3DEXPERIENCE®

Accelerate Science-Led Innovation for Competitive Advantage

BIOVIA Discovery Studio

Basic training course

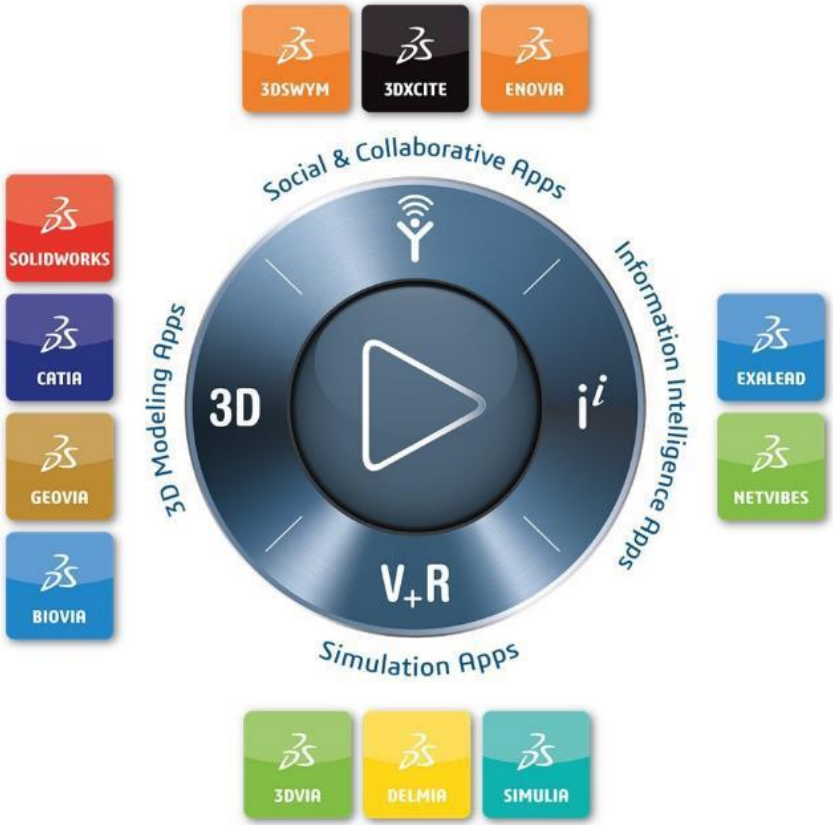
訊聯基因數位股份有限公司

分子數位中心

資深經理 陳冠文 (Gene)

訊聯基因數位

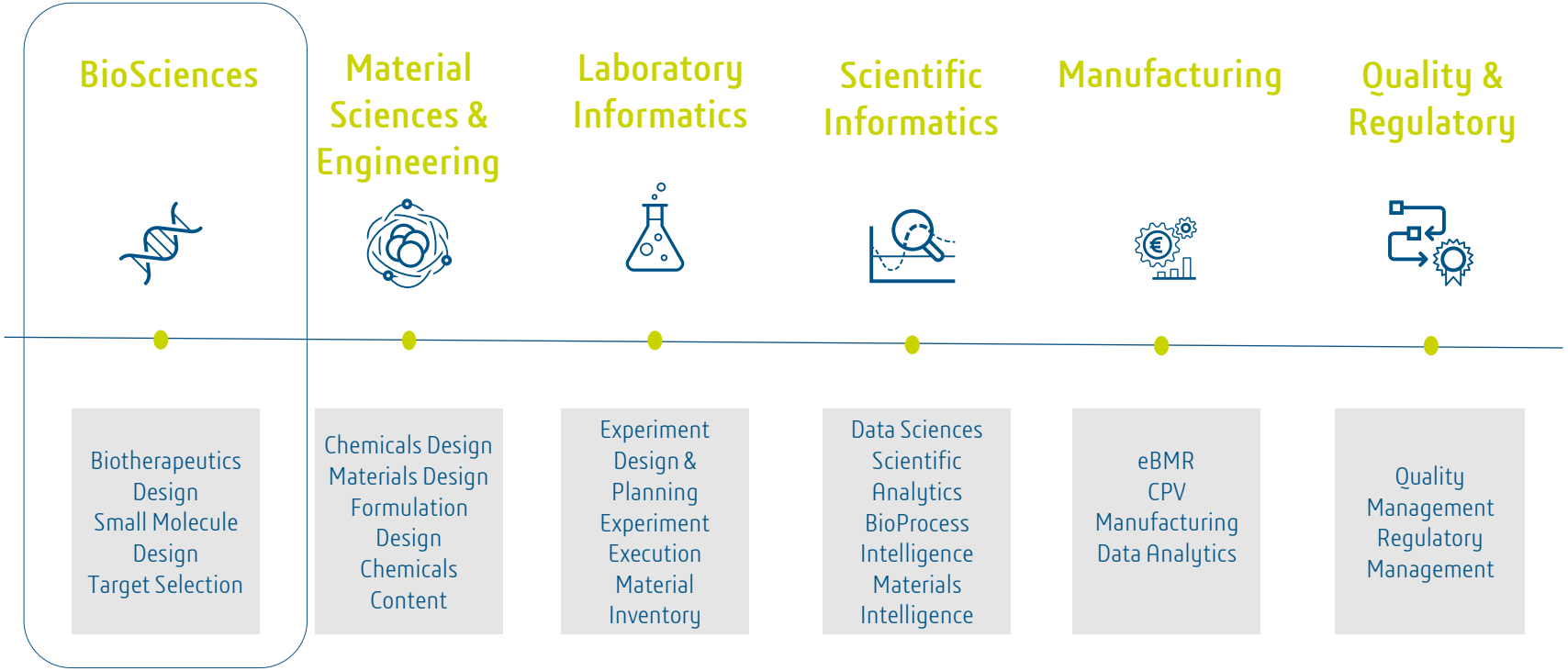
DASSAULT SYSTEMES | The 3DEXPERIENCE® Company



Copyright and Disclaimer

- Copyright © 2024 GGA corp. All rights reserved.
  - This presentation and/or any related documents contains statements regarding our plans or expectations for future features, enhancements or functionalities of current or future products (collectively "Enhancements"). Our plans or expectations are subject to change at any time at our discretion. Accordingly, GGA Corp. is making no representation, undertaking no commitment or legal obligation to create, develop or license any product or Enhancements.
  - The presentation, documents or any related statements are not intended to, nor shall, create any legal obligation upon GGA Corp., and shall not be relied upon in purchasing any product. Any such obligation shall only result from a written agreement executed by both parties.
  - In addition, information disclosed in this presentation and related documents, whether oral or written, is confidential or proprietary information of GGA Corp.. It shall be used only for the purpose of furthering our business relationship, and shall not be disclosed to third parties.

BIOVIA SCIENTIFIC PLATFORM



BIOVIA

DASSAULT SYSTEMES

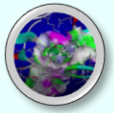
GENERATIVE THERAPEUTICS DESIGN PROCESS

Head of Research

Define project goals; track project progress

Molecular Modeler

Explore new targets; build and publish 3D predictive models



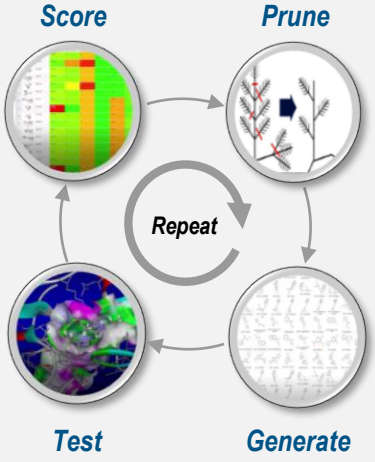
Data Scientist

Train and publish ML models

Virtual

Medicinal Chemist

Create virtual designs



Medicinal Chemist / Project Team

Assess synthetic accessibility

Decide which compounds to take to the lab

Automated re-training of ML models

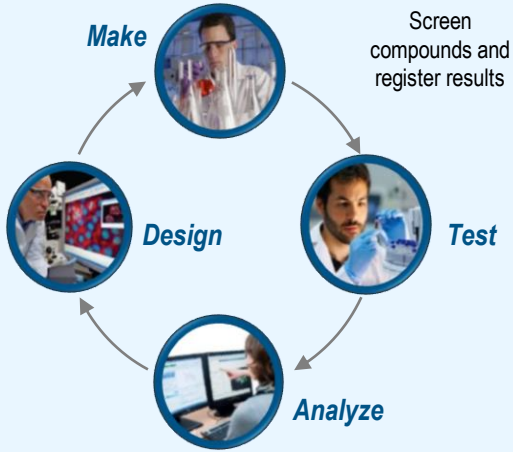
Real

Medicinal / Synthetic Chemist

Synthesize and register compounds

Biologist

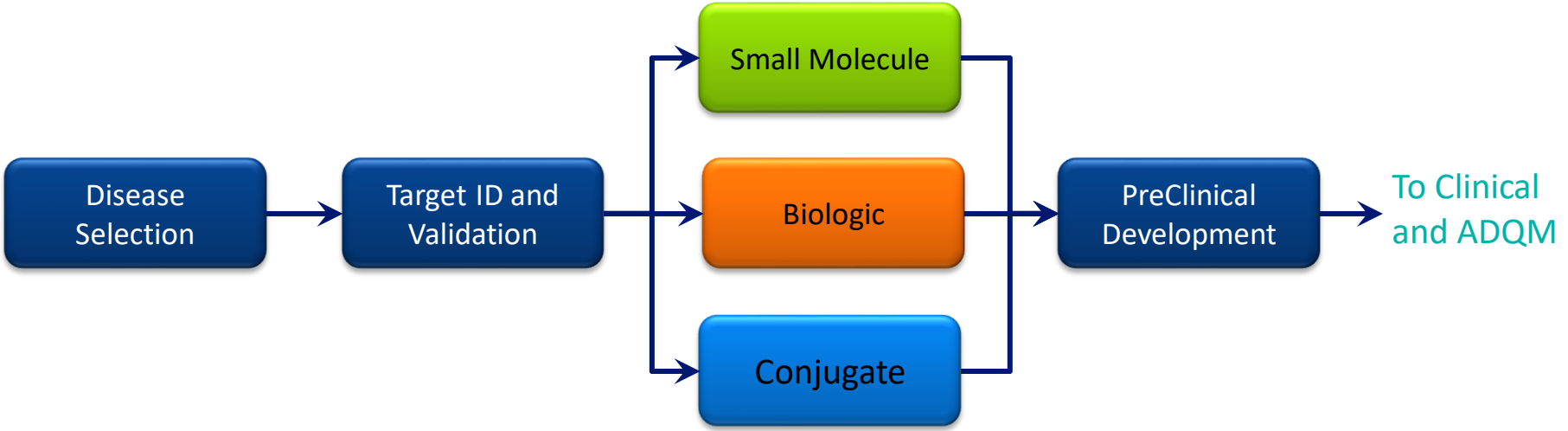
Screen compounds and register results



BIOVIA

DASSAULT SYSTEMES

Life Sciences R&D Top Level Workflow

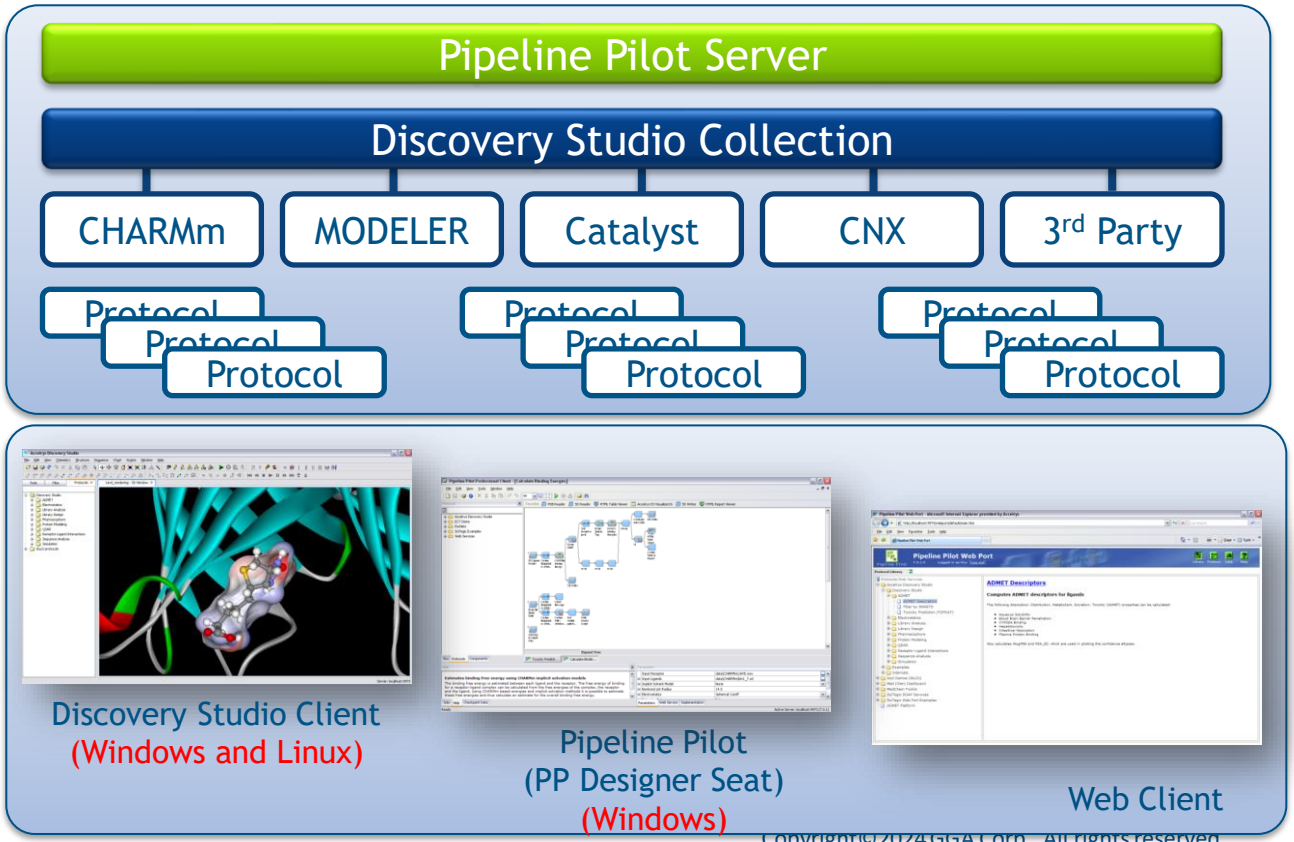


Discovery Studio Product Architecture

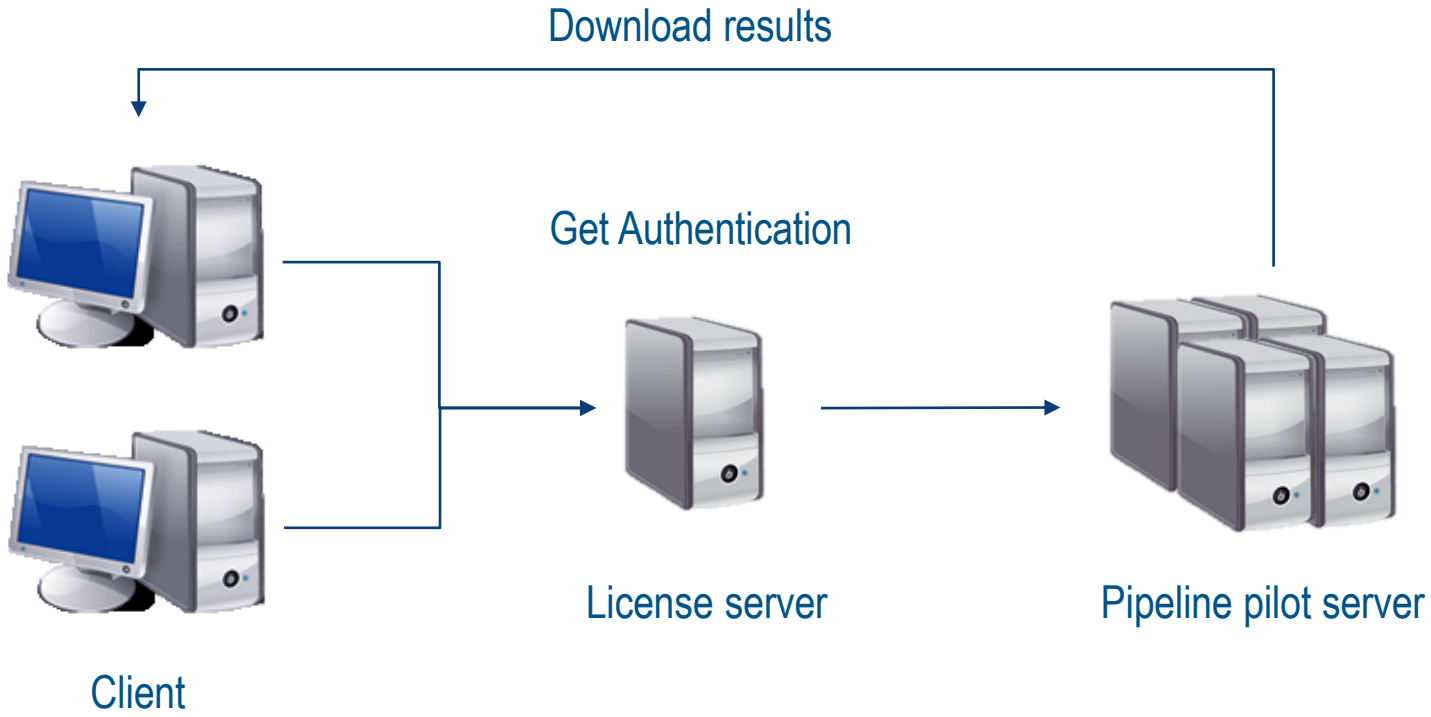
- Discovery Studio science runs on Pipeline Pilot:

- Server
  - AEP
  - PP Client
  - AEP Collections
  - DS collection
  - DS engines

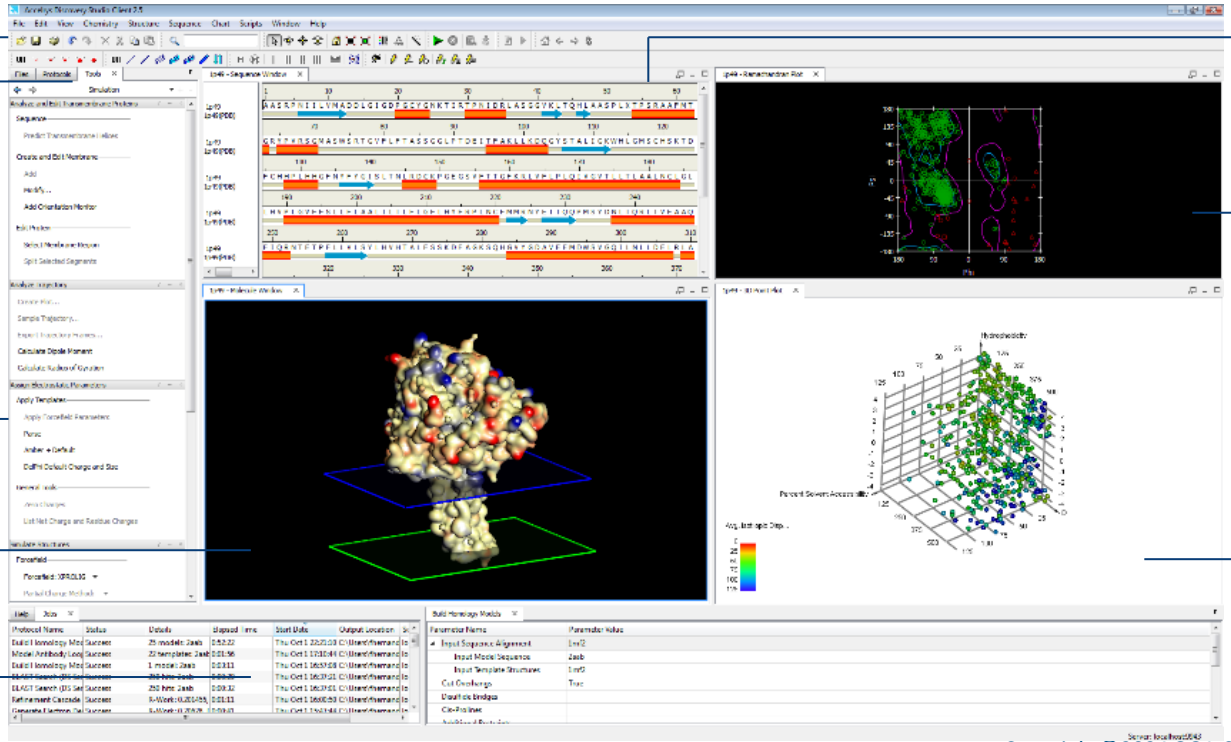
- Client(s)
  - DS client
  - PP Designer Seat
  - Web client



