BIOVIA

3DEXPERIENCE®

BIOVIA GENERATIVE THERAPEUTICS DESIGN ACCELERATING DRUG DISCOVERY WITH AI

Datasheet

DASSAULT SYSTEMES

Shortening the discovery cycle is key to improving productivity and enhancing competitiveness in today's Biopharmaceutical industry. BIOVIA Generative Therapeutics Design (GTD) is an artificial intelligence (AI) solution that automates the virtual creation, testing and selection of novel small molecules with a view to reducing expensive real-world testing. Research organizations can achieve true business transformation in molecular discovery by simultaneously improving lead quality and shortening discovery timelines, resulting in potential savings of millions of research dollars per program.

The BIOVIA GTD solution combines elements from five long-term and rapidly accelerating trends in drug discovery:

1. **Quantitative Structure-Activity Relationship Modeling (QSAR)** – QSAR modeling of empirical data on assays of interest using established and validated machine learning methods yields highly predictive estimations of biological activity for proposed molecules.
2. **Target Product Profile (TPP)** – Early attention to balancing all attributes required for a molecule to proceed from preclinical testing into human trials requires multi-parametric optimization at all stages of discovery (e.g., potency, safety, manufacturability, novelty, etc.).
3. **Active Learning (AL)** – This special case of Machine Learning tightly couples real and virtual activities, allowing a learning algorithm to query a user (or another information source) to label new data points with the desired outputs. In GTD, this means intentionally synthesizing molecules that will expand the domains of applicability of Machine Learning models.
4. **Generative Chemistry** – While enumeration of all molecules that could be drugs is not possible with reasonable resources, chemists can explore a large area of chemical space through iterative alteration of a starting molecule guided by a fitness function.
5. **Precompetitive Data Sources** – ChEMBL, PubChem, MDDR, etc. serve as valuable resources for identifying favored or disfavored chemical substructures for particular target families or ADME properties.

A VIRTUAL + REAL APPROACH TO DRUG DISCOVERY

The GTD discovery cycle combines virtual and real (V+R) activities in which the results of virtual generation and evaluation combine with real world synthesis and testing to allow active learning. Novel molecules advanced by the virtual generate-test-score-prune process move to the lab for synthesis and screening. Real world screening results allow the update of predictive models for subsequent cycles. Optimization continues until the target therapeutic profile is met. This iterative V+R cycle accelerates lead candidate design with improved quality, significantly reducing costs of experimentation and advancing only the most promising candidates to clinical trials.

THE BIOVIA GENERATIVE THERAPEUTICS DESIGN PROCESS

The BIOVIA GTD solution guides you through the steps of building Machine Learning models, defining a target product profile (TPP), executing generative designs including structure generation and filtering, multi-objective optimization based on Machine Learning, modeling/simulation methods and analyzing virtual results together with experimental data (V+R) as shown in Figure 1.

Machine Learning Model Building

When conducting a new study, a first step is usually building Machine Learning models for activity against your target. GTD automates the creation, validation and deployment of machine learning models through a user interface designed for domain experts such as medicinal chemists and biologists.

