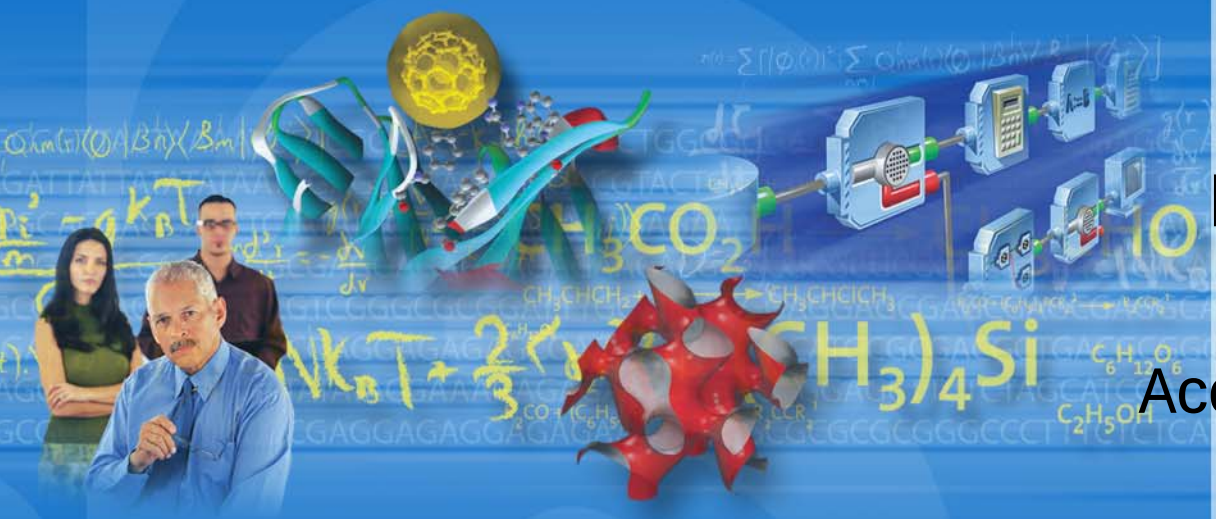


Discovery Studio 2.0

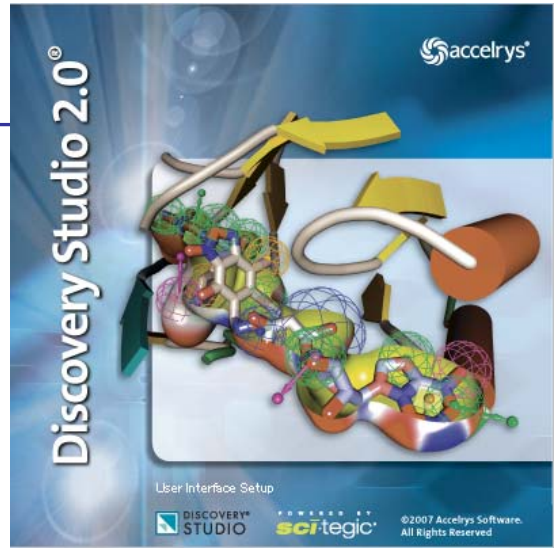
Accelrys Life Science Tool

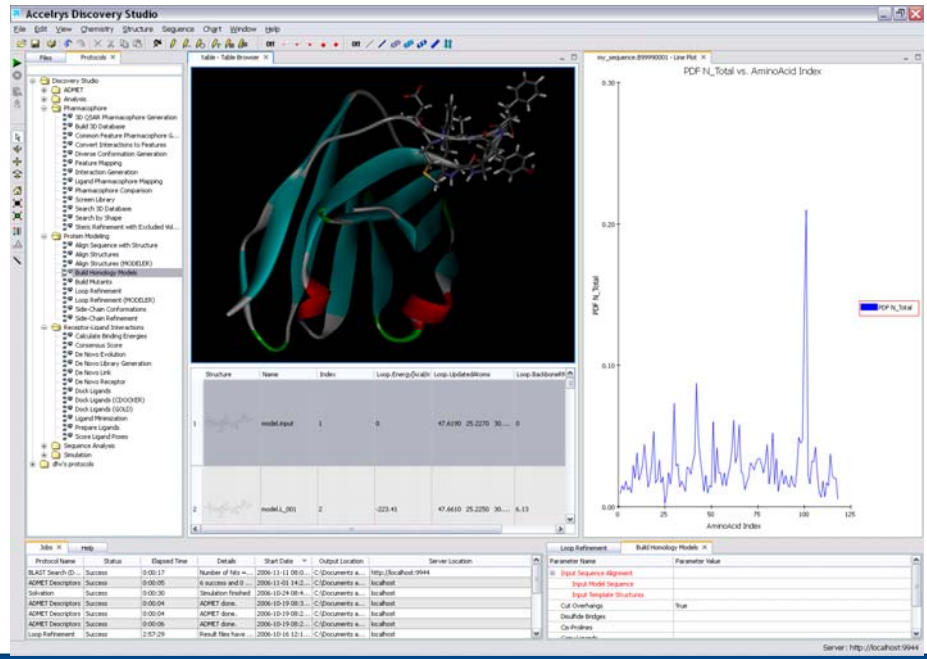
林進中

分子視算

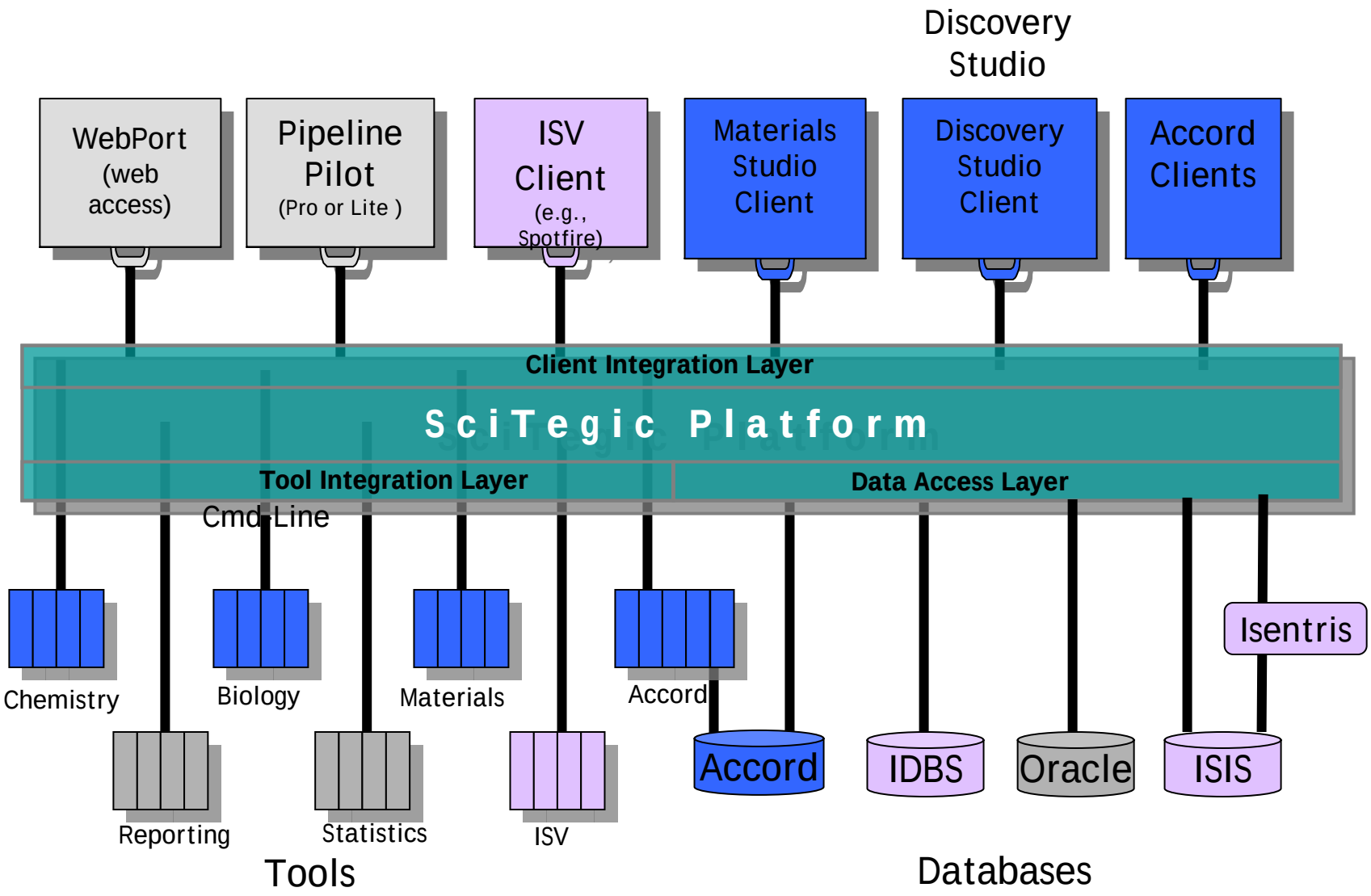
What is Discovery Studio?

- Discovery Studio is a complete modelling and simulations environment for Life Science researchers
  - Interactive, visual and integrated software
  - Consistent, contemporary user interface for added ease-of-use
  - Tools for visualisation, protein modeling, simulations, docking, pharmacophore analysis, QSAR and library design
  - Access computational servers and tools, share data, monitor jobs, and prepare and communicate their project progress
  - Windows and Linux clients and servers



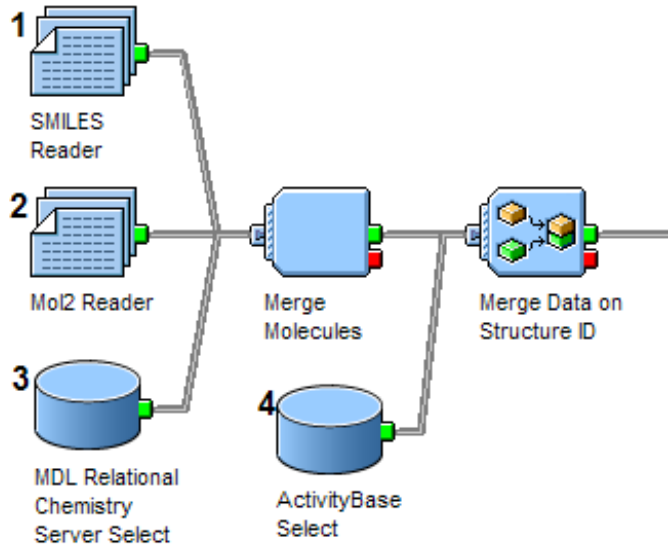


Accelrys Discovery Studio Application



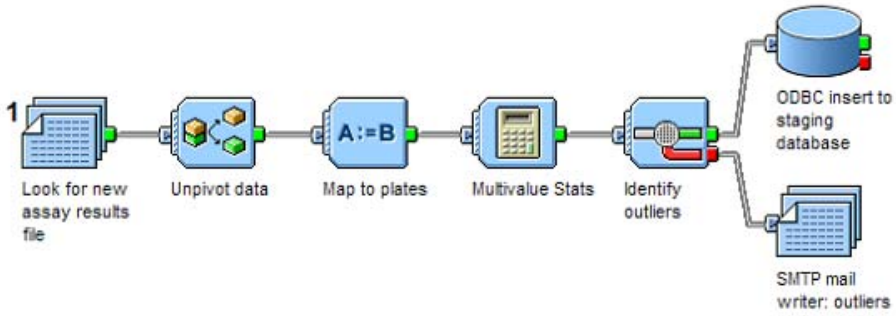
Pipeline Pilot - Data Processing and Integration

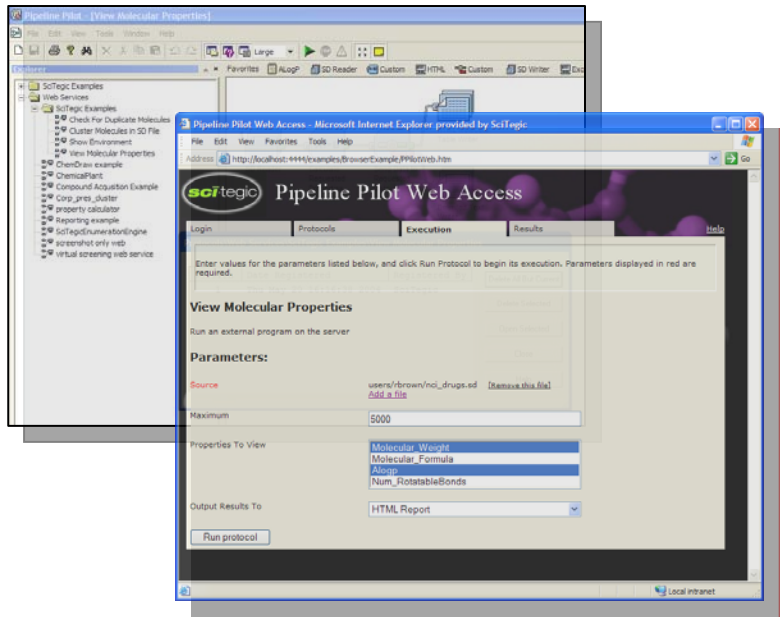
- Integration of data from multiple disparate data sources
- Integration of disparate applications
  - Third party vendors and in-house developed codes under the same environment



Pipeline Pilot - Data Processing and Integration

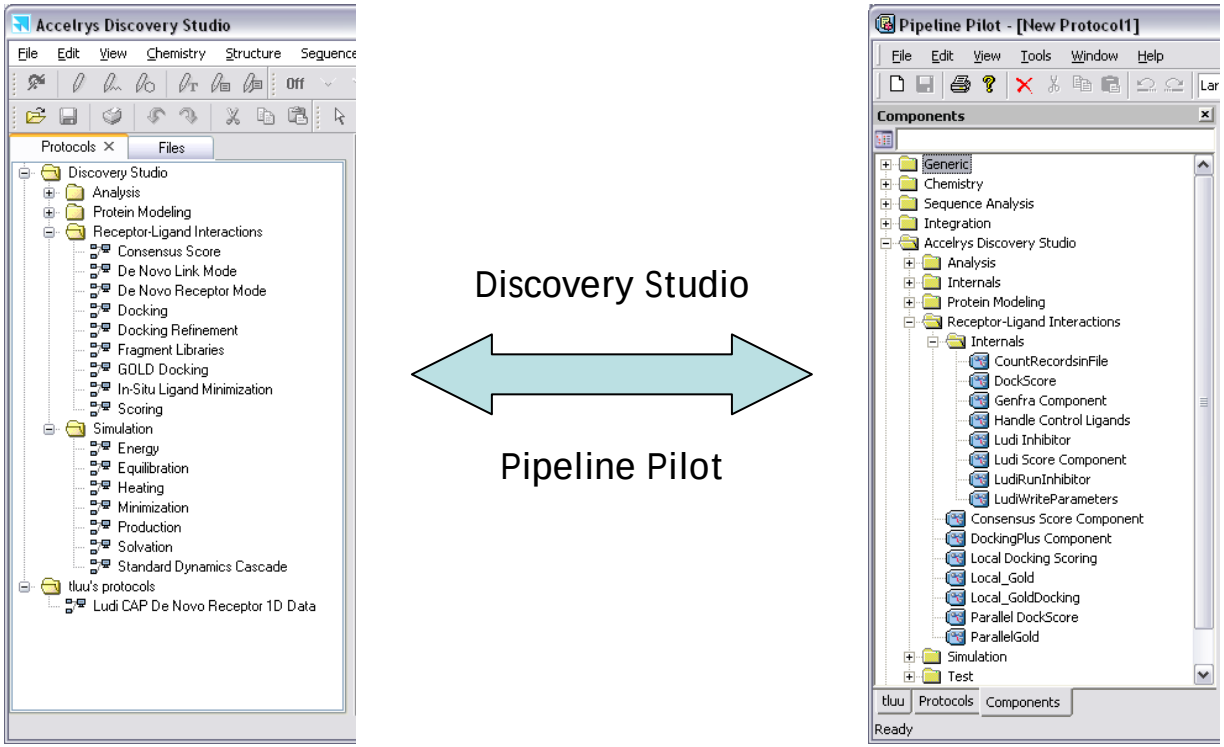
- Automated execution of routine processes
- Standardised data management
- Capture of workflows and deployment of best practice

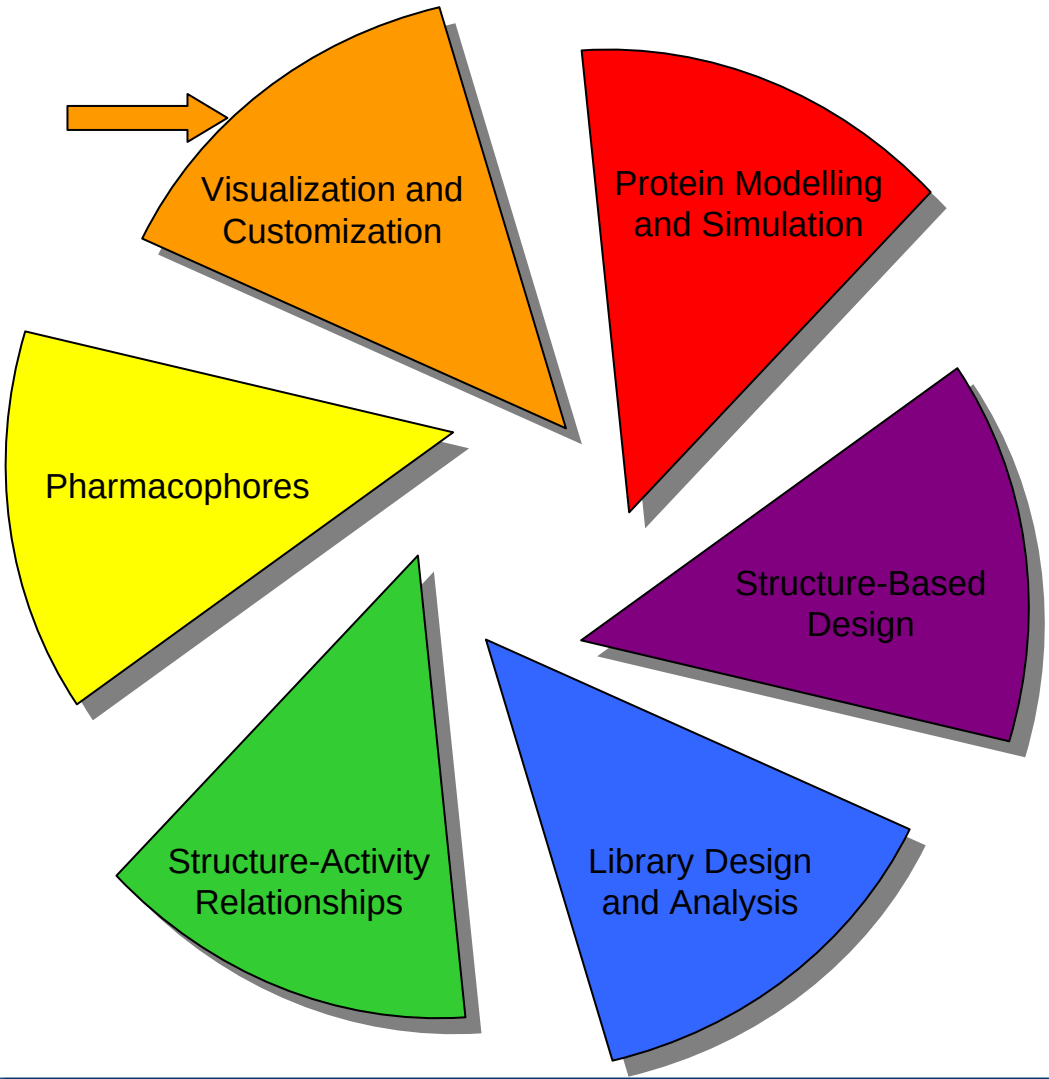




Interoperability

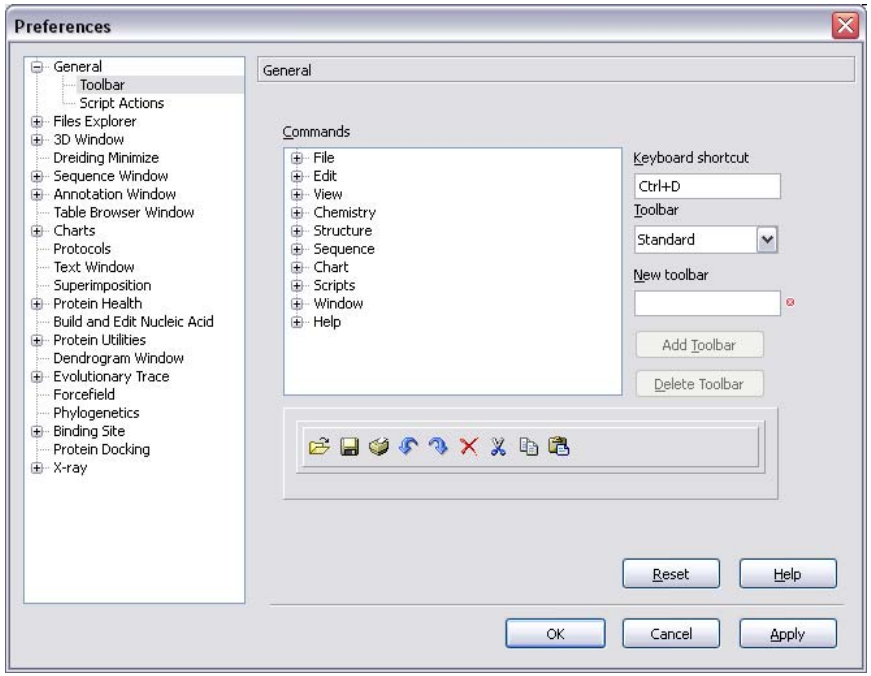
- Accessible thorough Discovery Studio or Pipeline Pilot
- Third party or user in-house codes can be made accessible through Pipeline Pilot Protocols in Discovery Studio (Example: Gold Docking)

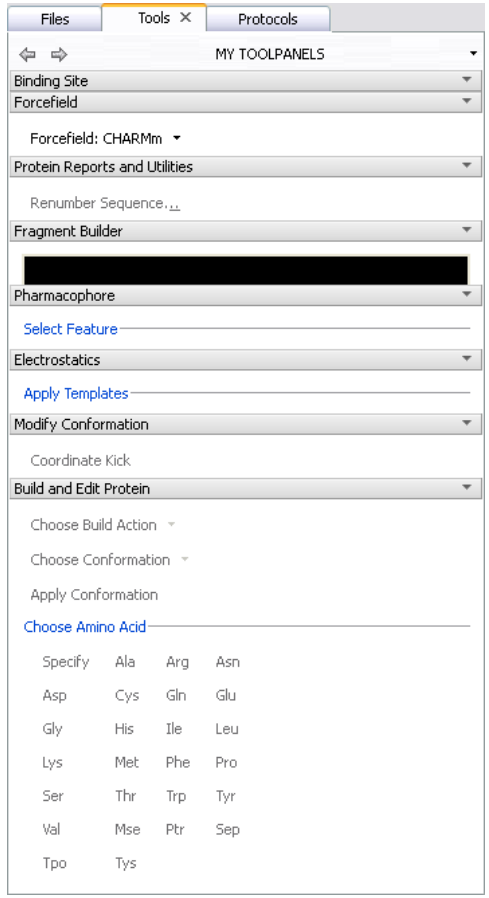



Visualization and Customization
Protein Modelling and Simulation
Pharmacophores
Structure-Based Design
Structure-Activity Relationships
Library Design and Analysis

Visualisation and Customisation

- Customise the interface by adding user defined toolbars, toolpanels* and shortcut keys

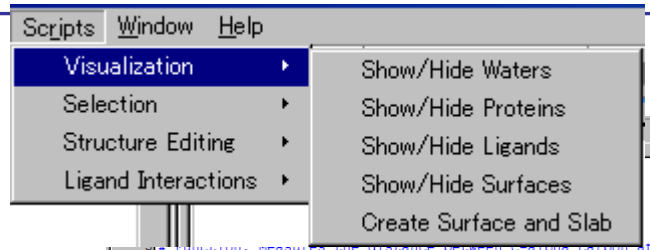


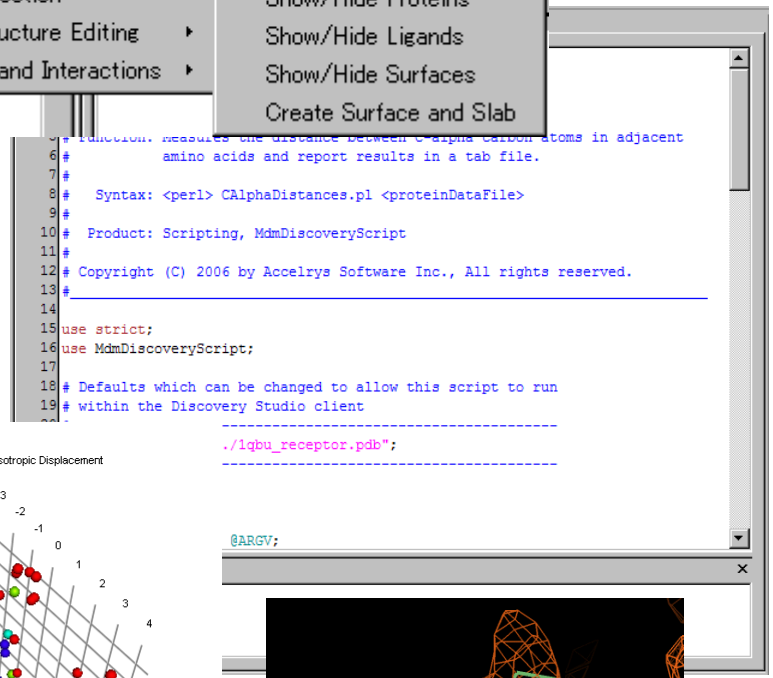


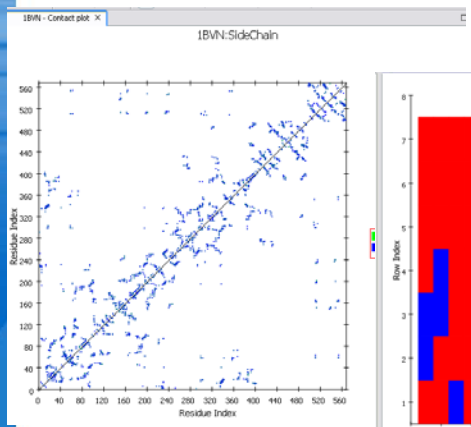
* DS Visualizer Pro only

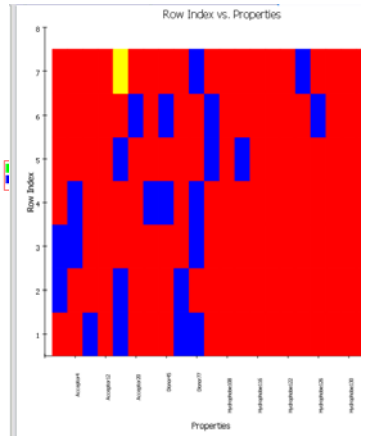
Visualisation in Discovery Studio

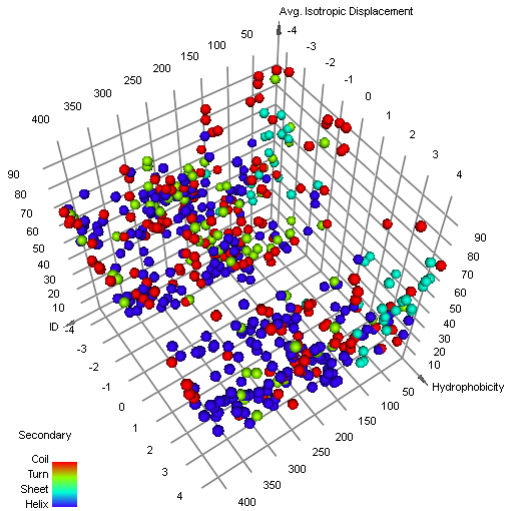
- Visualization
  - Visualize data (2D/3D Plot)
  - Script Interface
  - Protein, Molecular structure