Discovery Studio Product Architecture

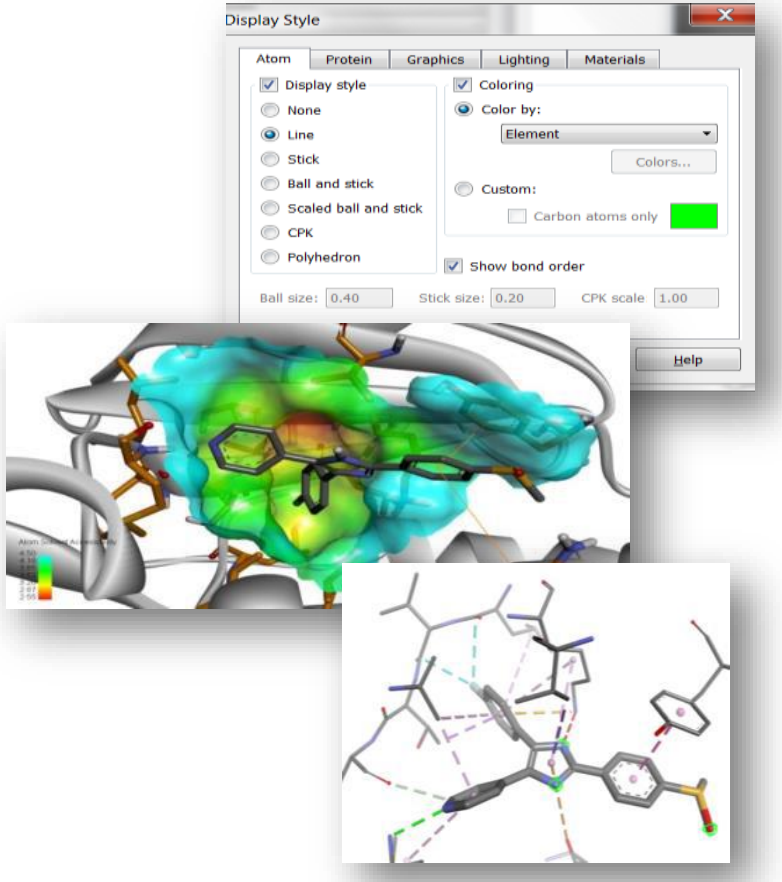


Supporting technology: Discovery Studio Client



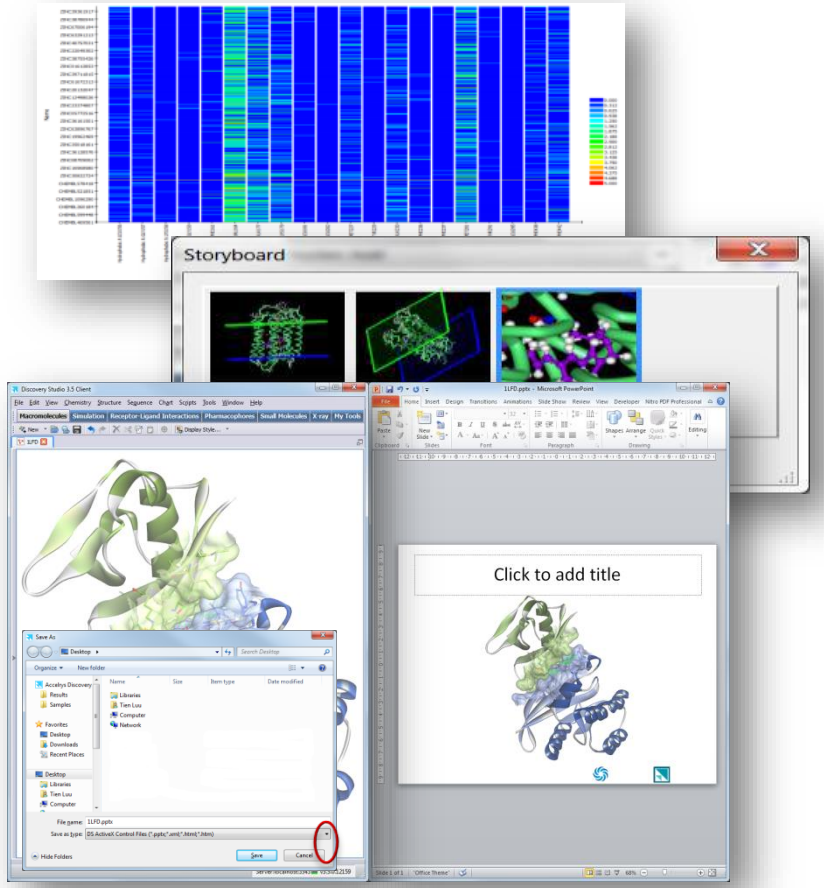
Discovery Studio Client

- General
  - Display Style
    - Ball and Stick, CPK, Solid Ribbon
    - Surface/Protein Surface
    - Shadow
    - Clipping Plane
  - Monitors
    - Distance, Angle, and Bump
    - Interactions
    - RMSD



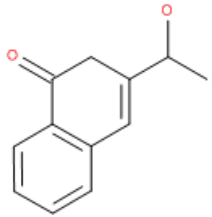
Discovery Studio Client

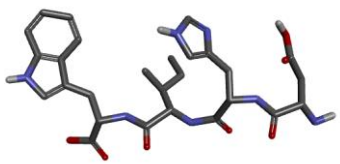
- General
  - Sequence view
  - Secondary structure prediction
  - Chart functions and data plots
    - Line plots, point plot, heat maps. etc
  - Collaboration, presentation functions
    - Story board, Active X
- Protocols
  - Automation and customization
  - Perl scripting



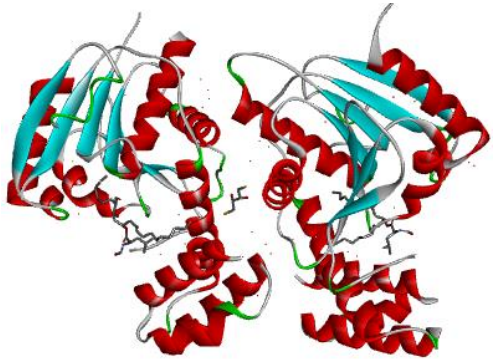
Data integration

- Open exist file
  - Chemical structure format
  - Sequence format
  - Protein structure format
- Download from database
  - Protein Data Bank (PDB)
  - NCBI Entrez Sequence Search
- Create new one
  - Small molecule, DNA, RNA, Peptide



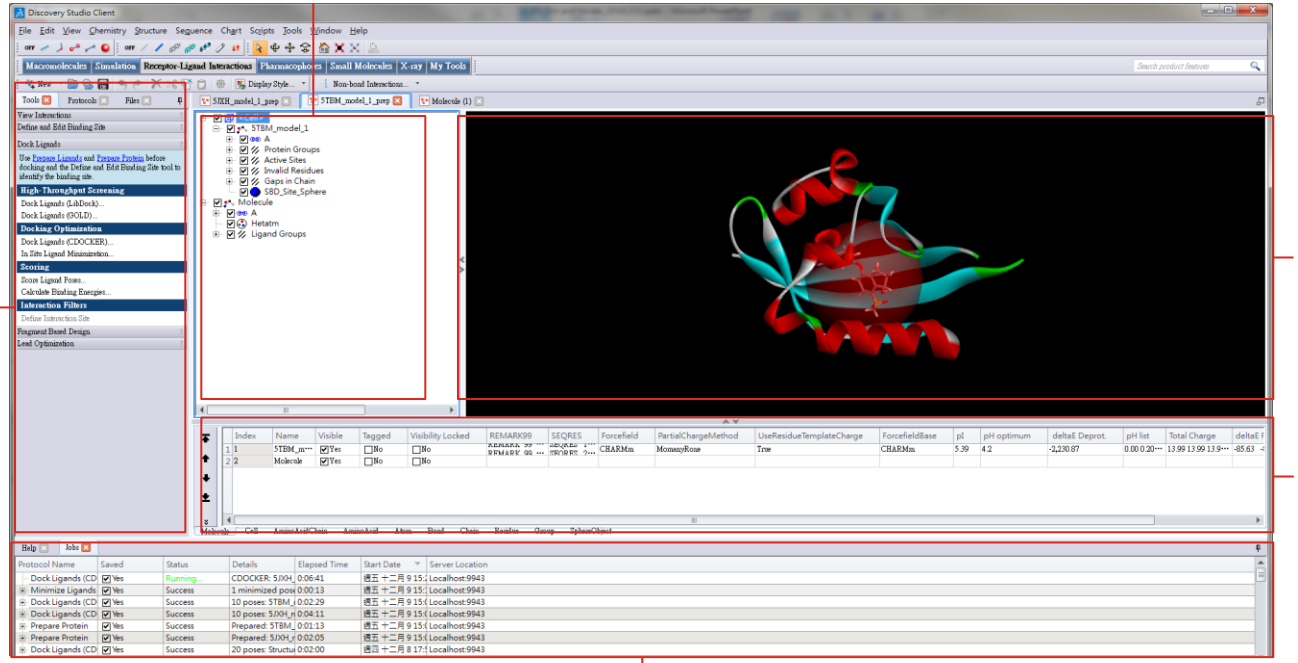


```
>sp|P05067|A4_HUMAN Amyloid beta A4 protein OS=Homo sapiens GN=APP PE=1 SV=3
MLPGLALLLLAAWTARALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTK
TCIDTKEGILQYCQEVYPELQITNVVEANQPVTIQNWCKRGRKQCKTHPHFVIPYRCLVG
EFVSDALLVPDKCKFLHQERMDVCETHLHWHTVAKETCSEKSTNLHDYGMLLPCGIDKFR
GVEFVCCPLAEESDNVDSADAEEDDSDVWWGGADTDYADGSEDKVVEVAEEEEVAEVEEE
EADDDEDDEDGDEVEEEAEEPYEEATERTTSIATTTTTTTESVEEVVREVCSEQAETGPC
RAMISRWYFDVTEGKCAPFFYGGCGGNRNNFDTEEYCMAVCGSAMSQSLLKTTQEPLARD
PVKLPTTAASTPDAVDKYLETPGDENEHAHFQKAKERLEAKHRERMSQVMREWEEAERQA
KNLPKADKKAVIQHFQEKVESLEQEAANERQQLVETHMARVEAMLNDRRRLALENYITAL
```