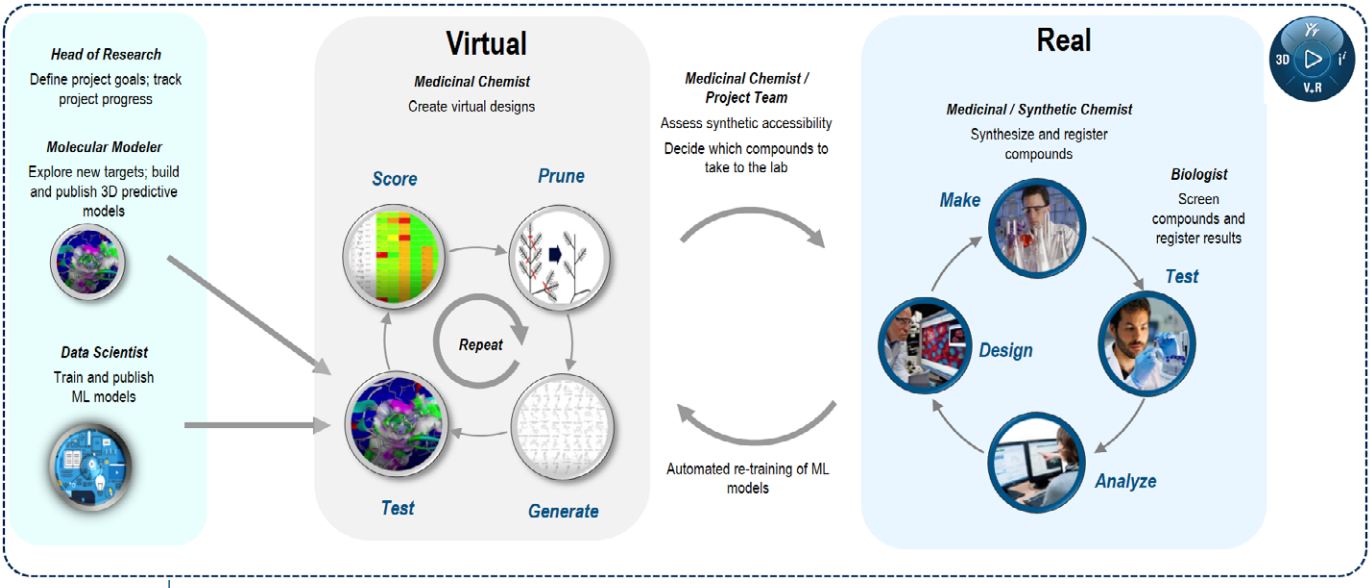


Figure 1: The discovery process is multi-disciplinary, involving different departments and specialists. The Dassault Systèmes 3DEXPERIENCE® platform optimizes both the speed and quality of these collaborative activities.

Capabilities:

- Use Machine Learning methods such as Bayesian and Random Forest that are pre-optimized for activity model building.
- Validate models by inspecting receiver operating characteristic (ROC) curves and classification statistics based on cross-validation.
- Publish models for use by team members.
- Update models when new data becomes available after active learning cycles.

Defining a Target Product Profile

When conducting a new study, the first step is to define your Target Product Profile (TPP), which represents the objectives to use when designing a new drug candidate. You can choose from built-in models or models you create yourself. Additionally, chemists can specify important preferences depending on project objectives. GTD will use these parameters to tailor-design compounds matching this TPP by performing a multi-objective optimization.

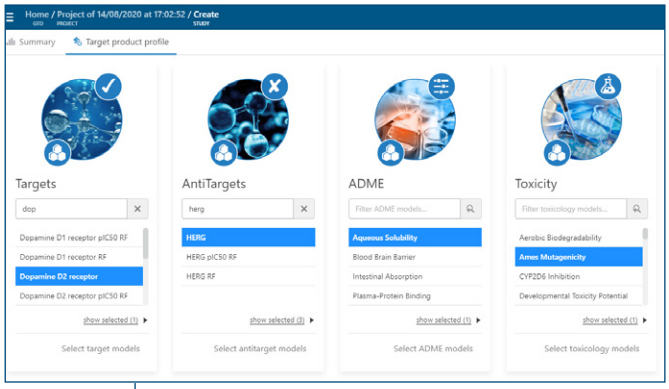


Figure 2: Define your target product profile by selecting pre-built models for common targets, anti-targets, ADME and toxicity properties.

Structure Generation and Filtering

GTD offers a variety of methods for generating virtual compounds. An evolutionary algorithm generates molecules for each input molecule using a variety of molecular transformations, including Matched Molecular Pairs (MMP), MedChem reactions, replacement of ring assemblies, replacement of molecular scaffolds, molecular variations (split, trim, rearrange, crossovers) and molecular mutations (change of atom and bond types, adding and deleting atoms, opening and closing rings). There are options to bias the generative algorithm to produce compounds similar to a target structure or ones that contain a certain scaffold.

Alternatively, you can generate structures by enumerating a chemical reaction using a selected set of reactants. As the number of total possible molecules is normally too large for a full enumeration, GTD enables a stochastic evolutionary process that samples the chemical space covered by the virtual library. The advantage of this approach is that it builds in synthesizability of compounds. Enumeration can also be based on Markush structures with a core molecule with Rgroup positions and sets of attachment fragments.