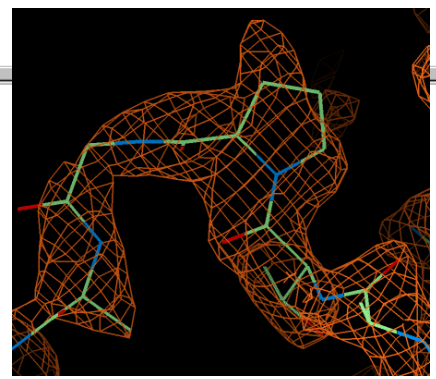












Visualisation and Scripting

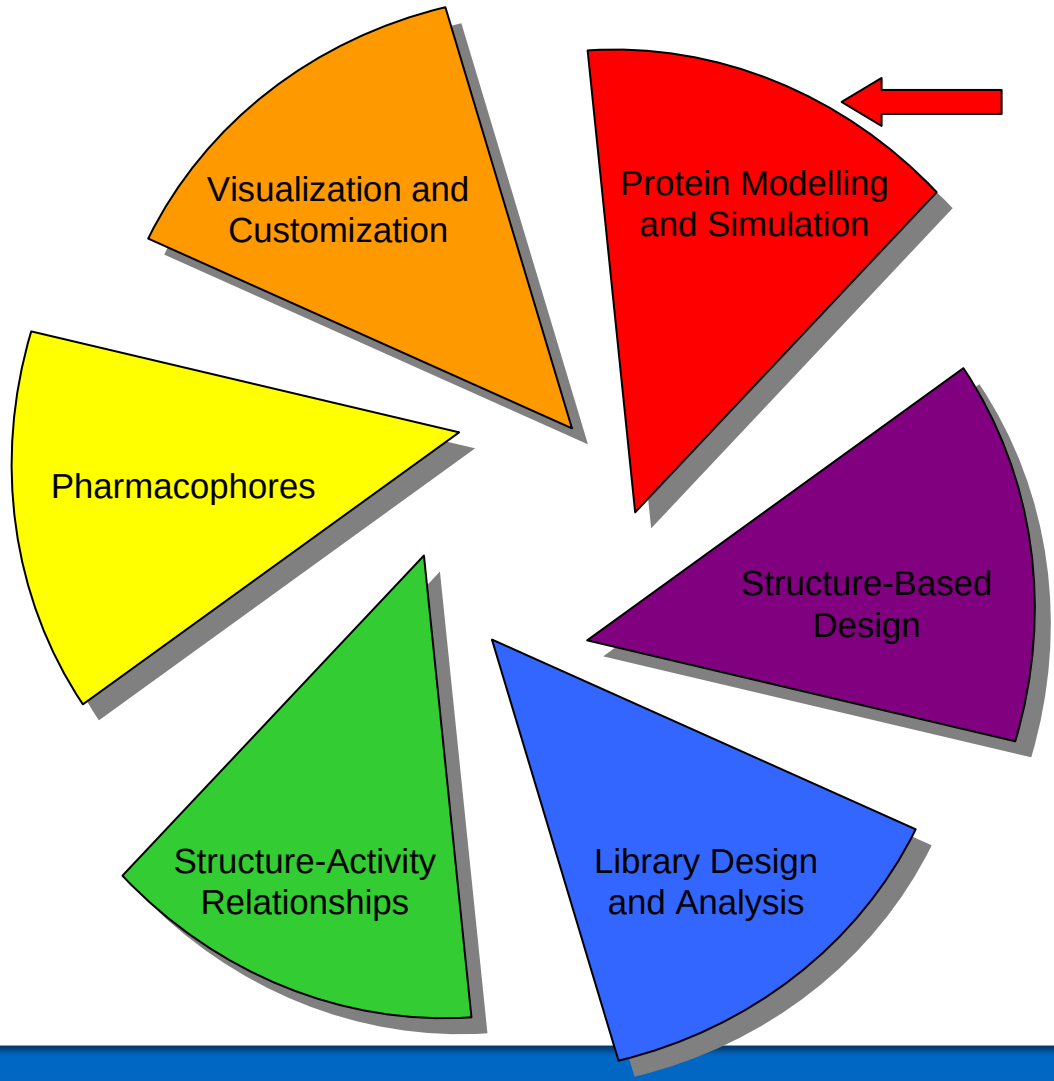
- Automate workflows through Perl scripting*
  - Accessible through command-line, DS Visualizer Pro client or Pipeline Pilot client
  - Requires DS Visualizer Pro
  - Access to rich Perl scripting language constructs
    - Objects, Modules, Sub-routines, Tests and branching, Loops & more
  - Loading, inserting and saving of files
  - Viewing of loaded files (when run inside of DS)
  - Scripts can be combined with other (both non-Accelrys as well as Accelrys) scripts

- Accelrys Community
  - http://accelrys.org/

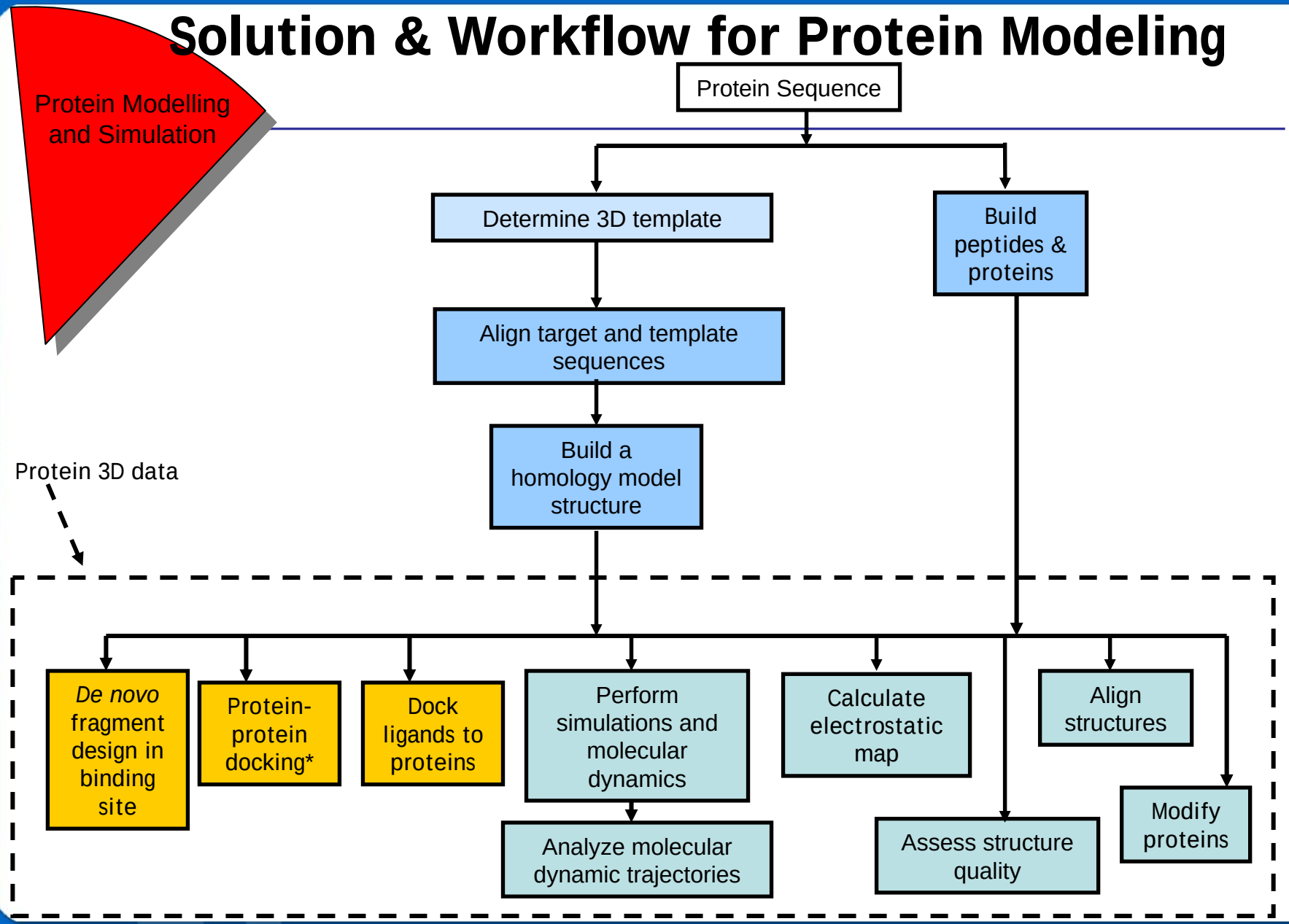
* DS Visualizer Pro only

Visualisation and Scripting

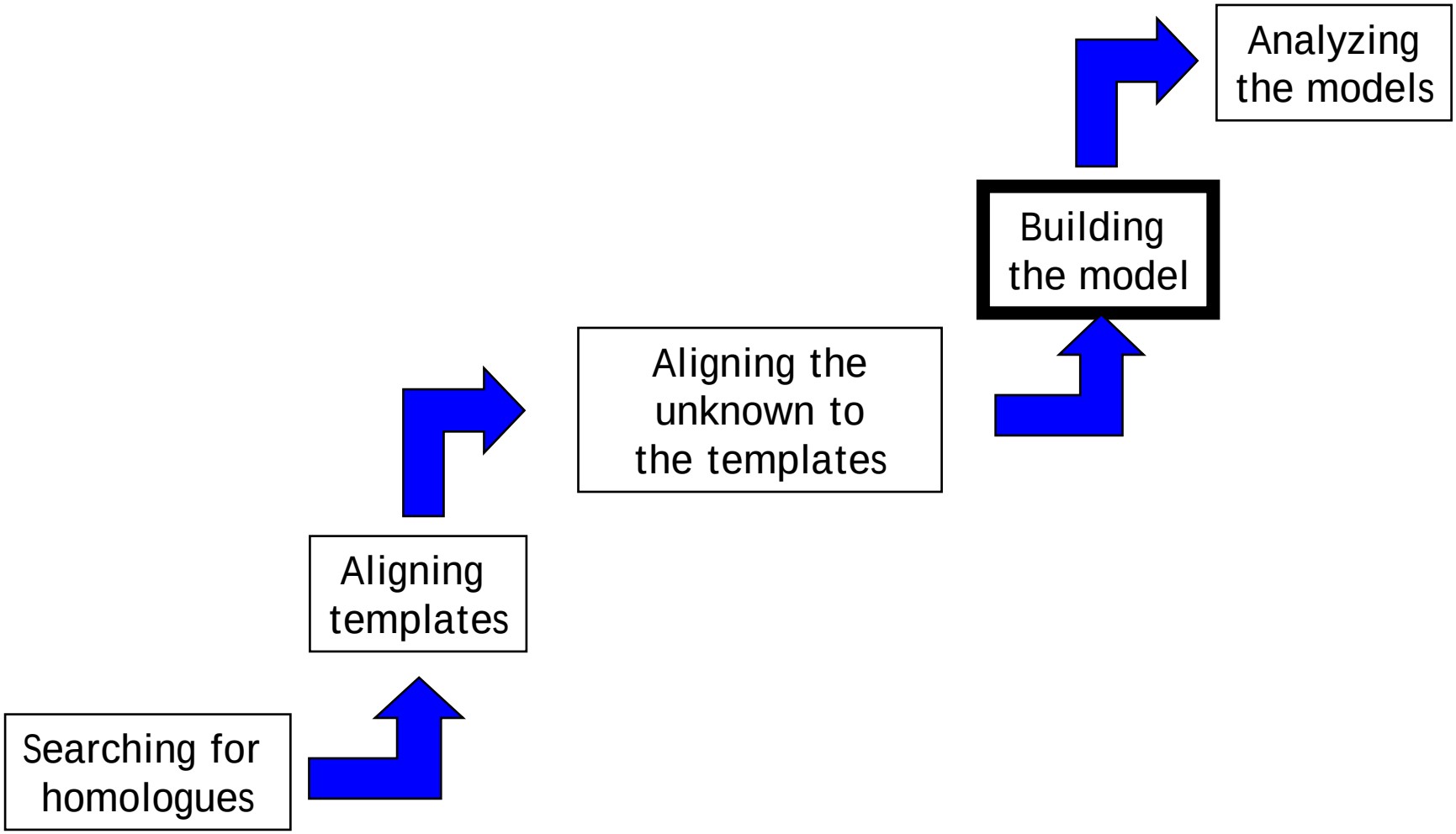
- Stand-alone scripting context
  - Script output (both error and normal) is displayed in command window
  - Can be run without launching the DS client
  - Full access to all scripting capabilities
  - Full access to Perl tools such as for debugging scripts
- DS Visualizer Pro client scripting context
  - Views on files opened by scripts
  - Views automatically updated when executing commands
  - Manipulating objects directly requires explicit update command
- Pipeline Pilot client scripting context
  - Full access to PP environment
  - Prepared for direct interactions between the scripts and the PP data stream


Visualization and Customization
Protein Modelling and Simulation
Pharmacophores
Structure-Based Design
Structure-Activity Relationships
Library Design and Analysis

Solution & Workflow for Protein Modeling



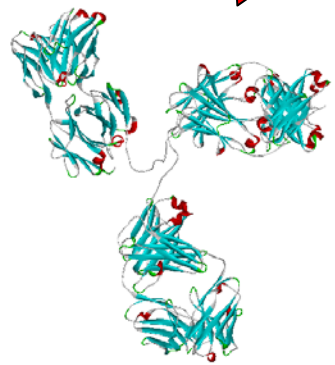
Structure Prediction by Homology Modeling



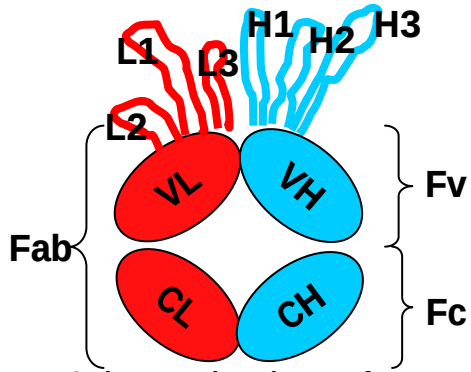
Generate Accurate Antibody Models

New

- Provide accurate antibody models of loops, side-chains to complete antibody-antigen complexes
- Determine the best antibody loop structures based on searching a pre-build canonical loop database
- Accurately predict the antigen-antibody binding site interactions using protein-protein docking
- Workflow*
  1. Sequence based search for homolog
  2. Sequence Alignment
  3. Building of Homology Model
  4. Analysis of Homology Model
  5. Identification of CDR loops
  6. Refinement of the loops



Refined 3D antibody model based on sequence alone

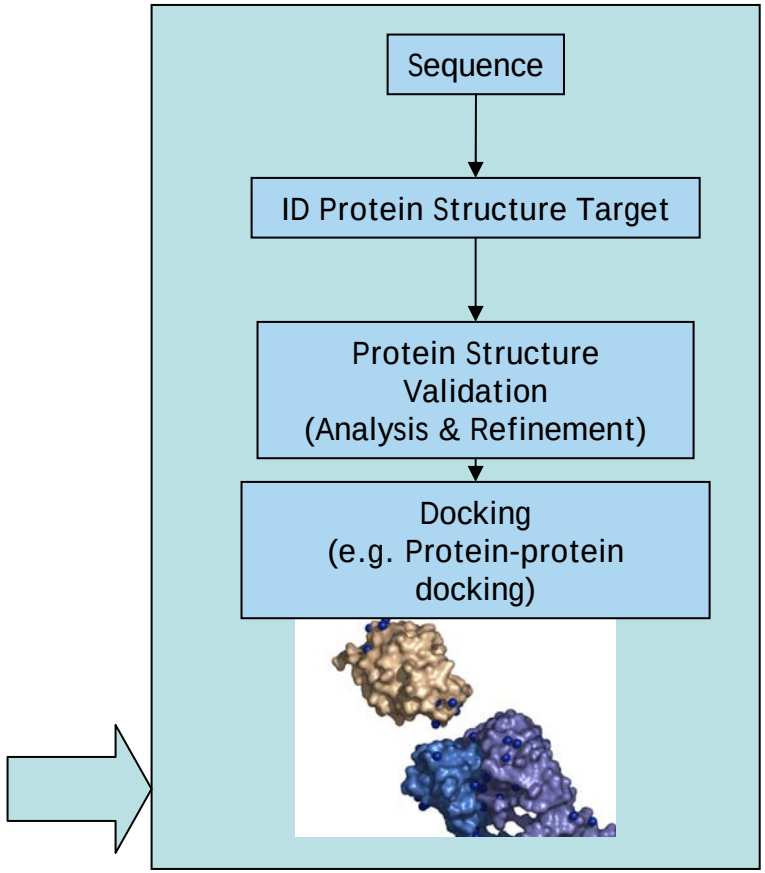


Schematic view of an Antibody Fab fragment with canonical loop region (H1-3, L1-3)

*Antibody Application Note: http://www.accelrys.com/reference/cases/studies/antibody_modeling_app_guide.pdf

Tools for Antibody Model Creation

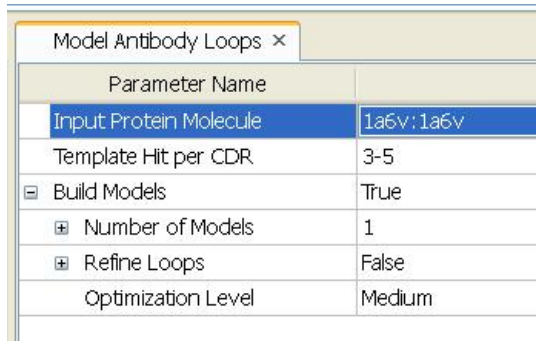
1. Utilise an integrated set of protocols and tools to model high quality antibody structures
   - Use pre-defined protocols for template searching, homology modelling, and loop and side-chain refinement
   - Combine sequence analysis, protein modelling, simulations and docking tools for complete structural analysis
   - Incorporate 3rd-party algorithms and custom protocols to create automated workflows from gene translation through a refined docked antigen-antibody complex



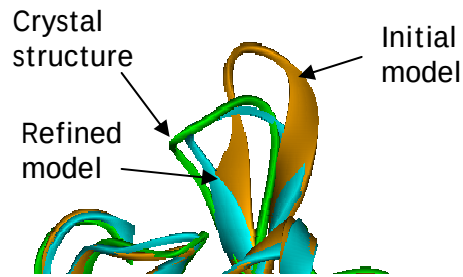
Example Antibody Modeling Workflow

Correctly Model Antibody Loop Structures

2. Determine the best antibody loop structures in the complementary determining region (CDR) based on searching a pre-build loop database
   - Easy-to-use protocol that automatically searches a database of canonical loop sequences and structures
   - Locates and generates several high quality antibody loop models
   - Provides alignment and scored models to easily rank and visualize
   - Builds models based on the top performing homology modeling algorithm, MODELER, developed by Professor Sali, UCSF



Pre-built 'Model Antibody Loops' protocol



Improve modeling of CDR loops using unbiased physics based methods**.

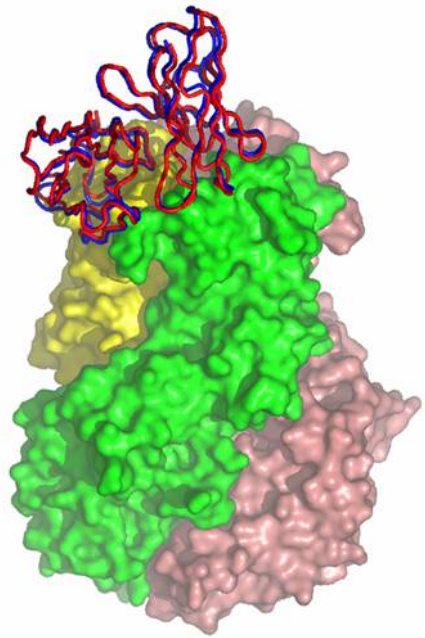
Over 30 groups used MODELLER at CASP7*, including the 11 of the top performers

* CASP, Critical Assessment of Techniques for Protein Structure Prediction : http://predictioncenter.gc.ucdavis.edu/

** Spassov, V., Flook, P., and Yan, L. "Looper: A Molecular Mechanics Based Algorithm for Protein Loop Prediction". *Manuscript in preparation.*

Locate Antigen-Antibody Interaction Site

3. Predict the precise docked structure of an Antigen bound to the Antibody
   - Quickly and accurately determine the structure of Antigen-Antibody complexes
   - Use easy-to-use pre-built protocols based on ZDOCK and RDOCK algorithms developed by Professor Weng from Boston University for protein-protein docking and re-ranking of docked hits



ZDOCK/RDOCK was used to correctly dock an antibody Fab fragment (red/blue) variable domain to a rotavirus VP6 virus coat protein (CAPRI* results)

* CAPRI: Critical Assessment of PRotein Interactions (Evaluation meeting for protein-protein docking methods) http://capri.ebi.ac.uk/)

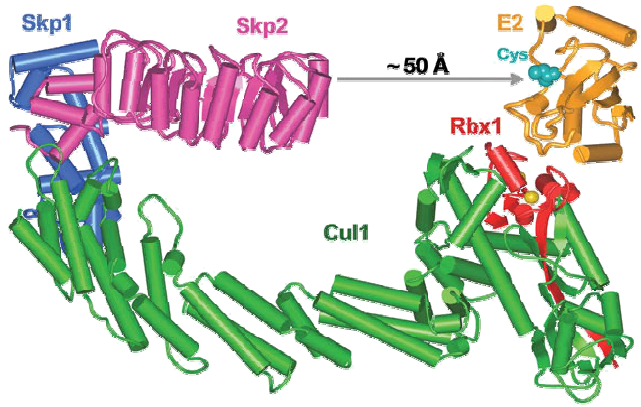
Selected Antibody Modelling References

- Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M., "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," J. Comput. Chem., 1983, 4, 187-217.
- Eswar, N., Eramian, D., Webb, B., Shen, M., Sali. A. "Protein Structure Modeling With MODELLER". Current Protocols in Bioinformatics John Wiley & Sons, Inc., , Supplement 15, 5.6.1-5.6.30, 2006
- Morea, V., Lesk, A., and Tramontano, A. "Antibody Modeling: Implications for Engineering and Design," METHODS, 2000, 20, 267.
- Spassov, V., Yan, L. and Flook, P. "The Dominant Role of Side-chain Backbone Interactions in Structural Realization of Amino-acid Code. ChiRotor: a Side-chain Prediction Algorithm Based on Side-chain Backbone Interactions," Protein Science, 2007
- Spassov, V., Flook, P., and Yan, L. "Looper: A Molecular Mechanics Based Algorithm for Protein Loop Prediction". Manuscript in preparation.