DS Client - Windows

Hierarchy window



Tools
Protocols
Files

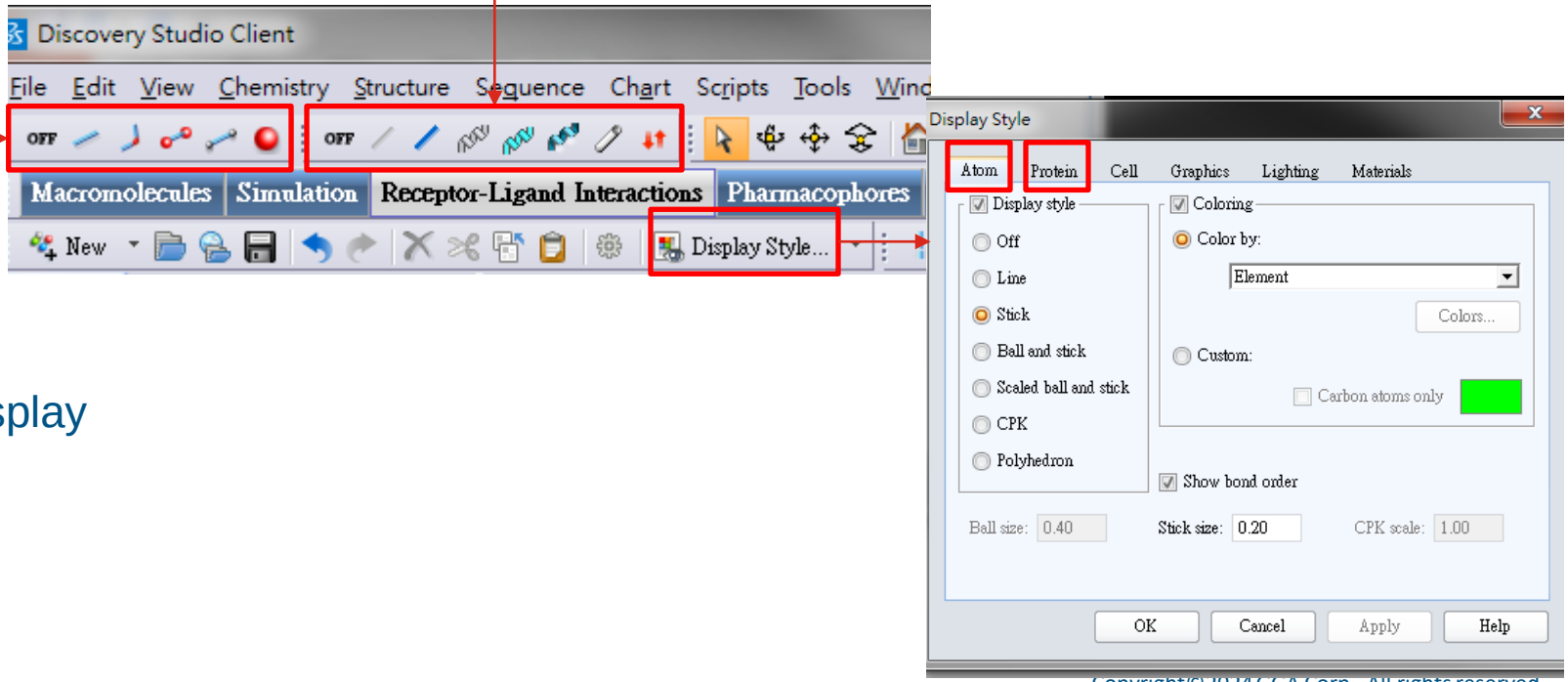
Molecule window

Table window

Job window

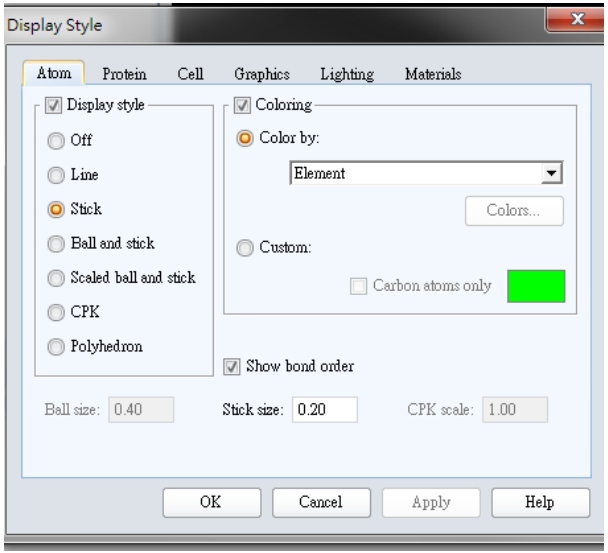
Display style

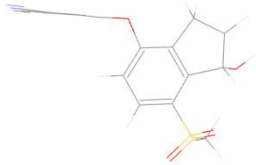
Protein display



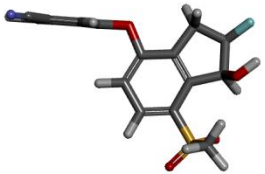
Atom display

Atom display

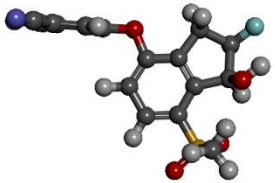




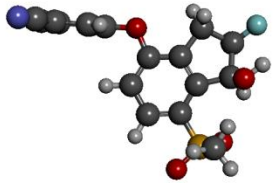
Line



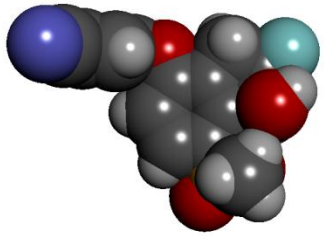
Stick



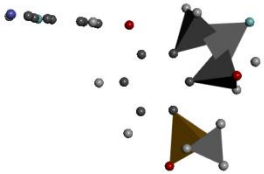
Ball and Stick



Scaled ball and stick

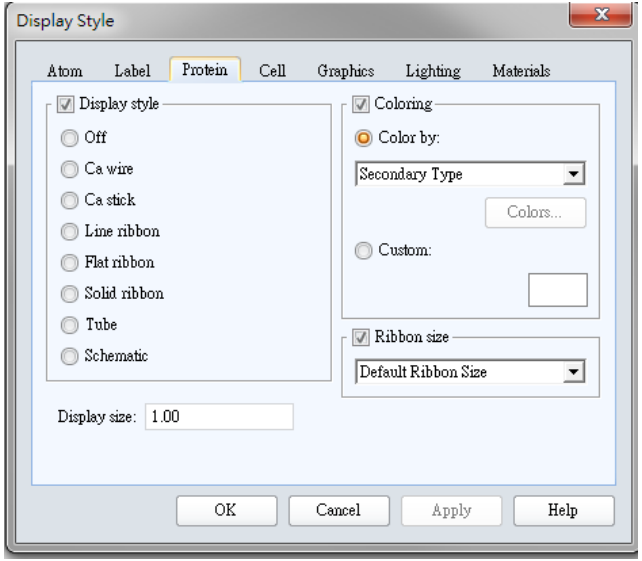


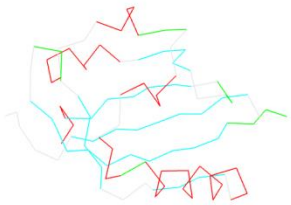
CPK



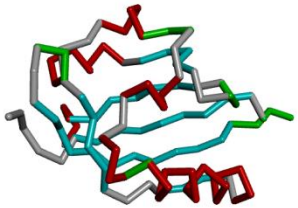
Polyhedron

Protein display

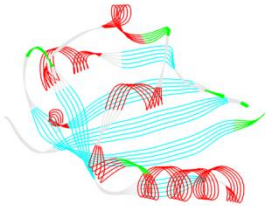




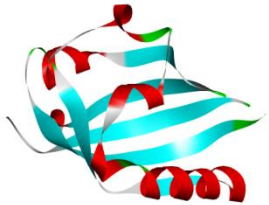
Ca wire



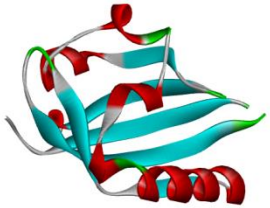
Ca stick



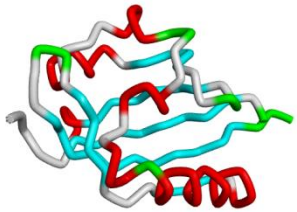
Line ribbon



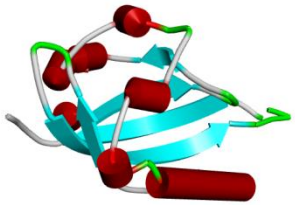
Flat ribbon



Solid ribbon

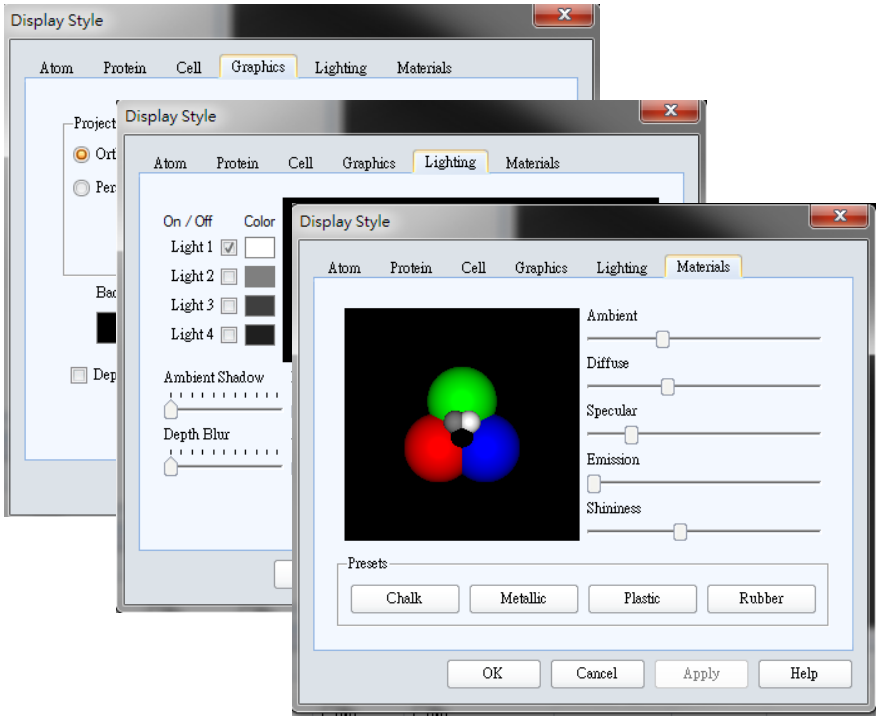


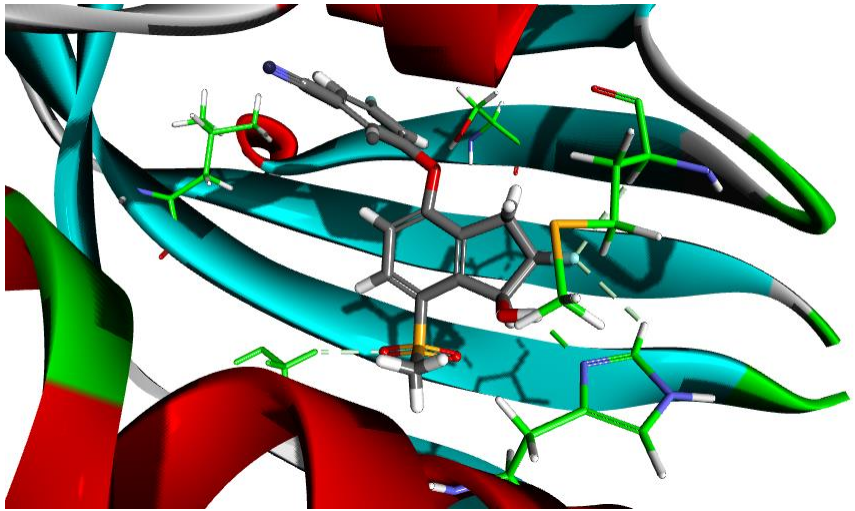
Tube



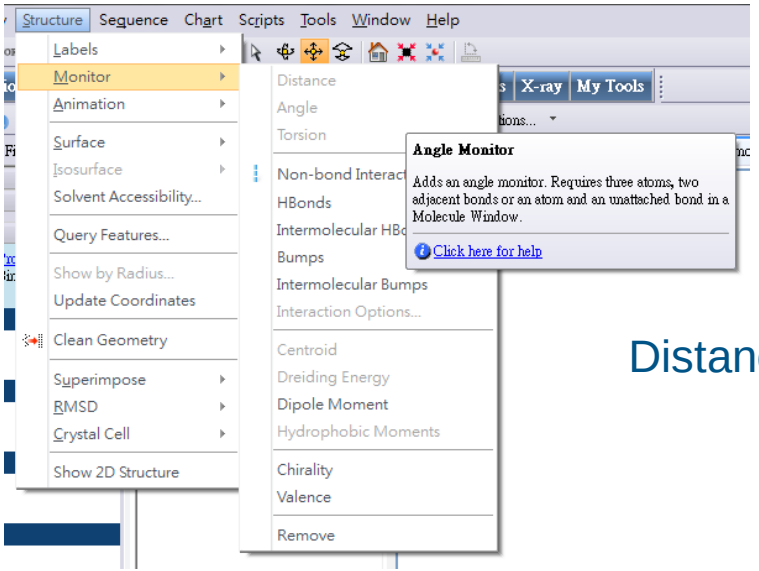
Schematic

Display style cont.

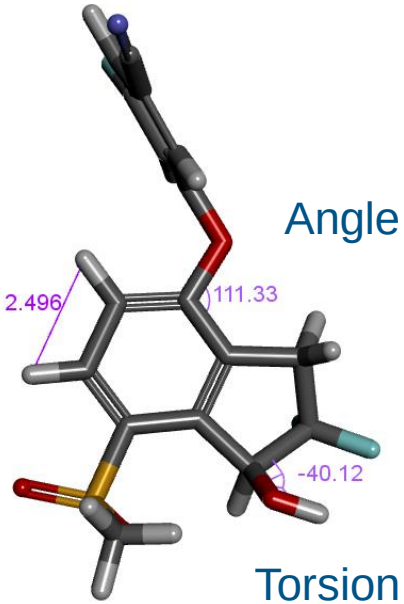




Monitor and Non-bond Interactions tools



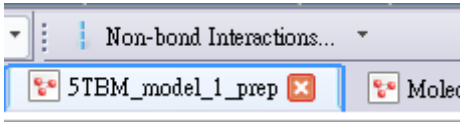
Distance

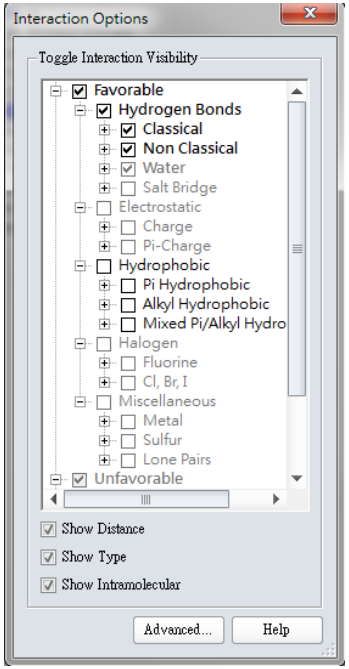


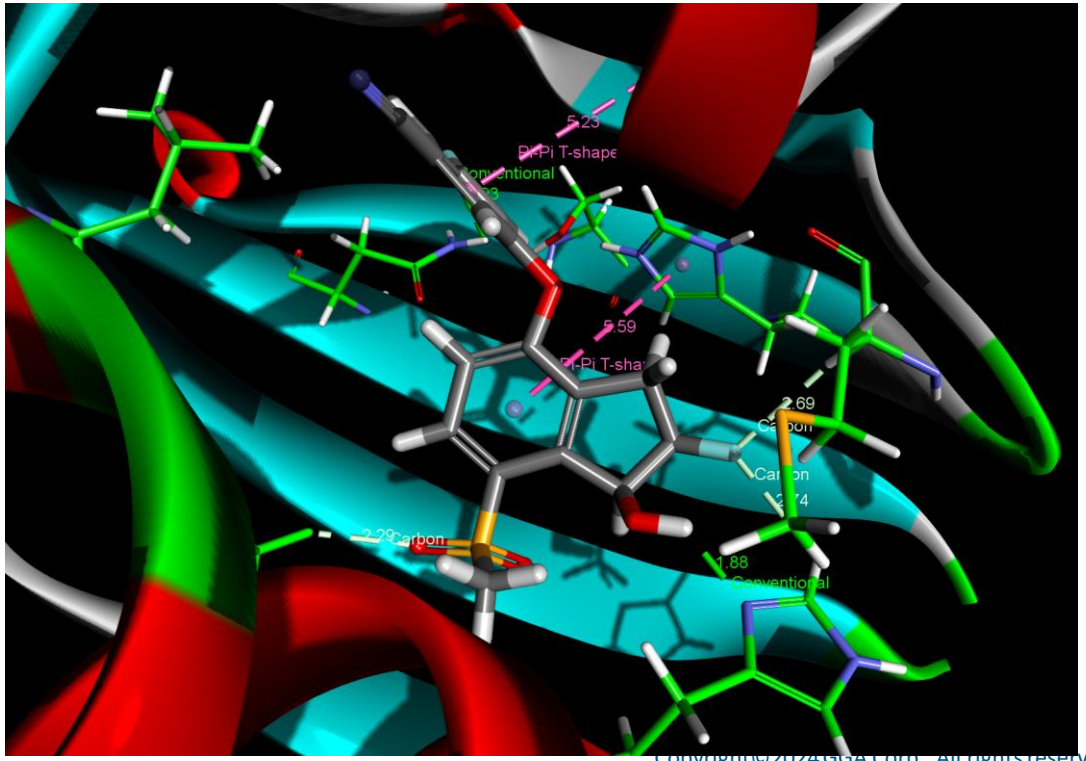
Angle

Torsion

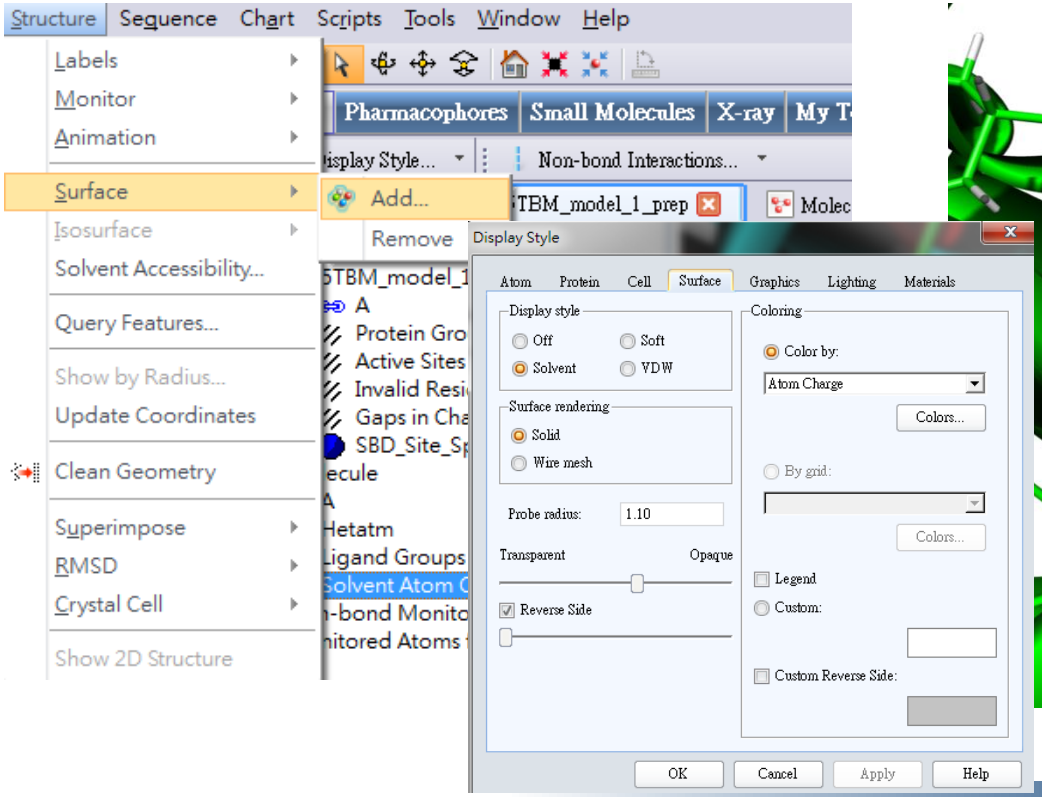
Monitor and Non-bond Interactions tools

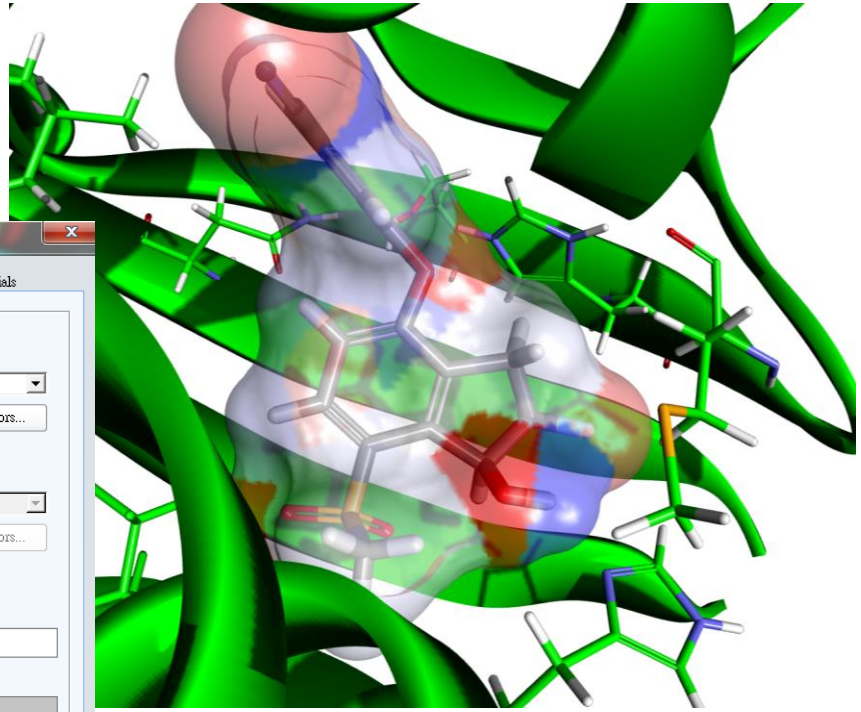




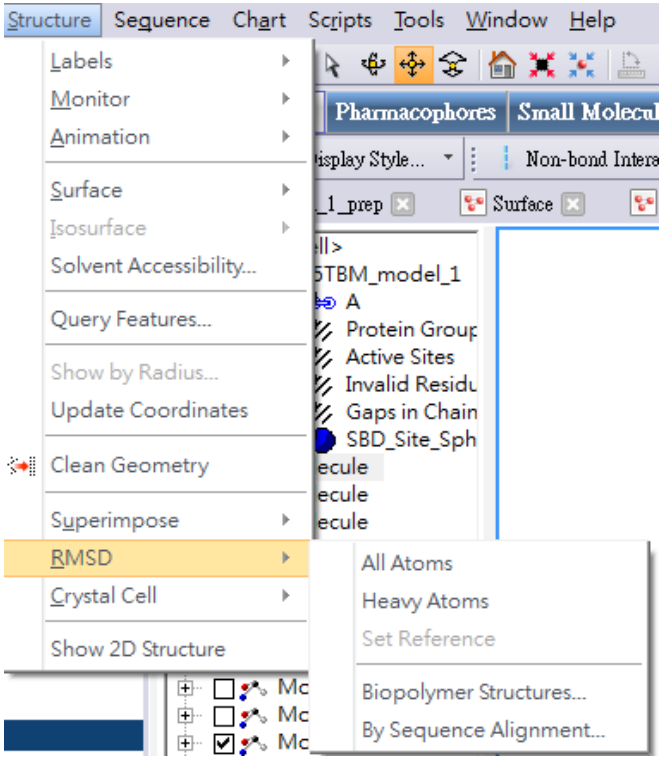


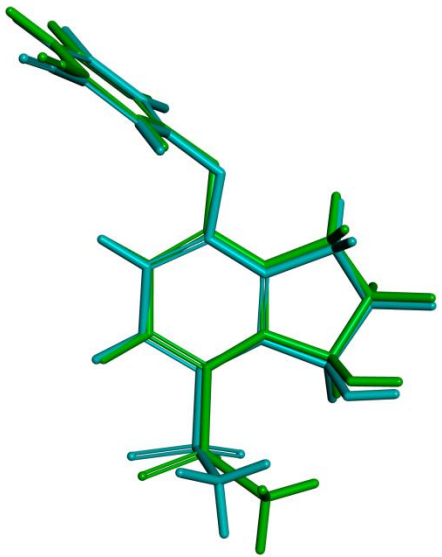
Surface





Overlay/Superimpose/RMSD



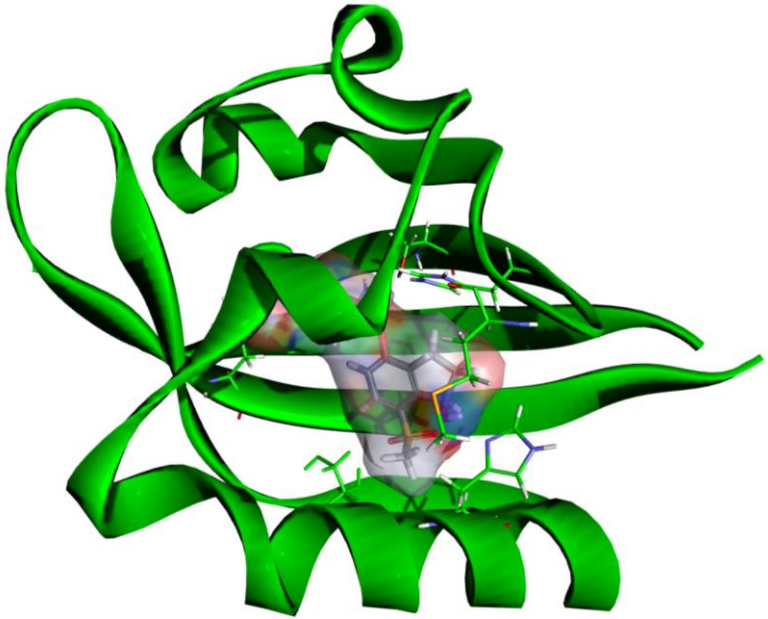


Heavy Atom RMSD to Molecule 2

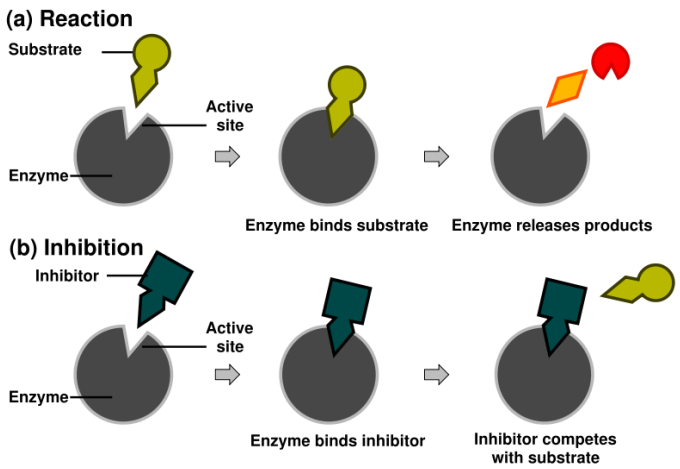
Name	Reference	RMSD (A)
Molecule 1	Molecule 2	0.4219

