Ultra High Throughput Screening using THINK on the Internet

Keith Davies
Central Chemistry Laboratory, Oxford University
Cathy Davies
Treweren Consultants, UK

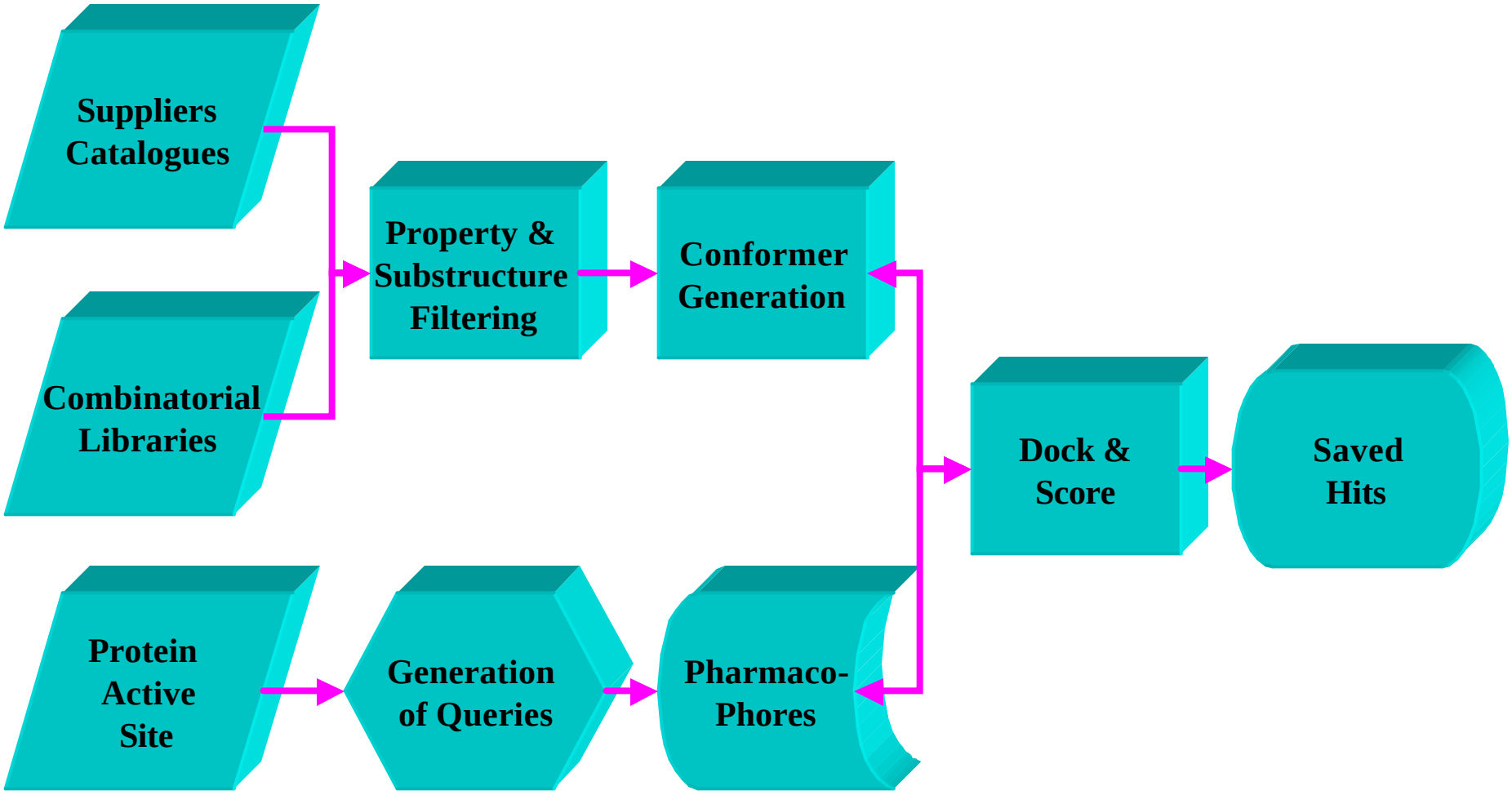
Blue Sky Objectives

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

CAN-DDO Project

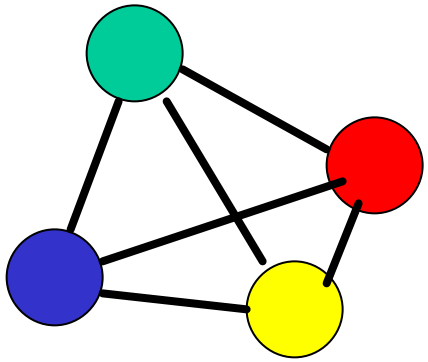
- Cancer Research Project Organised by
  - Oxford University
  - United Devices (www.ud.com)
- Used THINK Software as a Screen-saver
- Funded by NFCR and Sponsored by Intel
- 12 Targets 3.5 Billion Molecules
- Largest Computational Chemistry Project

