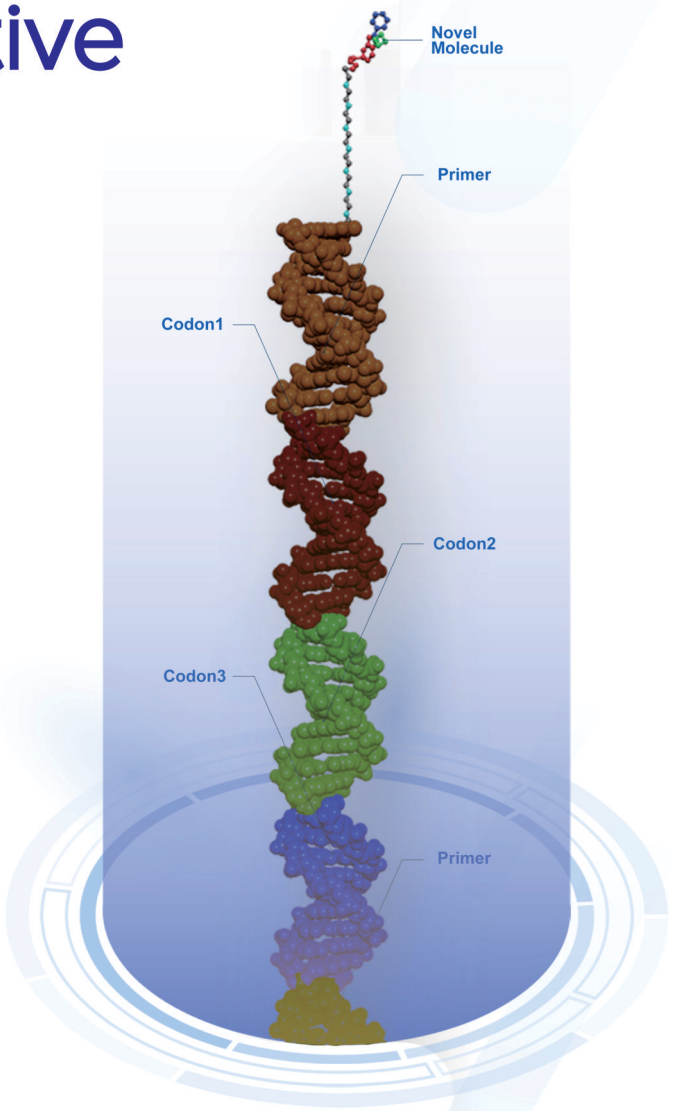
WHITEPAPER

DNA-ENCODED LIBRARY

Advancing Innovative Drug Discovery



Content outline of white paper

PRESENTED BY: HITGEN

PUBLISHED BY: FIERCE Biotech

Contents

Abbreviation

ADMET: Absorption, Distribution, Metabolism, Excretion and Toxicity

AI: Artificial Intelligence

ANN: Artificial Neural Network

ASMS: Affinity Selection-Mass Spectrometry

BB: Building Block

BLI: BioLayer Interferometry

BRD4: Bromodomain-containing protein 4

BRET: Bioluminescence Resonance Energy Transfer

CADD: Computer-Aided Drug Discovery

CRBN: Cereblon

CRYSPR: Clustered Regularly Interspaced Short Palindromic Repeats

Cys: Cysteine

DEL: DNA-Encoded Library

DUB: Deubiquitinylating Enzyme

DWAR: Data Warrior

ELISA: Enzyme-linked Immunosorbent Assay

FAC: Fluorescence-Activated Cell

FBDD: Fragment-Based Drug Discovery

FI: Fluorescent Intensity

FP: Fluorescence Polarization

FRET: Fluorescence Resonance Energy Transfer

GFP: Green Fluorescent Protein

GPCR: G Protein-Coupled Receptor

His: Histidine

HTS: High-Throughput Screening

IC50: Half Maximal Inhibitory Concentration

ITC: Isothermal Titration Calorimeter

Ka: Association Constant

KD: Binding Constant

Kd: Dissociation Constant

KO: Knock-out

LE: Ligand Efficiency

Lys: Lysine

MOA: Mechanism of Action

MW: Molecular Weight

NAA50: Nα-terminal Acetyltransferase 50

NAT: Nα-terminal Acetyltransferase

NGS: Next Generation Sequencing

PCR: Polymerase Chain Reaction

POI: Protein of Interest

PolyO: log10 Transformed Poisson Probability Product Score with Normalization

PROTAC: Proteolysis Targeting Chimeria

QSAR: Quantitative Structure-Activity Relationship

ROCK2: Rho Associated Protein Kinase 2

SAR: Structure-Activity Relationship

Ser: Serine

SPR: Surface Plasmon Resonance

SSR: Structure-Signal Relationship

tPSA: Topological Polar Surface Area

TRIC: Temperature Related Intensity Change

TSA: Thermal Shift Assay

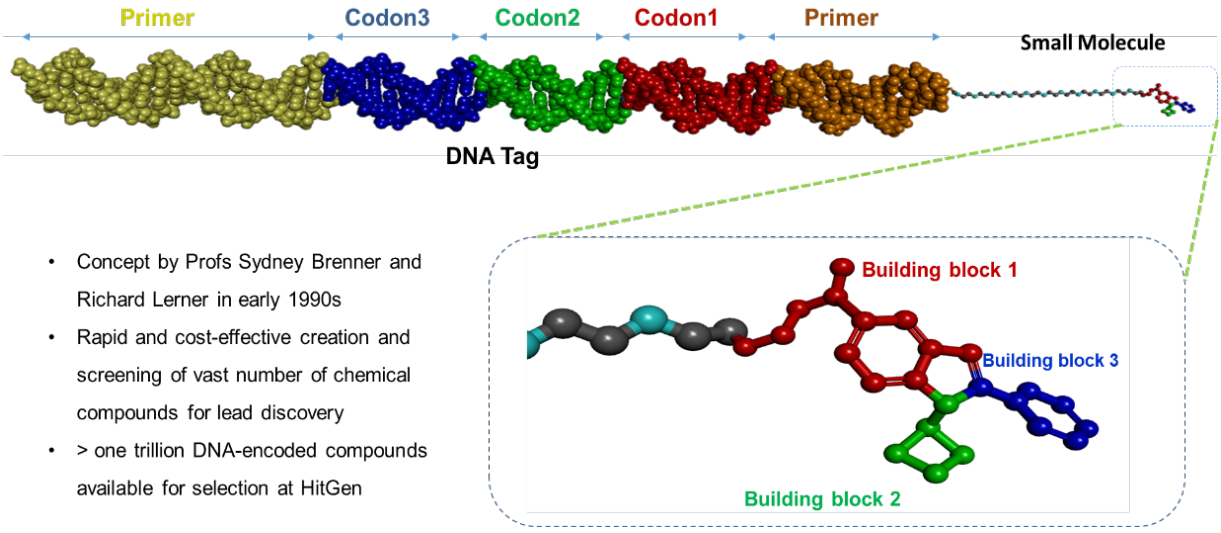
Tyr: Tyrosine

WB: Western Blot

1. Basic Principles of DNA-Encoded Library Technology

Since the introduction of initial concept by Brenner and Lerner in 1992 (Brenner, S., Lerner, R. A. Proc. Natl. Acad. Sci. 1992, 89, 5381-5383), DNA-encoded libraries (DELs) have evolved into a valuable platform for hit identification in early drug discovery. Compounds in DELs consist of building blocks serving as diversity elements, which are encoded by unique DNA sequences as identification "barcodes," and are assembled combinatorially in a split-and-pool fashion with DNA-compatible reactions. The DEL technology enables the fast and economic synthesis, screening and analysis of DNA-encoded collections of millions to billions of compounds against biological targets through a confluence of molecular biology, combinatorial chemistry, high throughput sequencing and advanced informatics techniques.

Compared to traditional high-throughput screening (HTS), DEL possesses multiple merits, including a much larger library size and more cost-efficient affinity screens. DNA-encoded library (DEL) platform is a fast-growing, cutting-edge hit discovery platform that is largely used in big pharma and biotechs. HitGen's formidable DEL platform that features over one trillion small, druglike molecules has been widely used by leading global pharmaceutical and biotech companies.



- Concept by Profs Sydney Brenner and Richard Lerner in early 1990s
- Rapid and cost-effective creation and screening of vast number of chemical compounds for lead discovery
- > one trillion DNA-encoded compounds available for selection at HitGen

Basic Construct of a DEL Compound

2. DNA-Encoded Library Design

2.1. DNA-Encoded Library Design at HitGen

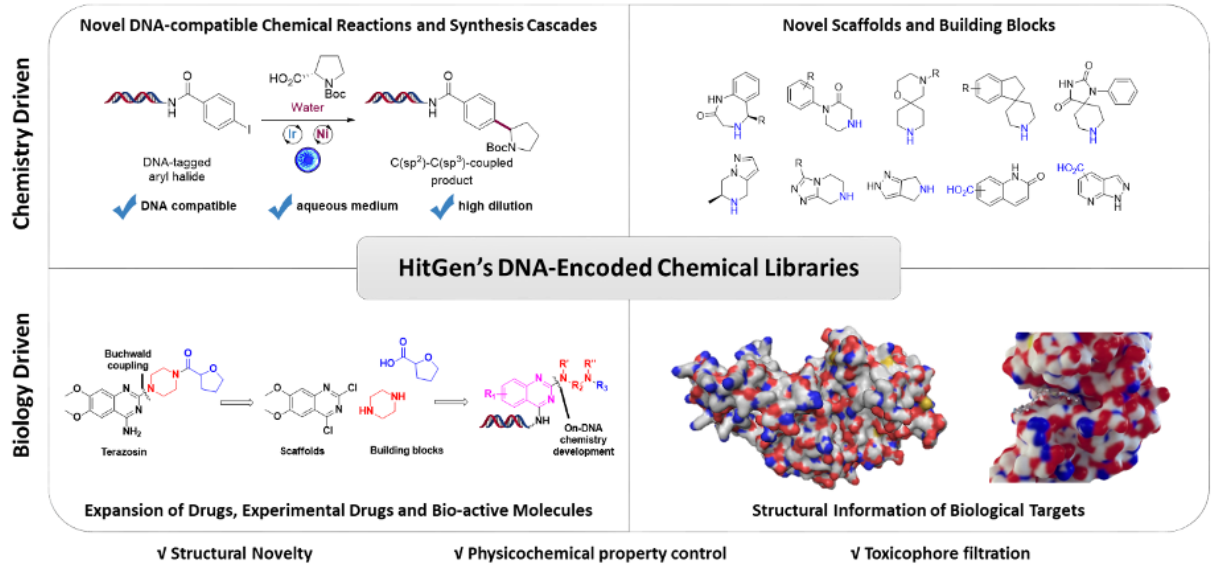
One of the fundamental steps of innovative drug research & development is the discovery of molecules that bind to a biological target of interest and exert the desired pharmaceutical effects. Library compound designed for biological target screening is critical for identification of high quality hit/lead compounds. DNA-encoded library technology has emerged and matured as a powerful tool for generating unprecedented number of novel molecules for drug discovery research. The design of DNA-encoded libraries at HitGen is considered from both chemistry-driven and biology-driven aspects, while physicochemical properties as well as empirical filters such as frequent hitters and toxophores are also applied to improve the drug likeness of library compounds.

Chemistry-Driven Library Design:

- Design of DNA-encoded libraries driven by employing novel DNA-compatible chemical reactions and synthesis schemes
- Design of DNA-encoded libraries driven by using novel and drug-like scaffolds and building blocks

Biology-Driven Library Design:

- DNA-encoded library design to cover and to expand chemical spaces of drugs on the market, drug candidates in clinical studies, bioactive molecules in preclinical stages.
- Library design based on structural information of specific biological targets or target types



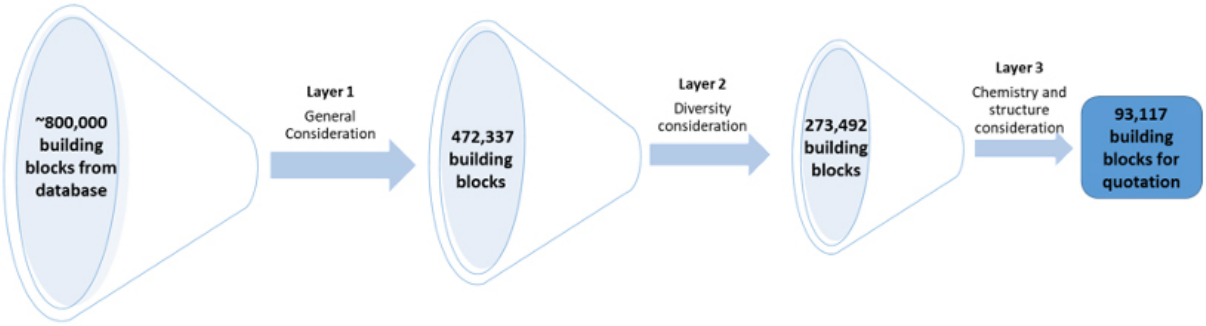
The Systematic Design of DELs at HitGen

2.2. Expansion of Building Blocks and Reagents, Methodology Used to Select Reagents

The physicochemical property, drug likeness and diversity of DNA-encoded library compounds are largely relied on chemical building blocks and therefore require diverse chemical building blocks to meet strict criteria, such as physicochemical properties, chemical space coverage, chemical reactivity, etc.

Commercial Building Blocks

We have gathered structural information of ~800,000 chemical building blocks from major building block suppliers worldwide and therefore conveniently select building blocks for HitGen DELs. As shown below, over 93,000 building blocks candidates were finally selected before final quotations by setting preconceived criteria of diversity, reactivity, drug-likeness, physicochemical properties, etc.



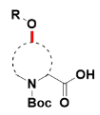
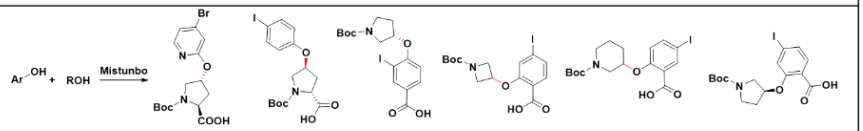
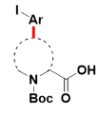
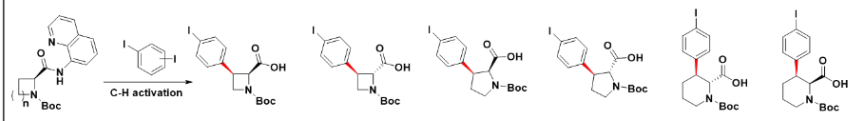
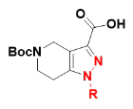
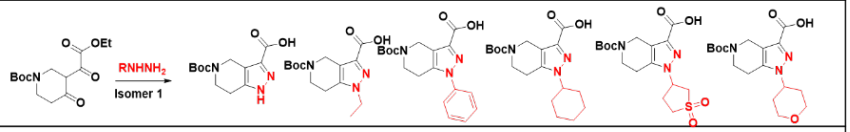
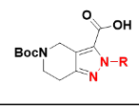
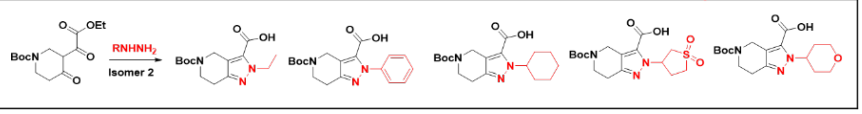
Layers	Layer 1	Layer 2	Layer 3
Steps	**Step 1.** Removing overlaps **Step 2.** BBs classified by reacting groups and removing BBs without DNA compatible reacting groups	**Step 1.** Pharmacophore diversity analysis[1] **Step 2.** 3D shape analysis[2] **Step 3.** Tanimoto analysis	**Step 1.** Removing PAIN structures[3] **Step 2.** physicochemical property control for each type of BBs **Step 3.** Removing BBs with unothogonal chemistry **Step 4:** ≤ 2 uncertain chiral centers

1. *Molecules. Mol. Inf.* **2013**, *32*, 133–138. 2. *J. Med. Chem.* **2009**, *52*, 6752–6756. 3. *J. Med. Chem.* **2010**, *53*, 2719-2740

Criteria of Reagent Selection from Commercial Sources for DEL Synthesis

Proprietary Building Blocks

Many scaffolds and building blocks are structually attractive. But not all these structures can be achieved directly by development of DNA-compatible chemistry. HitGen has built a dedicated team to design and synthesize proprietary building blocks for novel and high-quality library synthesis. The goal is to synthesize DNA-encoded library compounds which can't be accessed by readily available DNA-compatible chemistry (See examples below).

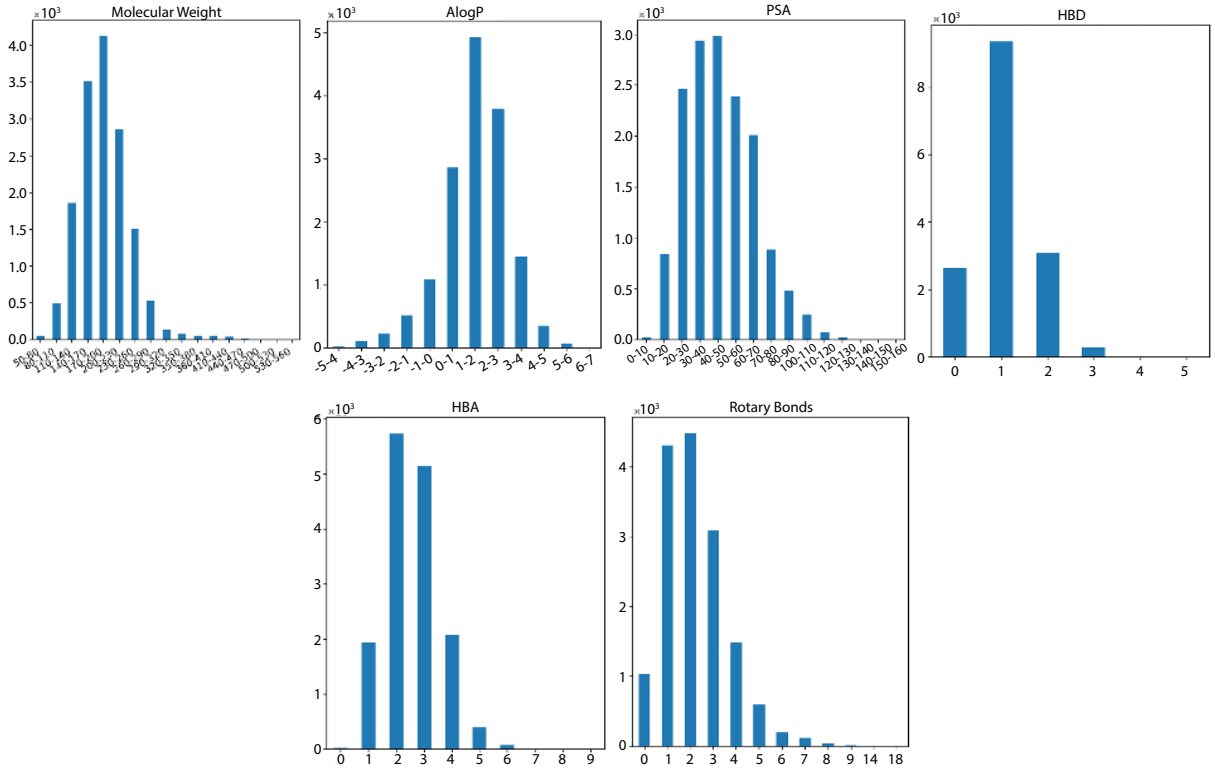
Issue to Tackle	Scaffolds	Examples of Scaffolds
C-O Bond Formation		
Sp2-SP3 Bond Formation		
Regioselectivity Issue in DEL Synthesis		
		

Examples of Scaffolds to Tackle Unachievable Bond Formation with DNA Compatible Chemistry

HitGen has accumulated over 30,000 building blocks and over 6,000 scaffolds using chemoinformatics tools, while approximately 20% of which are novel and can't be retrieved from Reaxy database. HitGen has been continuously increasing the building block collections for further increasing the nolvelty and diversity of library compounds.

Physicochemical Property

Building blocks, as key materials of DEL synthesis, are crucial for achieving diversity, novelty, and drug-likeness of final DEL molecules. For these purposes, physicochemical property of building blocks are considered as a critical factor when selecting building blocks for ordering or synthesis.



Physicochemical Properties of Reagents for DEL Synthesis

Structure Diversity

To address the structure diversity of building blocks, HitGen informatics group creates a comprehensive method to represent the building block diversity: pharmacophore feature pairs to capture potential interactions and normalized principle moments of inertia (NPMI) to measure

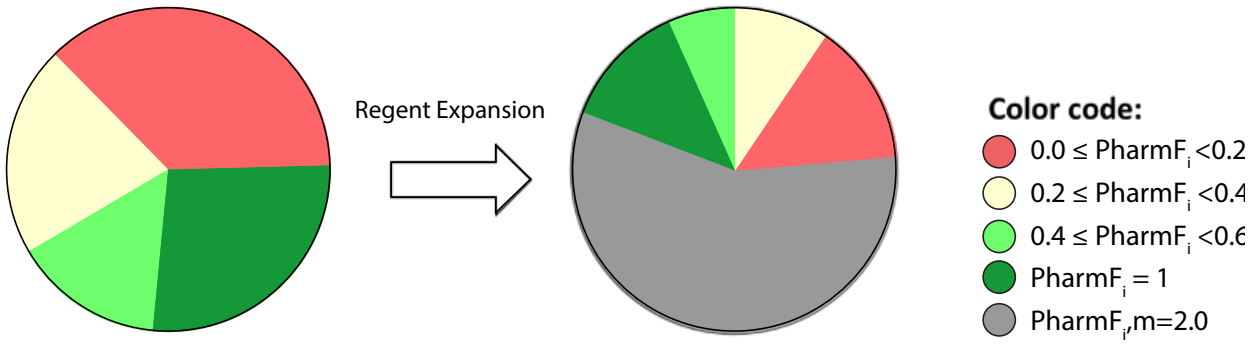
the chemical space coverage.

The attractiveness score of a fingerprint is defined by the pharmacophore fingerprint:

PharmFi pharmacophore score for fingerprint i

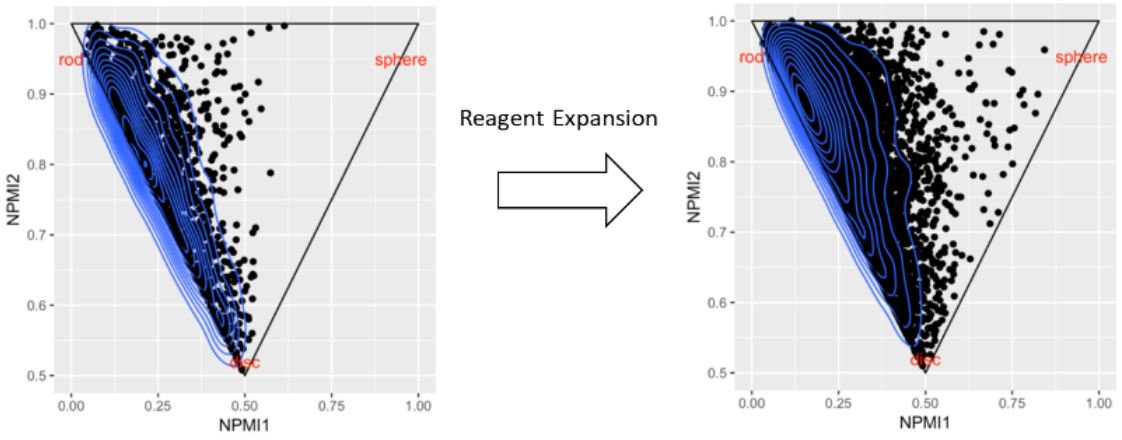
= 1/number of times fingerprint i occurs in the reference set

=2 if novel pharmacophore to reference



BB and Scaffold Pharmacophore Diversity Analysis at HitGen

Normalized principle moments of inertia (NPMI) used here to consider 'escape from flatland', the purpose is to improve PK properties as well as to access new chemical space.

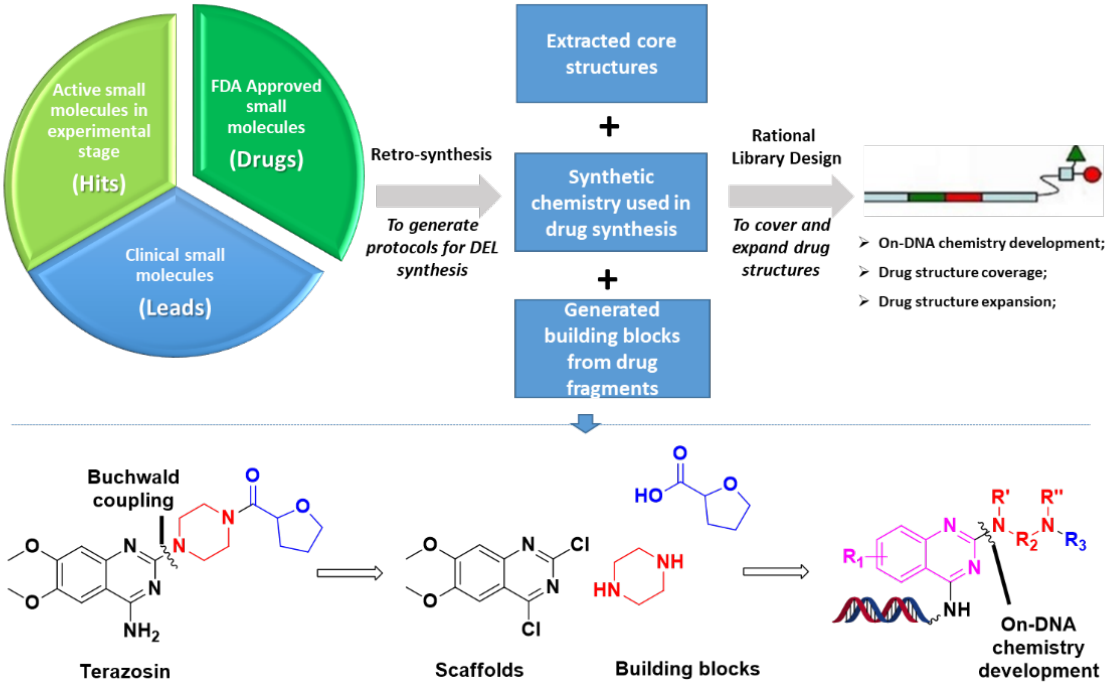


BB and Scaffold Shape Analysis at HitGen

Biological Relevance

Known active compounds are generally great sources for library compound design. We recently built a "biologically relevant fragment database" by extracting key fragments of active compounds from SureChEMBL compounds. The extracted fragments are attached with suitable functional

groups and are combinatorially combined together to encoded compounds by developing on-DNA chemistry. In this way, most of key fragments of known active compounds can be convered into library compounds.



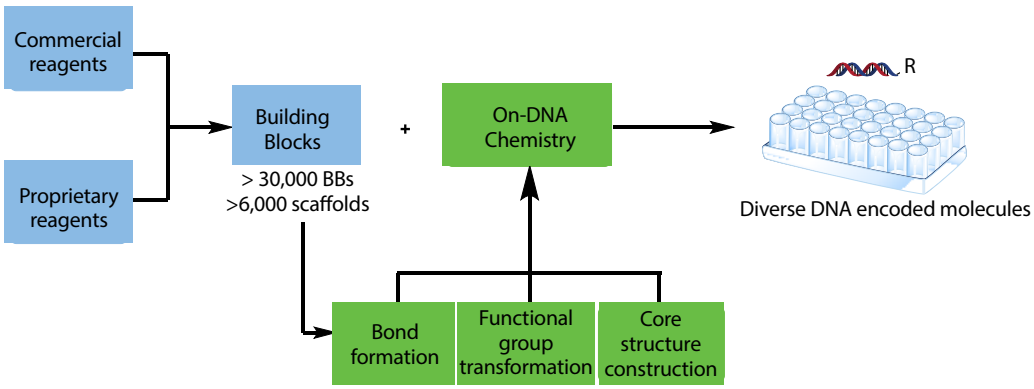
Coverage and Expansion of Bioactive Compounds in HitGen DELs

2.3. Expansion of Chemistry Used to Build DELs, Highlight Papers Published and Novel Molecule Structures

One of advantages of DELT is to enable the exploration of huge chemical space in a cost-effective and timely manner. The diversity of chemical structures of libraries is one of the key features of DELs. While DELs are synthesized by combinational assembly of building blocks (BBs) using a spilt-pool strategy in aqueous solutions, DNA-compatible chemistries are fundamental for constructing diverse and drug-like compounds. HitGen has made great efforts to expand the chemistry toolbox for DEL synthesis.

