynamic environments. For example, the translation of results from *in-vitro* assays to in-man clinical trials is expensive to conduct and often fails due to unforeseen safety issues or a lack of efficacy.

In comparison, the semiconductor and automotive industries (amongst others), have fully embraced digital technology across the iterative design, test, build and track product lifecycle. They have implemented automation technologies that allow for concise, reproducible and scalable manufacturing from digital blueprints. This has led, in part, to a large reduction in cost and the commoditization of these industries, whilst simultaneously allowing manufacturers to deliver an increased range of increasingly complex products.

By observing how these industries reduced cost whilst increasing product complexity, we gain new insights into how to conduct more resource-efficient biological research.



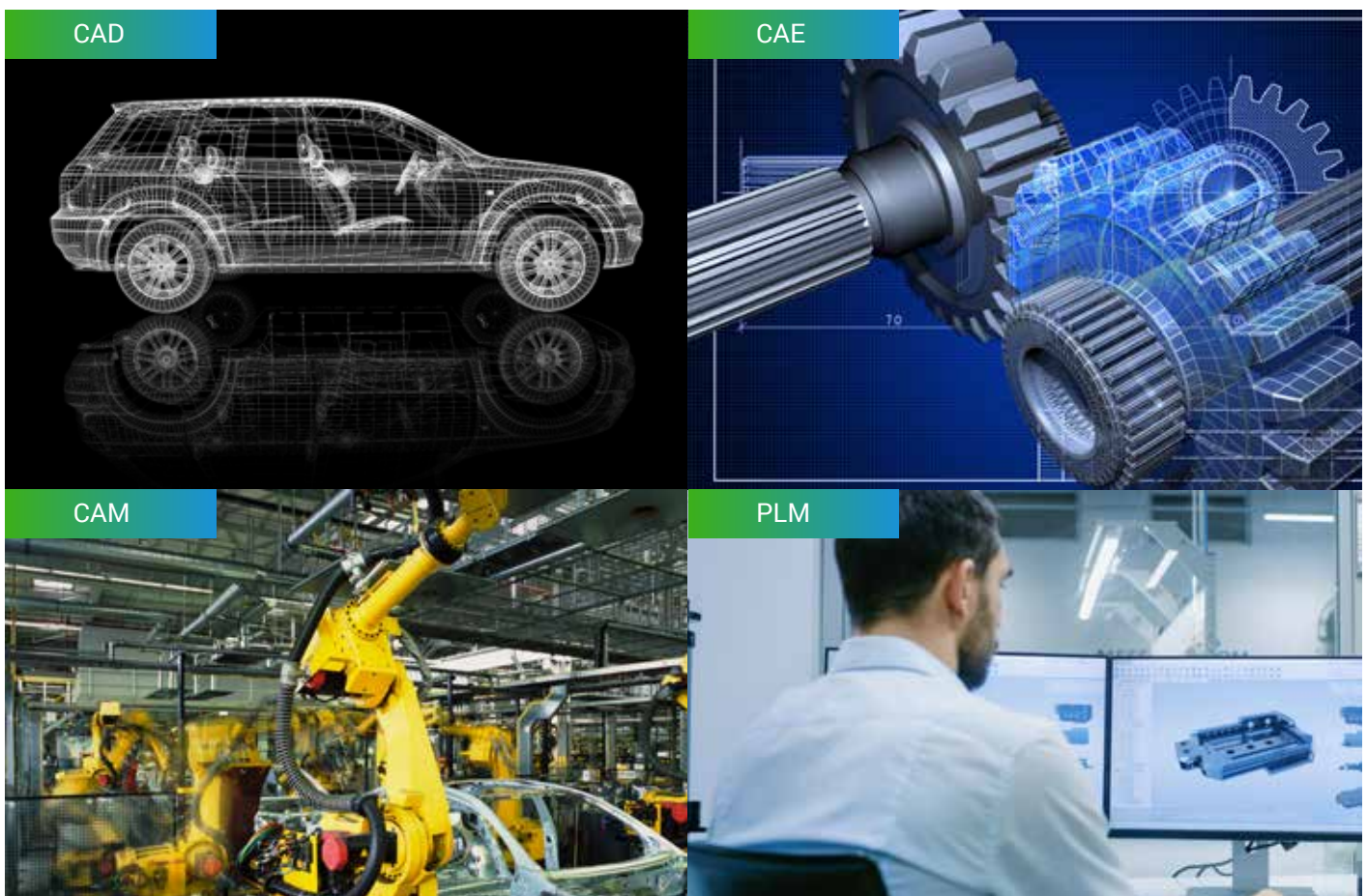
**Exhibit 2.** [Top] The automotive industry, lacking a monopoly amongst firms, is driven predominately by cost. Product differentiation occurs through design and improved features/performance. They have been able to deliver increasingly complex or high-performance products to consumers in part due to their adoption of design and manufacturing processes that lean heavily upon advanced digital tools. [Bottom] In comparison, the London lab of Alexander Fleming (Nobel Prize winner and the discoverer of Penicillin) and a modern biology lab, despite being taken more than 60 years apart, look remarkably similar.

# The automotive industry: reducing cost and time to market with integrated design-test-build software

Beginning in the mid 1960s, the aerospace/defense and then automobile industries were the first to introduce internal computer aided design (CAD) software to assist with the design of planes and cars.<sup>1,2</sup> In 1982, Autodesk, a company now synonymous with CAD/CAM, was founded and the first CAD software for PC was shipped.<sup>3</sup> Since then, an integrated set of digital tools comprising of computer aided design (CAD), 3D simulation/engineering (CAE), physical manufacturing (CAM) and asset/data tracking (Product Lifecycle Management: PLM) have been adopted across the automotive industry. This suite of tools augments the designers, allowing them to collaborate with engineers, to test safety features in-silico and to rapidly iterate designs based on consumer group feedback and/or simulation data.

A digital-to-physical transition in automobile manufacturing also occurs between CAD and CAM. This transition is enabled by CNC (Computer Numerical Code). Software such as Dassault Catia allows engineers to upload a CAD file, simulate its physical manufacturing and then use the same platform to manufacture the component (the software converting the design file into the CNC: a complex set of geometrical instructions that control each piece of manufacturing hardware).<sup>4</sup>

Modern day CAD/CAE/CAM systems often form part of an advanced integrated CIM (Computer Integrated Manufacturing) strategy coordinated via a PLM suite. However, the greatest impact upon the industry can be seen when looking back to the mid-1990s, when integrated tools were more embryonic. In 1988, GM had no dedicated CAD/CAM/CAE workstations. By 1995, it had 6500+ across the organization.<sup>5</sup> By 1998, Ford was already reporting >40% efficiency savings in engineering and a greatly shortened time to market (<24 months from concept to market) through use of CAD/CAE/CAM and PLM software.<sup>6</sup>

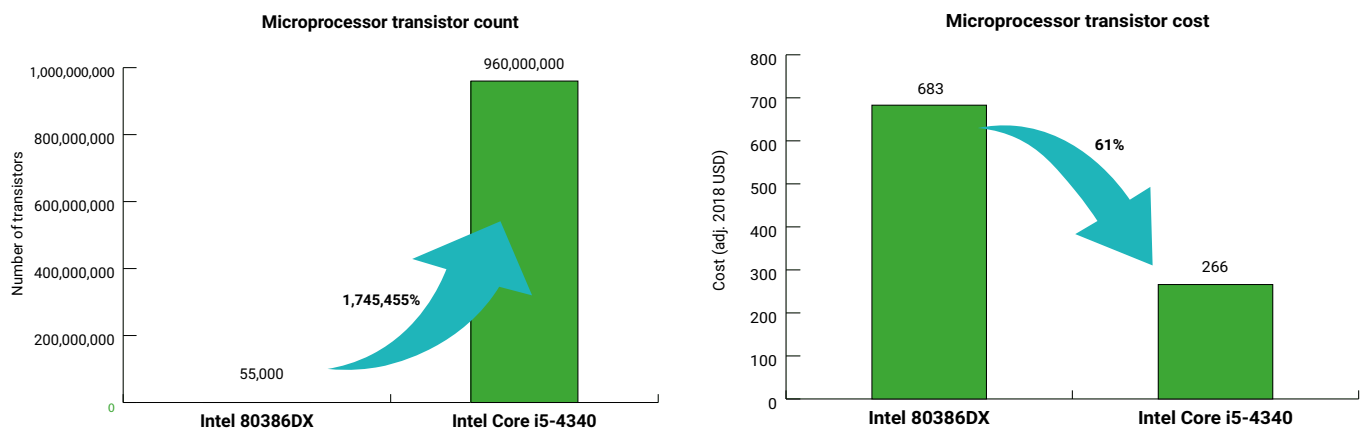
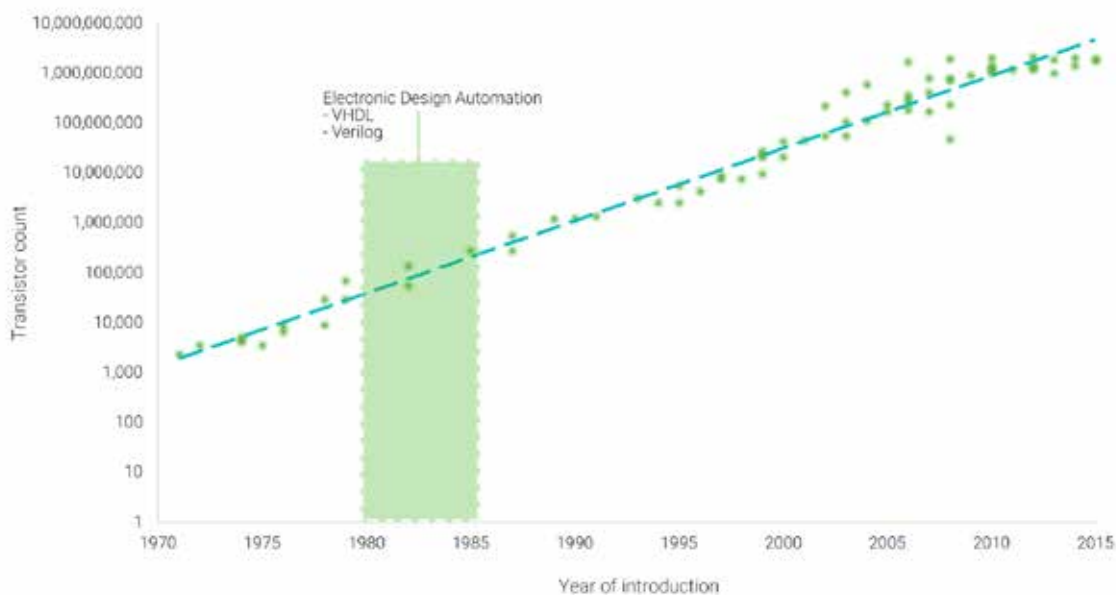


**Exhibit 3.** The stages of the automotive design (CAD), simulate (CAE), build (CAM) and track/integrate (PLM) life-cycle.