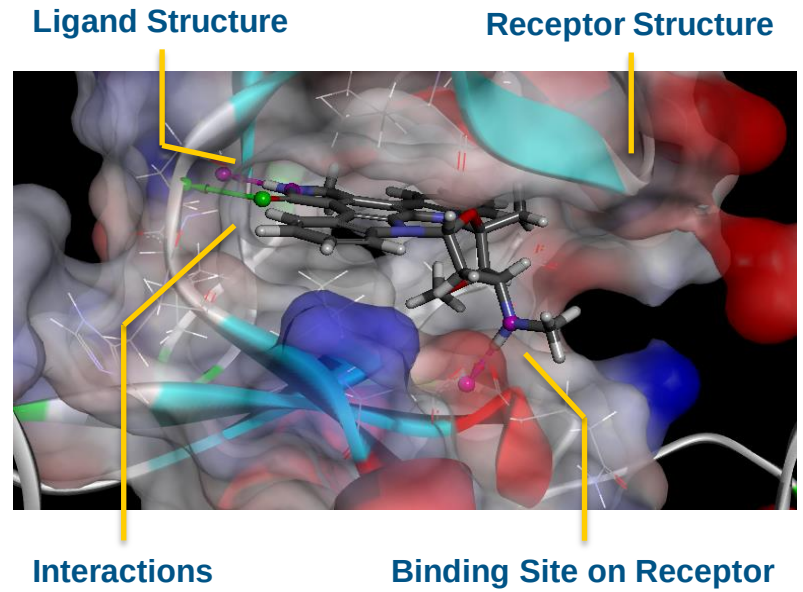
Hands-on

- PDB code: 5TBM (HIF2 alpha)
- Atom display:
  - His248, Met252, His293
  - 79A401 (PT2385)
- Protein display
  - Solid ribbon in Green color
- Surface
  - 79A401 (PT2385)

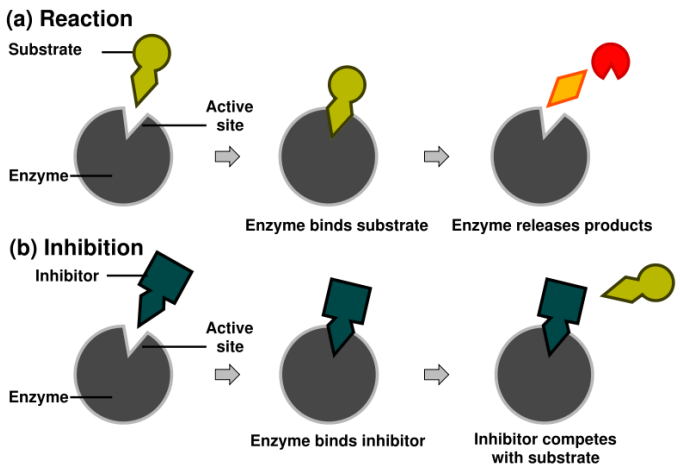


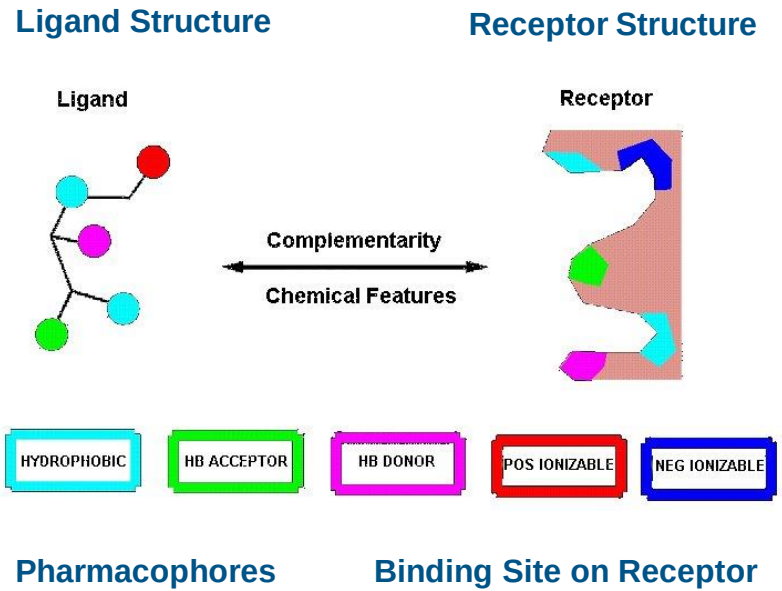
Docking & Scoring – Structure-Based Design



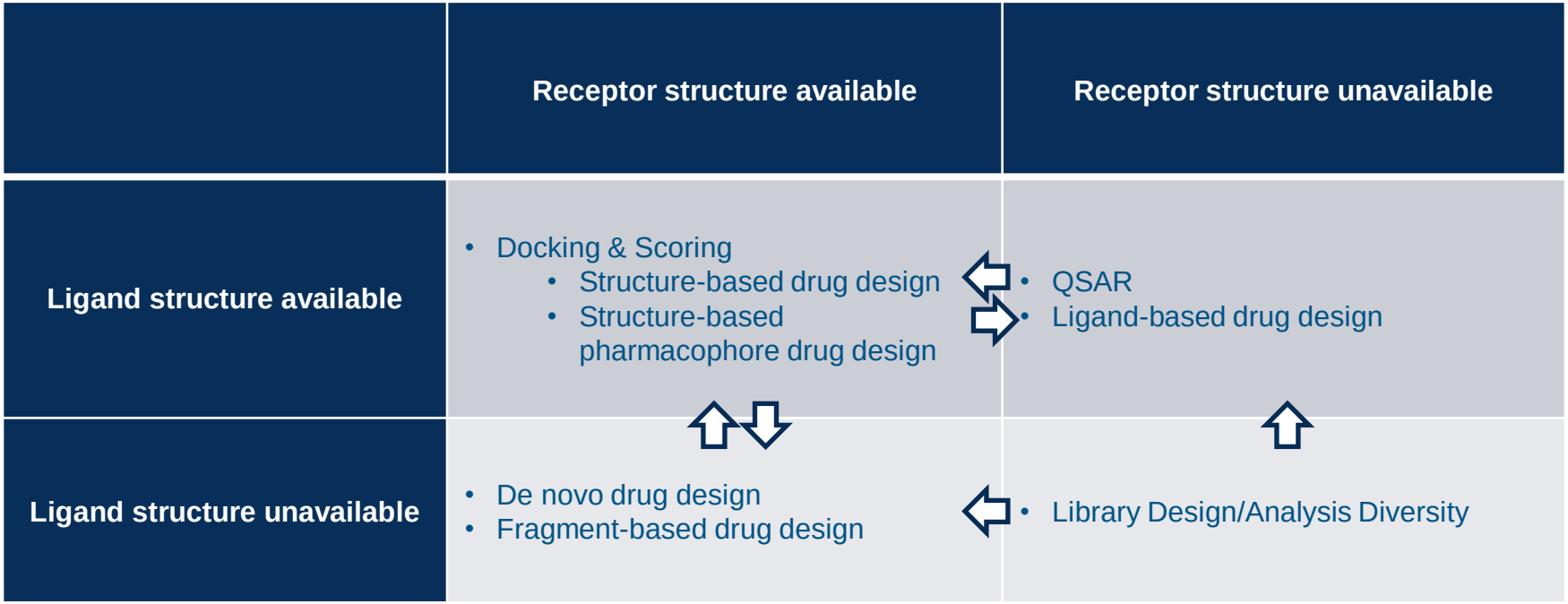


Docking & Scoring – Pharmacophore-Based Design





Strategy for Small Molecule Drug Design



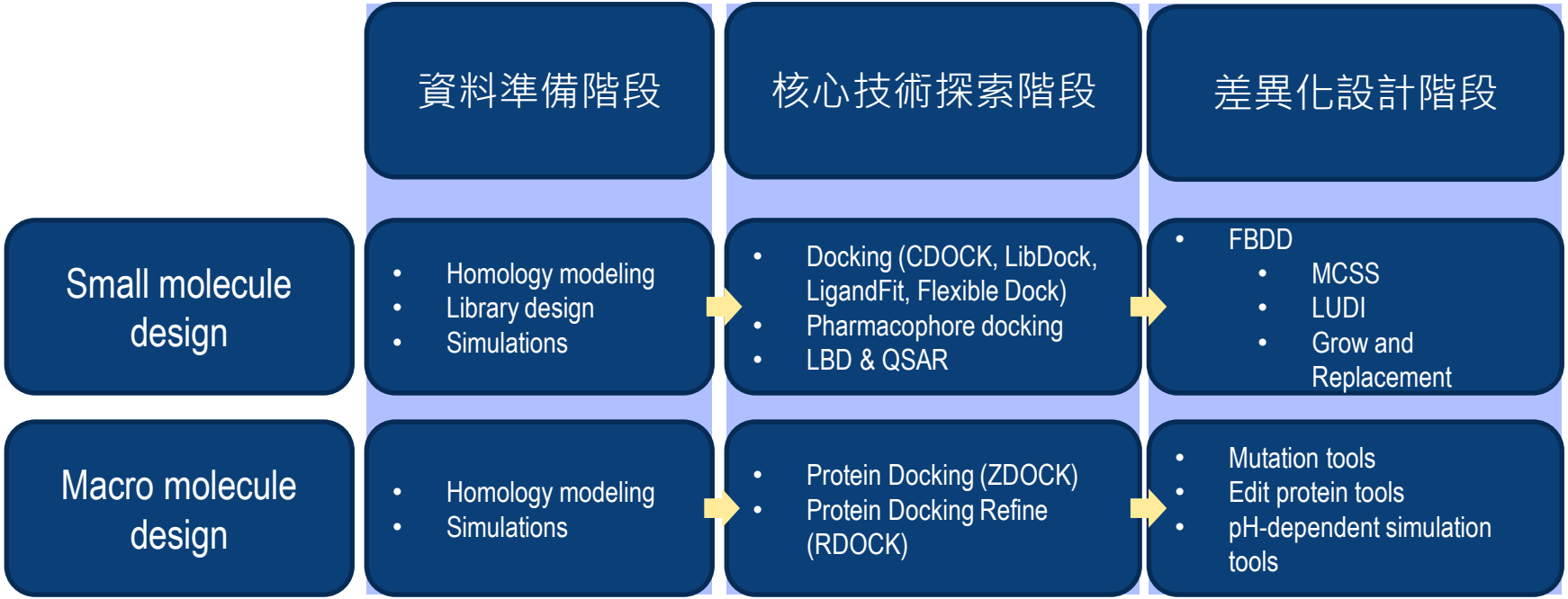
Strategy for Small Molecule Drug Design

	Receptor structure available	Receptor structure unavailable

Ligand structure available	• **Docking & Scoring**
	  • **Structure-based drug design**
	  • **Structure-based pharmacophore drug design**
		• QSAR
		• Ligand-based drug design

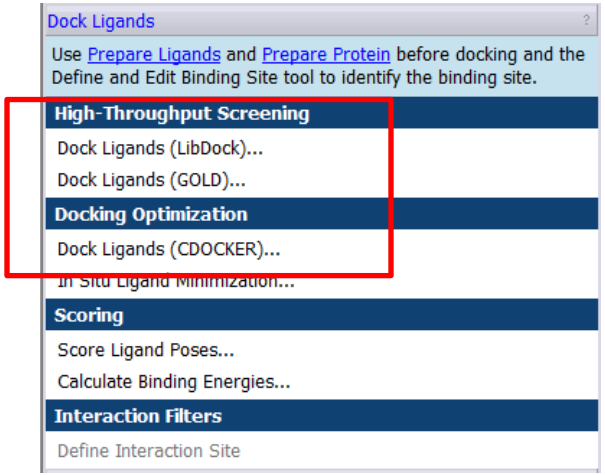
Ligand structure unavailable	• De novo drug design • Fragment-based drug design	• Library Design/Analysis Diversity

Product design stages

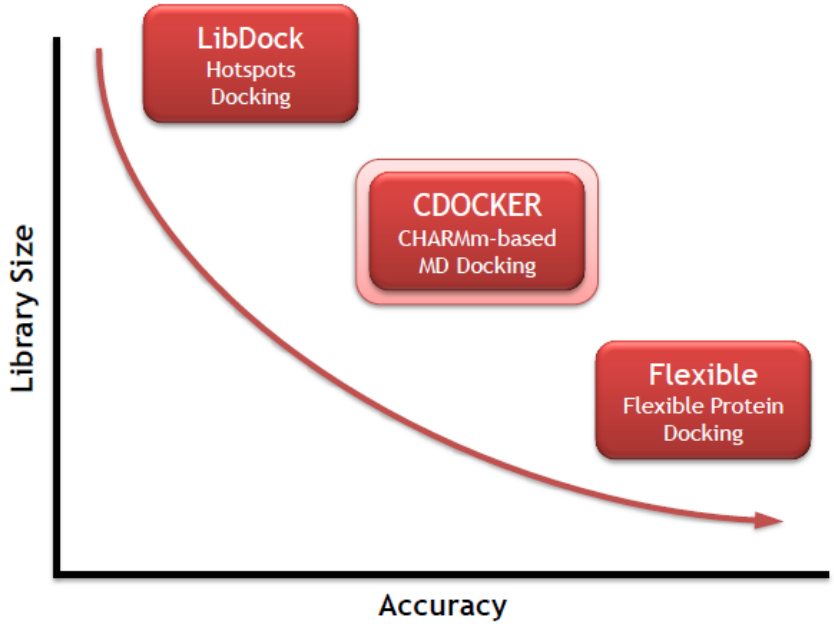


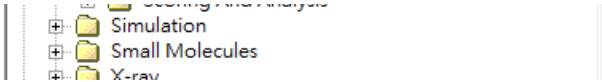
Docking from Tool Panel

- Docking tools in DS
  - Dock Ligands (CDOCKER)
  - Dock Ligands (LibDock)
  - Dock Ligands (GOLD)



- Docking tools in DS (Protocol)





Comparative performance of LibDock and CDOCKER on AstexDiverse dataset

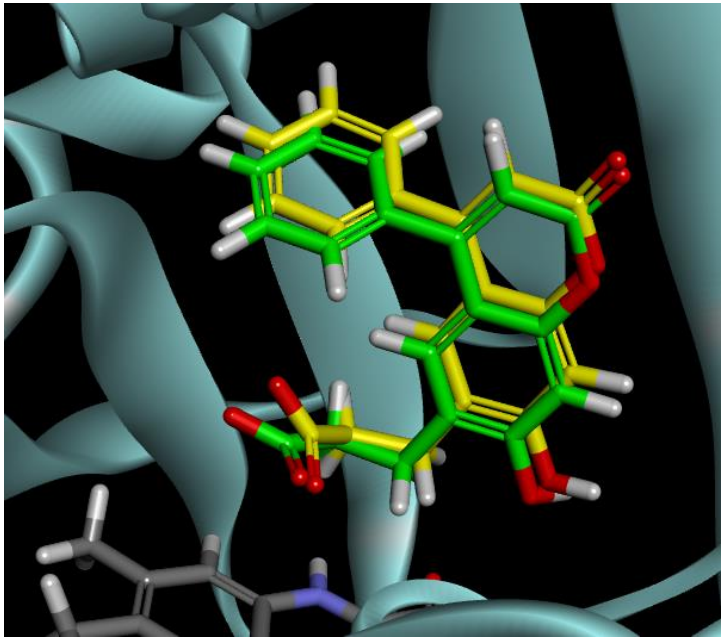
Method	% Docked Accurately (Best)	% Docked Accurately(Top)	RMSD Average (Best)	RMSD Average (Top)	Tim(min)
LibDock	91	50	1.2	3.8	0.5
CDOCER	94	79	0.8	1.5	5.0

LibDock is optimized for speed:
get accurate docked poses in seconds

CDOCKER is optimized for accuracy:
get significant improvement in rank-ordering of correct pose and RMSD to X-ray structure

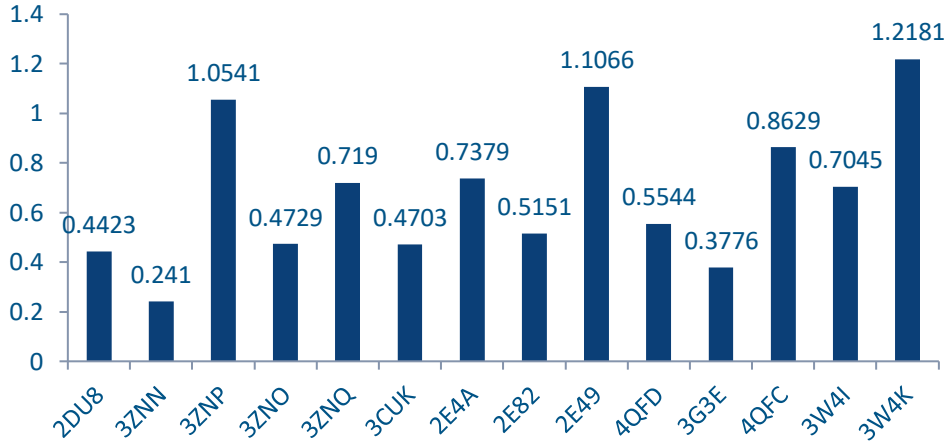
1. Hartshorn, et al. J. Med. Chem., 50 (4), 726 -741 (2007)
3 GHz

Docking Test for D-Amino Acid Oxidase Inhibitors



Green: Crystal conformation
Yellow: Calculated conformation

- 14 Crystal DAO Structures for the docking test
- Average RMSD: 0.6769Å



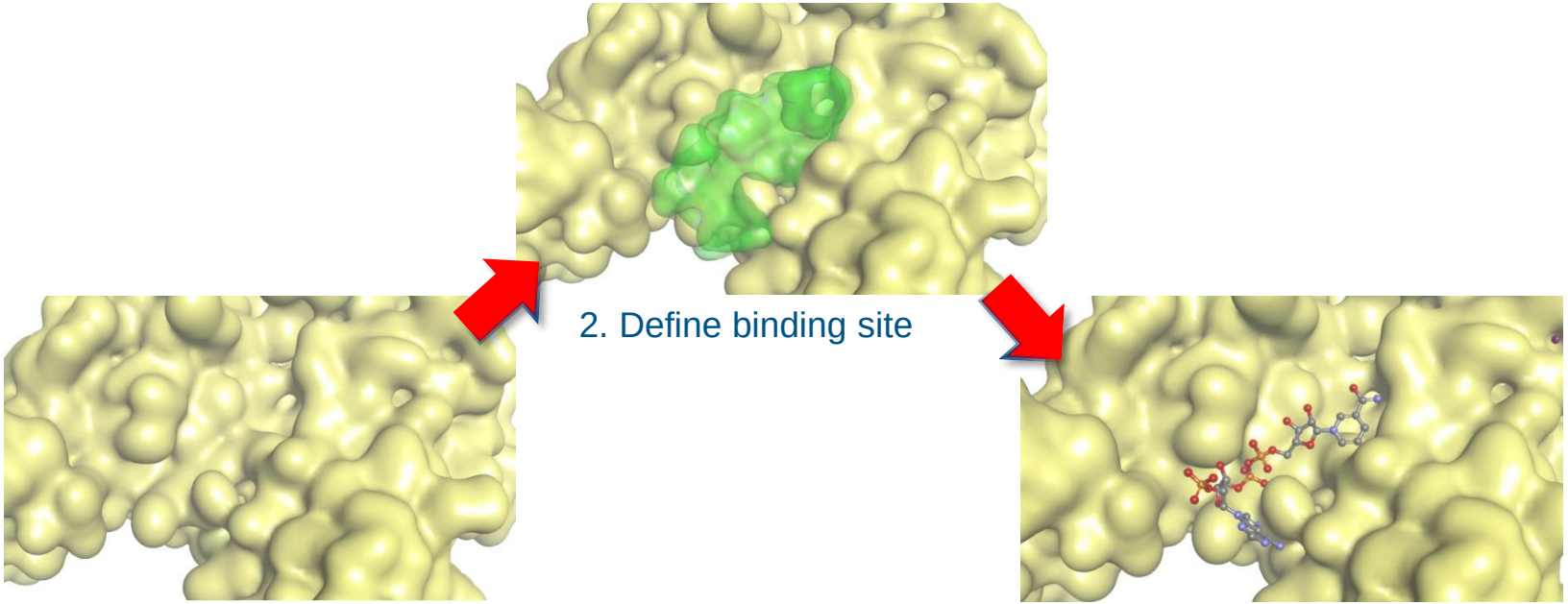
Structure-Based Drug Design

- Using knowledge of a receptor to guide design of new ligands
  - Structure of the receptor is known
  - Can use either an experimental or homology model

