Computational Procedures Applied to Medicinal Chemistry

SAPIENZA
UNIVERSITÀ DI ROMA

Tutor: Prof. Rino Ragno

Dottorando: Alberto De Petris

Overview

✓β-Secretase: Tridimensional Quantitative structure-activity relationship (3-D QSARs) and Assessment of Predictive Ability

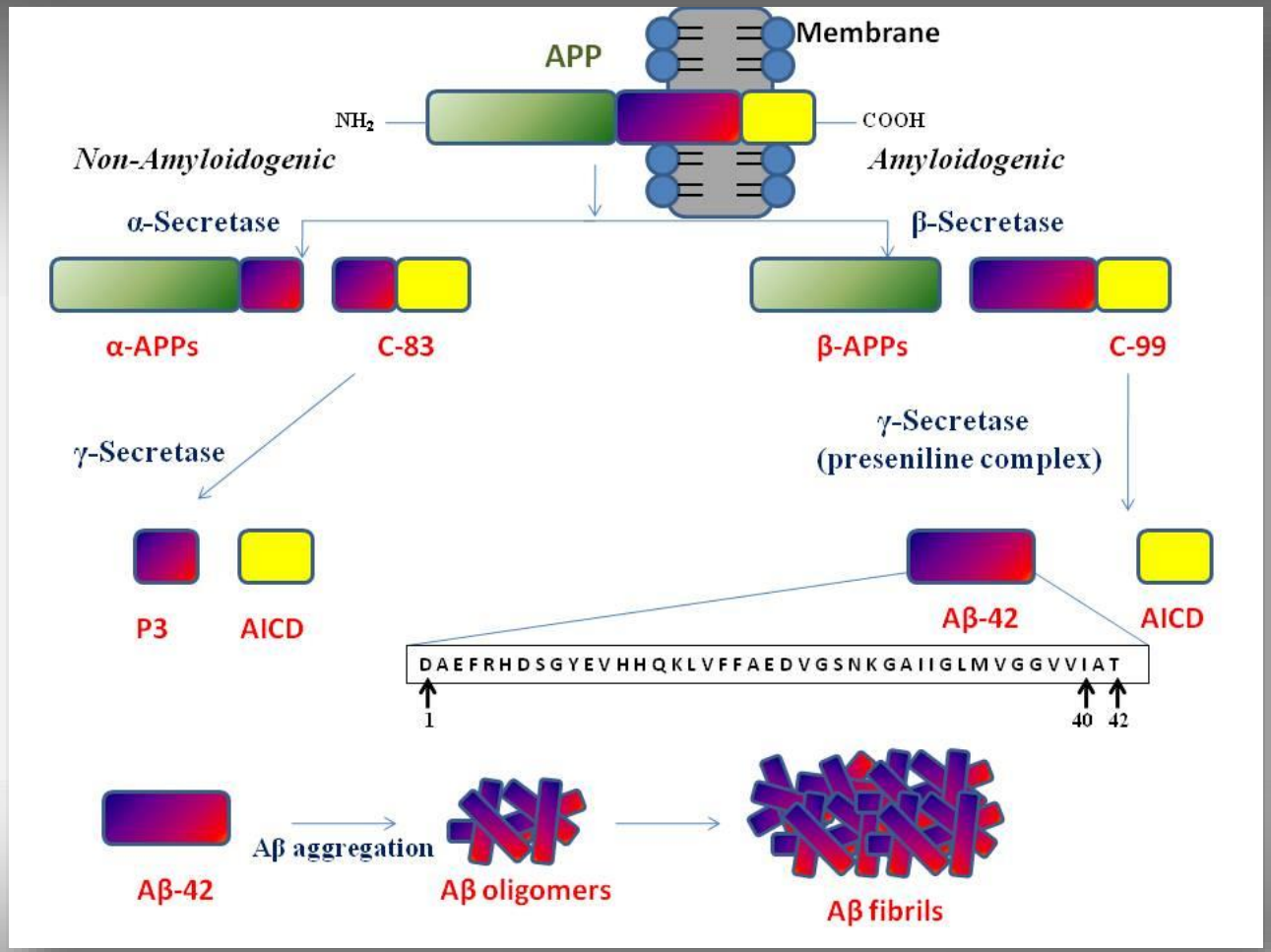
✓Density Functional Theory Based Ab Initio Molecular Dynamics Using the Car-Parrinello Approach

