SCIENTISTS AGAINST MALARIA

EXPERIENCE USING CESGA- FCSCL NORTHWEST SCIENCE CLOUD

Virtual Screening on a Pf-Kinase

Hugo Gutiérrez de Terán, PhD
Public Galician Fundation of Genomic Medicine (FPGMX)

Fundación Pública Galega de Medicina Xenómica

SERVIZO GALEGO de SAÚDE

CESGA

Scientists Against Malaria

Virtual Organisation for Drug Discovery

DouglasConnect — Roman Affentranger, Barry Hardy

EPFL École Polytechnique Fédérale de Lausanne — Christian Doerig, Matteo Dal Peraro

CERF by Rescentris, Inc. — Jeff Spitzner

University of Cincinnati Drug Discovery Center — Ruben Papoian, William Seibel, Sandra Nelson

Università degli Studi di Milano Facoltà di Farmacia — Alessandro Contini

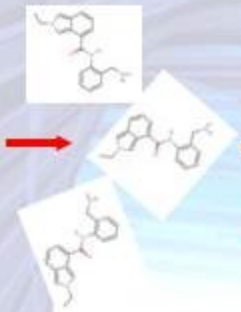
Monash University — Andrew Wilks, Isabelle Lucet

Fundación Pública Galega de Medicina Xenómica — H. G. de Teran

PHARMATROPE — Jeffrey Wiseman

inte:ligand Your partner for in-silico drug discovery. — Sharon Bryant

Synergy Collaboration Pilots



Screened Library

Data

CERF

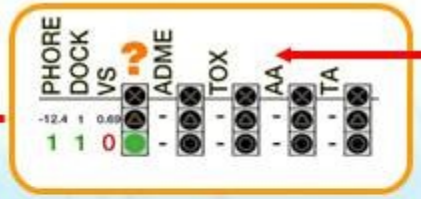
Refine Predictions

SYNERGY

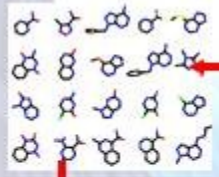
Data

Toxicity Predictions

OpenTox



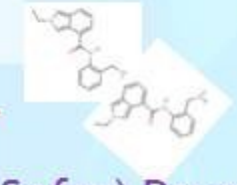
SYNERGY



Toxicity Assays

SYNERGY

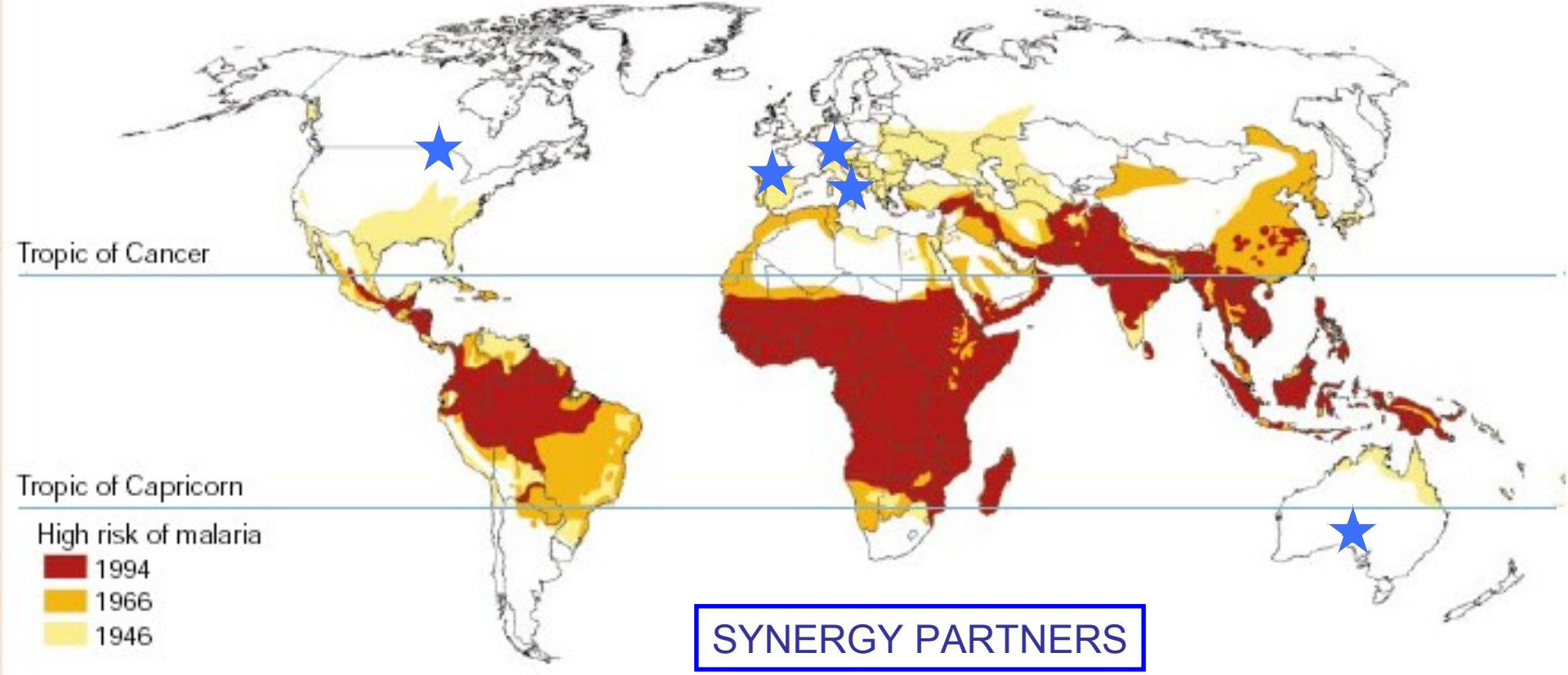
Decision Dashboard



(Safer) Drug Leads

OpenTox

- Third world disease
- 500 million clinical cases per year
- 1.5-3 million deaths per year (children bellow 5!)
- Number of cases constantly increasing
- Several therapeutic tools, but all of them generate resistances

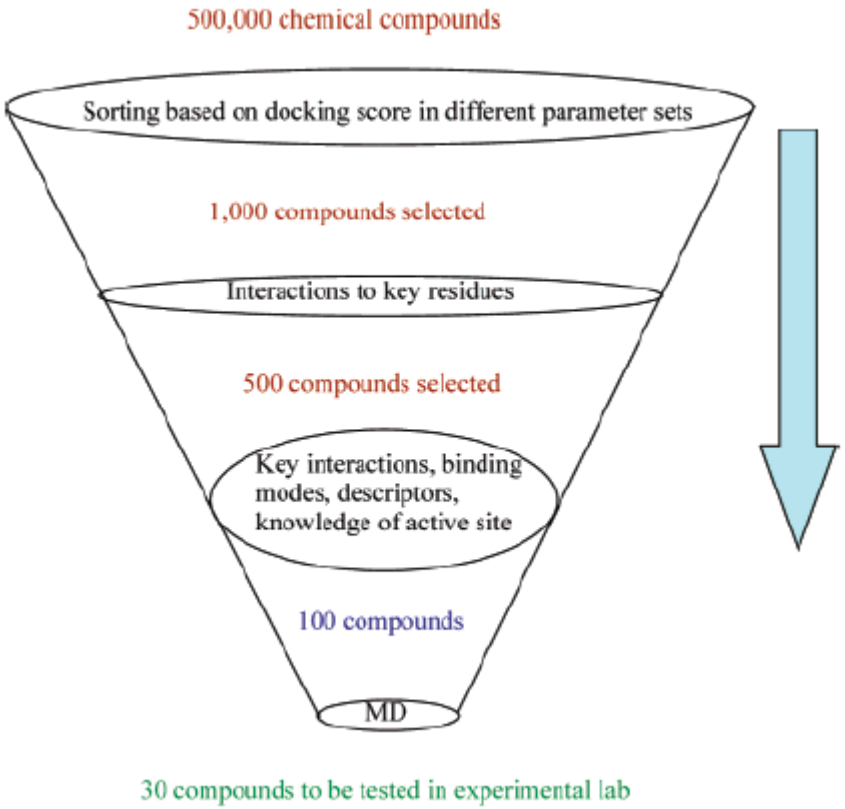


SYNERGY PARTNERS

Design of New Plasmepsin Inhibitors: A Virtual High Throughput Screening Approach on the EGEE Grid

