Finding Cancer Growth Inhibitors using the Internet

Keith Davies
Treweren Consultants

TC

Summary

- Virtual HTS
  - CAN-DDO
  - Find-a-Drug
- Preliminary Results
- Secondary Virtual Screen

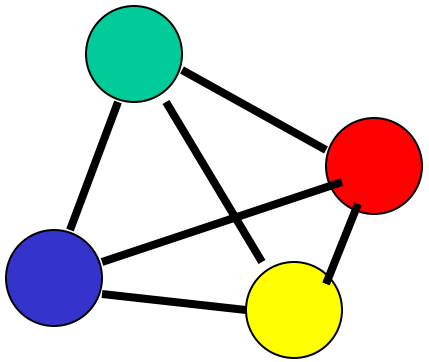
Blue Sky Objectives

- Reduce Development Failures
  - Real choice of leads candidates for follow-up
  - Better decisions
- Reduce Drug Discovery Research
  - Timescales
  - Costs

TC

Pharmacophores

- Centre Types
  - **H-bond Donor**
  - **H-bond Acceptor**
  - Acid
  - Base
  - **Positive Charge**
  - Negative Charge
  - **Aromatic Ring**
  - Lipophile
  - Lewis Base
  - Lewis Acid
  - 2 User Definable



- Options
  - 3 or **4 Centre yypes**
  - User definable bins
  - Occurrence frequency
- Output to file

TC

Conformer Generation

- Modes
  - Systematic
  - Random
  - Sample
- Contact Check
  - VdW
  - CPK (0.6*VdW)
- Bond Rotations
  - 3 Single (sp3-sp3)
  - 2 Conjugated
  - 4 Crowded
  - 6 Alpha (sp3-sp2)
  - 0 Amide (ie Off)
  - 0 Ring (ie Off)