✓A structure based virtual screening against Ornithine carbamoyltransferase 2, phaseolotoxin-insensitive from Pseudomonas Syringae pv Actinidiae

β-Secretase: Tridimensional Quantitative structure-activity relationship (3-D QSARs) and Assessment of Predictive Ability

SAPIENZA
UNIVERSITÀ DI ROMA

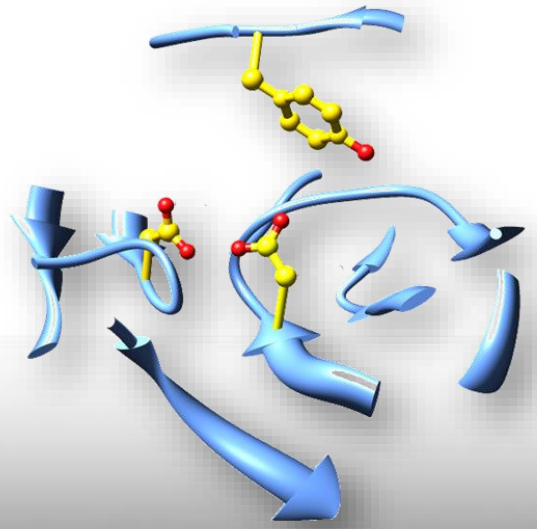
β site of APP cleaving enzyme



- **Progressive loss of memory**
- **Loss of cognitive ability**
- **Aphasia**

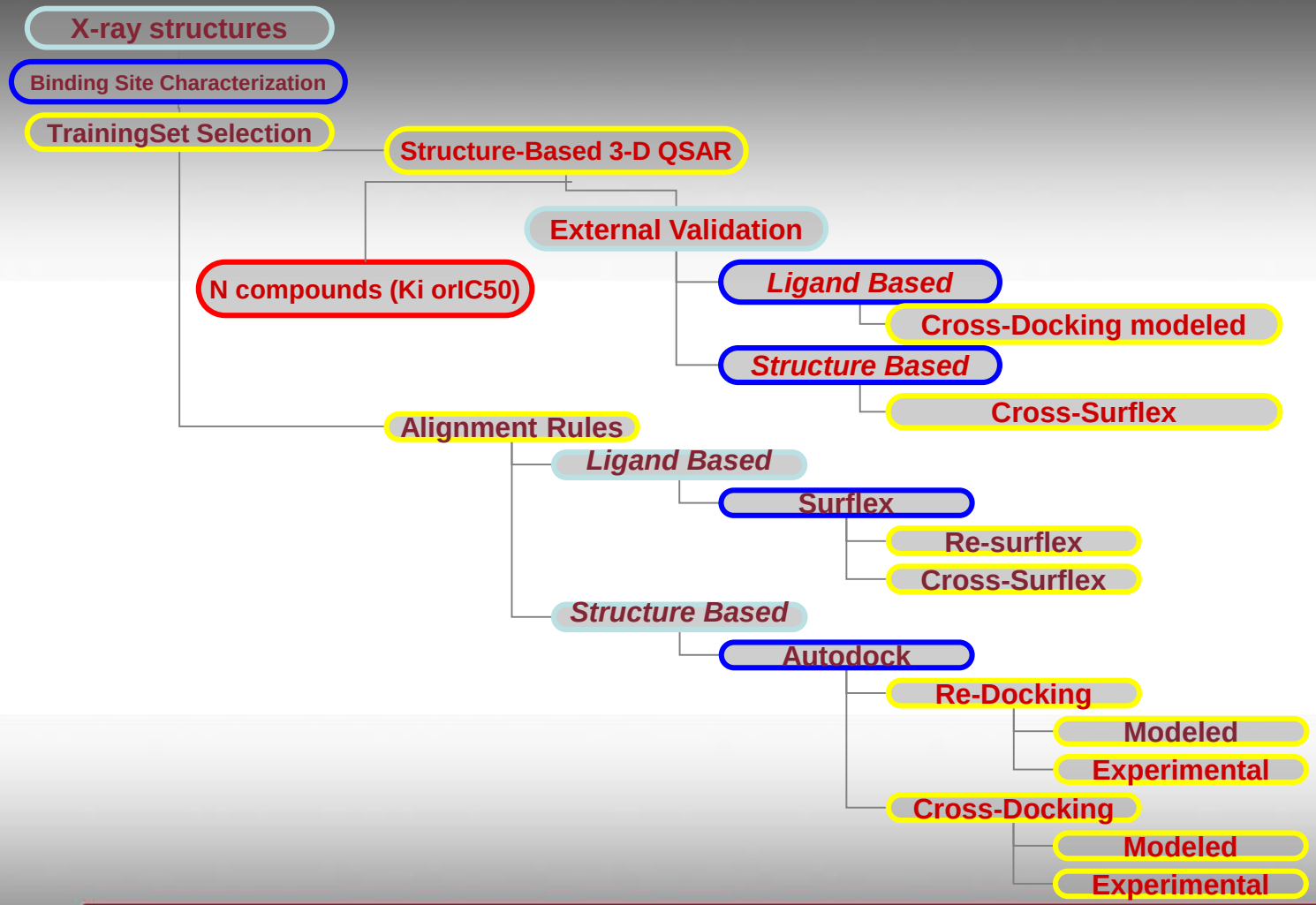
1. Park, S.-Y. *Archives Pharmacal Research* 2010, *33*, 1589-1609.
2. Taleb H. Al-Tel, R. A. A.-Q., Marco F. Schmidt *Journal of Medicinal Chemistry* 2009, 6484-6488.
3. P. van Tijn, W. K., M. W. Marlatt *Progress in Neurobiology* 2010, 1-16.

β site of APP cleaving enzyme



4. Lin Hong, G. k., Xinli Lin *Science* 2000, *290*, 150-153.
5. Rueeger H., Roundeau JM., Bioorg.Med Chem Lett., 2011, 21, 1942-7