CAN-DDO Virtual Screening



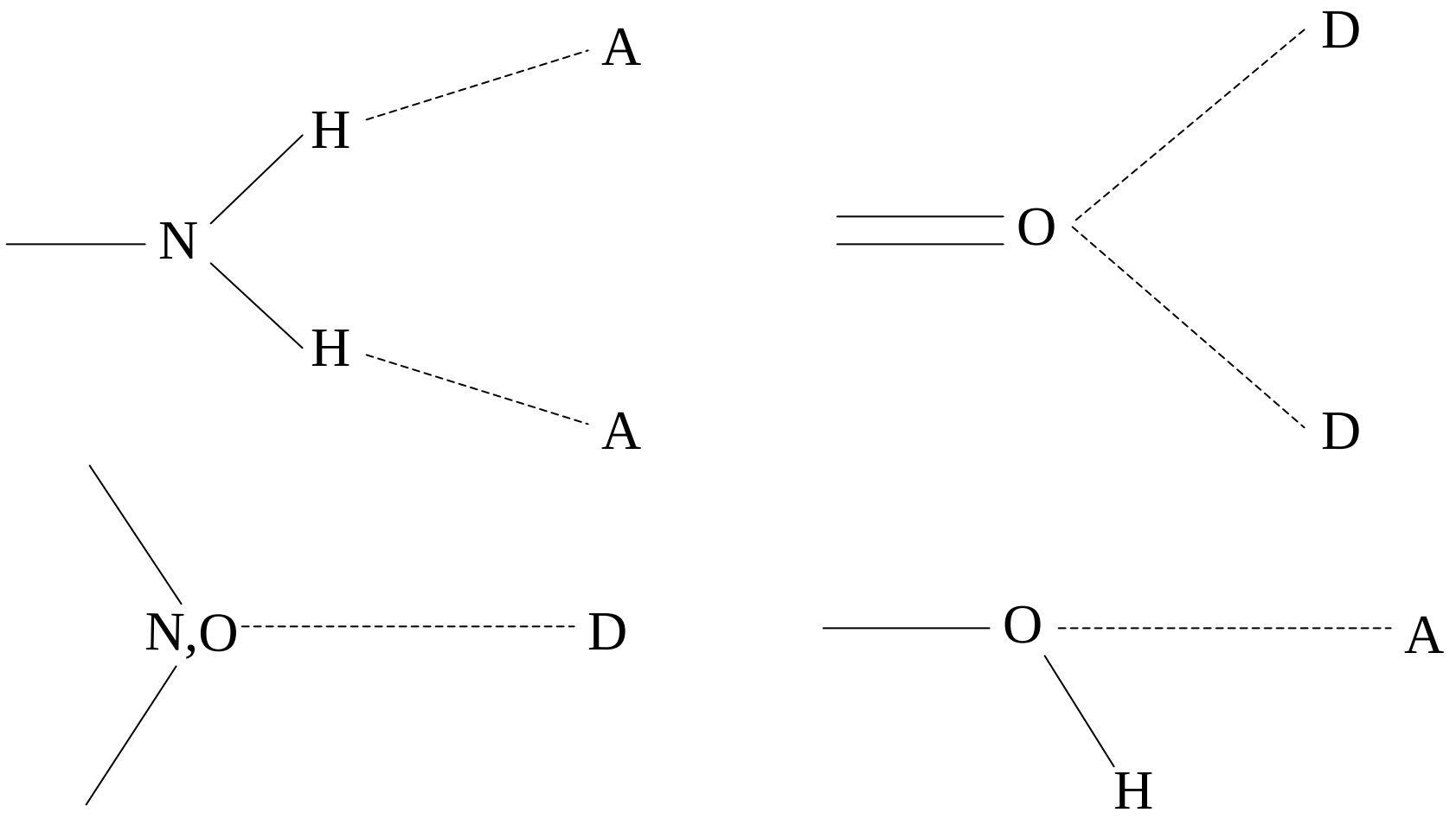
Pharmacophores

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



- Options
  - 3 or **4 Centre Types**
  - User Definable Bins
  - Occurrence Frequency
- Output to File