- There are over 30,000 BBs at HitGen selected from over 800,000 commercially available chemicals by considering structural diversity, pharmacophore representation, chemical reactivity and toxophore filtration. Most of those are mono-functional chemicals such as amines, acids, aldehyde, etc. Also there are bi-functional chemicals and tri-functional chemicals collected from commercial sources or designed and synthesized by a dedicated team at HitGen which could improve the diversity and novelty of DEL products.
- Since DELs are synthesized via combinational assembly of BBs in the presence of DNA tags, which make the reactions have to be performed in aqueous phase. With the advancement of synthetic chemistry, enormous bond formation reactions could be

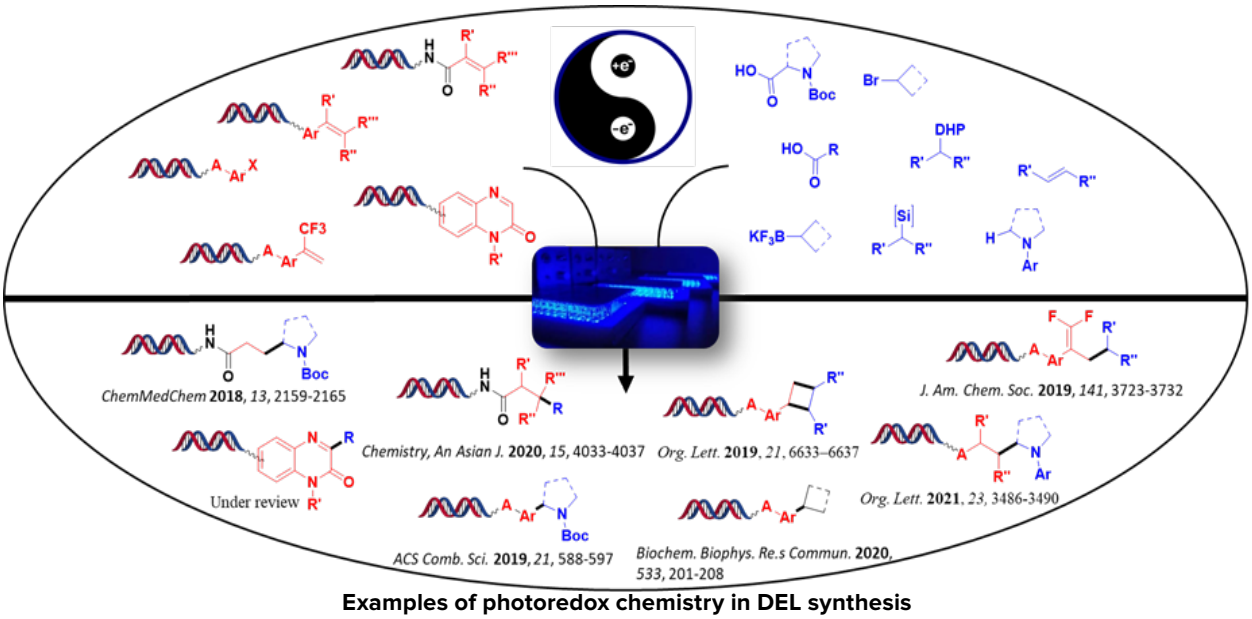
applied to construct novel structures. To explore the chemical space of HitGen DELs, on-DNA chemistry research has been carried out from aspects of bond formation, functional group (FG) transformation, core structure construction, etc. With continuous efforts, more than 100 types of on-DNA chemistry protocols have been established and applied into HitGen DEL synthesis.



Development of On-DNA Chemistry at HitGen

2.3.1 On-DNA Bond Formation

With the advancement of synthetic chemistry, bond formation reactions are widely applied by medicinal chemists to construct novel structures. However, many bond formation reactions are difficult to achieve in DEL construction, largely limited by DNA compatibility and solubility. With continuous efforts from our chemists, many on-DNA bond-formation chemistries have been achieved successfully, such as C-C, C-N, C-O, C-S, etc. It is well known that photoredox

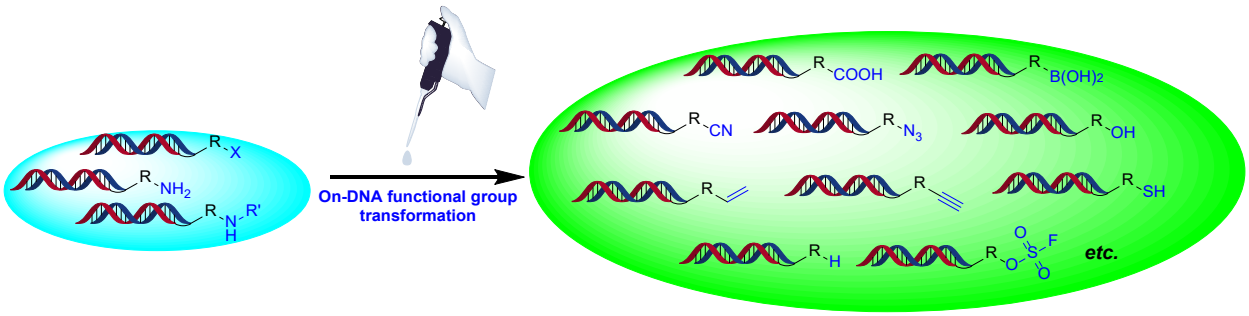


Examples of photoredox chemistry in DEL synthesis

catalysis is now an indispensable tool in organic synthesis due to its mild condition and wide substrate scope. Photoredox catalysis has been recognized as a powerful tool for discovery and development of novel chemistries, and was recently applied to on-DNA construction of C(sp3)-C(sp3) bonds. In 2020, as part of collaboration with Pfizer, the photoredox catalyzed [2+2] cyclization was published on Organic Letters (Kölmel, D. K.; Ratnayake, A. S.; Flanagan, M. E.; Tsai, M-H.; Duan, C.; Song, C. Org. Lett. 2020, 22, 2908–2913).

2.3.2 On-DNA Functional Group Transformation

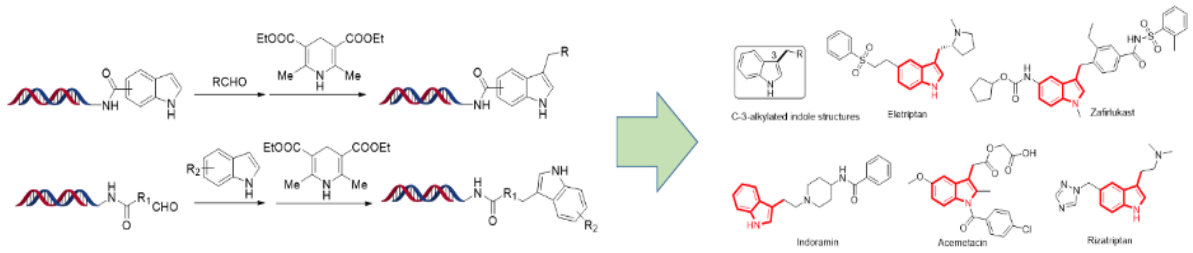
The numbers and diversity of these available BBs along with availability of DNA compatible chemistries are fundamental for constructing diverse and drug-like compounds. Mono-functional BBs, such as acids, amines, aldehydes, boronic acids, etc., are well-known reagent types used for DNA-encoded library synthesis. Such BBs are largely commercially available and represent vast chemical space as a portion of DECL molecules. Contrastingly, some varieties of bifunctional or trifunctional scaffolds tend to have limited commercial availability. Therefore, many endeavors have been made recently to transform widely available functional groups into others on DNA for subsequent derivitization in DECL synthesis to increase diversity.



On-DNA Functional Group Transformation

2.3.3 On-DNA Core Structure Construction

Privileged structures of drugs are of great interest for medicinal chemistry. HitGen's scientists have made a lot of achievements to develop on-DNA chemistry to bring key privileged core structures of drugs into DELs, such as C3-alkylated indoles (see below, Cai, P. W.; Yang, G. Y.; Zhao, L. Z.; Wan, J. Q.; Li, J.; Liu, G. S. Org. Lett. 2019, 21, 6633–6637), 2-aminobenzimidazole (Su, L. Q.; Feng, J.; Peng, T.; Wan, J. Q.; Fan, J.; Li, J.; O'Connell, J.; Lancia, D. R.; Franklin, G. S.; Liu, G. S. Org. Lett. 2020, 22, 1290–1294), Indazolone (Yao, Y. P.; Deng, Z. F.; Feng, J.; Zhu, W. W.; Li, J.; Wan, J. Q.; Liu, G. S. Org. Lett. 2020, 22, 16, 6277–6282), substituted pyrimidine (Wu, R. F.; Gao, S.; Du, T.; Cai, K. L.; Cheng, X. M.; Fan, J.; Feng, J.; Shaginian, A.; Li, J.; Wan, J. Q.; Liu, G. S. Chem Asian J. 2020, 15, 1–6), etc. Out of top 100 privileged structures appeared in FDA approved drugs, up to ninety have been covered into our DEL products.

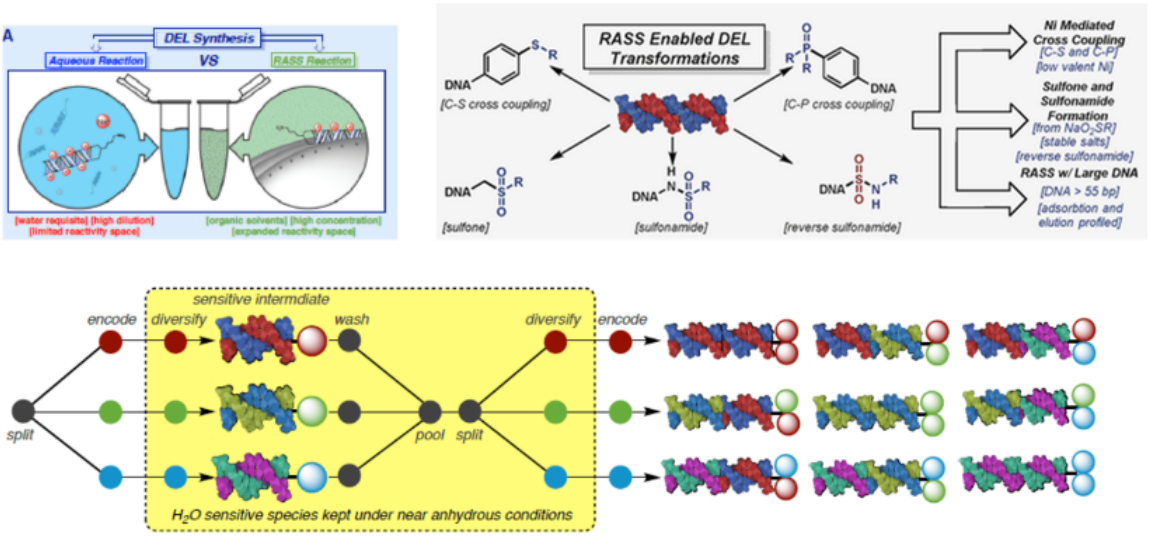


P. W. Cai; G. Y. Yang; L. Z. Zhao; J. Q. Wan; J. Li, G. S. Liu, *Org. Lett.* 2019, **21**, 6633

C3-Indole Functionalization Developed at HitGen

2.3.4 Beyond Aqueous Systems for On-DNA Chemistry Development

DNA-encoded libraries showed promise as a valuable technology for enabling the hit discovery process. Although there has been an increasing number of reports describing DNA-compatible transformations, the number of published chemical reactions on DNA is still quite limited, compared with traditional medicinal chemistry. Adapting organic reactions for use in dilute water presents many difficulties as many interesting bond forming reactions invoke water incompatible reagents or intermediates. Recently, Reversible Adsorption to Solid Support (RASS) have been demonstrated as promising methods to expand DEL reactivity and application scenarios (Angew. Chem. Int. Ed. 2020, 59, 19, 7377-7383). Many efforts have been dedicated into the above innovation at HitGen. DELT has already shown great promise in the generation of novel chemical leads, and has considerable scope to develop in terms of molecule types and the way these molecules are produced in the future.



On-DNA Chemistry Development on RASS

2.4. iDEL® Website for DEL Informatics and Protocols

What is iDEL® and Why iDEL®?

iDEL® is an online informatics platform aiming to promote the generation of innovative DEL ideas, interactions between customers and HitGen. Users can design, create DELs, share ideas, and retrieve DEL protocols on this platform. User's idea is protected as iDEL® is based on blockchain technology with good transparency and confidentiality, until the initiation of formal collaboration.

Service from iDEL®

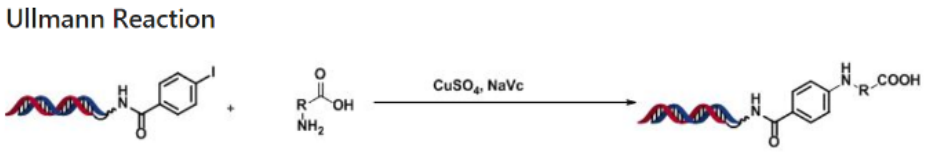
Online Retrieval of DEL Protocol

DEL Design, Interaction & Share

Project Cooperation

Online Retrieval of DEL Protocols (DEL Chemistry and Reaction)

iDEL® extracts the chemical reactions from DEL related publications and patents, and presents them in the module of DEL protocols. Users can easily retrieve information about DEL protocols from this platform and get updates on the pioneer work of DELs. Uses can easily extract the reactions and protocols by typing key words of reactions from the database.



Condition:

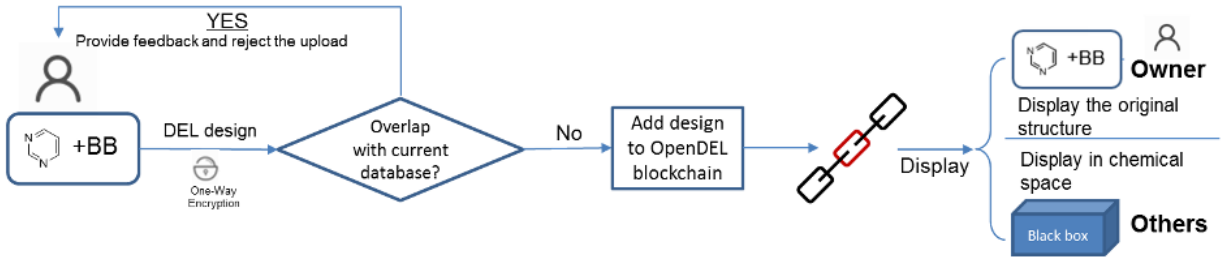
To a solution of DNA (1mM in 250mM BBS buffer pH 9.4, 1.0 eq.) was added the the solution of amino-acid(1M in 1M KOH aq., 1000eq) and was vortexed for seconds. The solution of CuSO4 (25mM in H2O, 4eq.) and NaVc(25mM in H2O, 8 eq) were added and then the...

Validation Summary :

(1) 71 acids(contain I, Br) were validated, 28 acids gave the yield >50%. (2) 304 amino acids were validated, only 32 α-amino acids gave the yield >30%.

DEL Design

The DEL design module offers free-accessed platform to users for the practice of library design. The design work by users is encrypted using blockchain technology which HitGen cannot directly access. System will automatically check if the design work by users overlaps with existing DEL design in the database. The design work by users will be stored in black box, and not disclosed to HitGen. If physicochemical properties and diversity of the newly designed DEL are attractive, HitGen may decide to unbox the DEL design idea. A smart contract will then be initiated. Consequently, users who design this DEL will receive a payment accordingly.

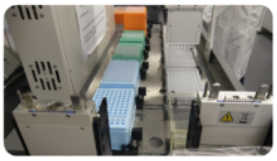
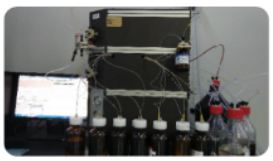
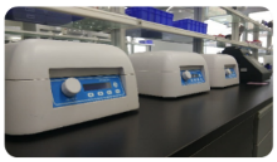
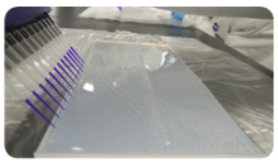


- iDEL® makes the DEL design accessible to all users who are interested.
- iDEL® protects user's intellectual properties via data encryption and smart contract.
- iDEL® provides a trading platform to allow secure exchange between DEL designers and HitGen.

3. DNA-Encoded Library Production

HitGen has built an integrated DEL platform enabling the design, synthesis and characterization of HitGen's proprietary DELs and custom DELs.

- Reliable in-house facilities for the full workflow of DNA-encoded library synthesis and quality control
- High-throughput DNA-encoded library synthesis and QC

Oligo Chemistry	DNA-encoded library chemistry	Encoding	QC and Purification
			
			

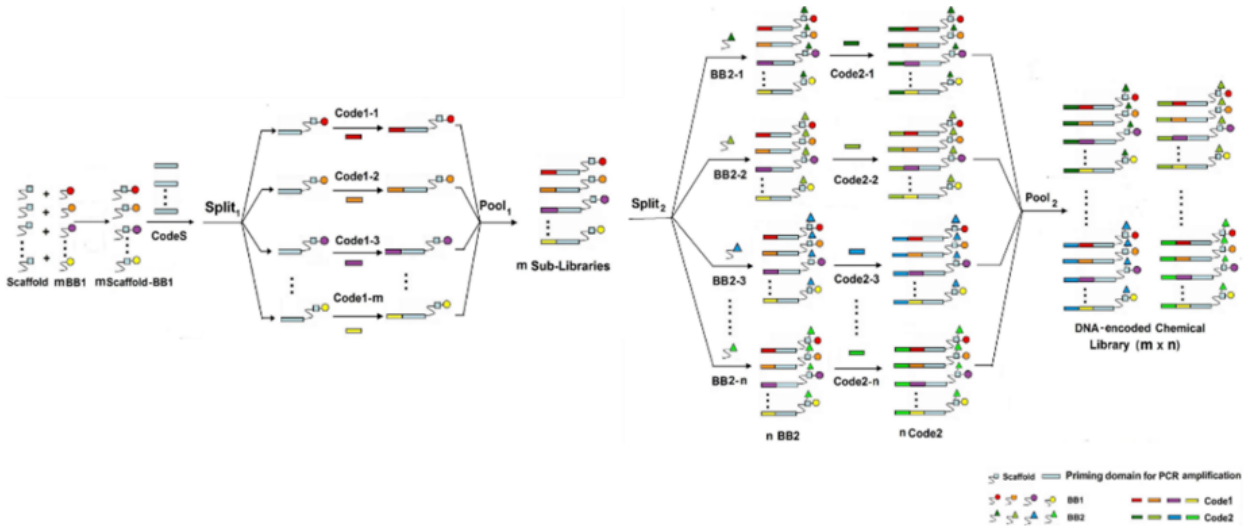
Integrated DEL Synthesis Platforms at HitGen

3.1. DEL Production Process and Quality Control

DEL allows the synthesis and screening of millions, or even billions, of encoded compounds in a cost and time efficient manner compared with using conventional methods. HitGen has developed a complete set of DEL production processes and quality control methods, which will ensure the DELs be constructed with higher efficiency and quality.

Production Process

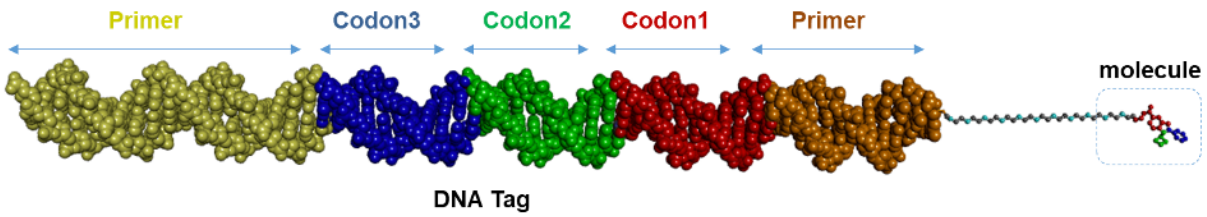
DEL technology comes from the merge of DNA encoding and combinatorial chemistry. A split-and-pool strategy involving iterative chemical synthesis and DNA encoding steps is currently the most widely used method for DECL preparation The typical DEL is assembled through a series of DNA encodings followed by chemical diversification. At the first step of the synthesis, chemical building blocks (BBs) are tagged with DNA barcodes. They are mixed together and then split into different portions for the next cycle of chemistry and ligation. A new set of building blocks is added to the chemical reaction, and then, corresponding DNA barcodes are attached and linked with previous DNA molecules. Thus, DNA encodes each step of chemical synthesis for each compound.



Scheme of Split&Pool Synthesis Strategy

Stringent and Multi-level Quality Control

In order to minimize the occurrence of false positives and to increase success rate of screening, it is necessary to ensure the quality of each library. The quality and process control including DNA tags, chemistry reactions and relationship between DNA tags and chemistry reactions are super important for final library quality.



DNA Tags

Quality of oligos:
- Identity
- Purity
- Impurities
- Quantification

Ligase:
- Activity
- Impurities

Encoding:
- Design of codon system
- Ligation efficiency
- Purification efficiency

Fidelity of library

Chemicals

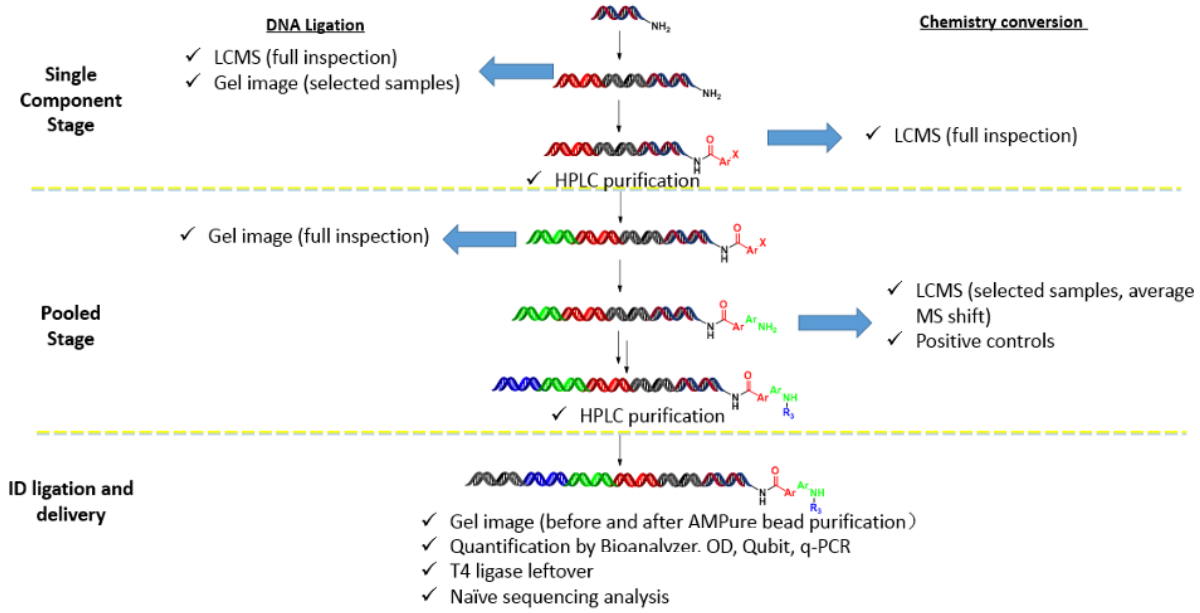
Quality of DEL Compounds:
- Drug-likeness
- Novelty
- Diversity

Quality of Chemicals:
- Identity
- Purity
- Salt form

Chemistry:
- Chemical reaction yields
- Identity of by-products
- Uniformity of chemistry results
- Chemistry modification or damage on DNAs

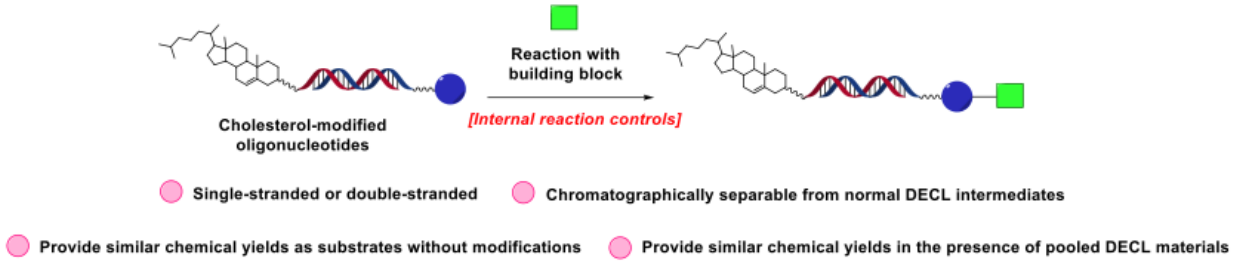
Quality Control of DEL Synthesis

During the process of DEL production, HitGen has applied quality control methods including UPLC-MS, HPLC, TOF-MS, electrophoresis and NGS to monitor chemistry or ligation reactions. The quality of final products are also further evaluated using our own quality control methods, such as electrophoresis, Qubit, OD, bioanalyzer and qPCR.



Key Quality Controls in DEL Synthesis

In addition to the conventional quality control methods mentioned above, to further improve DEL quality, a cholesterol-modified oligonucleotide is designed and developed as an internal probe by Janssen and HitGen team to evaluate the chemical transformation in the complex system in pooled stages (Cholesterol-Modified Oligonucleotides as Internal Reaction Controls during DNA-Encoded Chemical Library Synthesis; Bioconjugate Chemistry DOI: 10.1021/acs.bioconjchem.1c00045). As these cholesterol-tagged oligonucleotides are chromatographically separable from normal DEL intermediates, they can be directly monitored by mass spectrometry to track reaction progression within a complex pool of DNA.



Cholesterol-Tagged Oligonucleotides as an Internal Control in DEL Synthesis

Capability

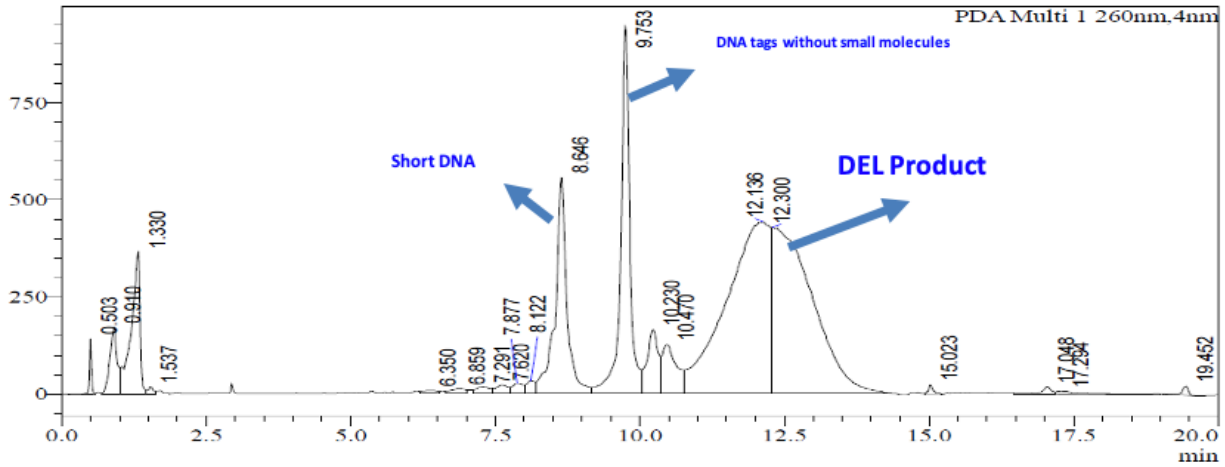
At HitGen, we have built the capacity to operate up to tens of thousands of splits in a single cycle. And it usually takes 1-3 months to deliver one library. With the implementation of automation for DEL production platform, the libraries are constructed with high efficiency and quality.

3.2. Analytical Capability for DEL Synthesis

HitGen has developed an integrated platform of analytical chemistry with advanced instruments, aiming to provide analytical services for building high quality libraries. It is built with a broad range of capabilities for the whole process of DEL production, including analysis on scaffold/building block synthesis, periodically building block QC check, codon QC, on-DNA chemistry development and characterizations during pool&split synthesis.

Analytical Capability for DEL Synthesis

Quality of DEL synthesis is analyzed and ensured at every stage in multiple approaches. As mentioned in the process and quality control of DEL production, MS analysis is one practical tool for monitoring the quality of DEL by analyzing the identity and quality of starting materials of building blocks and DNA tags, calculating the yield of DEL intermediates and products. Purification of starting materials, intermediates are performed with various analytical tools exhibited above to ensure their purities for DEL synthesis. Purification of DEL products is also executed to remove the unreacted materials without changing its chemical diversity for downstream analyses.



A Typical HPLC Chromatogram of a DEL Product with Unreacted Materials

3.3. Automation of DEL Synthesis

DNA-encoded library with high diversity usually requires large numbers of parallel reactions and split&pool operations. The efficiency and homogeneity of those operations will determine the physical quality of the DNA-encoded libraries. To ensure the library quality and free our scientists from tedious benchwork, HitGen has established an automatic synthesis platform for DNA-encoded library synthesis.

By working with several top vendors in automation and liquid handling, HitGen has customized the platform suitable for automatic, high-throughput and accurate DNA-encoded library synthesis operations including splitting, chemical reaction set-up, ligation reaction, pooling and precipitation which could be monitored online.

HitGen's automation platform is providing the support for both HitGen's DNA-encoded library synthesis and collaborators' library synthesis.