Predict Protein-Protein Binding Interfaces

New

- Experimental structure determination (e.g. x-ray crystallography, NMR) is slow and expensive
- In silico protein-protein docking provides faster, simple ways to find interactions between proteins
  - Locate the correct binding interface between two or more proteins rapidly and accurately
  - Fully understand protein interaction networks such as signal transduction pathways in a cell
  - Focus on small molecule, peptide or protein inhibitors or activators for drug discovery
  - Predict multiple protein assembly structures

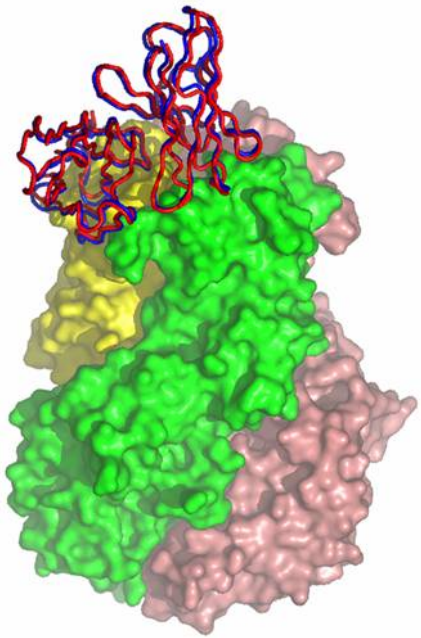


Determine protein-protein interfaces of large multiple protein complexes like the Ubiquination complex

Dock Protein (ZDOCK)

Provides Fast and Accurate Results

1. ZDOCK and RDOCK algorithms for protein-protein docking and refinement provide quick, accurate docked complex hits
   - Developed by Prof. Zhiping Weng, Boston University
   - ZDOCK is a fast, initial stage algorithm for unbound, rigid-body docking
     - An FFT-based method using a pair-wise shape complementarity function for identifying docked conformations
     - Scores hits based on atomic contact energies, desolvation and electrostatics parameters
   - RDOCK re-ranks (refines) the ZDOCK hits based a multi-staged CHARMm energy minimization method

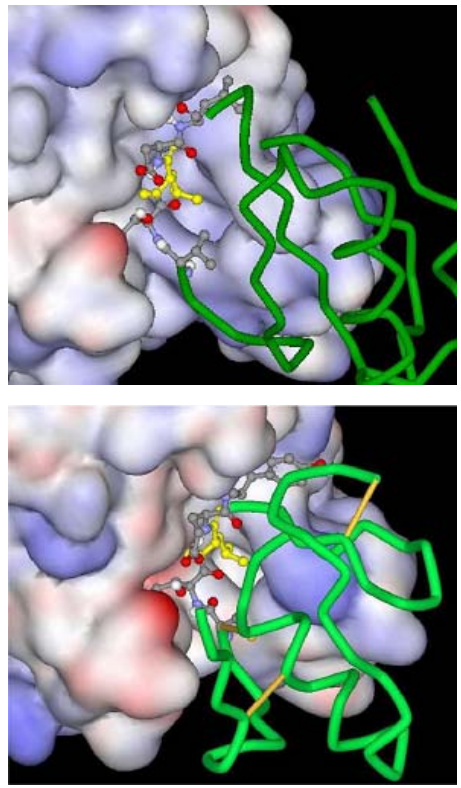


ZDOCK/RDOCK was used to correctly dock an antibody Fab fragment variable domain to a rotavirus VP6 virus coat protein

Dock Protein (ZDOCK)

Obtain Atomic Level Details of Binding Interfaces

2. Visualize atomic interactions between docked proteins
   - Understand interactions in signal transduction pathways and drug discovery
   - Determine residues and atoms involved in protein inhibition or activation
   - Confirm differences in binding affinities with different protein ligands
   - Provide a quantitative energetic measure of binding

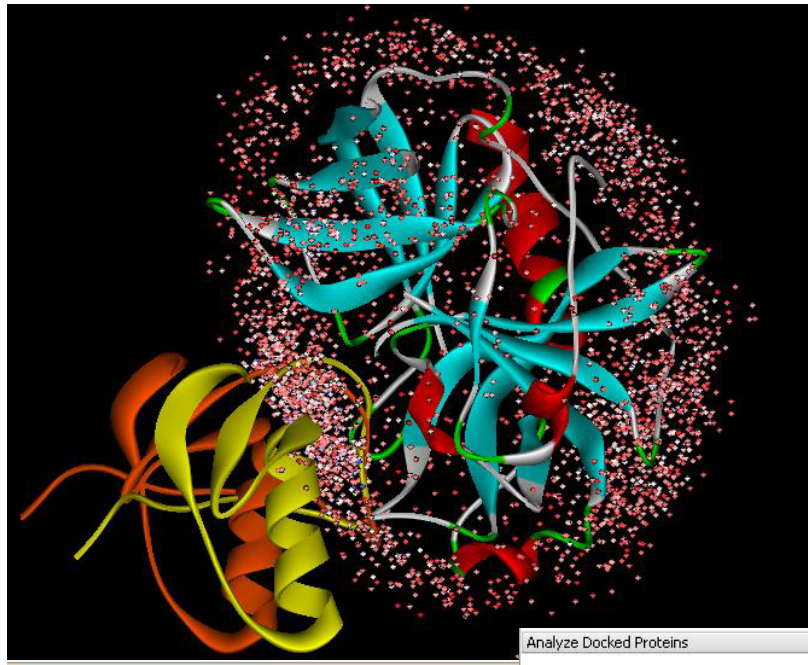


A protein-protein docking study of an alpha-lytic protease bound to 2 different inhibitors, Eglin C (top) & OMTKY3 (bottom). Eglin C show stronger binding strength versus OMTKY3, which correlate with *in vitro* binding studies.

Qasim et al. Biochemistry. 2006 Sep 26;45(38):11342.

Protein-Protein Docking in Discovery Studio

- Pre-defined, easy-to-use protocols for protein-protein docking, re-ranking and clustering of hits
- Comprehensive set of tools for analyzing results of docked hits
- Integrated platform for a complete workflow from sequence to refined docked complex
  - Combine sequence analysis, homology modeling, simulations and other docking tools
  - Incorporate 3rd-party algorithms and create complex scripts



Refine Docked Proteins (ROCK) | Dock Proteins (ZDOCK)

Parameter Name	
Input Receptor Protein	ZDockResults:1CGI_r
Input Ligand Protein	ZDockResults:1CGI_l
Ligand Blocked Residues	
Receptor Blocked Residues	
Angular Step Size	15
Filter Poses	
Receptor Binding Site Residues	
Ligand Binding Site Residues	
Distance Cutoff	10.0

Analyze Docked Proteins

Display
- Center Receptor/Ligand
- Show Full Docked Complex
- Select Binding Interface

RMSD
- Define RMSD Reference
- Calculate Binding Site RMSD (ZDOCK)

Clustering
- Show Cluster Center/Cluster
- Show Only Lowest Energy Pose
- Show All Poses in Cluster

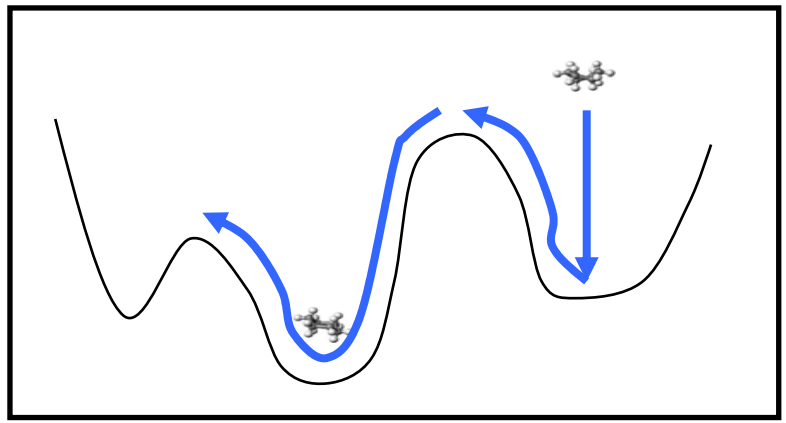
Receptor bound to crystal structure (orange) and best docked pose (yellow) ligands in similar location validating the docked results. Dots around receptor indicate ~2000 possible docked poses.

ZDOCK/RDOCK Selected References

- Methodology
  - Mintseris J, Pierce B, Wiehe K, Anderson R, Chen R, Weng Z Integrating Statistical Pair Potentials into Protein Complex Prediction. Proteins, in press (Accepted)
  - Pierce B, Weng Z (2007) ZRANK: Reranking Protein Docking Predictions with an Optimized Energy Function. Proteins 67(4), 1078-1086
  - Li L, Chen R (joint first authors), Weng Z (2003) RDOCK: Refinement of Rigid-body Protein Docking Predictions. Proteins 53, 693-707.
  - Chen R, Li L, Weng Z (2003) ZDOCK: An Initial-stage Protein-Docking Algorithm. Proteins 52, 80-87
  - Chen R, Weng Z (2003) A Novel Shape Complementarirty Scoring Function for Protein-Protein Docking. Proteins 51, 397-408
  - Chen R, Weng Z (2002) Docking Unbound Proteins Using Shape Complementarity, Desolvation, and Electrostatics. Proteins 47, 281-294
- CAPRI
  - Wiehe K, Pierce B, Mintseris J, Tong W, Anderson R, Chen R, Weng Z (2005) ZDOCK and RDOCK performance in CAPRI rounds 3, 4, and 5. Proteins 60(2), 207-21

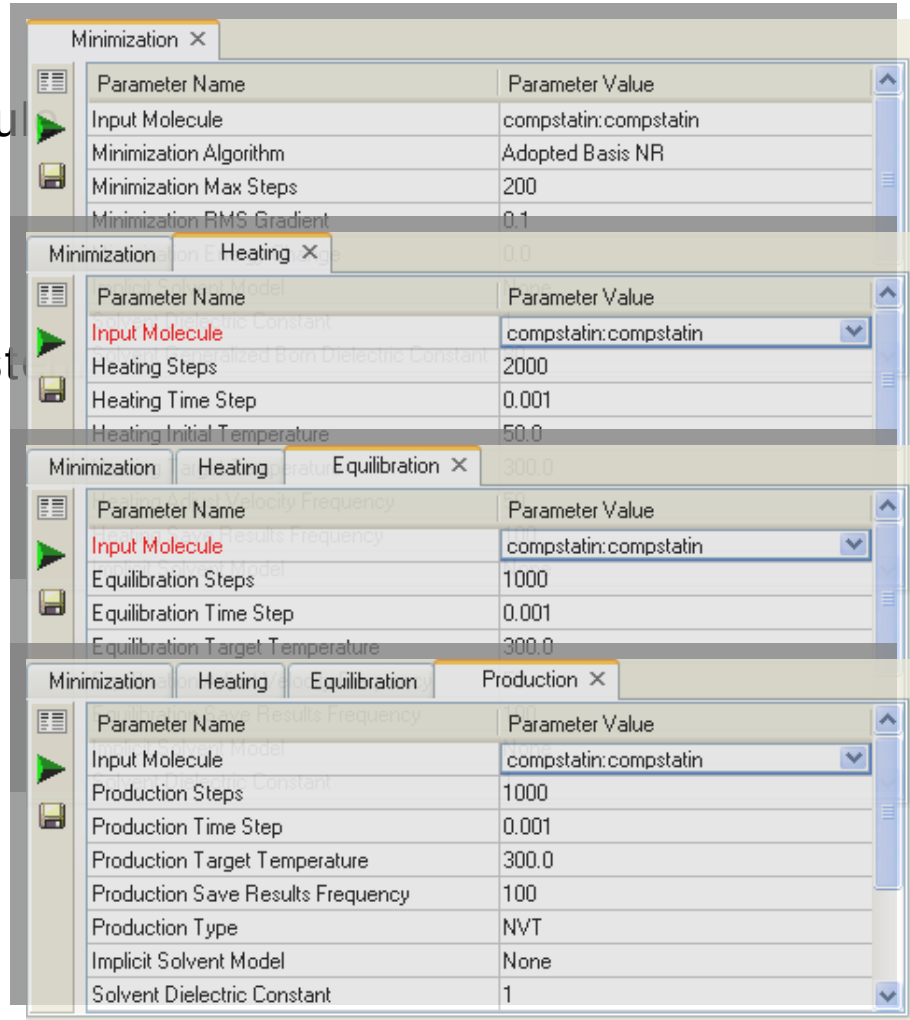
Molecular Mechanics/Dynamics (CHARMm)

- Structural changes of molecules
  - Conformational changes
  - Domain flexibility
  - Protein folding
- Refinement of docking experiments
  - Allow for induced fit
- Free energy changes
  - Calculation to binding energies
- Determination of thermodynamic properties
  - Enthalpy and entropy changes



Steps for an MD Simulation

- Prepare molecule
  - Read in a file or build the molecul
- Minimization
  - Essentially required
- Heating
  - Raise the temperature of the syst
- Equilibration
  - Ensure the system is stable
- Production
  - Collect your data
- Simulated annealing
  - Optional
  - Lowering the temperature



Protein Ionisation and pK

New

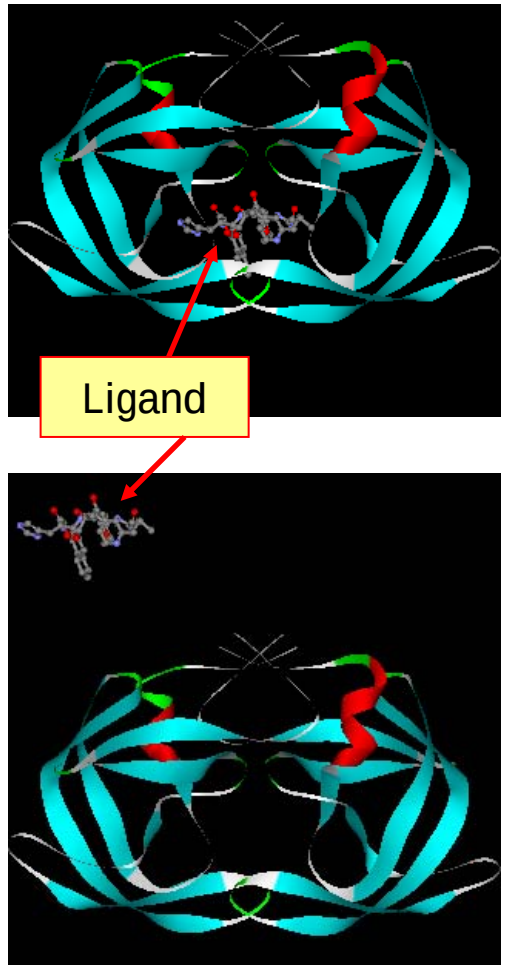
- Correct protonation states are critical for
  - Accurate docking and scoring of ligands to receptors
  - Stable, convergent molecular dynamics simulations
- Proper side-chain protonation usually ignored when computing electrostatic component of protein-ligand binding energy
  - Leads to inaccurate scoring of poses
- Accelrys solution
  - New electrostatics protocol ‘Calculate Protein Ionization and pK’
  - Based on CHARMm Generalized-Born methods
    - Faster and more accurate than existing Poison-Boltzmann methods
    - More accurate and rigorous than rule-based methods
  - Provides titration curves for specific residues using 3D environment in protein
  - Automatically set the protonation state of each residue under a specified pH
  - Calculate the pH dependent electrostatic energy
  - Automated and rigorous optimisation of symmetric acidic groups
    - Eg. Asp OD1 and OD2
  - Automated flipping of Asn and Gln O and N positions
  - Consistent CHARMm force field used throughout

Protein Ionisation and pK Prediction 1:

Accurate Protein-Ligand Binding Energies

- Estimate the electrostatic contribution to binding energy for protein-ligand and protein-protein docking
  - Protonation states of proteins may differ in docked and undocked form
    - This method could predict the protonation states accurately taking into account of the local environment changes upon ligand docking
  - Faster than DelPhi methods

Fast and accurate electrostatic component of binding energy due to consideration of proper protonation states

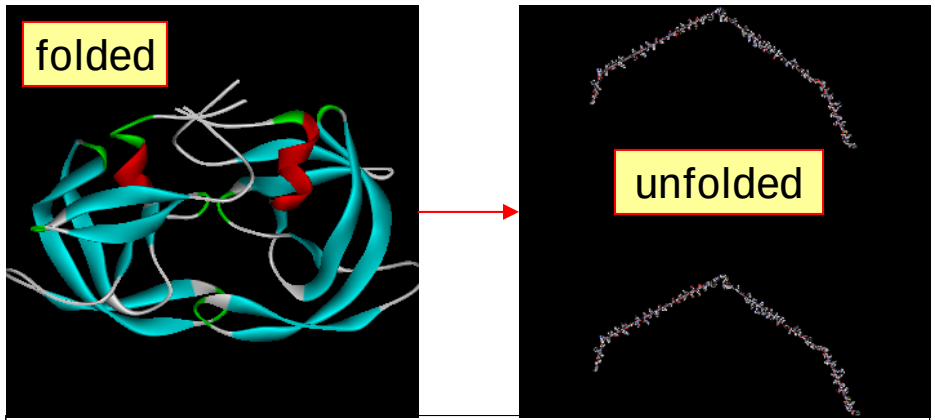


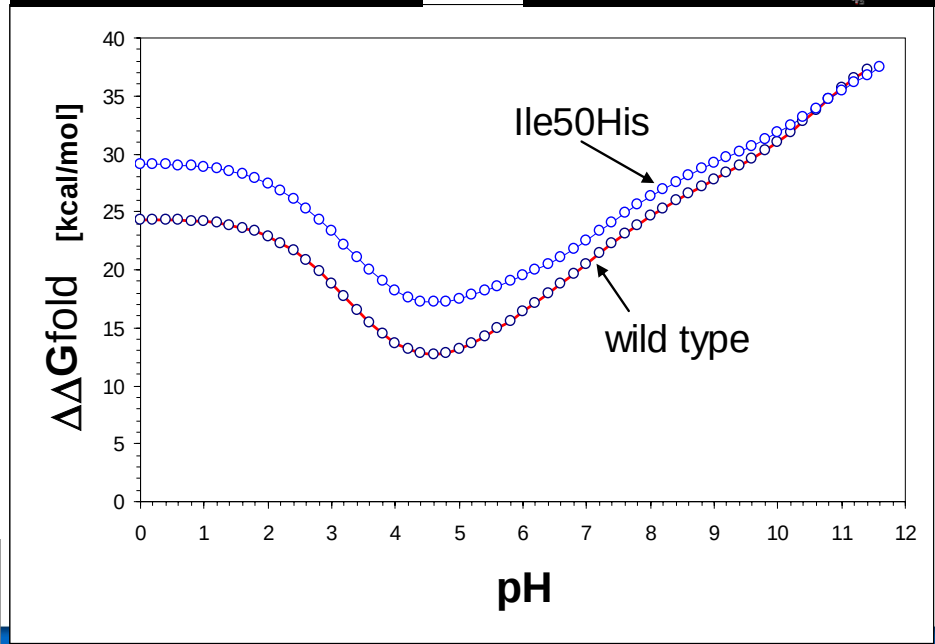
pH-Dependent binding energy of HIV protease

Protein Ionisation and pK Prediction 2:

pH Dependent Folding Energy

- Calculation done for HIV protease
- Optimal pH for protein stability
  - Experimental ~ 5
  - Predicted ~ 4.6
- Folding free energy change from pH 3.5 to 5
  - $\Delta\Delta G_{calc}$(pH=3.4->5.0) = 3.5 kcal/mol
  - $\Delta\Delta G_{exp}$ (pH=3.4->5.0) ~ 4.5 kcal/mol



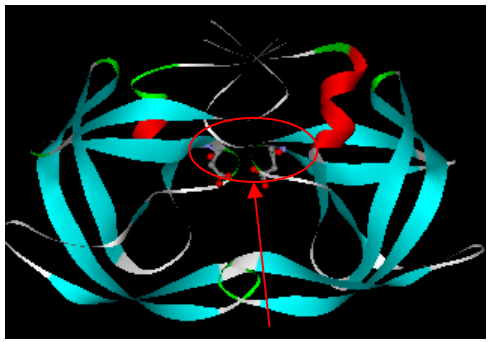


Optimal pH for stability critical for all biochemical assays

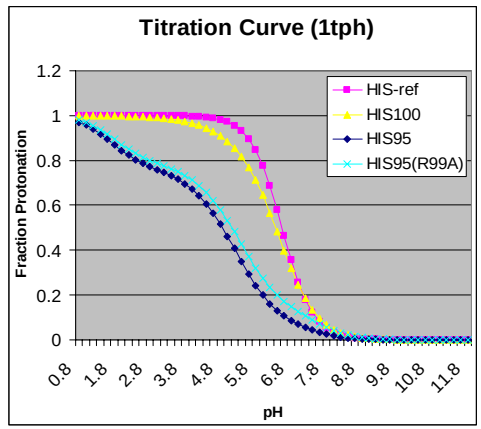
Protein Ionisation and pK Prediction 3:

Active Site Residue Prediction

- Predict potential active or post-translational modification sites based on analysis of titration curve for specific residues
- Abnormal curves indicate location



Standard pKa of an Aspartic Acid residue is 3.9. The pKa of the active site residue Asp25 in HIV protease is 6.9



The active site residue, HIS95, in the 1tph protein, shows an abnormal titration curve

Quickly identify key catalytic residues in proteins

Additional New Simulation Features in DS 2.0

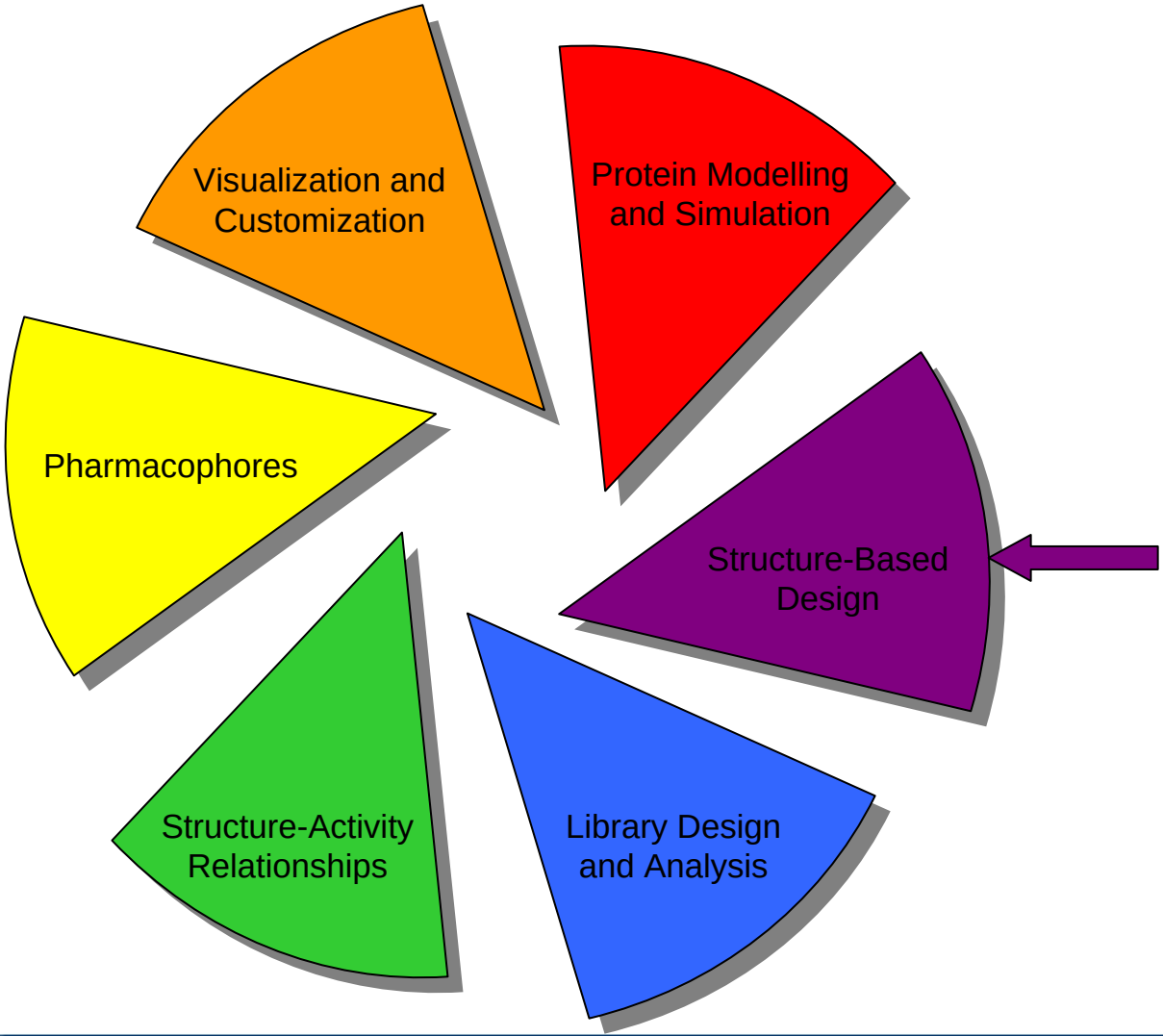
- Method for calculating vibrational entropy
- MMFF forcefield support
- Complete suite of trajectory analysis tools:
  - Radius of gyration (RGYR)
  - Clustering of trajectories
  - PCA on trajectories
  - Phi-Psi Time series
  - Radial distribution function
- CMAP1 support in charmm22, charmm27
  - a dihedral cross-term energy correction map that allows for an explicit 2D energy correction surface to be applied to any two dihedrals

Often leads to improved accuracy of MM-PBSA type scoring

Full support for trajectory analysis now available

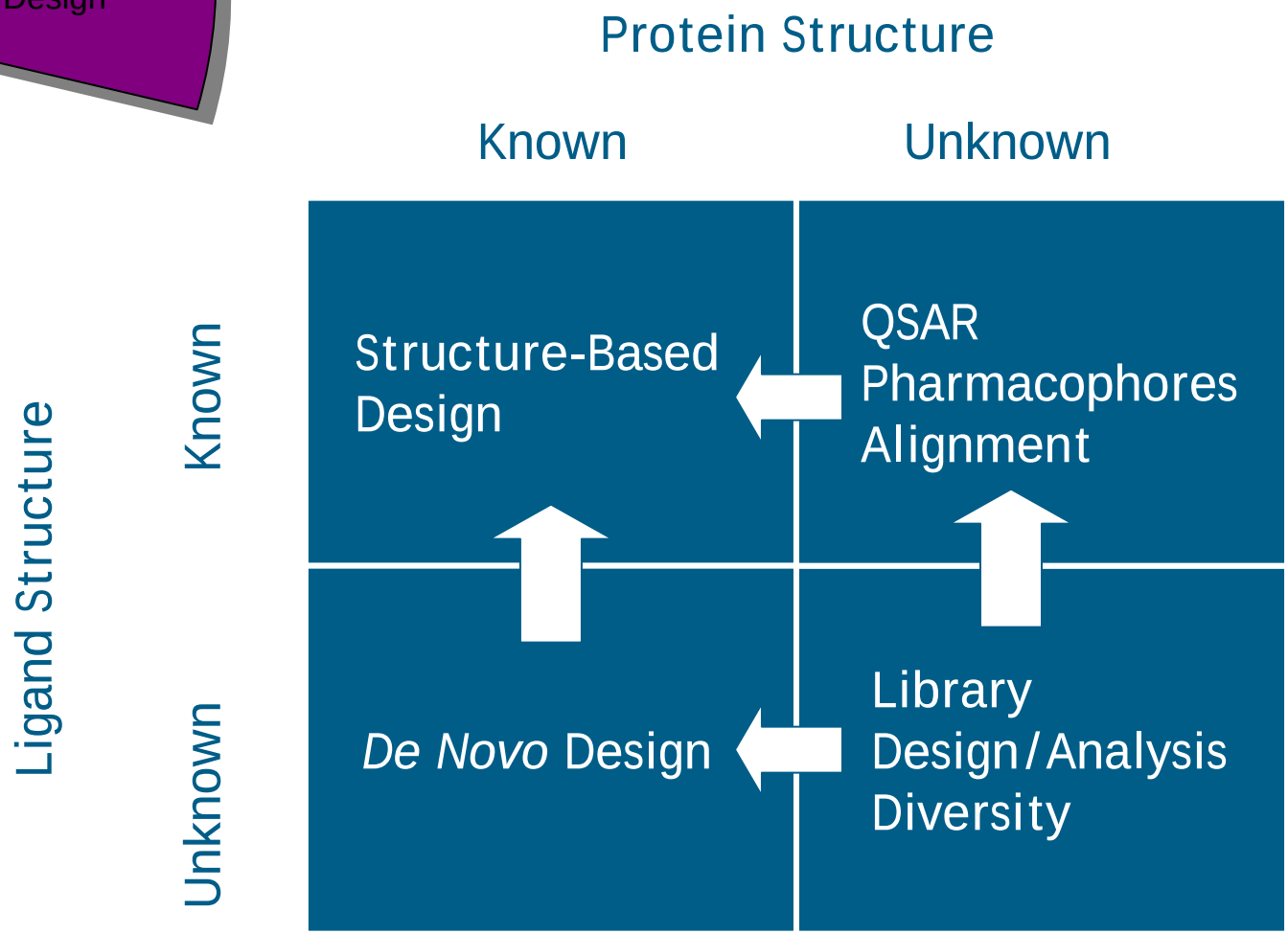
CHARMm developers have shown that CMAP improves stability of MD simulations