# Electronic design automation (EDA): Enabling Moore's law through abstraction

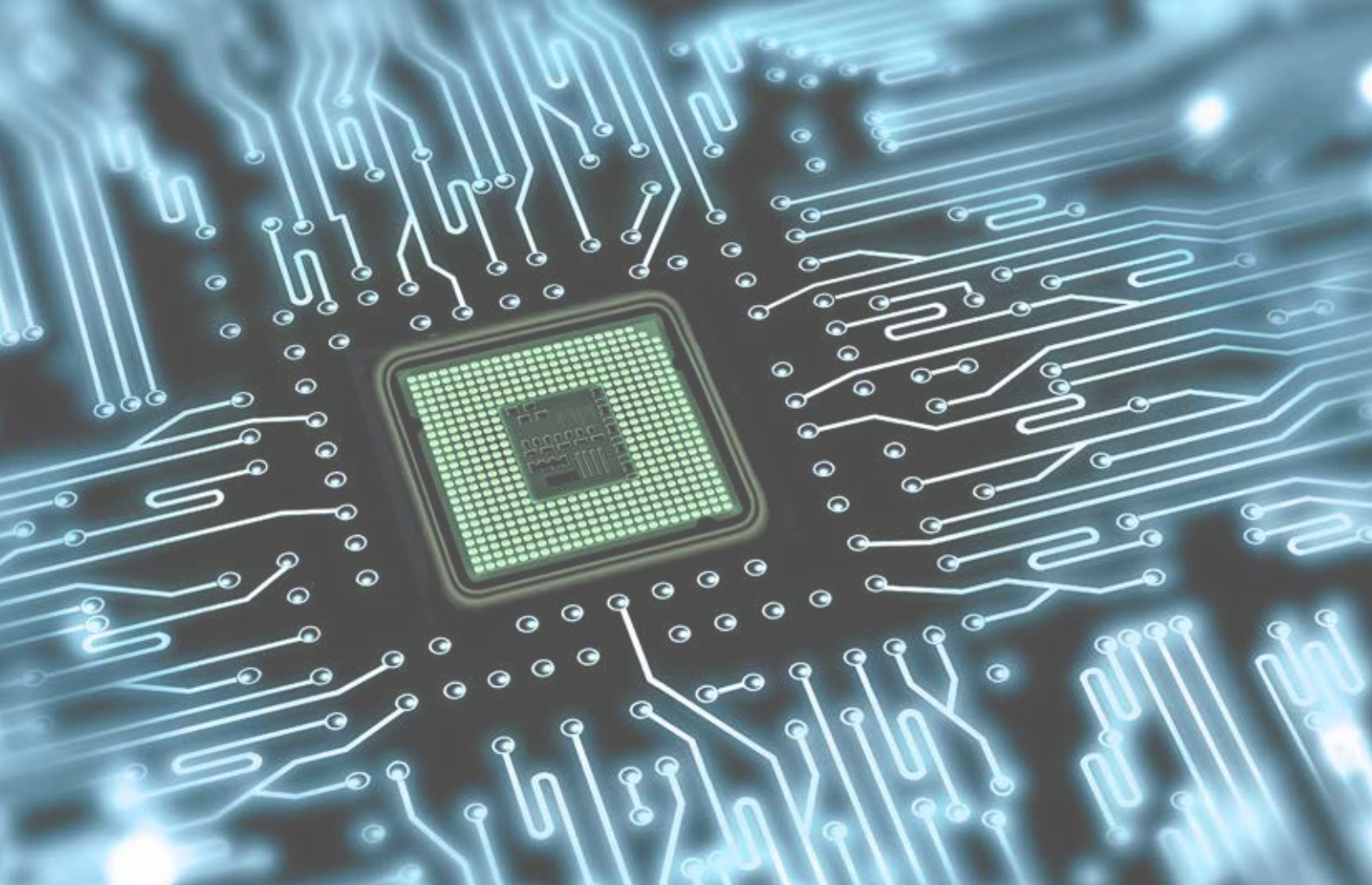
As the complexity of integrated circuits rapidly increased through the 1970s and 1980s, tools that relied upon artisanal and a purely graphical construction of integrated circuits were no longer fit for purpose. Instead, an ecosystem of software tools were introduced which allowed for the creation of integrated circuit designs in high level and abstracted programming languages (VHDL and Verilog), referred to as Very Large-Scale Integration (VLSI).<sup>7-9</sup> Abstraction, or the removal of extraneous detail, often at multiple levels (e.g. logic design, circuit design and layout), greatly increased the complexity of integrated circuits that could be designed. Furthermore, as integrated circuit manufacturing is very expensive, these tools included simulation/verification tools at multiple layers of abstraction that allowed users to verify the physical reliability of their designs prior to manufacturing. Once a design was finalized, it could be digitally transferred to a physical fabrication system and produced.

These tools, along with advances in low-cost and high-performance fabrication, allowed for the birth of computing revolution, and the observation of Moore's Law.<sup>10</sup> During this period, the increase in chip power/complexity, and the simultaneous reduction in cost/transistor, enabled powerful microprocessors to be installed in personal computers, smartphones, and other household systems; democratizing high-performance computing power for the wider population.



**Exhibit 4.** [Top] Moore's Law. [Bottom] In the 19 year gap between 1986 (Intel 80386DX) and 2015 (Intel Core i5-4340), based off the number of transistors in the microprocessor an approximate 1,745,455% increase in complexity was achieved. At the same time, a 61% reduction in cost to the consumer was also achieved.<sup>11</sup> During this period, integrated circuit R&D/production costs may have risen, but importantly for the increase in R&D/production cost, the companies were realising a >1X multiple return in integrated circuit complexity; leading to mass market products and economies of scale.



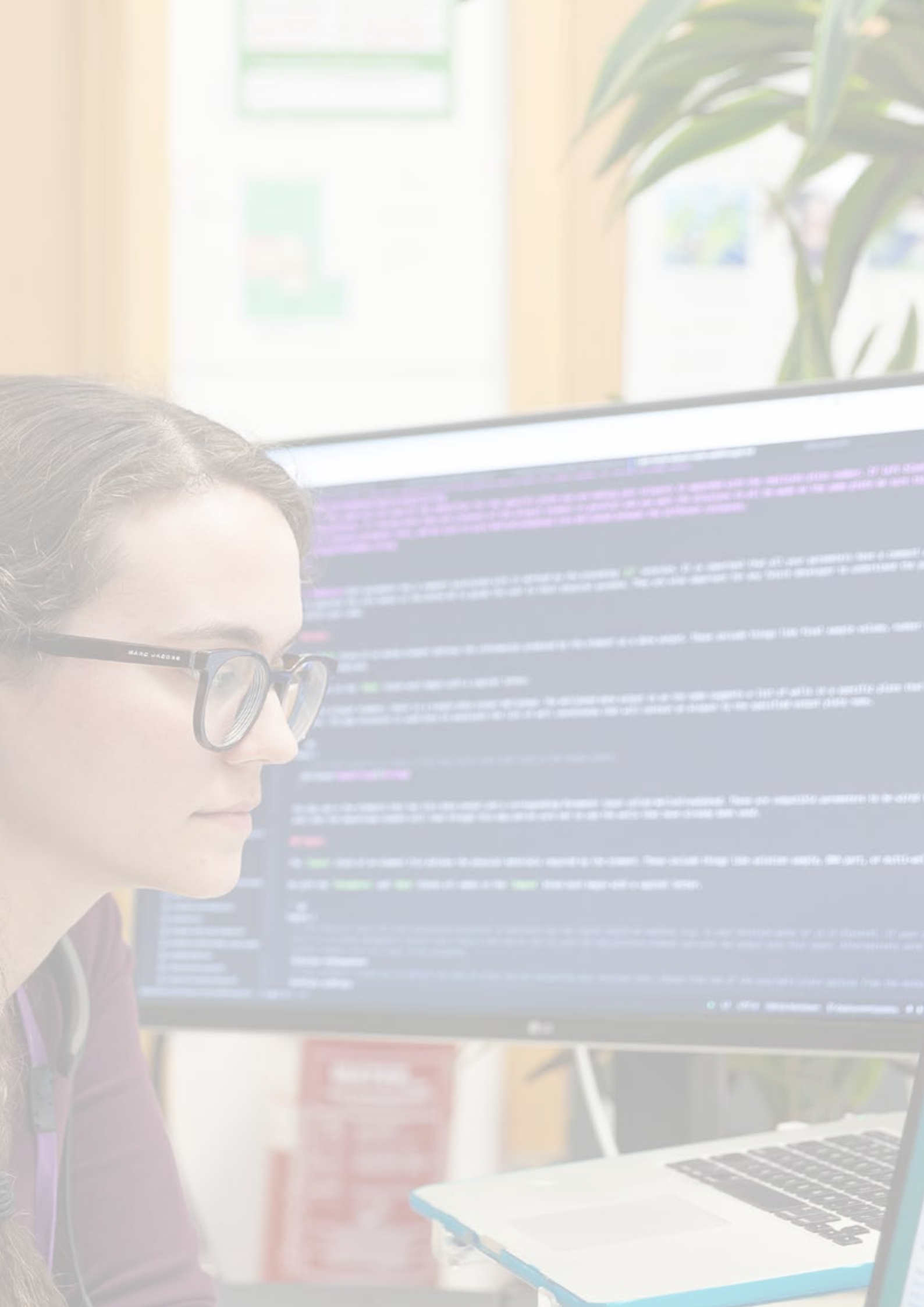


## The inevitable transition to general purpose solutions

Given the tremendous impact on productivity that integrated digital tools provided, they rapidly became the dominant working model for mechanical and electronic engineering. Today, no company operates without a skilled workforce that is proficient in modern digital design tools, as the productivity gap makes any other way of working untenable. Given the tremendous competitive advantage these tools provided, it also sparked heavy investment in the in-house development of these tools by companies, as well as adoption by and partnership with general purpose vendors.