Vinod Kasam,*,†,‡ Marc Zimmermann,*,† Astrid Maaβ,† Horst Schwichtenberg,† Antje Wolf,†
Nicolas Jacq,‡ Vincent Breton,‡ and Martin Hofmann-Apitius†

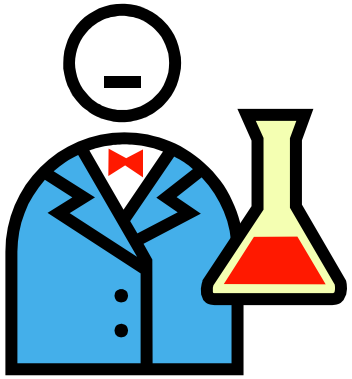
Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), 53754 Sankt Augustin, Germany, and LPC Clermont-Ferrand, Campus des Cézeaux, 63177 Aubière Cedex, France

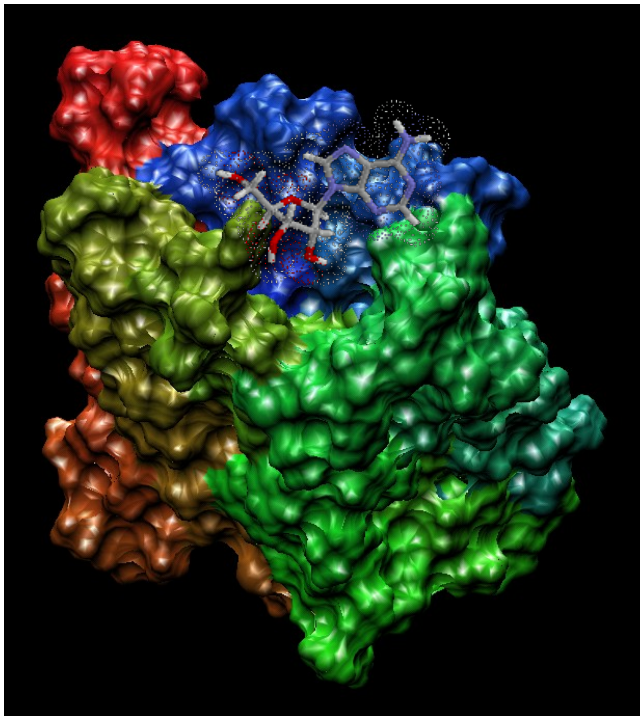


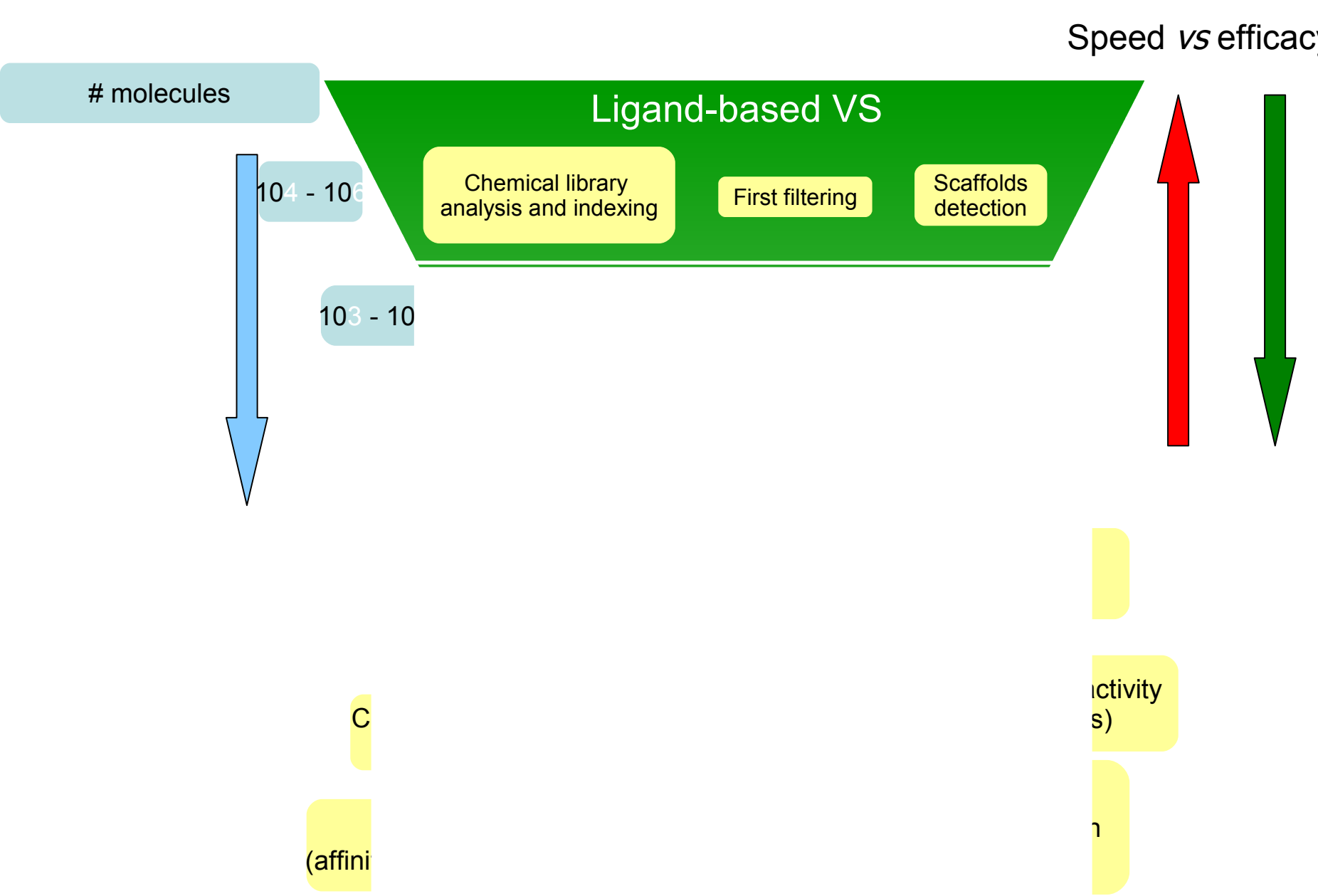
2005, an example of VS on the search for novel antimalarials