1. MacKerell, Jr. AD, Feig M, Brooks, III C. *J Comput Chem* 25: 1400-1415 (2004)

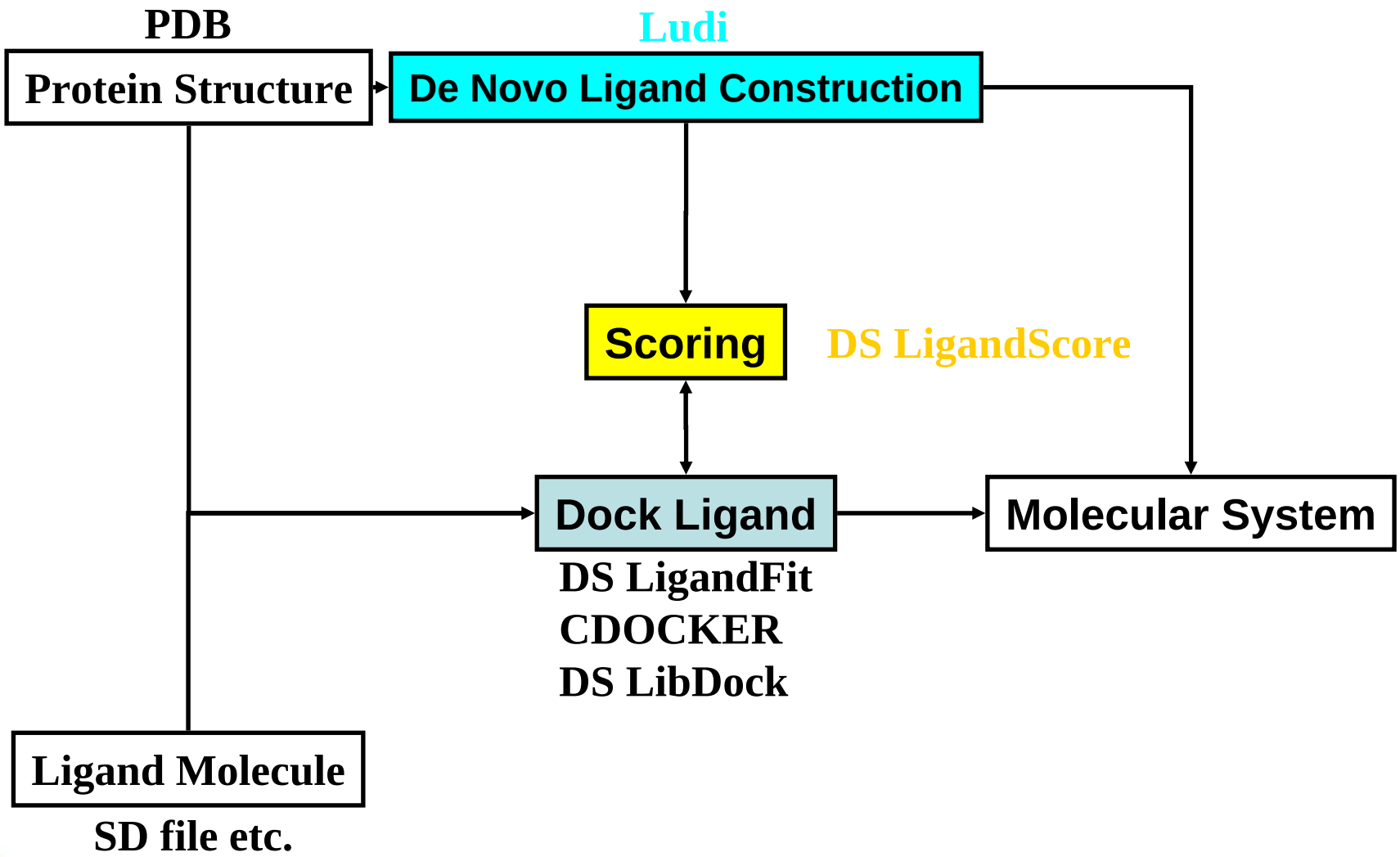

Visualization and Customization
Protein Modelling and Simulation
Pharmacophores
Structure-Based Design
Structure-Activity Relationships
Library Design and Analysis

Structure-Based Design

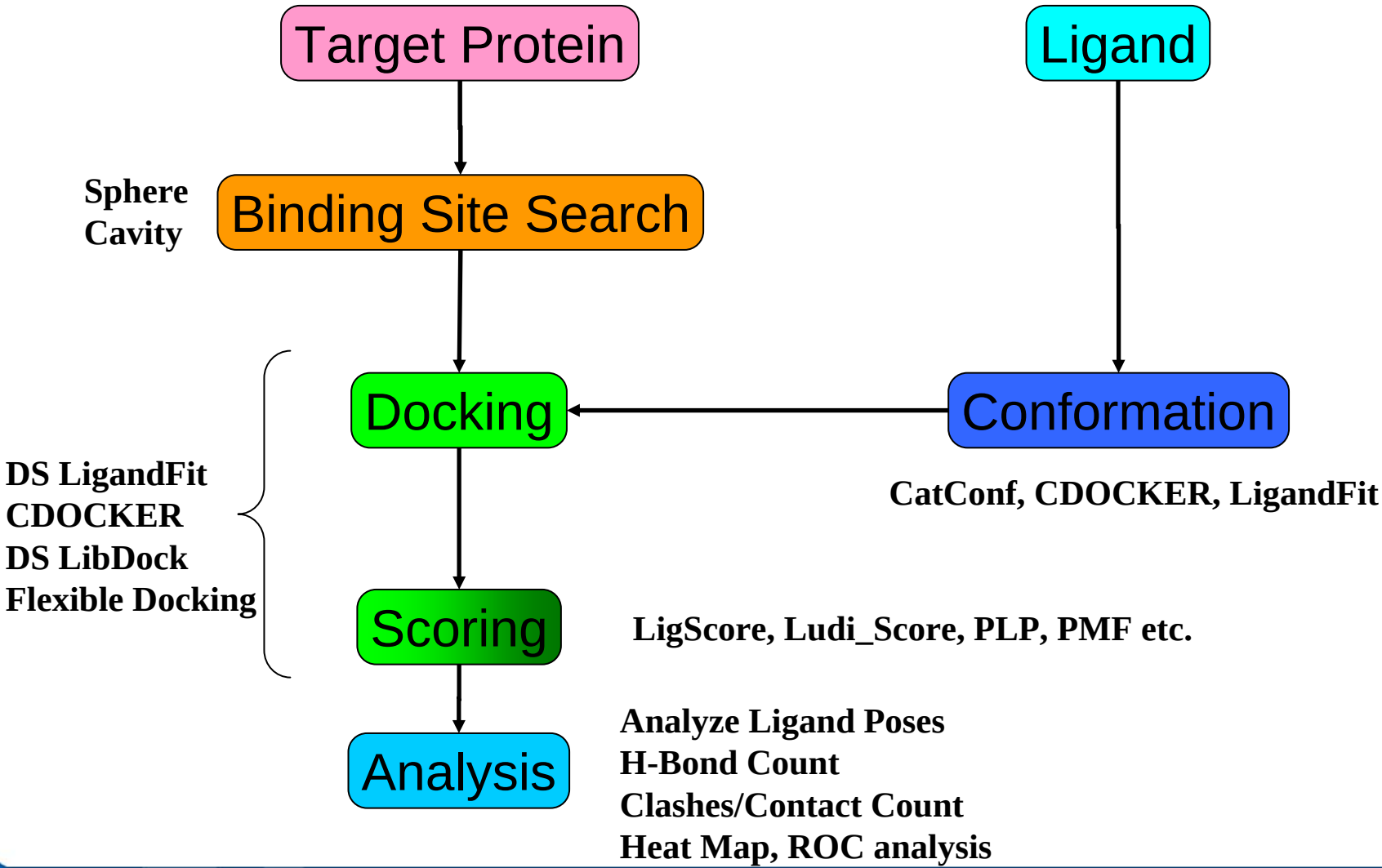
Drug Design Scenarios



Structure Based Design

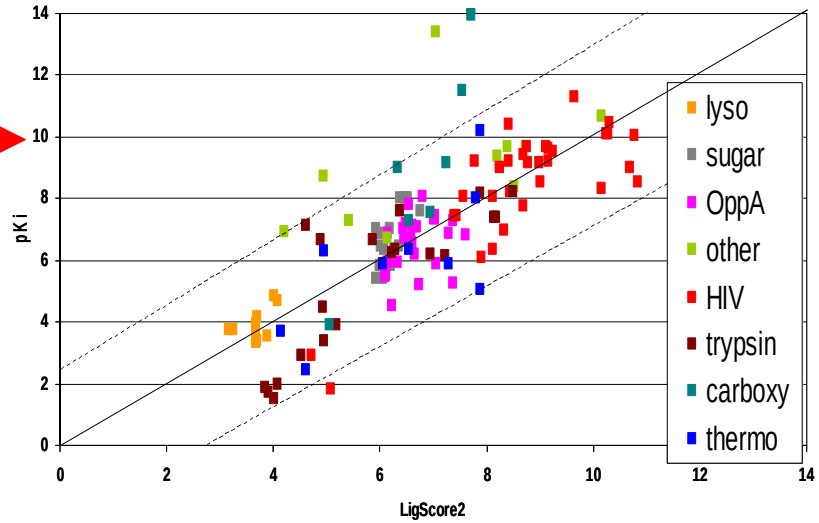


Docking WorkFlow

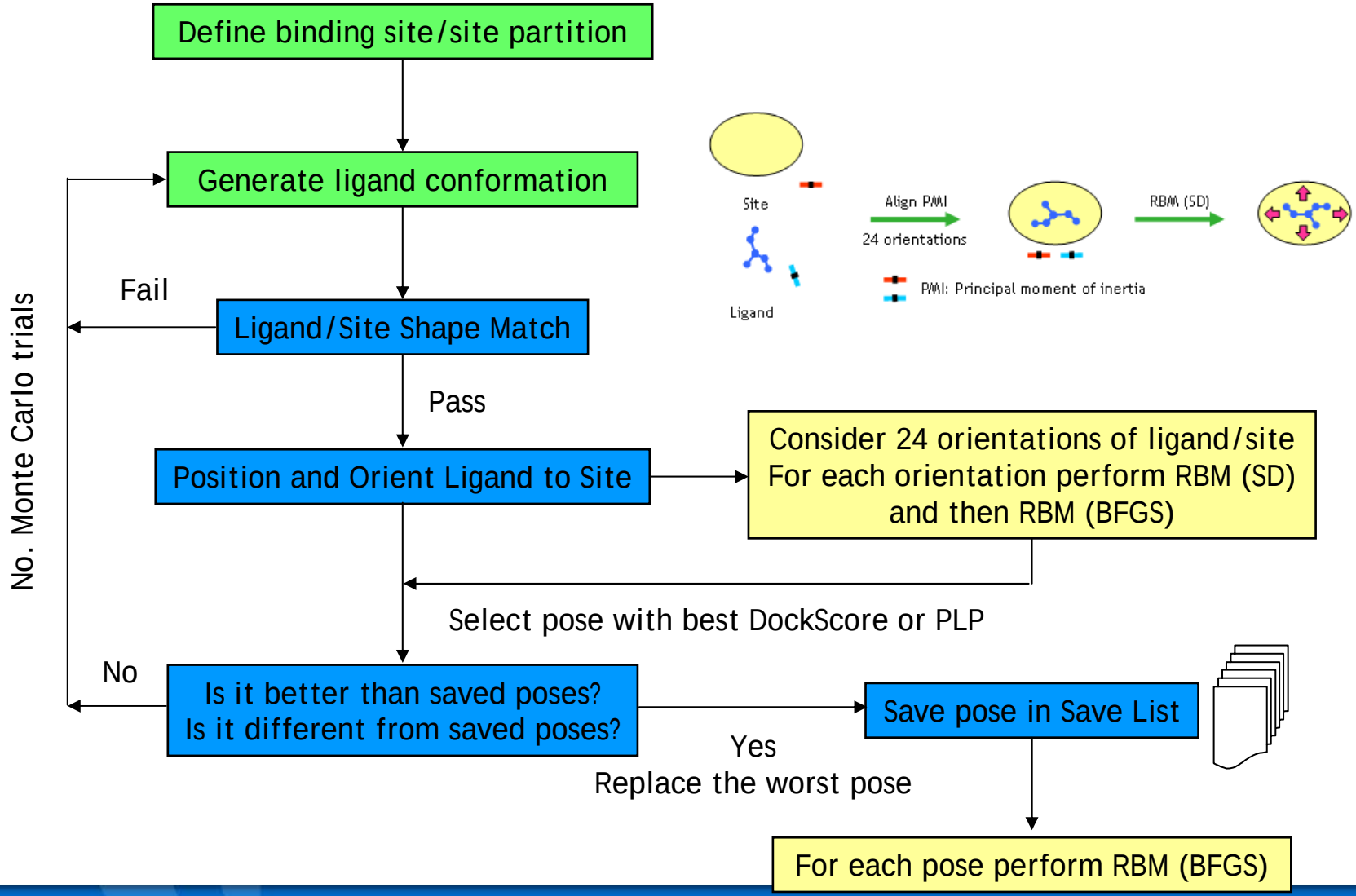


Scoring & Analysis

- LigandFit
  - DockScore = vdW + ele + “ligand_intra”
- CDOCKER
  - Potential ENERGY
- LibDock
  - PLP Like Score (pair-wise score)
- LigScore
  - vdW + surface_descriptors
- PLP
- PMF
- Ludi
- MM-PBSA/MM-GBSA
  - Energy Base
  - Rotation, Translation
- Consensus

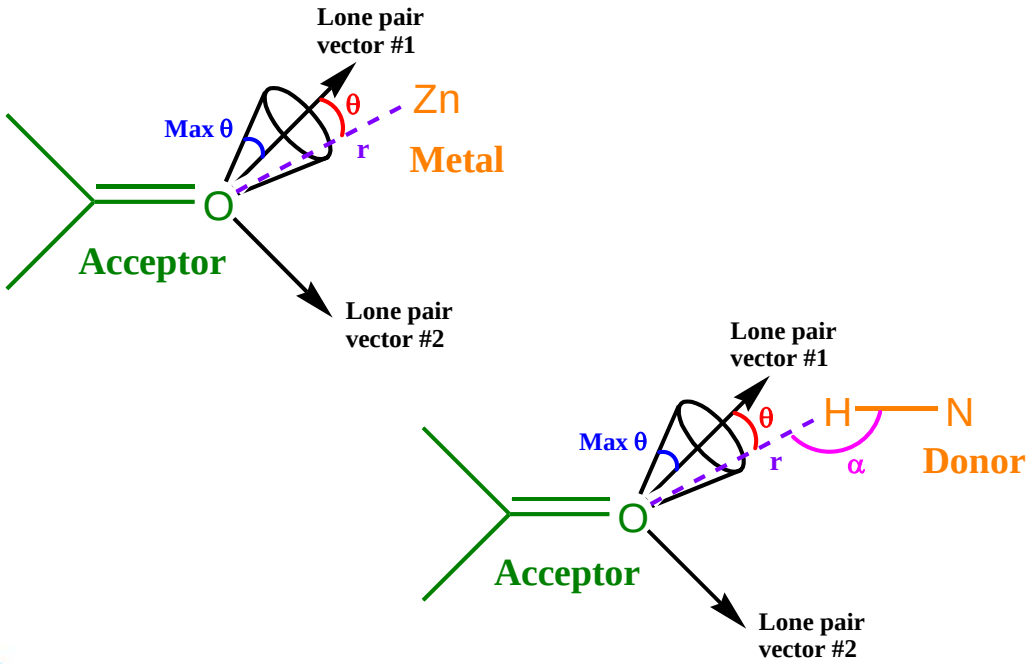


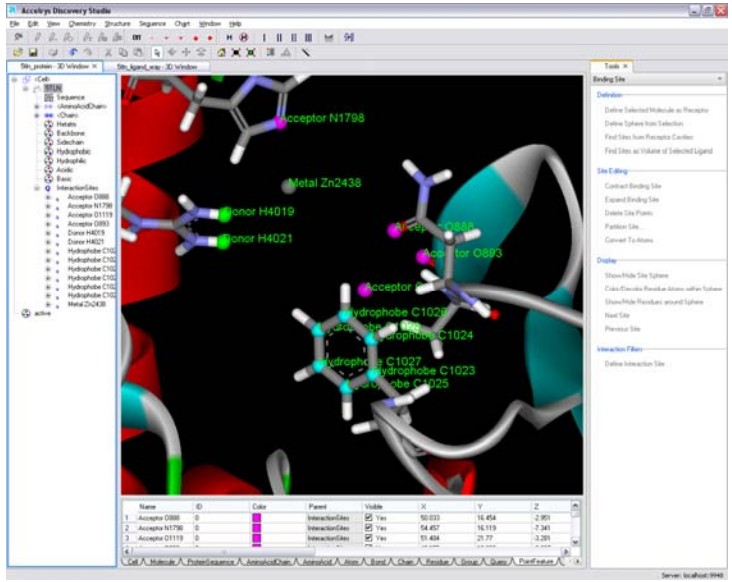
Dock Ligands (LigandFit) WorkFlow



Dock Ligands (LigandFit)

- Docking with interaction filters
  - ability to define receptor atoms that have one of these characteristics
    - H-bond donor
    - H-bond acceptor
    - hydrophobic contact
    - metal ion lone pair acceptor





Dock Ligands (LibDock)

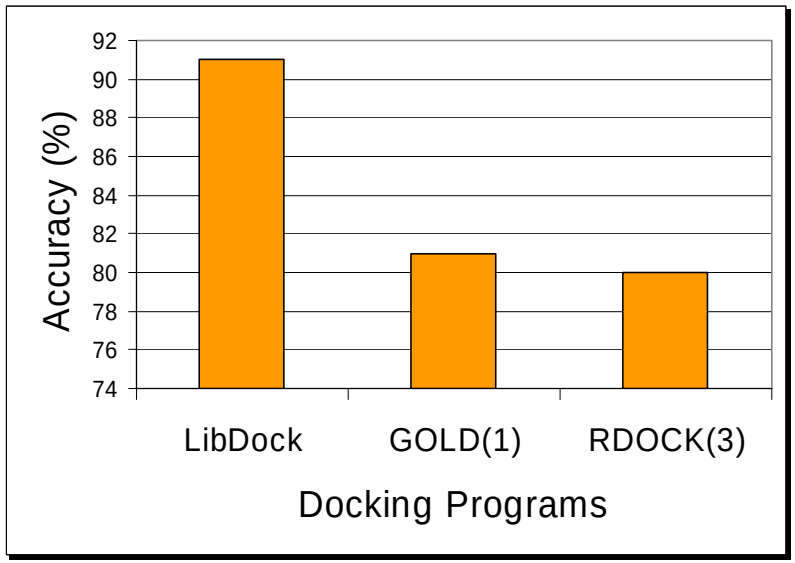
Fast and Accurate vHTS Solution

New

- Docking results on the newly available AstexDiverse[1] dataset
  - 85 receptor-ligand PDB entries comprising diverse receptor families
  - Docking step takes seconds per small molecule

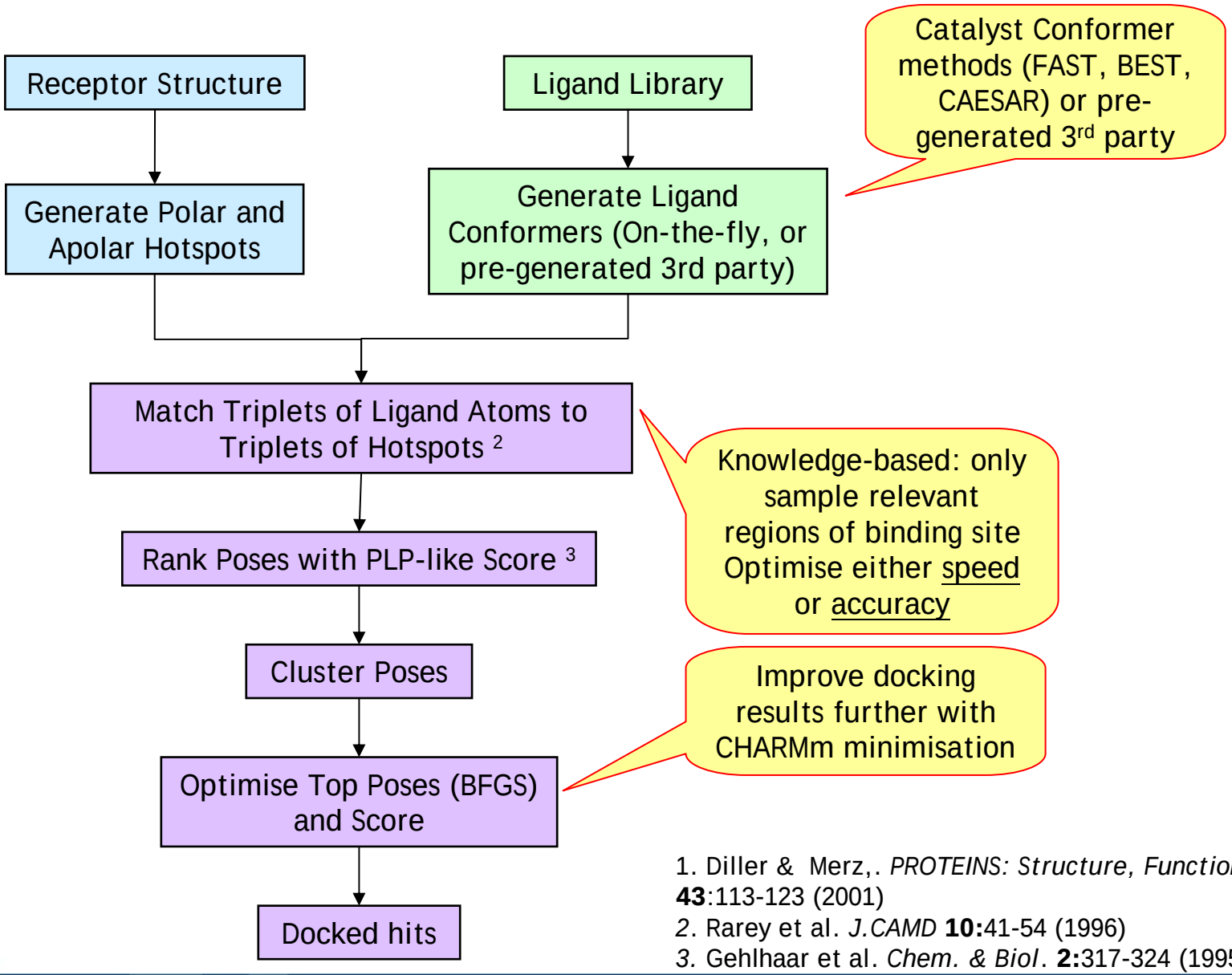
RMSD bin (Å)	LibDock % Docked Successfully	
	CatConf BEST	CAESAR[2]
<1	61%	67%
<2	91%	86%
Success Rate	91%	86%

% of AstexDiverse dataset docked successfully with LibDock and two conformation generation methods



Most Accurate results on AstexDiverse dataset available to date

1. Hartshorn, et al. J. Med. Chem., **50** (4), 726 -741 (2007)
2. Li et al, submitted to J. Chem. Inf. Model
3. Hubbard, R. CHI-SBD Conference, 2007



1. Diller & Merz,. *PROTEINS: Structure, Function and Genetics* **43**:113-123 (2001)
2. Rarey et al. *J.CAMD* **10:**41-54 (1996)
3. Gehlhaar et al. *Chem. & Biol*. **2:**317-324 (1995)

Dock Ligands (CDOCKER)

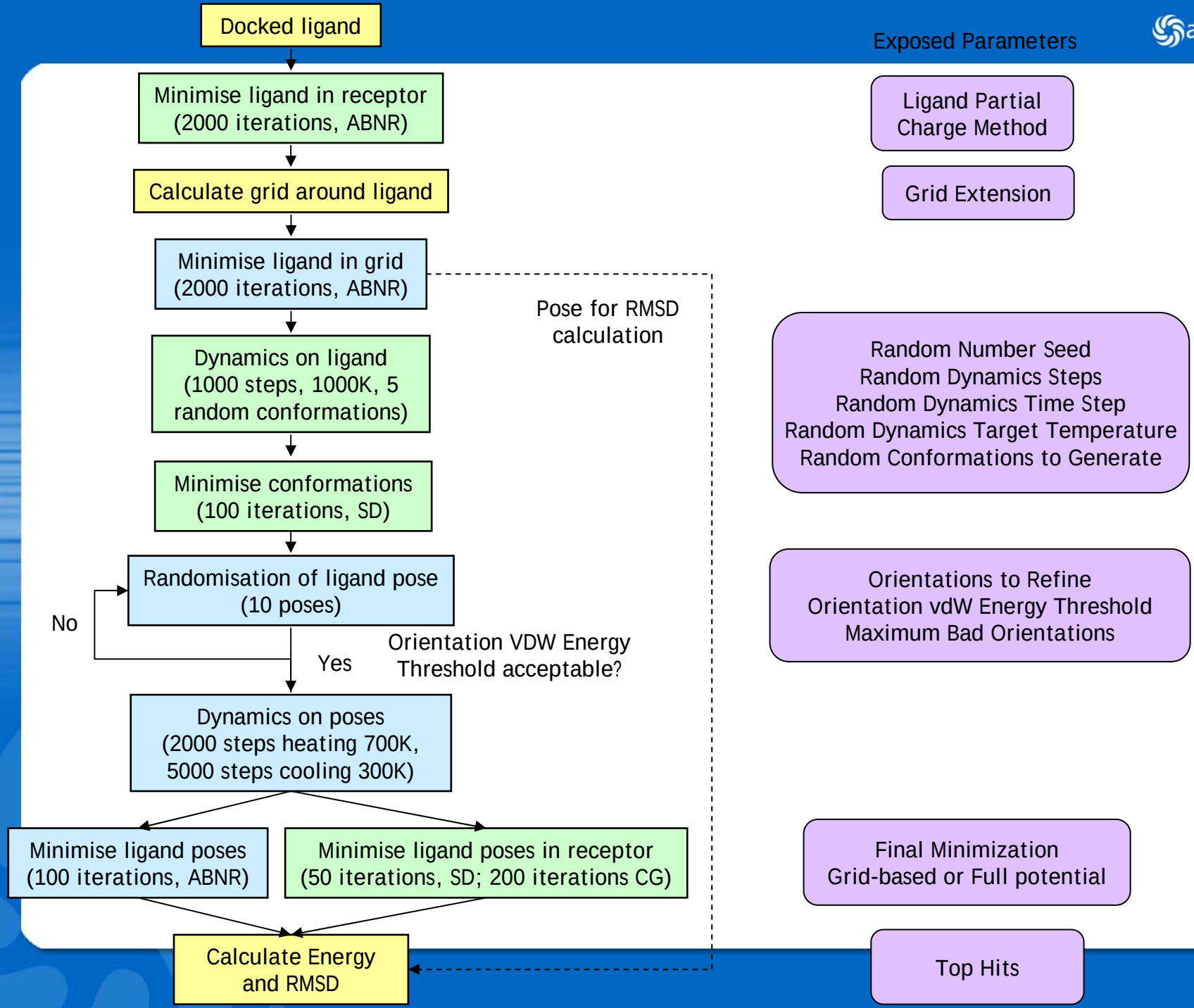
Accuracy Before Speed

- CHARMm-based docking/refinement algorithm
- Uses soft-core potentials and an optional grid representation to dock ligands into the receptor active site
  - High temperature MD to generate (5) starting conformations
  - Take each conformation and perform random rigid body rotations (10)
  - Minimise resulting structures (<=50)