6. Bursavich, M. G. R., D. H. *Journal of Medicinal Chemistry* 2002, 541-558.
7. Cooper, J. B. *Current Drug Target* 2002, 155-173.

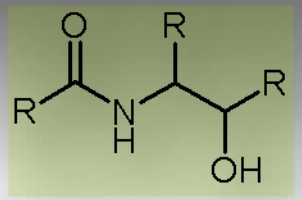
Workflow

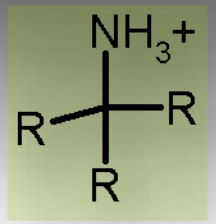

X-ray structures
Binding Site Characterization
TrainingSet Selection
Structure-Based 3-D QSAR
External Validation
N compounds (Ki orIC50)
Ligand Based
Cross-Docking modeled
Structure Based
Cross-Surflex
Alignment Rules
Ligand Based
Surflex
Re-surflex
Cross-Surflex
Structure Based
Autodock
Re-Docking
Modeled
Experimental
Cross-Docking
Modeled
Experimental

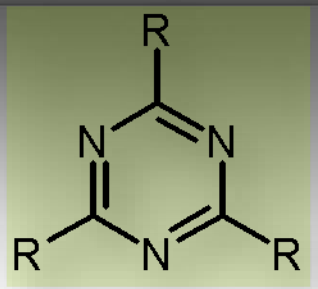
3-D QSAR Model

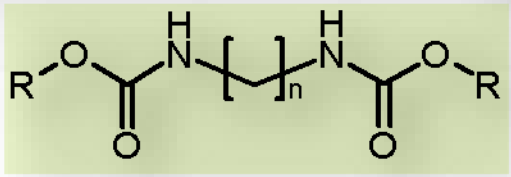
1. Training Set Selection
2. Molecular Models
3. Molecular Alignment
4. Molecular Interaction Field (AUTOGRID)
5. Statistical Processing of 3-D QSAR models (CRAN-R)
6. Results
7. Validation of 3-D QSAR model