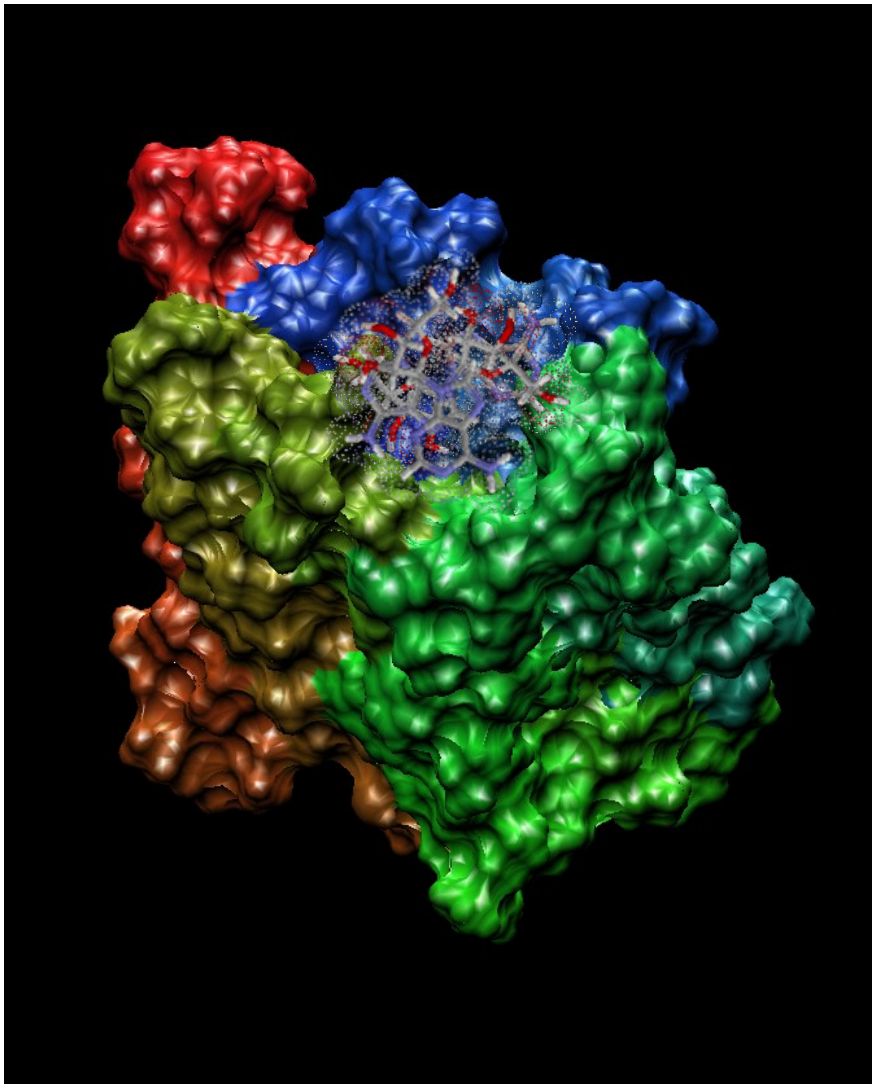
Concepts and computational techniques

- Protein-Ligand docking
- Binding free energy predictions
- Kinetic assays






Speed *vs* efficacy
molecules
Ligand-based VS
Chemical library analysis and indexing
First filtering
Scaffolds detection
10⁴ - 10⁶
10³ - 10
C
(affini
ctivity
s)
n



PROTEIN-LIGAND DOCKING

HOW MANY METHODS EXIST?

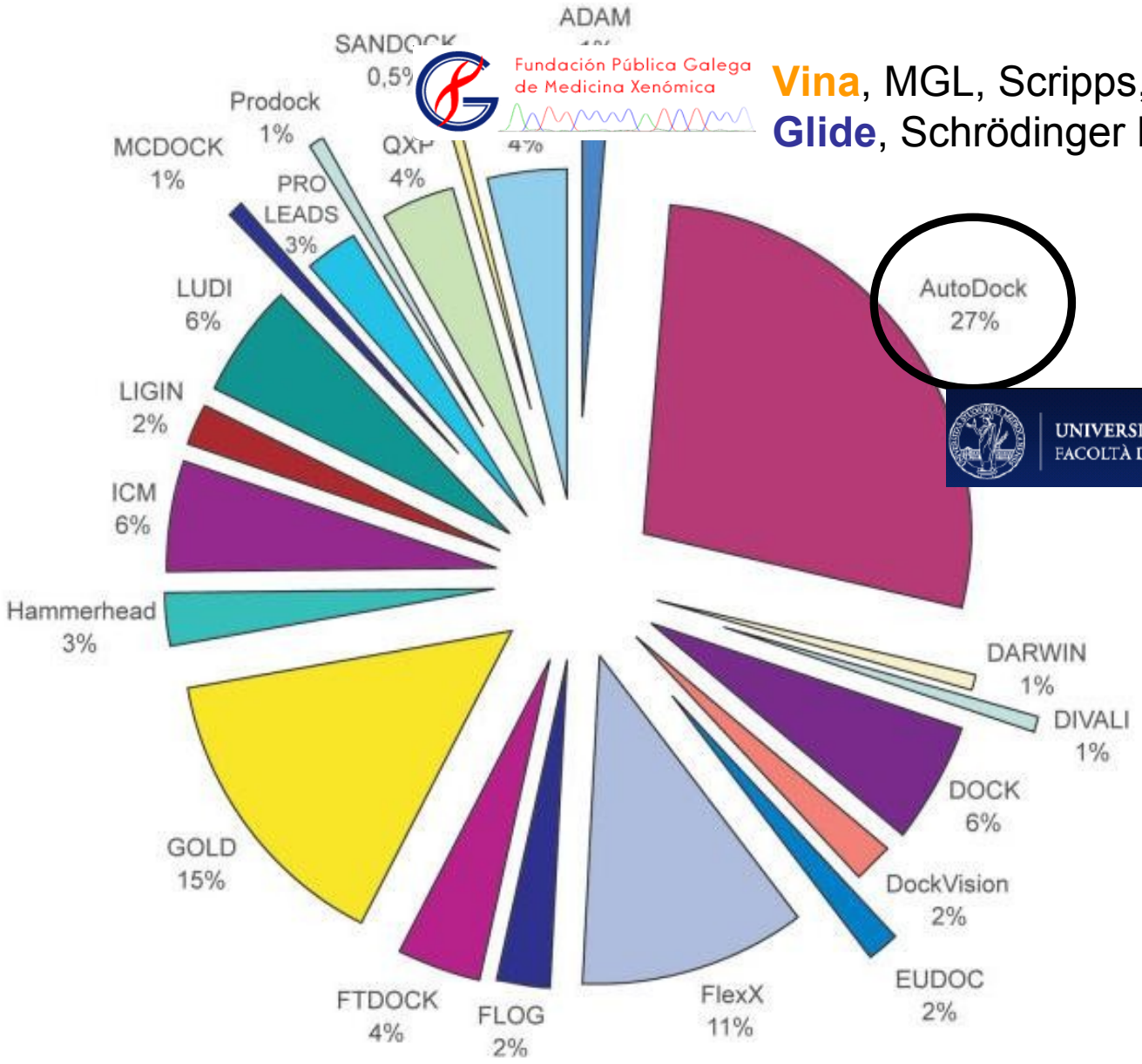
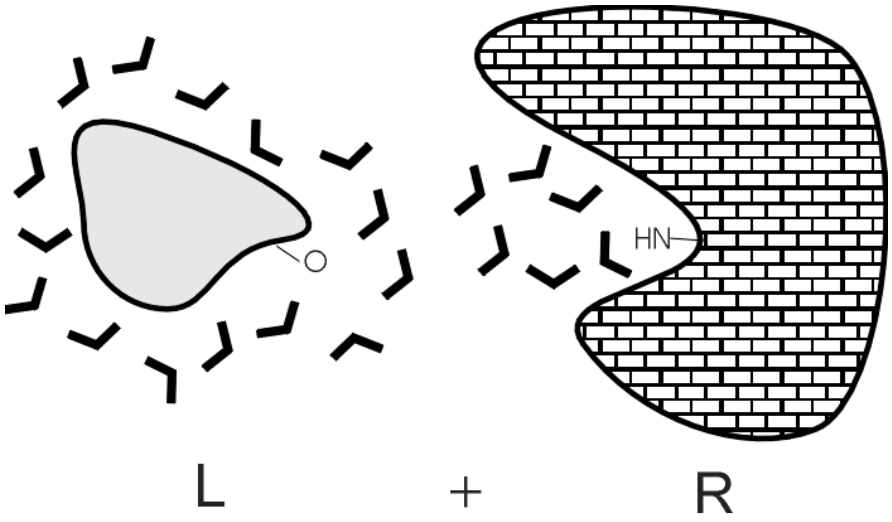


Fig. 2. Docking software—number of citations for some of the most common docking programs, analyzed from ISI Web of Science (2005) considering any of the original references as indicated in Table III.