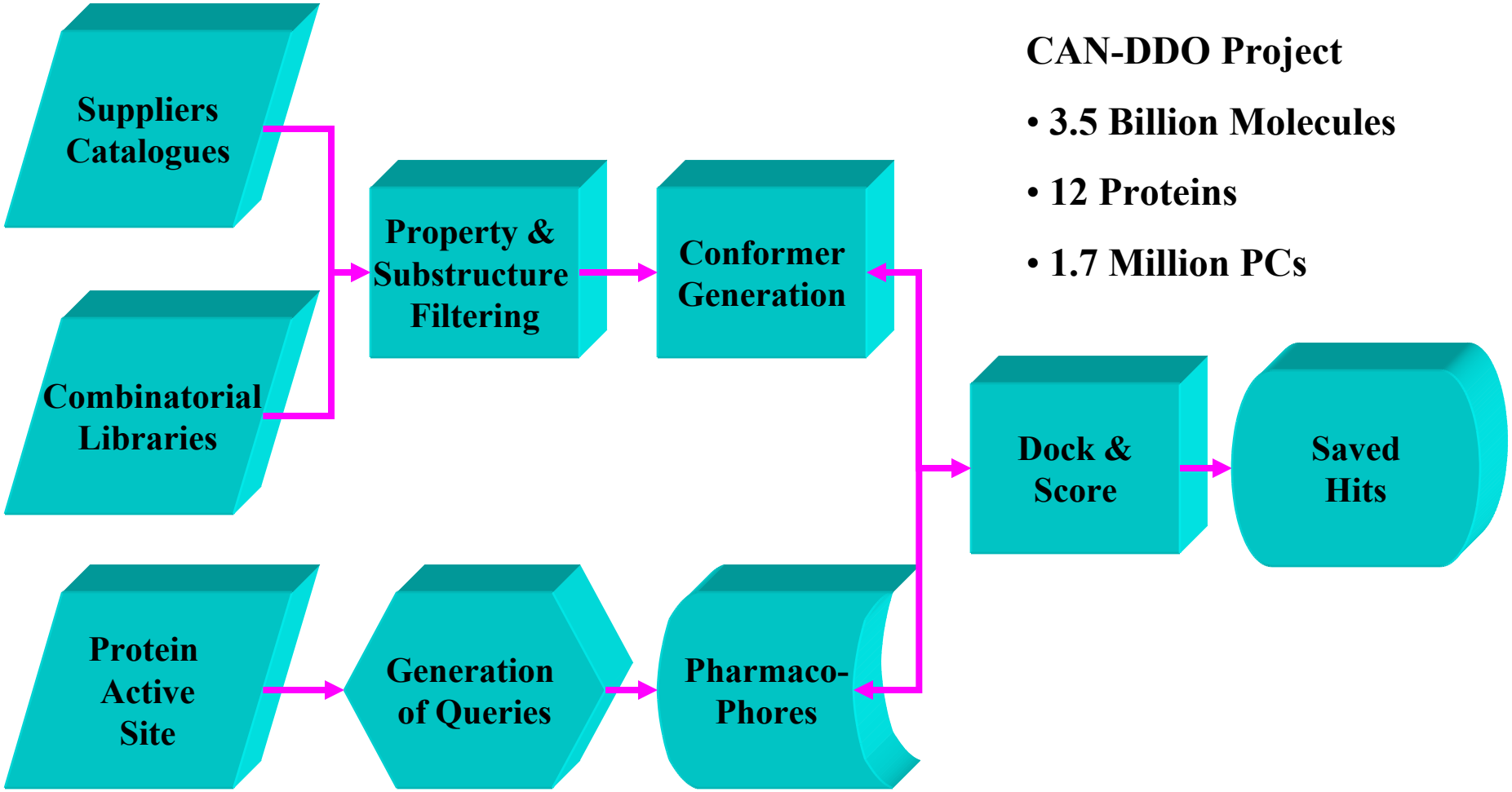
OMTC

CAN-DDO Project

- Cancer Research Project Organised by
  - Oxford University
  - United Devices (www.ud.com)
- Uses THINK Software as a Screen-saver
- Funded by NFCR and Sponsored by Intel
- 16 Targets 3.5 Billion Molecules
- Largest Computational Chemistry Project

TC

Virtual HTS



CAN-DDO Project

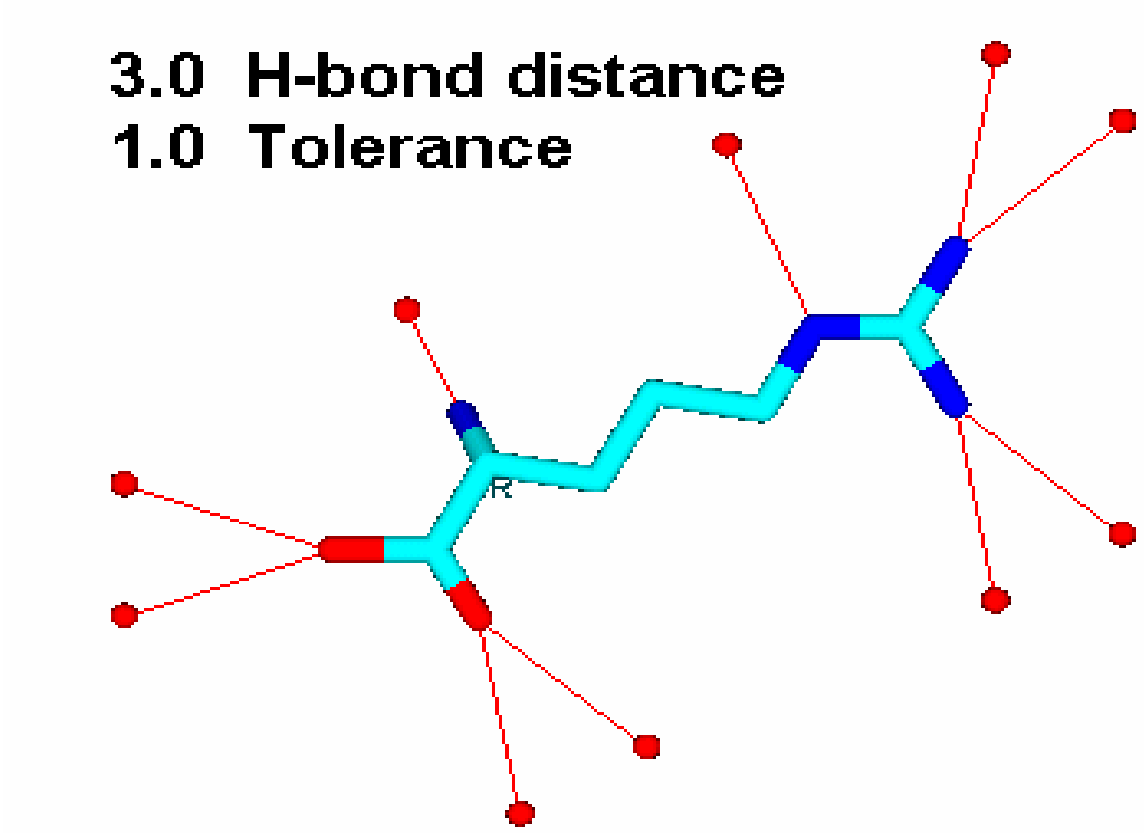
- **3.5 Billion Molecules**
- **12 Proteins**
- **1.7 Million PCs**

MTC

Cavity Identification

- Place Protein in 3-D Grid
- Remove Grid Points Inside Protein
- Remove Grid Points Inside 8 Å Sphere Positioned on the Grid Outside the Protein
- Binding Site
  - Minimum of 3 Grid Points
  - Minimum of 3 Residues