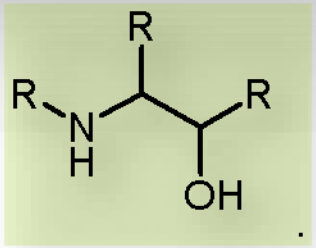
Training Set Selection

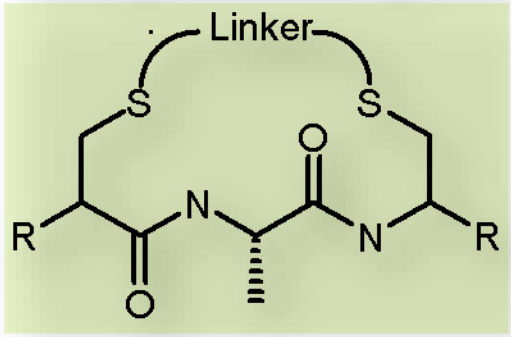




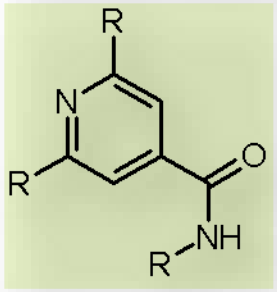


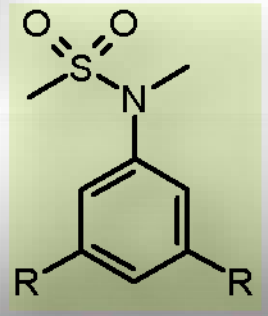


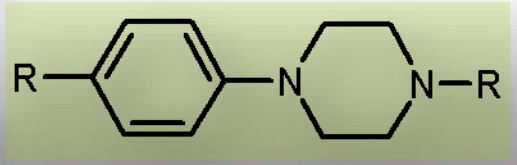


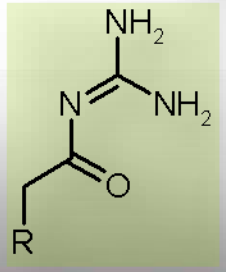


THE TRAINING SET







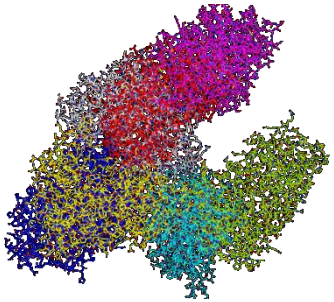


8. Berman H. M., W. J., Feng Z. *Nucleic Acids Res.* **2000**, 235-242.
9. Wenjin Yang, W. L., Yafan Lu *Journal of Medicinal Chemistry* **2006**, *49*, 839-842.
10. Miles Congreve, D. A., J. Alberto *Journal of Medicinal Chemistry* **2007**, 1124-1132.

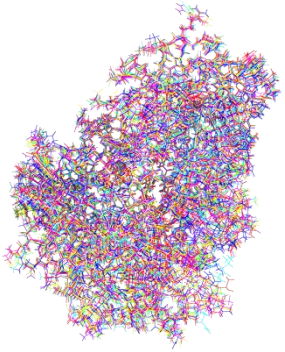
Molecular Models / Alignment

Minimization (AMBER9)