In both the automotive and aerospace industries, in the early days of CAD/CAM development and deployment, it was the proprietary in-house technologies that were internally developed and maintained (such as PDGS at Ford or TIGER 3D at Boeing) that represented the majority of the market. However, in both industries, as dedicated general purpose vendors emerged, it drove a gradual transition away from in-house solutions to the use of general purpose platforms. The majority of in-house solutions were spun out or replaced by the late 1980s. Most of the spun out solutions were eventually consolidated by general purpose vendor solutions, including arguably the most successful CAD/CAM solution, Dassault Systems CATIA, which originated as the in house CADAM technology from Lockheed.

Similarly, in the semiconductor industry, the early in-house solutions developed by AT&T Bell Labs, Intel and IBM were rapidly spun out and replaced with emerging pure-breed general purpose tools. The original cohort of dedicated vendor start-ups (e.g. Mentor Graphics, Cadence and Synopsys) became the foundation for the emerging industry that rapidly supplanted the in-house tools and spinouts of individual companies. The EDA industry has historically operated more as a pipeline than an integrated solution. This has enabled a steady flow of new technologies, each solving a portion of the design and manufacturing problem space, to rapidly become the best in breed solution for that portion of the overall solution. Well-defined industry standards have helped build interoperability across these best of breed technologies.

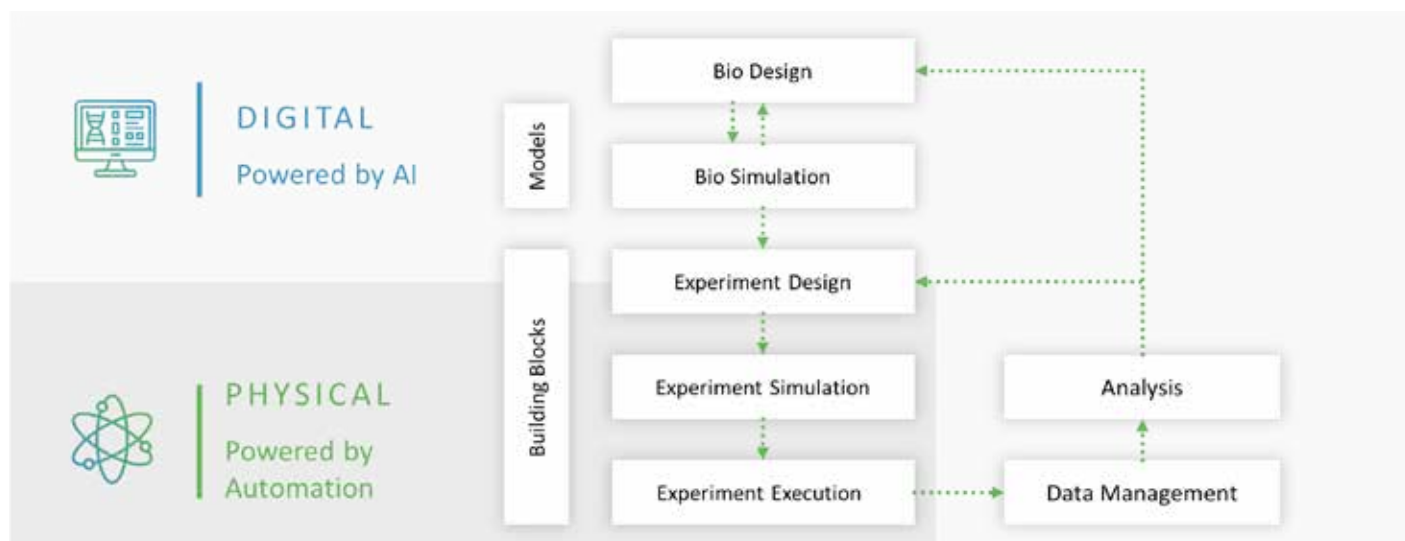


# Introducing Computer Aided Biology

Drawing on inspiration from EDA and the automotive industries, we introduce Computer Aided Biology (CAB). CAB is a conceptual framework and an emerging ecosystem of digital-to-physical research tools that augment human capabilities in biological research, allowing them to approach experiments of increasing complexity. This doesn't mean 'slowing' the research process. Instead, it means using digital tools to allow researchers to rigorously explore the dynamic biological space in a higher throughput, non-reductionist, and iterative manner. Computer Aided Biology is built around the logic that, given the tremendous complexity of biology, wet-lab validation will be required for the foreseeable future and is thus comprised of two domains: the Digital and the Physical.

The Digital, powered by artificial intelligence, includes software for the design and simulation of biological systems, as well as methods of collating, connecting, and analysing experimental data from wet lab experiments (e.g. CAD, CAE and PLM). The Physical, enabled by automation, includes systems that allow for the seamless transfer of biological designs into real 'wet lab' experiments via a protocol design, physical logistics simulation, and finally execution (e.g. CAM). The integrated physical execution allows for the creation of tagged, connected and structured datasets, ideal for advanced machine learning methods.

Due to the challenges of designing or engineering complex biological systems, as well as the concurrent challenges of programming complex experimental methods, leading products in the ecosystem leverage the synergistic benefits of abstraction with simulation of both biology and logistics to deliver error free bio-designs/experimental execution, without a high technical barrier to entry. This digital-to-physical workflow is akin to what previously occurred in electronic design automation, as well as within automotive design and manufacturing. In isolation, these tools create value, but value creation is significantly enhanced by connecting all aspects of the ecosystem together, digital and physical, to create an integrated and iterative loop of design, simulation, experimentation, and analysis.

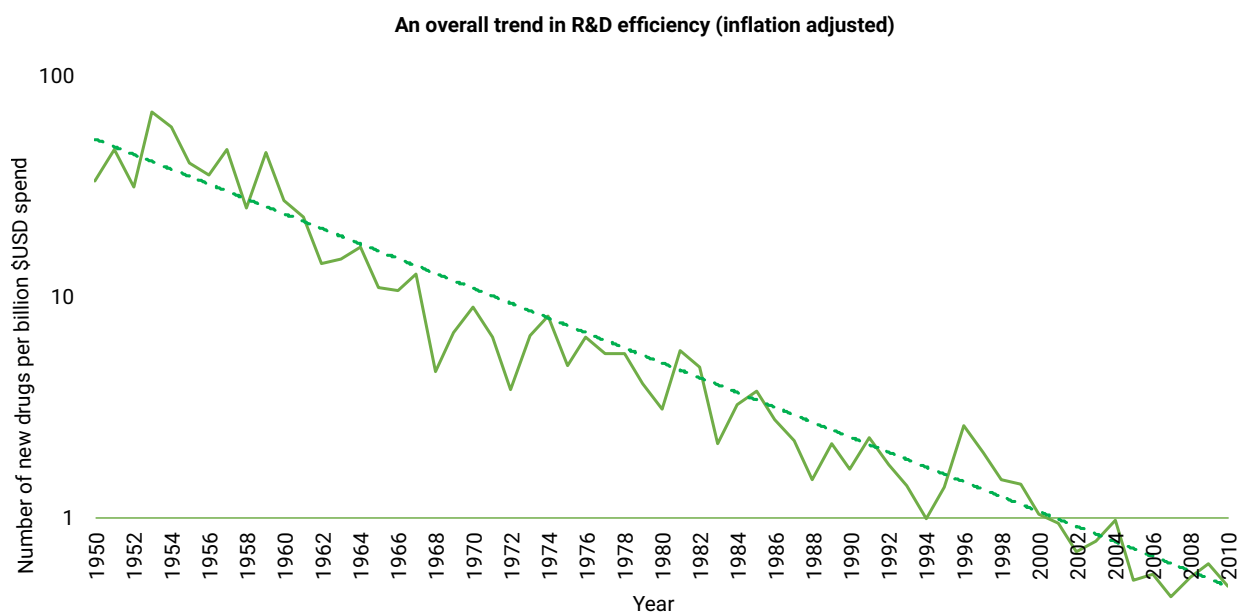


**Exhibit 5.** The Computer Aided Biology landscape, connecting the digital word with the physical; powered by AI and automation.



# Why Computer Aided Biology? Eroom's law and the biopharmaceutical industry.

Eroom's law (Moore's law backwards) is a term introduced in 2014 to describe a decades-long slide in pharmaceutical productivity.<sup>12</sup> Subsequent studies using different metrics have observed the same trend, with a 2017 study indicating that by 2020, the industry wide IRR (internal rate of return) on new R&D will hit 0%, an already unsustainable position moving forwards.<sup>13</sup>



**Exhibit 6.** Eroom's law. A schematic representation of a figure presented by Scannell *et al* in 2014.

Industry experts ascribe the blame to a variety of factors discussed in detail by Scanwell and Stott. Whilst all of the factors certainly contribute to the observed reduction in productivity, the research-brute force bias is the only one intrinsic to the conduct of the lab research function of the industry. All others are a factor of market risk (the availability of profitable indications with little on-market competition) or managerial/regulatory factors (throw money at it/cautious regulator/project prioritisation).<sup>14</sup>

Looking to reverse the slide in productivity, biopharma has looked to new biological methods and technologies to help it accelerate its R&D. However, the big breakthroughs in productivity haven't been observed, and the downward trend has continued. In hindsight, it is not surprising that these past technologies haven't delivered, for each time they simply reaffirmed the existing ways of conducting biological research: wasteful, and with an inherent reductionist 'brute-force research bias'. Unfortunately, this is also being played out as we speak with a large focus from technology investors and big pharma on 'AI for Drug Discovery', the majority of which seeks to create in-silico drug candidates with better pharmacological properties. This approach, targeting the earliest stages of the therapeutic R&D process, again fits into the 'brute-force' approach, and whilst useful, does nothing to change how biological research is conducted. This is not Computer Aided Biology. Fundamentally, it requires closing and automating the entire loop, not just point solutions.