- One approach is to identify potential ligands that can bind to receptor
  - High-throughput virtual screening
  - Rigid or flexible docking
  - Scoring of docked ligands

- May need to identify the binding site
  - Active site search

Illustration for Structure-Based Design and Analysis



1. Obtain receptor and ligand structures

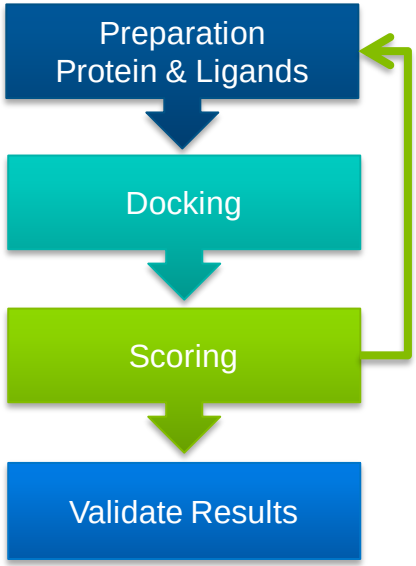
2. Define binding site

3. Dock ligand and validate results

Workflow of Virtual High Throughput Screening

- Standardize molecules
- Cluster subsets

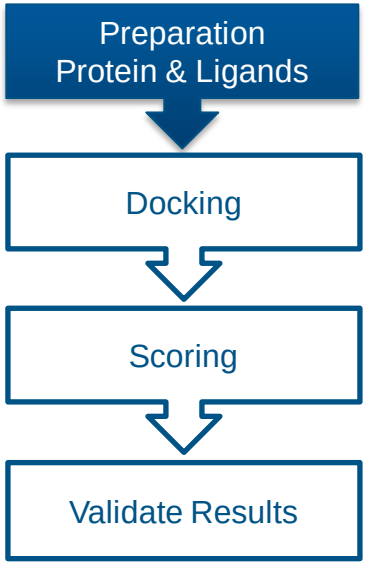
Docking Workflow

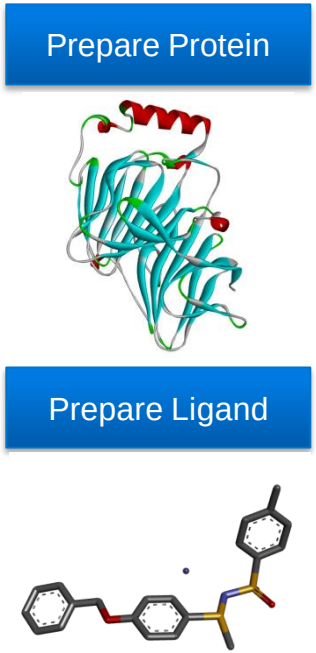


Repeat Cycle = Virtual High Throughput Screening (vHTS)

- In order to maximize the rate of success, the evaluation phase in each steps is very important.
- Requirements
  - Known active compounds and decoy
  - Bound molecules
  - Optimize the parameters of docking technique
  - Prioritized the molecules in library before proceeding the docking and scoring steps

Docking Workflow





Protein Preparation

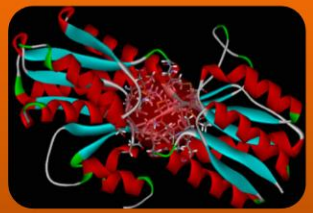
- Repair deficient structure
- Identify binding site
- Modify the protonation state

Ligand Preparation

- Add hydrogen atom
- Generate 3D structure coordinates
- Creating isomers
- Remove duplicates
- Valence modification
- Standardization charges

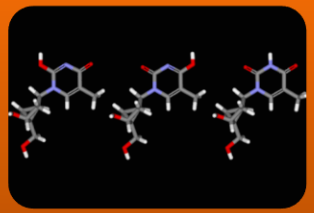
SBD: Input Preparation

Proteins



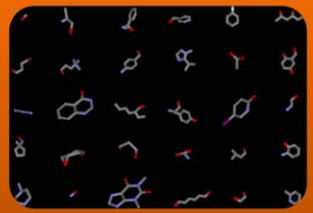
- Standardise atom names
- Insert missing atoms in residues
- Remove alternate conformations
- Insert missing loops
- Optimize short & medium size loops
- Calculate pK and protonate

Ligands



- Add hydrogens
- Calculate 3D coordinates
- Enumerate ionization states
- Ionize functional groups
- Generate tautomers and isomers
- Remove duplicates
- Fix bad valencies
- Standardize charges for common groups
- Retain largest fragment

Fragments



- Generate fragments using RECAP* rules
- Rule of Three filters

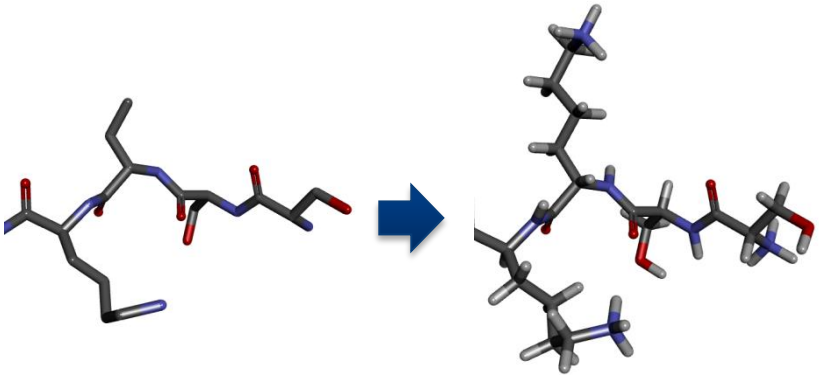
Selecting the Protein Receptor

- Choosing the protein receptor
- Probably have a receptor in mind already
  - Based on biological or medical problem
  - Reinforced by biological data
- Requires a three-dimensional structure of the receptor
  - X-ray crystal structure
  - NMR structure
  - Homology model
- Can be an apo form of receptor

Protein Preparation

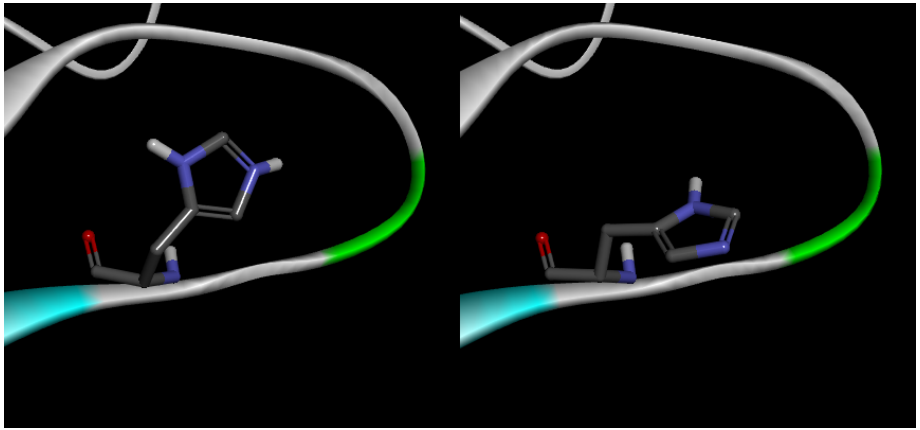
- Before any docking can be performed, the receptor must be properly prepared
  - Particularly a concern with PDB files

- Preparation includes having...
  - All residues completed
  - Correct chemistry
    - Correct bond orders
    - Correct atom valences
  - All required hydrogen atoms added
  - Correct formal charges



Protonation state

- Calculate pKa for each residue in different pH
- Modify the protonation (ionization) status

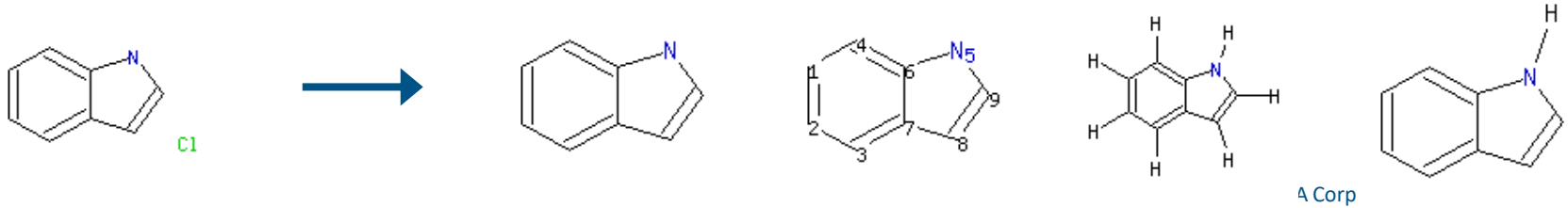


Example for His334 of Protein A in different protonation status

Protein A	pH7.2	pH8.0
HIS334	1	0
HIS376	1	0
HIS1253	1	0
HIS1540	1	0
HIS1575	1	0
HIS1591	1	0
HIS1937	1	0
HIS2391	1	0
HIS2452	1	0
HIS2455	1	0
HIS2488	1	0
HIS2500	1	0
HIS2625	1	0
HIS2826	1	0
HIS2865	1	0
HIS2898	1	0
HIS2946	1	0
HIS2998	1	0
GLU1018	0	-1

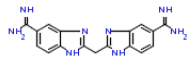
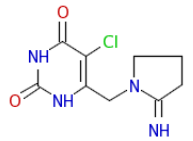
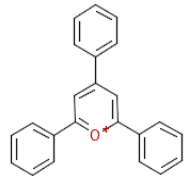
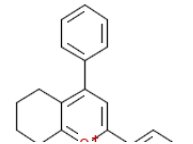
Standardize molecules

- Beautification
  - Keep/Remove largest/smallest fragments (option of Standardize Molecule)
  - Add/Remove atom numbers (found in Utilities)
  - Add/Remove hydrogens
  - Add hetero hydrogens (option of Add Hydrogens component)
  - Center molecule (option of Standardize Molecule)



Standardize molecules cont. - Strip salts

- *Strip Salts* will strip a parent molecule of its counter ions.
- Chemistry Data\Queries\Salts.sd contains any defined salt structure
- User defined salt queries can be added via parameter User Salts
- Further components: *Identify Salts* and *Generate Salts*

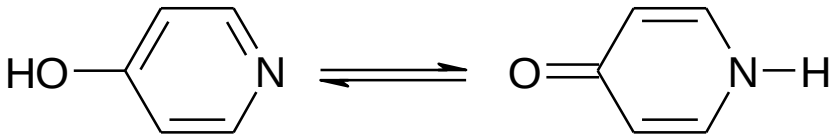
Molecule	CODE	RemovedSalts
	SPB 08110	HCl HCl HCl HCl H2O
	SPB 08153	HCl H2O
	TL 00001	[O-]S(=O)(=O)C(F)(F)F
	TL 00002	[O-]S(=O)(=O)C(F)(F)F

Ligand preparation

Performs the following steps, some of which can be controlled by the protocol parameters:

- Generate a canonical tautomer
- **Keep only the largest fragment**
- **Set standard formal charges on common functional groups**
- Kekulize the molecule
- **Enumerate ionization states at a given pH range (Optional)**
- Enumerate tautomers (Optional)
- Enumerate isomers (Optional)
- By default only unspecified bonds and atoms are enumerated
- **Remove duplicate structures (Optional)**
- Filter structures that violate Lipinski rules (Optional)
- Generate a standard 3D conformation (Optional)
- Catalyst is used to generate a reasonable 3D conformation

Tautomers



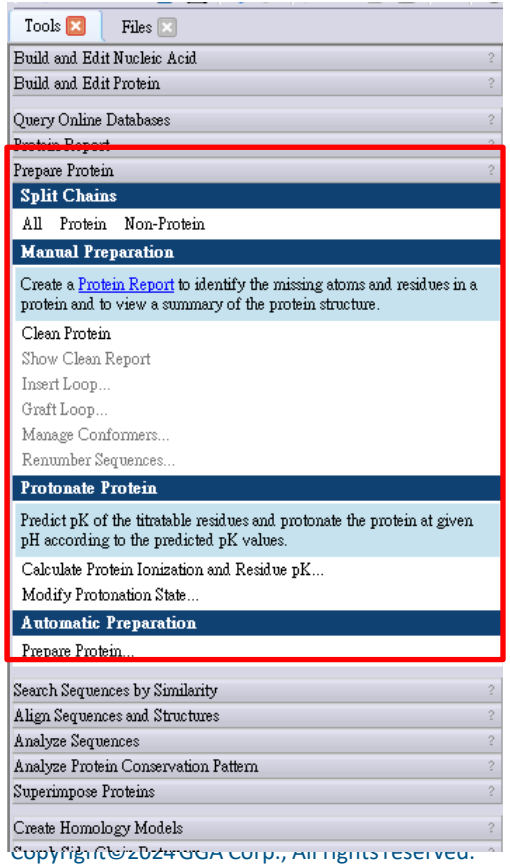
- Compounds whose structures differ markedly in arrangement of atoms, but which exist in easy and rapid equilibrium, are called tautomers.

  This means:

  – Possible duplicate molecules may be missed.
  – Different structures may have different values for calculated properties.

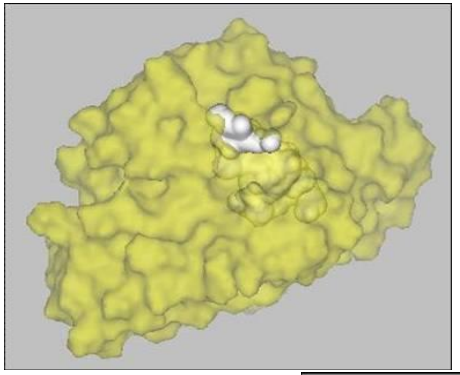
Protein Reports and Utilities Tools

- Allows you to:
  - Renumber sequences
  - Summarize information about the protein
  - Generate hydrophobicity plots
  - Split structures into distinct molecules
  - Clean protein molecules
  - Add missing atoms
  - Fix connectivity
  - Fix names
  - Define a template for a nonstandard amino acid



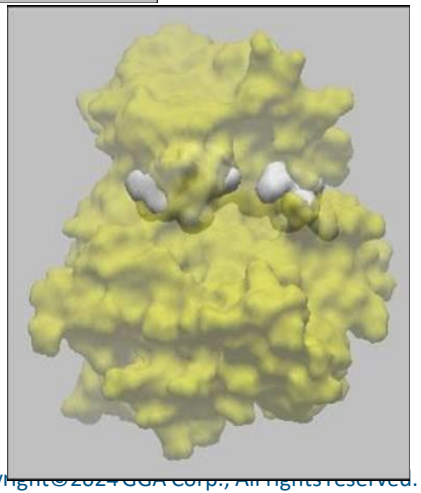
Binding Site Identification

- Liang et al. 1998 found small molecule binding sites to be:
  - Indentations, crevices, or cavities
  - And often the largest site is the true binding site
- Laskowski et al. 1996 reported an analysis of cleft volumes:
  - Often the ligand is bound in the largest cleft
  - Usually the largest cleft is considerably larger than the others
- Jones and Thornton 1997 found that protein-protein interaction sites tend to be:
  - Flat and hydrophobic



HSV-1 thymidine kinase

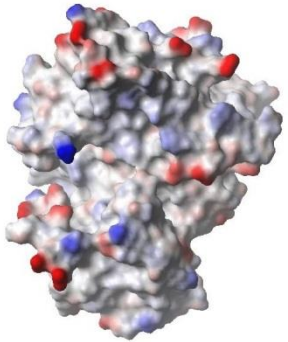
Abl tyrosine kinase



Binding Site Identification

- With experimental structure determination of ligand complexes...
  - Binding site is often identified
  - Analogues can show additional features
  - Use the position of known ligands to limit possible binding sites for new candidates

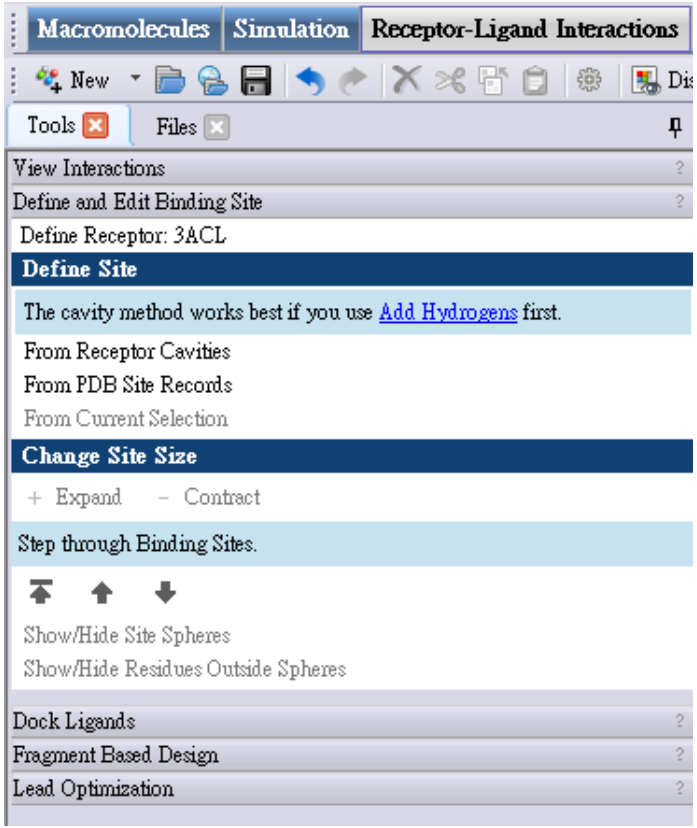
- With unbound proteins…
  - Binding site may not be obvious
  - Binding site must be sought



Binding Site Identification

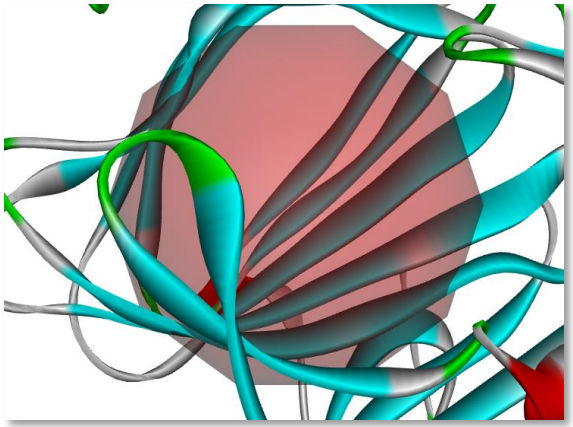
- When the binding site is unknown…
  - Search for cavities
  - Use experimental data
    - Site-directed mutagenesis studies
    - Cross-linking data
    - NMR results
  - Compare target to similar proteins

- Can be accomplished with Binding Site Tool panel

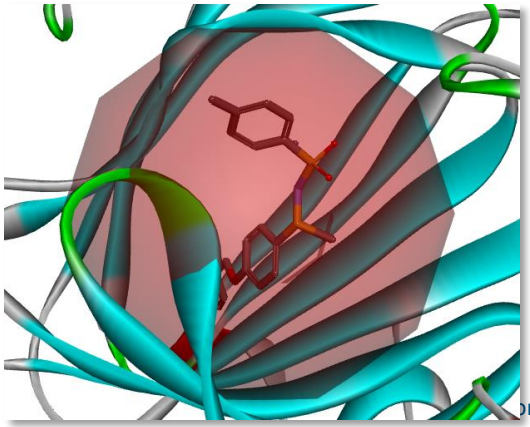


Site Search Approaches

- Protein Shape
  - Based on the shape of the protein only
  - Identifies cavities and crevices