New

- Improved sampling
- More parameters exposed
- Constrained docking (Pharmacophore)

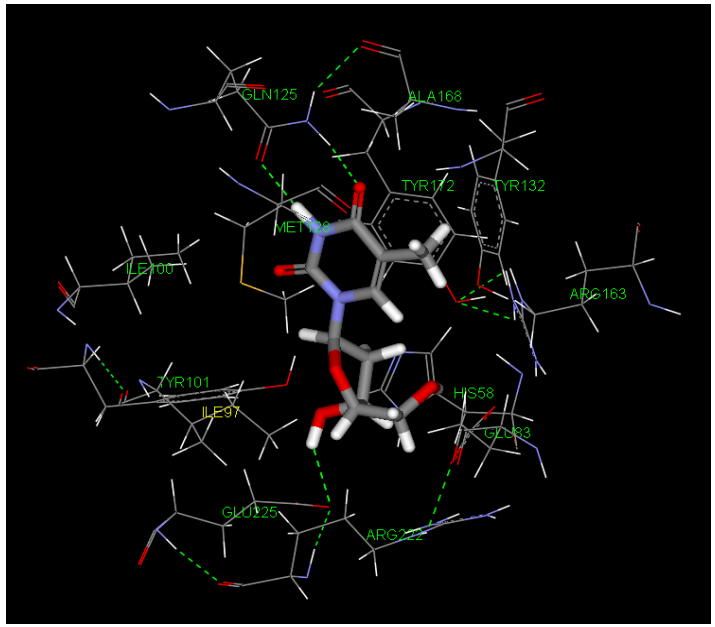
Use prior knowledge to reduce computation time


Docked ligand
Minimise ligand in receptor (2000 iterations, ABNR)
Calculate grid around ligand
Minimise ligand in grid (2000 iterations, ABNR)
Dynamics on ligand (1000 steps, 1000K, 5 random conformations)
Minimise conformations (100 iterations, SD)
Randomisation of ligand pose (10 poses)
No
Yes
Orientation VDW Energy Threshold acceptable?
Dynamics on poses (2000 steps heating 700K, 5000 steps cooling 300K)
Minimise ligand poses (100 iterations, ABNR)
Minimise ligand poses in receptor (50 iterations, SD; 200 iterations CG)
Calculate Energy and RMSD
Pose for RMSD calculation
Exposed Parameters
Ligand Partial Charge Method
Grid Extension
Random Number Seed
Random Dynamics Steps
Random Dynamics Time Step
Random Dynamics Target Temperature
Random Conformations to Generate
Orientations to Refine
Orientation vdW Energy Threshold
Maximum Bad Orientations
Final Minimization
Grid-based or Full potential
Top Hits

Flexible Docking

New

- Ignoring receptor flexibility during docking may lead to inaccurate poses
  - The problem is magnified in vHTS
- Accelrys solution
  - Docking into realistic receptor environment
  - Several low-energy initial receptor conformations used
  - Docking of ligand influenced by existing side chains in binding site
  - All flexible residues can be included in the computation
  - Library of receptor conformations can be saved and used for all ligands
  - Fully automated workflow uniquely tuned for vHTS applications



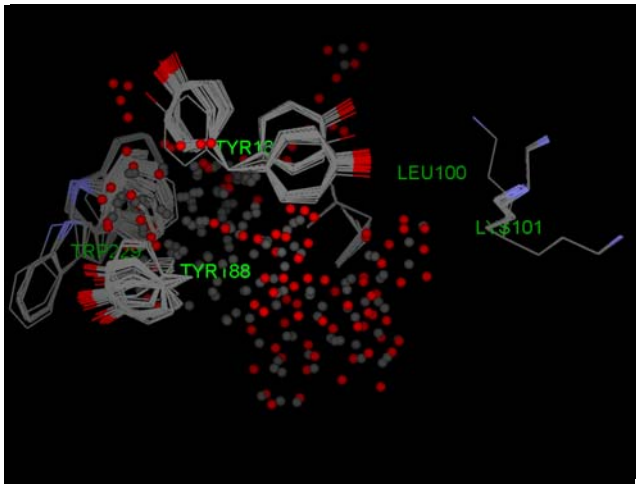
Complex network of interactions in Thymidine Kinase (PDB ID 1kim)

Flexible Docking Protocol

Step 1: Generate Receptor Conformations

Receptor

Generate Receptor Side Chain Conformations (n) (ChiFlex)



Side chain conformations identified by ChiFlex in PDB ID 1rev

- Generate reasonable low energy side chain conformations
- Side-chain/backbone interactions explicitly taken into account
- Any number of resides can be included
- Need to run only once per receptor binding site

Flexible Docking Protocol

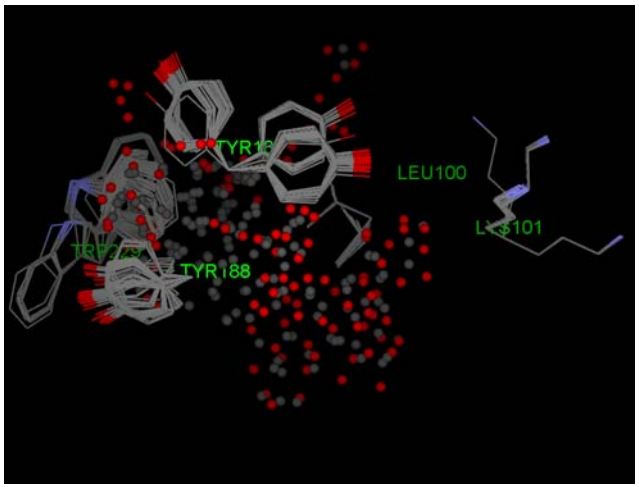
Step 2: Compute Protein Hotspots

Receptor

Generate Receptor Side Chain Conformations (n) (ChiFlex)

↓

Compute Hotspots on Conformation n



Receptor hotspots for PDB ID 1rev

Simple two-feature model has been shown to be accurate and robust for guiding docking [1]

Red = polar hotspots
Gray = apolar hotspots

Flexible Docking Protocol

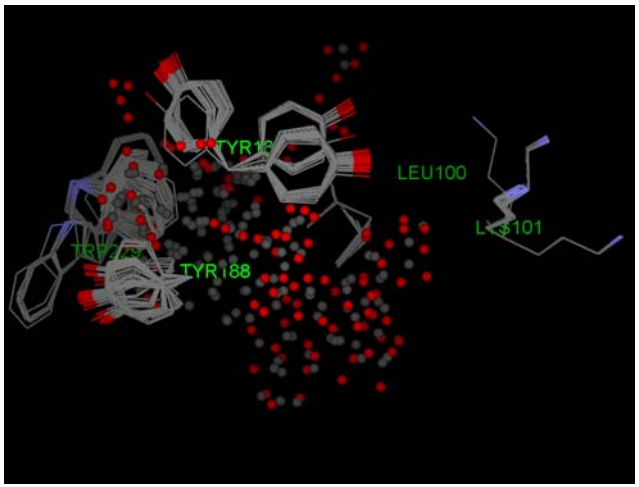
Step 3: Generate Small Molecule Conformations

Receptor

Generate Receptor Side Chain Conformations (n) (ChiFlex)

↓

Compute Hotspots on Conformation n



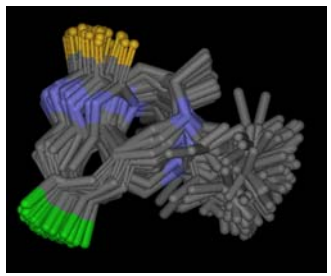
Receptor hotspots for PDB ID 1rev

Small Molecules

Generate Small Molecule Conformation (CatConf)

Several Methods Available: Optimise either speed or accuracy

- Generates diverse low energy conformers
- Fast (~ 50 compounds per minute)



Flexible Docking Protocol

Step 4: Fast and Efficient Docking to Hotspots

Receptor

Generate Receptor Side Chain Conformations (n) (ChiFlex)

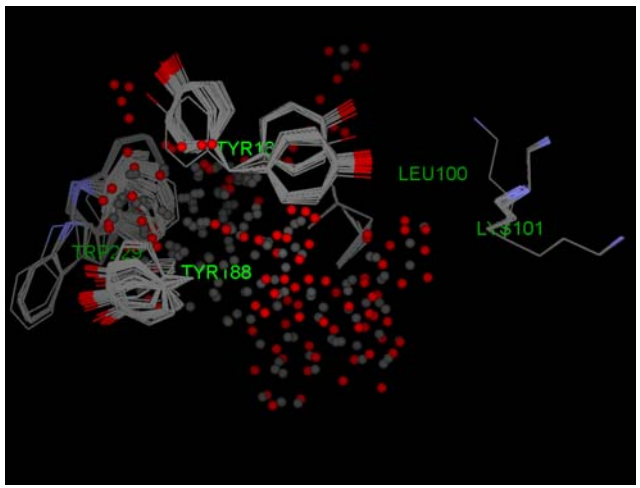
Compute Hotspots on Conformation n

Small Molecules

Generate Small Molecule Conformation (CatConf)

Dock to Hotspots (LibDock[2,3]) Retain m Poses

Knowledge-based docking: only sample relevant regions of binding site

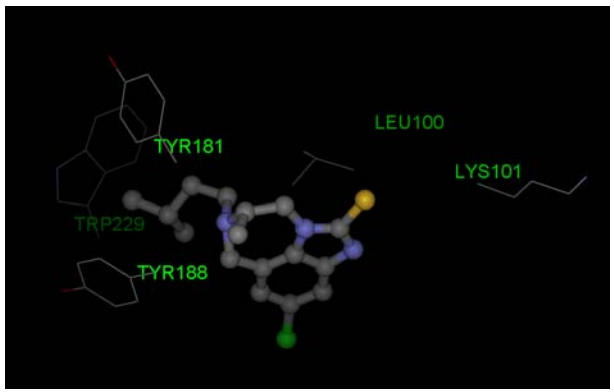


Ligand conformations docked to hotspots, PDB ID 1rev

- Triplets of hotspots matched to triplets of small molecule atoms
- Matches clustered and optimised

Flexible Docking Protocol

Step 5: Side-Chain Optimisation Around Docked Pose



Receptor

Small Molecules

Generate Receptor Side Chain Conformations (n) (ChiFlex)

Generate Small Molecule Conformation (CatConf)

Compute Hotspots on Conformation n

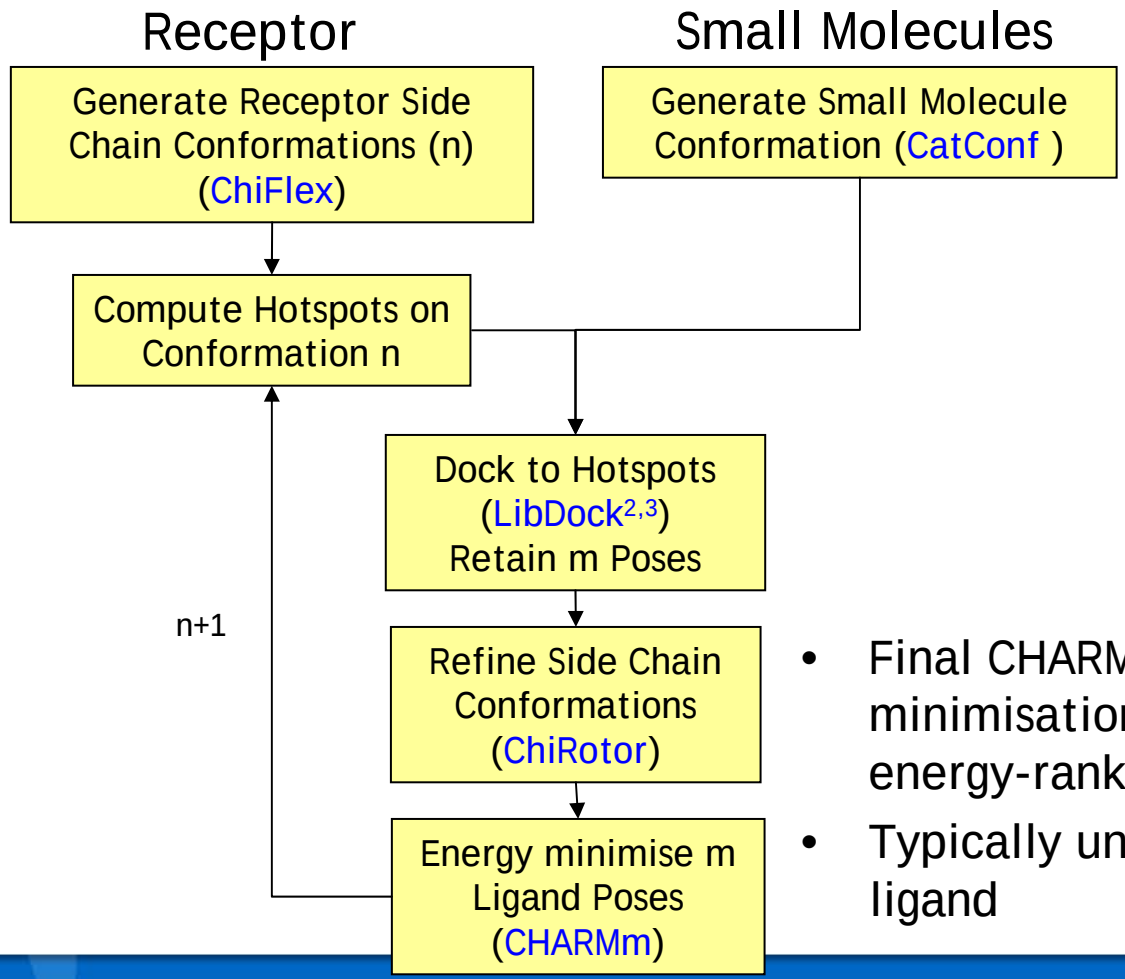
Dock to Hotspots (LibDock[2,3]) Retain m Poses

Refine Side Chain Conformations (ChiRotor)

- CHARMm-based refinement of side chain positions in the presence of docked molecule pose
- 1-3 minutes per pose

Flexible Docking Protocol

Step 6: Final Minimisation of Docked Pose



- Final CHARMm minimisation of pose for energy-ranking
- Typically under 1 min per ligand

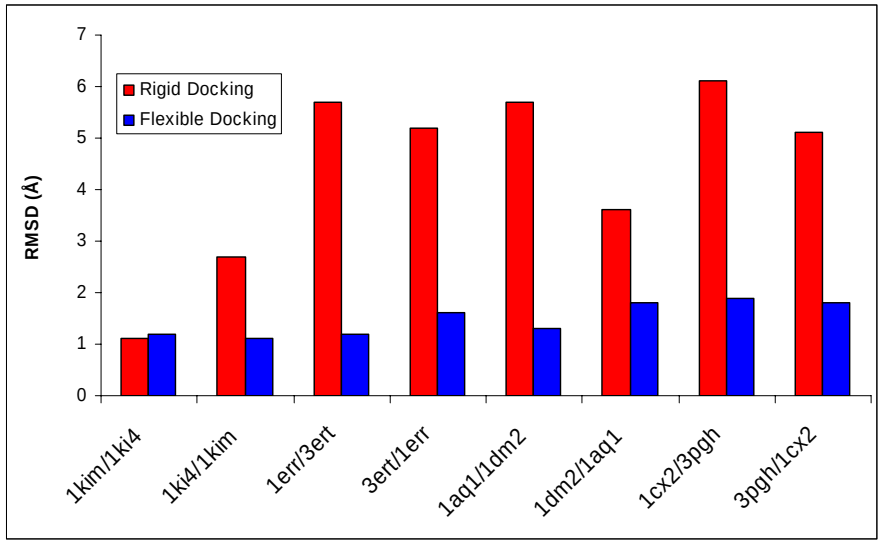
Flexible Docking: Key References

- LibDock
  - D. Diller and K. Mertz, PROTEINS: Structure, Function, and Genetics 43, 113–124 (2001)
  - D. Diller and Li, R. J. Med. Chem. 46, 4638- 4647 (2003)
- ChiRotor/ChiFlex
  - V. Z. Spassov, L. Yan, P. K. Flook, Protein Science 16, 1-13 (2007)
- CHARMm
  - B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. J. Comp. Chem. 4, 187-217 (1983)
  - A. D. MacKerell, Jr., B. Brooks, C. L. Brooks, III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. The Encyclopedia of Computational Chemistry, 1, 271-277 (1998)

Validation of Rational Flexible Docking

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

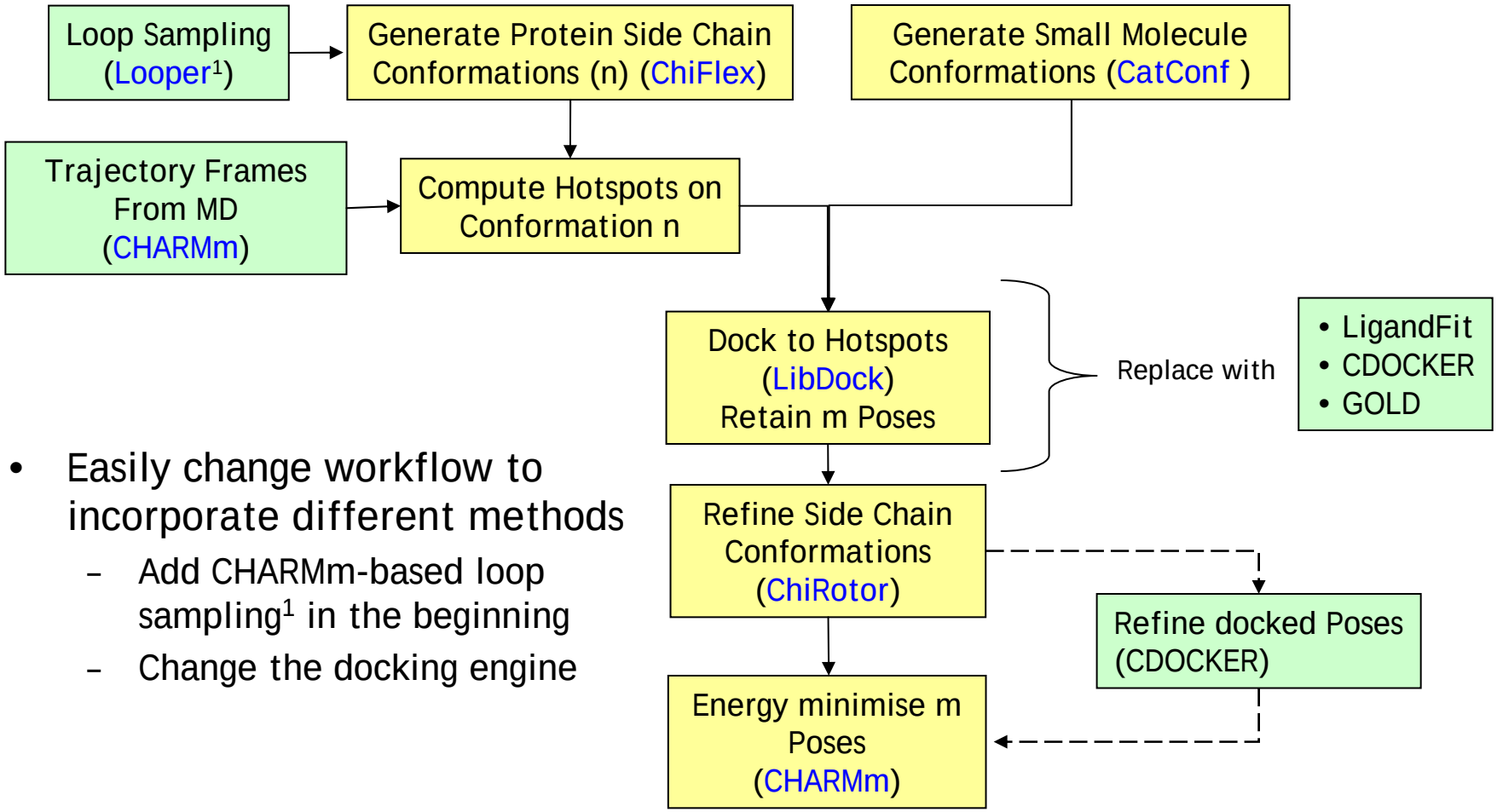
Receptor system (PDB IDs)	# of residues identified as flexible
Thymidine Kinase (1kim, 1ki4)	8
Estrogen Receptor (1err, 3ert)	7
CDK2 (1aq1, 1dm2)	9
COX2 (1cx2, 3pgh)	5



RMSD values compared to x-ray conformation for cross-docking experiments (1kim/1ki4 denotes 1kim ligand into 1ki4 receptor)

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

Flexible Docking: Unprecedented Customisability



- Easily change workflow to incorporate different methods
  - Add CHARMm-based loop sampling[1] in the beginning
  - Change the docking engine

[1] V. Z. Spassov, L. Yan, P. K. Flook, submitted in Protein Science

Accelrys Flexible Docking Methods

- Consistent force field (CHARMm) used throughout
- Minimal user intervention required
- Docking into a realistic environment
  - Docking of ligand influenced by existing side chains in active site
  - Prevents unrealistic poses
  - Accurate initial protein conformations
    - Realistic and based on side-chain-backbone interactions as well as side-chain-side chain interactions
- Library of receptor conformations can be saved and used for all small molecules being docked
- Customisable

Binding Site Analysis

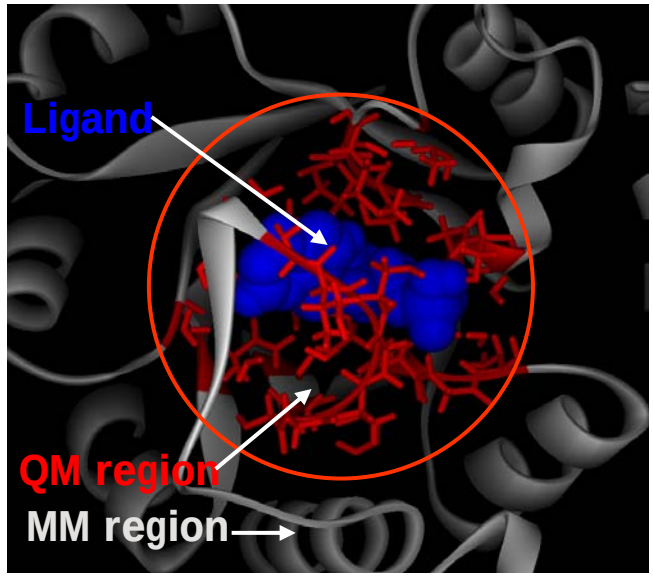
- Scientific Need: Rapid and intuitive analysis of receptor binding sites and receptor-ligand interactions

- Challenges
  - Individual tools exist in numerous programs
  - Numerous programs lead to high costs, switching times

- Solution: All the analysis tools in a single environment
  - Cluster and compare binding modes: quickly analyse thousands of poses
  - Analyse a set of x-ray structures/docked poses to identify similarity and diversity of ligands: gain insights into lead optimisation
  - Filter poses based on binding modes/interaction patterns: quickly pick relevant poses

QM/MM in Discovery Studio (In Planning)

- Increasing need for accurate minimization of protein-ligand complexes and accurate prediction of interaction energy
- Challenges
  - MM Force fields are limited in accuracy
  - QM methods are computationally expensive
  - QM/MM is an effective compromise
  - Treat the binding site region with QM
- Accelrys Solution:
  - QMera for QM/MM applications
    - QM portion = DMol3 (DFT)
    - MM portion = CHARMm
- Available Method
  - Optimise receptor-ligand geometries
  - Obtain accurate binding energy predictions for ligands
  - Optimise ligand charges in the field of the receptor