Generating Pharmacophores

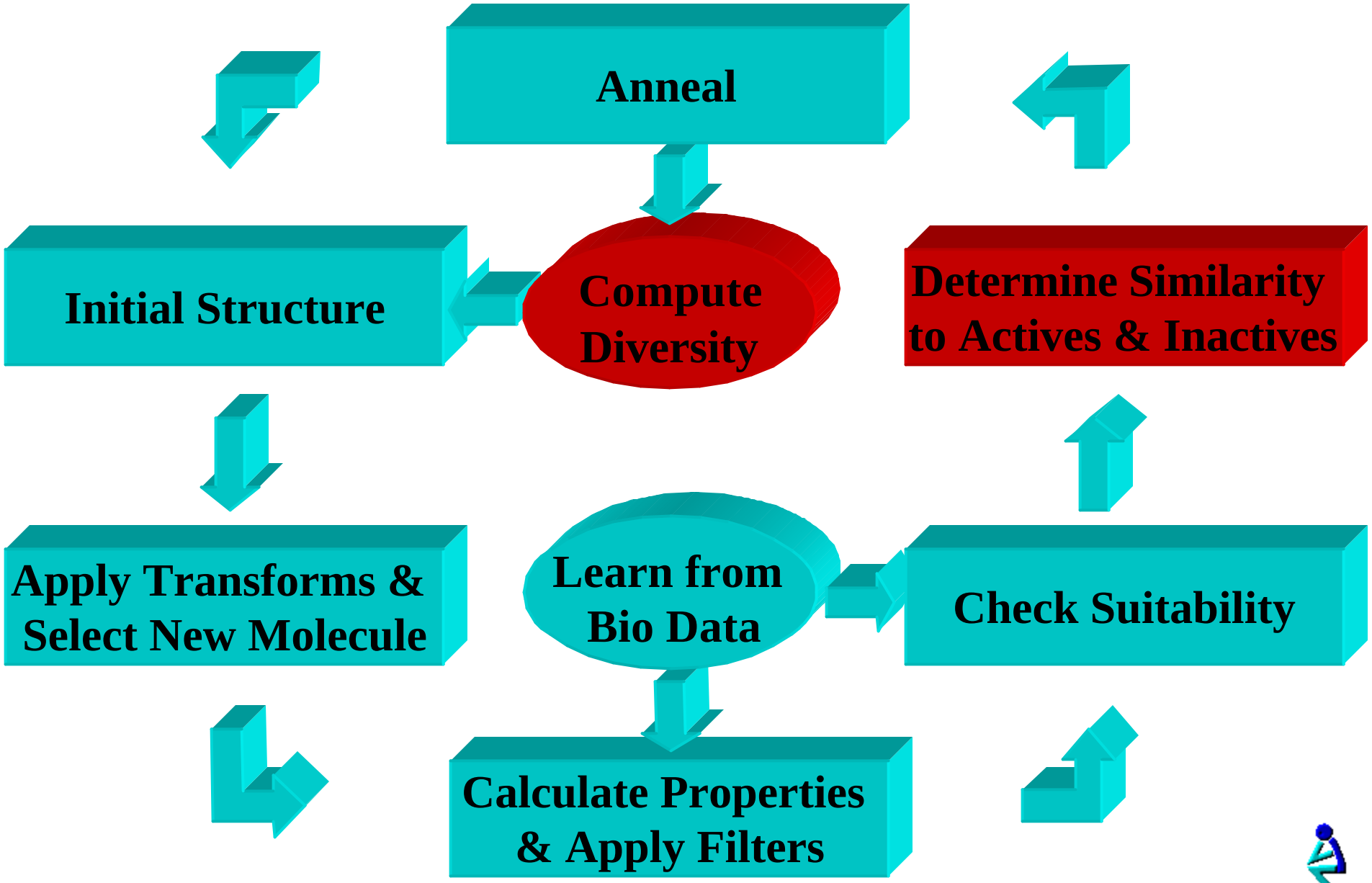


D=Donor; A=Acceptor (Not to scale)