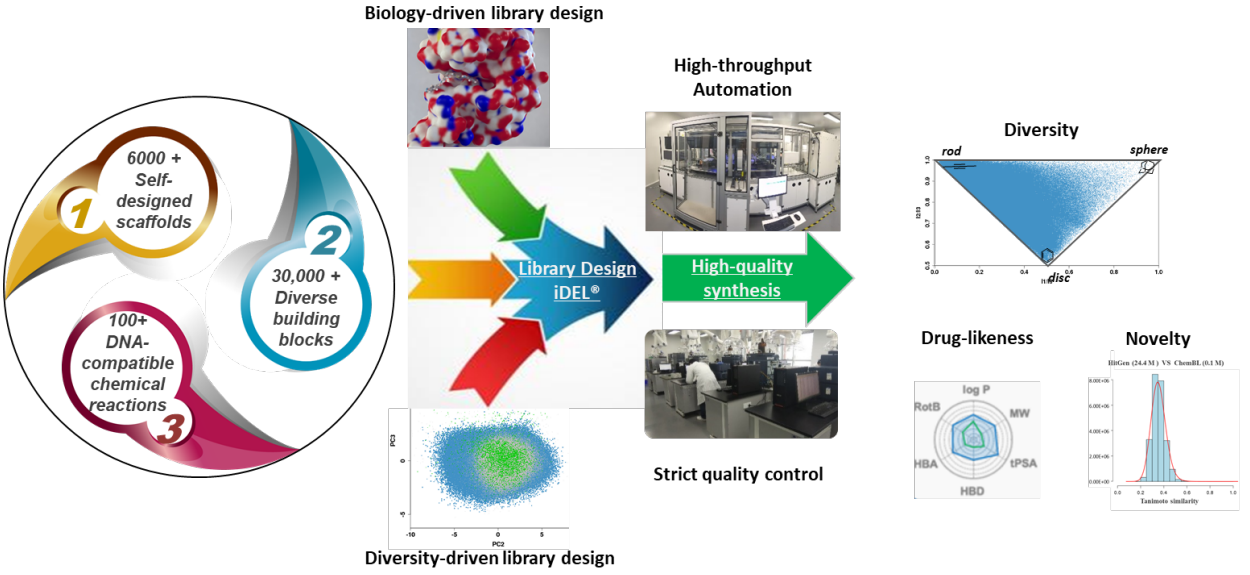
1. accurate and homogeneity aliquot for each sample

2. integrated systems with online monitoring and data recording using software and database

3. miniaturization of reaction systems with high efficiency

4. DEL Products and Enabling Tools

DNA-encoded chemical libraries (DEL) have evolved into a valuable platform for hit identification in early drug discovery. The DEL technology enables the fast and economic synthesis, screening and analysis of DNA-encoded collections of millions to billions of compounds against biological targets through a confluence of molecular biology, combinatorial chemistry, high throughput sequencing and advanced informatics techniques.



DEL Design and Synthesis at HitGen

Key Features of HitGen's DNA-Encoded Chemical Libraries:

- Over 30,000 reagents and 6,000 unique scaffolds to facilitate library design with novelty and diversity
- Applying cutting-edge chemistry and robust methodologies such as photoredox chemistry to library synthesis

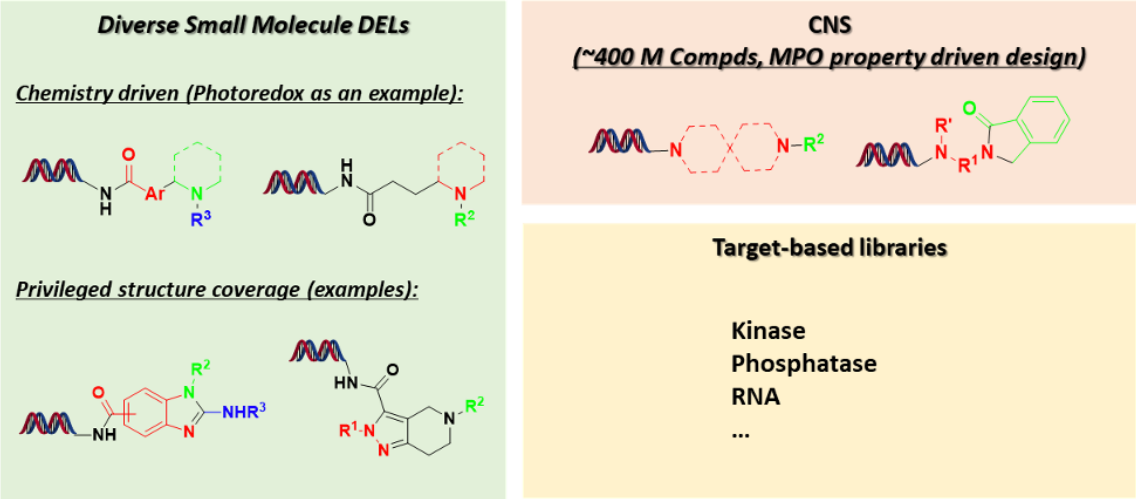
- Strict and multi-level quality control protocols and methodologies
- Over one trillion DNA-encoded compounds
- Diverse chemotypes with pharmacological interests, including small molecules, macrocycles, covalent molecules, protein degraders, encoded fragments, etc.
- Drug-likeness with well-controlled physicochemical properties
- Decent novelty compared with compound collections in ChemBL database
- Chemical structures with enhanced 3D shape

4.1. HitGen General DELs

The largest parts of HitGen's DNA-encoded libraries are diverse and druglike small molecules which represent super-large diverse chemical spaces. Except traditional diverse small molecules, HitGen team has also continuously made efforts on target-based library design and synthesis. Target-focused compound libraries are collections of compounds which are designed to interact with an individual protein target or, frequently, a family of related targets such as kinases and proteases. They are used for screening against therapeutic targets in order to find hit compounds that might be further developed into drugs. The design of such libraries generally utilizes structural information about the target or family of interest. Designing a library that is appropriate to the target or target types increases the likelihood of identifying better quality hits, which can reduce both timelines and overall cost of the drug-discovery process.

Small molecule libraries for CNS targets: HitGen specially designed CNS-focused libraries for discovery of new CNS drug candidates. The criteria for design of CNS library molecules which are able to penetrate blood-brain barrier (BBB) were low polar surface area, low degree of possible hydrogen bond formation and low ClogP values, etc.

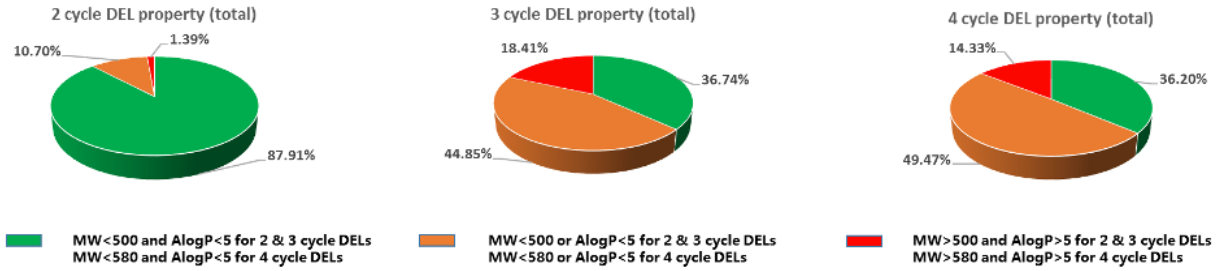
Small molecule libraries for specific biological targets: HitGen offers comprehensive support on discovery new hit compounds against specific target types by focused-target libraries.



HitGen Small Molecule DELs

4.1.1 Physicochemical Property of HitGen General DELs

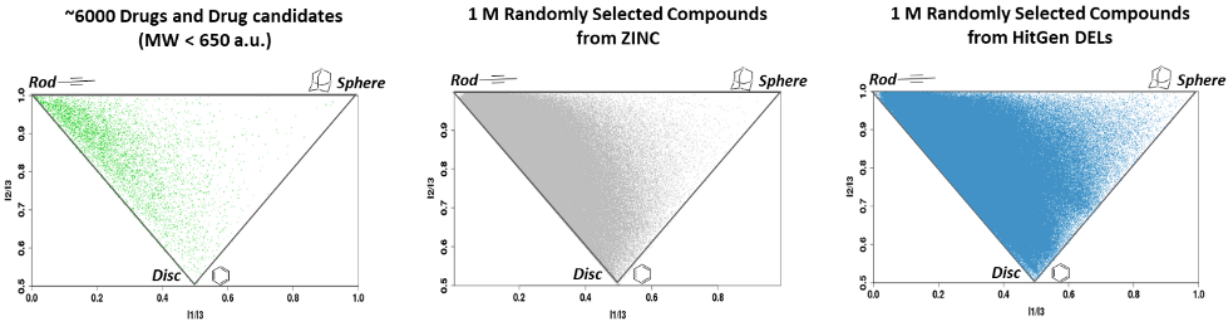
Compound physicochemical properties are known as key determinants for compound drug-likeness. Evaluation and control of physicochemical properties of library compounds are one of crucial decision-making points during library design and synthesis. Statically, 87.9% of molecules in 2-cycle libraries are in the range of MW<500 and AlogP<5, and 36.7% molecules of 3-cycle DELs are in the range of MW<500 and AlogP<5. By comparison, molecular weights of compounds in 4-cycle libraries are much higher than 2/3-cycle DELs, featuring 36.2% 4-cycle molecules are in the range of MW<580 and AlogP<5.



Physicochemical Property of HitGen General DELs

4.1.2 Topological Shape Analysis

The topology analysis provides characteristic values related to the topological structures of molecules. One million DNA-encoded compounds were randomly selected from HitGen DELs for topological shape analysis and compared with drugs, drug candidates and compounds in ZINC database. It's shown that HitGen's DNA-encoded compounds are more prone to sphere topology.



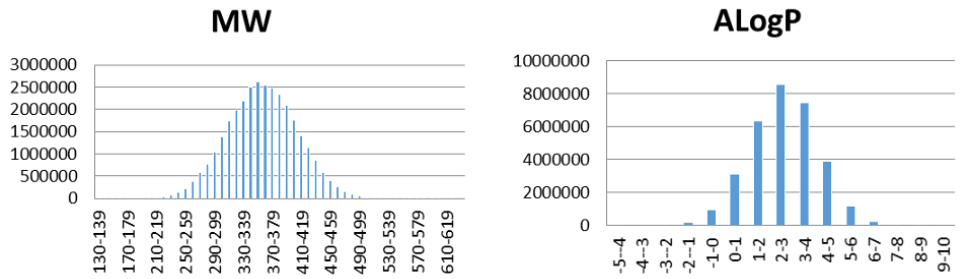
- Principal moment of inertia [1] calculations based on molecular topological shape
- Datasets include 1 million randomly selected structures respectively from HitGen DELs and ZINC, as well as approximately 6,000 drugs & drug candidates (M.W. < 650 a.u.).

Topological Shape Analysis

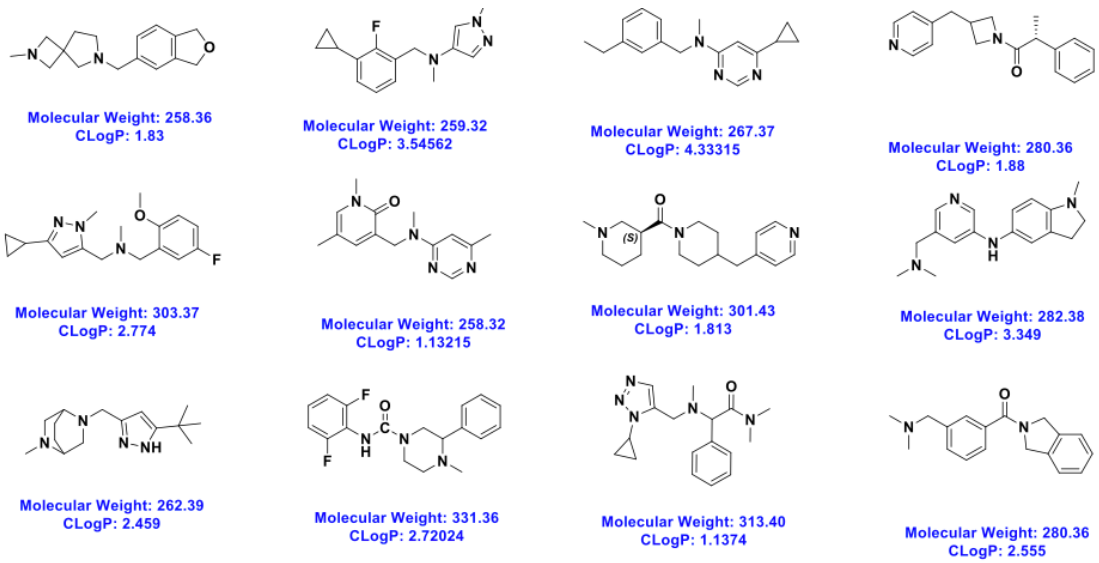
4.1.3 Small Molecules for CNS Drugs

Hundreds of millions of small molecules are designed and synthesized for discovery and development of CNS-focused biological targets. The libraries are aimed to facilitate the CNS

drug development projects and efforts in the search for novel neurotherapeutics. An average of molecular weight of chemicals in the library was around 350 dalton.



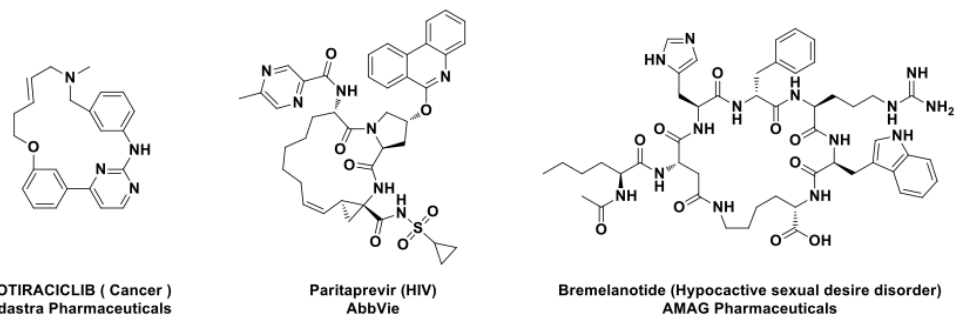
Distribution of Molecular Weight and AlogP of DEL Molecules Against CNS Targets



Representative Off-DNA Structures

4.1.4 Macrocyclic DELs

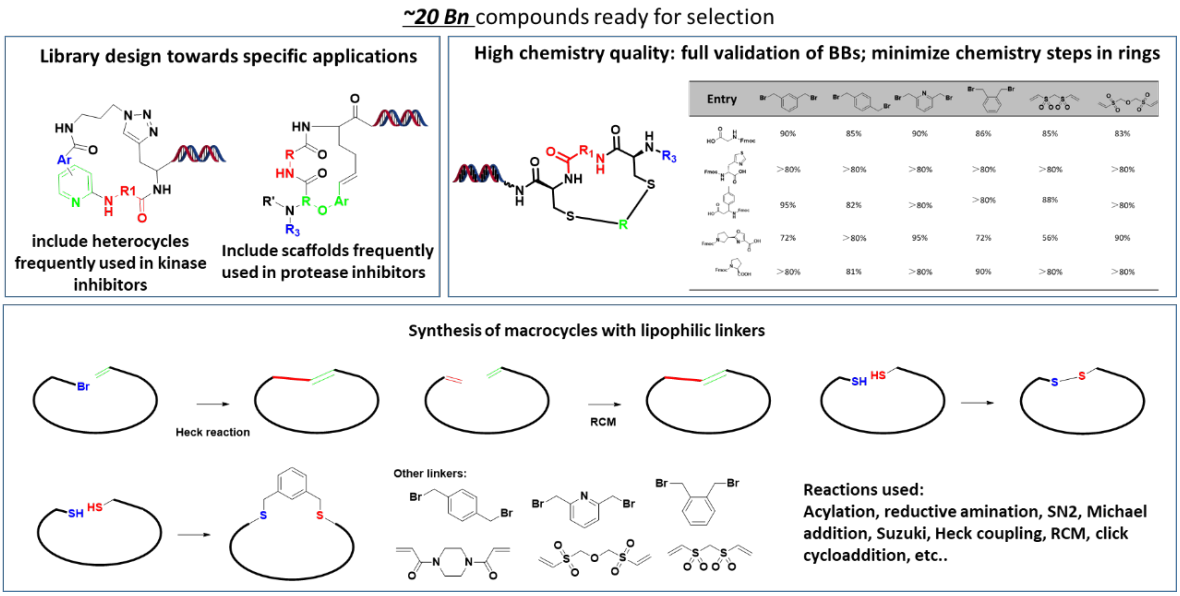
Macrocyclic compounds occupy a unique segment of chemical space and commonly found in natural products. In the last decade, this structural type has been successfully tested on many biological target classes, including both conventional targets (i.g kinase, protease) and some "challenging" targets with extended binding sites. (i.g. protein-protein interaction). Macrocyclic compounds can potentially boost both binding affinity and selectivity due to their size and complexity, which can engage targets through numerous and spatially distributed binding interactions. Furthermore, cyclization provides a degree of structural preorganization that can reduce the entropy cost of receptor binding compared to linear analogues.



Exemplified Macrocyclic Drugs

Our products and advantages

- More than 20 Bn macrocyclic compounds in stock.
- Critical considerations on macrocycle absorption, distribution, metabolism, excretion, and toxicity (ADMET).
- Reliable cyclization reactions validated at HitGen used in DEL construction.
- Good quality control during the DEL validation and production.



Examples of Ring Closure Strategies and Macrocyclic DELs at HitGen

HitGen has been attempting to build macrocylic libraries with good physicochemical properties, including controllable molecular weight, hydrogen bond donor and acceptor, etc.

Physical properties comparison between HitGen macrocycles and oral macrocyclic drugs in registered and clinical stages

Status	class	HBD	PSA	clogP	MW
Registered	Macrocycles	4 (3-5)	212 (200-225)	4.4 (3.0-5.8)	852 (805–901)
	Cyclic peptide	5	290	14.4	1203
	Macrolide	4 (3-4)	205 (194-215)	3.7 (2.7-4.6)	831 (798–864)
Clinical development	Macrocycles	3 (2-4)	171 (130-214)	6.9 (5.3-8.5)	775 (634–916)
	Cyclic peptide	5 (4-6)	277 (227-327)	11.3 (6.3-16.2)	1162 (946–1379)
	Macrolide	3 (2-3)	203 (190-215)	5.2 (1.2-9.1)	911 (755–1067)
HitGen	**All classes**	**4 (2-8)**	**210 (120-320)**	**2 (-5-5)**	**850 (500-1100)**

Physicochemical Property of Macrocyclic Drugs and Encoded Macrocyclic Molecules

4.2. HitGen Specialist DELs

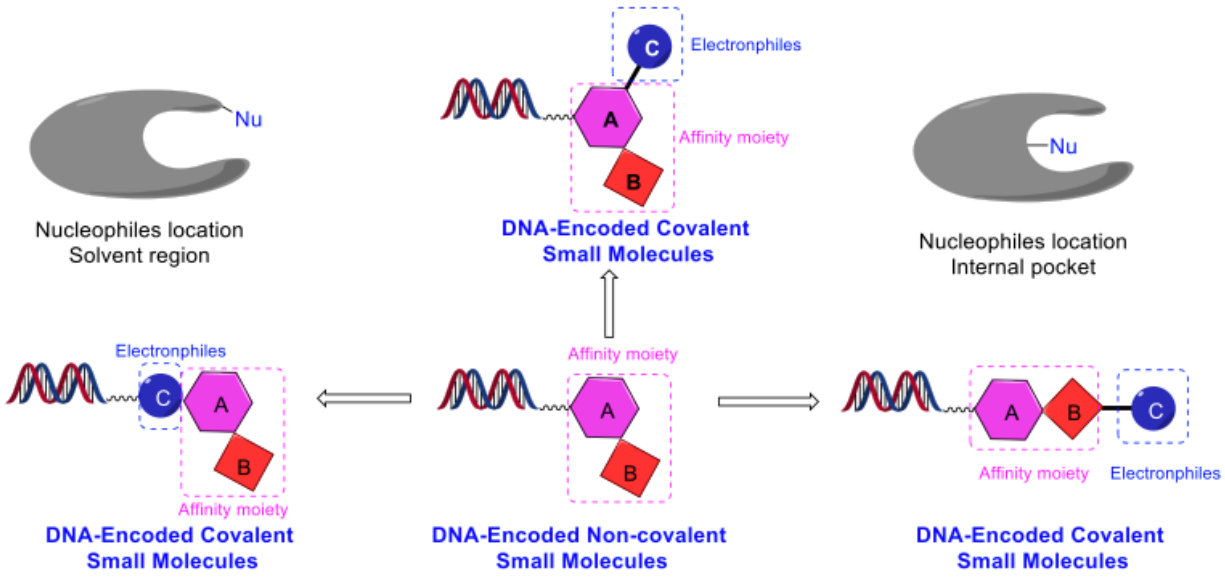
4.2.1 Covalent DELs

With the existing advantages of overcoming drug resistance, increasing the selectivity of the homogeneous target, and targeting the undruggable target, covalent inhibitors have made great achievements in the past decades. Encoded combinatorial library as a newly-emerging technology is admirable in identifying covalent binders in terms efficiency.

1. Over 40 covalent DELs and 15 Bn diverse DNA-encoded covalent molecules at HitGen
2. Majority of DNA-encoded covalent molecules are “Rule of 5” compliant
3. Over 100 warheads that may potentially target cysteine or lysine residues used in HitGen DELs

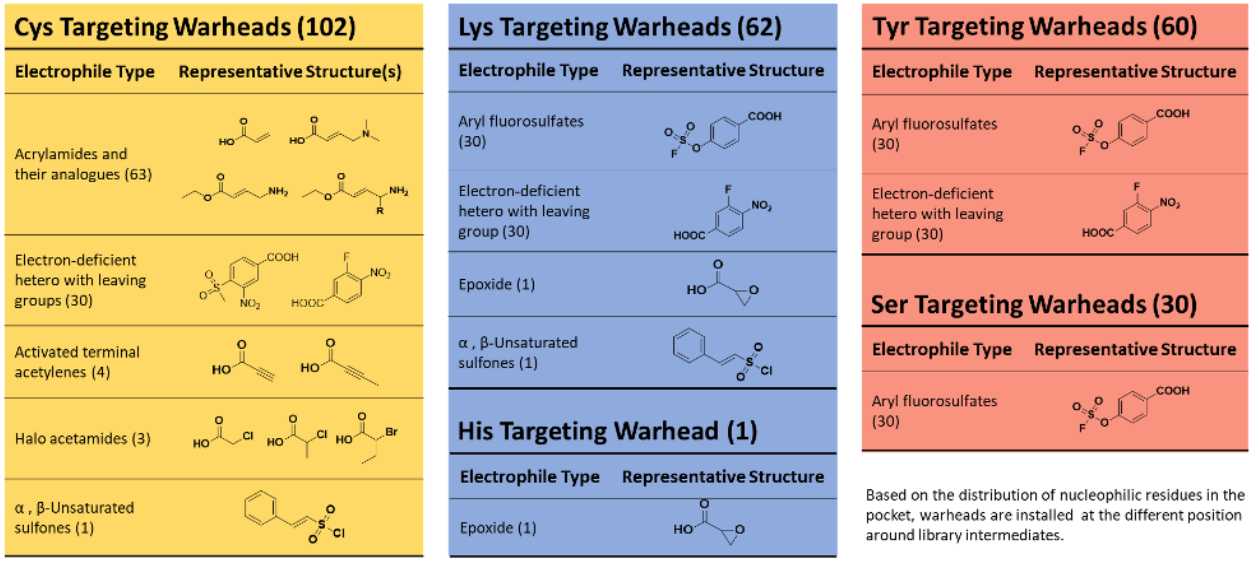
Library Design

Covalent binders in DELs are comprised of affinity moiety (Ki) and the warhead moiety (Kinactive). Diverse HitGen’s DEL intermediates and the covalent warheads collected from known covalent inhibitors were linked by covalent bonds via mild on-DNA chemical reactions. Considering nucleophilic residues are located at various positions of the binding pocket, three strategies for installing the warheads to the on-DNA intermediates were shown in below.



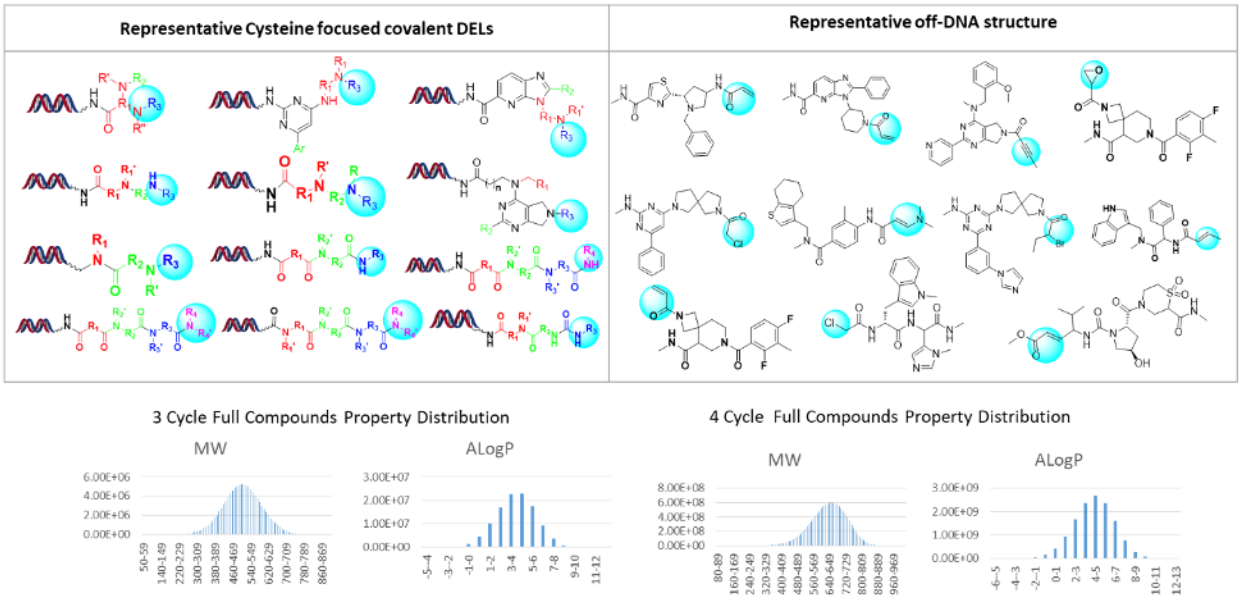
Covalent Library Design at HitGen

HitGen has developed a model system to evaluate the reactivities of DNA-tagged warheads against key amino acid residues, like Cys, Lys, etc.. With the protocols, hundreds of warheads are selected out for covalent DEL synthesis for increasing covalent library diversity.



Warheads Used in Covalent DELs

Representative covalent DEL generic structures, off-DNA structures and physicochemical property are shown below.

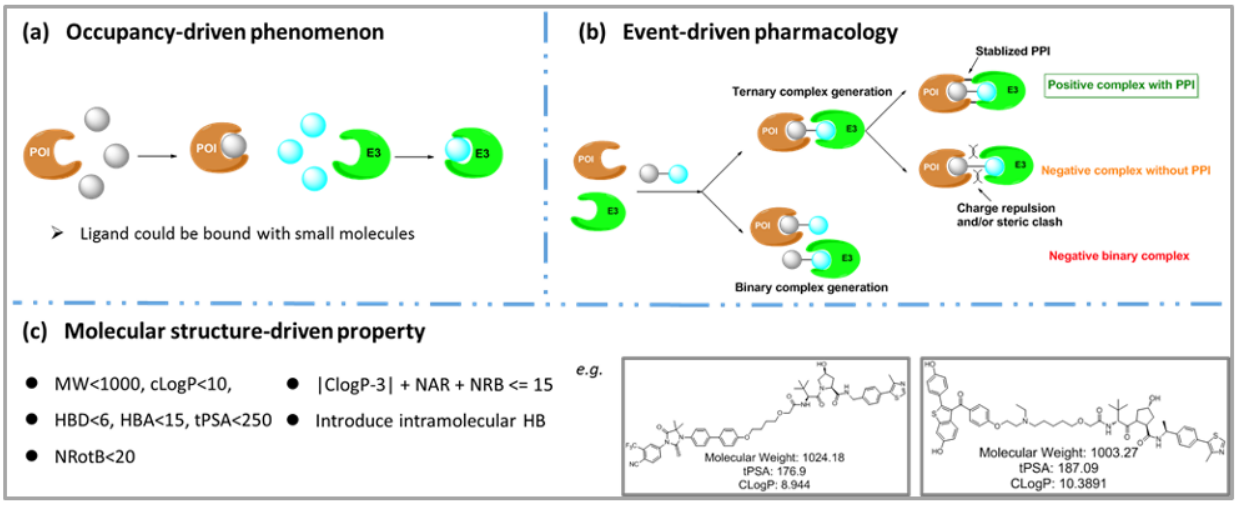


Exemplified Generic Structures, Off-DNA Structures and Physicochemical Property

4.2.2 DELs of Bi-specific Small Molecules_Protein Degraders as an Example

Multi-specific small molecules bind two or more different proteins and enable new therapeutic applications that cannot be replicated with conventional small molecules. Proteolysis targeting chimeras (PROTACs) are hetero-bifunctional molecules which incorporate a ligand for an intracellular target protein and an E3 ubiquitin ligase recruiting group, joined by a linker of a

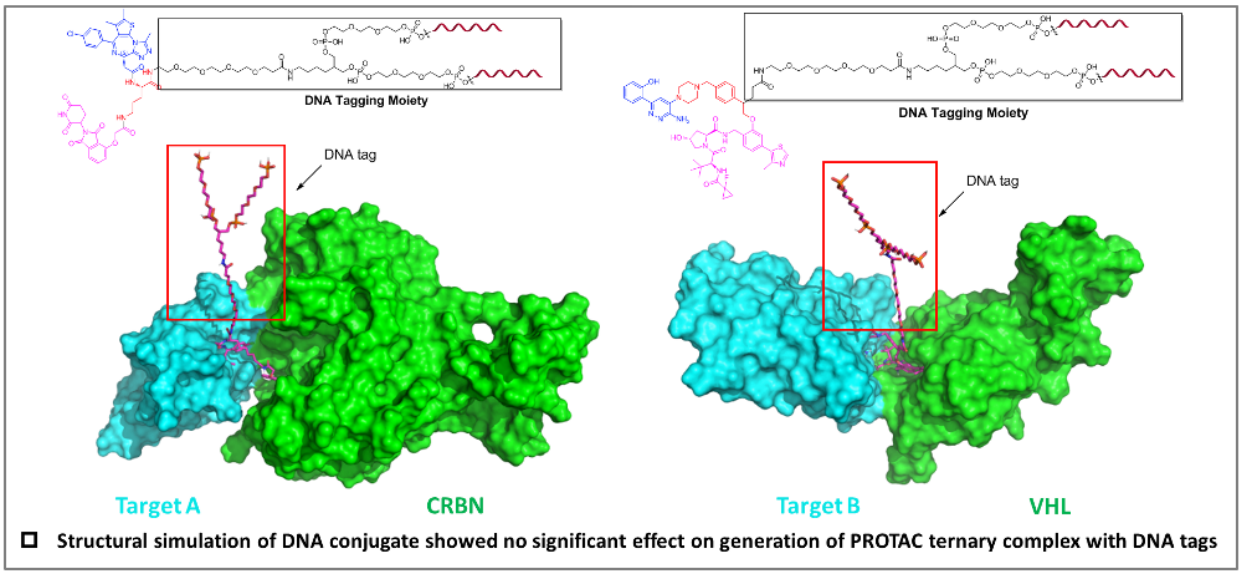
length appropriate to bring together target protein and ubiquitinating machinery and thereby elicit the ubiquitination of the protein of interest (POI) and its subsequent degradation in the proteasome. The methodology provides a powerful alternative approach to classical protein inhibition for therapeutic intervention for several reasons. Small molecule inhibitors need to achieve sustained target occupancy, typically requiring high systemic concentrations, while PROTACs offer a long-lasting effect by suppressing the target until resynthesis, which can take hours or days. While a classical inhibitor typically only blocks one function of a protein, the degradation of the protein perturbs all functions including allosteric regulatory sites and scaffolding/protein-protein interaction sites, which may lead to a more pronounced phenotype. In addition, since the PROTAC should be able to function when binding to any part of the target protein, the approach may provide additional opportunities to address less druggable proteins by allowing allosteric or even nonfunctional binding sites to be targeted. DELT is an ideal technology for the identification of protein degraders.