Centre Positions



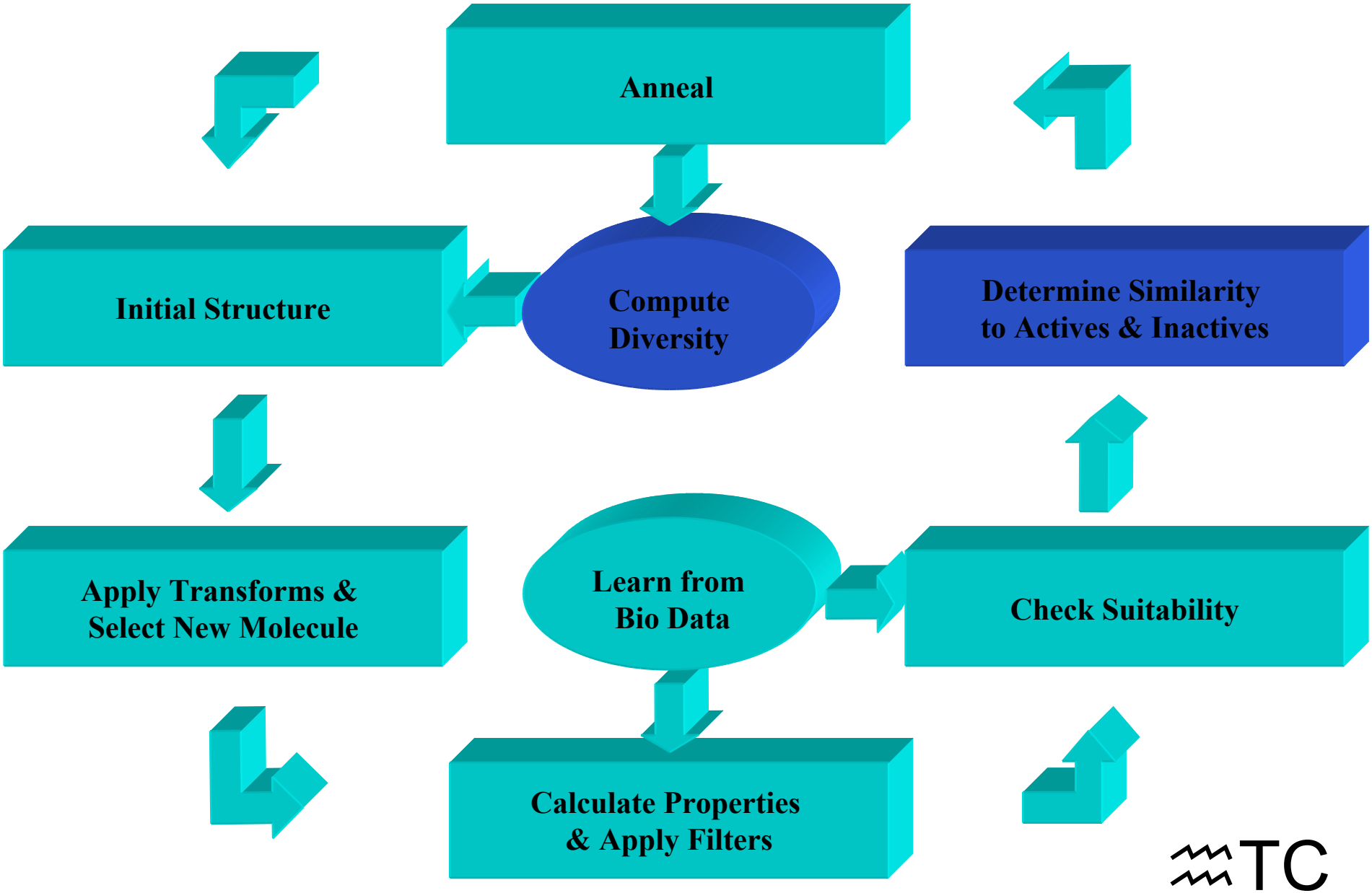
TC

Molecule Summary

- Reviewed
  - 1.4 Million Catalogue Molecules
  - 1 177 Million Library Molecules
- Drug-like
  - 0.4 Million Catalogue Molecules
  - 35 Million Library Molecules
- De Novo Derivatives
  - 100 for Each Molecule
  - Automatically Filtered
- Total of 3.5 Billion Molecules