Fundación Pública Galega de Medicina Xenómica

Vina, MGL, Scripps, CA (2007)
Glide, Schrödinger Ltd (2004)

UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI FARMACIA



$$K_d = \frac{[R][L]}{[RL]}$$

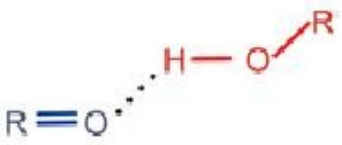
$$K_d = \exp\left(\frac{\Delta G_{binding}}{RT}\right)$$

$$\Delta G_{binding} = \Delta H - T\Delta S$$

Molecular interactions and Scoring functions

$$\Delta G_{binding} = \Delta H - T\Delta S$$

Tabla 1: Interacciones intermoleculares entre fármaco y receptor.

Tipo de interacción	*Geometría óptima*	*Ejemplo*
Electrostática o iónica	distancia: 2.8 Å	$=NH_2^+ \cdots {}^-OOC-$
Puente de hidrógeno	distancia: 2.7-3.1 Å ángulo: 120-180º	
Van der Waals	distancia: 3-4 Å	$CH_3 \cdots H_3C-$
Cambios entrópicos	-	Interacción hidrofóbica Energía conformacional

Scoring functions

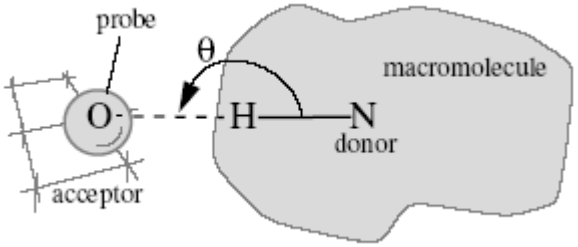
Empirical Scoring functions

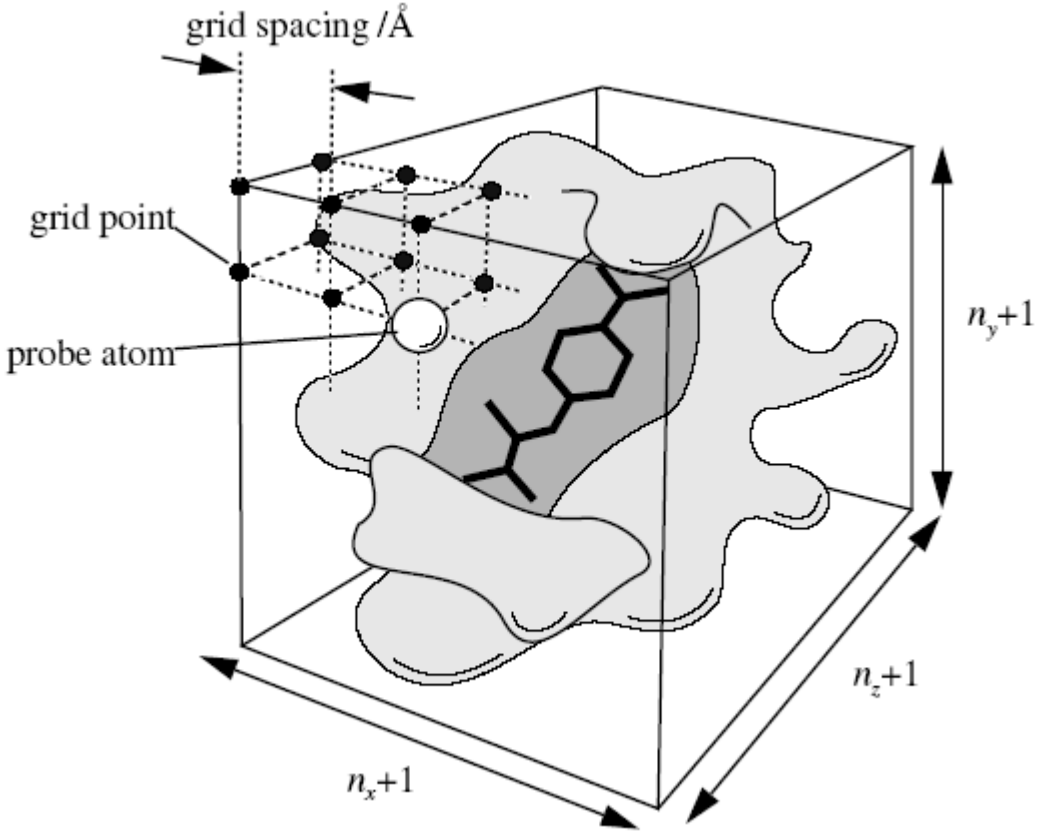
`Chemscore (Glidescore), Vina`

$$\Delta G_{binding} = \Delta G_{H-bond} + \Delta G_{metal} + \Delta G_{lipophilic} + \\ + \Delta G_{deformation} + \Delta G_0.$$

Coefficients obtained by multiple linear regression
Both Glide and Vina have been well trained

1. Ligand atom types? C, A, OA, ND (…)
2. Generate a grid map for each atom type (map.C, …)
3. Sample ligand conformations on the generated grid maps