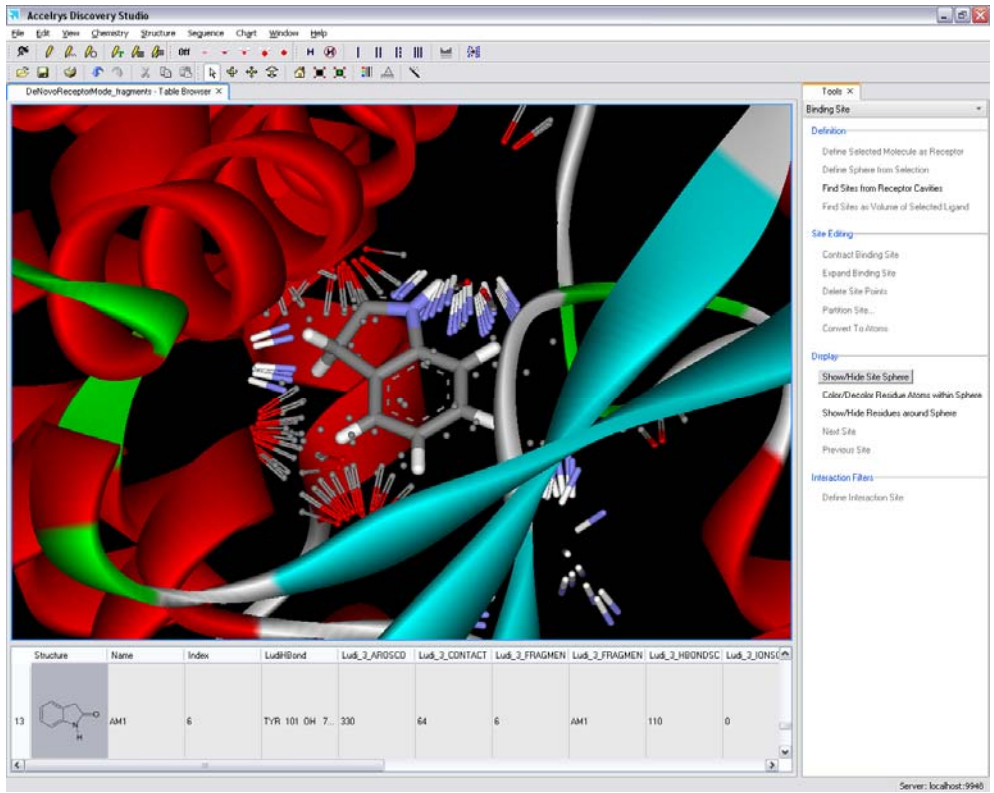


Combine CHARMm's robust forcefield with the accuracy of QM

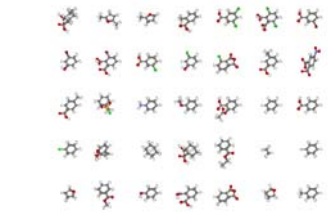
Structure-Based Design – *De Novo* Design

- Protocols
  - De Novo Receptor
  - De Novo Link
  - De Novo Evolution
  - De Novo Library Generation
- Tools
  - Binding Site
- Results analysis
  - Table Browser

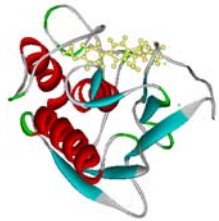
10 scoring functions now available for all protocols



De Novo Receptor/Link (Ludi)

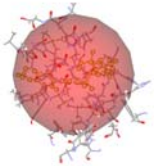


Protein
(& core ligand)



Ludi Run
Parameters

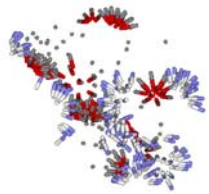
Define Active Site
(centre & radius)



Fragment library
(atom coordinates &
atom target)

Fit fragments to
interaction sites

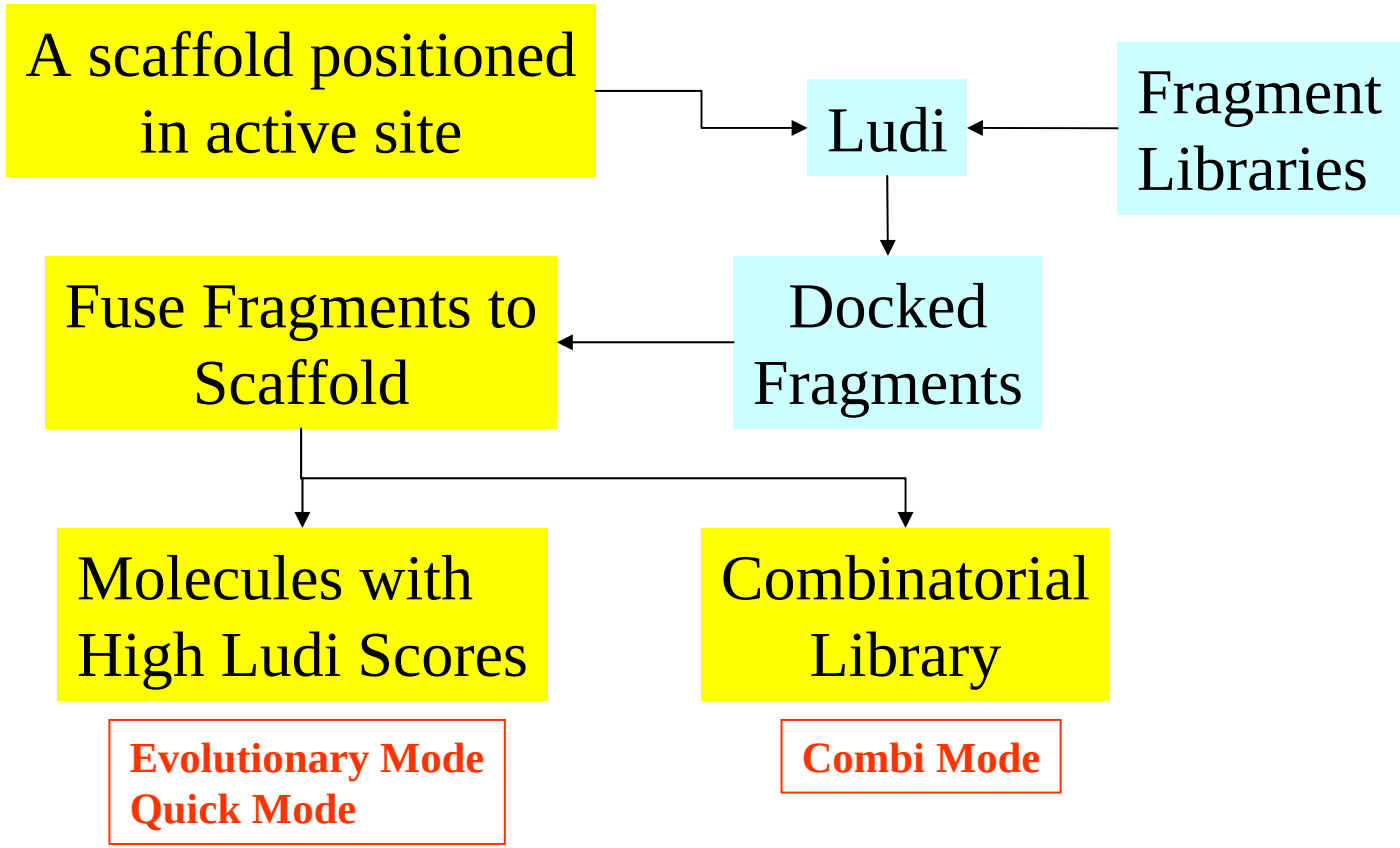
Generate
Interaction Site Map



Bump-checking,
Electrostatic interaction,
Buried polar groups

Score Ligands

AutoLudi workflow



Score and Analyze Ligand Poses

- Scoring functions
  - LigScore 1 and 2
    - LIGSCORE1 = vdW + C+_pol - $Totpol^2$
    - LIGSCORE2 = vdW + C+_pol - $BuryPol^2$
    - Dreiding or CFF
  - Piecewise Linear Potential (PLP1, PLP2)
  - Potential Mean Force (PMF)
  - Ludi
  - Jain
  - Poses can be filtered out according to score values
  - Consensus scoring

PMF04 added

Flexible protein handling

- Protocols that generate “flexible” protein data
  - Scoring traditionally scores multiple ligands with single, rigid receptor
  - Now scoring uses flexible atoms information in SD file

Score and Analyze Ligand Poses

- RMSD
- Hydrogen bonds
- Contact

New

New Analysis Methods

- Receiver Operating Characteristic (ROC)curves
  - Plot False Positives (x) Vs. True Positives (Y)
  - Much like a HitRate plot, but gives a system independent quality metric
  - Area Under Curve (AUC)
    - 1 perfect
    - > 0.9 Excelent
    - > 0.8 Good
    - > 0.7 Fair
    - > 0.6 Poor
    - <=0.6 Fail

Calculate Binding Energies

- Post-analysis of docked poses for ranking
- Calculation of the binding free energy of protein-ligand complexes Using continuum solvents
  - Can provide the electrostatic component of the binding free energy
  - Addition of an estimate of the non-polar contribution and the entropic components gives an estimate of total binding energy
    - Entropic component from normal mode analysis or quasiharmonic analysis
  - MM-PB(GB)SA

$$\Delta G_{bind} = G_{complex} - G_{ligand} - G_{protein}$$

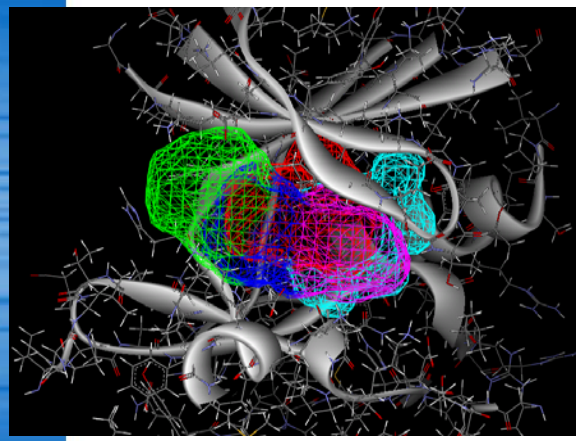
$$G = \langle G_{intra} \rangle + \langle G_{inte} \rangle + \langle G_{pol} \rangle + \langle G_{np} \rangle - T\Delta S$$

G_{intra} Intramolecular energy
G_{inter} Intermolecular energy
G_{pol} Polar contribution to solvation free energy
G_{np} Nonpolar contribution to solvation free energy

New translational and rotational entropy terms included for vacuum and distance-dependent implicit solvent models

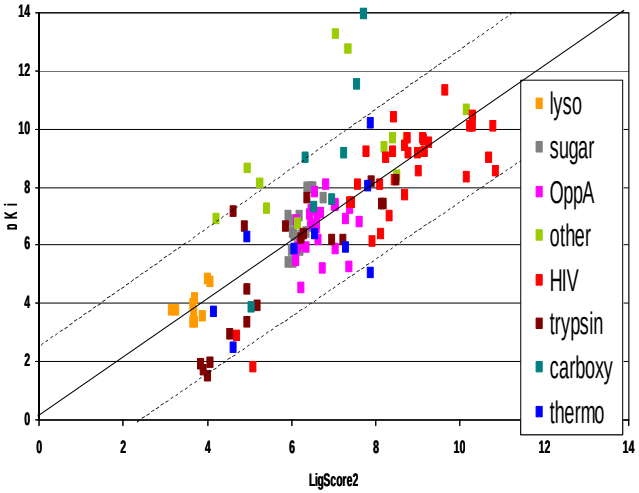
Applications of Structure-Based Design

Docking
screen ligands, find hits



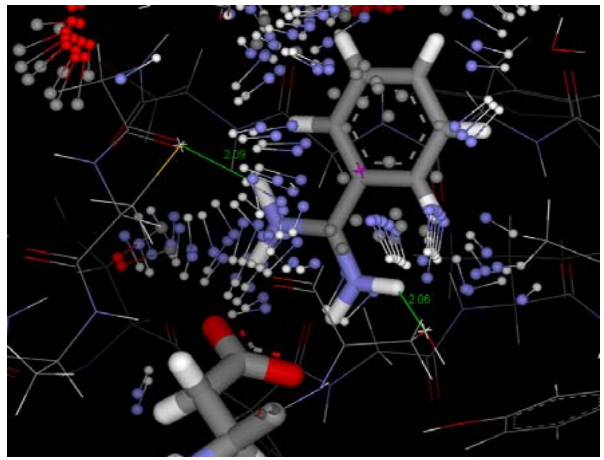
- Identify binding sites
- Explore binding modes for new ligands
- Dock ligand libraries

Scoring
understand potency, predict affinity

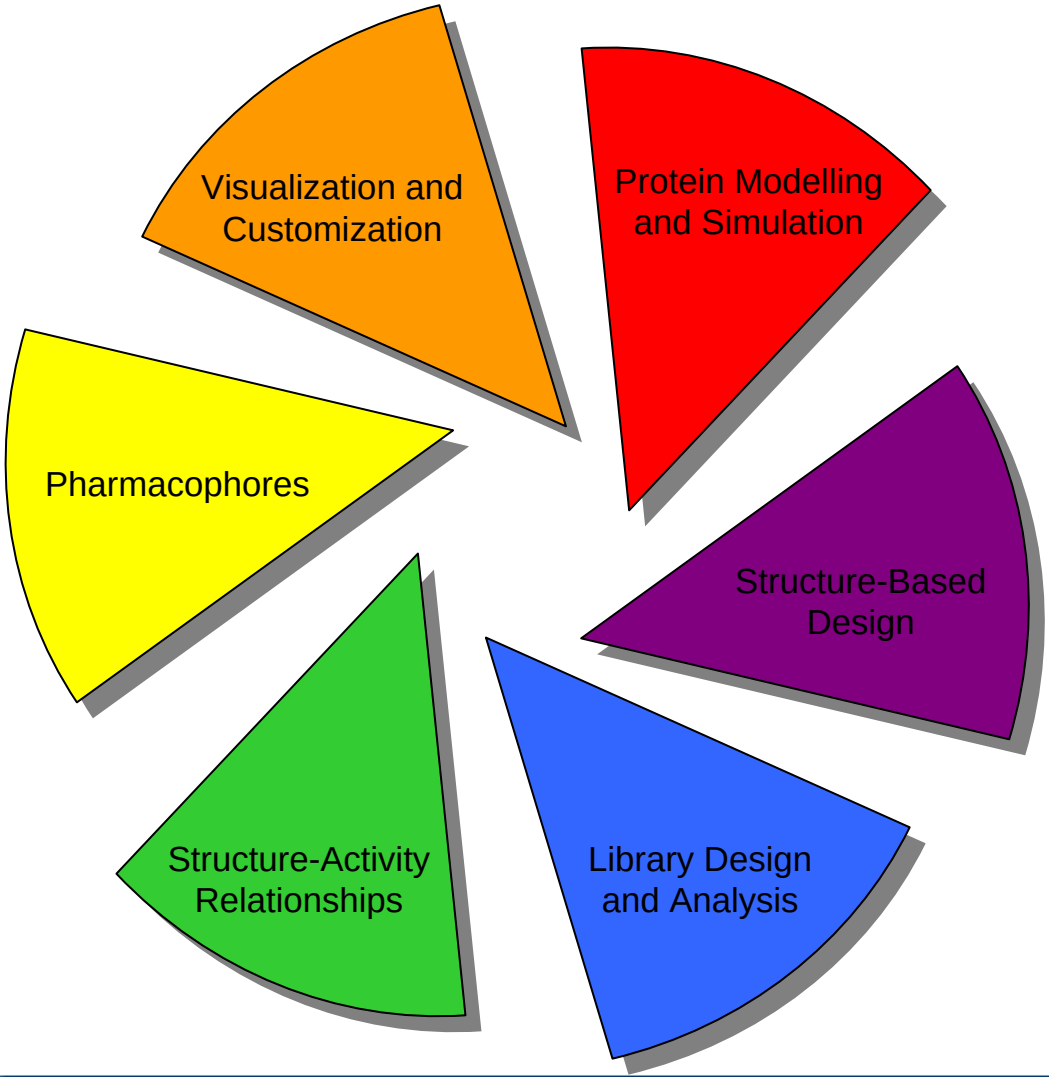


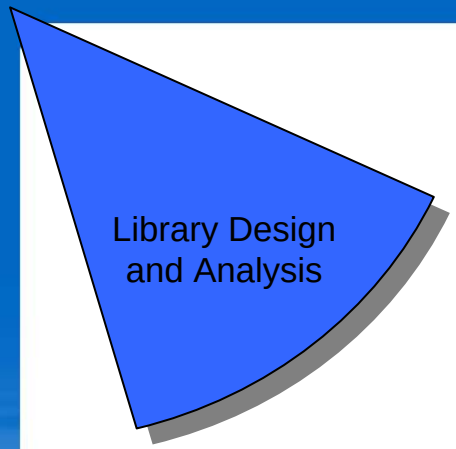
- Predict affinities of ligand-protein complexes
- Rank hits from searches

De Novo Design
design new ligands, optimize leads



- Define interaction sites
- Explore new analogs
- Design new scaffolds


Visualization and Customization
Protein Modelling and Simulation
Pharmacophores
Structure-Based Design
Structure-Activity Relationships
Library Design and Analysis



New

- Library Enumeration protocols
  - Enumerate Library By R-Group
  - Enumerate Library By Reaction
- Library Analysis protocols
  - Calculate Diversity Metric
  - Calculate Principal Component
  - Calculate Property Profile Penalty
  - Calculate Property Range Penalty
  - Create Property Profile

- Library Selection protocols
  - Cluster molecules
  - Find Diverse Molecules
  - Find Outlier Molecules
  - Find Similar Molecules
    - By Fingerprints
    - By Numeric Properties

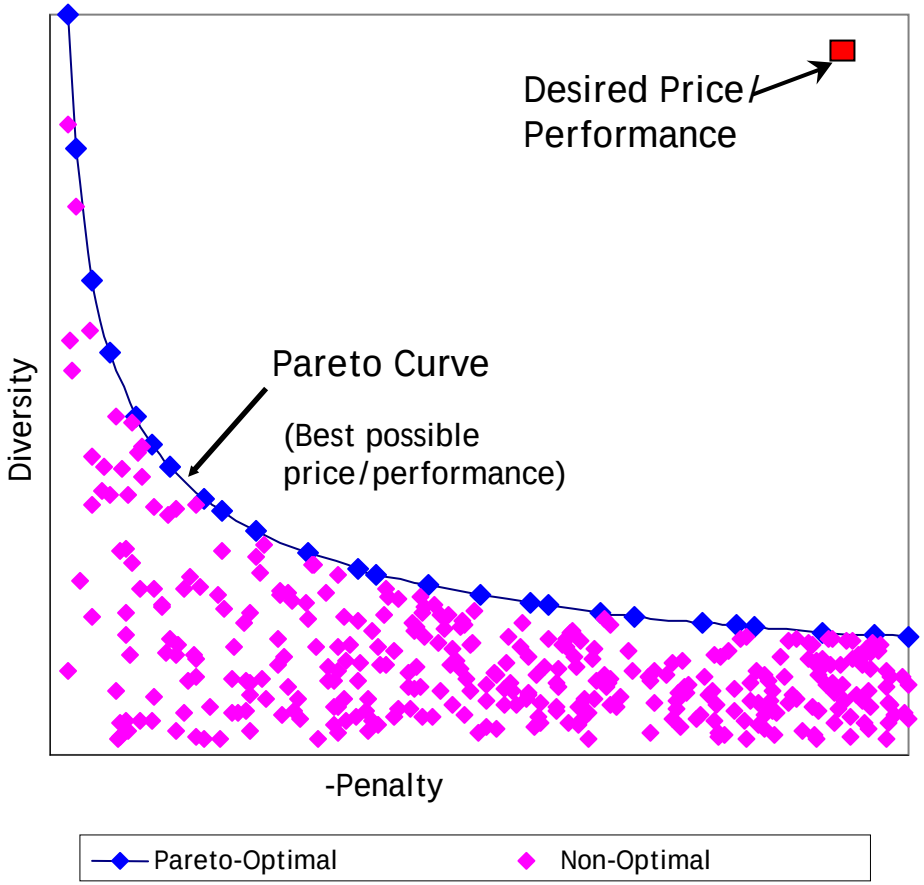
- Library Optimisation protocols
  - Optimize Combinatorial Library with Pareto Method
  - Optimize Subset Library with Pareto Method
  - Sort Data with Pareto Method
- Compare Libraries protocol

Pareto Optimiser

- Multi-objective optimisation
- Yields a family of solutions on the Pareto front
- Three protocols that use Pareto
  - Sort Data with Pareto Method
    - Simplest form of Pareto optimization
    - Done when you have all the information and can't create more
  - Optimize Combinatorial Library with Pareto Method
  - Optimize Subset Library with Pareto Method

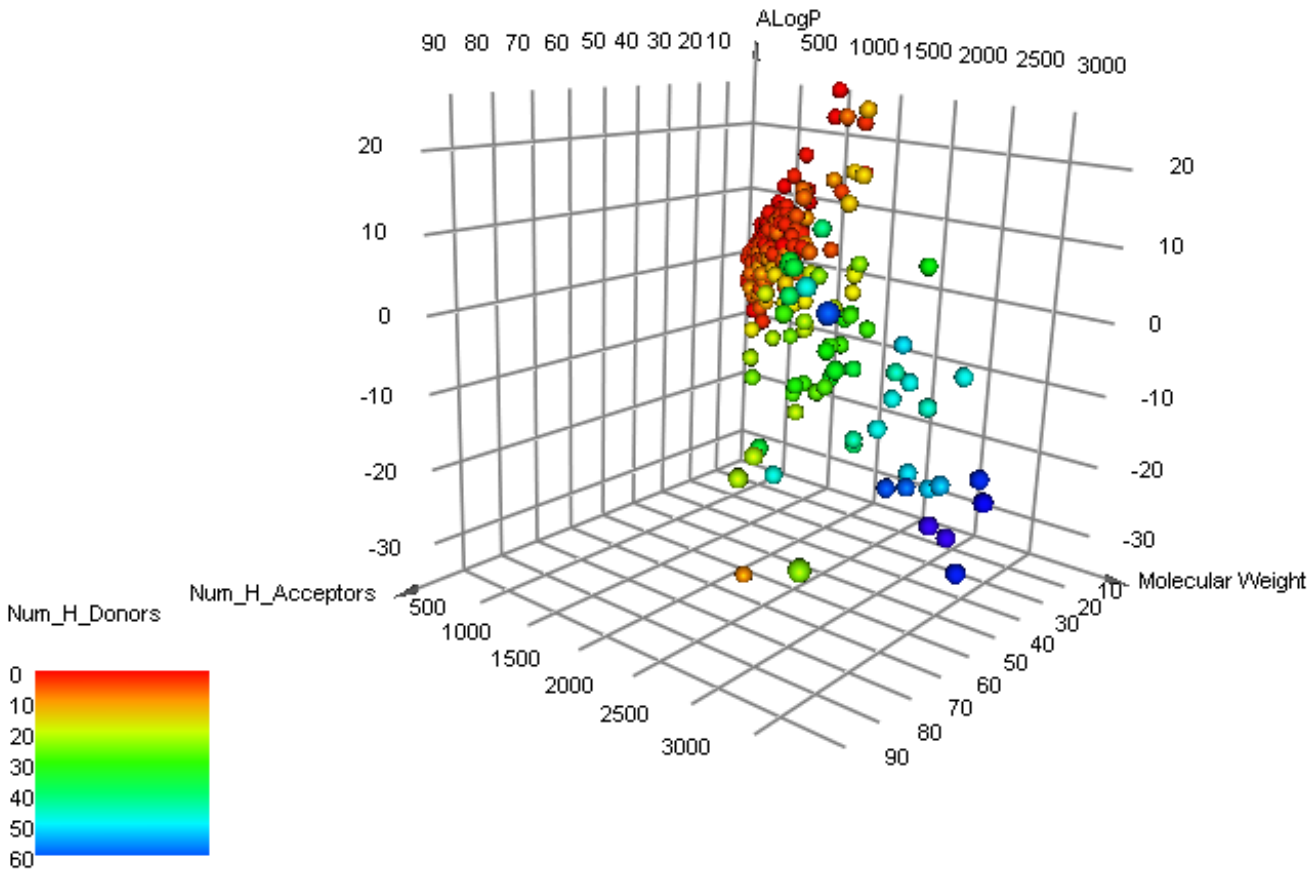
Trade-Off Optimisation

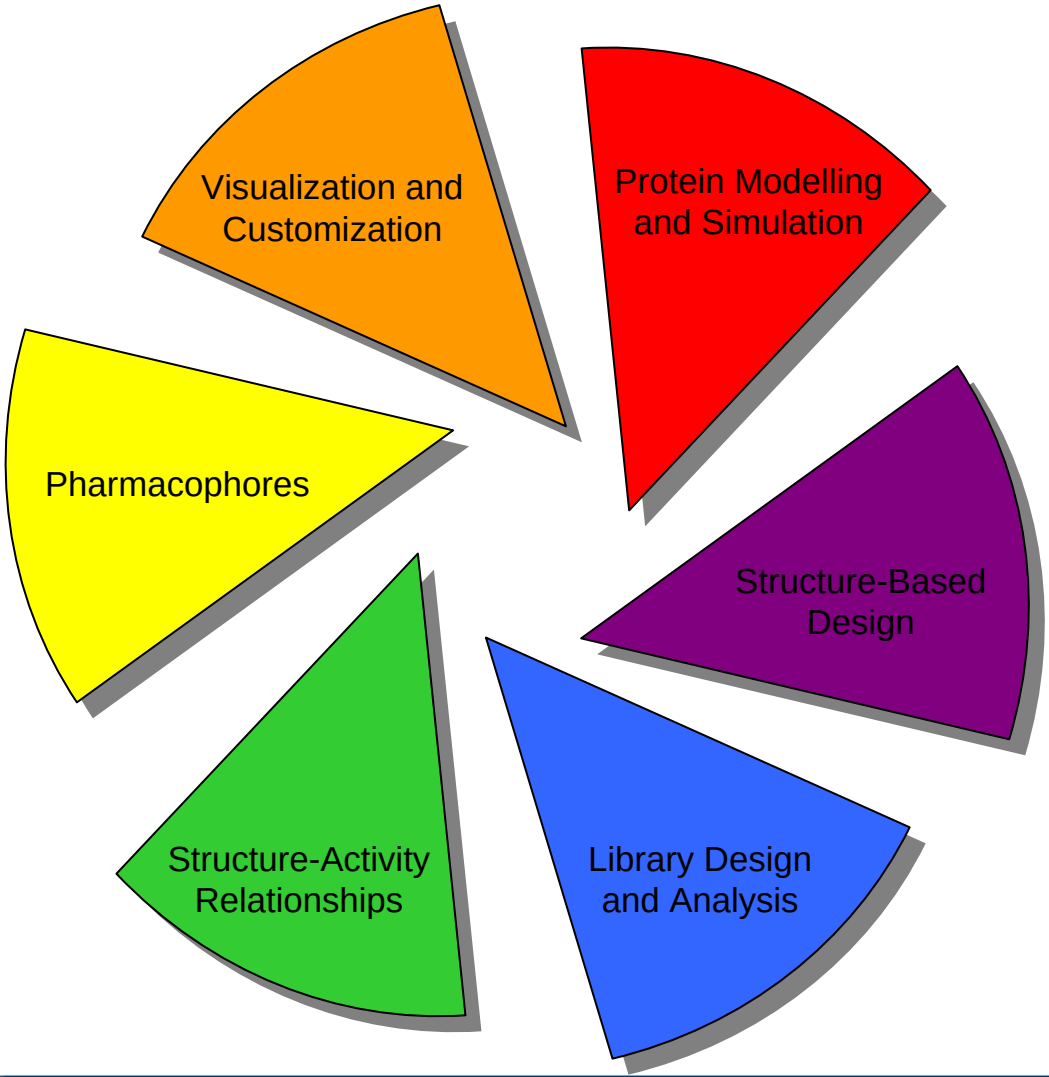
- Goal: Find system that give best trade-off of objectives
- Result: A set of systems on the Pareto curve or surface