TC

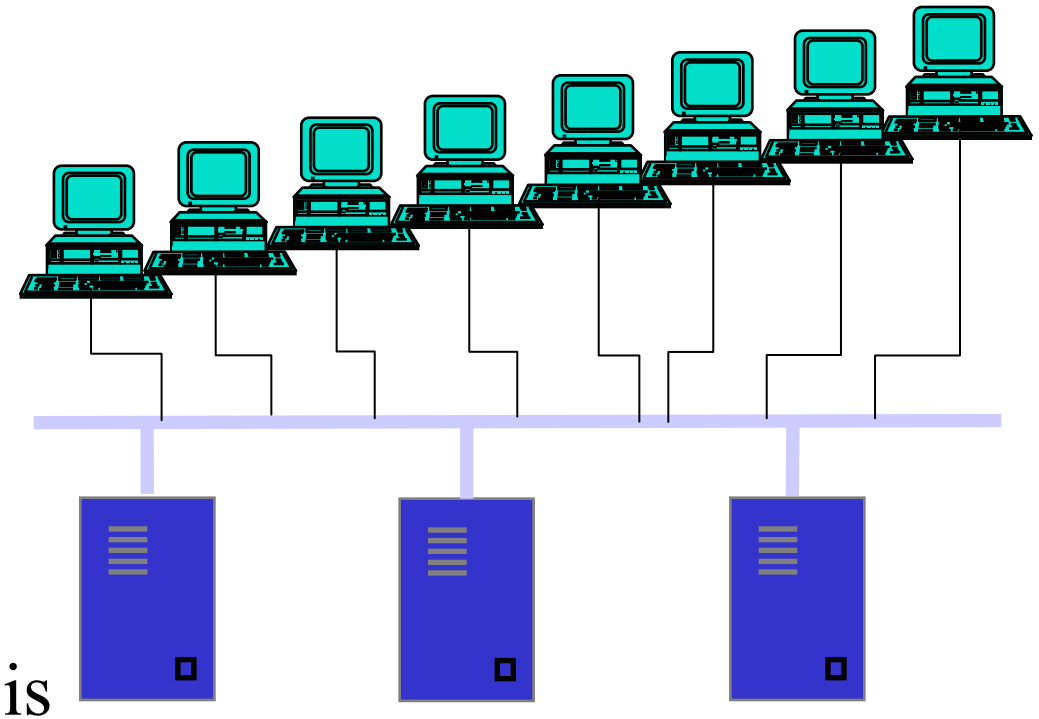
De Novo Structure Generation



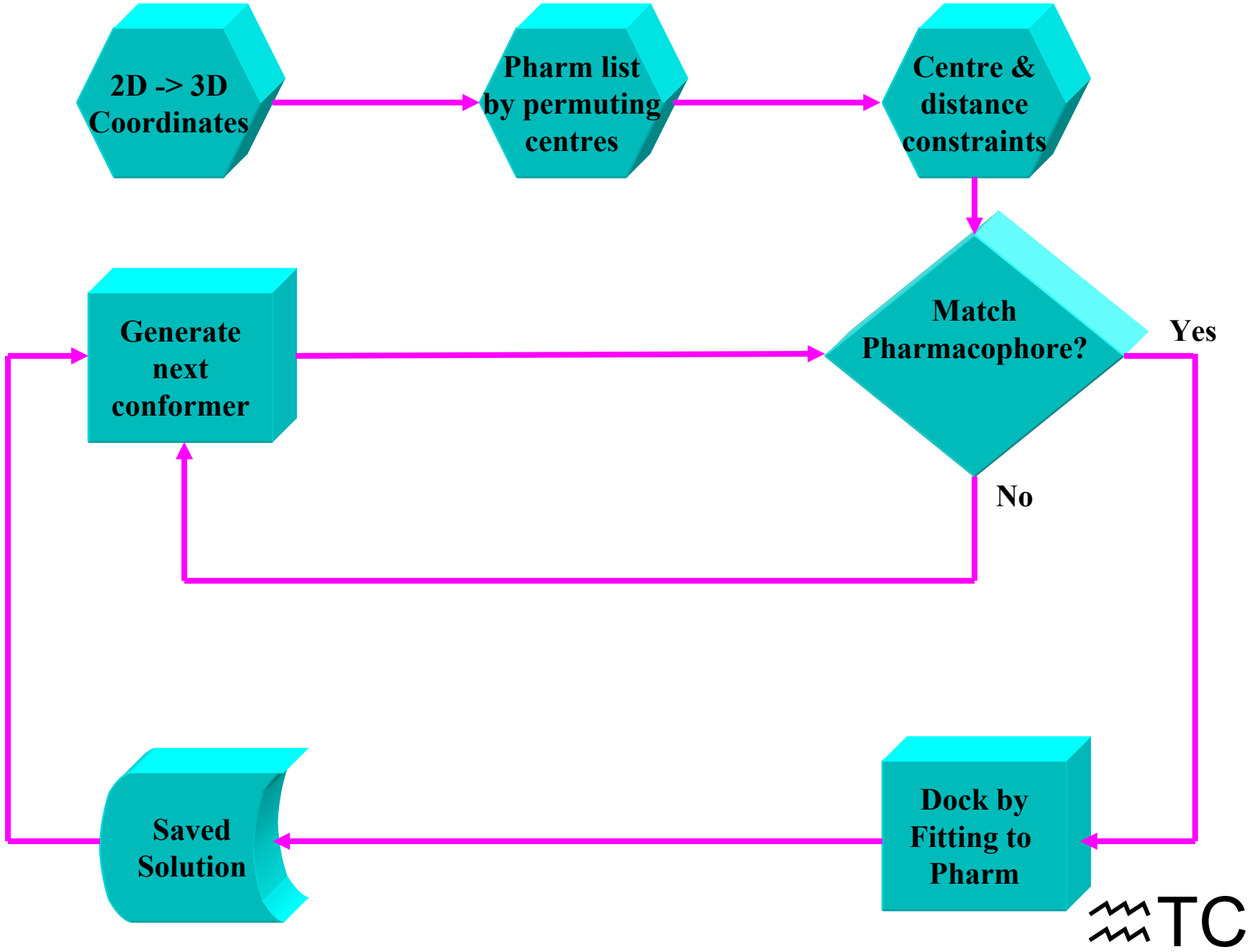
TC

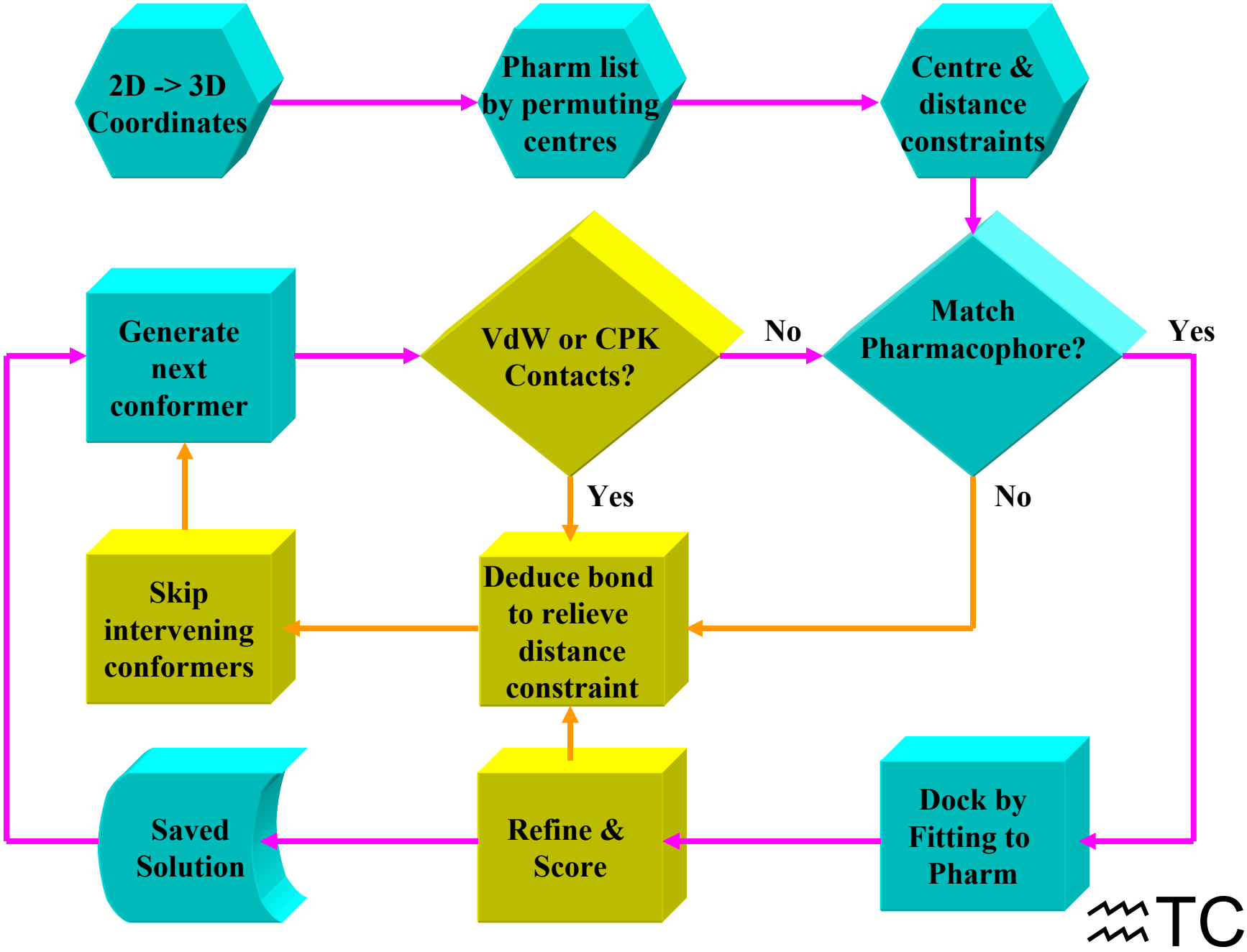
Find-a-Drug

- Multiple Servers
- Current Projects
  - Cancer
  - HIV
  - SARS
  - Bioterrorism
  - Multiple Sclerosis
  - Proteome



TC


2D -> 3D Coordinates
Pharm list by permuting centres
Centre & distance constraints
Match Pharmacophore?
Yes
No
Generate next conformer
Saved Solution
Dock by Fitting to Pharm
TC


2D -> 3D Coordinates
Pharm list by permuting centres
Centre & distance constraints
Generate next conformer
VdW or CPK Contacts?
No
Match Pharmacophore?
Yes
Yes
No
Skip intervening conformers
Deduce bond to relieve distance constraint
Saved Solution
Refine & Score
Dock by Fitting to Pharm
TC

Enhanced ChemScore

- X,Y,Z position, orientation and torsions refined
- Ranking Hits Based on ChemScore Equation

$\Delta G = \Delta G_0 + \Delta G_{hbond} * N_{hbond} + \Delta G_{lipo} * N_{lipo} + \Delta G_{bad} * N_{bad} + \Delta G_{rot} * N_{rot} + E$
where

ΔG_0 ΔG_{hbond} ΔG_{lipo} ΔG_{bad} ΔG_{rot} are constants (-5.48;-3.34; -0.117; 0.058; 2.56)
N_{hbond} is the number of qualifying interactions (on geometric criteria)
N_{lipo} is the number of lipohilic contacts
N_{bad} is the number of lipohilic-hydrophilic contacts (extension)
N_{rot} is the number of frozen rotatable bonds in the ligand
E is the VdW interaction energy and ligand torsional energy (extension)