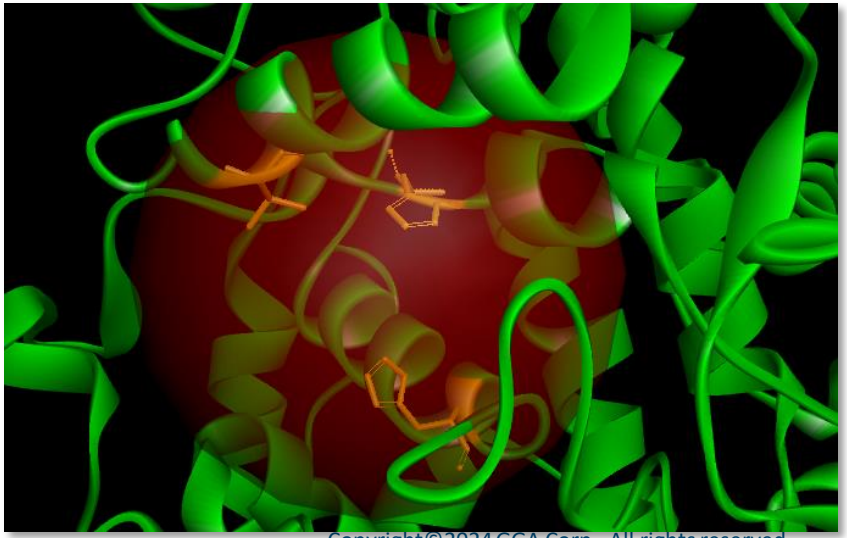


- Bound Ligand Volume
  - Requires presence of a bound ligand in protein
  - Identifies region around bound ligand



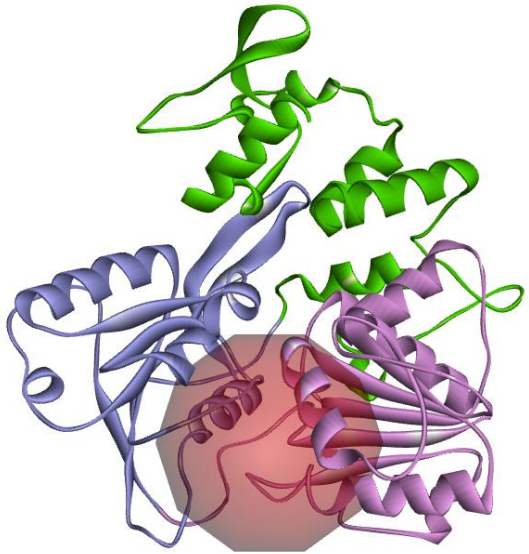
Site Search Approaches

- Use experimental data
  - Identified binding site from references
  - Site-directed mutagenesis studies
  - Compare target to similar proteins

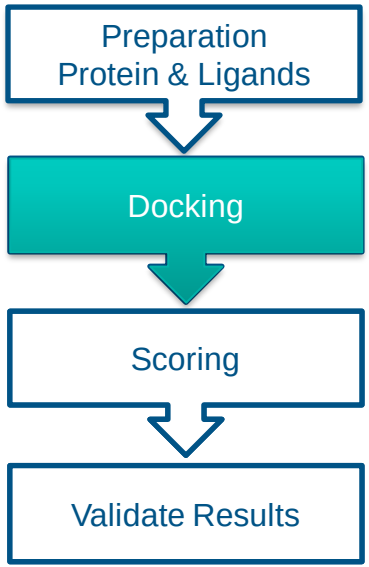


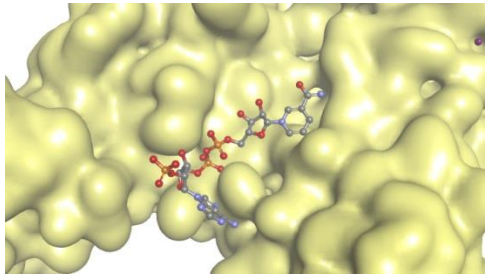
Demo

- Prepare Protein – 5JMT
- Prepare Ligand
- Define binding site
  - K200, T201, R202 (motif I, also called P-loop)
  - D285,E286 (Mn binding motif II)
  - Q455, R459, and R462 (motif VI)



Docking Workflow



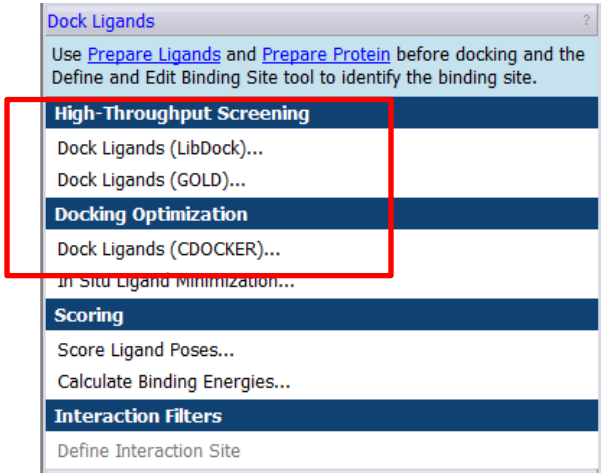


Docking

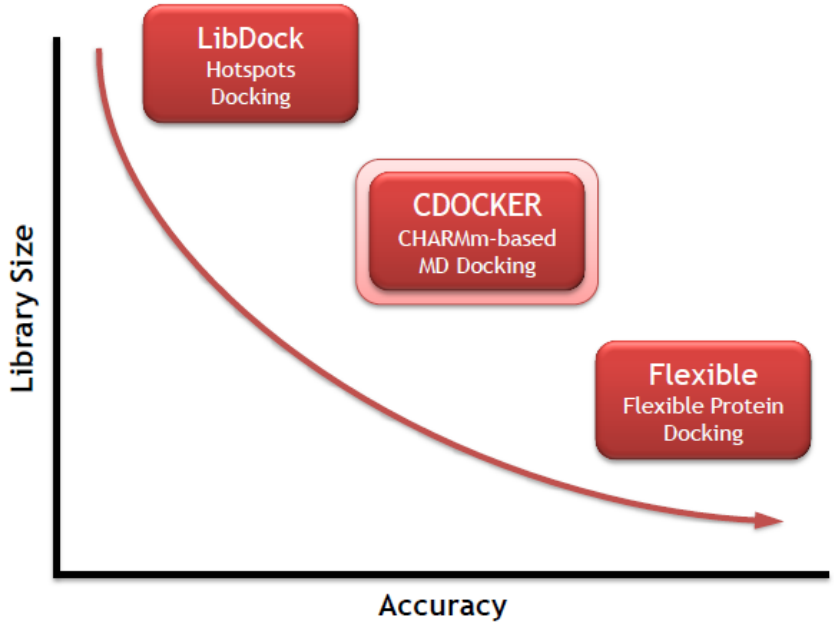
- Generate conformations
- Find a pose
- Docking tools in DS
  - Dock Ligands (CDOCKER)
  - Dock Ligands (LibDock)
  - Dock Ligands (GOLD, need extra license)
  - Dock Ligands (LigandFit)
  - Pharmacophore Docking
  - Flexible Docking

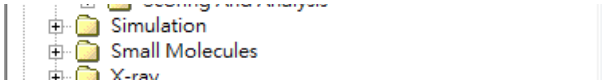
Docking from Tool Panel

- Docking tools in DS
  - Dock Ligands (CDOCKER)
  - Dock Ligands (LibDock)
  - Dock Ligands (GOLD)



- Docking tools in DS (Protocol)





Comparative performance of LibDock and CDOCKER on AstexDiverse dataset

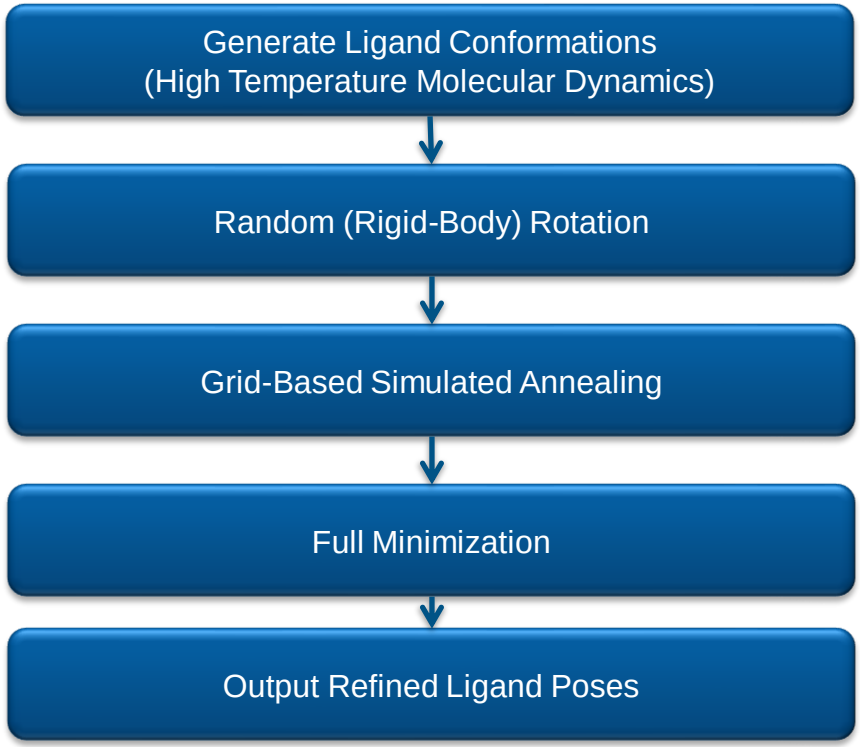
Method	% Docked Accurately (Best)	% Docked Accurately(Top)	RMSD Average (Best)	RMSD Average (Top)	Tim(min)
LibDock	91	50	1.2	3.8	0.5
CDOCER	94	79	0.8	1.5	5.0

LibDock is optimized for speed:
get accurate docked poses in seconds

CDOCKER is optimized for accuracy:
get significant improvement in rank-ordering of correct pose and RMSD to X-ray structure

1. Hartshorn, et al. J. Med. Chem., 50 (4), 726 -741 (2007)
3 GHz

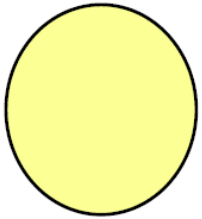
CDOCKER



- CDOCK is a grid-based molecules docking method.
- Ligand conformations are obtained by Molecular Dynamic (MD) methods

Wu G, Robertson DH, Brooks CL III, Vieth M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER - A CHARMm-based MD docking algorithm. *J. Comp. Chem.* **2003**, 13, 1549

CDOCKER : Ligand Fitting

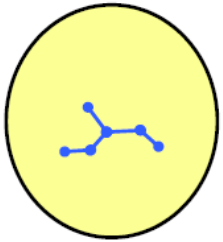


Site sphere

Ligand

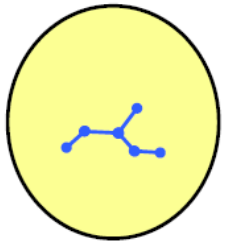
The conformation generated by high temperature MD

Sphere center



Ligand center

Simulated Annealing



Optimized Ligand

Small Molecule Docking using CDOCKER

- 41 protein-ligand complexes from the PDB
  - Structurally diverse set of ligands
  - All-atom representation used in published work[a]
- Grid-based approach used in Discovery Studio
  - Faster method
  - No significant compromise on accu

Success rates and CPU times for each algorithm

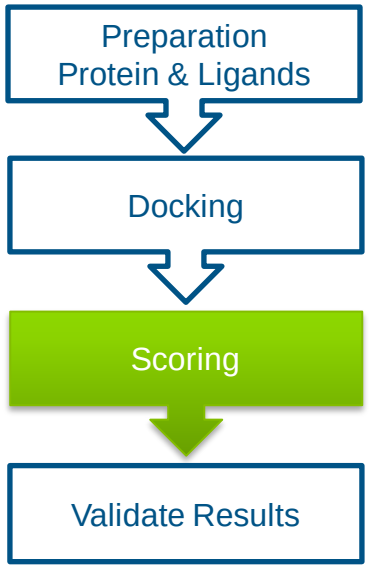
Docking Algorithm	success rate	av/median CPU times
GOLD[b]	46.3	824/708[c]
Dock[b]	51.2	200/114[c]
FlexX[b]	53.7	65/35[c]
CDOCKER (No Grid)[b]	82.9	2012/1740[c]
Discovery Studio CDOCKER (Grid)	78.1	630/580[d]

Single best docking run for each algorithm
(Success is defined as RMSD < 2Å from X-ray structure)

# of rot. bonds	total # of ligands	# of ligands docked correctly				
		Dock[b]	GOLD[b]	FlexX[b]	CDOCKER (No Grid)[b]	Discovery Studio CDOCKER (Grid)
< 8	20	18	14	16	19	18
≥ 8	21	3	4	6	15	14

a. Erickson et al. J Med Chem (2004) 47:45-55

Docking Workflow



Literature Scoring Functions

- Also known as Empirical/Knowledge-based scoring functions
- based on counting the number of various types of interactions between the two binding partners
- based on statistical observations of intermolecular close contacts in large 3D databases
- Scoring functions
  - LigScore 1 & 2
  - Piecewise Linear Potential (PLP) 1 & 2
  - Potential of Mean Force (PMF) & PMF04
  - Jain
  - Ludi 1, 2 , & 3

Energy-Based Functions

- Binding energy calculation method based on the following formula

$$E_{binding} = E_{complex} - (E_{ligand} + E_{receptor})$$

Scoring

- Scoring of docked poses is still a major challenge
- Aim of scoring:
  - Identification of the correct binding pose by lowest energy value
  - Ranking of protein-ligand complexes according to their binding affinities
- No single scoring function can correctly rank every protein-ligand complex
  - Relative contribution of different protein-ligand interactions may vary between structural families
- Use consensus scoring
  - Combination of several scoring functions

Types of Scoring Functions

- Empirical scoring functions
  - Derived from training sets of protein-ligand complexes with determined affinity data
- Force field-derived functions
  - Handle the ligand binding prediction with the use of potential energies (non-bonded interaction terms)
  - Could include solvation and entropy contributions
- Knowledge-based functions
  - Based on atom pair potentials derived from structural databases
  - Forces and potentials are collected from known protein-ligand complexes to get a score for their binding affinities

Consensus Scoring

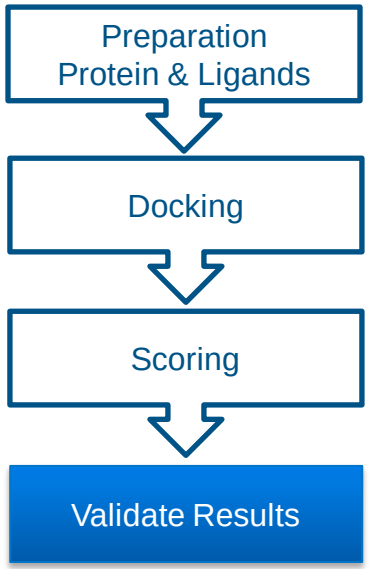
- Combination of several scoring functions
- The common top rankers get a higher consensus rank than single outliers
- False positives can be detected easier than one singular scoring function
- Advisable to use a well-suited scoring function as consensus score always presents an average value

Consensus Scoring

- 40% Example

Models	Score1	Score2	Score3	Score4	Consensus
mol1	12 | 0	55 | 0	43 | 0	241 | 0	0
mol2	22 | 0	46 | 0	113 | 0	**283** | 1	1
mol3	**112** | 1	**92** | 1	**221** | 1	**299** | 1	4
mol4	78 | 0	**82** | 1	182 | 0	251 | 0	1
mol5	**98** | 1	77 | 0	**193** | 1	263 | 0	2

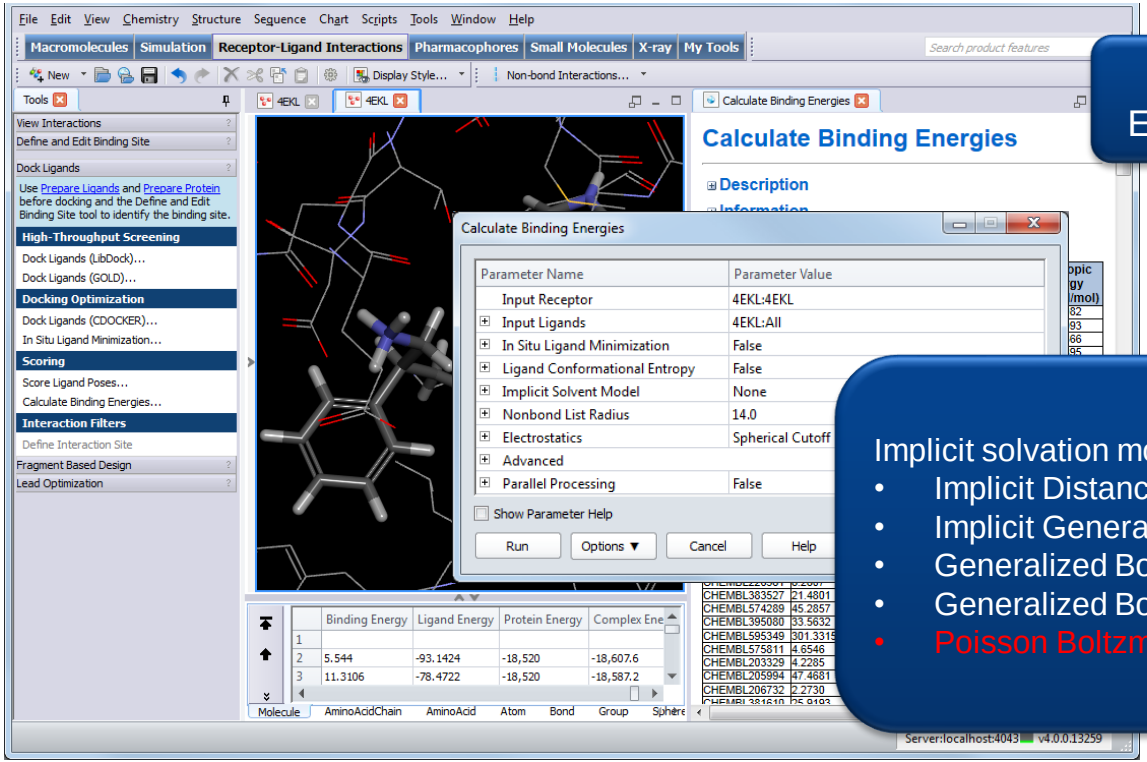
Docking Workflow



Search and filter for the hits

- Compare with experimental results
  - Key residues
  - RMSD with the reference structure
  - Salt bridge
- Calculate binding energy
- View the non-bond interactions between ligand and receptor
- View the non-bond interactions between ligand and receptor (2D diagram)

Binding Energies



Binding energy

$E_{binding} = E_{complex} - (E_{ligand} + E_{receptor})$

Implicit solvation model can be used:

- Implicit Distance-Dependent Dielectrics
- Implicit Generalized Born
- Generalized Born with Molecular Volume(GBMV)
- Generalized Born with a Simple Switching(GBSW)
- Poisson Boltzmann with non-polar Surface Area (PBSA)

Analysis - Analyze Ligand Poses

	Favorable Total	Unfavorable Total	Charge Total	Halogen Total	Hydrophobic Total	HydrogenBond Total	Other Total	Favorable A:LEU156	Favorable A:GLY157
1	19	0	4	0	9	5	1	0	0
2	17	0	2	0	3	10	2	0	0
3	16	0	0	0	10	5	1	0	0
4	15								
5	20								
6	14								
7	18								
8	12								
9	13								