Proteolysis Targeting Chimeras (PROTACs) as Hetero-Bifunctional Molecules

DELT-Based PROTAC Library

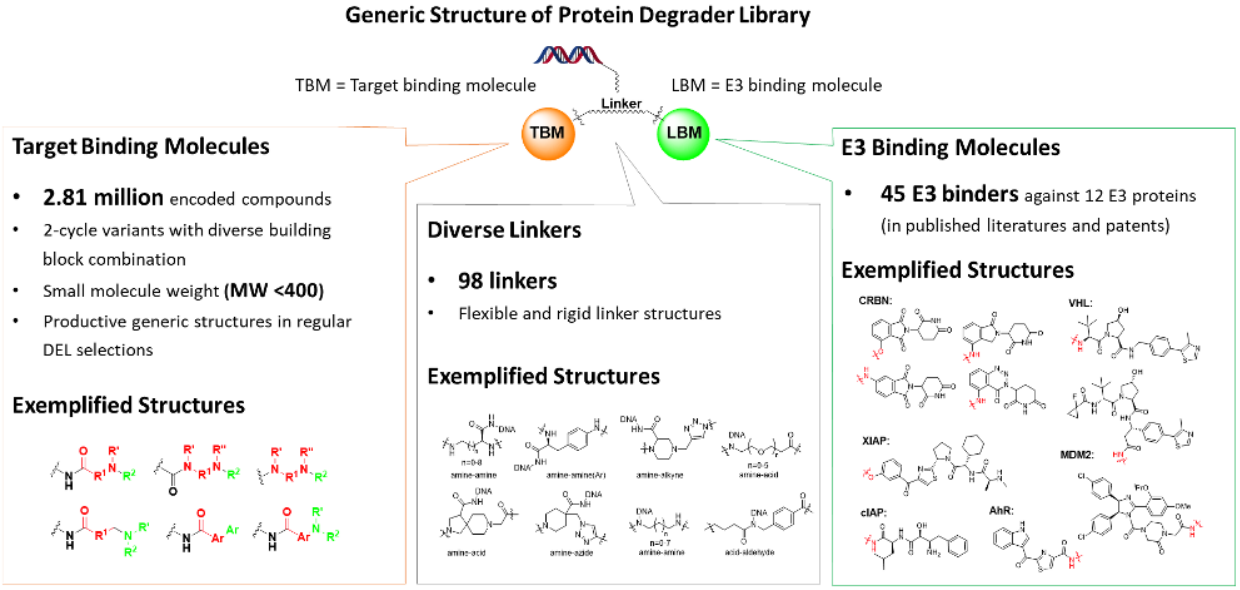
DELT-based PROTAC library could provide various strategy to find brand new PROTACs in different scenario. DEL technology with huge chemical space and molecular diversity will further accelerate the discovery of new PROTAC molecules.



DELT-Based PROTAC Library

HitGen PROTAC DELs

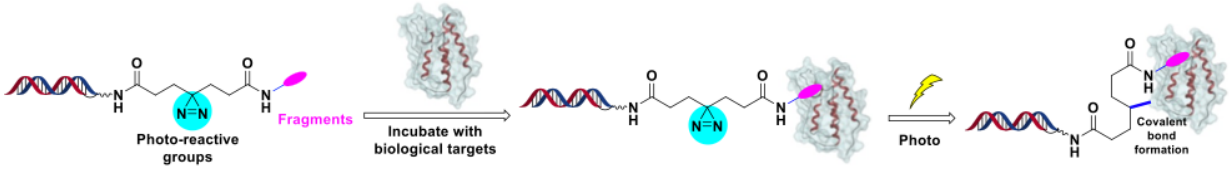
HitGen provides the PROTAC DELs to perform selections to discover novel POI ligand, novel E3 ligands, novel linkers and novel PROTAC molecules. Reported ligand and modified ligand of CRBN, VHL, MDM2, IAP, etc are also included in PROTAC DEL as E3 ligase binder moiety (LBM), combined with various types of linkers which is conducive to find new combination of E3 ligand or new type of PROTAC molecules.



4.2.3 Fragment DELs

Fragment-based drug discovery (FBDD) has become a mainstream and powerful technology for the identification of chemical hit matter in drug discovery. To date, FDA has approved 4 drugs and over 40 compounds are in clinical studies that can trace their origins to a fragment-based screening. Compared with the conventional HTS, using fragment library posess benefits including i) much greater efficiency of chemical spaces, ii) attractive physico-chemical properties and excellent starting point for lead series generation, iii) the lower costs of building, screening, and developing. Fragment DEL is based on DNA-encoded library technology, thus gains additional advantages:

- Small amounts of screened targets are needed
- No solubility concerns compared with traditional fragment screening
- DNA tagging points provide information of binding and subsequent fragment linking, growing and merging
- High-throughput affinity screening (datasets can be collected in 1-2 days for thousands of fragments)

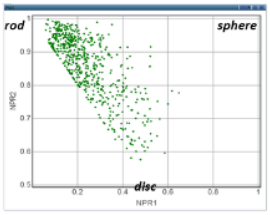


Principle of Fragment DEL at HitGen

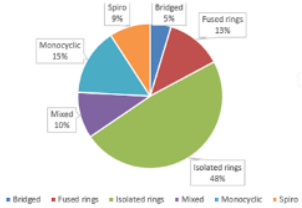
Fragment DEL developed with a careful selection of building blocks, which are widely commercially available. The key considerations include the “Rule of three” guideline, the removal of undesired functional groups (reactive, toxic, or ones prone to generate false positives) and the good fragment diversity.

Parameter	Range
MW	100 – 350
ALogP	≤ 3
HBD	≤ 3
HBA	≤ 4
RotB	≤ 4

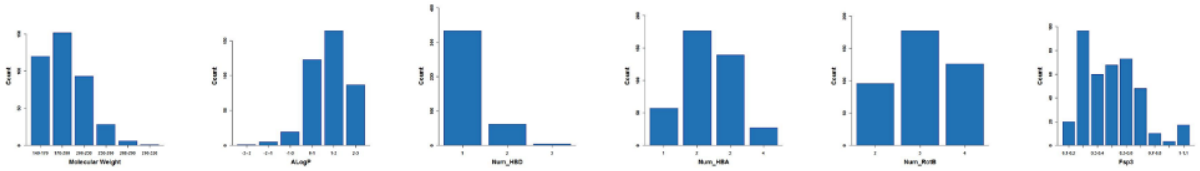
Physicochemical property control (**Ro3**) of fragments



Principal moment of inertia (PMI) plot of fragments



Ring systems of fragments

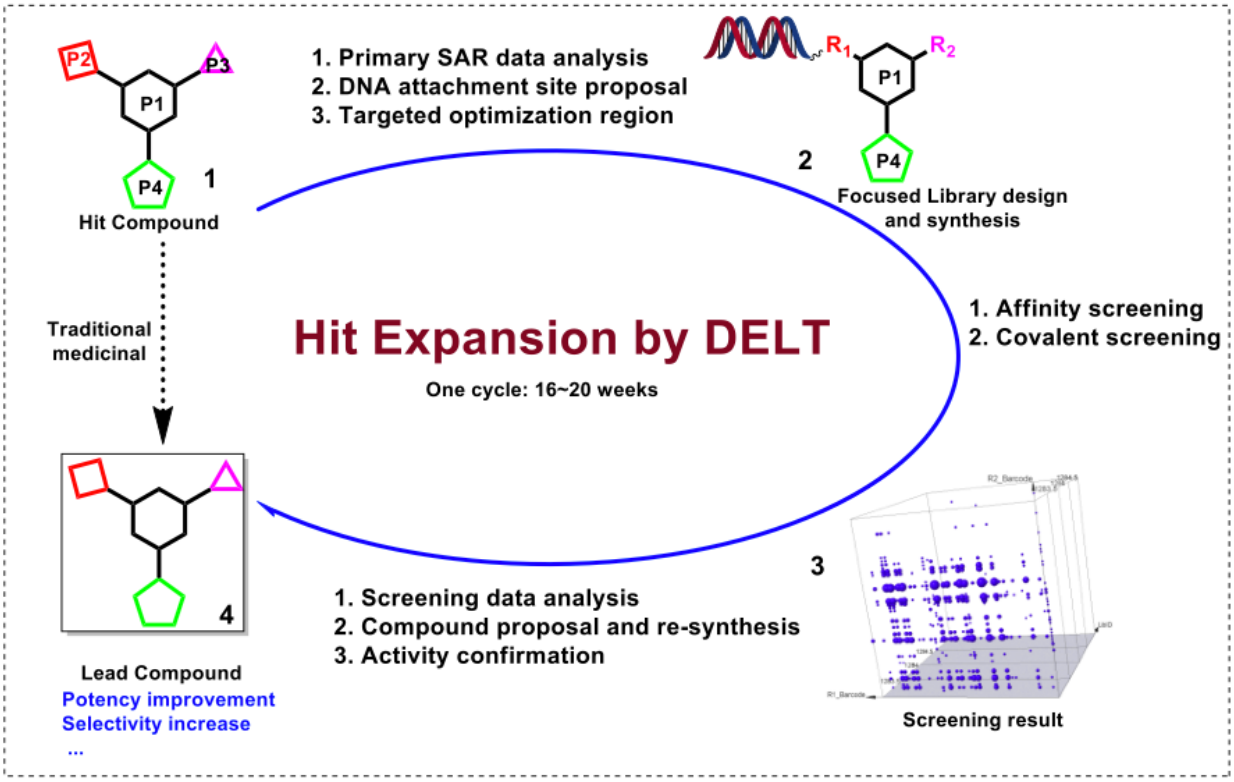


Physicochemical Property and Diversity of Encoded Fragments

4.2.4 Hit Expansion by DELT

Medicinal chemistry projects began adapting and using DNA-encoded library technology not only for the hit-discovery process to produce high-quality hits, but also for the lead optimization process. DELs with millions to billions of member around a hit or lead compound could be expanded to improve potency, selectivity or other drug-like properties.

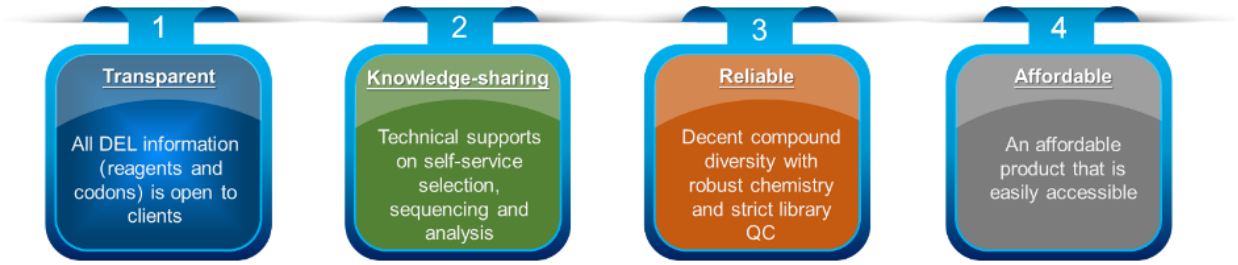
1. To obtain best fragment combination and cover broader chemical space
2. Focused DELs are designed and synthesized for specific target selection to optimize hit compounds in a timely manner
3. Focused DELs may provide rich SAR information to guide medchem optimization
4. The focused DELs can also be used for other homologous target screening



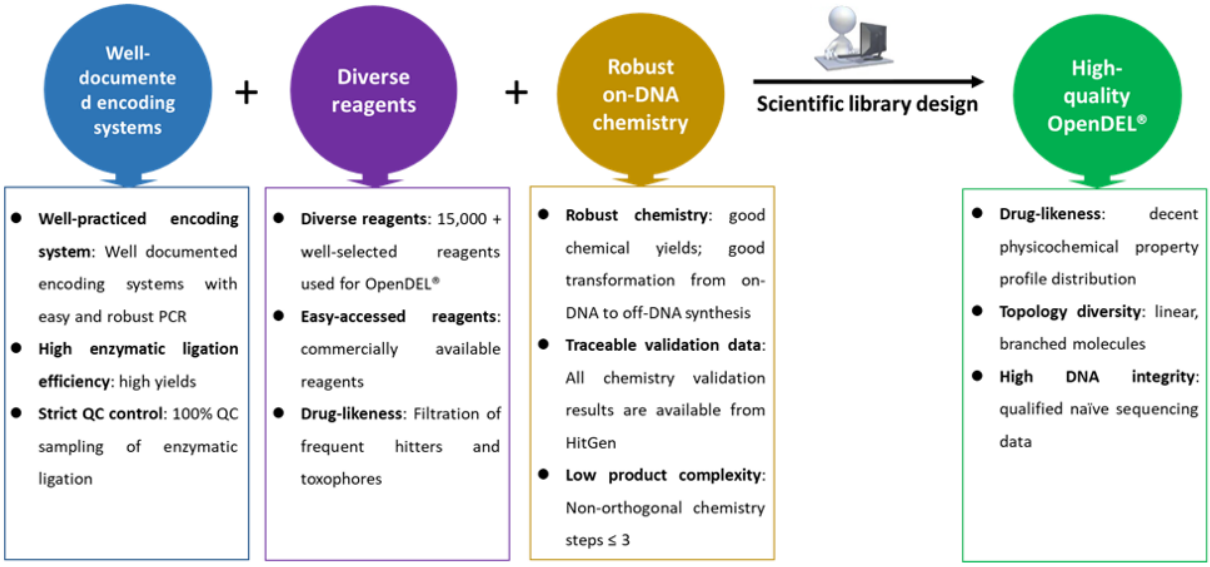
The Workflow of Hit Expansion by DELT

4.3. OpenDEL®

OpenDEL® as a self-serve product has previously been released by HitGen Inc. (HitGen) in 2015 (https://biopharmadealmakers. nature.com/documents/3606-hitgen-ltd-profile-jun-2015#sign-in-modal). After several years of evolution, it now contains small molecule DELs with high diversity and drug-like space, and helps to increase the possibility of finding potential hits in an efficient and cost-effective manner (see key features of openDEL® as shown below).



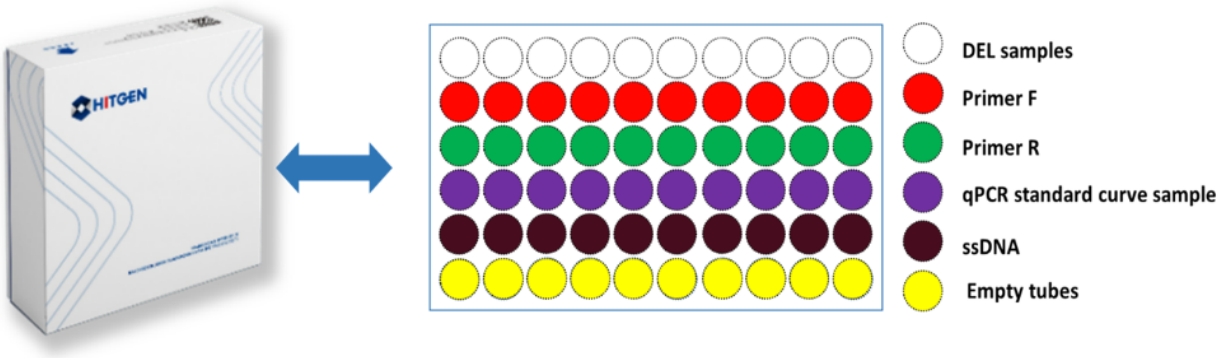
Key Benefits and Values of OpenDEL®



Criteria of OpenDEL® Products

What are included in OpenDEL®?

OpenDEL® contains DEL libraries for affinity selection and other materials necessary for DEL selections, including PCR forward primer (Primer F), PCR reverse primer (Primer R), qPCR standard curve sample, Salmon sperm DNA (ssDNA) and empty tubes. Detailed protocols for self-service selection are available and accessible.



HitGen's OpenDEL®

How does OpenDEL® work?

Target selection

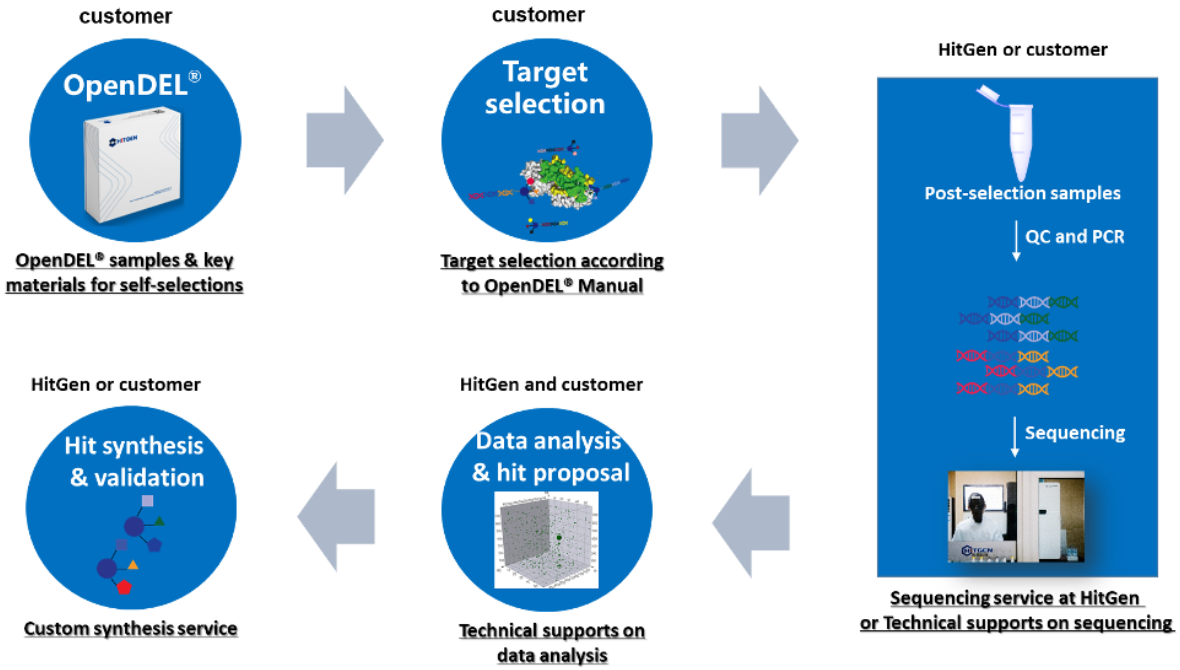
Researchers will determine the details of the affinity selection protocols, using their deep knowledge about their targets. HitGen provides the screening protocols and consultation services for free.

Hit generation

The post-selection samples can be decoded by NGS at HitGen or be decoded by Customers. HitGen will provides necessary technical supports.

Hit validation

The proposed potential hits through data analysis could be re-synthesized on-DNA for further confirmation of the binding ligands or off-DNA for activity validation via appropriate biochemical, biophysical or cellular assays.



4.4. DEL$_{EZ}$™- Makes DEL Synthesis Easier

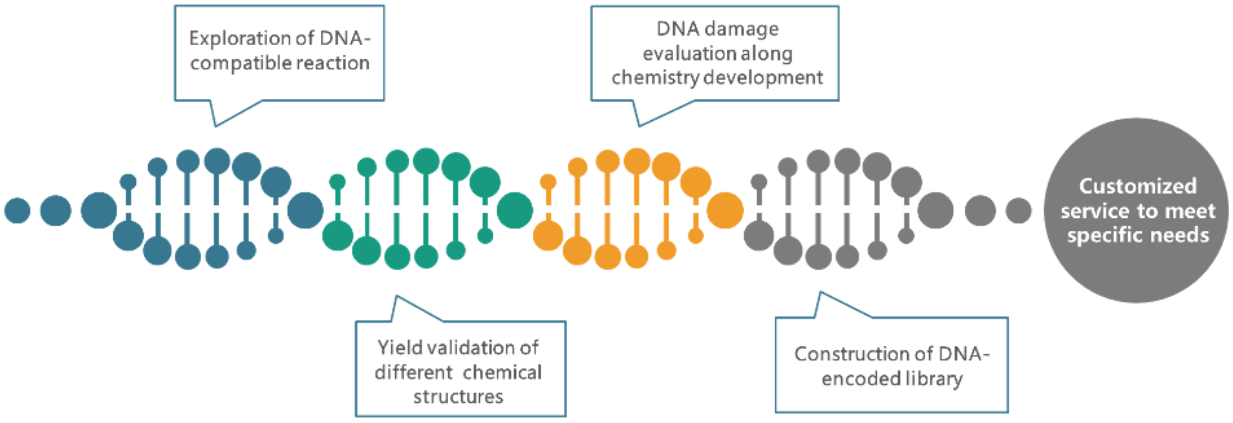
DNA-encoded library technology (DELT) has attracted more and more interests from both academia and industry in drug discovery due to its great potentiality in providing useful chemical starting points for medicinal chemistry.

In terms of DEL technology, the first issue comes to our mind is how to get a DEL containing structures with novelty, diversity and druglikeness. Although it has been widey acknowledged that DELs come from the merge of DNA encoding and combinatorial chemistry, where the most popular method to build DEL is “split and pool” process, there are still some factors impeding

most researchers from achieving DELs, such as the limited number and diversity of the building blocks or scaffolds used for the DEL construction, narrow range of chemistry compatible with DNA, lack of efficient quality control method in the whole process of DEL construction as well as quality inconsistency of raw material for both biological and chemical reaction.

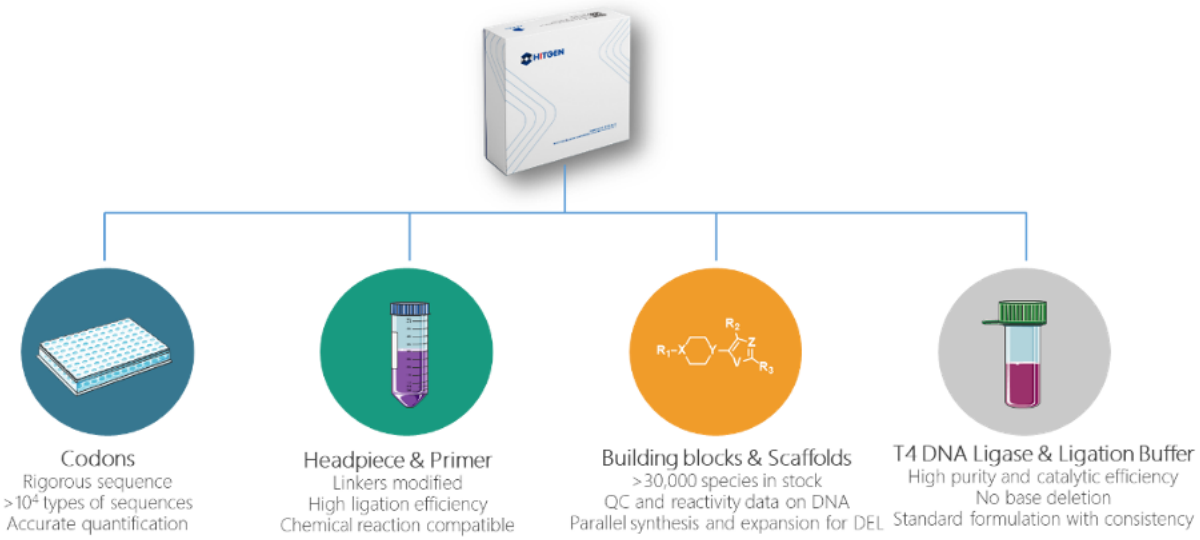
To lower the barrier of entry for DNA-encoded library chemistry and to facilitate the applications of DELT, DEL$_{EZ}$™, a DNA encoded library synthesis package is now available from HitGen. DEL$_{EZ}$™ is a supplementary product to typical DEL services and gives researchers an affordable and efficient solution to access DEL technology with HitGen's expertise on DNA-encoded library synthesis.

DEL$_{EZ}$™ enables researchers to develop their own chemistry for DELs and to assemble their own DELs based on their own expertise, experience in medicinal chemistry with the tried and tested supporting techniques, tools and resources in a timely and cost-effective manner. Having DELs containing desired structures ready, researchers can select potential ligands against biological targets with pharmaceutical interest.



Applications of DEL$_{EZ}$™ in the Whole Process of DEL Construction

Through the DEL$_{EZ}$™ service package, HitGen will provide tailored toolkits according to our in-house developed techniques, resources (Oligonucleotide Synthesis Platform, Protein Manufacturing Platform and Organic Synthesis Platform) and experience accumulated over many years to meet specific needs, and a detailed instruction manual for those entering this field recently.



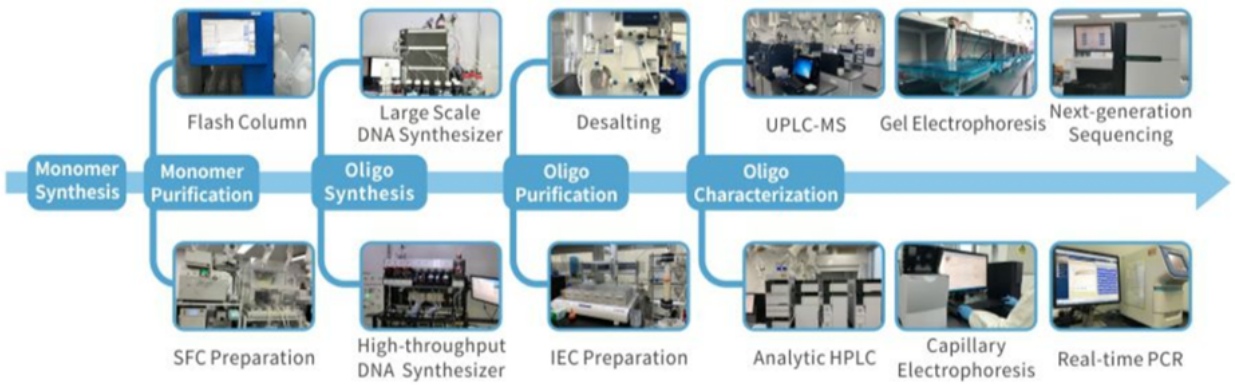
The Contents of DEL$_{EZ}$™ Package

Oligonucleotide Synthesis Platform at HitGen

HitGen has built an integrated oligonucleotide synthesis platform to provide DEL oligos and high-quality oligonucleotides, including headpiece, primers, codons, closing primers and customized oligonucleotides, small molecule-DNA conjugates.

The strict quality control system includes but not limited to high-throughput NGS to guarantee more than 95% perfect match ratio of each sequence, and advanced UPLC-MS analysis combined with HPLC purification to ensure more than 95% purity.

All the common primer products are in stock with large amount and ready to deliver.



Synthesis, Purification and Characterization System

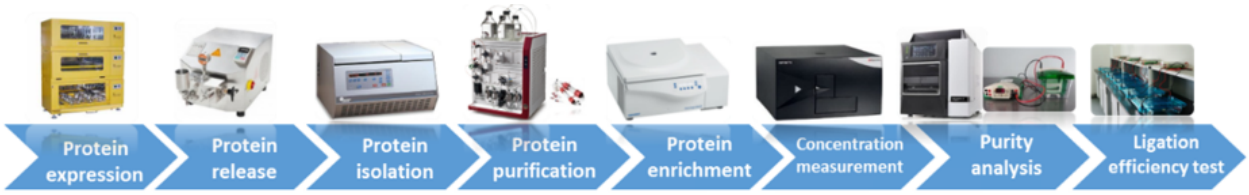
Protein Manufacturing Platform at HitGen

T4 DNA ligase as a “glue” in the construction of DNA encoded library could assist uniquely sequenced DNA to precisely label each chemical compound at a molecular level.

Protein manufacturing platform at HitGen could provide partners T4 DNA ligase with high purity,

stability and ligation efficiency.

The output of T4 DNA ligase could also completely satisfy the needs of DEL synthesis at HitGen and partner's site. It is in-stock with large amount in the long-term and ready to deliver.