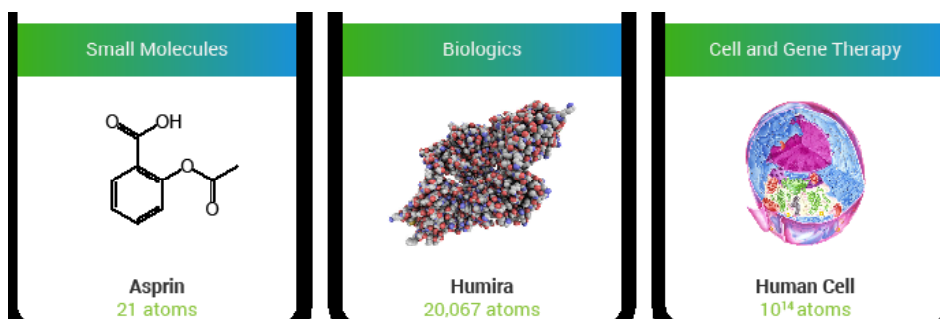
# Moving away from reductionism: Embracing complexity through Computer Aided Biology

Instead of adopting tools that reinforce existing ‘brute force’ approaches to drug discovery, the industry should instead transition to ways of working that focus around iteratively understanding the dynamic and multifactorial ‘biological parameter landscapes’ in which they discover and develop therapeutics. Some call this approach ‘Active Learning’. This way of working uses advances in machine learning, not just to screen/discover new therapeutics, but also to draw insights from the vast inflows of data from across the presently siloed experimental, preclinical, and clinical sources. It also leverages abstraction to allow lab researchers to quickly design complex biological networks, or experiments, which can then be simulated prior to execution.<sup>15</sup>

One of the central tools that will help Biology move from ‘brute force’ to iterative understanding will be ‘Experimental Design’ or ‘Design of Experiments’. Changing multiple parameters at once to understand the dynamics and emergent properties of biological landscapes is an approach more akin to Experimental Systems Biology (an approach that seeks to understand biological systems as multifactorial networks of interacting parameters), rather than the traditional view of molecular biology that uses a reductionist ‘one factor at a time’ (OFAT) approach to interrogate biology. Thus, whilst our theoretical frameworks now often treat biology as a system, our experimental methods have lagged, necessitating this new Computer Aided Biology framework.

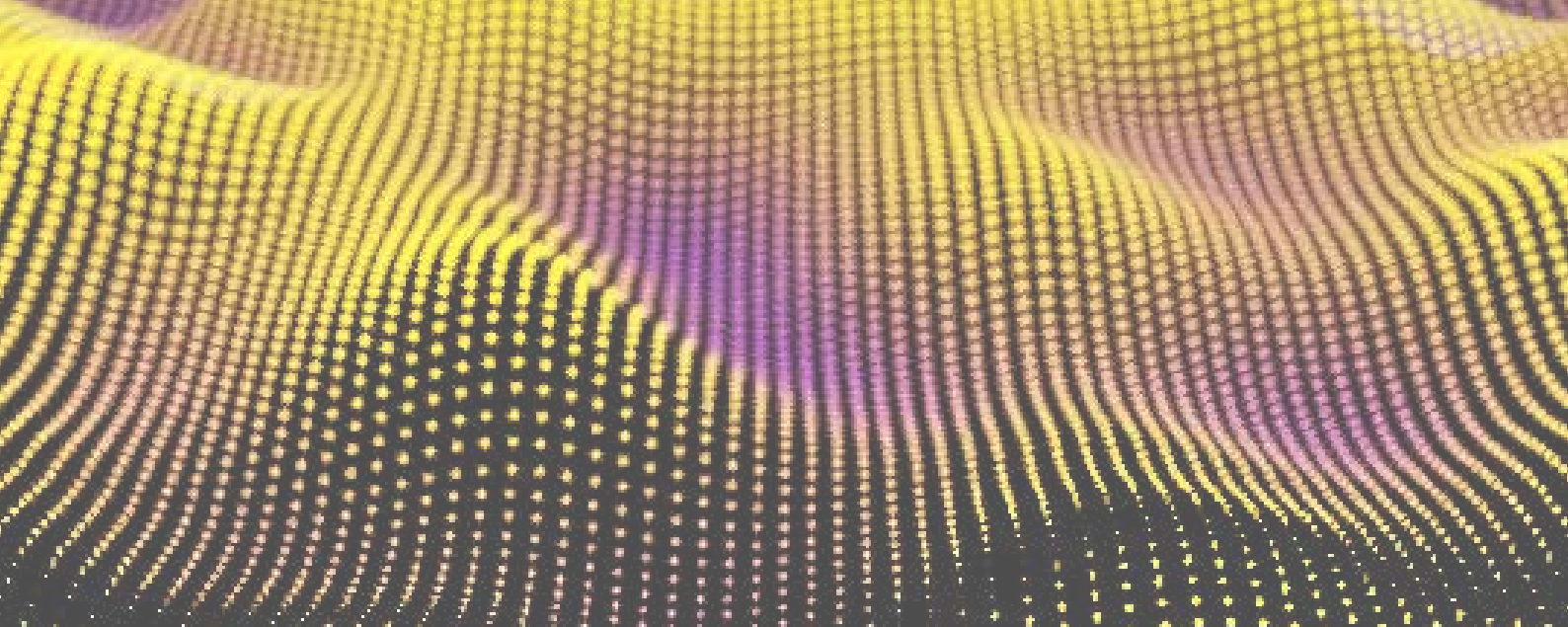
Without the creation of tools across the CAB ecosystem, this approach will not succeed because its success is reliant upon fast, iterative build test cycles, connected and structured data sets, and complexity reducing abstractions. It should be stated that this new, iterative way of conducting drug discovery is not actually new and is similar to the old ways of iteratively doing drug discovery. The difference here is that by using new tools and technologies, we no longer need to sacrifice speed to adopt this approach.<sup>16</sup>

The need to adopt these ways of working is also growing, with biopharmaceutical companies now moving into complex curative advanced living therapies, including CAR-T cell therapies. These therapeutics are often produced on a per-patient basis and involve manufacturing a clinically valid product from variable donor material. These factors introduce a prohibitive financial cost to the patient/healthcare system. This necessitates a new approach, one with a focus on the complete and early understanding of the ‘biological parameter landscape’, e.g. to enable the use of leading indicators for product end-point quality. The development and scaled-out production of these new therapies is enabled by the rapid adoption of abstracted genetic design software, inline analytics, lab automation, and the application of data integration/machine learning.



**Exhibit 7.** The increase in complexity, from small molecules through to patient specific cell therapies.

The semiconductor and automobile industries do not suffer from this same brute-force approach to R&D. Instead, they utilize a suite of digital and physical tools to navigate their respective product’s landscapes, optimizing for acceptable surrogates for success such as performance, user experience, or safety.



## Challenges to adopting Computer Aided Biology

The transition to Computer Aided Biology (CAB) will not happen overnight. That is because the framework doesn't advocate treating the symptom, but instead advocates fixing the cause: the way we work with biology. Despite the benefits that this new way of doing biology will deliver, there remains challenges to its adoption:

**Human capital and culture.** As the adoption of new ways of working emerge, so will a skills gap within the biological research workforce. Current skills (such as proficiency with manual experimental techniques) will be superseded by an increased demand for data science and coding skills, along with a renewed focus on a fundamental understanding of biology. This will have a transformative affect upon the productivity of the industry but may lead to resentment from late adopters. Scientists, by using the 'augmentation' tools of the CAB framework, will be at the center of this transformation. However, in the future this may lead to a bimodality in the scientific workforce, with it comprised of the 'creative thinkers' (who are proven scientific experts with an ability for drug discovery) and the 'technicians' (who are responsible for QC/QA/support for the swathes of new automated equipment). The middle-tier of worker may be replaced by data scientists. Overall though, this new landscape will free up research scientists to focus on the 'why' instead of the 'how'. Importantly, CAB still places the scientists at the heart of this future vision, with their innate capabilities augmented by digital tools. In addition, many universities are now placing a stronger focus on digital skills as part of postgraduate education, and many tools are developing highly abstracted user interfaces so that even technology-adverse scientists can begin to participate in this new way of doing biology.

**A need for a common language.** In this integrated future vision of the ecosystem, value will lie both in the isolated ROIs (return on investments) of each of the tools and in the interconnectivity between them. To realize an aggregated ecosystem wide ROI, the emergence of better standards for design files, analytical data, and environmental data is needed so that the individual tools can communicate via a common language of collaboration. In addition, these tools must also speak a common language with their users: the language of biology.

**Timelines and hype.** To date, the CAB ecosystem hasn't suffered the same 'hype' as AI for drug discovery, which is a positive factor. There is no magic bullet for changing the way biology is conducted. Instead, it is about building slowly by winning hearts and minds and delivering a strong ROI at each stage of growth. This makes tracking ROI a key component of success, as results will be measured over longer periods.

# Realising value with Computer Aided Biology

Our Computer Aided Biology framework links into a wider industry movement towards digital transformation. By focusing on changing the way biological research is conducted, rather than a point solution, CAB will be as useful in early stage discovery as in manufacturing (albeit with varying weightings on differing parts of the framework).

Within biopharma, data management platforms in the CAB ecosystem may be expected to increasingly integrate with tools that aggregate data from genomics initiatives, clinical trials, public data repositories and real-world data. This allows for an organisational wide, 360 degree approach to patient stratification, drug R&D, manufacturing and post approval monitoring. In this future vision of increasingly precision therapeutics, CAB will be an essential component that allows industry to rapidly and confidently iterate to complex precision products, faster and cheaper than is currently possible.