When designing compounds for a specific therapeutics application, it is important to control their property profile (drug-likeness, stereochemistry and flexibility), removing undesirable substructures and building in synthetic feasibility.

Capabilities:

- Explore chemical space using a combination of well-proven and novel methods.
- Include reaction-based/R group-based enumeration.
- Filter virtual compounds for undesirable substructures based on universal, company and project-specific criteria.
- Incorporate design preferences based on 'drug-likeness' profiles tailored to a therapeutic area.
- Prioritize based on synthesizability considerations.

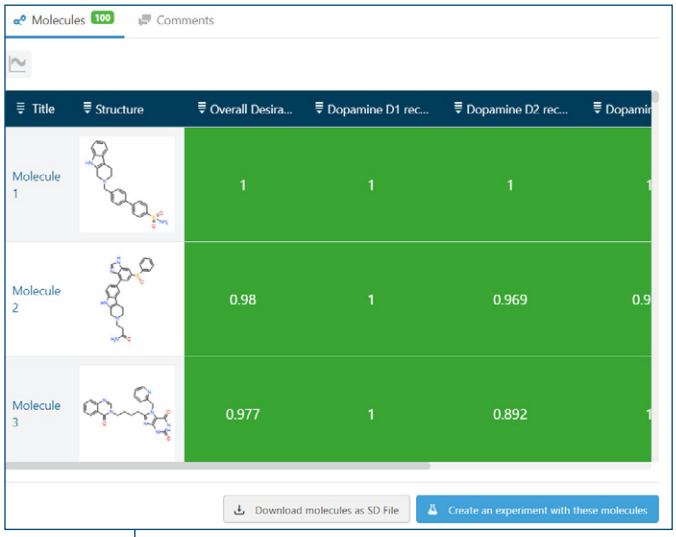


Figure 3: Analyze molecular structure and property profile for highest scoring virtual compounds.

Multi-objective Optimization

One of the challenges involved in searching for lead compounds is understanding the trade-offs to make when optimizing multiple design criteria. GTD allows chemists to design improved lead compounds by maximizing an overall desirability function based on individual desirability profiles based on the TPP. For each objective in the TPP, the chemist uses the profiles to specify how good is "good enough" to meet the project goals. GTD displays model prediction distributions and empirical probability values (e.g., the probability a compound will be active if the model score exceeds a certain value) to aid in this process. Alternatively, GTD provides reasonable default desirability profiles based on the TPP.

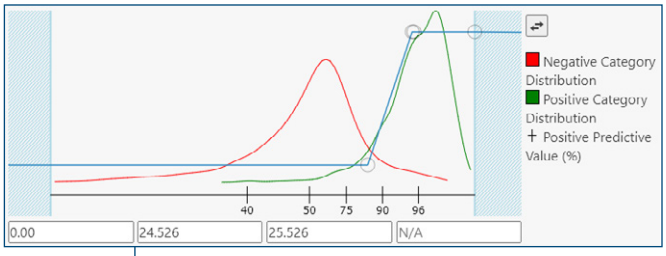


Figure 4: Tune desirability profile for each optimization objective based on characteristics of the Machine Learning model.

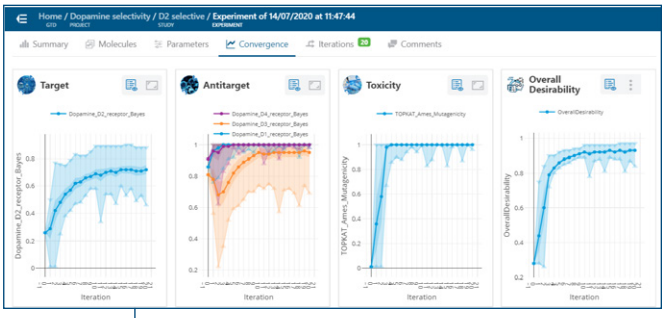


Figure 5: Monitor the optimization progress for individual objectives.

GTD then uses an evolutionary algorithm to drive compounds to maximum desirability, typically generating millions of compounds over multiple cycles in order to end up with a small optimal set. The system uses clustering or a Pareto method to maintain diversity, ensuring that the solution set is not a set of nearly identical compounds with only slight variations from each other. After completing the optimization (a "virtual experiment"), the chemist can view the convergence charts and resulting compounds in order to decide whether to synthesize and assay some of the resulting structures, or to adjust the optimization parameters and perform another run.