Protein Manufacturing Platform

The advantages of DEL$_{EZ}$™

a) Scalability: flexible split number and reaction scale based on the diversity and large stock of codon and building blocks;

b) High quality: strict quality control system and practical application promise the consistency and stability of the product;

c) Time-saving: comprehensive instruction and technical support effectively avoid unnecessary experimental exploration;

d) Cost-effectiveness: one-stop service of raw material saves purchasing cost and time;

Product features:

- Flexible package to meet different needs;
- Minimized input and fast production cycle;
- Detailed procedures that are easy to follow and accessible to most labs.

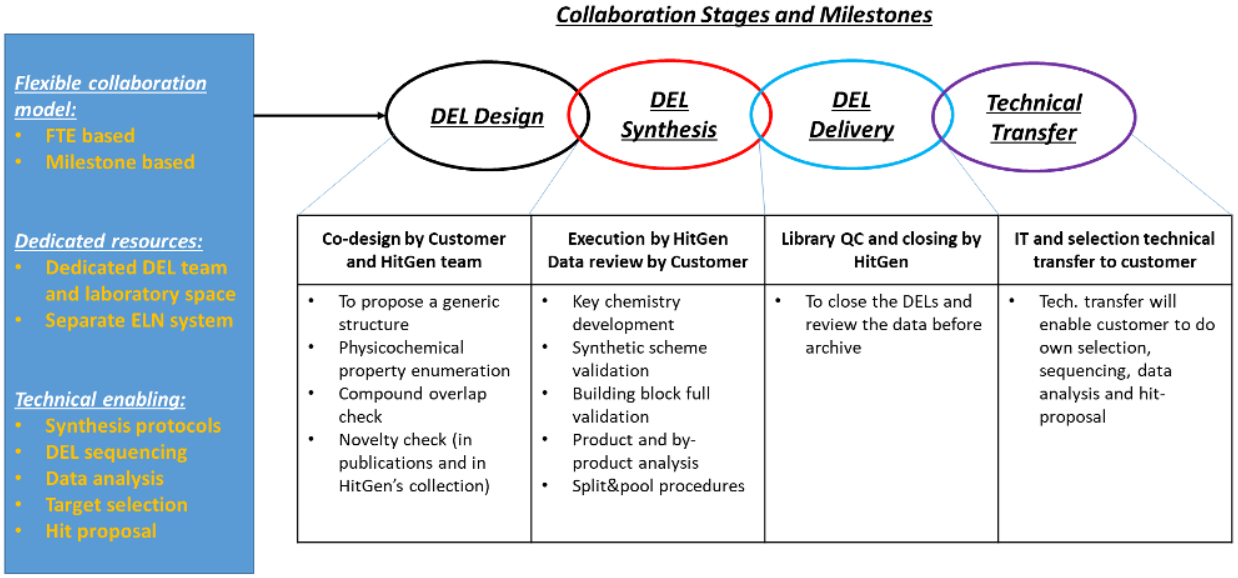
4.5. Custom DELs

With HitGen's expertise in library design, on-DNA chemistry development and library synthesis, HitGen offers specialized custom DEL synthesis services tailored to our customer's requirements. HitGen uses its technique, protocol, know-how and resource to build and deliver high quality libraries to customers. Four partners (Pfizer, JnJ, MSD and FORMA) were publically announced in the collaboration of custom DEL design and synthesis.

Business Model and Work Scope of Custom DELs

- HitGen uses its technique, protocol, know-how and resource to build and deliver high quality libraries for customers.
- Libraries generated in this collaboration can be selected against any Customer targets.

- One synthesis campaign will provide samples for hundreds or thousands of affinity selection experiments
- All the library information including synthesis procedures and QC data will be transferred along with library samples.
- Tech. transfer will enable customer to do own selection, sequencing, data analysis and hit-proposal.



5. DEL Selection Methodology and Strategies

Hit Identification using DNA Encoded Library (DEL) screening is a fast-growing and gradually mature technology. It has many advantages over traditional approaches: 1) easily accessing trillions of physically existing compounds; 2) saving in protein/target consumption; 3) short duration in hit identification and validation; 4) more cost effective.

DEL screening is based on the molecular interaction (a biophysical process) between DEL compounds and target, where compounds with more engagement to the target will be selected, therefore DEL screening process should be precisely called DEL selection.

5.1. DEL Selection Principle and Strategies for Finding Functional Hits

DEL Selection Principle

DEL selection uses DNA barcode as a tracking system to report which molecules are interacting with the target and in what degree. The readout of DEL screening is the corresponding DNA sequences and abundancy, assuming the input of all the compounds are the same or at least in a comparable number (e.g. 10^5 copies as the starting point). The abundancy of a given species is driven by either binding affinity or off-rate of the compound. DEL compounds can be recognized

by their DNA tags in a non-physically separated system, in other words, all the DEL compounds can be selected in one single vial. (every compound in traditional HTS requires a single well during screening)

In DEL selection, target concentration is much higher (μM range) to drive the equilibrium towards the formation of Target-Compound complex, the direct readout is the identity and abundancy of Target-Compound complex, rather than the compound concentration (see equation below), therefore much less compound concentration is needed (10^5 per compound or lower). DNA sequencing can precisely distinguish every single compound by the DNA tag and the throughput of 2nd generation of DNA sequencing can be 1 billion copies or more; therefore, DEL selection approach allows billions to trillions of compounds are evaluated in a single vial.

$$\text{Target} + \text{Compound} \underset{\text{Kd}}{\overset{\text{Ka}}{\rightleftharpoons}} \text{Target-Compound}$$

Besides driving the equilibrium to complex formation, the excess amount of target allows its interaction with all the possible compounds simultaneously, making DEL selection extremely efficient in terms of less total target consumption, less compound consumption and shorter screening duration.

Strategies of DEL Selection for Finding Functional Hits

DEL selection identifies binders which may not necessarily result desired functional hits. However, when DEL selection includes multiple samples, it is possible to distinguish compounds just binding to the target or having potential function(s). For example, the identification of ROCK2 inhibitor uses the following experiment design (also called Selection Plan):

No.	Target	Library	Supplement	Purpose
1	His-ROCK2	HitGen DEL	-	Identify all the possible binders
2	His-ROCK2	HitGen DEL	ROCK2 Inhibitor*	Identify all the binders with active site blocked by ROCK2 inhibitor
3	No Protein	HitGen DEL	-	Identify all the beads binders

*Excess amount of ROCK2 inhibitors are used to make sure that ROCK2 active site is saturated.

For Sample 1, all the possible binders are identified; For Sample 2, it identifies binders that interact with binding pockets other than active site. The subtraction of Sample 1 binders to Sample 2 will result compounds interacting with ROCK2 active site; whereas the overlapped compounds possibly are allosteric functional binders or non-functional binders. Sample 3 is to remove the binders interacting with beads.

Interrogation of Mechanism of Action in DEL Selection

Different from traditional high throughput screening, DEL selection provides the capability

of differentiating binding modes of hits, extending the triage of hits from affinity/potency to binding modes and mechanisms of action. For target of pharmacological interests, binding sites/modes differentiation can be achieved by incorporating binders/substrates/cofactors or including protein constructs with different domains in the Selection Plan. For enzymes with multi-substrates or special catalytic mechanisms, such as bi-bi compulsory ordered or Ping-Pong reaction mechanism, addition of substrates in the selection plan offers the opportunity of identifying binders to enzyme intermediates representing a distinct structure. In addition, counter targets can be utilized in the DEL selection to choose selective compounds and save downstream efforts in making and testing compounds for selectivity. This is especially useful for kinase targets as most ATP site binders lack selectivity among kinase family, where a potent inhibitor could hit many other kinases in the kinase panel.

5.2. DEL Selection Target Requirements

DEL Selection for Purified Proteins

- DEL selection targets are usually recombinant proteins with tags (tag is used to immobilize protein onto solid support for separating from compound mixture and washing).
- His, biotin, Strep, GST, FLAG and Fc tags are all compatible (the first 3 tags or dual/multiple tags are preferred).
- For endogenous protein extracted from biological samples, biotinylation by commercial kit is recommended.
- The tag in the protein construct or chemical biotinylation should not impair target activity and functionality.
- Commercial carboxylic acid or CNBr-activated beads can be used for direct immobilization of protein without tags, but it should be noted that the immobilization efficiency is dependent on unique tag-free proteins and loss of activity may still occur.
- Recommended protein purity is >90% with a clear determination of the multimeric status.
- The required protein amount is dependent on the selection group design. For a typical 50 KD protein with 6 groups in selection plan, 1 mg is sufficient.
- The protein can be provided by collaborators, purchased from commercial sources, or prepared by HitGen.

DEL Selection for DNA-binding Proteins

A frequently asked question is whether DNA binding proteins can be used in DEL selection, and the short answer is yes. HitGen can distinguish compound binding and DEL DNA binding during the process of data analysis. Strategies like supplementing DNA blocking sequences are also useful. On the other hand, mutant proteins with lower DNA binding affinity or excluding the DNA binding domain in the construct are also effective options. Transcription factors, DNA polymerase, DNA/RNA helicase have all been successfully screened at HitGen. As long as the target is not degrading DNA, it`s compatible with DEL selection.

DEL Selection for Receptors on Overexpressed Cells

For multi-transmembrane targets like GPCRs, ion channels or transporters, recombinant expression is of great challenge because the natural cellular environment is required to maintain proper protein conformation. However, DEL selection is not limited to purified proteins, and selection with live cells expands the utility of DEL for challenging targets that can`t be purified in their active forms. For targets expressed on the cell membrane, the direct interaction of DEL molecules and target protein is accessible. Transfection of DEL into the cells and employment of cell penetration peptide (J Am Chem Soc. 2019 Oct 30;141(43):17057-17061) provide strategies for intracellular target DEL screening. Cell lysate-based DEL selection is another option, where a tag designed in the target or a corresponding antibody can be utilized for immobilization of targets.

In cell-based DEL selection, high target expression level and proper negative control cells are required. For targets which are overexpressed, control cells with empty vector are recommended. For the targets having high endogenous expression level, knock-down or knock-out cells are proper negative controls. DEL selection is a binding process driven by the concentration of target protein, and higher expression level could facilitate the enrichment of signals. An absolute quantification of the target protein on cells can sometimes be tricky, but a careful assessment of the expression density is required to ensure successful enrichment of target binders. A direct comparison of expression level and enrichment ratio in cell-based DEL selection was reported (ACS Comb Sci. 2015 Dec 14;17(12):722-31). Selection for cell membrane targets without overexpression or high endogenous expression level is reported by introducing a complementary DNA to the target to increase the interaction between target and DEL molecules (Nat Chem. 2021 Jan;13(1):77-88).

5.3.DEL Selection Compound Input, Endpoint and Process Control

DEL Input

DEL input refers to the copy number of individual DEL molecule used in the beginning of the first round of selection. As the DEL technology gradually matures over the past decades, the size of DEL(s) is constantly increasing, reaching trillions for HitGen DELs. DEL selection requires a small volume of incubation and therefore limits the copy numbers of individual DEL molecule for solution viscosity reason. In the literature, 10^5, 10^7 copies were used in most DEL selection cases, yet as low as hundreds of copies have also been reported. We systematically compared the DEL input and selection performance and recommend to use no less than 3.3×10^4 in DEL selection. Please refer to our publication for more details (SLAS Discov. 2020 Jun;25(5):523-529).

DEL Selection Endpoint

For classical DEL selection, it is a multi-round process with the output of previous round as input for subsequent round(s). For each round, the DEL molecules with binding affinity to target protein stay and non-binders are washed off, therefore, signal-to-noise ratio of binders increases with progressive rounds of selection. Three rounds of selection are often able to achieve good signal-to-noise ratio and desired number of DEL compounds for sequencing. With multi-billion to trillion size DELs, the selection endpoint is usually set as the total copy number of DEL molecules reaching 10^7~10^9, the magnitude that 2nd generation DNA sequencer can easily interrogate. There is a trade-off between performance and cost: output with higher copy

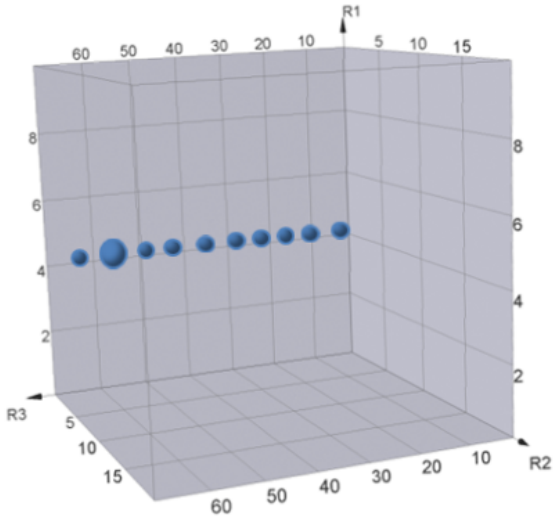
number would exceed the capacity of sequencing while lower output would exhibit weaker signals and lower the possibility of finding hits.

How to Control DEL Compound to Reach Desired Endpoint?

When DEL input and endpoint are given, the key parameters to control are incubation time, washing stringency and elution conditions. Although different targets retain DEL compounds in different degrees (meaning the number of compounds retained on target for a given round of selection), statistically 0.1%~1% of DEL compounds are recovered for most of the targets. Usually 3 rounds of selection typically result in the proper endpoint defined above.

5.4. Affinity Hits Identification Algorithm

The results of DEL selection can be either identification of single compound or very often compound clusters where compounds sharing same scaffold and partial building blocks. At HitGen, we call a compound cluster as a Feature that usually recognized as a line or a plane in a cubic plot. For example, a line feature consisting 10 molecules with common scaffold, R1, R2 building blocks and variable R3 building block is shown in the following cubic plot. The bubble sizes reflect the sequence counts of the DEL compounds. In order to quantify enrichment strength and effectively identify features, we invented a proprietary hit identification algorithm, PolyO.



A Line Feature in A 3D Cubic Plot

What does PolyO stand for?

The full term of PolyO, “log10 transformed Poisson probability product score with normalization”, suggests that this algorithm is derived from a mathematical model with regards to probability

theory and Poisson distribution. Its short term o0oooooooo (PolyO) is a vivid representation of a feature in a cubic view.

How do we calculate PolyO?

PolyO algorithm calculates the feature enrichment score in three steps. Firstly, a probability score is assigned to each compound based on Poisson distribution. The choice of lambda λ in Poisson distribution is related to the sequencing depth (average number of times a molecule being sequenced). Compounds with high sequence count will be assigned with a low probability, indicating its significance in a selection experiment. Secondly, the raw PolyO score of a feature is calculated as the product probability of all individual molecules on that feature, which denotes the joint probability of all independent compounds being enriched in the same experiment. Lastly, this raw score is normalized by the DEL size and the estimated background noise, in order to make this score comparable across DELs and samples.

How PolyO Helps Identify Affinity Hits?

Although it is controversial whether the copy count correlates with the affinity level of a molecule, by reviewing historical projects, we found that the confirmation rate is strongly correlated with the PolyO score in carefully planned experiments. PolyO score applies normalizations based on sequencing depth, DEL size, and etc. It is capable of detecting weak features that are often ignored at the sequence count level. Meanwhile, this scoring system shows no preferences over DELs with higher dimensions (e.g., 4-cycle DELs over 3-cycle and 2-cycle DELs), making the selection outcome less likely to be overwhelmed by large compounds at signal level.

6. DEL Selection and Hit Identification Process

6.1. Target Activity Confirmation and Pre-Selection Studies

For compound confirmation and future translation reasons, we need to pursue DEL selection experiments under the natural physiological conditions (i.e. correct folding and proper behavior) for any given targets. As DEL selection is an affinity-based ligand identification process conducted in aqueous buffers, ensuring the proper binding capability of the target is a must-have prior to the actual selection.

HitGen has developed a series of criteria to achieve high success rate where a careful assessment regarding the target activity confirmation and selection feasibility are introduced as Pre-Selection studies. We highly recommend targets with high purity and desired multimeric status prior to selection experiment (the target is not necessarily a homogeneous protein; it is acceptable to screen a heterogeneous complex with structure clarified, proved to be stable, and with activity).

Target Activity

The general idea of the target activity confirmation is to make sure the target behaves properly as the dataset provided by the target provider. Biochemical and/or biophysical assay(s) will be involved in the target activity confirmation.

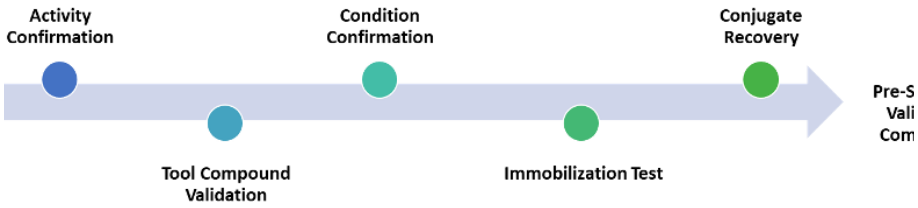
Whether the target is an enzyme or a structural protein, reference compound(s) will be a useful tool in activity confirmation. The consistency of affinity between target and tool compound(s) between HitGen and the provider will shed light on the readiness of the target and its feasibility for DEL selection.

Pre-Selection Studies

HitGen has developed a delicate methodology to ensure success in DEL selection while minimizing chance of false positive hits. A standardized path to meet the "Go/No-Go" point prior to the actual selection is to evaluate the feasibility of adapting the activity confirmation tests into the selection condition. A thorough assessment of each components will be conducted aiming to accomplish a well-characterized selection buffer in which the target would behave the same or at least similarly. This assessment normally will be established through a careful comparison of activity results between different conditions.

For challenging targets, HitGen recommends to use DNA-tagged reference compound(s) (termed "conjugate(s)") to validate the protein and selection process prior to the finalization of selection parameters. The successful recovery of the conjugate from a mini DEL selection helps solidify the DEL selection conditions. Of course, the conjugate can be used in the formal DEL selection as a positive control.

To conclude, pre-selection studies help build confidence for the initiation of DEL selection by established knowledge that the target shows good activity under selection condition, and the immobilization will not impede the binding between target and DEL compounds.



Scheme for Pre-Selection Studies

6.2. DEL Selection against Multiple DELs (Pooling)

HitGen has constructed hundreds of different types of DELs and the total size of the library has expanded over one trillion. In general, hundreds of libraries will be pooled to one or several DEL packages for target DEL selection in consideration of target usage efficiency, target type, DEL type, optimal DEL population as well as DEL solubility. HitGen offers highly efficient and flexible DEL pooling strategies with assistance of automatic or semi-automatic instrumentation, as a bonus, highly customized pooling strategies are also available.

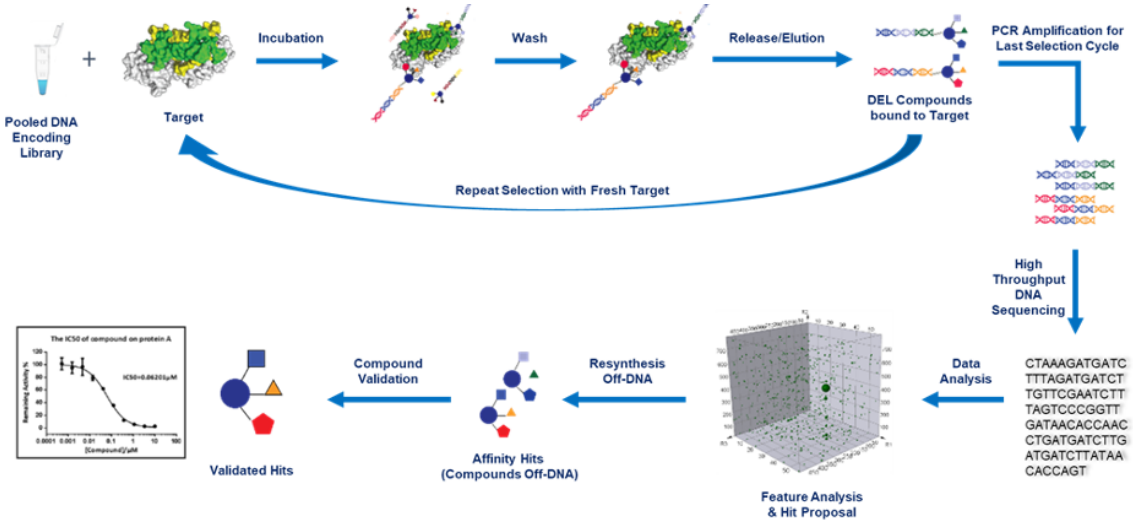
Key Instruments for DEL Pooling and Selection Aliquots

6.3. Various DEL Selection Processes for Different DEL Types

Regardless of the library type, DEL selection involves the following general steps: target-DEL incubation, unbound molecule removal, and bound molecule recovery. Each step is of great importance in effectively identifying hits. HitGen provides various DEL selection options involving different types of libraries, and each type of selection follows distinct procedures. Next-Generation Sequencing (NGS) is utilized for the sequencing process. In order to achieve high data usage efficiency, proper range of copy number in each sample is critical. To achieve this, we monitor total copy number of the elution with Quantitative Real-time PCR (qPCR) and make decision how many rounds of selection run are needed.

Classical DEL Selection

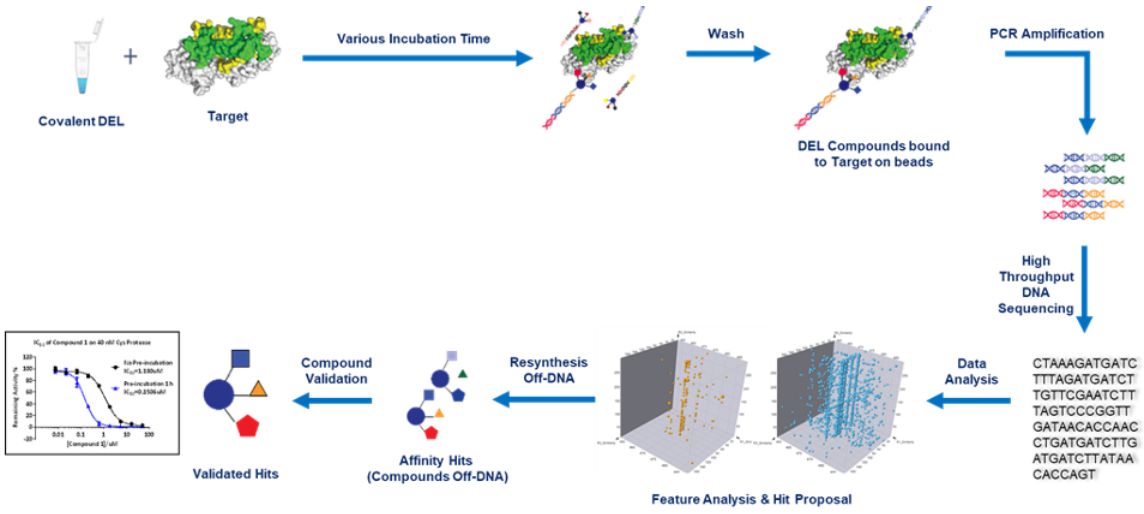
Classical DEL selection (also called Non-covalent DEL selection) is considered to be a "standard" selection in terms of the practical procedures due to reversible binding nature, time independent binding behavior between target and DEL compounds. Classical DEL selection takes multi-round selection where compounds get several rounds of equilibrium, therefore yielding highly reproducible results with proper enrichment fold and optimal copies for DNA sequencing. In particular, the target-library incubation step will be conducted with consistent mixing/pipetting to ensure sufficient exposure, unbound molecules will be removed by rinsing the target-DEL complex for several times. Denaturation of target protein releases the bound compounds (termed as "elution") and the selected compounds are recovered, quantified, and serve as the input for the next round of DEL selection where fresh target is used. The repeating cycle ends when the selection endpoint is met and the recovered compounds will be forwarded to PCR and DNA sequencing.



Scheme for Classical/Non-covalent DEL Selection

Covalent DEL Selection

Covalent DEL selection is to have DEL compounds bind to target first and then the embodied electrophilic warheads react with nucleophilic residues (Cys, Lys, Tyr, Ser, His) on the protein. When reacted, the compounds will be covalently linked to protein and cannot be washed off or released during protein denaturation step. This is the most significant difference comparing to classical DEL selection. Since the covalent reaction is irreversible and time dependent, longer incubation period would definitely favor the enrichment of covalent DEL molecules at target site. We often include several incubation time periods in covalent DEL selection to distinguish covalent binders versus strong non-covalent binders. A series of stringent wash,

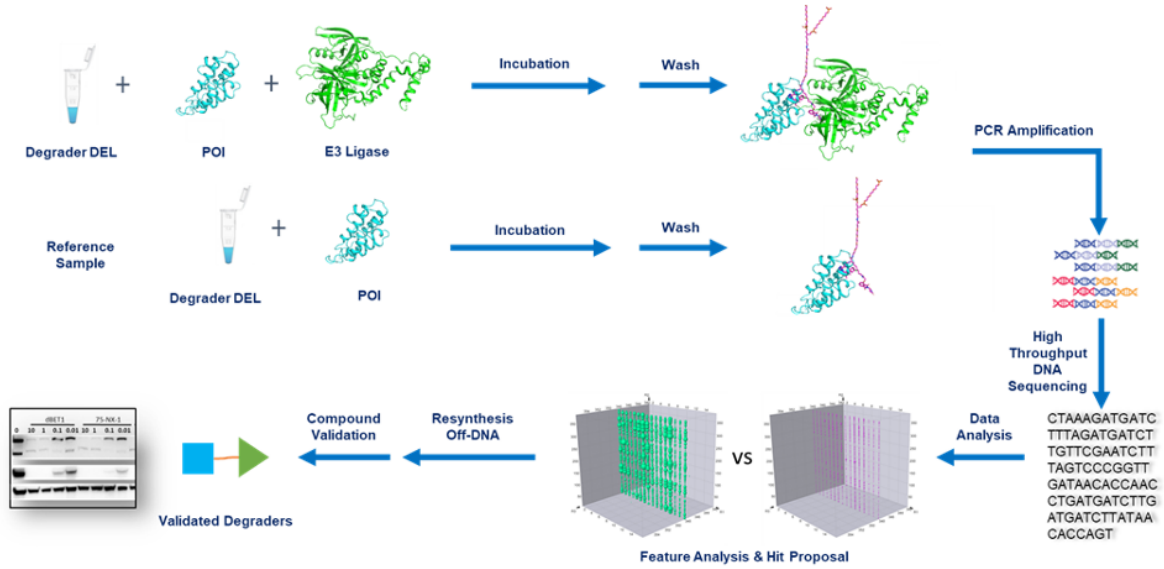


Scheme for Covalent DEL Selection

where denaturing of target is involved, will be applied after the incubation, this is not only capable of unbound molecule removal but also capable of trimming weak/reversible binders. Consequently, only those covalently bound DEL molecules will be enriched, isolated, and then identified within one round of selection.

Protein Degrader DEL Selection

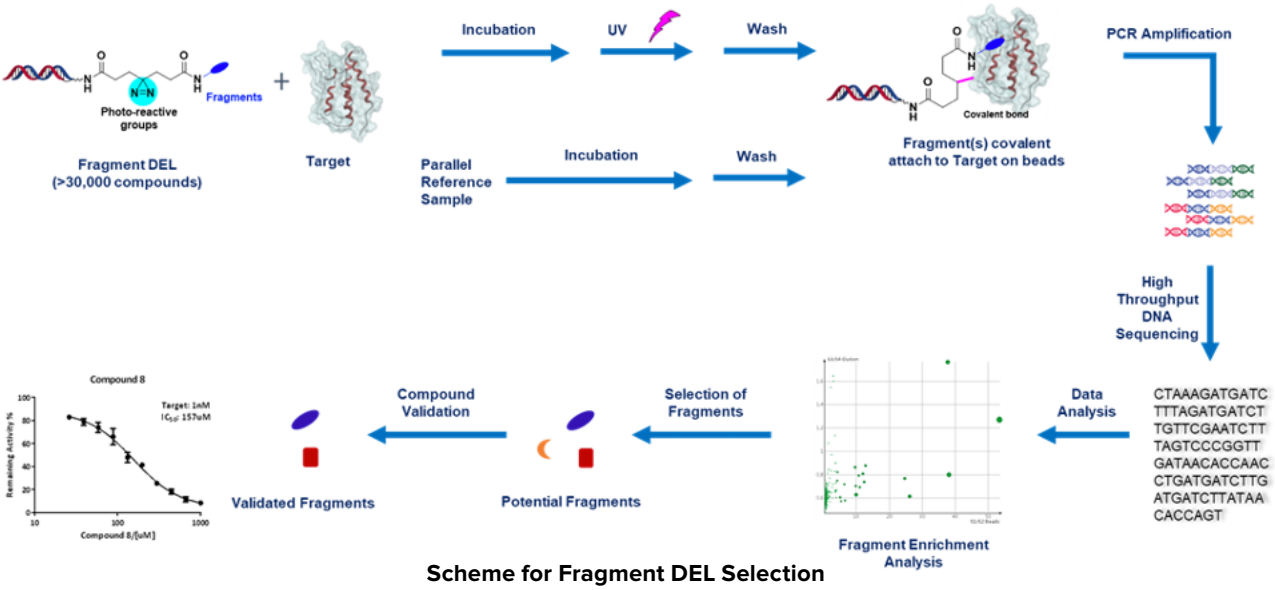
Protein degrader discovery can be fueled by harnessing the power of DEL and HitGen has developed several strategies for protein degrader discovery where either the degraders or the binders for POI (protein of interest) can be pinned down within one selection campaign. Co-incubation of protein degrader DELs and a mixture of POI and E3 ligase will preferentially isolate ternary complex of POI and E3 ligase bridged by DNA encoded degrader, where the degrader candidates can be directly identified. As a control, we usually perform parallel selection with only POI but the identical incubation, washing and recovery steps. For details regarding the direct identification of degraders, please refer to our case study of novel BRD4 degrader discovery. Alternatively, the degraders can be obtained by the selection of POI to identify the potential binders first (by classical DEL selection) and extend these binders via known linkage point with proper linker and E3 binder(s).