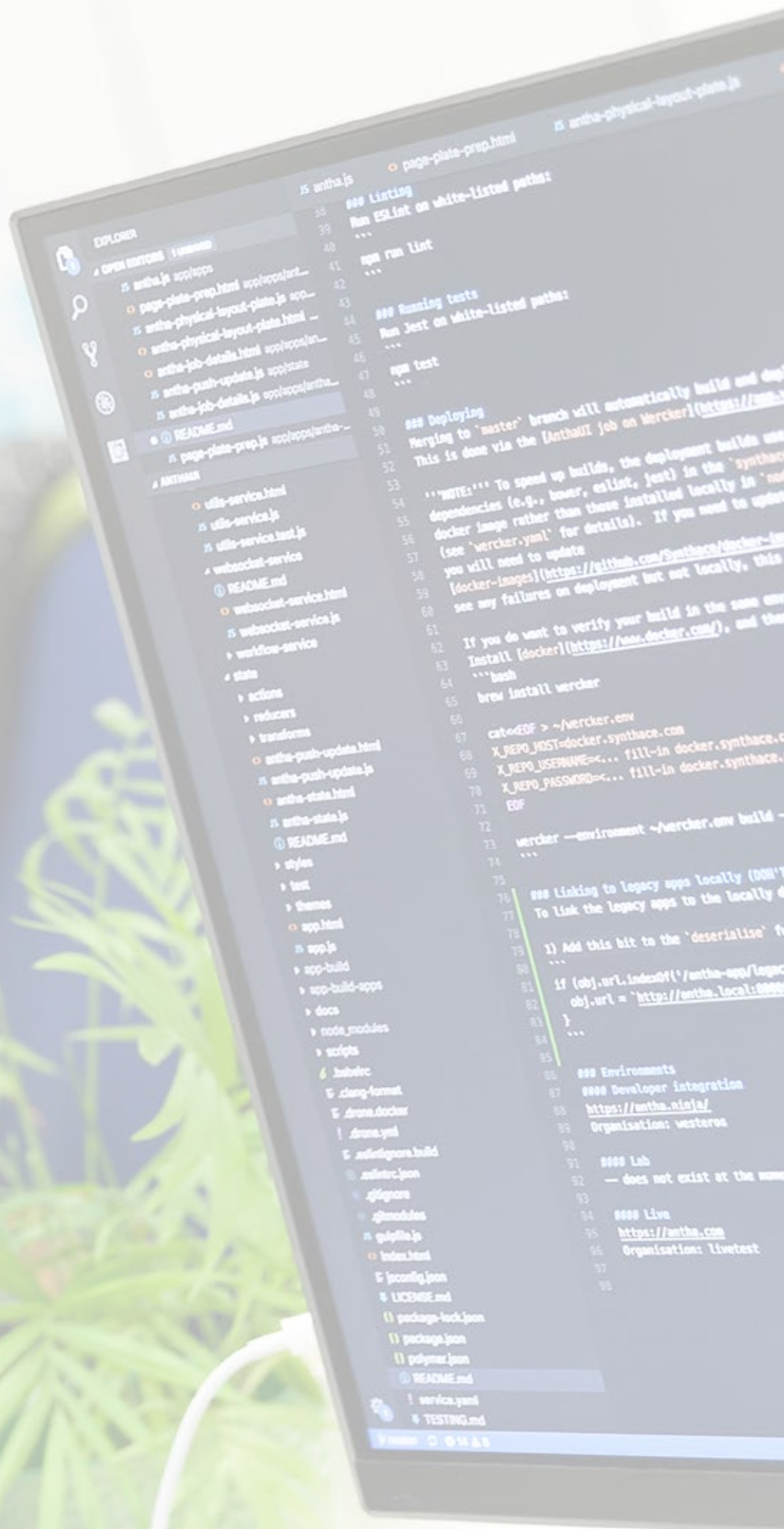
For decision makers within biopharmaceutical companies, this vision may sound exciting, but issues often arise in that this end-state requires a long-term commitment, organisational wide buy-in, and significant long-term investments of time and resource. So how can CAB provide immediate benefits to the researchers? In our experience, working with leading companies across biologics, advanced therapies and agrochemicals, leveraging a CAB strategy realises tangible benefits along three thematic trends:

**Cost reduction.** Cost reduction can take many forms, including reducing the costs of the raw materials used (e.g. using them more efficiently), a reduction in FTE (full time employee) cost to achieve the same levels of understanding, and an increase in reproducibility. Sample provenance is also increased as samples can be tracked, in some cases from 'needle to needle,' and linked seamlessly to associated design and data files across the organization. At a higher level, the aggregation of multiple data streams, in addition to an increased volume of high-dimensional experimental data, should lead to a deeper understanding of why drugs fail. This will allow for the rational design of better drugs and better clinical trials.

**Improved product quality and quantity.** High-dimensional experiments, often too burdensome or out-of-scope to conduct early in a discovery/development program, can become more widely used due to high throughput integrated digital design and physical execution. This, along with integrated analytics, can lead to a more robust exploration of therapeutic/biological design space and thus, new adaptive ways of producing biological materials. The upshot of this iterative optimisation approach is an improvement in product quality (reduction of impurities, better thermal properties, etc.) and a higher yield.

**Time saving.** Experimental automation saves time through increased walk-away time and 24/7/365 operation. Linked with data integration and design tools, it enables the scientists to avoid spending time structuring data or programming complex designs, which when coupled with simulations of biological designs and experimental systems, reduces the likelihood of expensive errors occurring. Timesaving also includes aspects such as an accelerated technology transfer between process development and manufacturing, as well as via the cloud sharing of optimized digitally defined method files both internally (e.g., across teams) and with partners (e.g., a CDMO).

As an example of this, when Synthace worked with a leading pharmaceutical company, "[Synthace's software] increased throughput by an order of magnitude and enabled an estimated 25% reduction in overall vector construction timelines, a 33% reduction in costs, while dramatically reducing manual steps and increasing robustness."



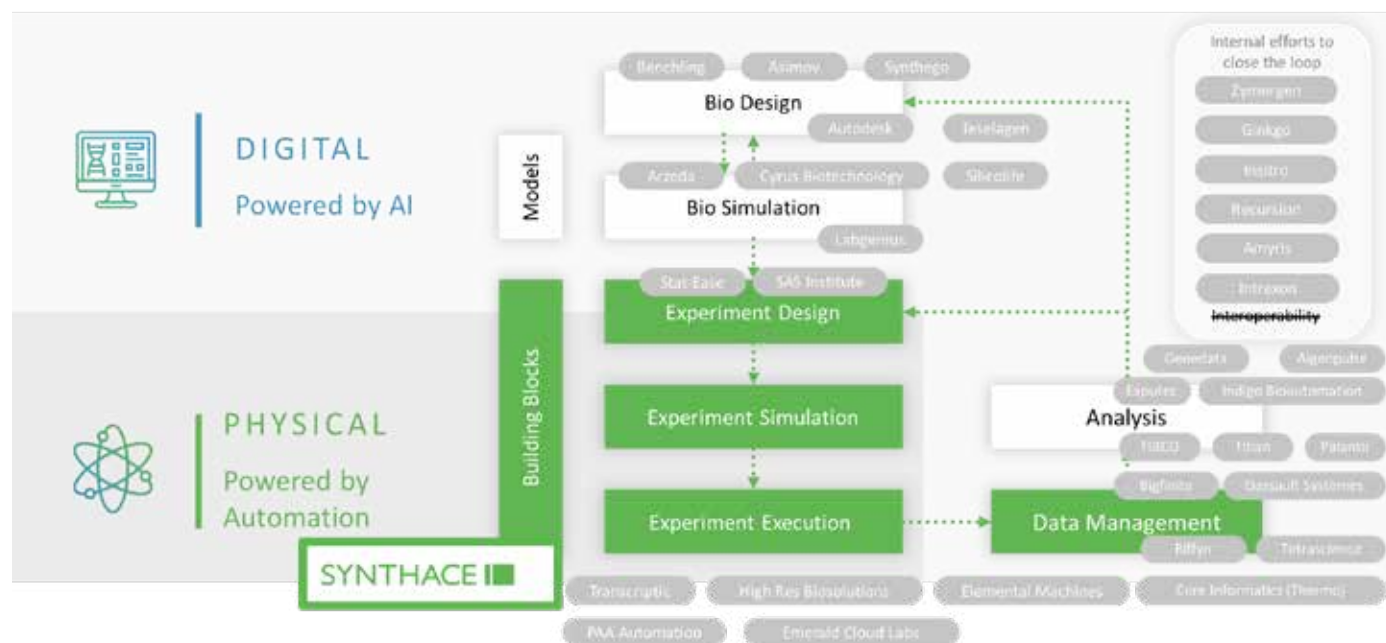


# Navigating the Computer Aided Biology ecosystem

The term Computer Aided Biology, first introduced by Synthace in Q2 2018, is not the first attempt at describing this ecosystem. Others have recently introduced terms such as the 'Synthetic Biology Stack' (Q3 2018) or 'Digital Biology' to describe a selection of companies including reagent providers. Each of these approaches fails to capture that it is a fundamentally new way of conducting biology that is needed, and that creating tools that simply fit into existing inefficient frameworks isn't the solution to the long-term challenges of biological research. The term Engineering Biology has also been used frequently, with it being the approach taken when adopting the tools of the Computer Aided Biology ecosystem.<sup>17</sup> This is akin to Electronics Engineering being empowered by Electronic Design Automation as it transitioned to Computer Engineering, or Automotive Engineering being linked to the CAD/CAE/CAM tools used in vehicle production.

There are a range of companies in the Computer Aided Biology ecosystem: equipment manufacturers creating IoT integrations for their hardware, data management platforms, and biodesign packages. Some of these packages are well established (e.g. GeneData), whilst others are very nascent.

A selection of companies in the Computer Aided Biology ecosystem are shown below. This abbreviated ecosystem map indicates that there is a congestion of companies in the Bio-Design/Simulation space, as well as in Data Management/Analysis. In comparison, there are relatively few companies working on the digital-to-physical transition; an area that requires the integration of abstracted experimental design with simulation, deep equipment integration and cross-platform method management and execution abilities.