Scientists Against Malaria

Virtual Organisation for Drug Discovery

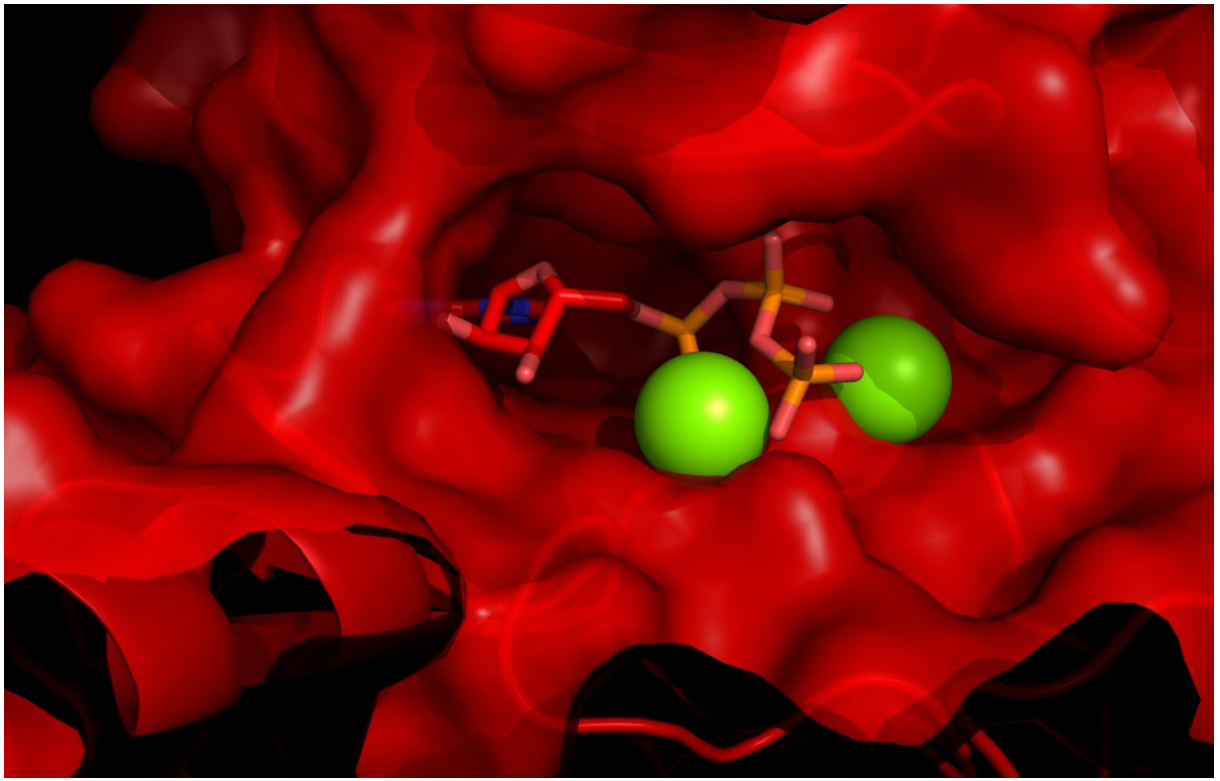
1. Target selection
2. Chemical library construction
3. Virtual screening (I): Molecular docking
4. Consensus scoring

Scientists Against Malaria

Virtual Organisation for Drug Discovery

1. Target selection
1. Chemical library construction
2. Virtual screening (I): Molecular docking
3. Consensus scoring

- Target selection: P. *falciparum* Kinase.
  - Reliable homology modeling,
  - Available for enzymatic assays



Structure of Pf Kinase selected as the protein target.
Binding site defined by ATP(sticks) / Mg+2 (spheres)

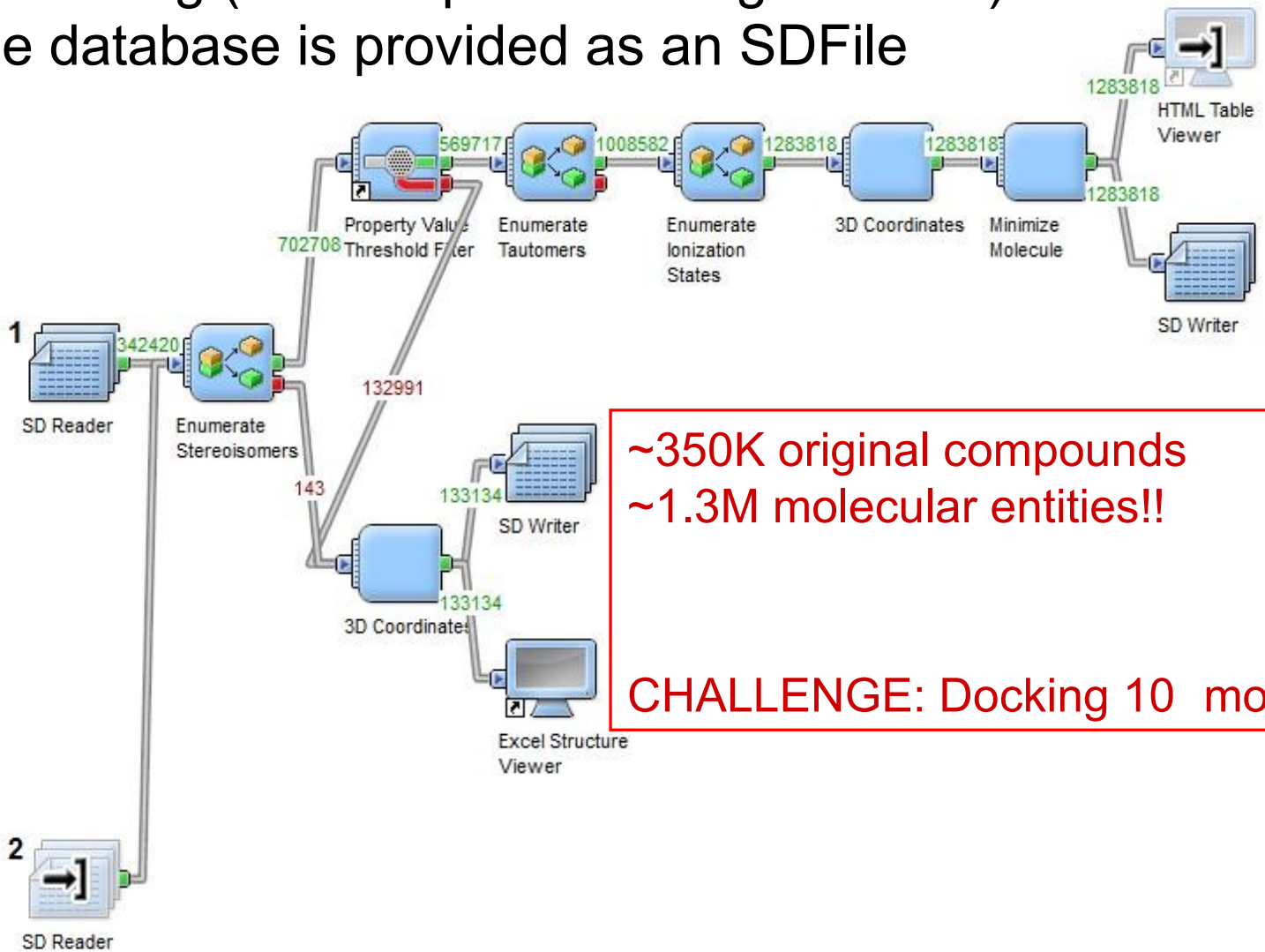
Scientists Against Malaria

Virtual Organisation for Drug Discovery

1. Target selection
1. Chemical library construction
1. Virtual screening (I): Molecular docking
2. Consensus scoring

The Chemical Databse in U. of Cincinatty

- Pipeline Pilot. Generation of all possible *xxx*omers
- No filtering (look for pharmacological tools)
- The database is provided as an SDFile



~350K original compounds
~1.3M molecular entities!!