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

De Novo Structure Generation



Filtering (1)

- **Properties**
- Substructures
  - Unstable
  - Reactive
  - Undesirable (eg toxic)

Centres	2:9
Mass	150:800
XSA	20:240
Rot-Bonds	<10
Conformers	<1000000

Filtering (2)

- Properties
- Substructures
  - **Unstable**
  - Reactive
  - Undesirable (eg toxic)

NOC	HNO
[ND]=O	[NITR]
N[SSP2]	O[SSP2]
OO	SS
NN	N#N
N=N=N	N[CAK]O
N=C=O	N=C=S
N=C=N	O[HAL]
N[HAL]	S[HAL]
OC[HAL]	NC[HAL]
C=COH	

Filtering (3)

- Properties
- Substructures
  - Unstable
  - **Reactive**
  - Undesirable (eg toxic)

CC(H)=C(H)C=O
C =COH
C=CNH
O=C[HAL]
N=C[HAL]
S=C[HAL]
[HAL]C[HAL]
HOCOH
O=COC=O
COS(=O)(=O)C
COS(=O)OC

Filtering (4)

- Properties
- Substructures
  - Unstable
  - Reactive
  - **Undesirable** (eg toxic)

[M]	[Si]N
[Si]O	SH
C1XC1	C1SC1
SC#N	C(=O)S
NP	PS
C=P	P[HAL]
CN#C	PC#N
O=CC#N	OCC#N
NC#N	CC(H)=O
PP	

Conformer Generation

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

Site Searching

- Generation of Complementary Centres
- Ligand Pharmacophores Matched to Site
- Ranking Hits Based on ChemScore Equation

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

where

ΔG_0 ΔG_{hbond} ΔG_{lipo} ΔG_{rot} are constants (-5.48;-3.34; -0.117; 2.56)

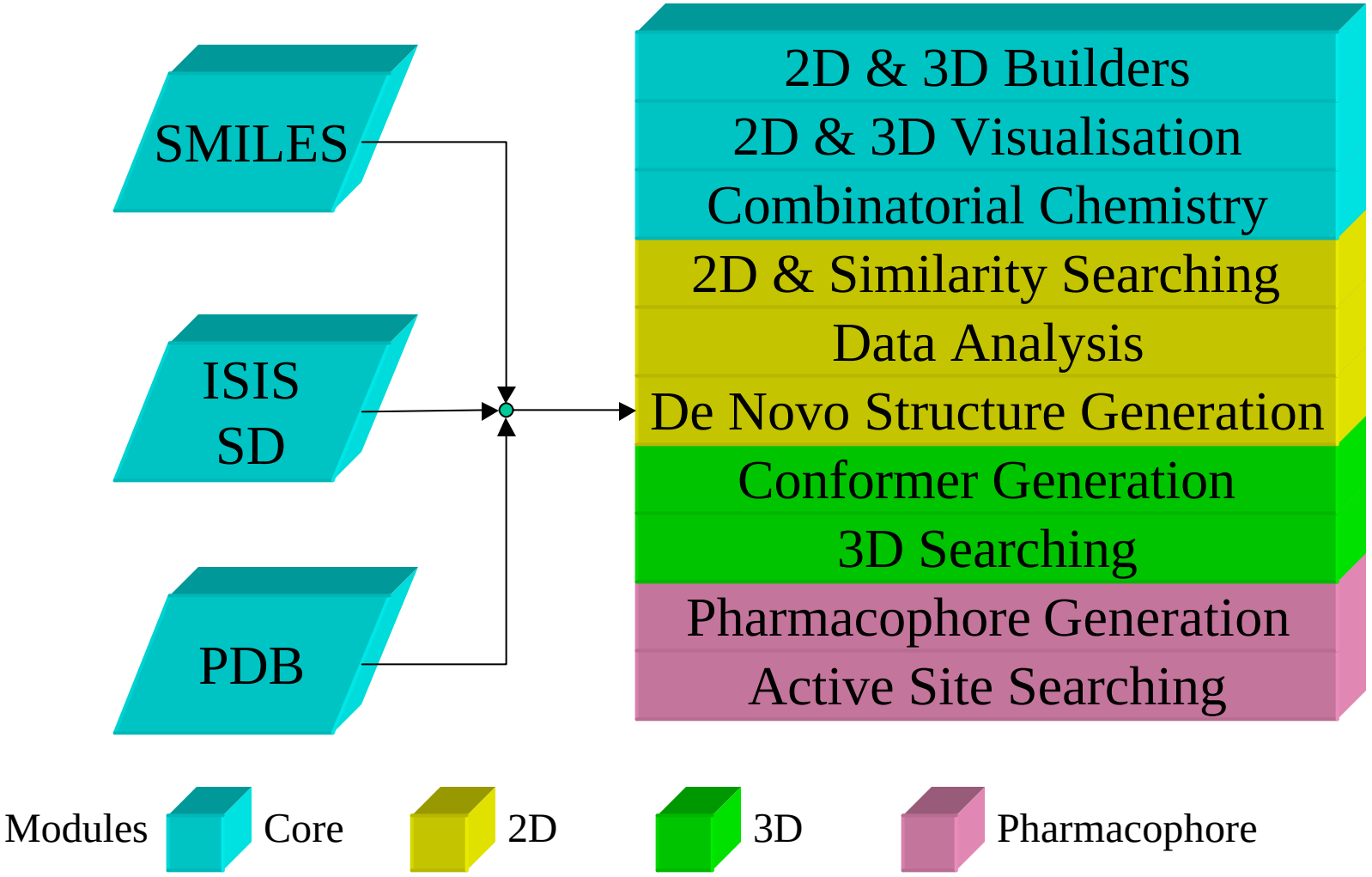
N_{hbond} is the number of qualifying interactions (on geometric criteria)