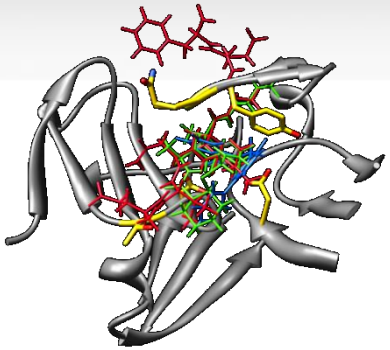
GEOMETRIC AND STRUCTURAL OPTIMIZATION



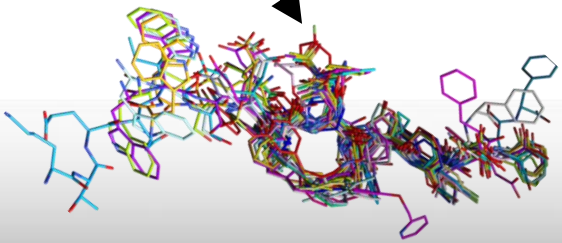
BACE-1/Inhibitors complexes from Protein Data Bank



Aligned Complexes



Binding Conformation

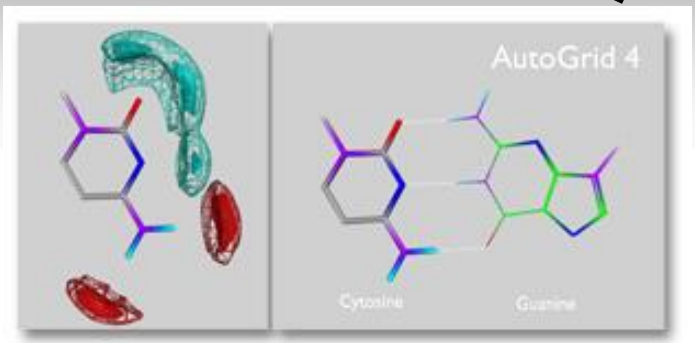


Structure Based Alignment

11. Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Huang, C. C.; Ferrin, T. E. *BMC Bioinformatics* 2006, *7*, 339.
12. Pearlman D.A., C. D. A. 1995.
13. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Simmerling, C.; Wang, B.; Woods, R. J. *J Comput Chem* 2005, *26*, 1668-88.

Molecular Interaction Field / 3-D QSAR Model

Visualization and Analysis



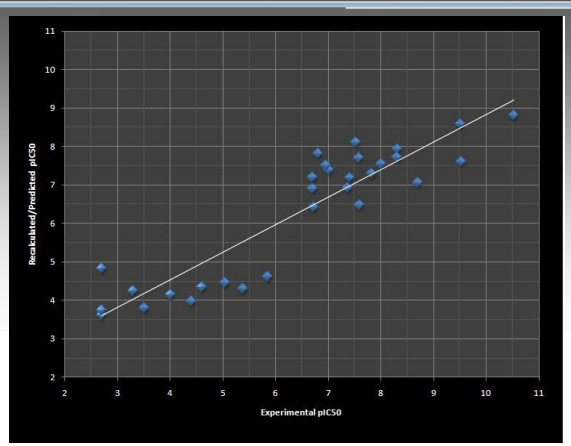
UCSF CHIMERA
an Extensible Molecular Modeling System

N	GRID probe	V	PC	r^2	q^2	$SDEP_{cv}$
30	A (Aromatic Carbon Atom)	68445	3	0.98	0.84(K5FCV)	0.87
30	OA(H Bond Donor)	68445	3	0.98	0.85(KF5CV)	0.88

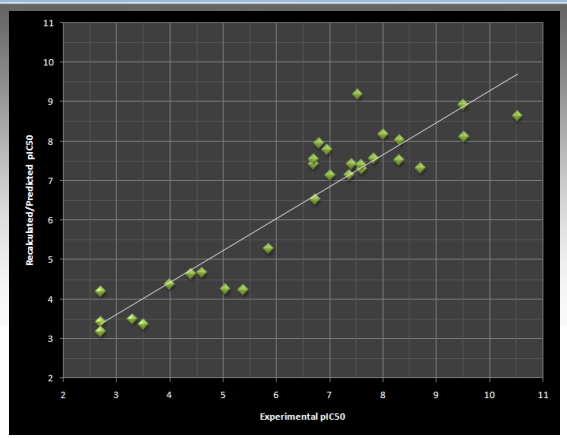
N= number of compounds in the training set; V=number of variables; PC= optimal number of principal components analysis; r^2= conventional square correlation coefficient; q^2= cross-validation correlation coefficient; SDEP= cross-validated standard error of predicting using the KFCV method; K5FCV= K-Fold (5 random groups) Cross-Validation Method