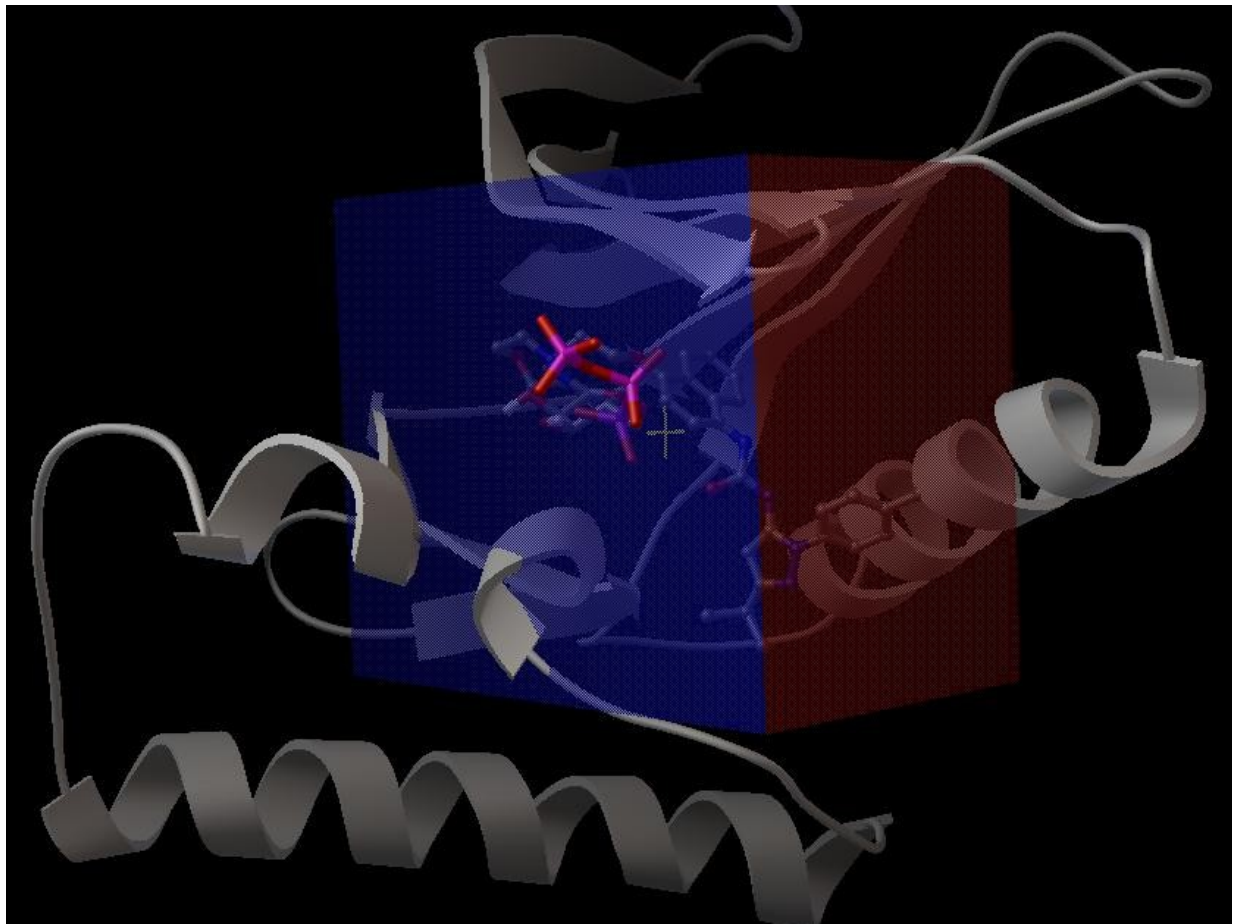
CHALLENGE: Docking 10 molecules

Scientists Against Malaria

Virtual Organisation for Drug Discovery

1. Target selection
2. Chemical library construction
1. Virtual screening (I): Molecular docking
1. Consensus scoring

- ADT (Autodock Tools).
  - Standard Kollman charges.
  - United atom model (only polar Hs).
  - Grid box: using as a reference the ATP binding site.
  - Generous search Space.



Vina.conf

Grid Space: 0.375Å
Grid points: 60x50x36
9 dockings / molecule

Database Preparation

SDFile (3D)

UC code
InChI strings

SDFile
<code>
UCxxxxxxx
<InChI>

Openbabel
(Hs add)

Mol2 File (1.3M entries, 4Gb)

scripting
(split)

25073 directories
50 single mol2 each

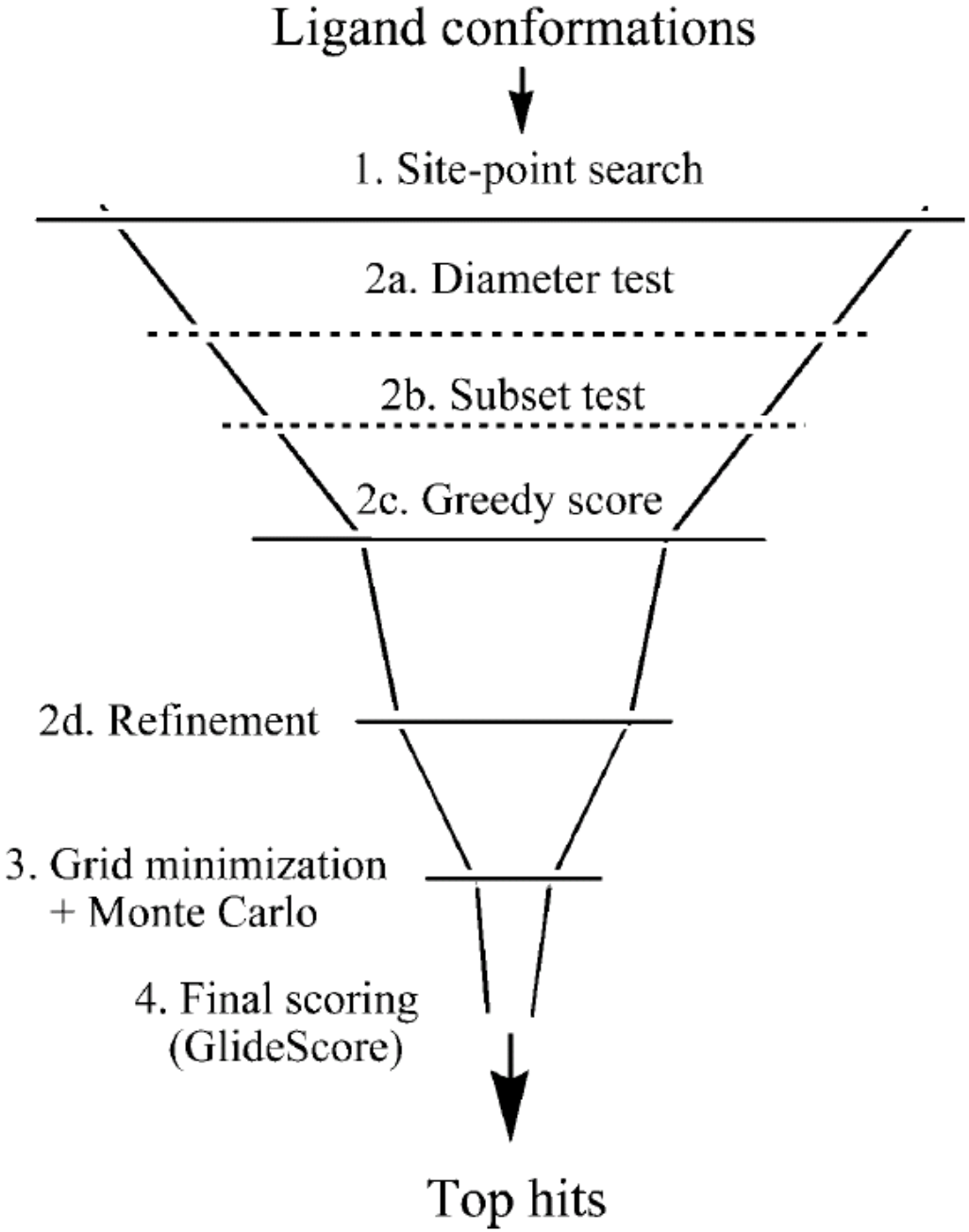
ADT
(mol2 to pdbqt)

25073 directories
50 pdbqt each
50 “vina.conf” each

Ready to the cloud!!