N_{lipo} is the number of lipohilic contacts

N_{rot} is the number of frozen rotatable bonds in the ligand

E is the VdW interaction energy and ligand torsional energy

THINK Architecture



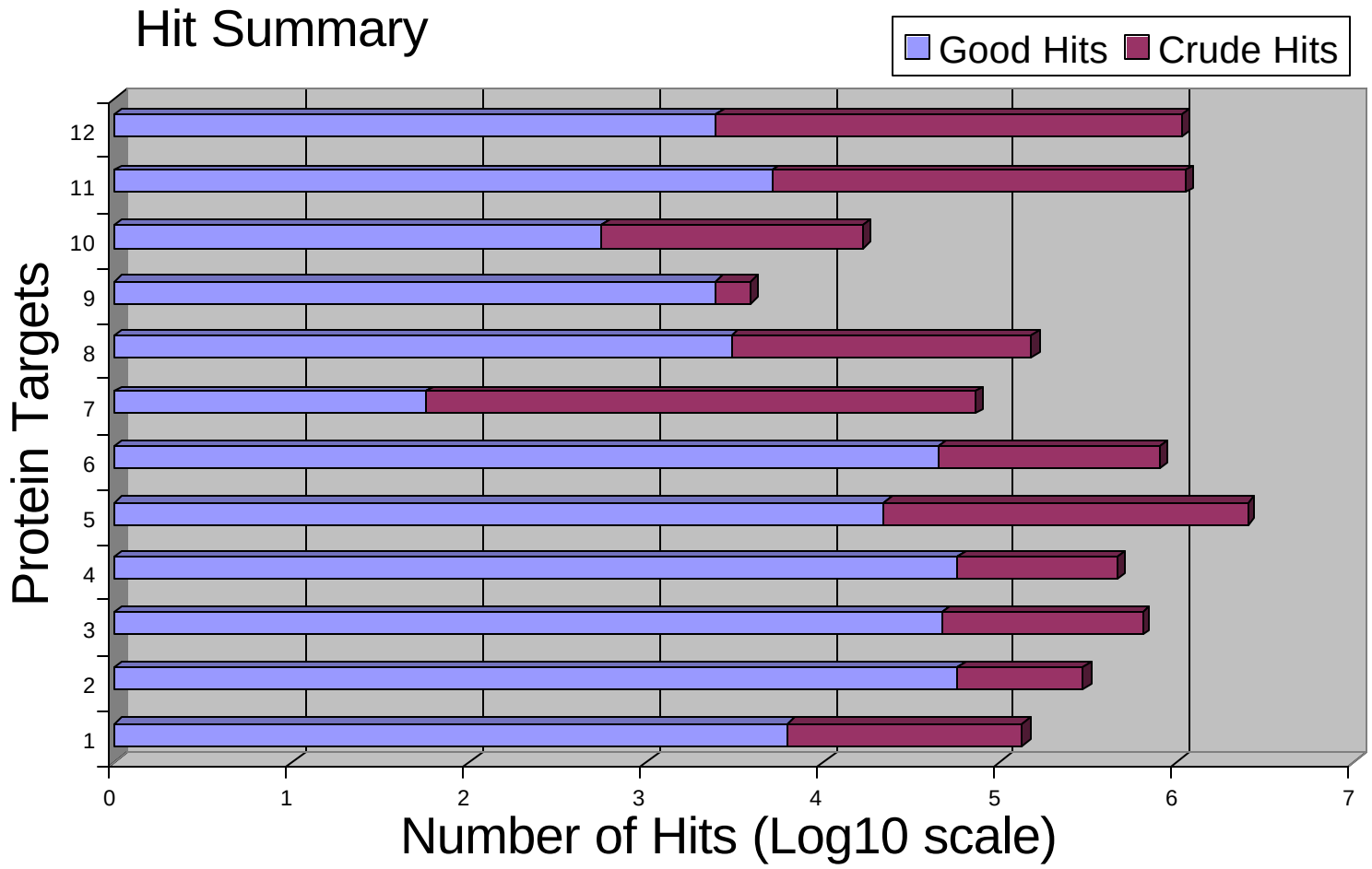
Data Bandwidth Issues

- Protein Query
  - PDB Format
  - Centres
  - Site Residues
  - 50-100Kbytes
- Molecule Data per Work Unit
  - 100 Molecules per Job in SMILES
  - 5-10 Kbytes
- Results
  - SMILES on Internet