Scheme for Protein Degrader DEL Selection

Fragment DEL Selection

Fragment-based Drug Discovery (FBDD) is of great potential yet the relatively low affinity against target tends to be one inevitable obstacle for this technique. Combining traditional fragment library with fine-tuned DNA coding and a photoactivatable crosslinker enables fragment-based DNA-encoded library to be a promising path to drug-like leads with versatile chemistry capacity and high diversity. The experimental process is very similar to that of covalent DEL selection except that the reactive headpiece is photoactivatable, meaning that the anchoring of fragments is achieved by introducing a photo-crosslinking step at the end of target-library incubation. Stringent wash will remove the non-specific binders due to the intrinsically low affinity of the fragments and only those fragments with proper proximity will remain bound via crosslinking and therefore be identified.

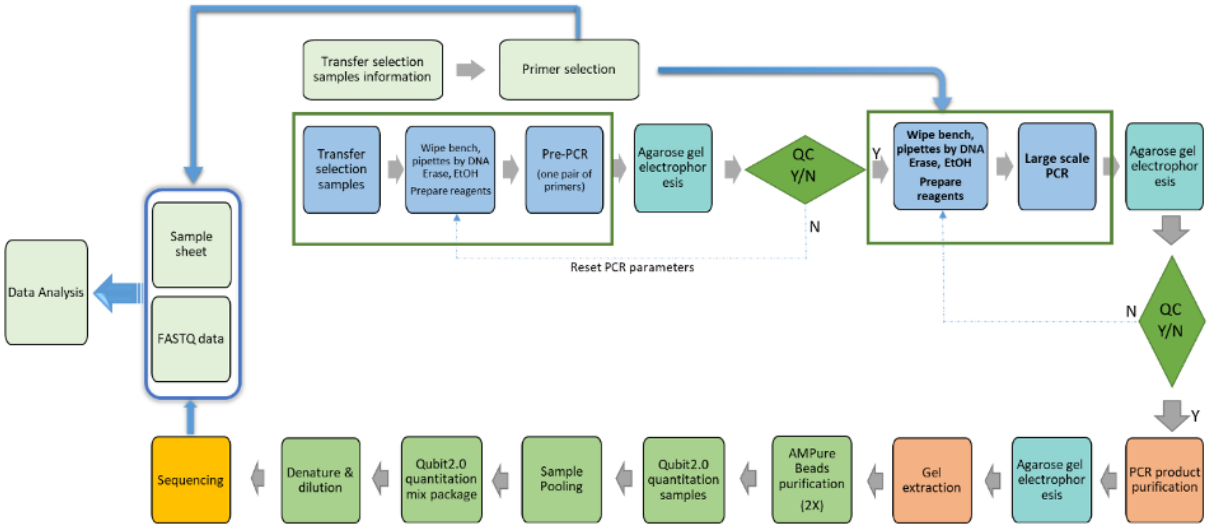


Scheme for Fragment DEL Selection

6.4.PCR and DNA Sequencing

PCR and DNA sequencing are the subsequent steps for DEL selection process. Ligands enriched in the selection procedure are identified by PCR amplification of the corresponding DNA-tag and by the subsequent sequencing of up to hundreds of millions of PCR amplicons by the Illumina sequencing platform. The relative abundancy of library codes found in the sequencing experiment correlates to those small molecules that are retained on the target protein in the selection experiment.

One Step PCR, a HitGen patented PCR method (patent application number 201811151077.3), is used to prepare the sequencing libraries of DEL selection samples. Compared to traditional sequencing library preparation kit, One Step PCR reduces the procedure from 4 steps to 1 step and modify protocols to fit DEL products sequencing strategy, leading to a very high sequencing output and quality.



2nd Generation Sequencing Process

HitGen currently has two Illumina sequencing instruments, Hiseq2500 and NovaSeq6000, providing high quality, flexible, and timely DNA sequencing. Specifically, it offers flexibility in sequencing capacities from 115M to 6400M reads and robust sequencing quality with average Q30≥90%. Based on different DEL selection endpoints and PCR amplified products, we select proper flowcell (6 different types of flowcells available) for DNA sequencing. In house DNA sequencing usually finishes in one day and there is zero delay in sequencing data analysis by our automatic data process tool.

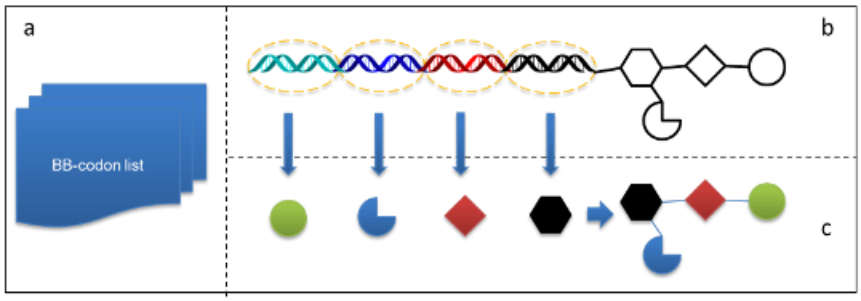
Flexible DNA Sequencing Options

Platform	Hiseq2500		Novaseq6000			
Flowcell (FC)	Rapid	V4	SP	S1	S2	S4
# of FC Lanes	2	8	2	2	2	4
Read Length (bp)	SR116+8	SR116+8	SR116+8	SR116+8	SR116+8	SR116+8
Sequencing Time*	16h	48h	12h	12h	15h	20h
Total Reads	300M	1600M	800M	1600M	3200M	10000M
Used Reads	115M	1024M	512M	1024M	2048M	6400M
Q30	≥85%	≥85%	≥85%	≥85%	≥85%	≥85%

* Sequencing time refers to single flowcell.

6.5. DNA Sequence Data Deconvolution

Data deconvolution is a process of converting selection readout (in the form of DNA sequences) to chemical compounds that these DNA tags encode. It includes several attainable but meticulous processes. To ensure the correct transection of the data stream, our platform takes all input information derived from a centralized database that stores the information of all codons, reagents, and their encoding relationships. Firstly, according to the building block and DNA codon relationship lists (bb-codon lists, exported from the database) (Figure a), the valid combinations of DNA codons are identified for all decodable sequences (Figure b to c). We use alignment-free algorithms with some major adjustment to allow maximum 1 bp mismatch/deletion in each codon region. Therefore, the algorithm can process sequence data in a super-fast manner while also tolerate to some level of sequencing errors and poor codon qualities. To keep track of all DEL compounds, a unique compound index is assigned to each DEL molecule based on its DEL number and the indices of the codons in each cycle. Once the decoding is done, we can obtain a list of compound indices and their copy counts. We use our proprietary feature identification algorithm (see "Affinity Hits Identification Algorithm" section) to further convert the copy counts to PolyO scores, a more balanced scoring function for feature identification. The primary structures of these enriched compounds will be enumerated based on the synthetic route of the corresponding DEL (Figure c) for downstream analysis.



Sequence Decoding Rationale

6.6. Automation in DEL selection, Data Analysis Report

Automation in DEL Selection

Procedure of DEL selection includes protein immobilization, DEL incubation, wash unbound binder and elution by denaturing target protein. Affinity resin and magnetic beads are widely used for protein immobilization to sufficiently separate enriched DEL molecules on protein and un-enriched DEL molecules in DEL solution.

In most cases, DEL selection at HitGen is performed with automated sample purification system (KingFisher Due prime and KingFisher Flex), which can automatically process 1~96 samples by moving magnetic beads (not liquids) among different wells of 96-well deep plate for different selection steps. The automated DEL selection can save time by removing manual steps and guarantee consistent results. For detailed selection automation, please refer to a well written paper of DEL selection using Kingfisher: Nat Protoc. 2016 Apr;11(4):764-80.

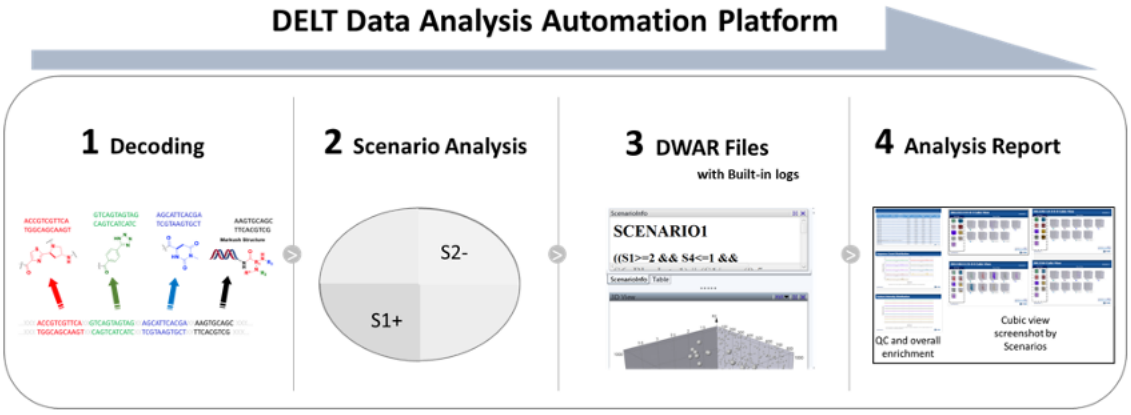
KingFisher Due prime

KingFisher Flex

Automation in DEL Selection Data Analysis Report

Generally, DEL selection results contain hundreds of millions of (chemical) data points for each selection campaign. Fully exploitation of selection results is the most important step in the entire DEL selection process. Library-by-library feature examination is a simple approach to visualize and analyze the DEL selection result, however it becomes very challenging for surveying and rank ordering compounds against hundreds of DELs.

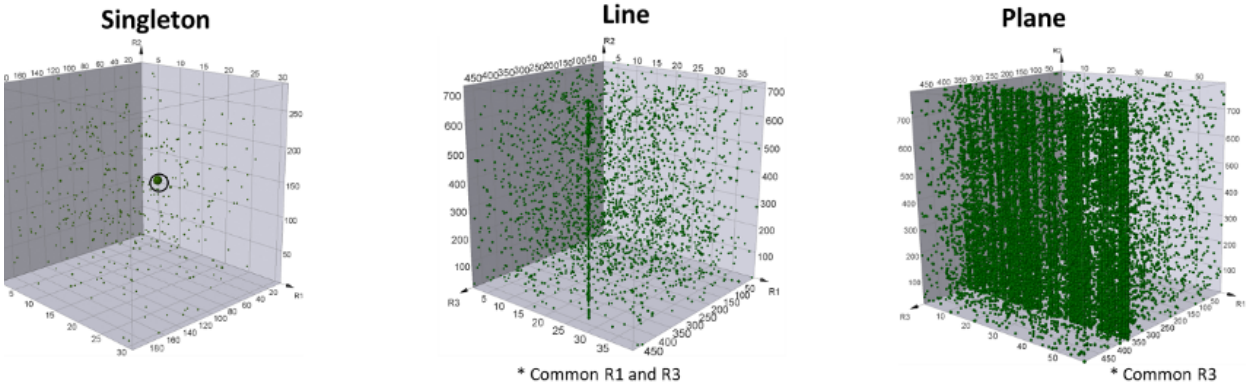
To improve the efficiency of the data processing and analysis, also to establish better support for data-driven decision making, HitGen has built a DELT Data Analysis Automation Platform. This platform includes the sequencing data demultiplexing, decoding, enrichment calculation, feature identification, automatically generation of DWAR files with built-in logs, and ppt/web version data analysis report. This platform significantly shortens the data process and allows project team to review the preliminary data shortly after the sequencing completes. The workflow of DELT Data Analysis Automation Platform is shown below:



Automation in DEL Selection Data Analysis and Feature Report

6.7. Hit Proposal and Off-DNA Synthesis

As per described, HitGen's proprietary algorithm will efficiently help with the process of locating significantly enriched signal patterns (also known as "features") in the selection campaigns. A feature could be a chemotype series with varying number of compounds. The lower limit of this number is a singleton with only one highly enriched species, while the higher limit of that could be over 100,000 depending on the favored moiety combinations, library size or other factors. For the majority of features, they usually contain over 100 compounds while a classical selection normally would not be only dealing with one feature. The representative features plots are shown as below.



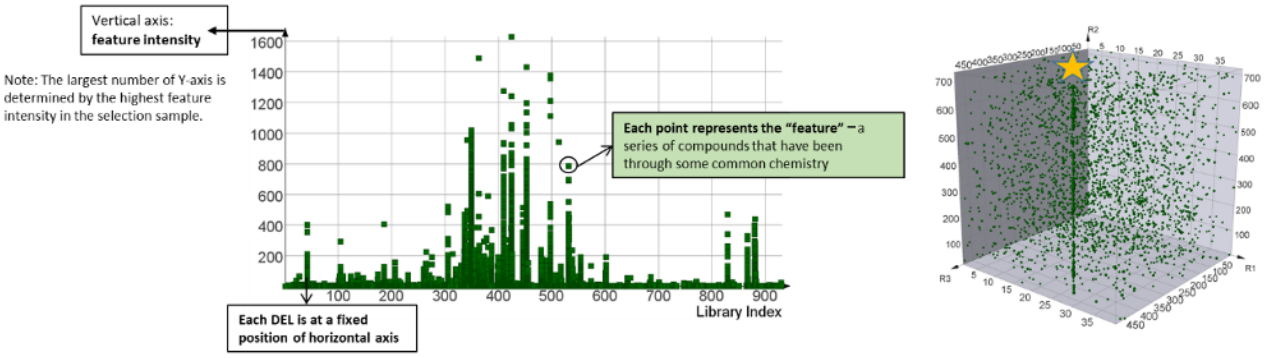
Different Types of Features

Hit nomination is a relatively complex process and multiple factors are considered to select structurally diverse, attractive and active compounds. The major considerations of hit nomination are listed as below:

• Frequent-hitter database: a complete record of historical features, so any feature that has

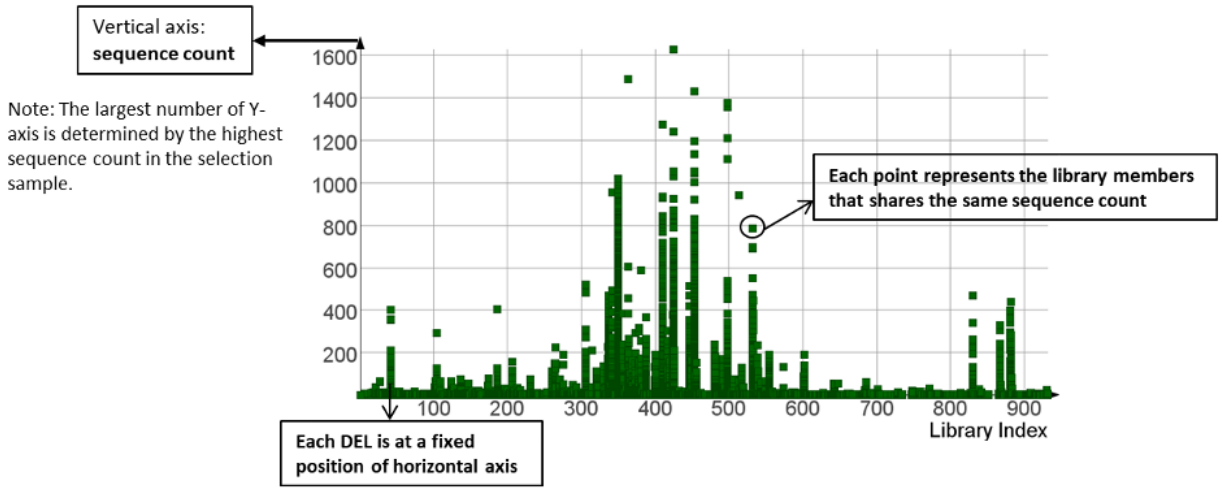
been enriched in more than one project will be flagged as potential frequent hitters and thus deprioritized.

- Property distribution: a multidimensional analysis of the physiochemical properties of enriched DEL species to prioritize features in terms of the “drug-likeness”.
- Feature intensity: a normalized parameter to measure signal strength and judge how preferred a given feature or chemotype is to the target against the background noise and other features, Normally, the higher the feature intensity is, the better the signal of the compound series is enriched.



Feature Intensity Distribution

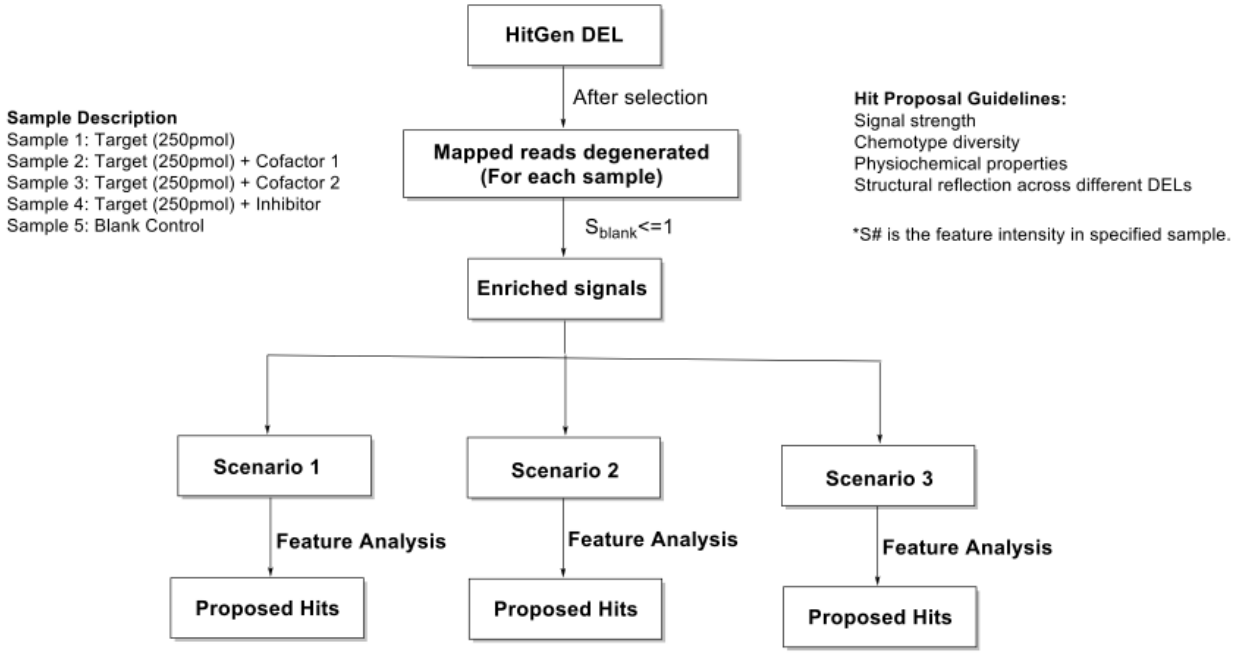
- Sequence count: the readout of the copy number for a particular DNA barcode as a library species after PCR, and represents the enrichment of individual species. As different chemistries and yields are involved in the DNA-encoded library construction, sequence count could only be regarded as in positive correlation to affinity on the level of overall probability. Besides,



Sequence Count Distribution

for a single compound to be considered as significantly enriched in a typical HitGen DEL selection, the sequence count of that species has to be no less than 2.

- Cross confirmation: the signal enrichment comparison between chemically related libraries and the overall trends of enriched structures to a) help with the prioritization of features, i.e. features with relatives enriched will be deemed as the higher priorities; b) give more guidance for hit-to-lead optimization in the future.
- Synthesis feasibility: the easiness of the synthesis, also a criterion to prioritize the features and compounds to maximize the outcome with limited efforts.
- Chemotype diversity: to incorporate different core structures in the hits nomination list if possible, so that the project team can have a more comprehensive understanding of the selection results and higher chance to hit the target.



Flowchart of Hit Nomination

The hit nomination is conducted collectively by DEL selection team and project team (from customer), a representative flowchart as shown above helps the team to streamline the process.

HitGen is capable of top performance of synthesis in the industry with experienced chemists that have worked on over hundreds of projects and advanced equipment and instruments at service. Typically, 5 mg is offered for every compound and for up to 100 compounds, with 6~8 weeks' lead time. Gram-scale of synthesis is also available as needed. On the top of competitive price range and efficient synthesis, the following items can be provided if requested:

- Synthesis status update: biweekly update to check up on the progress of synthesis;
- Synthetic route report: a quick report of employed synthetic routes;
- Synthetic procedures: a comprehensive report of the experimental procedures and work-up details;
- Analysis data: ^{1}H-NMR, LC-MS, HPLC.

6.8. Biophysical and Biochemical Assays for Hits Validation

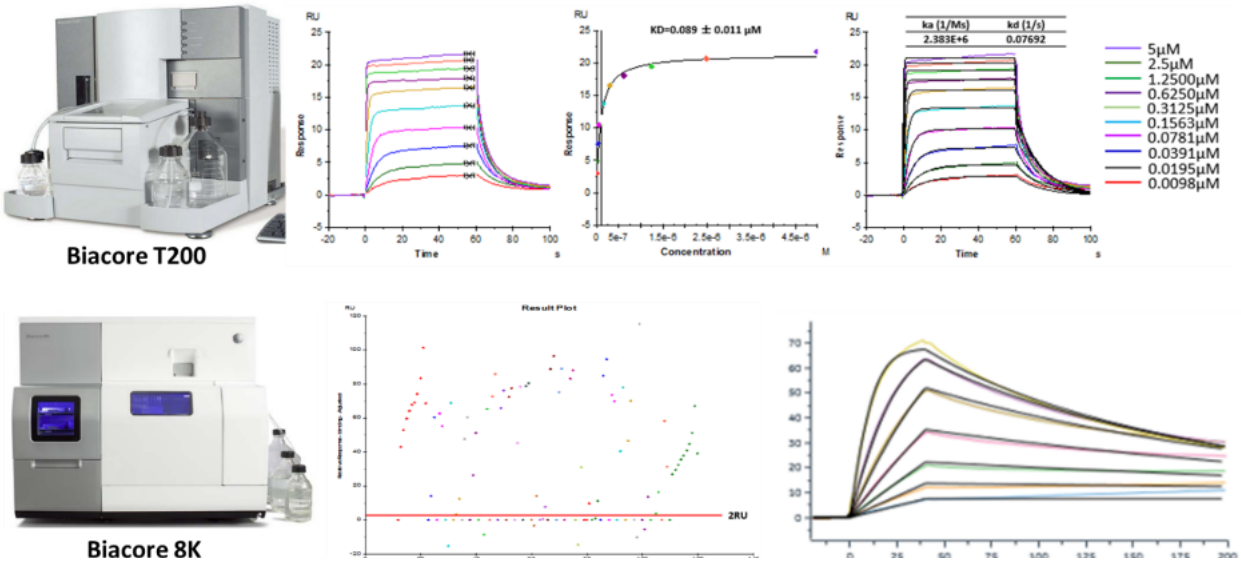
Biophysical Assays

HitGen recognizes application of biophysical methods and technologies an important role in different stages of drug discovery and development, especially during hit validation process. Robust improvement in methodology and instrumentation makes rapid and high throughput biomolecular interaction characterization achievable at HitGen with unprecedented quality. By establishing a comprehensive biophysical testing platform, HitGen is capable of providing accurate, real-time, and label-free interaction analysis data including both affinity and kinetic information, i.e. association rate constant (ka), dissociation rate constant (kd), equilibrium dissociation constant (KD), entropy/enthalpy change regarding intermolecular interactions.

Surface Plasmon Resonance (SPR)

Cutting-edge Biacore™ series SPR platform has become a rapid, highly-sensitive, and label-free tool for hit validation as well as hit-to-lead optimization.

- Biacore™ T200 and Biacore™ 8K guarantee efficient and accurate data generation with minimized hassle and maximal throughput.
- The platform offers both affinity and kinetic information with thorough analysis in a single run.

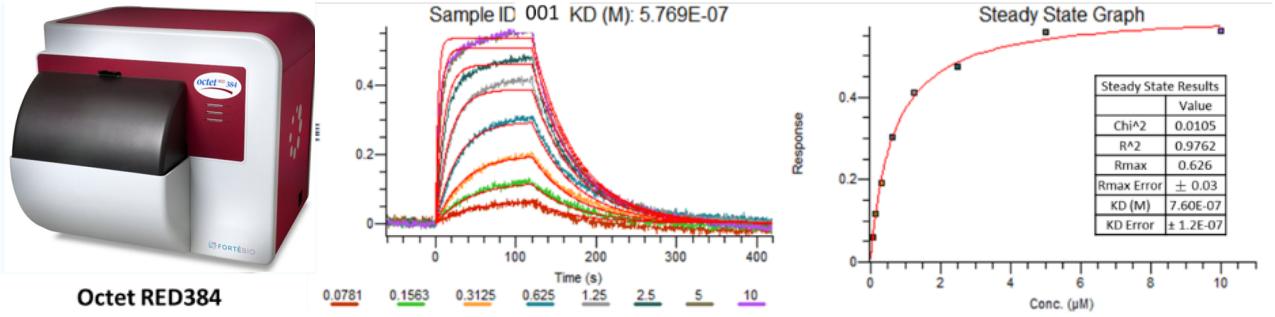


- Comprehensive understanding of biophysics and methodology enables HitGen to offer not only precise interpretation of the data but also customized guidance for hit optimization and drug discovery.

BioLayer Interferometry (BLI)

BioLayer Interferometry (BLI), as an optical analytical technique, is capable of measuring kinetics and affinity of biomolecular interactions in a real-time fashion.

- Octet RED384 System is designed for the analysis of biomolecular interactions in real time in the determination of kinetic constants.
- With 16 biosensors, Octet allows a variety of conditions to be monitored, and compared in real-time. This enables multiple concentrations, different antibodies, epitopes, clones, inhibitors, controls or different media types, and multiple layers of protein binding to be studied and monitored simultaneously.



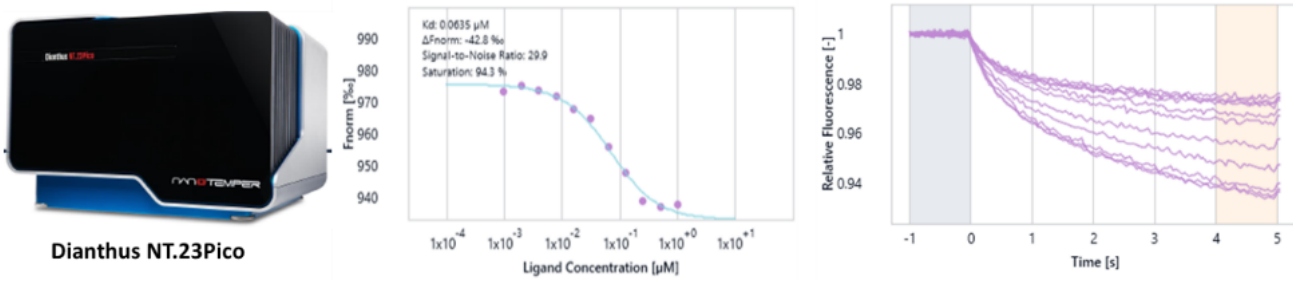
Octet RED384

- Highly sensitive and high capability of Octet facilitates hit validation process and therefore shortens the timeline for novel drug discovery.

Temperature Related Intensity Change (TRIC)

Temperature Related Intensity Change (TRIC) based Dianthus NT.23Pico, by measuring a fluorescence intensity change upon a precise laser induced heat, provides fast decision-making on the selection of right candidates for lead optimization.

- Dianthus NT.23Pico generates easy-to-interpret affinity ranking tables and histograms with great capacity: a full 384-plate containing different ligands and targets can be monitored and analyzed within 1 hour.



Dianthus NT.23Pico

• Little to no limitation of the sample form makes it possible to analyze aqueous solution, plasma, and cell lysate.

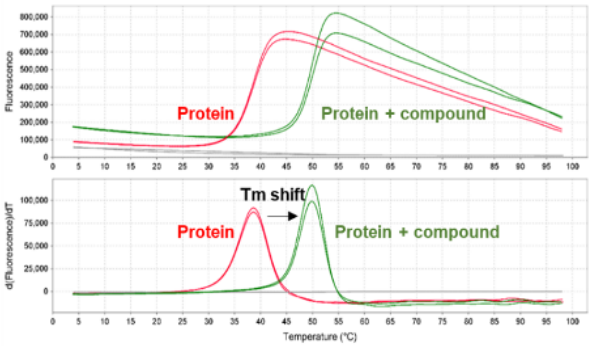
Thermal Shift Assay (TSA)

Melting temperature (Tm) shift measures thermal stability of target protein under various conditions and therefore offers a quick method to determine the binding between target protein and ligands.

• TSA measures the shift of Tm in the presence of a ligand, indicating whether the ligand stabilizes/de-stablizes the target.

TSA-StepOne Real-Time PCR System

TSA-QuantStudio™ 6 Flex Real-Time PCR System



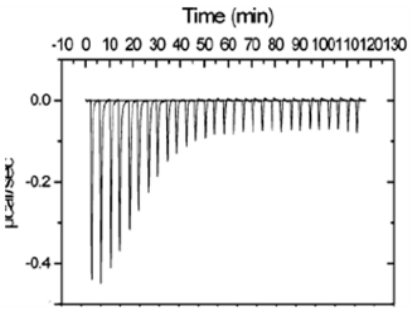
• HitGen equips both 96-well and 384-well real-time PCR systems which is capable of high-throughput TSA study.