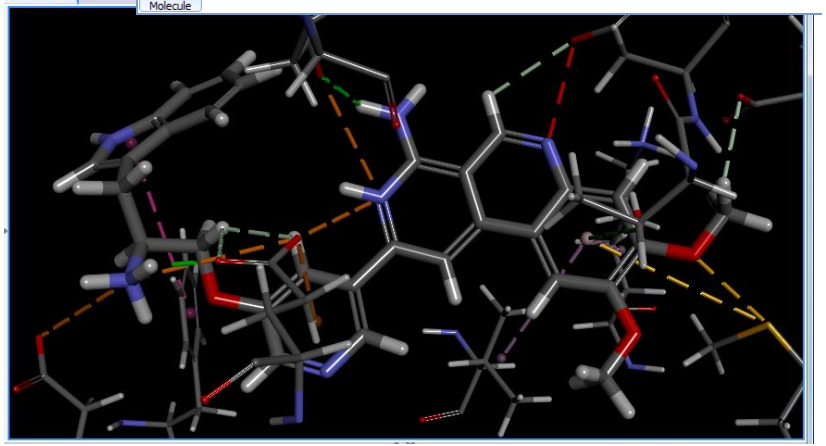
Molecule

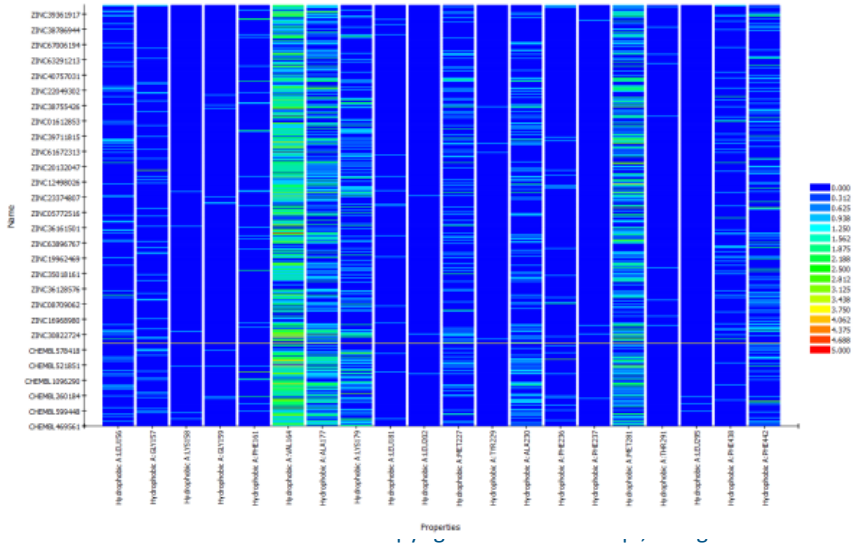
otor-Ligand Interaction

4EKL

	Hydrophobic A:PHE438	Hydrophobic A:PHE442	HydrogenBond A:LEU156	HydrogenBond A:GLY157	HydrogenBond A:LYS158	HydrogenBond A:GLY159	HydrogenBond
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0
4	0	0	0	0	1	0	0
5	0	0	0	0	2	0	0
6	0	1	0	0			
7	0	1	0	0			
8	0	0	0	1			
9	0	0	1	0			

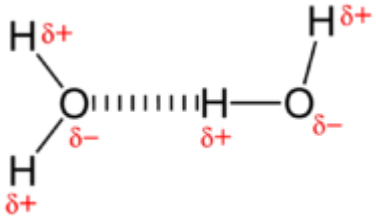
Molecule

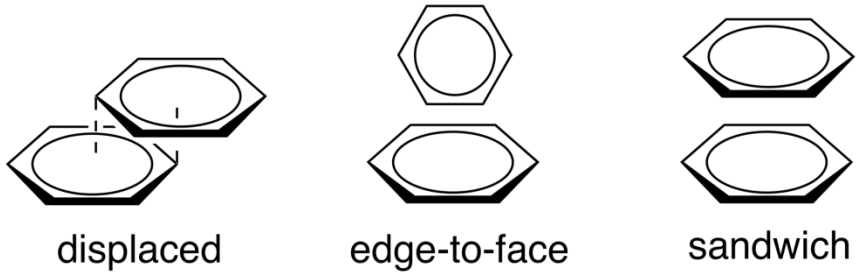




Non-covalent (Non-bond) interactions

- Hydrogen bond interactions
- Electrostatic interactions
- π-effects
- Van der waals forces
- Hydrophobic effects





bonds strenghts

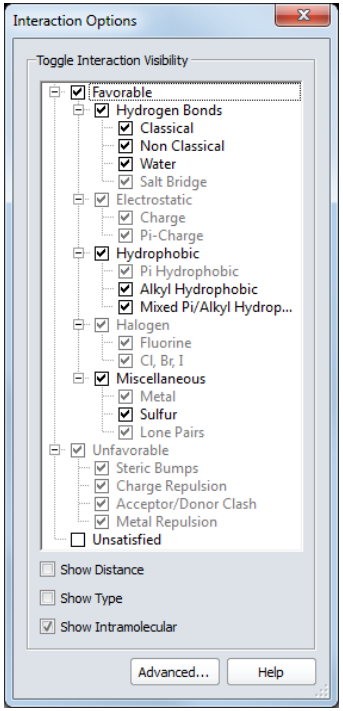
Covalent bond

Type	Energy
C-O bond	81 kcal/mol
C-C bond	86 kcal/mol
C-H bond	103 kcal/mol
C=C bond	143 kcal/mol
C=O bond	165 kcal/mol

Non-Covalent bond

Type	Energy
Hydrophobic	<10 kcal/mol
Hydrogen bond	2-30 kcal/mol
Electrostatic	1-20 kcal/mol
π-π aromatic stacking	0-10 kcal/mol
Van der Waals	0.1-1 kcal/mol

Analysis - Non-Bond Interactions



Favorable (*See below*)

Unfavorable
- Steric Bumps
- Charge Repulsion
- Acceptor-Acceptor clashes
- Donor-Donor clashes

Unsatisfied
- Hydrogen bond donor
- Hydrogen bond acceptor
- Charged atoms

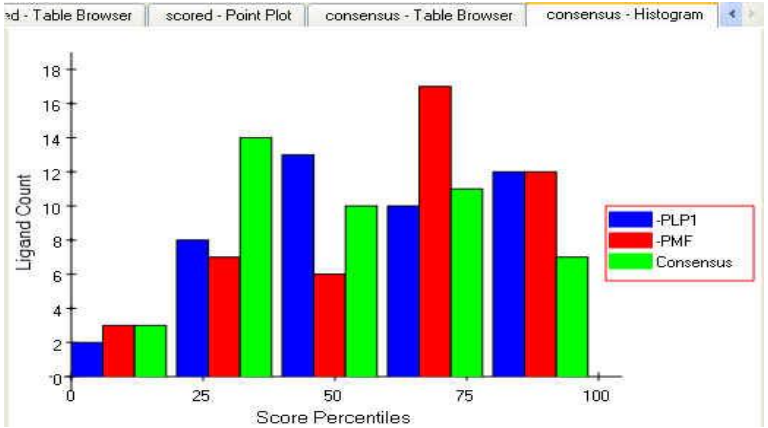
- Charge
  - Attractive Charges
  - Salt Bridge
  - Pi-Cation
  - Pi-Anion
- Halogen
  - Halogen (Fluorine)
  - Halogen (Cl, Br, I)

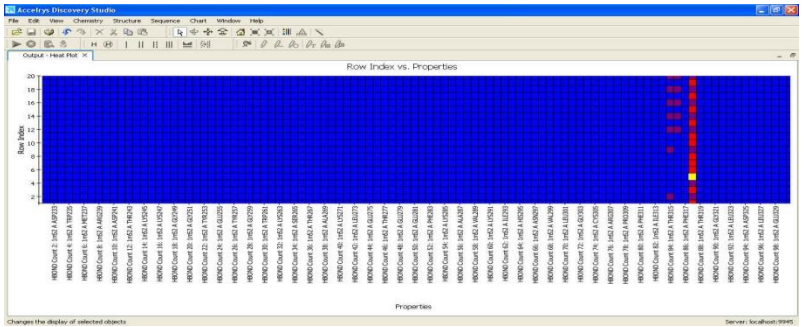
- Hydrophobic
  - Pi-Pi Stacked
  - Pi-Pi T-Shaped
  - Amide-Pi Stacked
  - Alkyl
  - Pi-Sigma
  - Pi-Alkyl
- Other
  - Metal-Acceptor
  - Pi-Sulfur
  - Sulfur-X
  - Pi-Lone Pair

- Hydrogen Bond
  - Conventional Hydrogen Bond
  - Carbon Hydrogen Bond
  - Pi Donor Hydrogen Bond
  - Water Mediated Hydrogen Bond
  - Water Hydrogen Bond
  - Salt Bridge

Pose Analysis

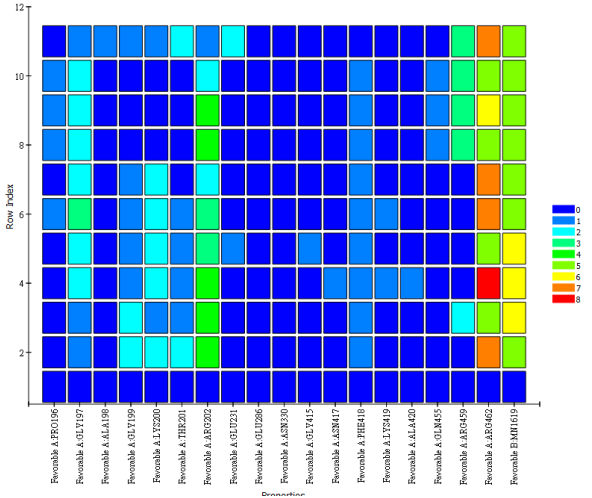
- Plot Charts of Scoring
  - Simple Line Plot
  - Histogram
  - Hit Rate Plot
- Run Analyze Ligand Poses protocol
  - Hbonds
  - Contacts
  - Heat Maps
- Optional energy minimization of each pose
  - Ligand Minimization Protocol
- Fits saved to a molecule table
  - Can be exported to an SD file



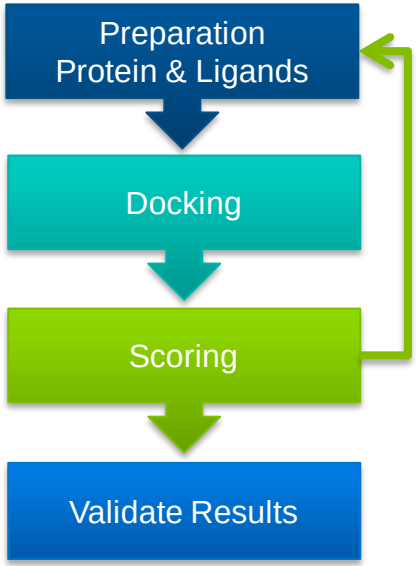


Demo

- Analyze ANP poses by Analyze Ligand Pose protocol
- Draw the heat map by Chart tools
- Find the poses which interact with
  - K200, T201, R202 (motif I, also called P-loop)
  - D285,E286 (Mn binding motif II)
  - Q455, R459, and R462 (motif VI)
- Find the best pose



Docking Workflow



Repeat Cycle = Virtual High Throughput Screening (vHTS)

- In order to maximize the rate of success, the evaluation phase in each steps is very important.
- Requirements
  - Known active compounds and decoy
  - Bound molecules
  - Optimize the parameters of docking technique
  - Prioritized the molecules in library before proceeding the docking and scoring steps

SBD: Fragment-Based Design Methods

Probing the Binding Site

Adding Fragments

Modifying Scaffolds

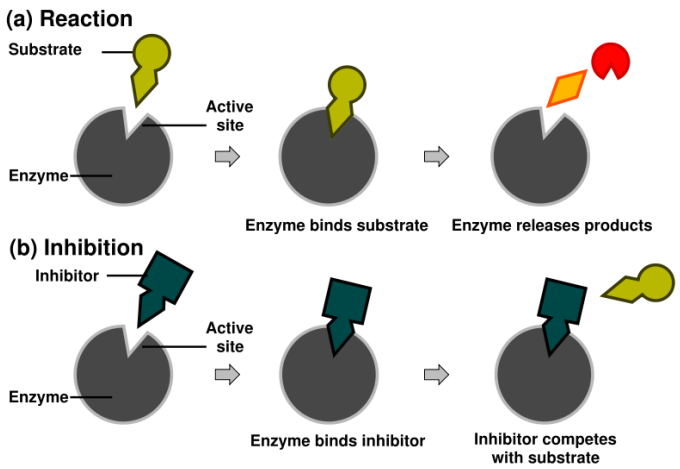
Alternative fragment-based methods available
- Fragment Based Pharmacophores

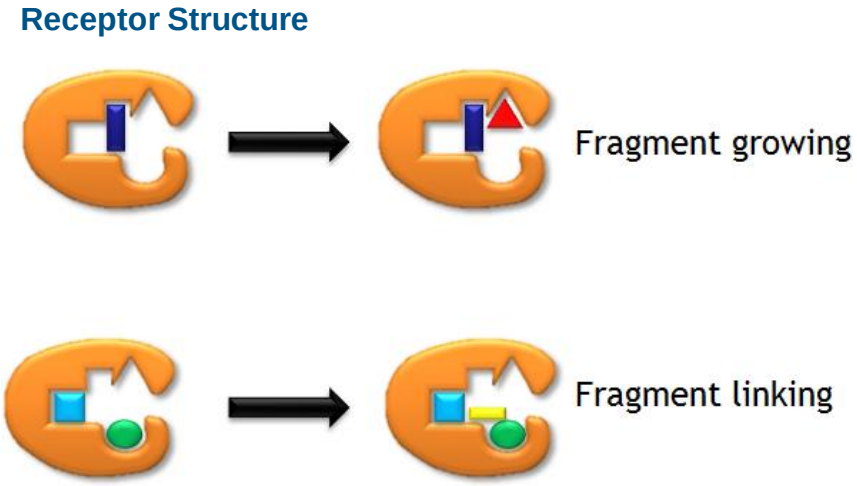
Adding Fragments

SBD: Fragment-Based Design Methods

- GROW
  - Reaction-based in situ ligand enumeration
  - E.g., Amide synthesis, Esterification, Hiyama, Kuyama, Negishi, Stille, Suzuki, Williamson Ether
  - Pre-filtered sets of reagents selected from ACD
- REPLACE
  - Fragment based in situ isostere replacement
  - E.g., scaffold-hopping, R-group replacement
  - Pre-filtered set of 1.5M fragments generated from SCD

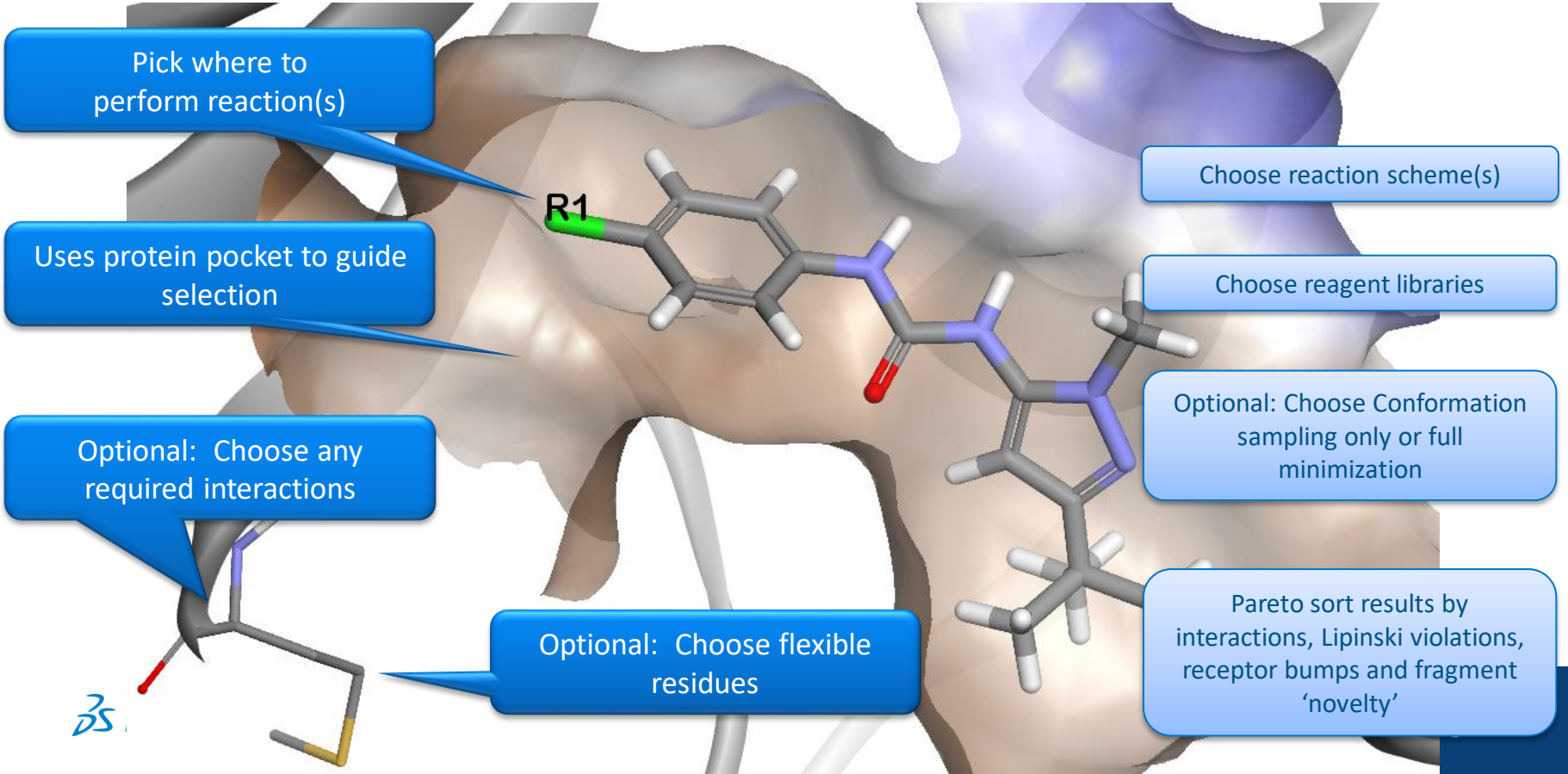
Fragment-Based Drug Design





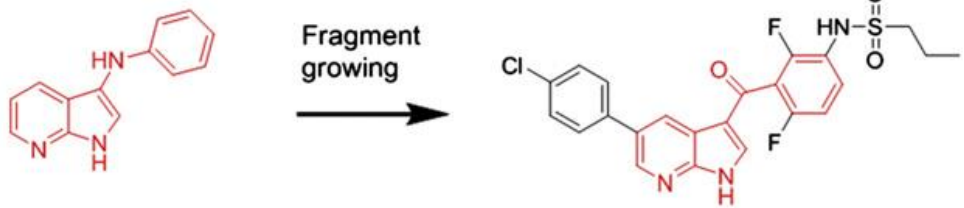
Generate a novel compound from existed scaffold

GROW: Reaction-based *in-situ* Ligand Optimization



Fragment-based design of the BRAF inhibitor vemurafenib.

First fragment-based drug (Zelboraf) approved in 2011!