  - ISIS SD on Intranet

Performance

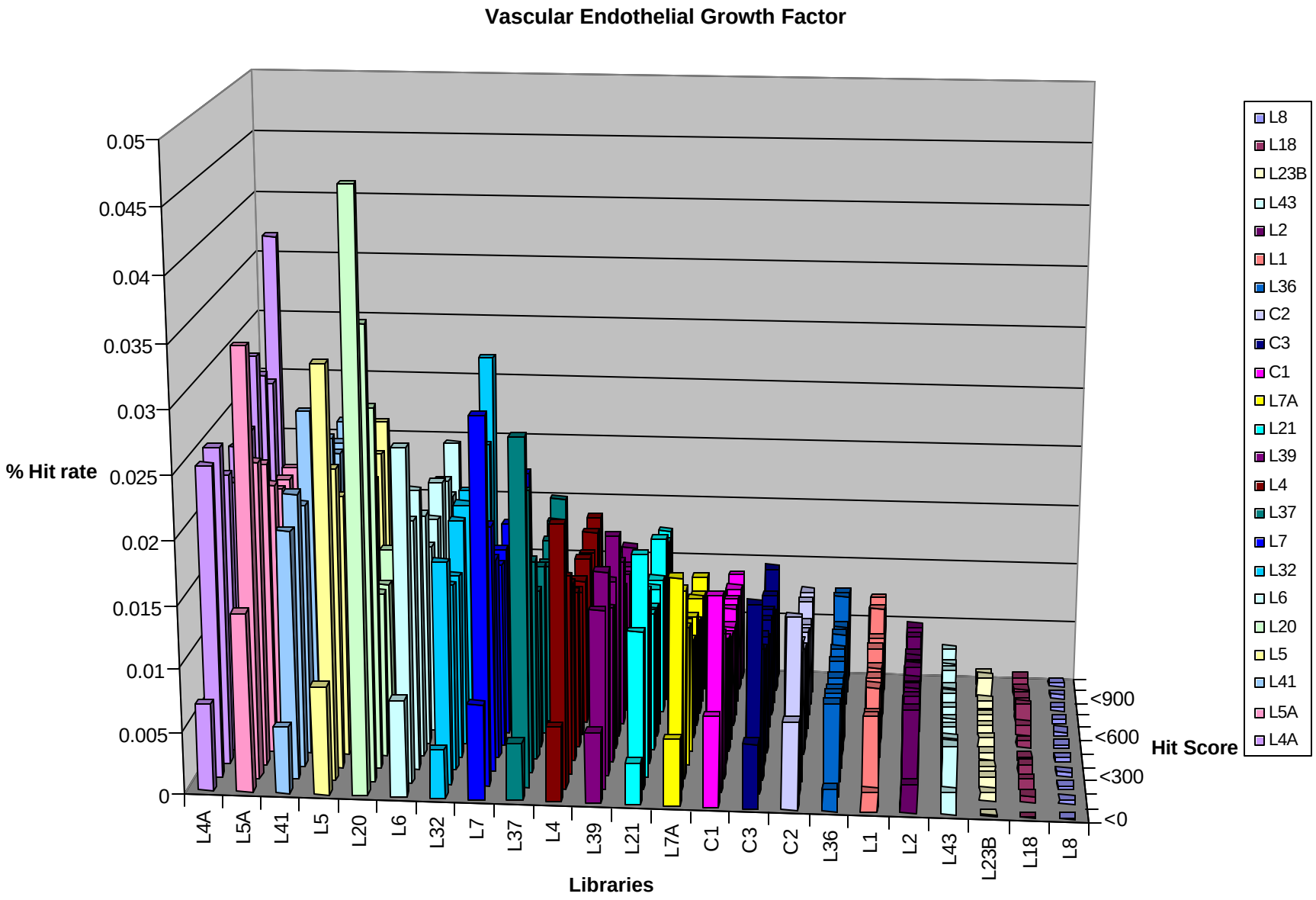
- THINK 1.03 (used for CAN-DDO)
  - 42,000,000,000 molecules
  - 126,000 years
  - 900 molecules per CPU day (excluding redundancy)
- THINK 1.2 (current release)
  - Optimised with assistance from Intel
  - Up to 40 times faster
  - More centres useful for larger sites
  - Refinement of docked geometry
  - About 250,000 molecules per 2GHz CPU day