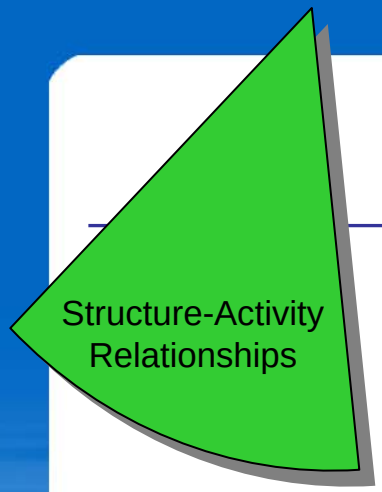


Libraries on Pareto (Tradeoff) Curve have the best possible combinations of properties

Library Design and Analysis Visualisation



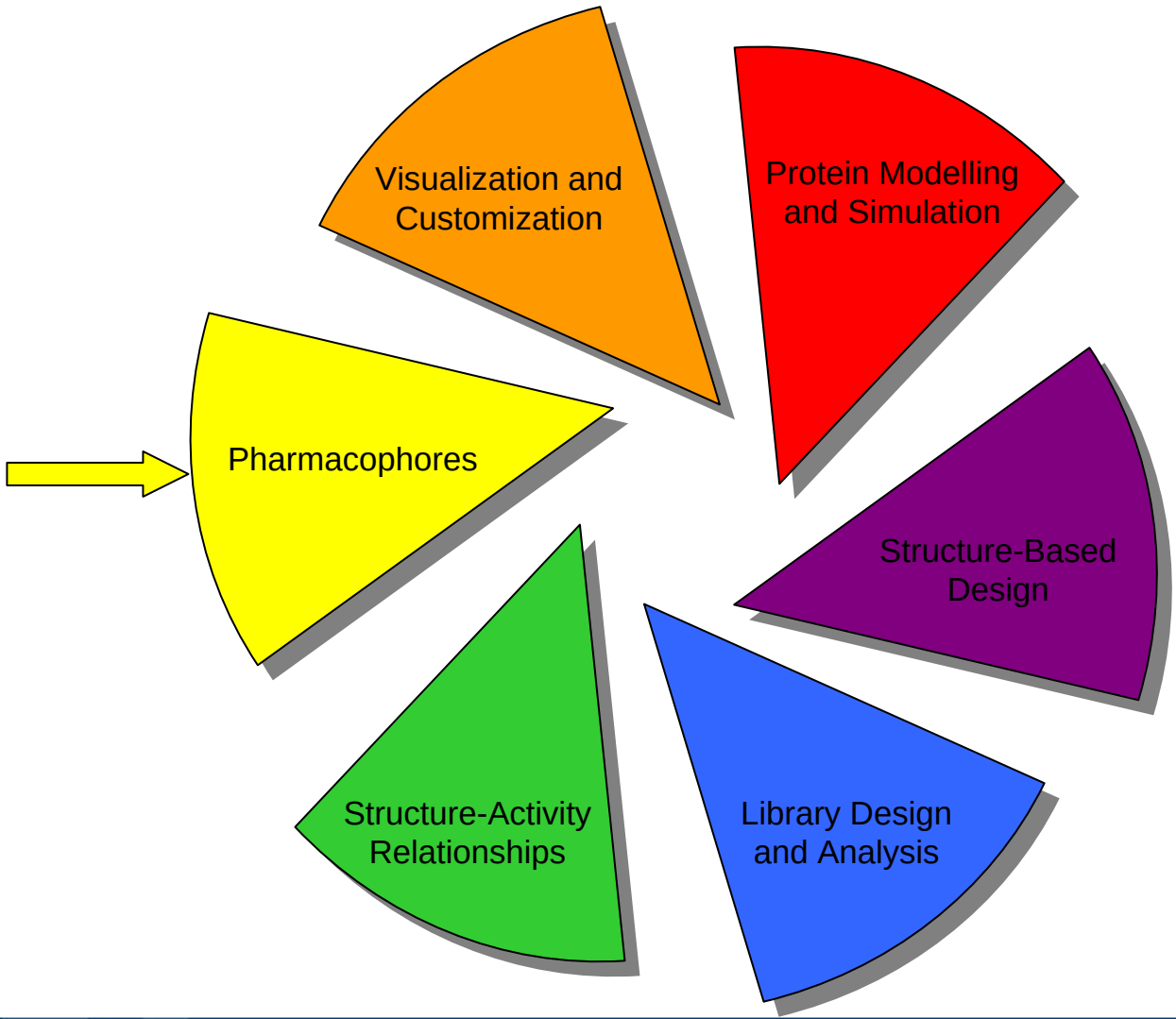

Visualization and Customization
Protein Modelling and Simulation
Pharmacophores
Structure-Based Design
Structure-Activity Relationships
Library Design and Analysis



New

- Property calculation protocols
  - Calculate Properties
  - Calculate all calculable properties in Pipeline Pilot
  - Calculate semiempirical QM descriptors (VAMP)
  - Calculate density functional QM descriptors (DMOL3)
- Structure-activity model building protocols
  - Create Bayesian Model
  - Create Multiple Linear Regression Model
  - Create Partial Least Squares Model
  - Create Back Propagation Neural Network (BPNN) models
  - Create Genetic Function Approximation (GFA) model

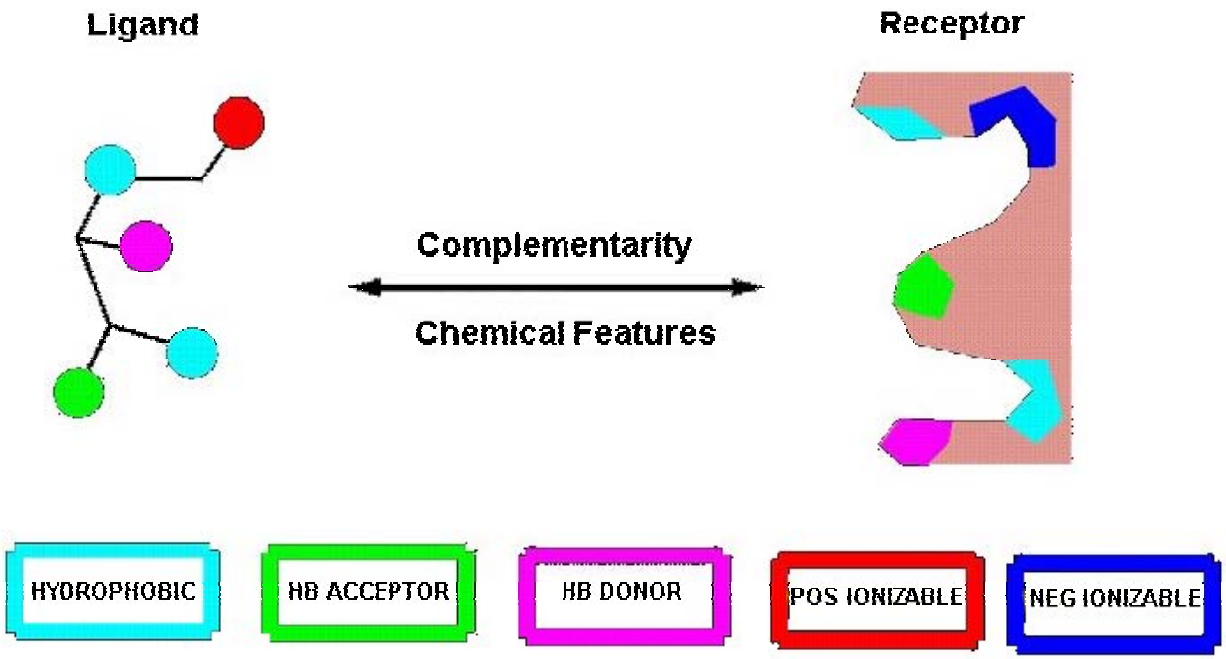
Includes multi-objective Pareto optimiser


Visualization and Customization
Protein Modelling and Simulation
Pharmacophores
Structure-Based Design
Structure-Activity Relationships
Library Design and Analysis

Central Assumption

"The central assumption of structure-based design is that good inhibitors must possess significant structural and chemical complementarity to their target receptor."

From Kuntz, *Science* (1992) 257, 1078-1082



A More Modern Definition…

For Ligand---

"The molecular framework that carries (*phoros*) the essential features responsible for a drug's (*pharmacon*)biological activity."

From P. Ehrlich, *Dtsch. Chem. Ges.* **1909** *42*, 17

Thus, the pharmacophore…

"A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response."

Wermuth *et al*. *Pure Appl. Chem.* **1998** *70*:1129-43

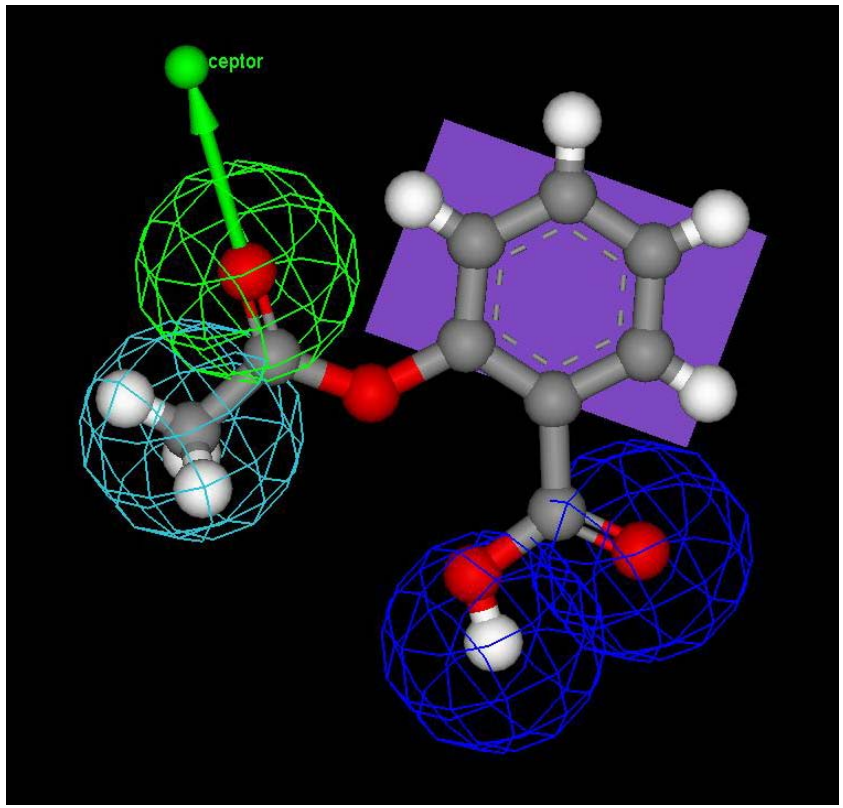
Pharmacophore Modeling in Discovery Studio

New

- Discovery Studio provides protocols and tools that allow you to:
  - Generate pharmacophores from ligands
  - Diverse conformation Generation (FAST, BEST, CAESAR)
  - Generate pharmacophores from proteins (SBP)
  - Score ligands against a pharmacophore
  - Search databases using pharmacophores
  - Create conformer models
  - Create databases
  - Fragment Based Pharmacophore
  - Ligand Profiler
  - Customised Features

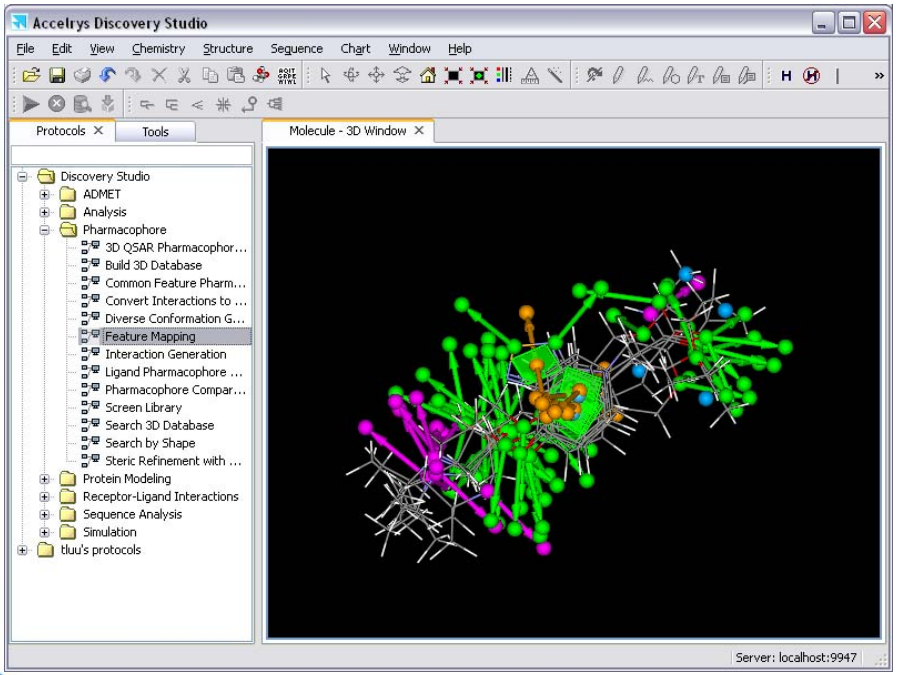
Pharmacophore Models

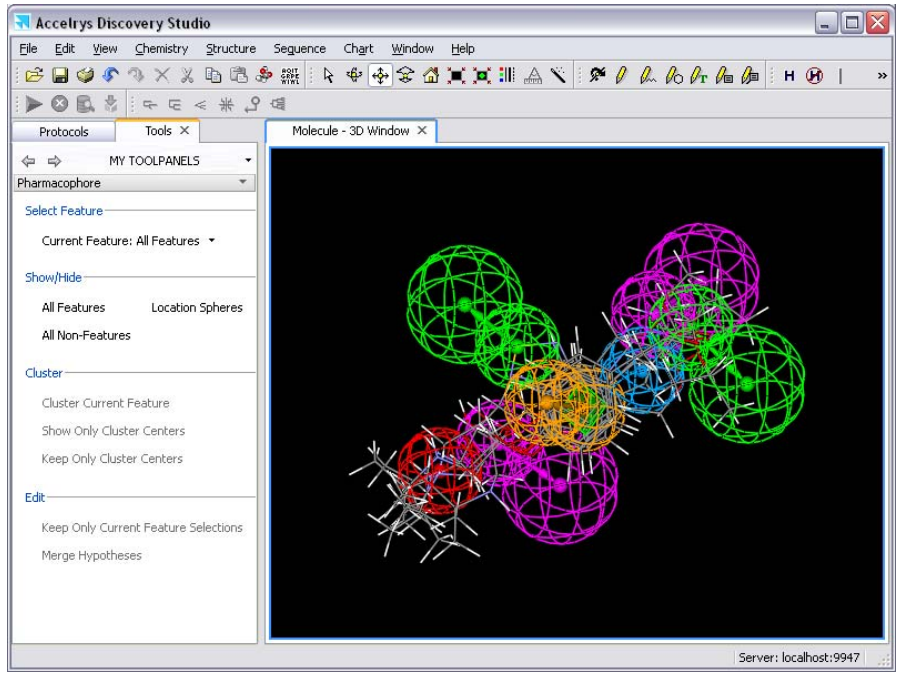
- 1D
  - Physical/Biological properties
- 2D
  - Substructures
- 3D
  - Chemical features
    - Hydrophobic or charged groups
    - H-bond donors/acceptors
- Shape-Based



Alignment Based Pharmacophores

- Create a pharmacophore from
  - X-ray crystal stucture
  - Set of pre-aligned structures
- Ability to inspect and bias feature type selection and location based on known ligand and protein information

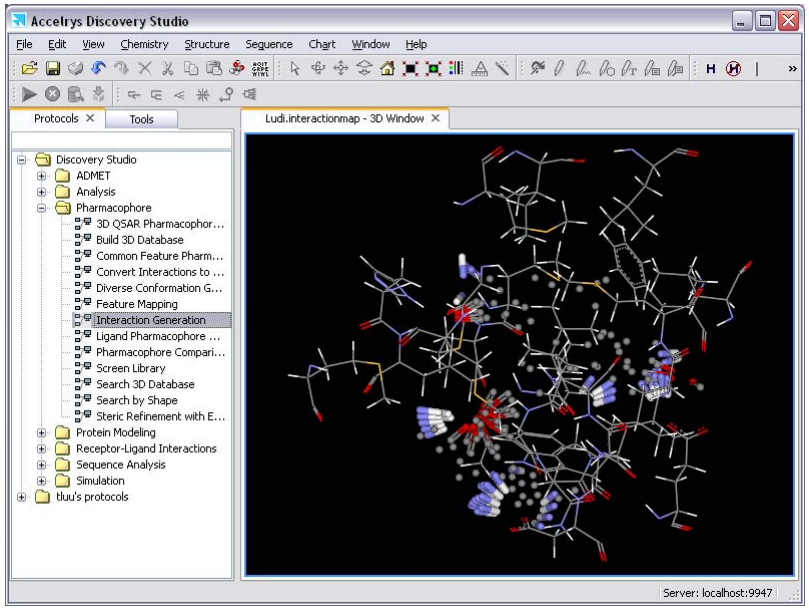


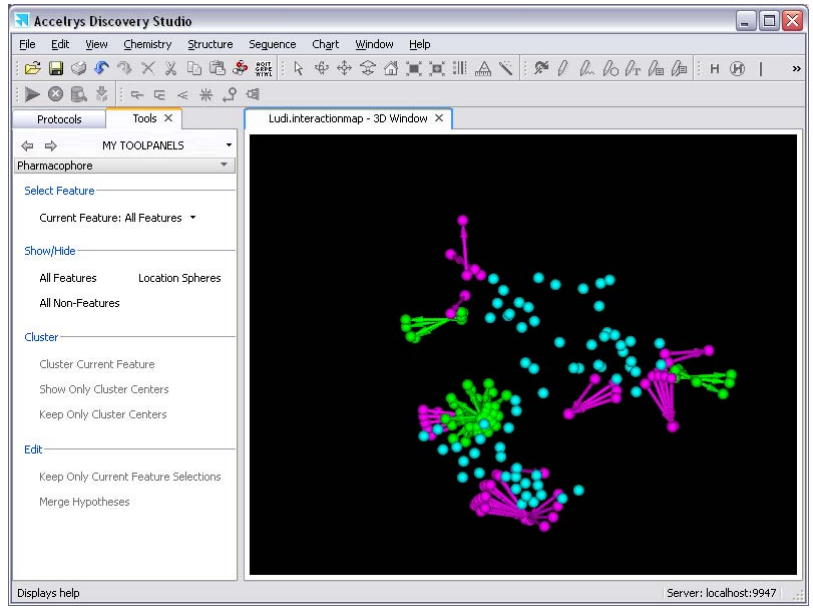


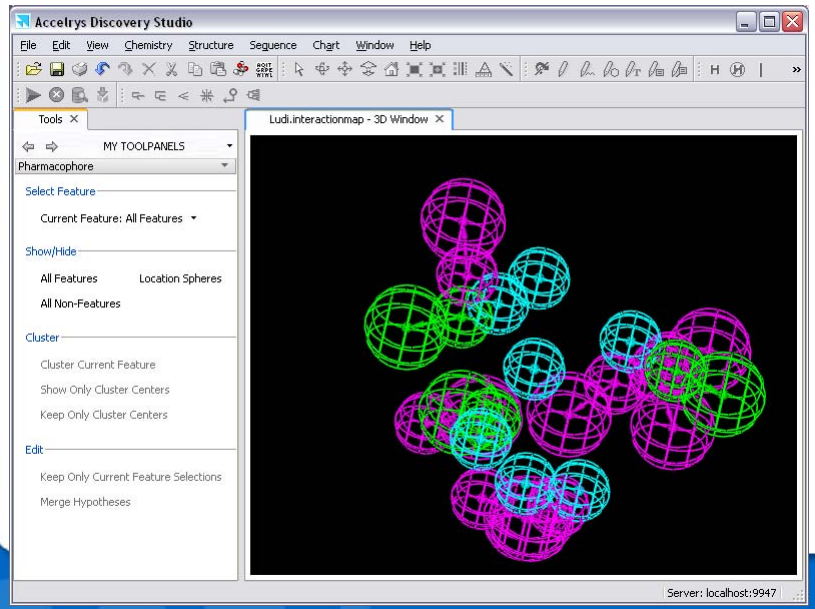
Structure Based Pharmacophores

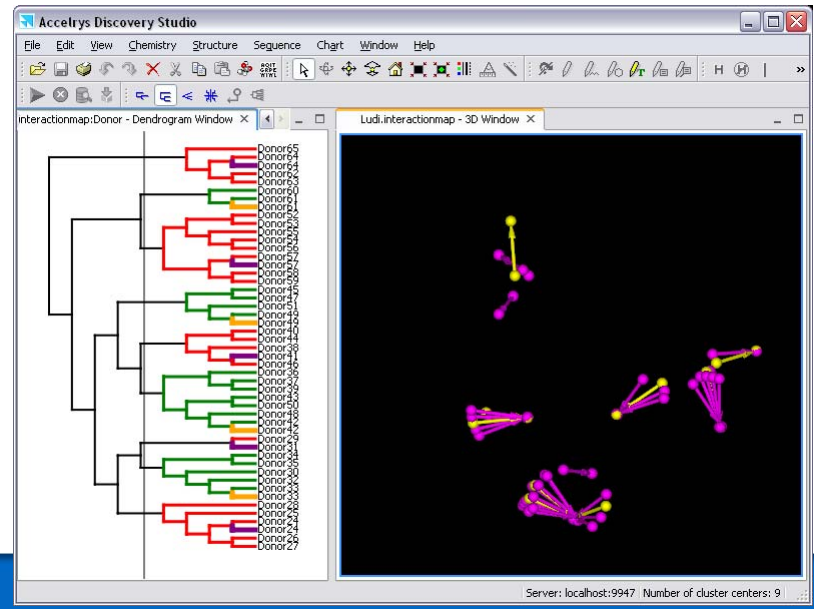
- Utilise known or suspected protein active site to select compounds most likely to bind within the active site
  - Complementary to Docking/Modelling
  - Maximises use of binding information

1. Generate interaction map (donor, acceptor, hydrophobes)
2. Cluster and select pharmacophore features
3. Screen compound databases (SD)
4. Score, rank, and analyse hits
5. Identify significant sets of 3D queries

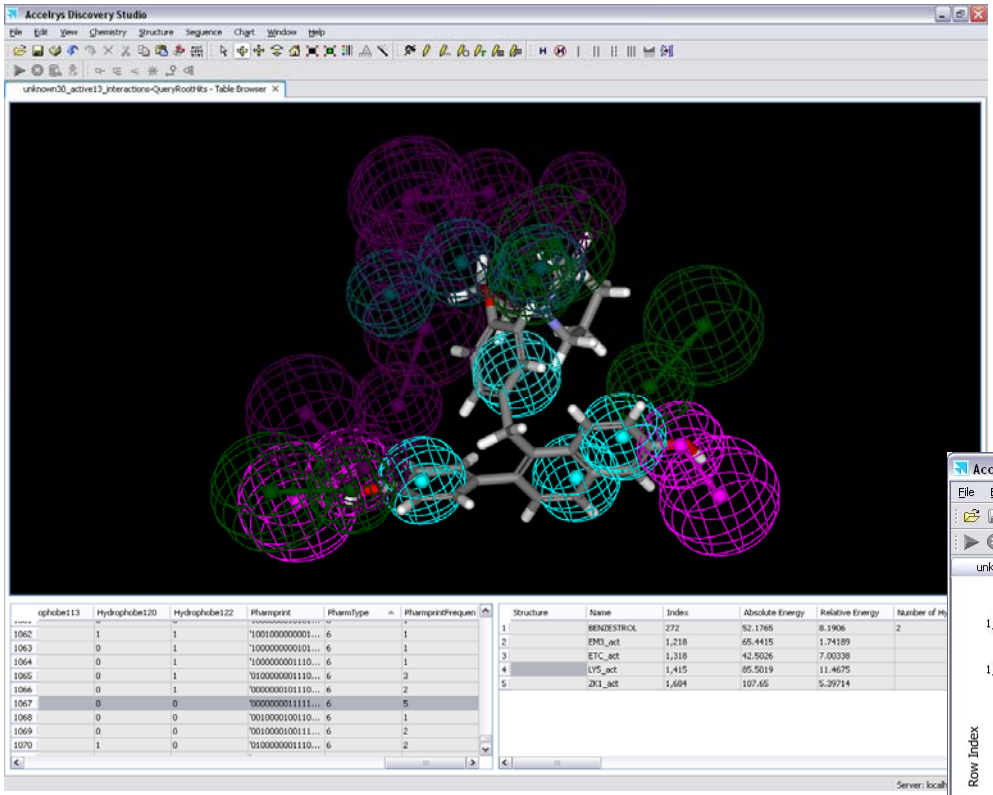

Accelrys Discovery Studio
File Edit View Chemistry Structure Sequence Chart Window Help
Protocols Tools
Ludi.interactionmap - 3D Window
Discovery Studio
ADMET
Analysis
Pharmacophore
3D QSAR Pharmacophor...
Build 3D Database
Common Feature Pharm...
Convert Interactions to ...
Diverse Conformation G...
Feature Mapping
Interaction Generation
Ligand Pharmacophore ...
Pharmacophore Compari...
Screen Library
Search 3D Database
Search by Shape
Steric Refinement with E...
Protein Modeling
Receptor-Ligand Interactions
Sequence Analysis
Simulation
tluu's protocols
Server: localhost:9947


Accelrys Discovery Studio
File Edit View Chemistry Structure Sequence Chart Window Help
Protocols Tools
Ludi.interactionmap - 3D Window
MY TOOLPANELS
Pharmacophore
Select Feature
Current Feature: All Features
Show/Hide
All Features
Location Spheres
All Non-Features
Cluster
Cluster Current Feature
Show Only Cluster Centers
Keep Only Cluster Centers
Edit
Keep Only Current Feature Selections
Merge Hypotheses
Displays help
Server: localhost:9947