14. Autogrid4, The Sc ripps Resea rch Ins titute, Molecula r Graph ics Labora tory
15. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J Comput Chem* **2004**, *25*, 1605-12.
16. http://rm.mirror.garr.it/mirrors/CRAN/, The R Project for Statistical Computing

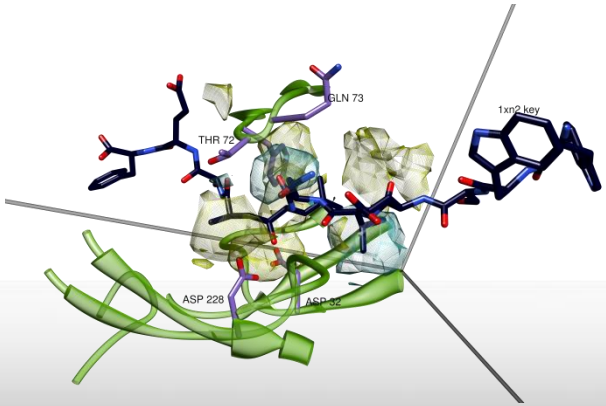
Model Interpretation



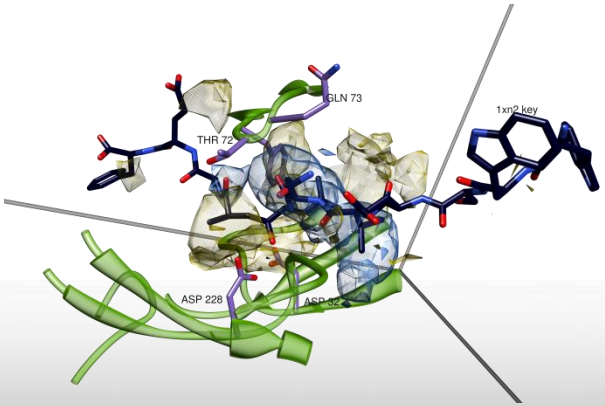
Experimental vs Predicted pIC50 Plot for A Probe



Experimental vs Predicted pIC50 Plot for OA Probe



Autogrid/R Coeff Contour Map (A Probe)



Autogrid/R Coeff Contour Map (OA Probe)

External Validation

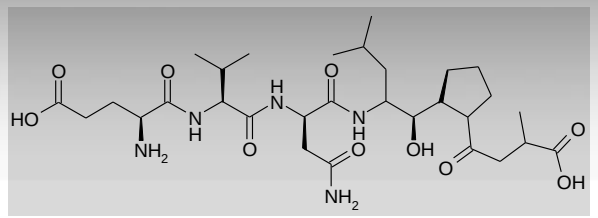
Along with the internal validation (**cross-validation**), the predictive capabilities of the model were estimated by the selection of a **test set** of 366 compounds, extracted from 30 different review and reporting several class of BACE-1 inhibitors not used in generating the model. The activities of compounds were predicted using the **AUTOGRID/R method**. The predicted biological activities are in good agreement with the linear correlation lines obtained from the training set, which confirm that the model can be used in estimating the activities of BACE-1 inhibitors sharing a similar molecular skeleton and binding mode. The test set contain the maximal **diversity** in terms of molecular structures and bioactivities, in order to reproduce the different chemical hallmarks of the training set.