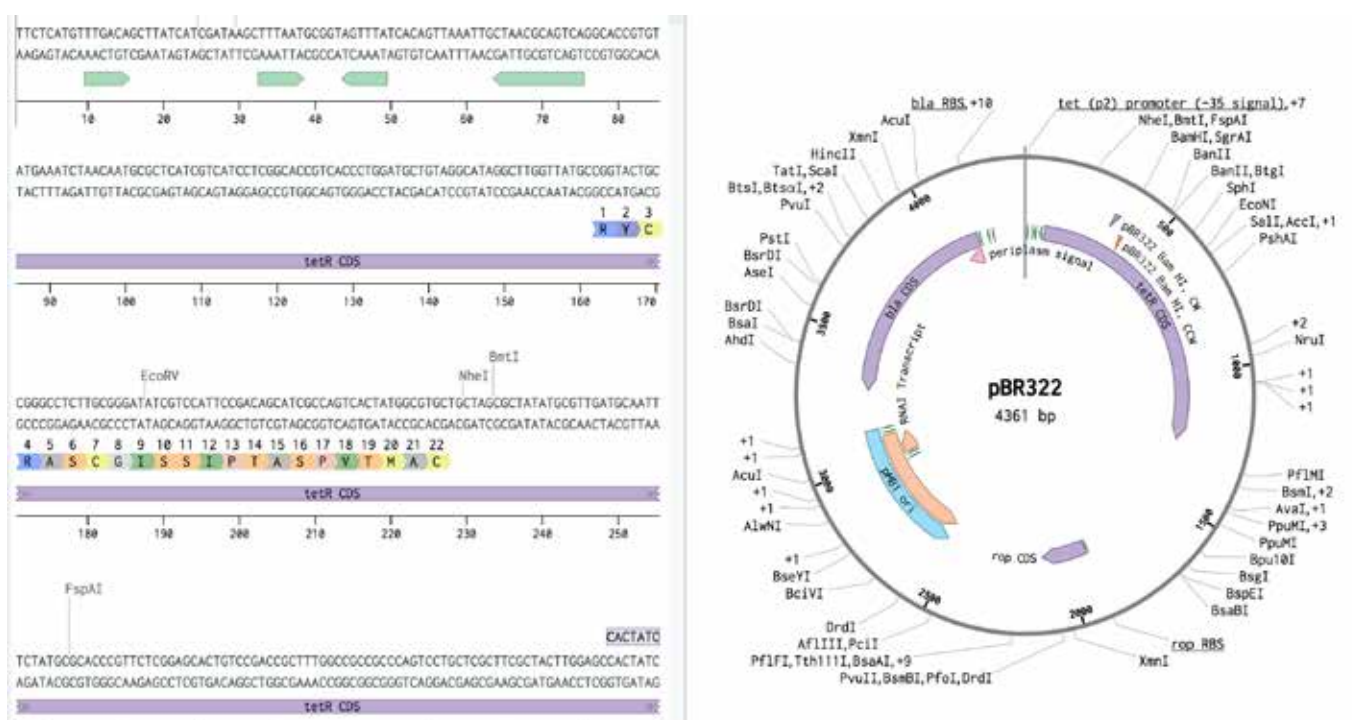
**Figure 8.** An abbreviated Computer Aided Biology landscape. A more complete view of the landscape also includes hardware manufacturers, who build the tools upon which the wider ecosystem rests.

What follows is a more detailed overview of each of the thematic sections of this ecosystem, as well as a discussion of the role of the wider ecosystem stakeholders who fund, support or manufacture equipment.

# Bio-Design & Simulation (Bio-CAD/CAE)

Bio-design (or Bio-CAD) tools often use visual abstraction to enable researchers to compose complex genetic constructs in-silico. Similar to EDA, these tools often contain in-built error checking (biological simulation) to avoid costly failures at the genetic construct synthesis stage. Other companies are developing software packages that allow for the design and simulation of de novo proteins, as well as other important biological macromolecules. Some companies in the design space also link seamlessly with ELNs, LIMS, or entity registration packages. This allows researchers to design biological systems with components that are already in their inventories.

Many of these tools are created by companies whose central business is the sale of reagents (CRISPR kits, Oligonucleotides, Enzymes etc.). These tools are used to provide either a competitive product advantage (e.g. a better enzyme), or as sales enablement to facilitate the customer in purchasing the correct biological construct. More companies, realizing the value of 'Design' software, are now beginning to focus on selling access to their software.



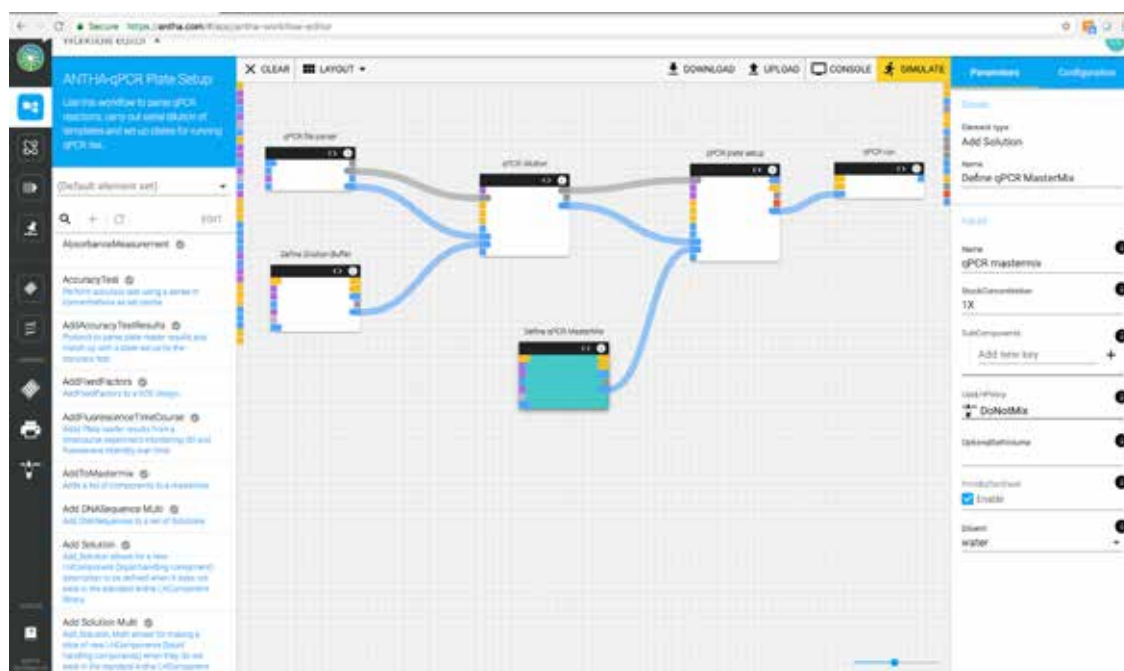
**Exhibit 9.** An example from the Bio-CAD section of the ecosystem, Benchling is a leading ELN/LIMS platform with integrated Bio-CAD tools for CRISPR.

Examples of companies in this space include **Benchling** (ELN/LIMS and design), **Synthego** (gRNA design and verification), **Desktop Genetics** (CRISPR gRNA design software and reagents), **Autodesk** (Genetic constructor and simulator), **Asimov** (in-cell logic design), **Pepticom** (peptide design), **Lattice Bioautomation** (bio-design), **Teselagen** (DNA design), **Arzeda** (computational enzyme design), **Lab Genius** (AI driven protein evolution), **Cyrus Bio** (protein design and simulation), **Codexis** (protein engineering), **Enzbond** (computational enzyme design) and **Silico Life** (computational metabolic engineering).

# Experimental Design, Simulation & Execution (CAM)

In-silico simulated digital blueprints need to be transformed into a physical reality, tested, and the data fed back into the system so that they can be iteratively improved. This is where experimental design, abstraction, simulation and execution come to bear.

Whilst end-users of automation hardware typically focus on increasing throughput, this is an important metric within the existing 'brute force' paradigm. Far more important is that, in the near future, increased adoption of flexible automation equipment will allow scientists to conduct far more complicated experiments more rapidly, with the data generated being created in a more standardized, clean and structured format. This is a crucial component to opening up biology to the power of advanced machine learning. In this respect, removing the experimental error of manual intervention and providing traceability has far more power than simply just increasing throughput.



**Exhibit 10.** The workflow editor of the Synthace product, Antha, allows users to rapidly create complicated processes that integrate physical operations and code. The individual blocks, or elements, represent physical or digital protocols. Ultimately, this creates a fully digital representation of the entire biological workflow. This is an advanced example of abstraction, as the system automatically error checks/simulates the process to ensure that complicated experiments run as desired.

Flexibility is also an important part of this aspect of the ecosystem. Whilst rigid, high-throughput work cells or 'cloud labs' (such as those controlled with software from PAA, HighRes Bio, or Transcriptic) allow researchers to automate simplistic and repetitive experiments, they reinforce the existing 'brute force' approach to biology, with a focus on high-scale and low-variation experiments. Flexible platforms allow researchers to rapidly conduct more complicated bespoke experiments, often by utilising their existing lab hardware. Examples of companies across this aspect of the ecosystem include:

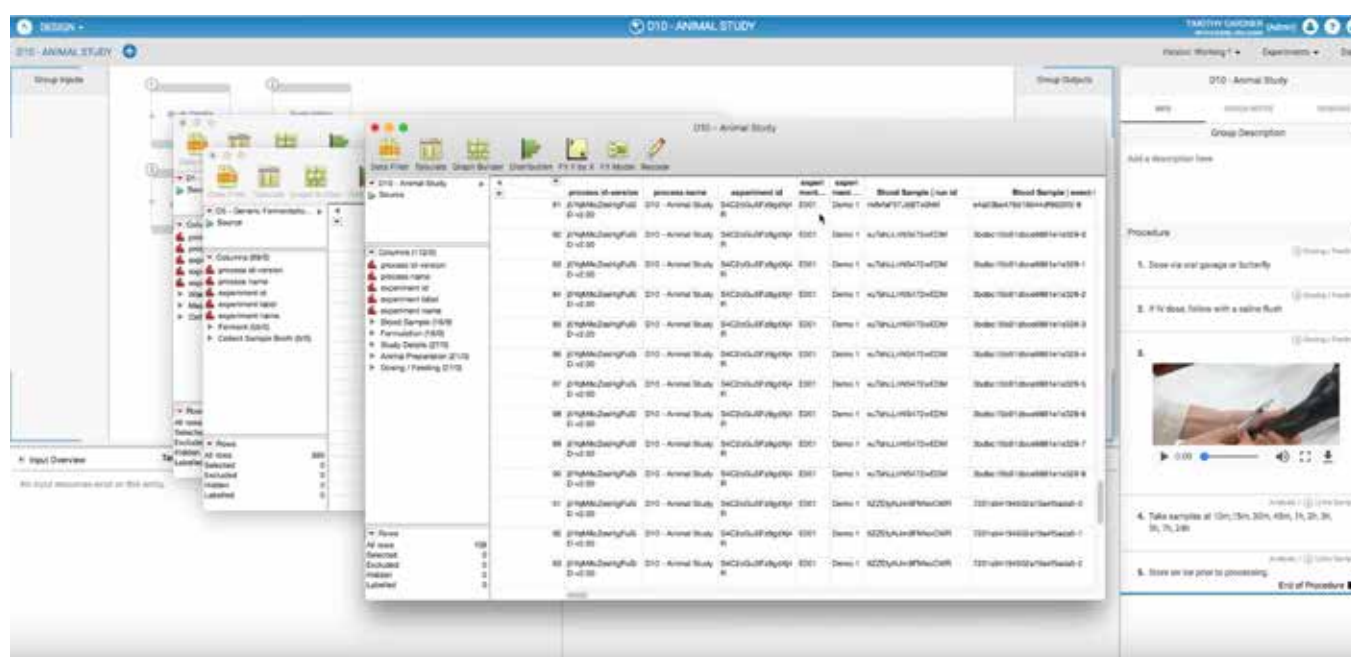
Experimental Design: **SAS Institute** (JMP) and **Stat-Ease**.