Summary Results



Results to Date

- 7.2 Million Preliminary Crude Hits
- Post-Processing (using LigandFit)
  - Eliminate more false positives
  - Optimise X,Y,Z position, rotation and torsions
- Oxford University Spin-off - Inhibox

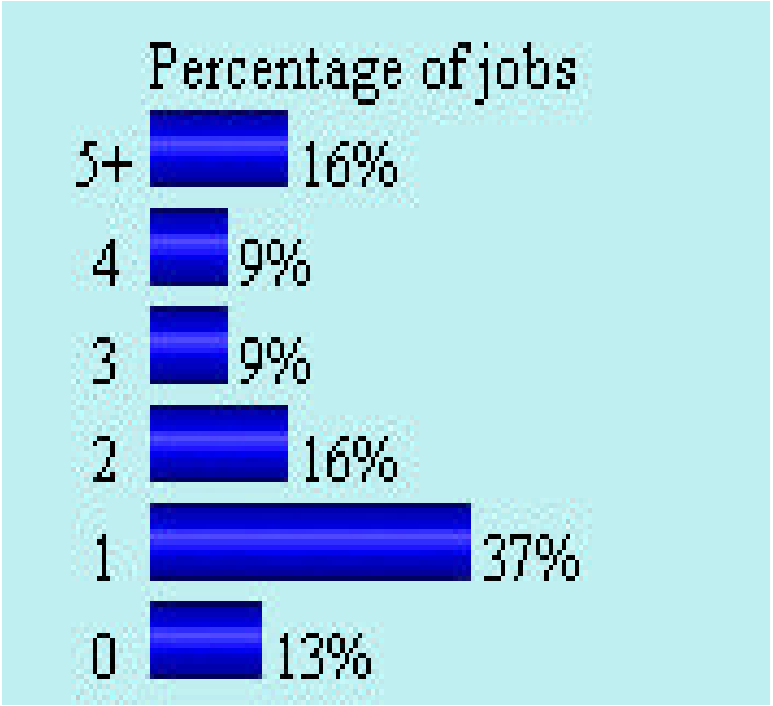

Vascular Endothelial Growth Factor
% Hit rate
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
L4A
L5A
L41
L5
L20
L6
L32
L7
L37
L4
L39
L21
L7A
C1
C3
C2
L36
L1
L2
L43
L23B
L18
L8
Libraries
<900
<600
<300
<0
Hit Score
L8
L18
L23B
L43
L2
L1
L36
C2
C3
C1
L7A
L21
L39
L4
L37
L7
L32
L6
L20
L5
L41
L5A
L4A

Member Statistics

- Motivating Volunteers
  - Feedback & Recognition
  - Counting Jobs, Molecules, Hits etc
- Teams
  - Promoting Membership
  - Sense of identity and dependence
- Points & Rewards
  - Donations to Charities eg “Make a Wish”