- 14 days computation time
  - Most time spent on huge molecules
- Analysis of the results (bash & perl scripts)
  - 2650 molecules failed docking: presence of metals or other non-drug like atoms (Si, B, …)
  - All compounds are retained in the best docking pose (out of 9 independent runs per ligand).
  - All have negative ΔG values.
  - The number of compounds filtered depend on the threshold considered:

Threshold	# Compounds	% of Database
ΔG <-12 kcal/mol	1,726	0.13
ΔG <-11 kcal/mol	7,909	0.62
ΔG <-10 kcal/mol	43,547	3.39
ΔG <-9 kcal/mol	213,324	16.62
ΔG <-8 kcal/mol	596,134	46.43
ΔG <-7 kcal/mol	984,646	76.70



Generate ligands' conformations
(once, local computation)

Initial shape-based docking
(1st filter)

Fine-tune docking (GlideScore)

Molecular Mechanics (OPLS-AA)
Monte-Carlo Sampling

Scoring, ranking (Glide Score)

- Virtual Screening Workflow (VSW). Consists on 3 steps:
  - **Glide HTVS** (High Throughput Virtual Screening). Then recover best 10% compounds (only if negative scoring poses)
  - **Glide-SP** (single precision). Then recover best 10% compounds (only if negative scoring poses)
  - **Glide-XP** (extra precision). Then recover best 10% compounds (only if negative scoring poses). Impact/OPLS minimization+rescoring.

The method has pros and cons:

↑ Fully automated

↑ 3 rounds of docking, increasing quality parameters

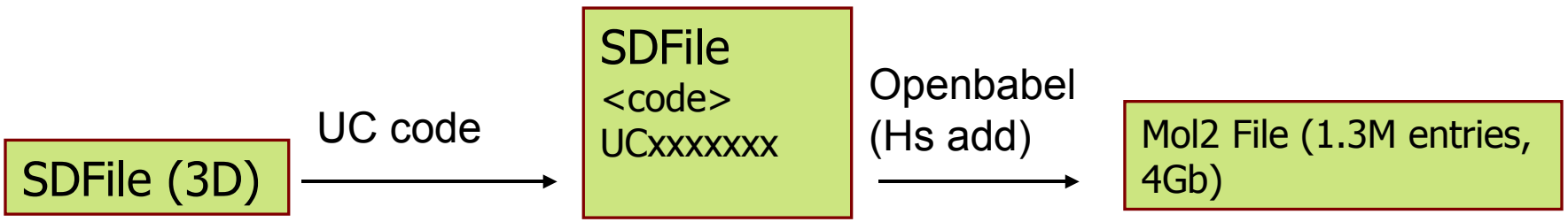
↑ Easiness of analysis. Conversion to different output formats (SDF, csv text files, graphical analysis with Maestro)

↓ Black box: not easy to recover intermediate information

- Protein Preparation: Maestro, Protein Preparation wizard.
  - Optimization of H-bond network
  - His/Gln/Asn tautomers, His/Gln flipping,
  - Small protein minimization.

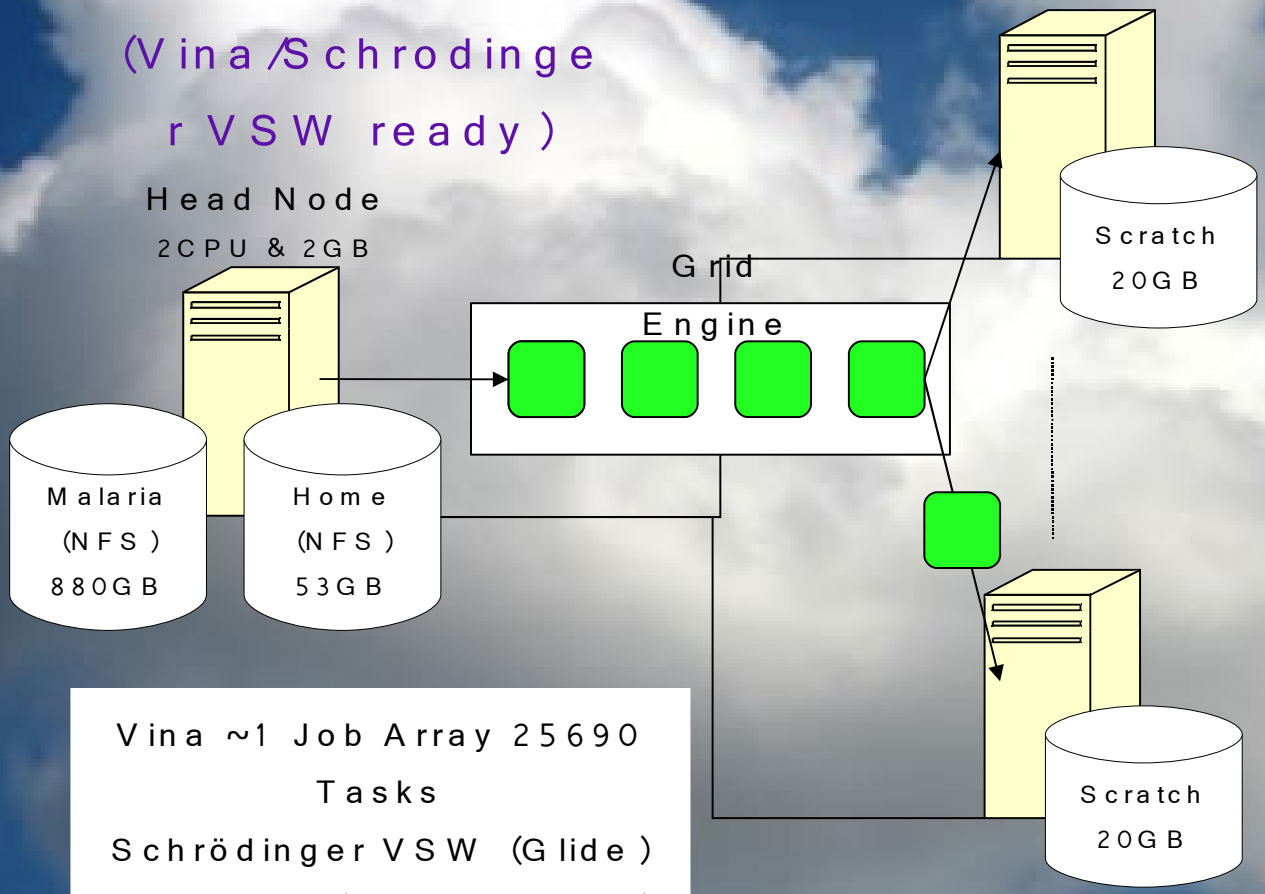
- Glide box is generated on the basis of the ligand present (in this case ATP).

Simpler ligand database preparation!



- Resources:
  - 500 tokens **granted from Schrödinger** (started 20/09/2010)
  - 64 processors of a grid distributed cloud of processors (Collaboration between CESGA and CYL HPC centers, Spain)
  - 4 days computing:
    - 2 h processing the library
    - 17 h Glide-HTVS (~1.3M compounds)
    - 10 h Glide-SP (~130K compounds)
  - 6.4 Gb information (150% of initial database)
  - 2 days of data transfer and analysis

Virtual Cluster
(Vina/Schrodinger VSW ready)