In-lab democratized experimental execution, design and simulation: **Synthace** (abstracted lab automation), **Radix Bio** (abstracted lab automation), **Beckman Coulter** (in-house software for their own hardware), **Opentrons** (protocol builder and execution for their own hardware) and **Unchained Labs** (in-house software for their own hardware).

Workcells/Cloud Labs: **Transcriptic** (workcell automation), **Emerald Cloud Labs** (cloud lab), **PAA** (workcell automation) and **High Res Bio** (workcell automation)

# Data Management & Analysis

Data management and analysis software includes a suite of tools designed to integrate varied data streams from a variety of sources, including internal databases and equipment, tools to track dataflows across locations, and tools to extract biological insight from data. Some tools in this space utilise physical hardware to act as an IoT hub for lab meta data such as temperature and utilisation rates (such as Tetrascience, iLabService and Elemental Machines) whilst others (such as Riffyn) operate with existing data lakes. They add value by allowing for the easy collating, structuring and analysis of data. Many solutions are SaaS applications, with some still requiring the use of cumbersome Excel file imports. Many also leverage the rich open source analysis packages available through programming languages such as R and Python.



**Exhibit 11.** The Riffyn Science Development Environment allows researchers to connect siloed data lakes with one another, allowing for insight to be gained by tracking data flows throughout a workflow. The software also integrates with analysis and visulation platforms such as Tableau.

Examples of companies in this space include:

Data Integration and Analysis Platforms: **Indigo Bio-automation** (mass spec data integration), **Aigenpulse** (data integration and analysis), **Exputec** (data Integration and analysis), **Vium** (data Integration for animal experiments), **Palantir** (company-wide data integration), **Riffyn** (data integration) and **BigFinite** (data management and analysis for manufacturing)

IoT platforms: **Tetrascience** (lab IoT and data integration), **Cubus Lab** (lab IoT), **iLabService** (lab IoT and data integration) and **Elemental Machines** (lab IoT).

In addition, there are several companies and many academic organisations looking to create an in-house closed loop of design-test-analyse-repeat. In all cases, this currently serves as an aspect of these companies' competitive advantage and to date hasn't been effectively democratised or spun out into the wider ecosystem (such as CAB aims to do). Examples in drug discovery include: **InSipro** and **Recurssion Pharmaceuticals**, and in 'Synthetic Biology': **Amyris**, **Zymergen**, and **Ginkgo Bioworks**.



# The role of equipment manufacturers & ecosystem enablers

An aspect that has been consciously missed from the CAB ecosystem map is equipment manufacturers, a selection of which are shown below. Liquid handling is at the core of modern biology, and can be thought of as the physical connections that link different analytical and product-producing steps in a workflow. By integrating liquid handling into the CAB ecosystem, like Synthace have done, structured datasets can be easily generated, and complex multifactorial assay preps done faster. Furthermore, with the high cost of these machines, companies such as OpenTrons and Gilson are building lower-cost automation systems to enable wider access to these technologies.



**Exhibit 12.** Examples of equipment manufacturers including Liquid Handlings, Analytics/Experimental Automation (e.g. Flow Cytometry, HPLC and Bioreactors) and high throughput/low flexibility workcells.

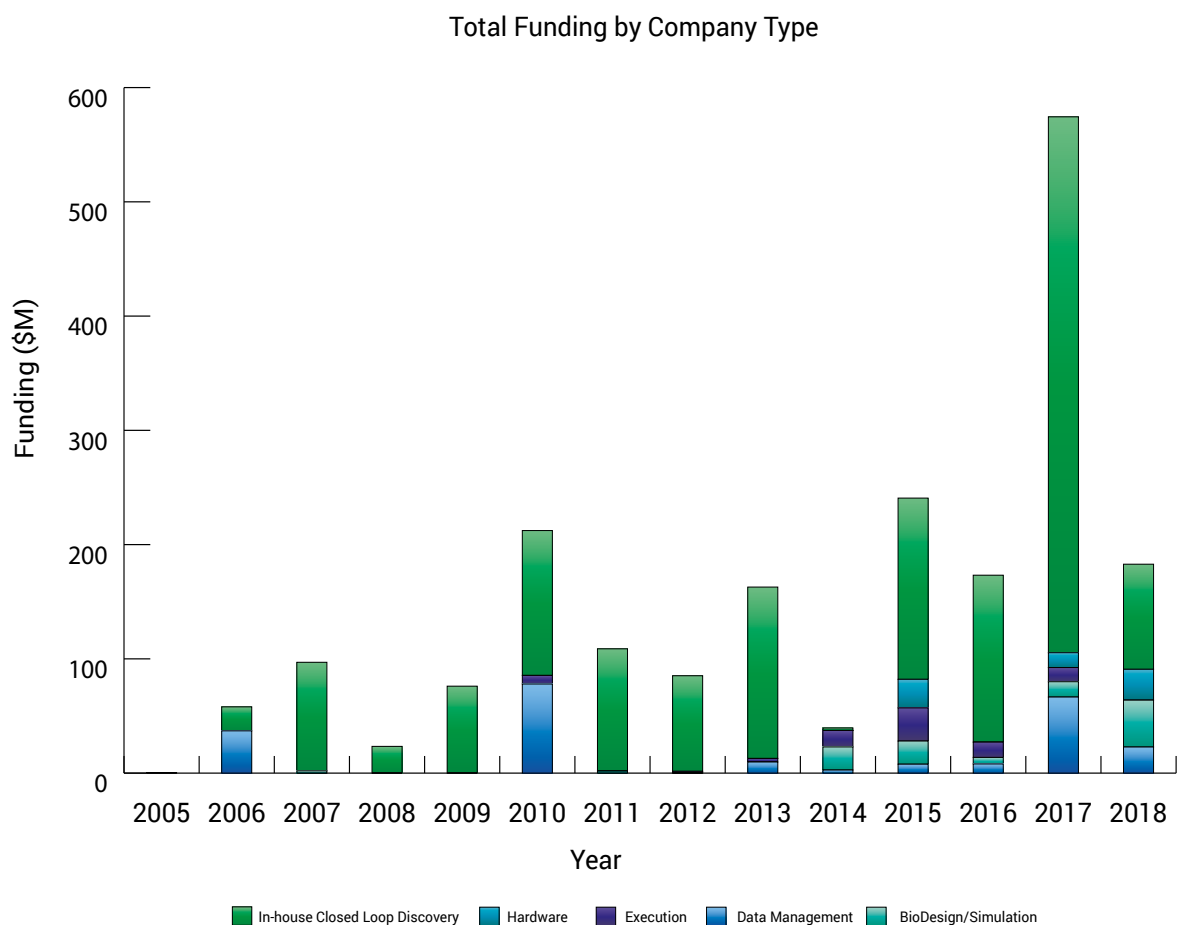
The Computer Aided Biology ecosystem has long been aligned with the Synthetic Biology community, with organisations such as SynBioBeta, iGEM and SynBiCite all having assisted companies from the CAB ecosystem. In addition, coalitions focused on the 'Lab of the Future' such as the Pistoia Alliance have also been supporting initiatives, with several foundations also creating standards to allow for the vendor agnostic integration of equipment and data sources. Examples of this include SILA and the Allotrope Foundation, although to date many have had a slow start to developing real-world support for their standards.



**Exhibit 13.** Other ecosystem participants in the CAB landscape including facilitators, standards creation organisations and start-up accelerators.

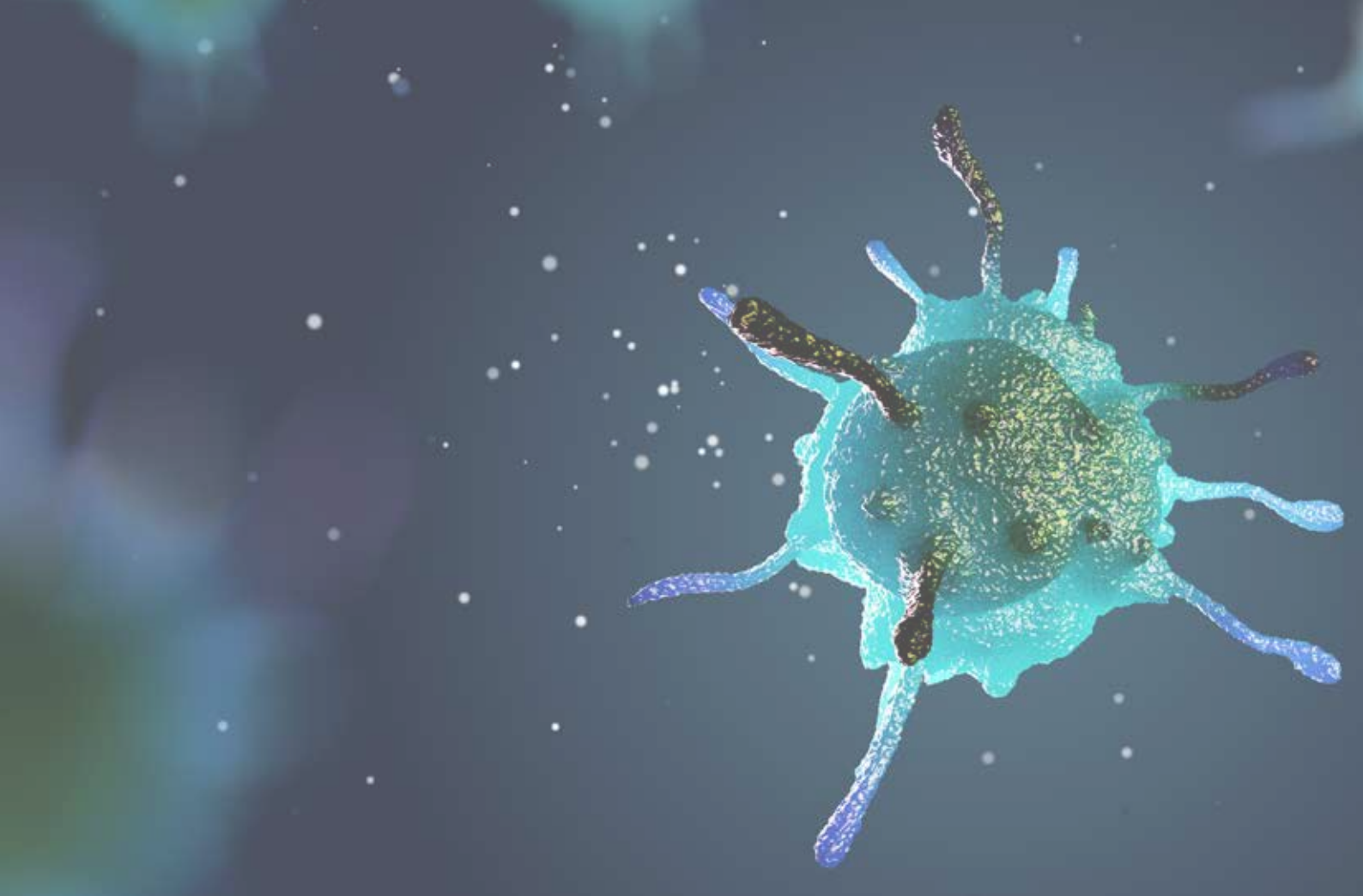
# Computer Aided Biology: funding acceleration & diversification