TC

Performance

- THINK 1.03 (used for CAN-DDO)
  - 42,000,000,000 molecules
  - 126,000 years
  - 900 molecules per CPU day (excluding redundancy)
- THINK 1.24b
  - Optimised with assistance from Intel
  - Up to **100** times faster
  - More centres useful for larger sites
  - Refinement of docked geometry
  - About **500,000** molecules per 3GHz CPU day

Validation

- Reproduce Ligand-Protein Crystal Structures
  - RMS deviation of non-H atoms
  - Docking score
- Dock Actives
  - Used for developing scoring functions
- Prediction
  - Enrichment over random
  - Percentage of false positives

Selection Criteria

- Possible Kinases Cancer Targets
- X-ray Crystal Structures in PDB
- Resolution (1.9-2.8)
- All Atoms (1IAN Cα only)
- Ligand Flexibility <=10 Rotatable Bonds (excludes 1GAG, 1IR3, 2FGI, 5TMP, 1LCK)
- 21 Structures Processed

TC

PDB ID	Score	RMS	Notes
2KI5	-52.2 (-39.8)	5.88 (5.32)	3 A S
1QHI	-78.4 (-73.5)	1.03 (0.98)	3 A
1STC	-105.7	0.59	2
1E8Z	-86.8	1.77	2
1AQ1	-87.0	0.89	2
1AGW	-60.0 (-53.9)	3.45 (0.65)	2
1FGI	-83.8 (-69.4)	0.56 (0.52)	2
1FVV	-81.3 (-73.3)	0.61 (0.56)	3
1FVT	-69.8 (-58.5)	1.44 (0.72)	3

A All Site Points
S Single Bond Increment
C Conjugated Bond Increment
2 Two Centre Fit
3 Three Centre Fit
W Water Site Point

TC

PDB ID	Score	RMS	Notes
1DI8	-55.2 (-52.9)	0.98 (0.70)	2 W
1DI9	-39.5	1.97	2 W C
1YDR	-60.1 (-43.9)	3.00 (2.33)	2
1YDS	-55.5 (-48.1)	2.80 (0.76)	2 W
1YDT	-36.8	2.86	2
1PME	-28.4	1.30	3 W*2 C
4ERK	-17.9 (-11.8)	7.83 (3.33)	2
1IEP	-90.6 (-84.8)	0.73 (0.69)	3
1FPU	-47.1	0.75	3 C
1DM2	-54.0	0.61	3
1CKP	-43.2 (-43.0)	1.99 (1.10)	2 C
1QCF	-58.8	1.33	2

Validations

- Trace Mode
  - Drug-like Molecules
  - Number of Interactions in “Pharmacophore”
  - Volume, Area and Accessibility Constraints
  - Conformational Increments
  - Contacts & Refinement Threshold
- Observations
  - Most Ligand-Protein Crystal Complexes Validate
  - Best Ligands Score <-50kJ/mol

Configuration Options

- Reproducing Known Complexes
  - Slow and rigorous
  - Sometimes < 4 centre
  - Aim for RMS < 2.0 and interaction geometry
- Virtual HTS
  - Fast and more approximate
  - 4 Centres for greater selectivity
  - Accept some omissions
  - Avoid false positives

Find-a-Drug Cancer Results

PDB Code	Protein	Number Tested	Number Active
1FLT	VEGFr	3	3
821P	RAS	48	4
1C1Y	RAF	3	1
1FGI	FGFr-1	44	6
1E7U	PI3K	47	5

TC

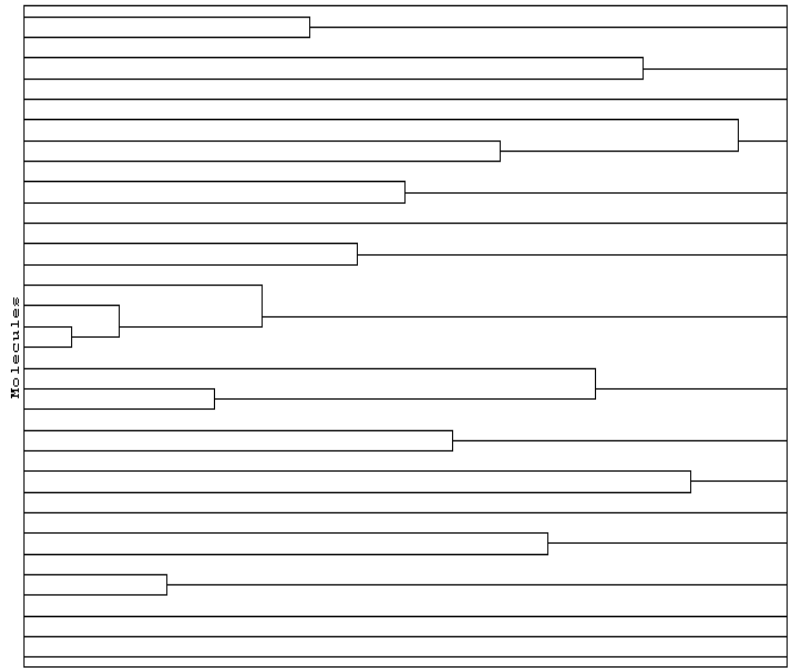
Computational Issues

- Motivation and Fabrication
  - Points awarded to measure member contribution
  - Encourages fabrication of results
- Use Redundancy to Identify
  - Erroneous results
  - Compiler dependency
- CPU Dependency
  - Rounding errors
  - Overclocking

What use are >1,000,000 Hits?