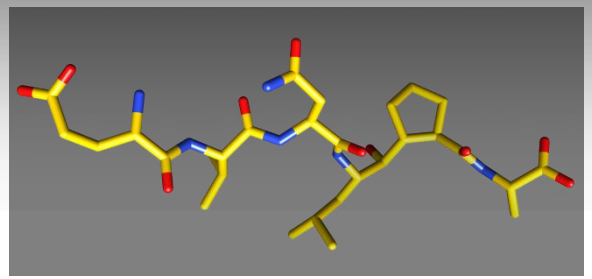
N	GRID probe	PC	SDEP
366	A (Aromatic Carbon Atom)	3	0.838
366	OA(H Bond Donor)	3	0.834

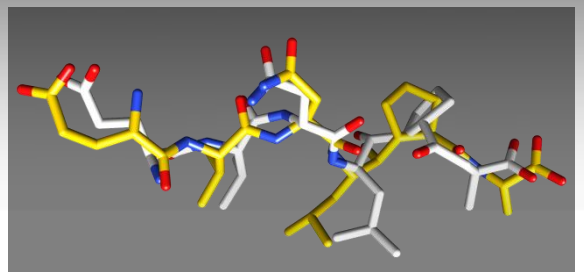
N= number of test sets; PC= optimal number of principal components analysis;
SDEP= cross-validated standard error of prediction;

Alignment Rules

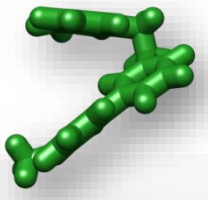
Ligand-Based Method

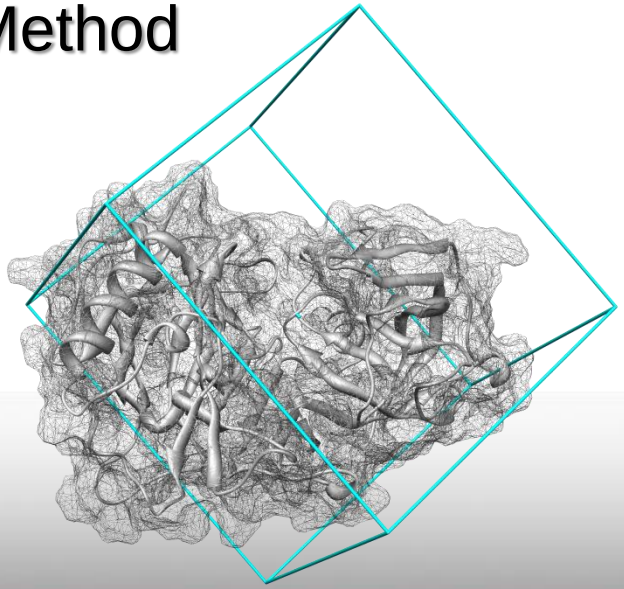


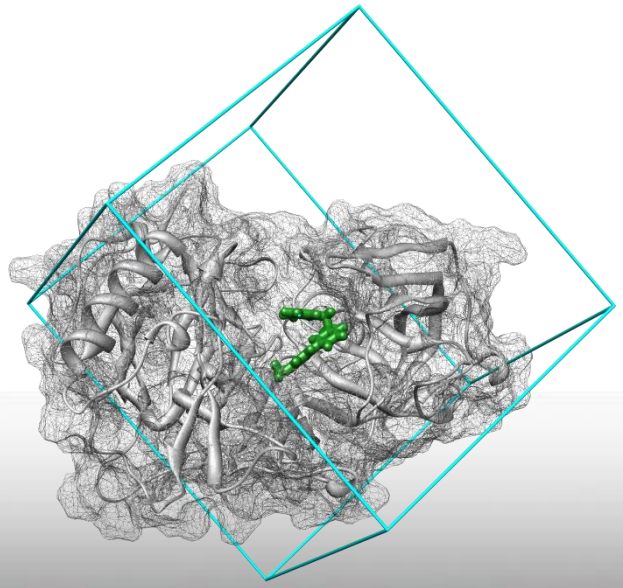




Structure-Based Method







17. Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem. 2004, 47, 947–961*
18. Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit. 1996, 9, 1–5*