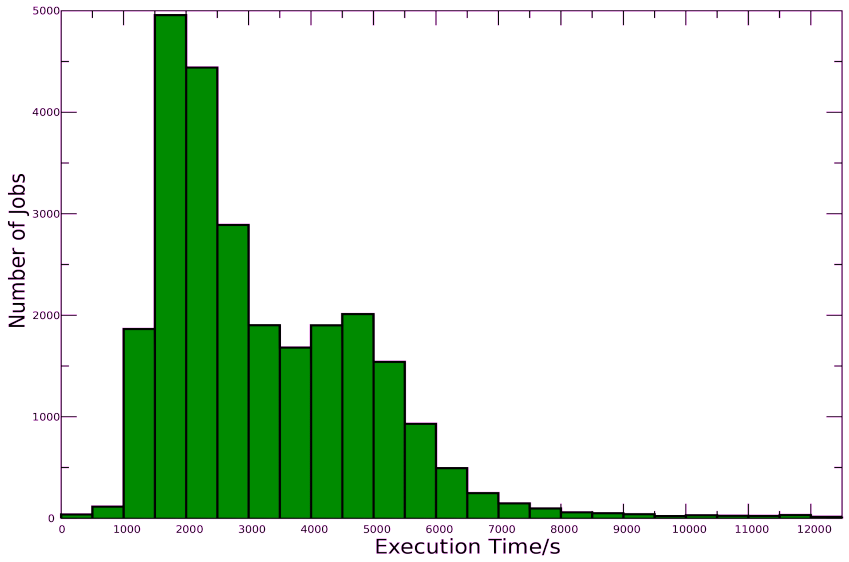
Vina ~1 Job Array 25690 Tasks
Schrödinger VSW (Glide)
~191 jobs (HTVS-SP-XP)

	Cores	Total Execution Time/s	Jobs	Average Job execution time/s	Efficiency (%)
VINA	322	1214530	25690	3412	22.4
VSW	64	331016	191	96390	86.9

VSW already has a efficient job manager

Vina: 131 jobs exceed 12500 s.

Some jobs reach near 700000 s.



Vina support SMP parallelization

+

Efficient job grouped algorithm is needed

↓

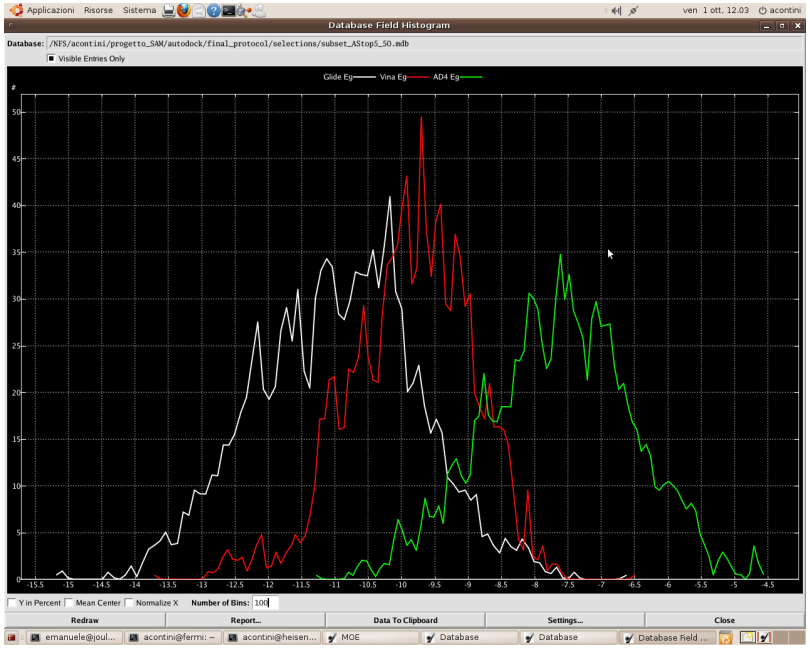
Efficient vina job manager to be developed

Scoring functions

- AD4: Statistical, force-field-based, with entropy and solvation energy terms (Huey, *J Comp Chem*, 2007).
- Vina: Statistical scoring function (Trott, *J Comp Chem*, 2010)
- Glide: Empirical + force-field-based (Friesner, *J Med Chem*, 2004)

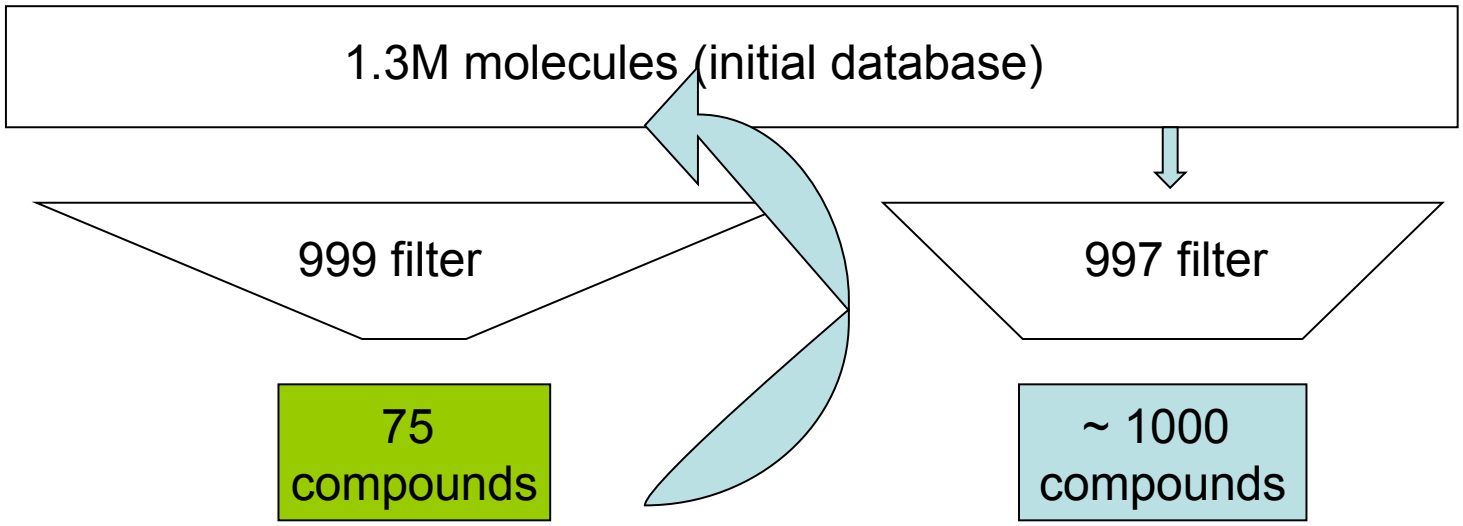
Filtering rules (I): Simple rules

- The 999 rule: ALL the three docking programs must provide $\Delta G \leq -9$ kcal/mol
  - VERY RESTRICTIVE! Selects 102 molecules, **75 unique compounds**
- The 997 rule: $\Delta G_{Glide} \leq -9$ kcal/mol + $\Delta G_{Vina} \leq -9$ kcal/mol + $\Delta G_{AD4} \leq -7$kcal/mol
  - SOFT rule, but still needs that all the 3 docking programs agree. 2442 molecules, but **996 unique compounds**



Distribution of scores (histogram of frequencies) for the different docking programs: **Glide**, **Vina** and **AD4**

Experimental Screening Results



1_{st} biological assay (under development)

Looking forward!

Scientists Against Malaria

Virtual Organisation for Drug Discovery

DouglasConnect — Roman Affentranger, Barry Hardy

EPFL École Polytechnique Fédérale de Lausanne — Christian Doerig, Matteo Dal Peraro

CERF by Rescentris, Inc. — Jeff Spitzner

University of Cincinnati Drug Discovery Center — Ruben Papoian, William Seibel, Sandra Nelson

Università degli Studi di Milano Facoltà di Farmacia — Alessandro Contini

Monash University — Andrew Wilks, Isabelle Lucet

Fundación Pública Galega de Medicina Xenómica — H. G. de Teran

PHARMATROPE — Jeffrey Wiseman

inte:ligand Your partner for in-silico drug discovery. — Sharon Bryant