Capabilities:

- Use desirability profiles to configure each optimization objective.
- Graphically adjust each profile based on probabilities determined by applying the Machine Learning model to known compounds.
- Increase the "weight" of the most important objectives (e.g., activity) to help ensure that they are met during the optimization.
- Specify the overall goal of the optimization, whether immediate compound improvement or exploration of chemical space to improve the models for later compound improvement.

Virtual plus Real (V+R) Analysis

Medicinal chemists can use GTD to combine their unique expertise and intuition with the algorithmic power of Artificial Intelligence. By analyzing the structure and characteristics of proposed compounds, a chemist can tailor design criteria to steer subsequent virtual iterations. In addition, as part of the active learning process, combining virtual and real (experimental) data can accelerate decision-making.

Capabilities:

- Chart any combination of properties to investigate trade-offs when designing with multiple objectives.
- Select virtual compounds with desirable wet-lab characteristics and seed subsequent virtual iterations.
- Compare against historic virtual and real (V+R) results for proposed virtual compounds.

Figure 6: Chart any combination of properties to understand trade-offs between properties and evolution during virtual optimization cycles.

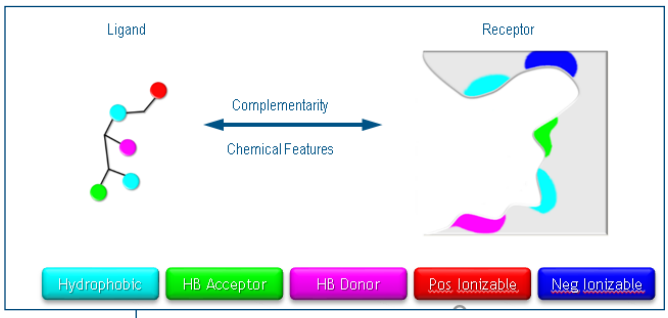


Figure 7: Pharmacophore models can help improve virtual design by incorporating 3D knowledge of the active site.

Modeling and Simulation

3D modeling and simulation methods can enhance the accuracy of predictions for drug potency, efficacy and selectivity, leading to higher quality virtual compounds. One particularly useful method uses 3D pharmacophores to score virtual compounds based on models built from knowledge of the active site. Other methods known to increase the accuracy of the prediction include molecular docking, free energy perturbation and multi-site lambda dynamics.

Capabilities:

- Build 3D pharmacophore models based on known active molecules.
- Use these models as an additional scoring function in the multi-objective optimization process.
- Publish custom-built 3D models for use by team members.

Knowledge Management

BIOVIA GTD is based on the 3DEXPERIENCE platform, which helps to track results generated by both 'virtual' and 'real' studies. GTD can also interface with auxiliary corporate data stores for compound registration, inventory and screening.

Ready access to both positive and negative data allows scientists to reinforce predictive models and track the lineage of compounds from ideation into development. This 'digital continuity' is critical for tracking intellectual property (IP) produced in a discovery organization.

By leveraging data from historic compounds, scientists can reduce experimentation, saving considerable time and money, while also rapidly identifying entities with known liabilities. Scientist can also collaborate more easily by being able to identify internal experts who have worked with similar compounds. Further, GTD allows users to explore the chemical space more efficiently leveraging existing knowledge from virtual and real compounds across departments and throughout history.

This comprehensive, knowledge-based drug discovery solution dramatically increases the impact of an AI-driven approach, truly helping to transform the efficiency of discovery efforts.

CLOUD DEPLOYMENT

BIOVIA GTD is available on the cloud providing organizations with a secure, validation-ready research informatics solution. As a cloud configuration, this solution is completely web-based and exists outside your firewall. The solution helps avoiding costly in-house development efforts and minimizes Information Technology (IT) overhead costs. It frees up internal resources by eliminating the need for internal IT staff to manage applications and servers, keep track of upgrades, maintain performance and manage security. Virtual machines can be quickly spun up or down according to changing project needs. The solution delivers secure access via SSL encryption, helping to ensure system security regardless of user location. Authorized users anywhere in your company, anywhere in the world, can access this secure research informatics solution at any time.