Since 2005, over \$2Bn USD has been invested in the CAB ecosystem, with companies attempting to create an internal ‘closed loop’ discovery model accounting for >75% of all publicly announced funding raised (Gingko, Recursion, Amyris, Intrexon and Zymergen). In comparison, companies focused primarily on software or tools to democratize access to these capabilities have raised less. Examples include: Synthace (execution, \$19.6M USD), Benchling (design, ELN and LIMS, \$27.4 USD) and Riffyn (data management, \$9.9M USD).



**Exhibit 14.** Funding into computer aided biology companies between 2005 and 2018. Major financing events include the 2010 Codexis IPO; the 2015 Gingko and Zymergen financing, the 2016 Gingko financing and the 2017 financing of Gingko and Zymergen. Since 2014, more and more companies developing software for Bio-Design, Data Management and Experimental Execution have begun to attract financing.

Interestingly, there has been little interest in this space from traditional biotechnology venture capital, with F-Prime (Benchling) and Sofinnova Partners (Synthace) being two examples of this. Instead, the majority of financing comes from deep technology focused investors such as OS Fund, DCVC, Andreessen Horowitz, Founders Fund and Benchmark capital. Late stage financings of Zymergen, Gingko and UnChained Labs, have also begun to attract funding from private equity and asset managers including ICONIQ, Baillie Gifford and Allen & Company.



## Computer Aided Biology: Summary & future evolution

CAB is an emerging ecosystem of tools that augment human capabilities in biological research, as well as a framework for understanding how modern biological research should be conducted. This framework was created by comparing modern biology (which is cost inefficient, reductionist, and error-prone) to two industries (semiconductor and automotive) which successfully leveraged digital tools to enable them to manage the complicated workflows needed to embrace complexity, while simultaneously reducing cost to the consumer and shortening R&D and production cycles.

CAB is comprised of two domains: The Digital and the Physical. The Digital, powered by artificial intelligence, includes software for the abstracted design and simulation of biological systems, as well as methods of collating, connecting, structuring and analysing experimental data from wet lab experiments. The Physical, enabled by automation, includes systems that allow for the seamless transfer of simulated biological designs into real 'wet lab' experiments via abstracted experimental design, simulation and execution. An abbreviated ecosystem of market participants has also been compiled, which shows a clear acceleration in the ecosystem.

# References

- <sup>1</sup> CAD software history CAD CAM computer aided design <http://www.cadazz.com/cad-software-history.htm> (accessed Sep 9, 2018).
- <sup>2</sup> CAD software history CAD CAM computer aided design 1980 to 1985 <http://www.cadazz.com/cad-software-history-1980-1985.htm> (accessed Sep 9, 2018).
- <sup>3</sup> Company <https://www.autodesk.co.uk/company> (accessed Sep 9, 2018).
- <sup>4</sup> HSM | CAM Software Integrated Into SOLIDWORKS And Inventor | Autodesk <https://www.autodesk.com/products/hsm/overview> (accessed Sep 9, 2018).
- <sup>5</sup> Engineering revolution: CAD/CAE advancements changing vehicle development. <https://www.wardsauto.com/news-analysis/engineering-revolution-cadcae-advancements-changing-vehicle-development> (accessed Sep 9, 2018).
- <sup>6</sup> Networking the Organization: Ford's CAD/CAM/CAE/PIM Strategy : Automotive Design & Production <https://www.adandp.media/articles/networking-the-organization-ford's-cad-cam-cae-pim-strategy> (accessed Sep 9, 2018).
- <sup>7</sup> Open-Source VLSI CAD Tools: A Comparative Study [ilin.asee.org/Conference2010/Papers/A1\\_Liu\\_Anan.pdf%0A](http://www.asee.org/Conference2010/Papers/A1_Liu_Anan.pdf%0A) (accessed Sep 10, 2018).
- <sup>8</sup> Terman, C. J. Simulation Tools for VLSI; Springer, Boston, MA, 1987; pp 57–103.
- <sup>9</sup> EDA Tools for VLSI Design [http://users.ece.utexas.edu/~mcdermot/eda\\_tools.htm](http://users.ece.utexas.edu/~mcdermot/eda_tools.htm) (accessed Sep 10, 2018).
- <sup>10</sup> Driving the Industry: 50 Years of Moore's Law <https://www.intel.co.uk/content/www/uk/en/silicon-innovations/moores-law-technology.html> (accessed Sep 9, 2018).
- <sup>11</sup> Chronology of Microprocessors (1980-1989) <http://processortimeline.info/proc1980.htm> (accessed Sep 10, 2018).
- <sup>12</sup> Scannell, J. W.; Blanckley, A.; Boldon, H.; Warrington, B. Diagnosing the Decline in Pharmaceutical R&D Efficiency. *Nat. Rev. Drug Discov.* 2012, 11 (3), 191–200.
- <sup>13</sup> Pharms-broken-business-model-an-industry-on-the-brink-of-terminal-decline/ <https://endpts.com/pharmas-broken-business-model-an-industry-on-the-brink-of-terminal-decline/> (accessed Sep 9, 2018).
- <sup>14</sup> The Brute Force Bias | In the Pipeline [http://blogs.sciencemag.org/pipeline/archives/2012/03/12/the\\_brute\\_force\\_bias](http://blogs.sciencemag.org/pipeline/archives/2012/03/12/the_brute_force_bias) (accessed Sep 10, 2018).
- <sup>15</sup> Van Regenmortel, M. H. V. Reductionism and Complexity in Molecular Biology. Scientists Now Have the Tools to Unravel Biological and Overcome the Limitations of Reductionism. *EMBO Rep.* 2004, 5 (11), 1016–1020.
- <sup>16</sup> Csete, M. E.; Doyle, J. C. Reverse Engineering of Biological Complexity. *Science* 2002, 295 (5560), 1664–1669.
- <sup>17</sup> Endy, D. Foundations for Engineering Biology. *Nature* 2005, 438 (7067), 449–453.



# Acknowledgements & disclaimer

The material in this report has been prepared by Synthace Ltd to serve as an educational overview of the Computer Aided Biology ecosystem. This information is given in summary form, based upon internal Synthace research, and does not purport to be complete. Efforts have been made to attribute rights to any images used, as well as to faithfully represent the development of the ecosystem based upon publicly available data.

Cover Image: Royalty Free License, Adobe Stock, © artemegorov

Exhibit 1: © Synthace Ltd

Page break 1: Royalty Free License, Adobe Stock, © ustas

Exhibit 2: Royalty Free License, Adobe Stock, © caifas, © ashtproductions, © phonlamaipphoto, © Archivist

Exhibit 3: Royalty Free License, Adobe Stock, © ArchMen, © Gorodenkoff, © Rainer, © V. Yakobchuk

Exhibit 4: © Internal Synthace graphs with data from cited sources and Wikipedia (transistor count)

Header 1: Royalty Free License, Adobe Stock, © Edelweiss

Page break 2: © Synthace Ltd

Exhibit 5: © Synthace Ltd

Exhibit 6: © Synthace Ltd based of literature data (cited)

Exhibit 7: © Synthace Ltd and Alamy Stock (Royalty Free Licenses):

Header 2: Royalty Free License, Adobe Stock, © Maksim Kabakou

Page break 3: © Synthace Ltd

Exhibit 8: © Synthace Ltd

Exhibit 9: Screenshot from Benchling product ([www.benchling.com](http://www.benchling.com))

Exhibit 10: © Synthace Ltd Antha screenshot

Exhibit 11: Screenshot from public Riffyn YouTube video

Exhibit 12: © Synthace Ltd

Exhibit 13: © Synthace Ltd

Exhibit 14: © Synthace Ltd based off CrunchBase data

Header 3: Royalty Free License, Adobe Stock, © 7activestudio

Authors: Dr Tim Fell, Mr Sean Ward, Dr Markus Gershater, Mr Mark Watson, Dr Peter Crane & Mr Robert Wiederhold.

Contact: [info@synthace.com](mailto:info@synthace.com)

Date: 1st October 2018

The authors thank Mr Matthew Redpath, Mx Bobby Brill, Ms Morgan Spicer (Freelance Designer) and Mr Chris Downie (Freelance Designer) for support/assistance in preparation of this document.



LBIC, 2 Royal College St, Kings Cross, London, NW1 0NH  
+44 (0) 20 7554 5877 | [info@synthace.com](mailto:info@synthace.com)  
[synthace.com](http://synthace.com)