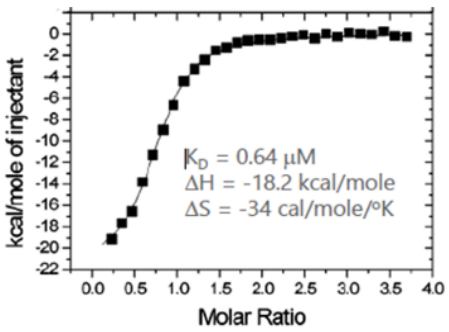
Isothermal Titration Calorimeter (ITC)

By directly measuring heat absorbed or released during the binding event, isothermal titration calorimeter (ITC) provides accurate reads of binding parameters including binding affinity (KD), stoichiometry (n), enthalpy (ΔH), and entropy (ΔS) in a single experiment.

• HitGen equips MicroCal PEAQ-ITC, which enables in-solution biomolecular interaction with high sensitivity, low volume, and free of labeling.

MicroCal PEAQ-ITC





Biochemical assays

Biochemical assays provide extensive knowledge in biomolecular functions, which has been a routine and reliable component in target characterization, hit validation, and hit-to-lead

optimization. HitGen has built profound experience of developing of functional biochemical assays during years of exploration in drug discovery.

- Kinase
- Hydrolase
- Protease
- Deubiquitinylating enzymes (DUBs)
- Phosphatase
- Dehydrogenase
- Transferase
- Ligase
- DNA/RNA polymerase
- Protein-protein interaction
- Protein-DNA/RNA interaction

HitGen has comprehensive capabilities in the development of functional biochemical assays for various target classes with great coverage:

Tecan Spark 20M Microplate Reader

Gilson PLATEMASTER P220

Echo 655 Liquid Handler

BMG PHERAstar FSX Microplate Reader

Multidrop™ Combi Reagent Dispenser

Liquid Handling and Microplate Reading Capabilities

With comprehensive instrumentation and methodology, HitGen can provide a full package of different assays including:

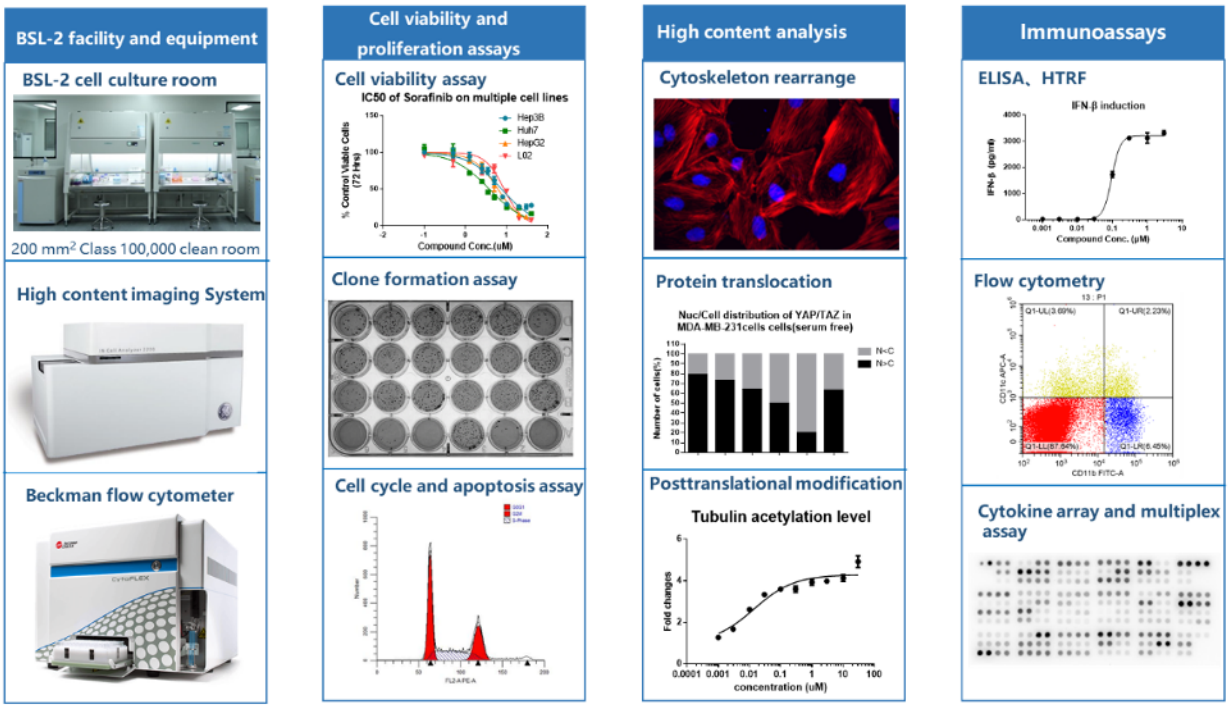
- Colorimetric assays (absorbance, etc.)
- Fluorometric assays (FP, FRET, FI, calcium flux, etc.)
- Luminescent assays (ELISA, Alpha-tech, BRET, Nano-BRET, Glo assay, etc.)
- Gel-based assays (pulldown, WB, etc.)

Alone with the standard biochemical assays, HitGen is also glad to provide highly customizable assay package including HTS screening for different types of ligands.

Cellular Assays

Cellular assays evaluate compound activity under their natural biological environment, providing better biological relevance. It ensures protein target(s) retaining their native conditions with physiological concentration/ cycling of substrate/ cofactors and assesses the compound permeability and engagement with desired target(s). The failure in translation of activity from biochemical assay to cellular assay is usually attributed to lack of cell permeability and complexity in cellular context. HitGen equips with BSL2 labs for routine cell culture/cryo preservation and offers a variety of cellular assay covering a wide range of assay types.

- FACs (target engagement, cell cycle, apoptosis)
- Cell engineering (stable cell line generation, CRISPR KO, siRNA/ shRNA KD)
- High content imaging
- Cell viability and proliferation
- Reporter gene assay (GFP, luciferase)
- Cell signaling assay (GPCR, Cytokine signaling)



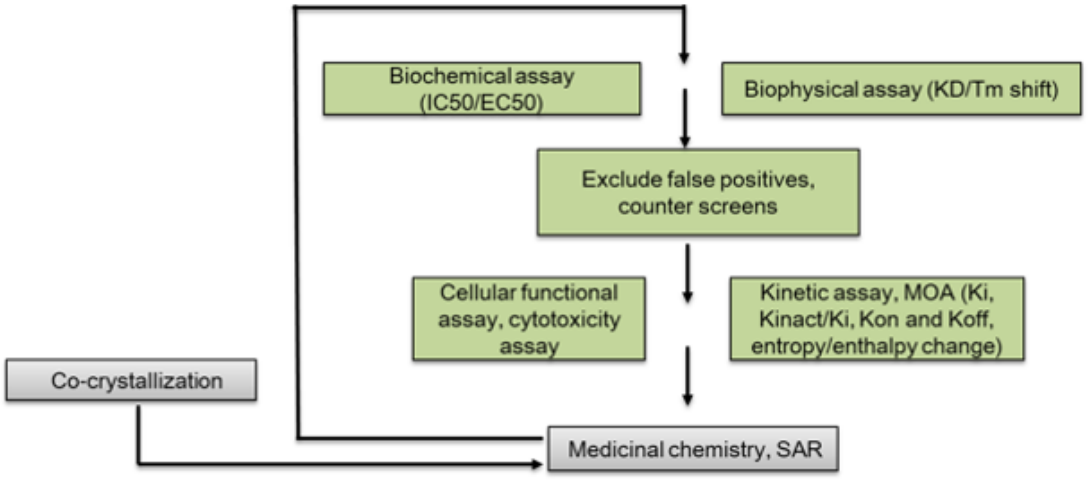
Cellular Assay Capabilities

Assay Development

A careful assessment of compounds requires a broad range of assays for determination of the potency, mechanism of action (MOA) and kinetics to enable the rank and profiling of the potential hits. An integrated assay cascade is a process with multiple steps (shown below) including potency determination by different biochemical and biophysical assays, counter screens, kinetic/MOA study and downstream cellular assays to ensure that the results are

biologically relevant. Appropriate evaluation of the drug candidate offers a full data package for various decision points and help selecting compounds progressing to next stage in drug discovery.

Various targets would require various characterization under specific, carefully designed QC package as a matter of nature. HitGen delivers fully customizable QC and selection plan with a versatile assay and selection cascade, which enables a thorough understanding and detailed achievable measurements. Scientist at HitGen would work closely with our partner to work through the cascade, and to provide intellectual input towards assay development, selection plan, hit validation, and troubleshooting.



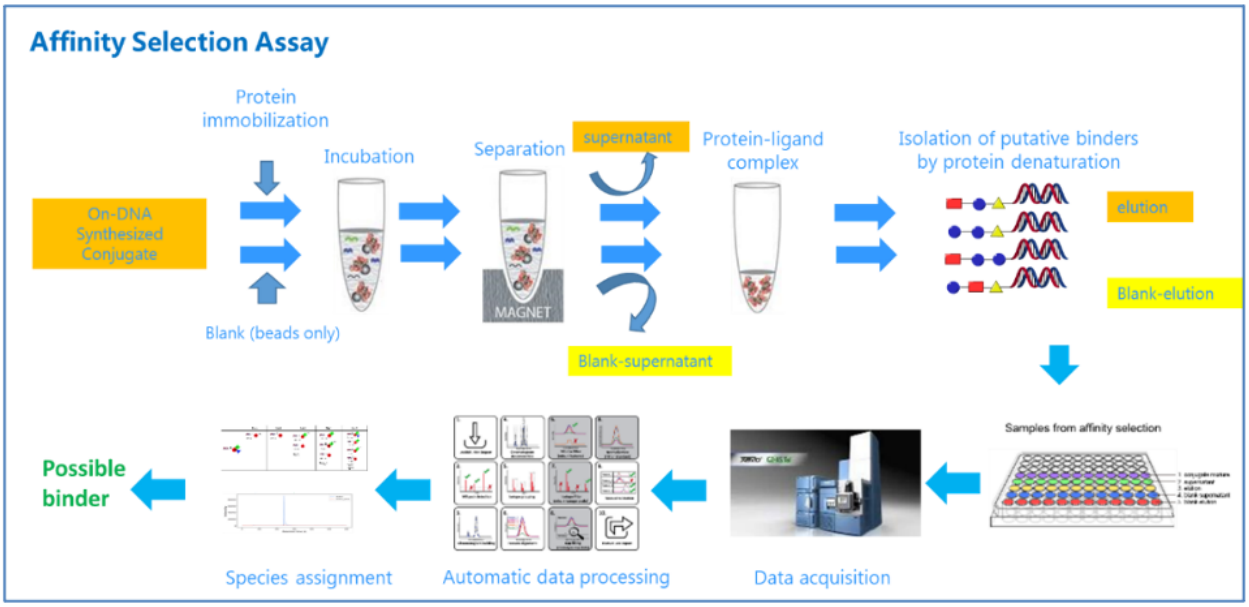
Assay Cascade of Compound Validation Cycle

6.9. ASMS for Feature Compound Identification

Affinity selection-mass spectrometry (ASMS) is an efficient and convenient high throughput screening approach for identifying potential ligands from a complex mixture. It assesses the binding abilty of candidate molecules to immobilized protein, whereas a mixture is incubated with target protein and then the potential ligands bonded to the receptor will be eluted and detected by MS after removal of the supernatant containing the components without binding. The advantage of this technique is that all the possible binders are detected as a whole without separation. It accelerates the process and reduces the cost of affinity screening. ASMS is a valuable complement to traditional drug discovery technologies, especially in the discovery for hit identification.

The DEL affinity screening results are presented as signal enrichment cubic view for feature nomination. The chemical structures of selected features from the cube are deducted according to their hit-index, and proposed for resynthesis for hit identification. Beside conducting off-DNA synthesis directly, ASMS with on-DNA validation is a powerful alternative approach for compound identification and validation. The on-DNA conjugates of deducted chemical

structures are resynthesized with the same condition as in DEL production. This mixture of on-DNA products and their byproducts are incubated with the same immobilized protein using affinity selection assay for selection of the components that are binding to target protein. These components are then identified using MS, and assigned to species of the on-DNA conjugates.



ASMS is used in on-DNA validation in order to select possible binders with high confirmation rate. This approach takes both products and byproducts into consideration by mimicking the reactions during DEL production. We are exploring different kinds of cleavable linkers, which is to minimize potential interference by DNA tags during affinity selection.

Workflow of ASMS

Our on-DNA ASMS workflow contains the 4 steps as shown below. The on-DNA conjugates of

Workflow	DEL_DV Feature Analysis	On-DNA Synthesis	Affinity selection & MS analysis	On-DNA work summary and off-DNA recommendation
Deliverables / Output	☐ List for on-DNA synthesis: proposed/selected features to do on-DNA synthesis; feature/hit index, BB information and availability Check	☐ Yield and identity of product mixture: overall yield analysis in each step of desired product and any possible byproduct	☐ QTof result for mixtures before/after AS ☐ Data interpretation of MS analysis ☐ Identified binders after ASMS	☐ List for off-DNA synthesis

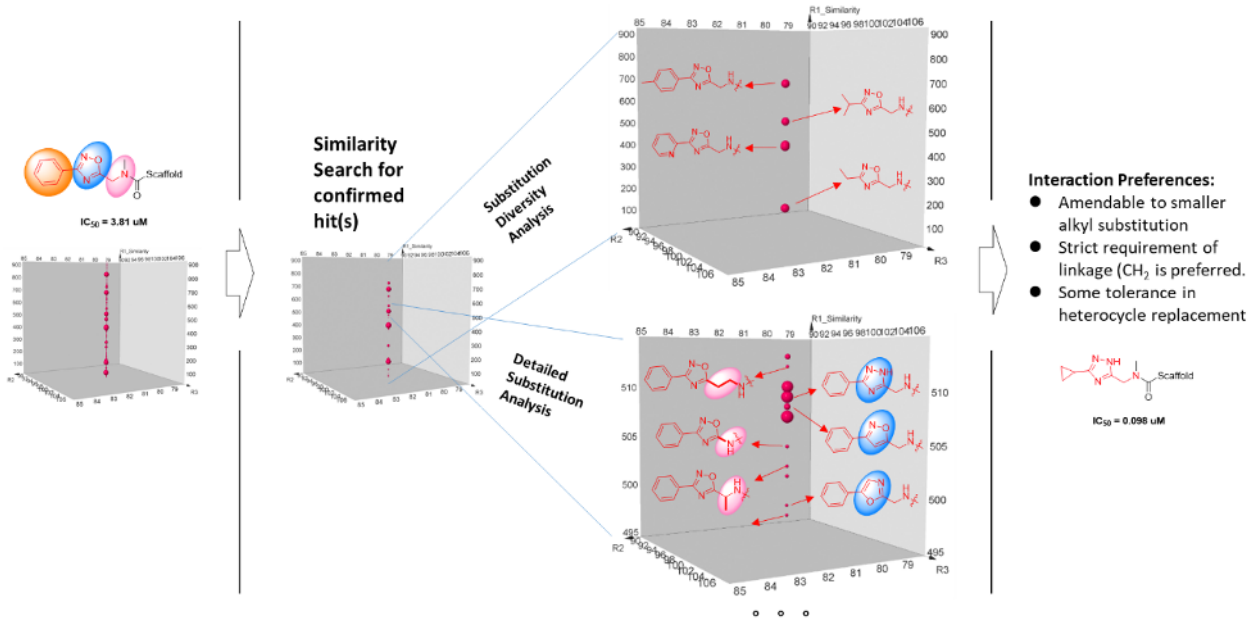
deducted chemical structures are synthesized with the same condition as in DEL production. This mixture of synthesized products and their byproducts are incubated with the same immobilized protein under the same condition as in the affinity selection. The putative binders in the elution are then analyzed using QTof and assigned to species of the on-DNA conjugates. A list of possible binders for resynthesis is finally delivered.

A confirmed ligand attached to a short strand of DNA is used as positive control if available. A short strand of DNA is added as an internal standard to validate the system and adjust the signal response. Experimental parameters including detection threshold in the affinity selection assay is provided prior to collaboration.

6.10. DEL Results Driven Hit-2-Lead Optimization

One of the major advantages of DELT is the rich information coming with the affinity selection. The tremendous library is able to probe the target's pocket and provide feedback to the project team with the preferred traits and negative characteristics of the ligands.

This step of data mining can be initiated at any time, but it is most useful when the off-DNA synthesis of representative compounds has been proven positive in the bioactivity tests. The validated hits are good starting points for structure-signal relationship analysis to shed some light on the direction for hit-to-lead optimization.



Structure-Signal Relationship (SSR) Analysis

The concept of structure-signal relationship (SSR) analysis originates from structure-activity relationship (SAR) analysis and signal strength (sequence count for individual compound) is used in the place of actual bioactivity data. Since the structure-signal relationship analysis is largely

conducted on the same feature which entails the compounds to be analyzed are involved in highly similar, if not the same, chemistry, the difference of sequence count should be mostly due to the result of natural selection of the target. As aforementioned, sequence count positivity correlates to bioactivity.

The practice of structure-signal relationship (SSR) analysis is in the same nature of structure-activity relationship analysis. The twist is to use the logic of hit-to-lead optimization to find the matching structures and their corresponding sequence count, and then summarize the trend this set of data supports as exemplified in the above figure.

Take the above as an example, if the optimization of a certain moiety is up for consideration, then the analogues of that moiety paired up with the corresponding sequence count (proportional to the dot size) should be collected using similarity search. Ideally, the analogues should cover most of the starter-change in the traditional hit-to-lead optimization process:

- Minimal changes on the phenyl ring: small changes, e.g. extra methyl and the introduction of heteroatom are well tolerated, so the first conclusion is phenyl moiety allows for modifications;
- Drastic changes of substitution: isopropyl and ethyl are representative exemplars to testify the second conclusion that phenyl moiety could be replaced with smaller groups if not completely removed;
- Variations of linkage and other parts of the moiety can all be explored in the same way as illustrated above.
- If enough evidences are collected, it's reasonable to make a jump from the existing species to virtual compounds, e.g. different forms of truncations.

Taken together, SSR analysis is an efficient way to explore and confirm the strategies to accelerate the process of hit-to-lead optimization.

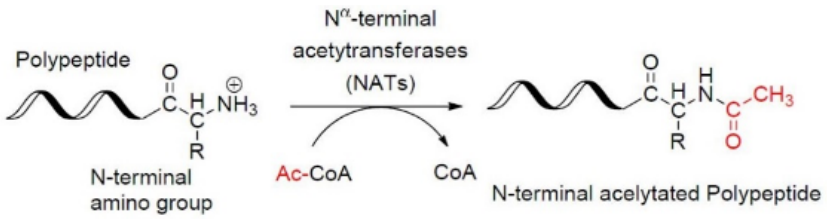
7. DEL Screening Applications-Case Studies

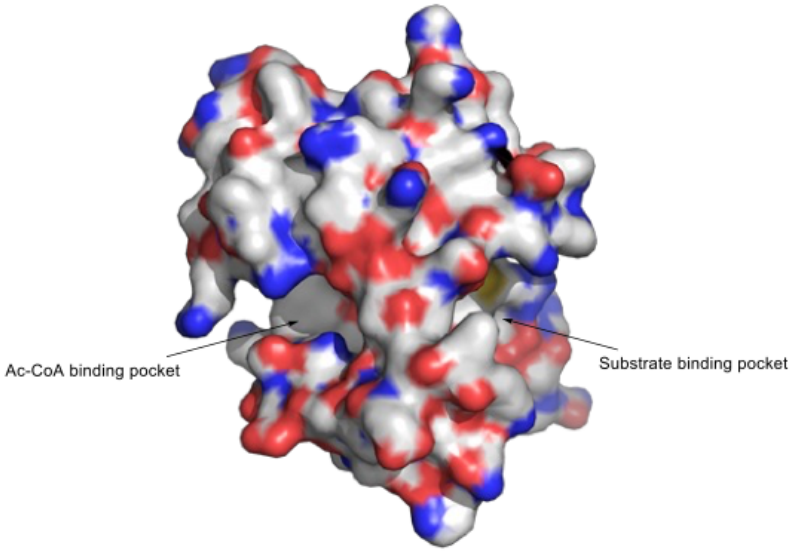
7.1. Identification of NAA50 Inhibitors by DEL Selection*

(*Collaboration Project with Pfizer, ACS Med. Chem. Lett. 2020, 11, 1175–1184)

Protein NAA50 and its Function

The N-terminal-acetylation of a protein can affect its nuclear import and export and can also act as a degradation signal to control the protein's cellular stability. The Nα-terminal acetyltransferase (NAA50) enzyme is a member of the Nα-terminal acetyltransferase NAT protein family. It coexists with NAA10 and NAA15 in the NatE complex and is responsible for the enzymatic function of the complex. NAA50 is also found to be essential for normal sister chromatid cohesion and chromosome condensation. Therefore, an inhibitor of the NAA50 enzyme might have therapeutic applications in oncology indications. The enzymatic catalysis and protein structure are shown below.

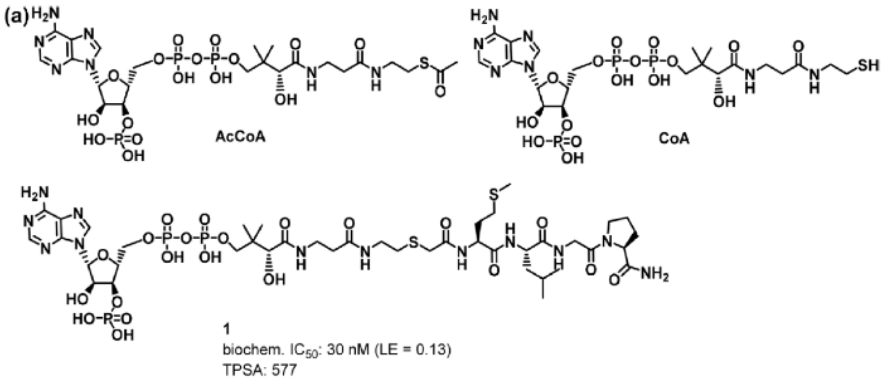




NAA50 Known Inhibitor and Objective of DEL Selection

Compound 1 is designed by studies of the NAA50 biochemical mechanism which indicated formation of a ternary complex between the AcCoA cofactor, an appropriate protein substrate (tetra-peptide MLGP), and the enzyme. Although compound 1 is a potent NAA50 inhibitor, the molecule is not particularly efficient due to its large molecular weight (ligand efficiency (LE)10 = 0.13). In addition, its high molecular weight (MW = 1,223) and high polarity (tPSA = 577 and cLogP = −4.1) likely prevents facile permeability across cell membranes and may thus compromise the use of the molecule as a robust in vitro tool compound.

The DEL selection is to identify potent and selective NAA50 inhibitors with improved physiochemical properties relative to compound 1 (i.e., reduced molecular weight and tPSA; increased logD).



DEL Selection Plan

In a typical DEL selection plan, we set up 3 samples (1. protein alone, Apo NAA50 in this case; 2. protein and inhibitor in saturated concentration; 3. Blank control). However, during the studies of NAA50, we realized that NAA50 protein involves conformational change in the catalytical process. By comparison of selection results of Sample 1 and 2 aforementioned, we are not able to effectively identify inhibitors. Consequently, we included two additional samples by the addition of AcCoA and CoA, seeking for compounds binding to the transition state. The final Selection Plan are summarized below.

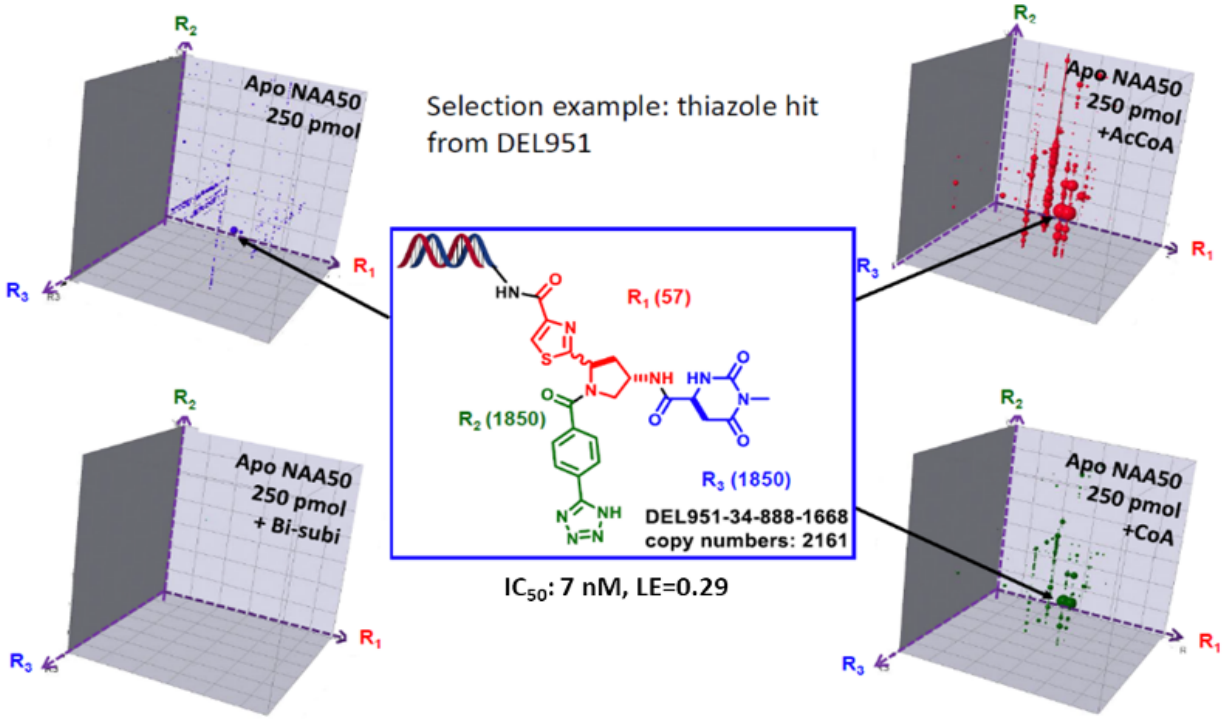
No	Target	Library	Supplement	Purpose
1	His-NAA50 (250 pmol)	HitGen DEL	-	Identify all the possible binders
2	His-NAA50 (250 pmol)	HitGen DEL	AcCoA	Identify all the binders in the presence of AcCoA
3	His-NAA50 (250 pmol)	HitGen DEL	CoA	Identify all the binders in the presence of CoA
4	His-NAA50 (250 pmol)	HitGen DEL	Compound 1	Combine with Sample 1,2,3 to identify inhibitors
5	No Protein	HitGen DEL	-	Identify beads binders

Representative DEL Selection Results

DEL selection is typically presented in a cubic layout, where the axis represents the corresponding building block and bubble size represents the sequence count of each compound. If the compounds have more engagement to the protein (high binding affinity or slow off-rate), the sequence counts will be presented as bigger bubbles. As we can see from the first four samples (blank sample 5 is now shown), Apo NAA50, Apo NAA50 + AcCoA, Apo NAA50 + CoA and Apo NAA50 + Compound 1 (also called Bi-subi) resulted different enrichment pattern, proving that NAA50 has conformational changes during the catalysis. Selection of compounds binding to all the protein stages and competing with Compound 1 obviously will yield NAA50 inhibitors. Such a selected compound structure is shown in the figure below, where the compound is built in the

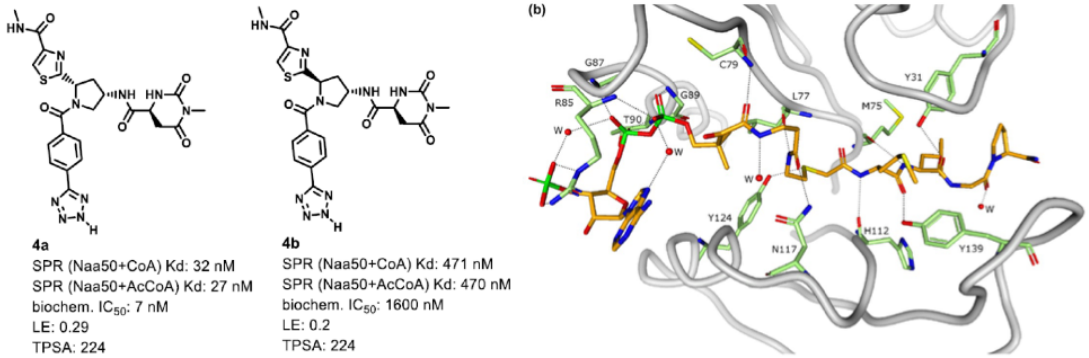
DEL as a racemic mixture.

The selected compound DEL951-34-888-1668 is resynthesized for validation.



DEL Compound Validation

The two isomers of the selected compound DEL951-34-888-1668 were synthesized and tested by SPR in the presence of CoA and AcCoA and biochemical assay. The chiral isomer 4a has been found as a very potent inhibitor with improved MW, Ligand Efficiency, and tPSA. The interaction of compound 4a and NAA50 has been further confirmed by co-crystal structure by Pfizer (pdb code: 6WFN).