- More Choice Means Better Decisions
- Analyse and Select by
  - Families
  - Clustering
  - Ligand-Protein Interactions
  - Medicinal Chemistry Rules
  - Binding Mode
  - Snugness of Fit

Clustering

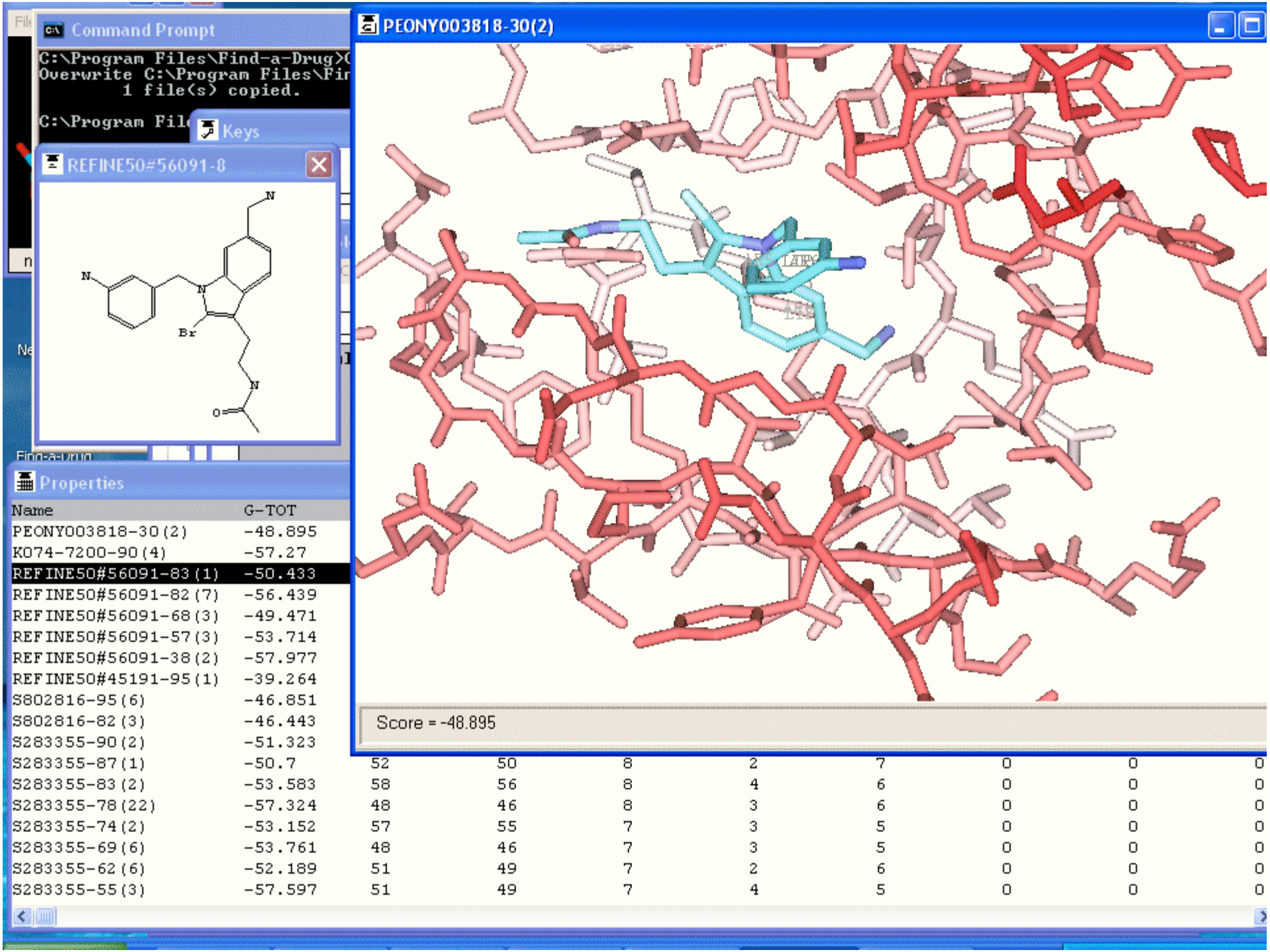


- Advantages
  - Idiot Proof
  - Functional Groups
  - Provides Representatives
- Disadvantages
  - Approximate for large hit lists
  - No Receptor or Synthetic Justification

TC

Observations of Clustering

- All Molecules in a Cluster must be Similar
- Not all Similar Molecules in the same Cluster
- Derivatives are not necessarily Similar
- Similarity does not relate to Synthetic Route
- Clustering is “Not a thinking man’s choice”