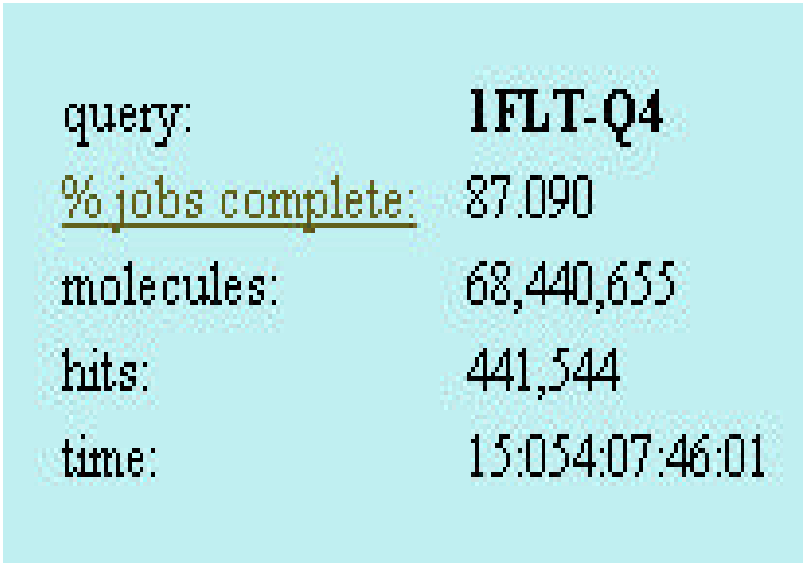
Calculation Redundancy

- Validation
  - Processors
  - Tampering
- Non Returns
  - Member Attrition
  - Job Failures
- Job Scheduling
  - Transaction Model
  - Multiple Servers



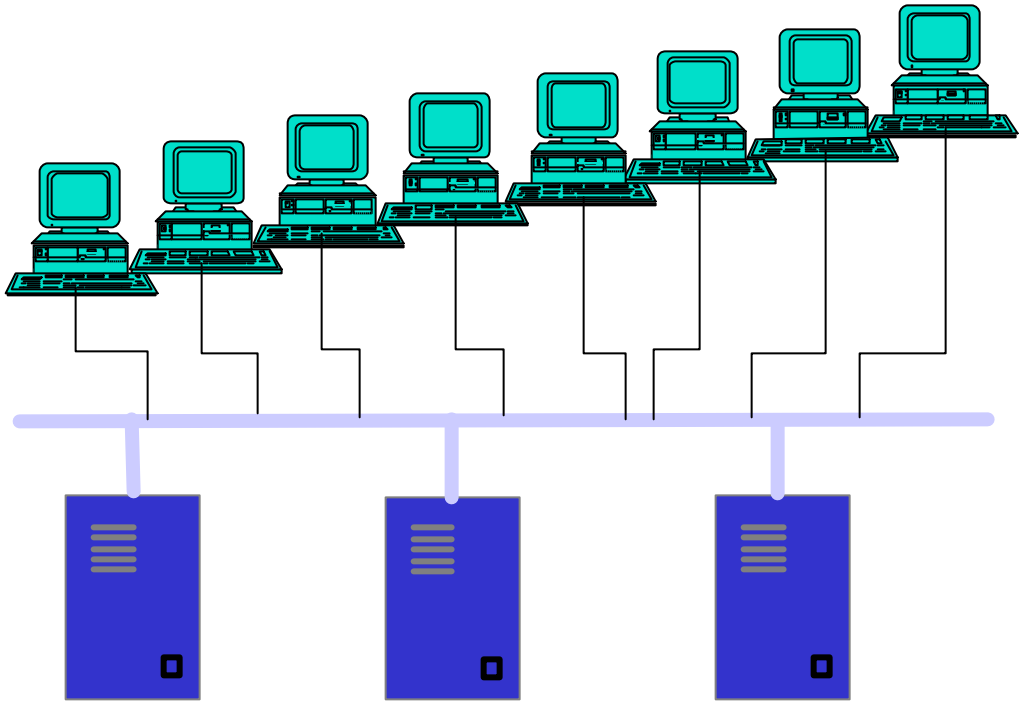
Project Progress

- Meaningful Feedback
- Open Communication
- Project Comparisons



Find-a-Drug

- Multiple Servers
- Future Projects
  - Cancer
  - HIV
  - Bioterrorism
  - Proteome



Acknowledgements

- United Devices, TX
- NFCR
- Graham Richards, Oxford University