7.2 Novel BRD4 degrader discovery using DEL technology

Advantages of using DEL Selection for Protein Degrader Discovery

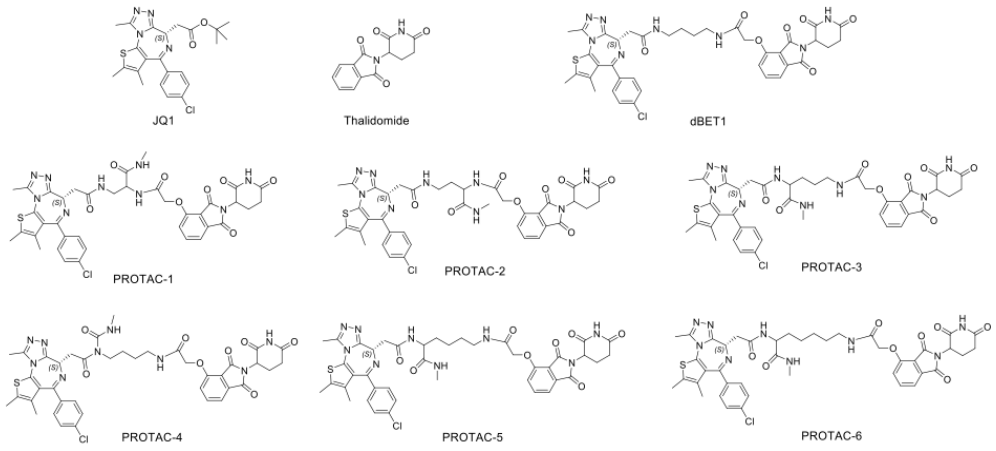
DEL selection is an affinity-based approach to recognize compounds interacting with the target, including the compounds modulating target function(s) or simply binding to the protein. The architecture of a DNA Encoded Compound is very similar to proteolysis targeting chimeras (PROTAC) as shown in the following figure. Both DEL compound and PROTAC molecule require covalent linkage of two molecules with known attachment points that have minimal impact to the binding.

More importantly, DEL selection is able to identify affinity binders for Protein of Interest (POI) and E3 ligases, therefore it has advantages in creating stronger intellectual properties and exploring therapeutic benefit from novel E3 ligase(s) by their tissue distribution differences.

In a traditional DEL selection, where target is seeking for compound, the direct readout of DEL Selection is the absolute number of DNA sequences (ideally proportional to number of compounds). In principle, when both POI and E3 ligase are used in the DEL selection, we are able to identify compounds binding to both proteins, using POI and/or E3 ligase as control(s).

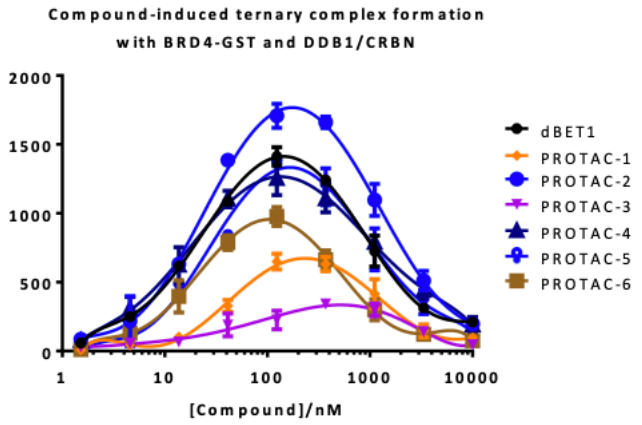
POI-Compound-E3 Ligase Ternary Complex Stability is the Key for Protein Degradation

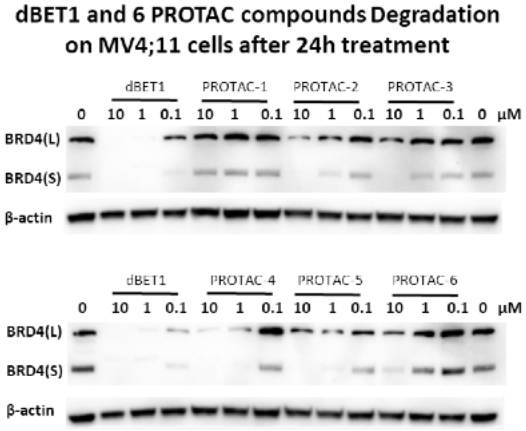
In order to prove the above theory, we have used JQ1 as the BRD4 binder, thalidomide as CRBN binder, linkers with different lengths to test whether we are able to distinguish the ternary binding stabilities. Specifically, six PROTAC molecules and dBET1 (a well-known BRD4 degrader, Biochem Biophys Res Commun. 2018, 497(1):410-415.) were made and their IC50 values were measured (compound structures and BRD4 inhibition are shown below). From the IC50 values, we are not able to know which one(s) bind to both BRD4 and CRBN to form more stable complex.



Compound	dBET1	PROTAC-1	PROTAC-2	PROTAC-3	PROTAC-4	PROTAC-5	PROTAC-6
BRD4 IC50 (nM)	43	130	92	88	50	72	120

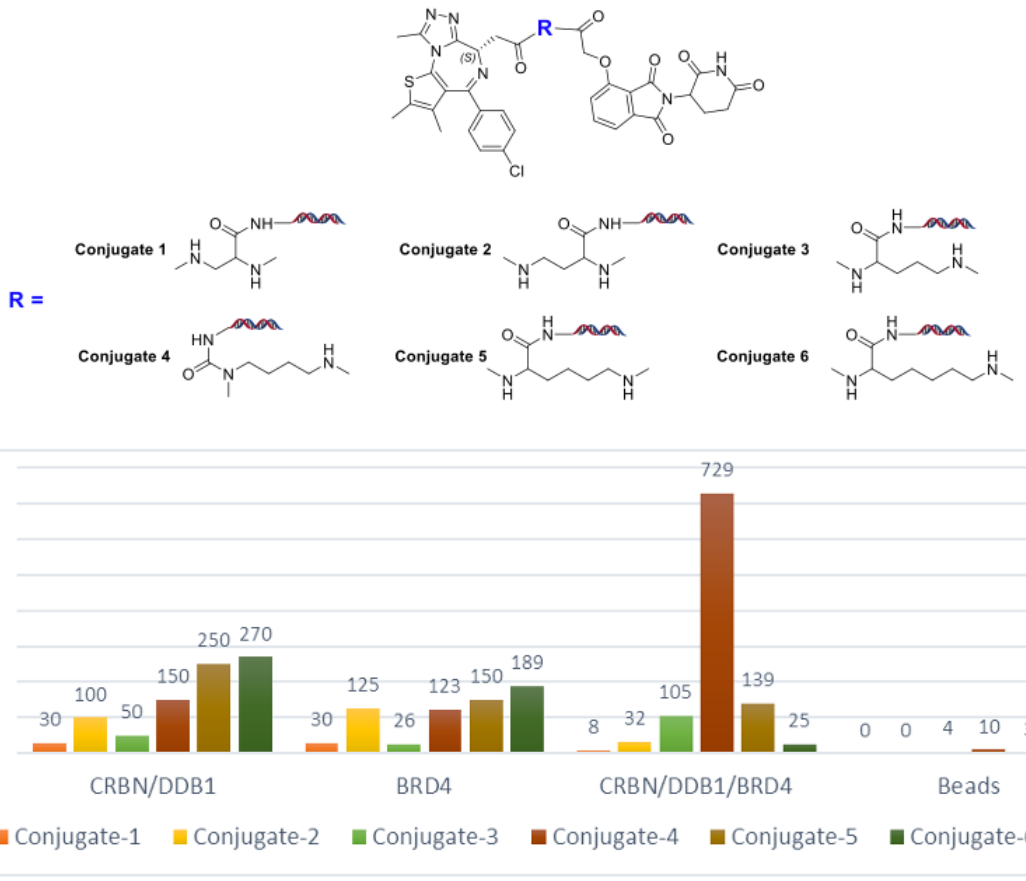
Then these 6 PROTAC molecules and dBET1 were evaluated with a FRET based ternary complex formation assay, they clearly showed different capabilities in ternary complex formation. BRD4 degradation experiment in MV4;11 confirmed the differential degradation of these 6 PROTAC molecules (Ternary complex assay and BRD degradation assay are shown below).





DEL Selection Approach Positively Correlates to Complex Stability

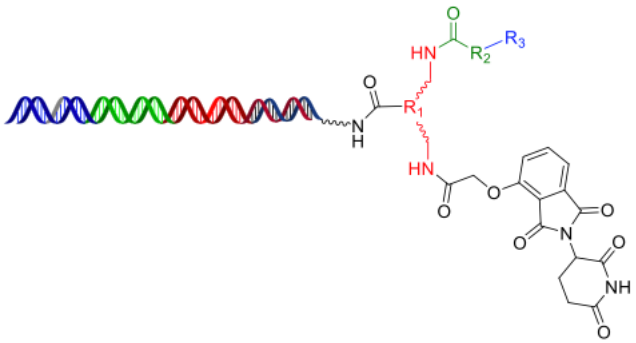
In order to confirm that DEL selection is able to distinguish complex stability, we made 6 corresponding PROTAC-DNA conjugates (conjugate structures are shown below) to see if different recovery can be observed from a DEL selection. In the "Conjugate Recovery Experiment", Conjugate 1~6 were pooled into DEL library; a DEL selection process was conducted and the corresponding DNA sequence counts were compared (shown below). Conjugate-4 is readily recognized as the optimal degrader molecule since it has significantly higher sequencing count, which is corresponding to number of compounds binding to CRBN-BRD4 complex. The length is consistent to optimal linker length in dBET1. On the other hand, recovered DNA sequences from CRBN and BRD4 samples did not give the correct trends.



Spike-in conjugates and Their Recovery from CRBN:BRD4 DEL Selection

Discovery of Novel BRD4 Degraders using HitGen PROTAC DELs

HitGen Protein Degrader DELs (described in the previous chapter) include multiple E3 ligase binders, various linkers and diverse chemical structures for POI binding. The following general structures shows one of the Protein Degrader DELs, where thalidomide is to target CRBN and R1, R2, R3 are linker and POI binding variation, respectively.



Representative CRBN Focused Protein Degrader DEL

Protein Degrader DEL Selection Plan and Objectives

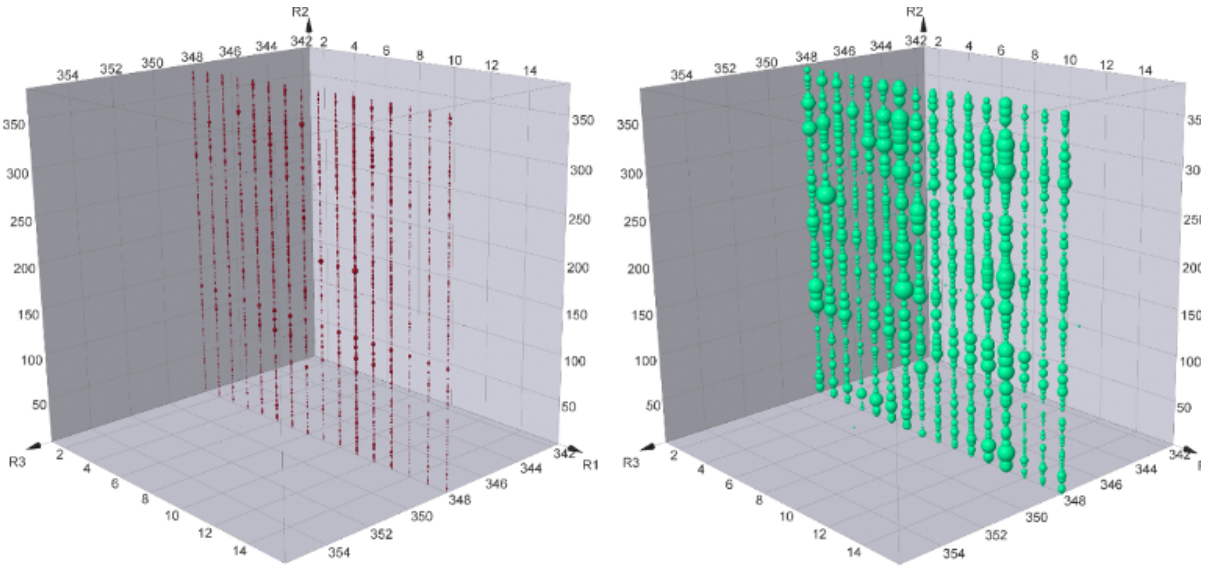
For this specific DEL, thalidomide has been approved clearly to bind CRBN, therefore DEL Selection against CRBN alone was not included. The Selection Plan is outlined below.

No.	Target	Library	Purpose
1	BRD4	Degrader DEL	Identify BRD4 binders
2	BRD4 & CRBN	Degrader DEL	Identify binders simultaneously bind to BRD4 and CRBN
3	No Protein	Degrader DEL	Identify beads binders

The aim is to identify any novel compounds binding to both BRD4 and CRBN over BRD4, subsequently proving that the PROTAC molecules with strong signals in Sample 2 result better BRD4 degradation.

Protein Degrader DEL Selection Results

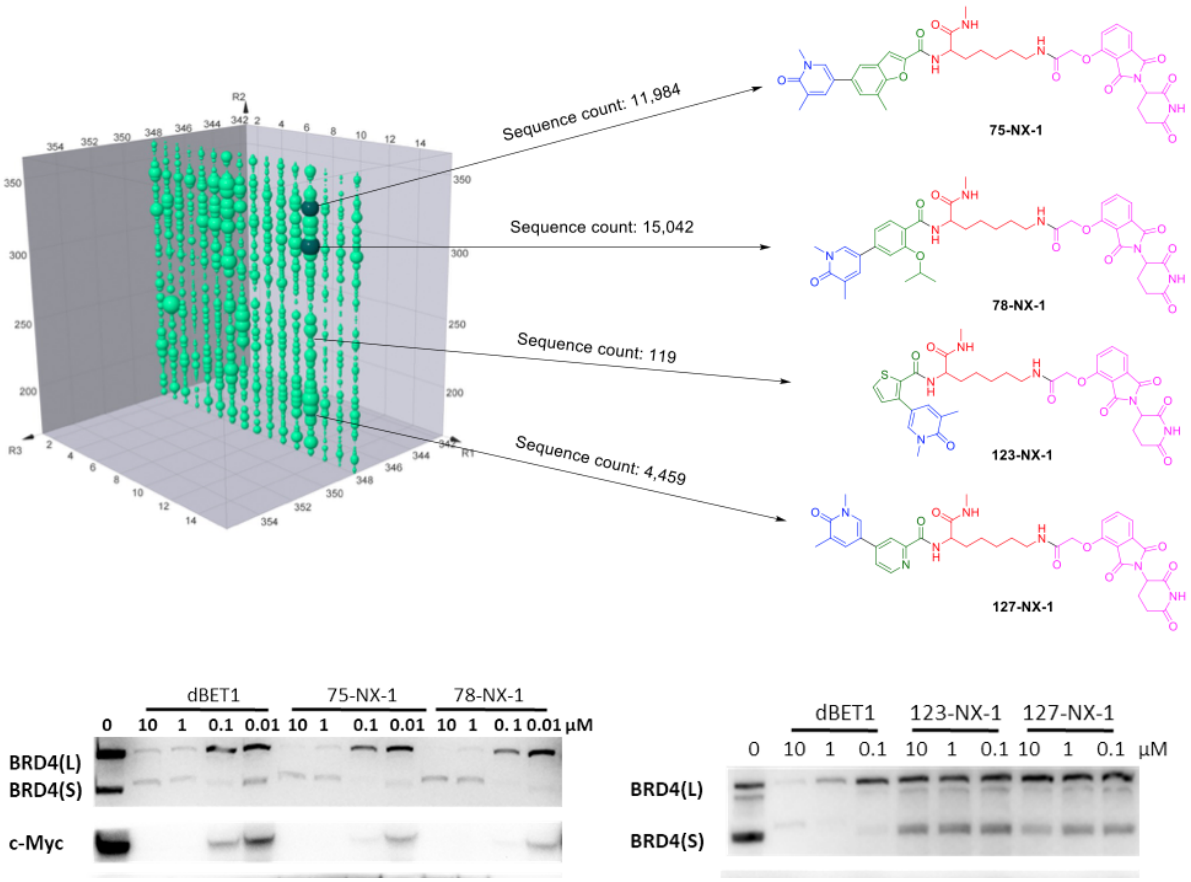
For clarity, a specific plane with given R3 in the cubic presentation of DEL compound enrichment was compared between Sample 1 and Sample 2 (Blank Sample not shown).



Enrichment Comparison for Sample 1 vs Sample 2. Sample 1 only use BRD4 as the protein for DEL Selection; Sample 2 use both BRD4 and CRBN for DEL Selection.

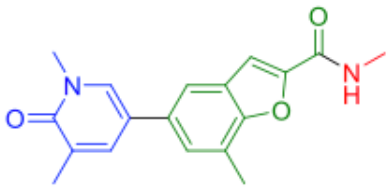
Plane R3=348 clearly suggests that many compounds on this plane are preferentially bind to both proteins (Sample 2) rather than just BRD4 (Sample 1). Meanwhile, there are also many compounds bind to both samples similarly. As shown in the following figure, all the compound structures are completely different from reference JQ1 or dBET1.

Presumably, the compounds with bigger bubbles (higher sequencing counts) form more stable complexes than that with smaller bubbles. In order to prove this, we have selected 4 representative compounds with different sequence counts in Sample 2 (11984, 15042, 119, 4459) for Off-DNA synthesis and protein degradation evaluation (Western Blot). These 4 compounds are named as 75-NX-1, 78-NX-1, 123-NX-1, and 127-NX-1, respectively. Using dBET1 as a reference, compound 75-NX-1, 78-NX-1 with higher sequencing counts was observed with very similar BRD4 degradation in MV4;11 cell; whereas compound 123-NX-1 almost did not show BRD4 degradation. Compound 127-NX-1 did not show BRD4 degradation either. In other words, the sequence counts positively correlates with protein degradation in MV4;11 BRD4 degradation (Western Blot results shown below). Similarly, the sequence counts of different compounds with different linkers are also positively correlates with BRD4 degradation (not shown).

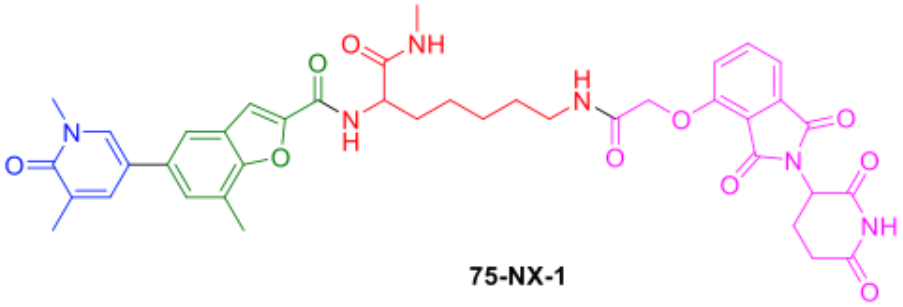


BRD4 Degradation by Newly Identified Degraders (MV4;11 cells were treated with compounds for 24h and Western Blot were performed)

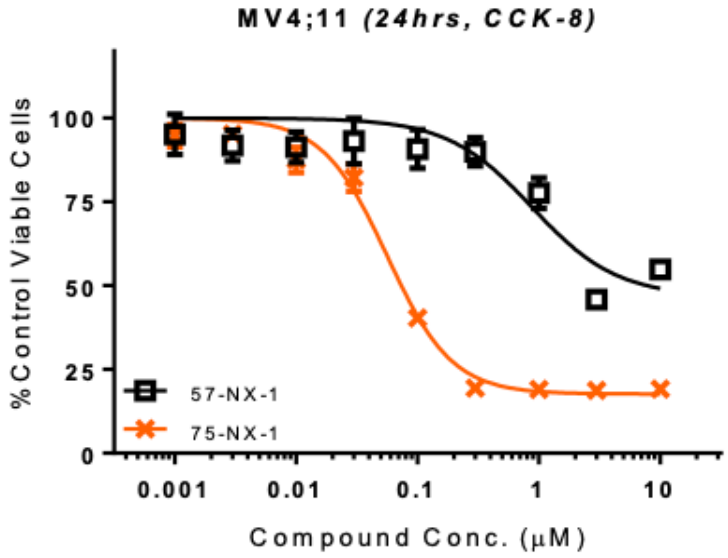
BRD4 degrader 75-NX-1 and its BRD4 binder/inhibitor (57-NX-1) are evaluated in anti-proliferation assay in MV4;11 (CCK-8), the results showed that compound 75-NX-1 has much higher activity than 57-NX-1 (Compound 57-NX-1 was found as a BRD4 inhibitor).



57-NX-1

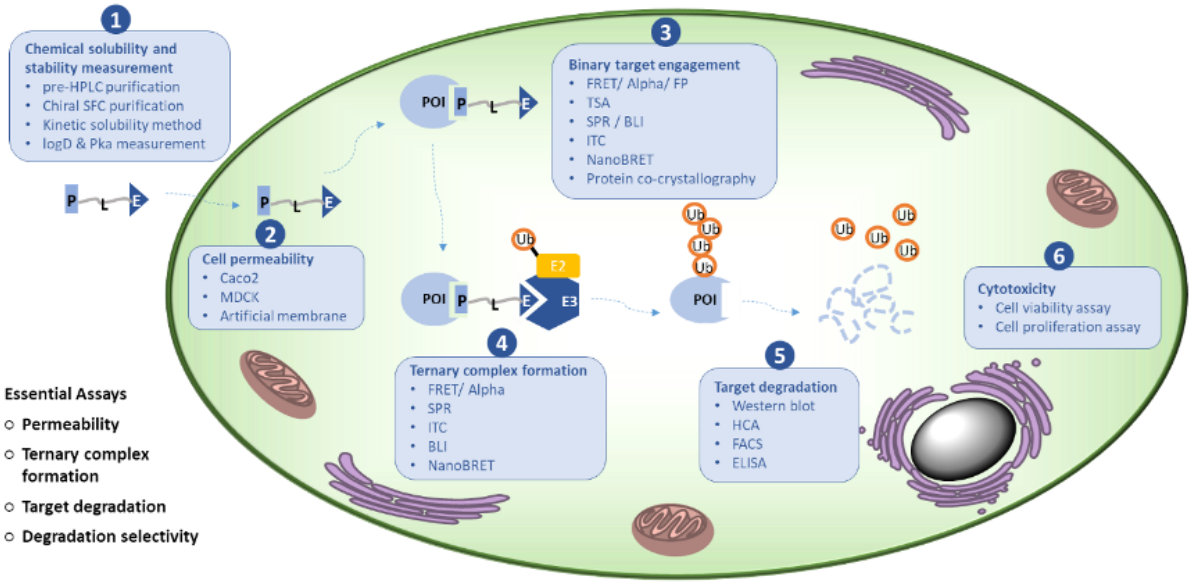


75-NX-1



Anti-proliferation Assay for PROTAC and its Precursor

Taken the above together, this experiment suggests that DEL selection is an effective approach to discover or optimize degrader molecules by identification of stable POI:Compound:E3 complexes. We are fully aware that there are additional factors could affect the protein degradation such as degrader solubility, lipophilicity, orientation and etc. as shown in the following figure, but there is no doubt that DEL selection approach provides a high throughput method to evaluate ternary complex formation and stability (an essential step for protein degrader optimization).



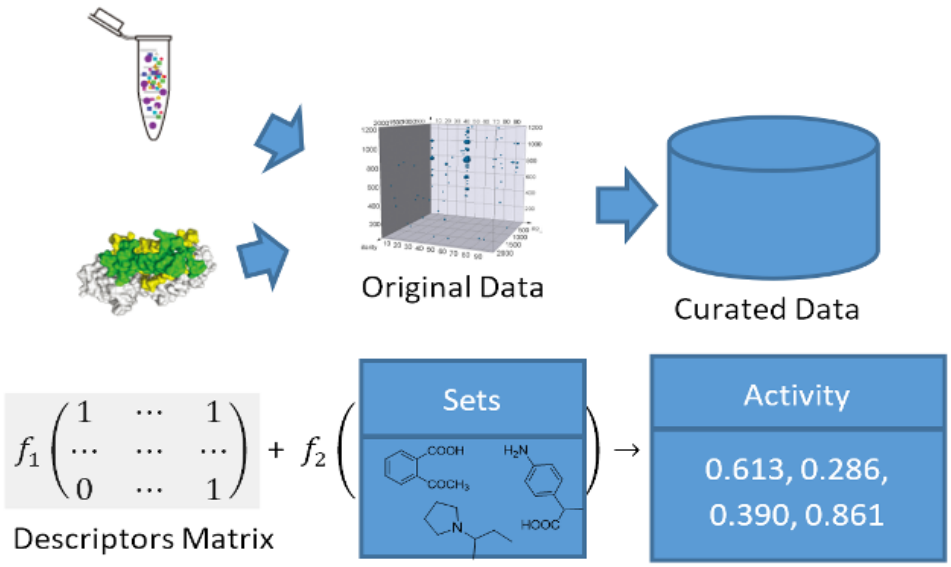
Protein Degradation Related In Vitro Assays

8. Accelerating Drug Discovery through AI Applied to DELs

DELs as an important hit finding tool, has been successfully used across a broad range of targets of varied classes by multiple pharmaceutical, biotech, and academic groups. With the power of NGS technology, this technology can generate millions to billions of data points simultaneously. Such high quality and quantity data can not only provide the active hit by directly examining the output, but also can discover hits outside of the DEL chemical space with the help of AI technology.

Constructing AI Models based on DEL Sequencing Data

It is a common method to use AI/QSAR models to predict the bio-activity of compounds, in the field of CADD. Data composed of the binding information of thousands of molecules to a target protein is a necessity to build powerful AI-based QSAR models. The abundant enrichment signal data generated by DEL screening makes it possible to build AI models to predict bio-activity of compound, explore unlimited chemical space in a time- and cost-effective manner and accelerate the drug discovery process. Applying AI algorithm to construct a model is a sequential process that requires design of AI architecture, data preparation, and model training & validation.



Scheme of Constructing An AI Predictor Based on DEL Screening Data

Design of AI Architecture

It is crucial to design an appropriate model architecture according to the task. This step includes the choice of a suitable algorithm and setting sensible initial values for hyper-parameters. In general, artificial neural networks (ANN) are given priority to complete the task. The architectural parameters for a neural network include, but are not limited to, the choice of the number of neurons and layers (and their type), learning rate and decay, regularization parameters, and the presence of connections between neurons or adjacent layers.

Data Preparation

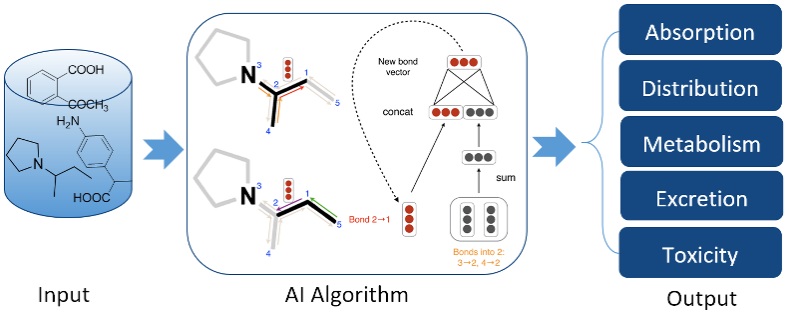
Preparing quality and quantity of data for AI is one of the most important and often the most time-consuming parts of data analysis, so a good data management system is very necessary. In order to efficiently collect and process the R&D data, HitGen has developed a customized intelligent data platform which can help AI group to prepare the high quality data efficiently.

Model Training & Validation

Once the initial architecture and the data sets have been established, we proceed to model training and evaluation. The training step aims to search a set of parameters with the objective of reducing/minimizing the prediction error. The final AI model should have the ability to express the underlying relationship between the molecular representations and practitioners' own specific purposes.

Prediction of ADMET Properties

The main reasons for research and development failure are the efficacy and safety deficiencies which are related in part to absorption, distribution, metabolism and excretion (ADME) properties and various toxicities (T) or adverse side effects. At HitGen, we developed an ADMET evaluation platform to help us select lead candidates and optimize lead compounds to clinical candidates.



ADMET Prediction Platform at HitGen

What Do We Provide?

Trained AI Model

The AI model trained on DEL screening data and that can be directly used to predict the bio-activity of external compounds, and explore unlimited chemical space at low cost.

AI Prediction

HitGen uses AI models to predict the bio-activity value of the compound provided by the customer.

Novel Molecule Generation Based on DEL Hits

Scaffold/Core hopping based on hit proposal or molecule with bio-activity, intensively explore adjacent chemical space. AI-aided drug design has the potential to change the way of drug discovery. The approach reduces the human effort involved, reduces the total time needed to arrive at a solution, and lets us explore wider chemical space.

HitGen's DEL/AI-aided drug discovery platform is an automation and informatics driven solution for the drug discovery, we have established the concept of building AI model based on DEL screening data, including design of the AI architecture, data preparation, and model training & validation.

9. Summary and Outlook

DNA-encoded libraries (DELs) have come of age and emerged to become a powerful technology platform for ligand discovery in biomedical research and drug discovery. DELs have been widely adopted in the pharmaceutical industry and employed in drug discovery programs worldwide. As this technology becomes more broadly accepted, a number of hit compounds originated from DELs have been reported in scientific literature, and several of them have already been progressed into clinical trials. We anticipate that DEL will increasingly be used in parallel with alternative hit-finding methodologies such as HTS and fragment-based lead discovery.

The key to the future success of DELs as true discovery modalities is to go beyond just making unprecedented molecules or finding affinity hit compounds against purified targets. We present

our views and perspectives on the present challenges and future directions for the development and application of DELs as indicated below. Ultimately the future success of the DELT will be reliant on a community-wide joint collaboration including both industry and academia to advance the technology.

► **Expanding Molecular Diversity**

- More reactions: broader range of coupling reactions, organic solvent reactions, etc
- More building blocks
- Different types of molecules

► **Improving Selection**

- Moving from affinity to function
- Rich DEL data mining and prediction of active compounds in expanded chemical space
- Selection Efficiency and Flexibility

► **Broadening applications**

- Expand to more target types: RNAs, Transcription factors, Phosphatases, and more
- Expand to broad biological systems or MOAs: cell surface protein, intracellular targets, protein degradation
- Focused library for lead optimization