Name	G-TOT							
PEONY003818-30(2)	-48.895							
K074-7200-90(4)	-57.27							
REFINE50#56091-83(1)	-50.433							
REFINE50#56091-82(7)	-56.439							
REFINE50#56091-68(3)	-49.471							
REFINE50#56091-57(3)	-53.714							
REFINE50#56091-38(2)	-57.977							
REFINE50#45191-95(1)	-39.264							
S802816-95(6)	-46.851							
S802816-82(3)	-46.443							
S283355-90(2)	-51.323							
S283355-87(1)	-50.7	52	50	8	2	7	0	0	0
S283355-83(2)	-53.583	58	56	8	4	6	0	0	0
S283355-78(22)	-57.324	48	46	8	3	6	0	0	0
S283355-74(2)	-53.152	57	55	7	3	5	0	0	0
S283355-69(6)	-53.761	48	46	7	3	5	0	0	0
S283355-62(6)	-52.189	51	49	7	2	6	0	0	0
S283355-55(3)	-57.597	51	49	7	4	5	0	0	0

Medicinal Chemistry Rules

- Impractical to Review 000's of Hits
- Synthetic Knowledge and Experience
  - Chirality
  - Synthetic difficulty
  - Poor *De Novo s*uggestions
  - Additional toxicity and metabolism alerts
- Select Leads from Hits
- Need Families not Lone Actives

Binding Mode

- Defined by Pharmacophores
- List of Interacting Residues
  - Ligand atoms mapped by to protein atoms
  - Serial numbers in hits file
- Favour Hits which Interact with the Same Residues as Known Actives

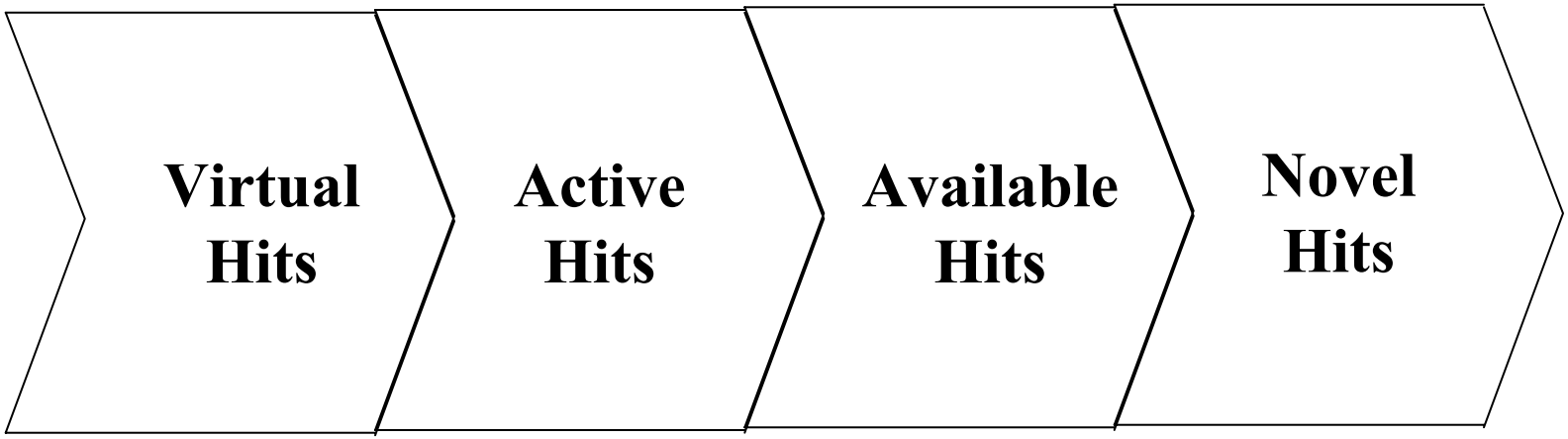
Hole Count

- Measure of Snugness of Fit
- Use 1 Angstrom Grid
- Calculate Difference of
  - Number of Holes in Free Protein
  - Number of Holes in Ligand-Protein Complex
- Hole Count is Negative for Snug Fits

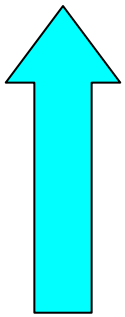
Secondary Virtual Screen

- Consider Rich Families Only
- Rejection by
  - Chiral Centre Count
  - Unacceptable Functional Groups
  - Extended Property Ranges
- Numerical Combination of
  - Docking Score
  - Pharmacophore Overlap
  - Hole Count
- Use False Positives as a Test Set

Find-a-Drug Progress



- **Virtual HTS**
- **Secondary Virtual Screen**



- **Future Possibilities**
  - **P450 Virtual Screen**
  - **Pharmacophores**

TC

FGFr-1 Progress (1FGI)

- Virtual HTS
  - 13.6% True Positives (6/44)
- Enrichment
  - 125 (6 actives+2/977 from ChemHTS-1)
  - 512 (6 actives+13/9729 from ChemHTS-1)
- Secondary Virtual Screen
  - 29.4% True Positives (5/17)

TC