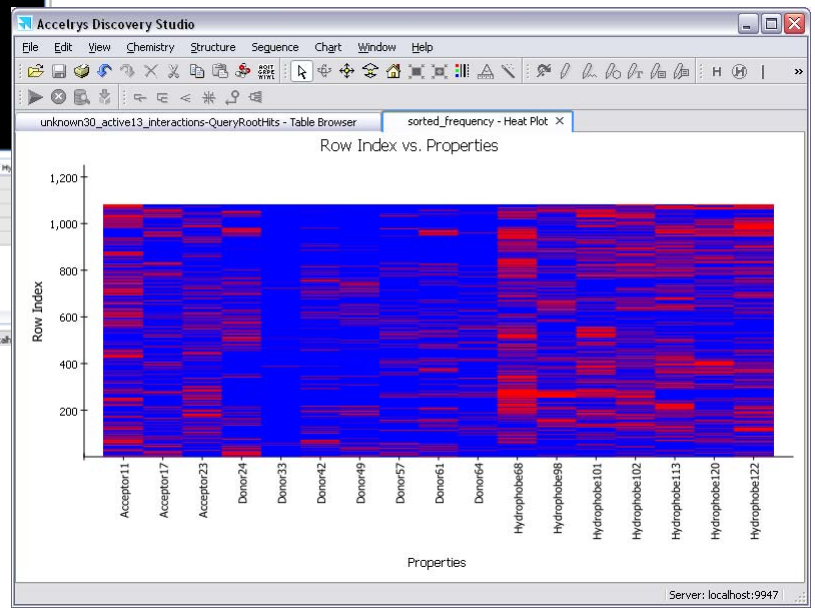

Accelrys Discovery Studio
File Edit View Chemistry Structure Sequence Chart Window Help
Tools
Ludi.interactionmap - 3D Window
MY TOOLPANELS
Pharmacophore
Select Feature
Current Feature: All Features
Show/Hide
All Features
Location Spheres
All Non-Features
Cluster
Cluster Current Feature
Show Only Cluster Centers
Keep Only Cluster Centers
Edit
Keep Only Current Feature Selections
Merge Hypotheses
Server: localhost:9947


Accelrys Discovery Studio
File Edit View Chemistry Structure Sequence Chart Window Help
interactionmap:Donor - Dendrogram Window
Ludi.interactionmap - 3D Window
Server: localhost:9947 Number of cluster centers: 9

Structure Based Pharmacophores



- Results
  - Fit Value
  - Pharmprint
  - PharmprintFrequency
  - PharmType
- Heat map of mapped features



- Dim features that are not mapped
- Table browser
  - Group By…
  - Represent By…

Ligand Profiling

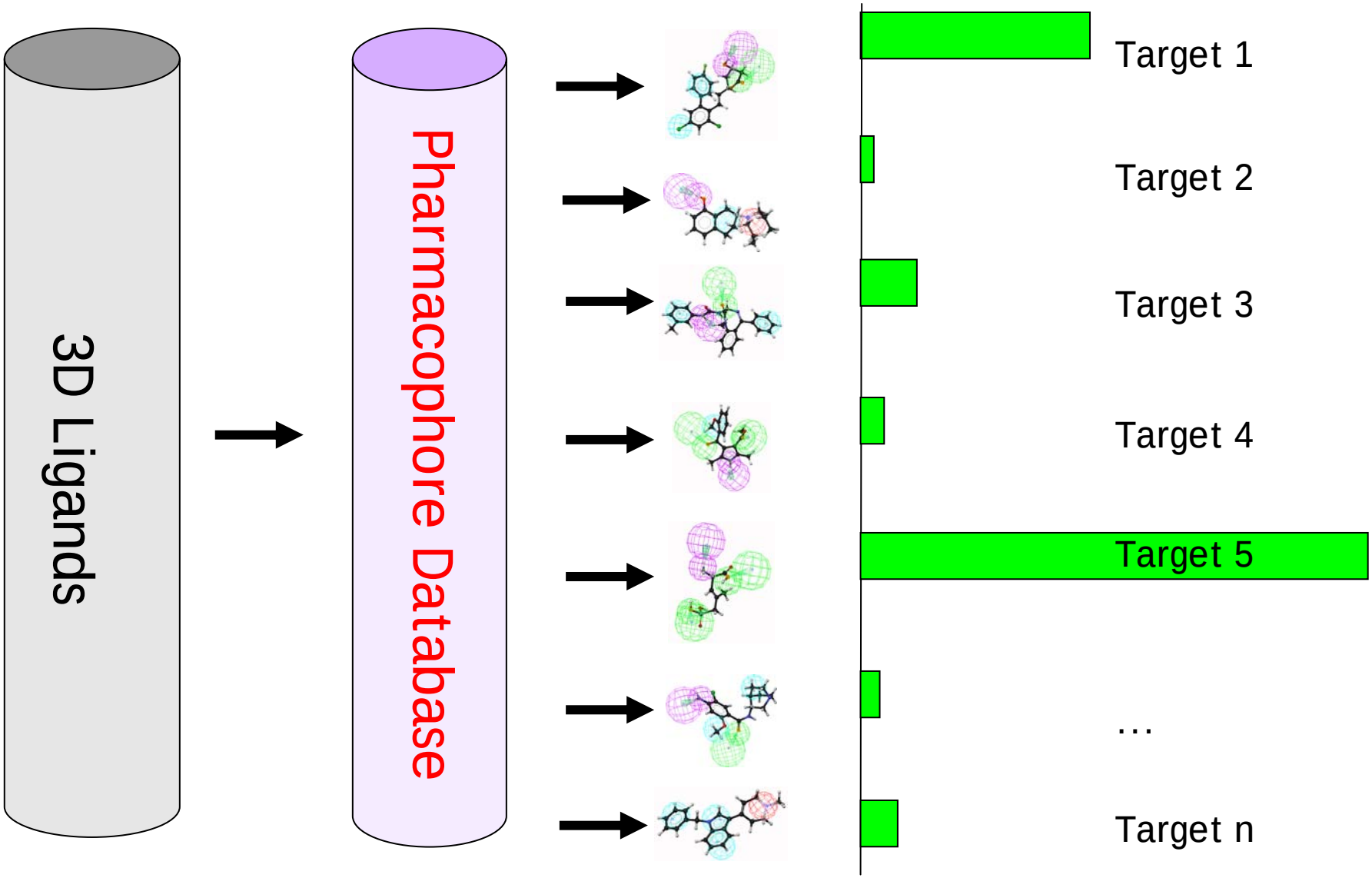
- Rapidly screen millions of ligands against thousands of proteins to identify additional drug targets or side-effects

  100,000 ligands each tested against

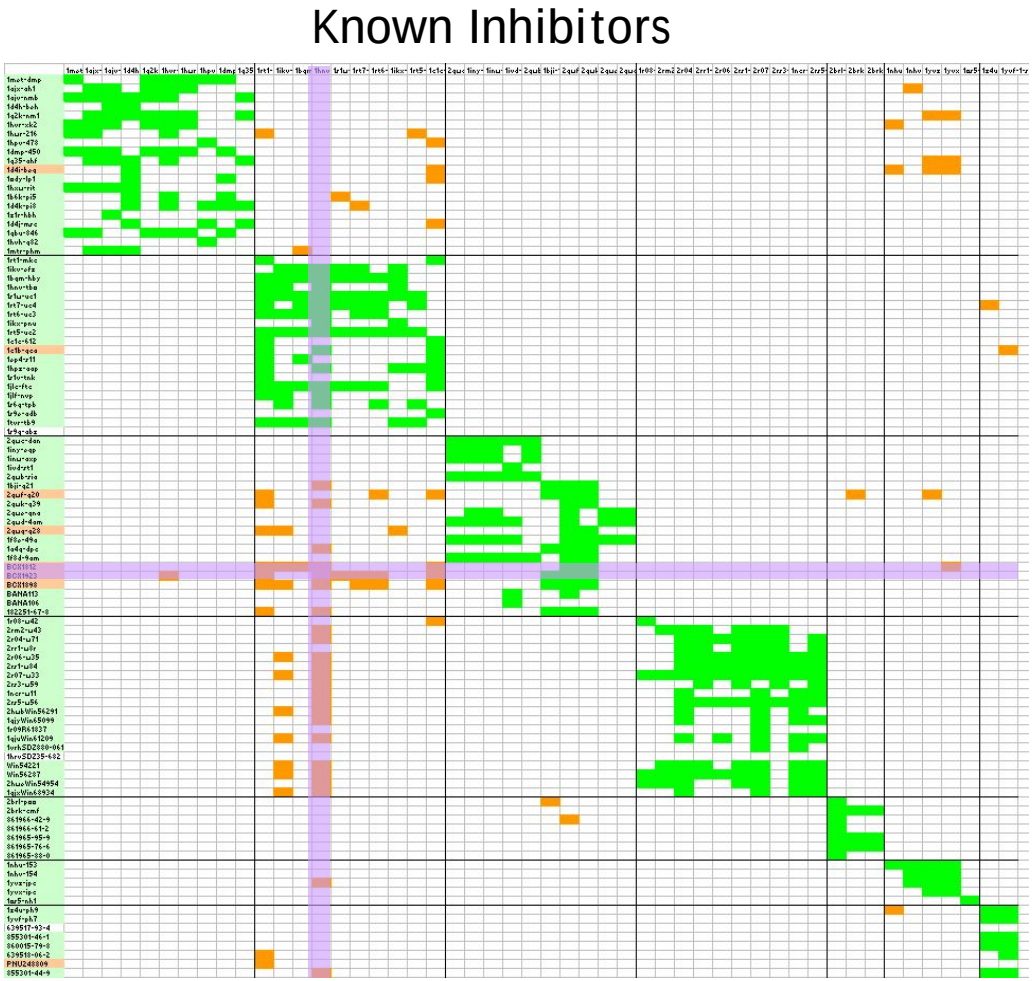
  2,000 different protein binding sites

  100,000 × 2,000 = 200,000,000

- Ligand Profiler: Data analysis and database creation
  - Powerful, easy to use data analysis tools
  - Build custom profiling databases using in-house data
- Pharmacophore Profile Database
  - Approximately 2,000 structure-based pharmacophores (50% with shape descriptors) covering ~200 targets
  - Highly relevant therapeutic areas
  - Compounds selected from the PDB and literature


3D Ligands
Pharmacophore Database
Target 1
Target 2
Target 3
Target 4
Target 5
…
Target n

Easy to Interpret Cross-Reactivity Heat Map



- Positive Control
  - Pharmacophore models can accurately predict their corresponding inhibitors
- Negative Control
  - Pharmacophore models can discriminate matching ligands from others
- Verticals highlight promiscuous inhibitors
- Horizontals show poor models

Enumerate from Fragments

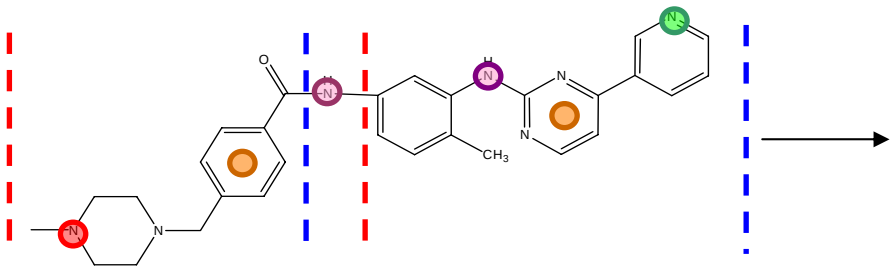
- *De Novo* Pharmacophore fragment-based design and optimisation
  - Lead optimisation for existing known compound
  - Explore higher molecular diversity space; build focussed libraries

1. Divide molecule into fragments
2. Build feature/shape pharmacophore queries from each fragment
   - Shared feature preserved on both sides of a detachment point
3. Search fragment database
4. Join fragments to generate *de novo* library

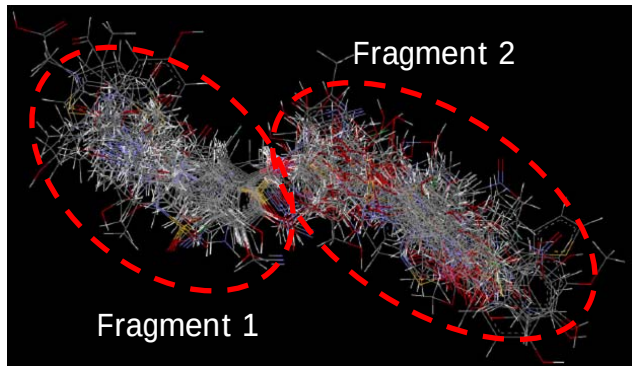
De Novo **Pharmacophore Lead Optimiser**

- Fast and Efficient Method for Creating Novel Drug Candidates
  - Automated rapid linear screening and combination of fragments followed by optimisation of new compounds

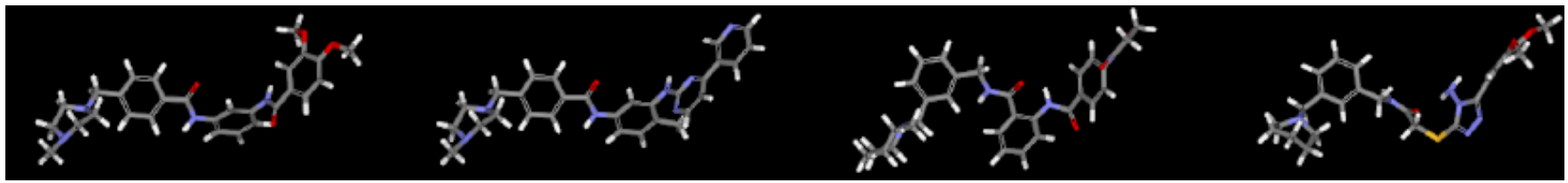
1. Fragment the pharmacophore



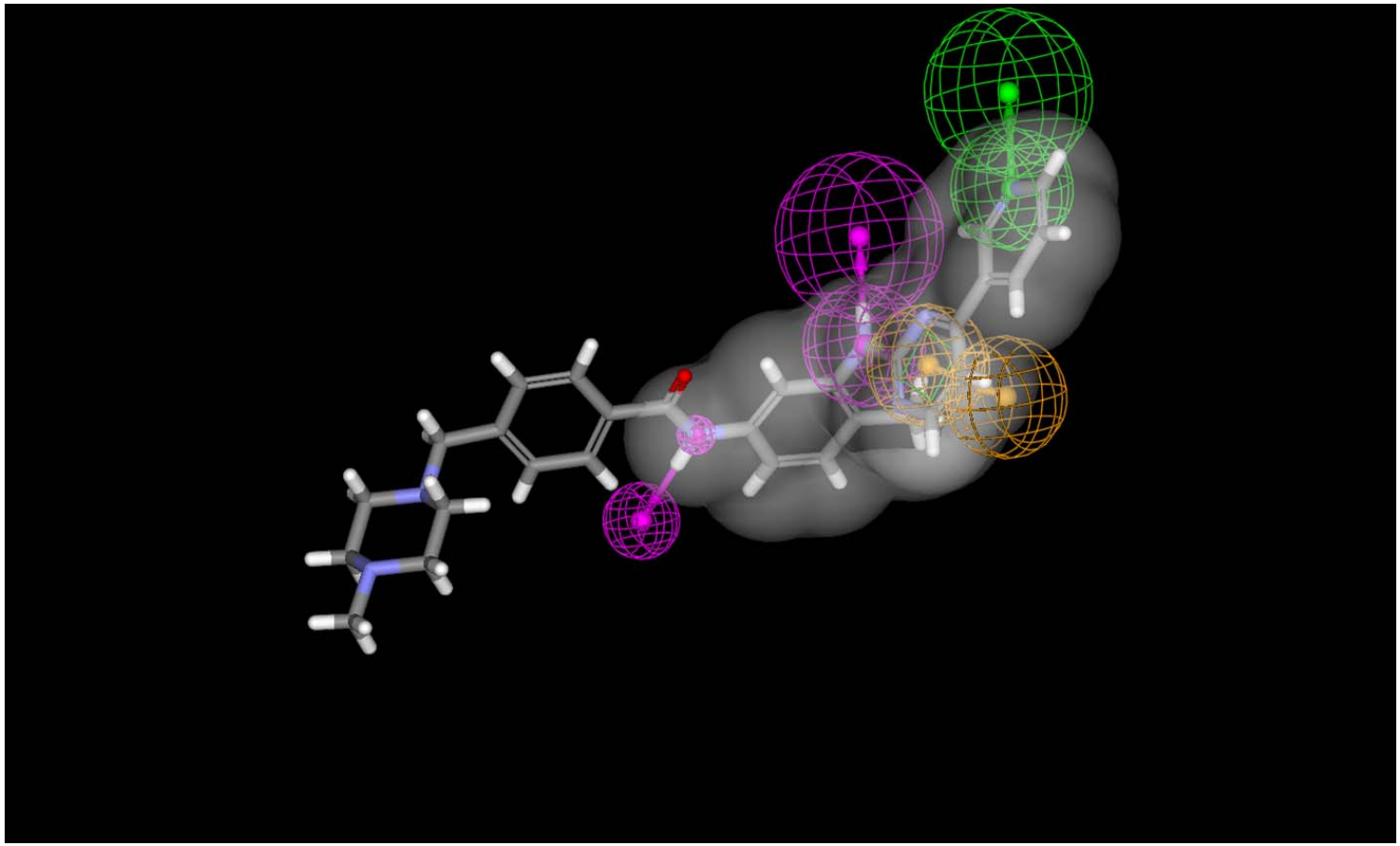
2. Search for fragment matches



3. Select **new analogs**



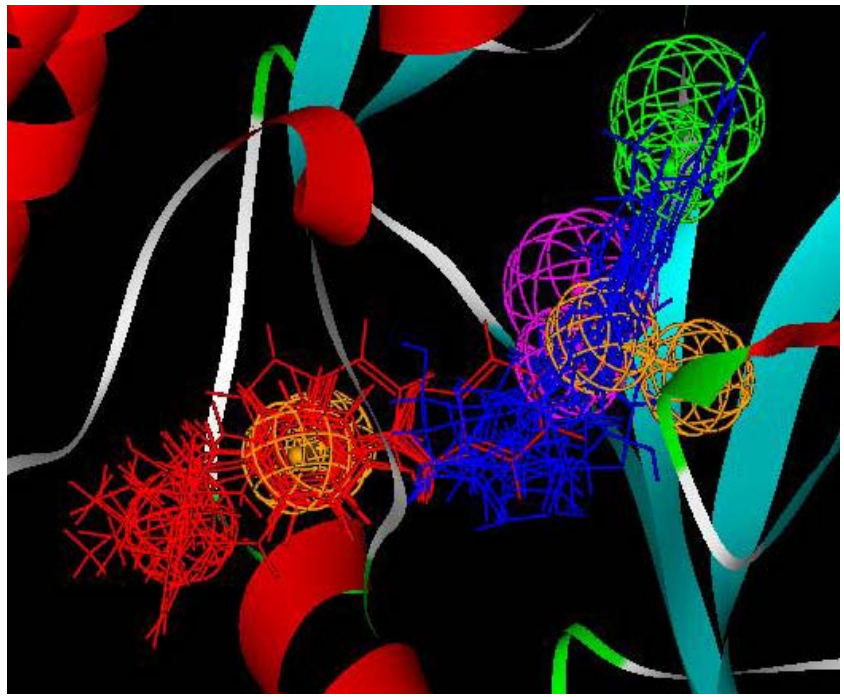
Gleevec



Gleevec: Results

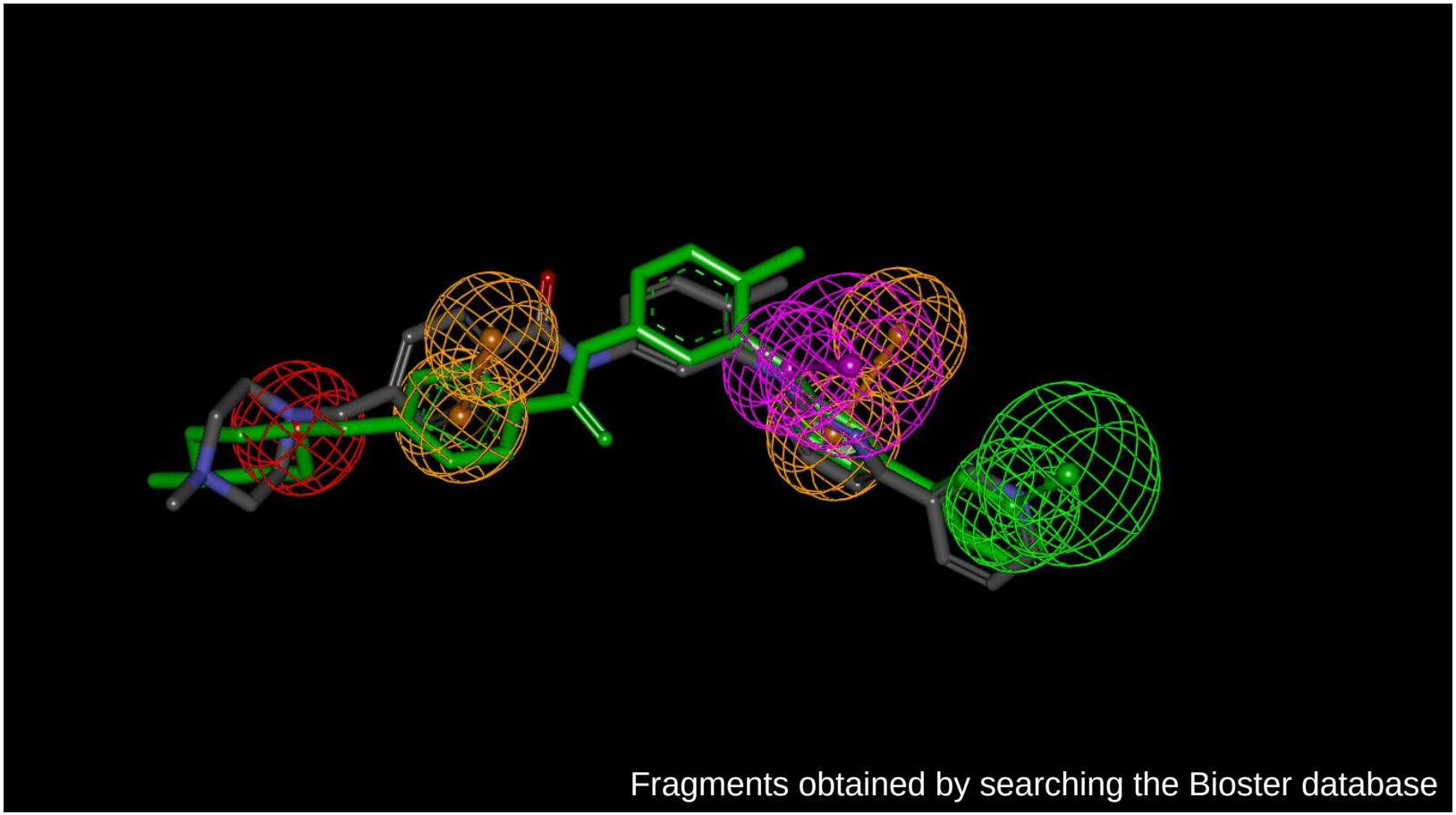
- Five-feature query returns 20 hits
- Total of 73,112 combinations from fragment search results using default (low) shape similarity cut-off

Query	Hits CAP (220,902)
Frag1	296
Frag2	247
5 features	20



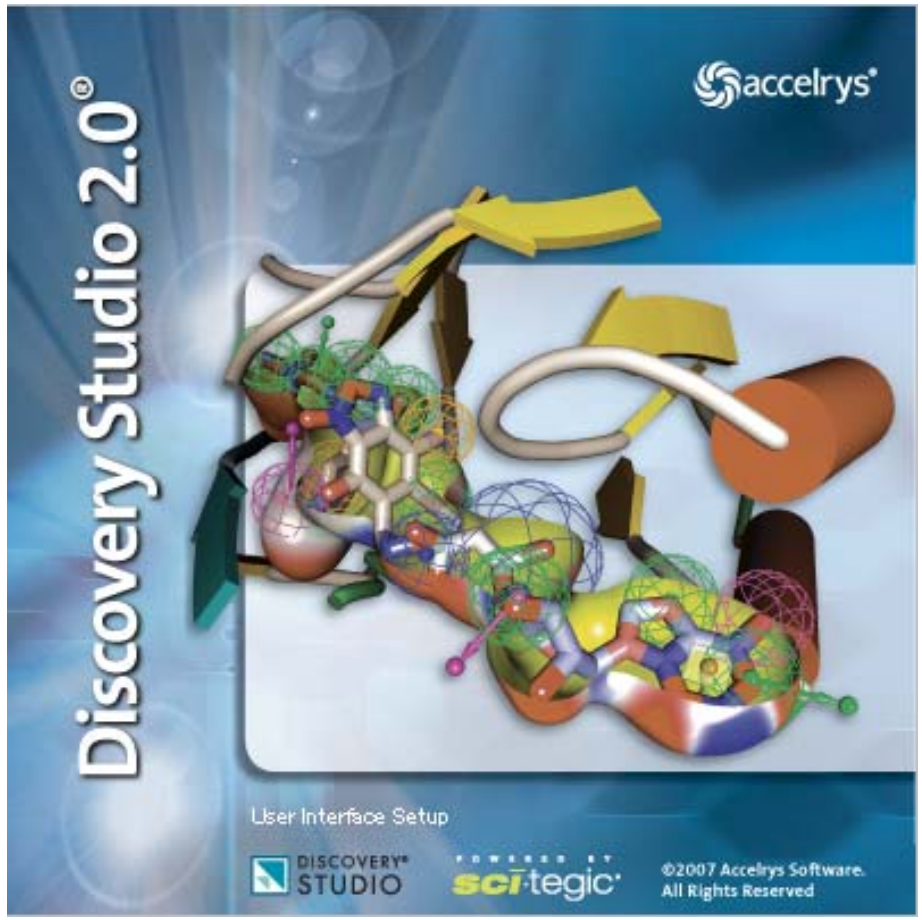
Can We Recover Gleevec?

- Only if the right fragments are available



www.accelrys.com

(Makes science faster !!)



Demo & Hands-on

- Basic features in Discovery Studio
- Molecular modeling based on NMR experiment data/constraints (by DS CHARMM)
- Adding NMR experimental data for Protein-Ligand interaction (by DS LigandFit, DS CHARMM)
- Protein-Protein interaction (by DS ZDock)