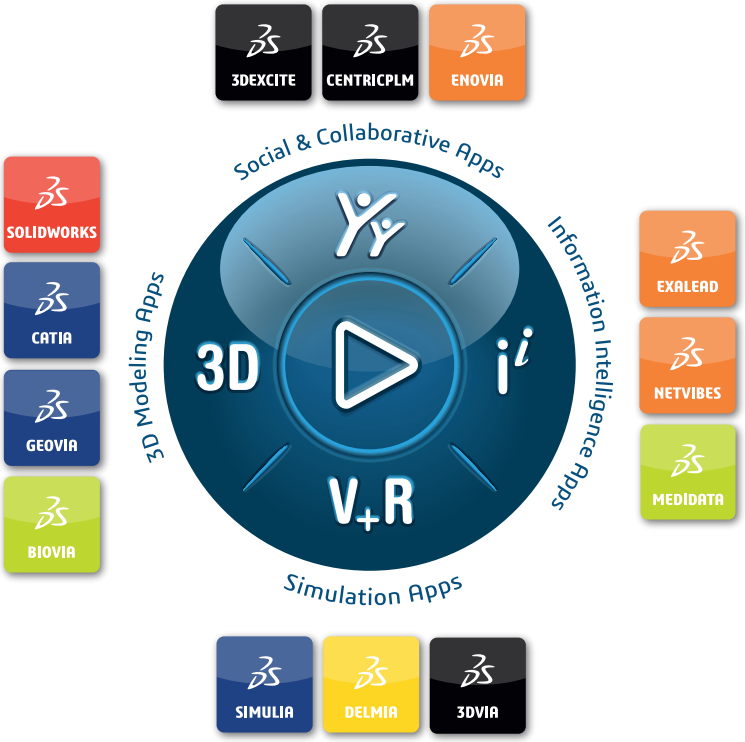
LEARN MORE

SOLUTION BENEFITS

The following benefits* can be realized:

BIOVIA Solution Capabilities	Potential Improvements
1. AI-Driven Virtual Lead Optimization • Generative molecule design algorithms • High accuracy statistical models and simulations • User interface that enhances understanding of results	↑ R&D capacity and speed ↓ Number of lead optimization projects (22%-28%) ↓ Number of compounds to synthesize (30%-50%) ↓ Cost per project (16%-24%)
2. Integration of Virtual and Real experimentation • Active learning approach to selection of molecules for synthesis • External collaboration environment for projects with CROs • Reduce complexity by implementing an automated V+R cycle	↑ Data and model accuracy ↓ Experiment cycle time ↓ Communication bottlenecks
3. Aggregation of Organizational Knowledge • Library of models for all targets you have investigated • Tracking of decisions, rationales and contributing data	↑ Cross-function alignment ↑ Data driven go/ no-go decisions ↓ Reporting/summarization efforts
4. Simplified Research Informatics Systems • Avoid costly in-house development efforts • Adapt deployment according to changing project needs	↓ IT overhead and costs ↓ Implementation duration (70-80%) ↑ User adoption rates (30-40%)

*Estimates based on customer use cases and validation studies



Our **3D**EXPERIENCE® platform powers our brand applications, serving 11 industries, and provides a rich portfolio of industry solution experiences.

Dassault Systèmes, the **3D**EXPERIENCE Company, is a catalyst for human progress. We provide business and people with collaborative virtual environments to imagine sustainable innovations. By creating 'virtual experience twins' of the real world with our **3D**EXPERIENCE platform and applications, our customers push the boundaries of innovation, learning and production.

Dassault Systèmes' 20,000 employees are bringing value to more than 270,000 customers of all sizes, in all industries, in more than 140 countries. For more information, visit **www.3ds.com**.

Europe/Middle East/Africa
Dassault Systèmes
10, rue Marcel Dassault
CS 40501
78946 Vélizy-Villacoublay Cedex
France

Asia-Pacific
Dassault Systèmes K.K.
ThinkPark Tower
2-1-1 Osaki, Shinagawa-ku,
Tokyo 141-6020
Japan

Americas
Dassault Systèmes
175 Wyman Street
Waltham, Massachusetts
02451-1223
USA

DASSAULT SYSTEMES | The **3D**EXPERIENCE® Company