4

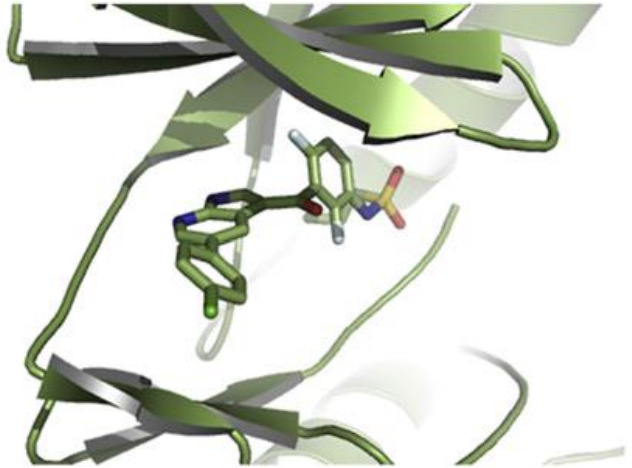
Unselective, weakly potent fragment hit (IC_{50} > 100μM)

PIM-1 IC_{50} ~ 100μM

5 (Vemurafenib)

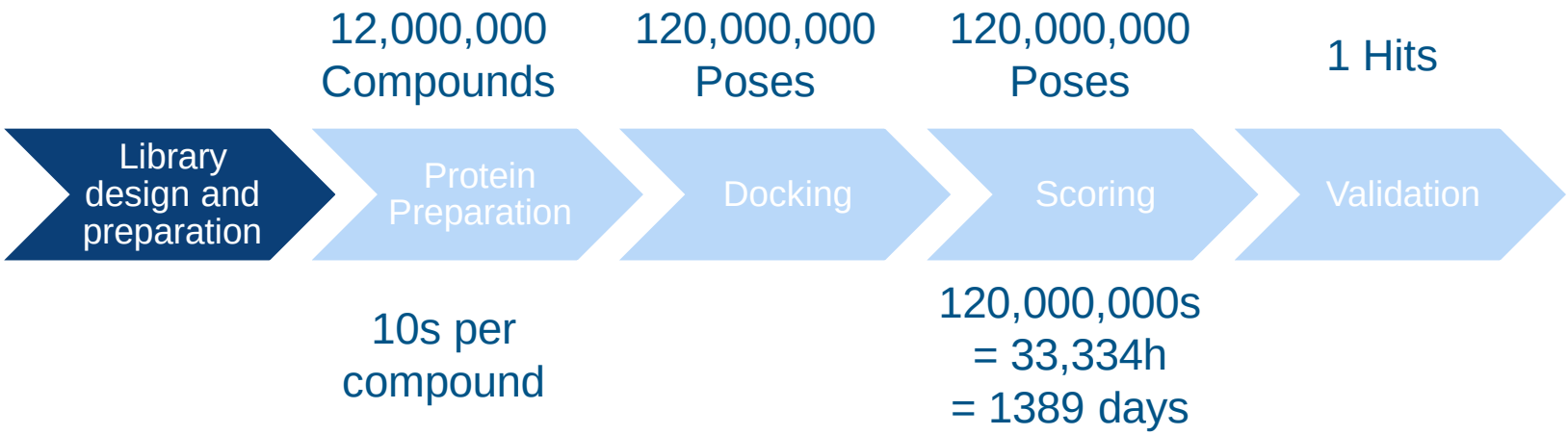
BRAF (V600E) IC_{50} = 31nM
High degree of selectivity against other kinases

PIM-1 IC_{50} > 100μM

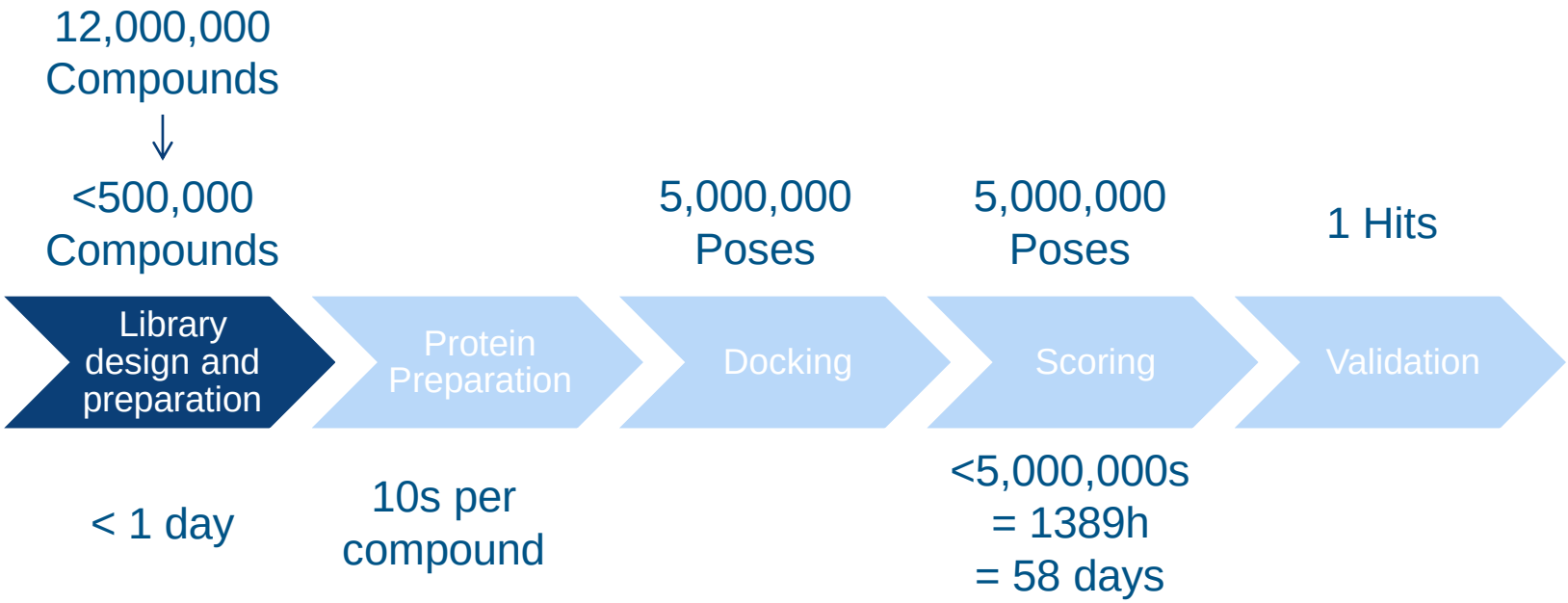


Swen Hoelder, Paul A. Clarke, Paul Workman, 2012

Workflow of Virtual High Throughput Screening



Workflow of Virtual High Throughput Screening



Compound databases

- BIOVIA database
  - **A**vailable **C**hemicals **D**irectory – 12,138,856 unique compounds
  - **S**creening **C**ompounds **D**irectory – 10,852,222 unique compounds
  - MDDR – 239,064 registered structures with bioactivity data
  - Toxicity – 172,542 registered structures
  - **C**omprehensive **M**edicinal **C**hemistry – 9603 registered structures
  - Metabolite – 71,359 molecules within 119,425 reactions

ACD

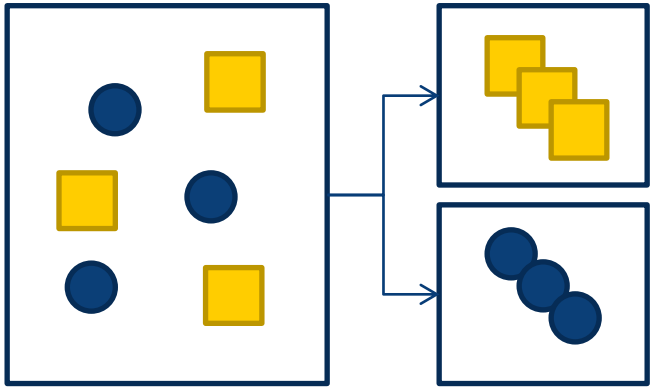
	Index	Name	Visible	Tagged	Visibility Locked	MDLNUMBER	CAS_TEXT	supplier.coden	CLOGP
1	1	Molecule	☑ Yes	☐ No	☐ No	MFCD00000001	645-96-5	ABCR ACC	2.29
2	2	Molecule	☐ No	☐ No	☐ No	MFCD00000002	7188-38-7	ABCR	
3	3	Molecule	☐ No	☐ No	☐ No	MFCD00000003	14542-93-9	ABCR ACC	
4	4	Molecule	☐ No	☐ No	☐ No	MFCD00000004	10540-91-7	ABCR	

ABCR
ACC

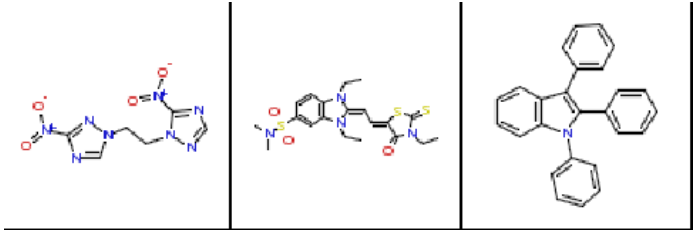
Total Products	Unique Compounds (3D Models†)	New Chemicals	Removed Chemicals*	Storage Requirements (ACD)			
				RCG Format		Direct Format **Top: Direct 8.0, Bottom: Direct 9.x**	
				Oracle Tablespace	Unzipped Dump Files	Oracle Tablespace	Unzipped Dump Files
28,539,359	**12,030,775 (11,988,413)**	**664,325**	**707**	**58.8 GB**	**38.0 GB**	**84.8 GB** **80.2 GB**	**51.5 GB** **48.7 GB**

Cluster subsets

- Number of molecule cluster
- Diversity of molecule cluster
- Countable property of molecule cluster
- ….

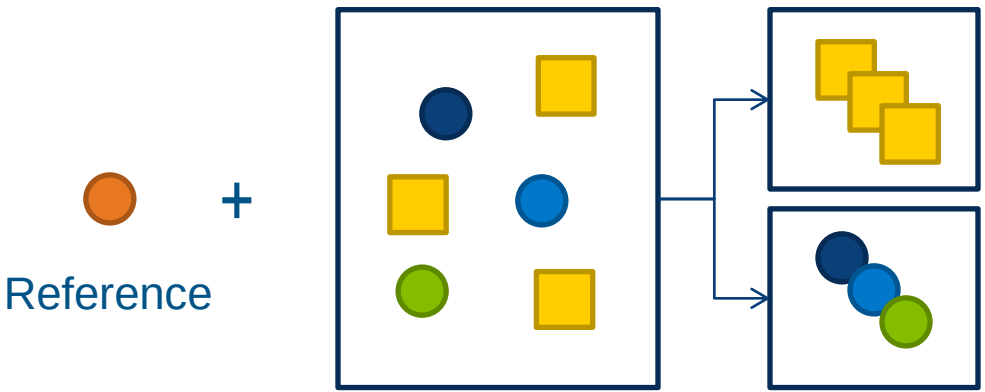


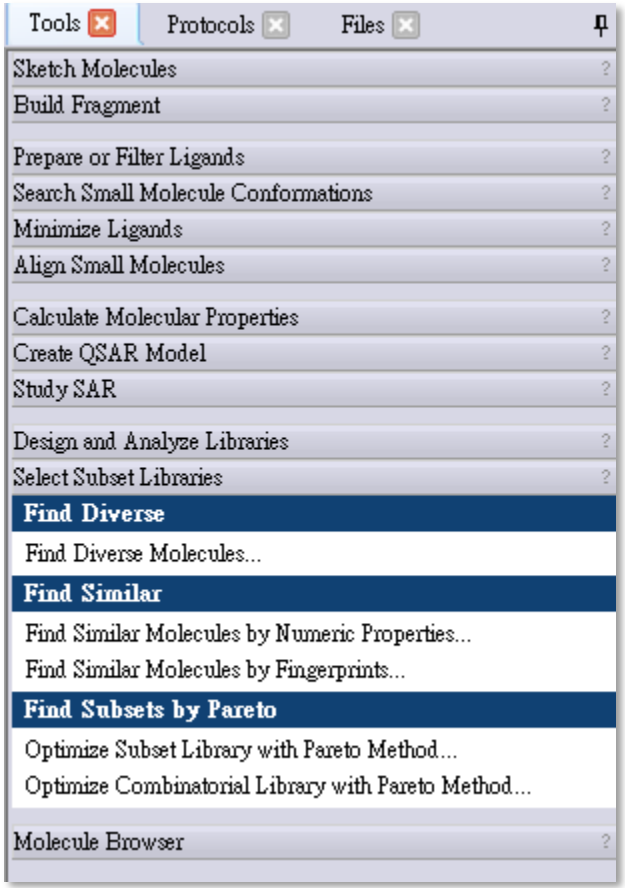
Save time for the sequential analysis steps



Cluster subsets cont.

- Diversity of molecule cluster
- Countable property of molecule cluster





Flexible Docking

Generate Receptor Side Chain Conformations (ChiFlex algorithm)

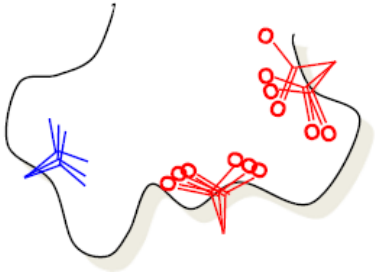
Generate Receptor Hotspots for Receptor Conformations

Generate Ligand Conformations (On-the-fly methods, or pre-generated)

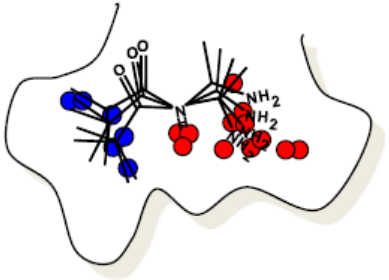
Match Conformations to Hotspots

Refine Protein Side Chain Conformations (ChiRotor algorithm)

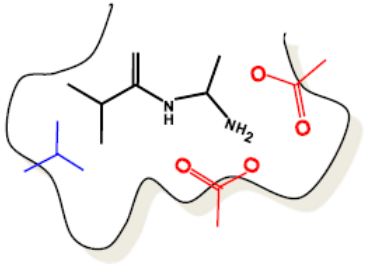
Simulated Annealing and Energy Minimisation of Poses



LibDock



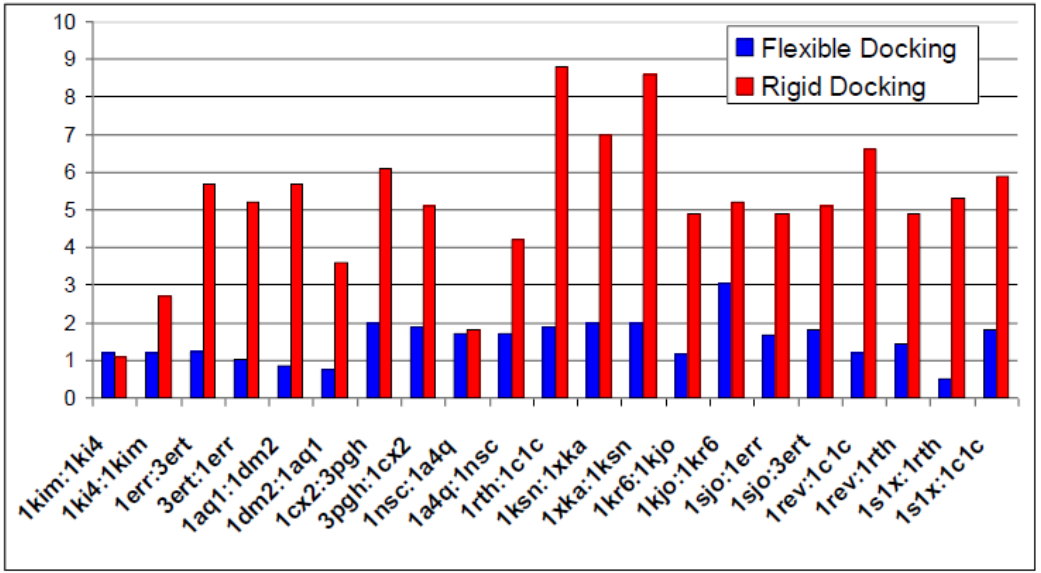
CDOCKER



oska J, Spassov VZ, Maynard AJ, Yan L, Austin N, look PK, Venkatachalam CM. Fully automated olecular mechanics based induced fit protein-ligand ocking method, *J. Chem. Inf. Model.* **2008**, 48, 1965-973.

Validation of Rational Flexible Docking

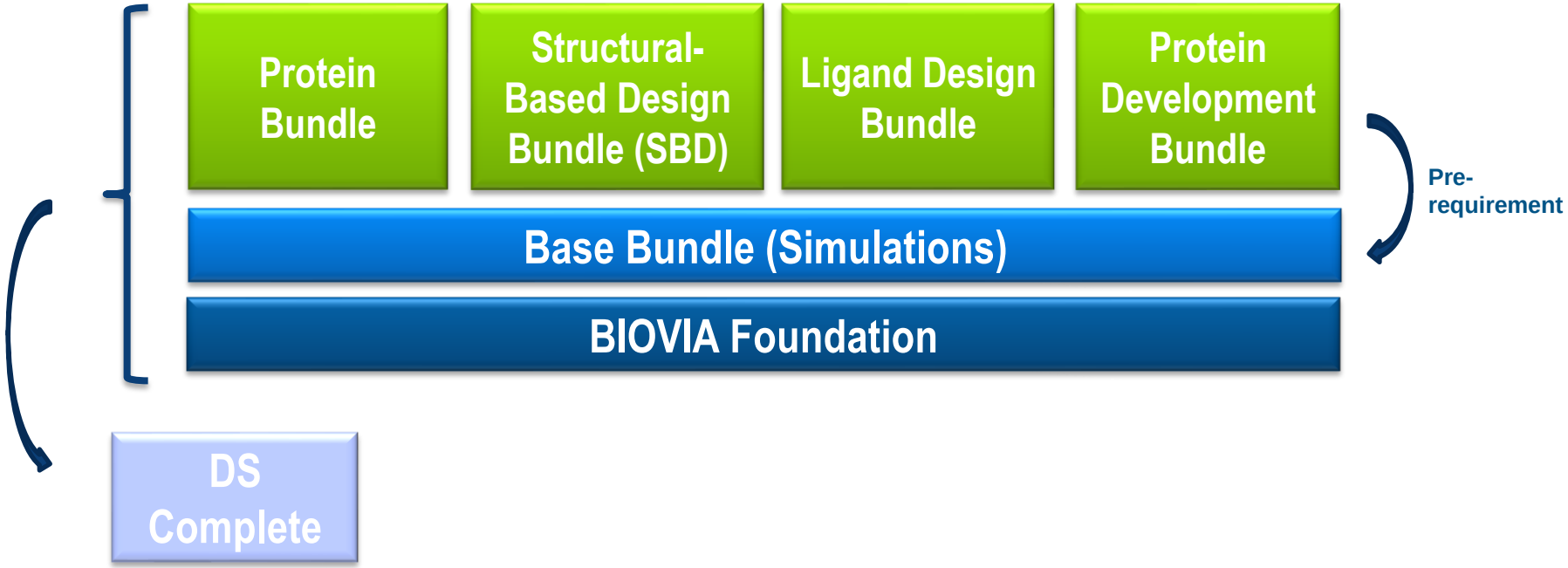
Cross-docking: Dock ligands into an alternate conformation of the same receptor



RMSD Values compared to X-ray conformation for cross-docking experiments (1kim:1ki4 denotes 1kim ligand docked into 1ki4 receptor)

CHARMm-based sampling successfully captures receptor movements induced by a non-native ligand

Bundle solution



Critical Development Issues

Too many projects to support

- Lots of projects to support
- Multiple applications needed to complete each study
  - Time consuming and not automated
- Whole process manually repeated, as new results released

Too many different applications

- Research requires multiple scientific applications
  - Each with own data formats and procedures
- Each study is time-consuming and can't be automated

Diverse, complex data to work with

- Variety of disparate complex scientific data needed for each study
- Need to analyse data together
  - Interactive, 3D, charts, tables, etc.

Topic and requirement relations

Topic	Requirement	Solution
Protein Engineering	• 蛋白質功能研究 • 蛋白質純化 • 基因工程預測工具	• Base Bundle • Protein Bundle
New Drug Discovery (Small molecule)	• 合成化學預測工具 • 新藥篩選 & 老藥新用 • 小分子化合物作用機制研究(天然物、複方作用機制)	• Base Bundle • SBD Bundle • Ligand Design Bundle
New Drug Discovery (Protein)	• DNA、RNA、Peptide、Antibody、Protein藥物作用機制研究 • 巨分子藥物設計與篩選	• Base Bundle • Protein Bundle • Protein Development Bundle

Q&A

訊聯基因數位

For more information please contact…

台北市114內湖區新湖一路36巷28號

中華民國 台灣

Ph: (02) 2795 1777 x 3666 Fax: (02) 2793 8009

msc-support@gga.asia

www.gga.asia