CrossDocking Results

PDB	BD (rmsd)	BD (BE)	BC (rmsd)	BC (BE)	BF (rmsd)	BF (BE)
1fkn	2.59	-6.98	6.68	-3.37	2.45	-4.8
1m4h	1.16	-7.21	1.16	-9.87	1.16	-7.21
1tqf	2.55	-11.49	2.5	-9.62	2.13	-9.14
1w51	2.92	-9.91	2.92	-7.86	0.92	-8.99
1xn2	0.58	-13.46	0.58	-13.46	0.58	-13.46
1xn3	0.75	-9.01	0.75	-9.01	0.75	-9.01
1xs7	2.32	-8.5	4.57	-6.47	2.32	-8.5
1ym2	0.64	-11.55	0.64	-5.93	0.64	-11.55
1ym4	1.15	-11.7	2.96	-6.14	1.15	-11.7
2b8l	2.28	-10.87	3.23	-6.49	2.28	-5.06
2b8v	1.17	-13.15	1.17	-13.15	1.17	-13.15
2f3e	3.84	-10.29	3.84	-10.29	1.67	-6.63
2f3f	0.61	-13.15	3.24	-8.75	0.61	-13.15
2fdp	0.34	-17.26	4.92	-8.04	0.34	-17.26
2g94	2.21	-10.55	3.05	-7.87	1.83	-8.03
2hiz	2.81	-9.1	3.02	-6.26	2.16	-7.15
2hm1	3.14	-8.8	3.58	-7.41	1.78	-7.15
2iqg	0.69	-14.09	0.69	-6.78	0.69	-14.09
2irz	2.87	-12.33	5.02	-9.6	0.92	-10.75
2is0	2.99	-15.38	2.57	-10.66	2.37	-7.8
2of0	1.92	-8.39	1.92	-8.39	1.75	-6.69
2ohk	1.67	-9.87	2.09	-8.45	1.67	-7.68
2ohl	1.74	-5.98	1.74	-5.98	1.74	-5.98
2ohm	3.63	-6.98	2.42	-6.17	2.42	-6.17
2ohn	0.8	-8.01	0.8	-8.01	0.8	-8.01
2ohq	3.04	-8.54	3.04	-8.54	1.81	-8.35
2ohr	1.3	-8.18	1.3	-8.18	1.3	-8.18
2ohs	2.17	-8.31	2.17	-8.31	1.53	-7.94
2oht	2.14	-9.53	3.9	-8.31	2.14	-9.53
2ohu	4.57	-10.85	3.27	-9.91	2.03	-7.03

BD = Best Docked, BC = Best Cluster, BF = Best Fitted, BE = Binding Energy

: <2

: 2 < n < 3

: >3

The 3-D QSAR server:
http://www.3d-qsar.com/

www.3D-QSAR.com
THREE-DIMENSIONAL QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS SERVER

SAPIENZA UNIVERSITÀ DI ROMA

by www.RCMD.it

Log in or Sign Up

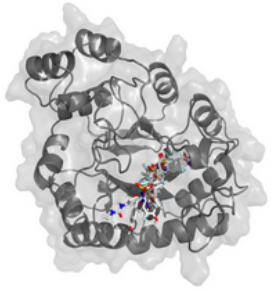
Home | Launch Predictor | Datasets | About | Change Log

A 3-D QSAR MODELS DATABASE for Virtual Screening

Substance libraries, composed of million of compounds, are tested in the pharmaceutical industry with robotic systems. Furthermore, the follow-up of the "hits" is often so expensive that essentially only large companies can use this method. Virtual screening helps to decide which compounds to screen, which libraries to synthesis and which compounds to purchase from an external supplier reducing the overall cost associated to the discovery and development of new drugs.

Main classes of virtual screening methods are:

- Similarity search (ligand-based virtual screening)
- Identify a common 3-D pharmacophore, then do a 3-D database search
- Train a machine learning technique
- Protein-ligand docking



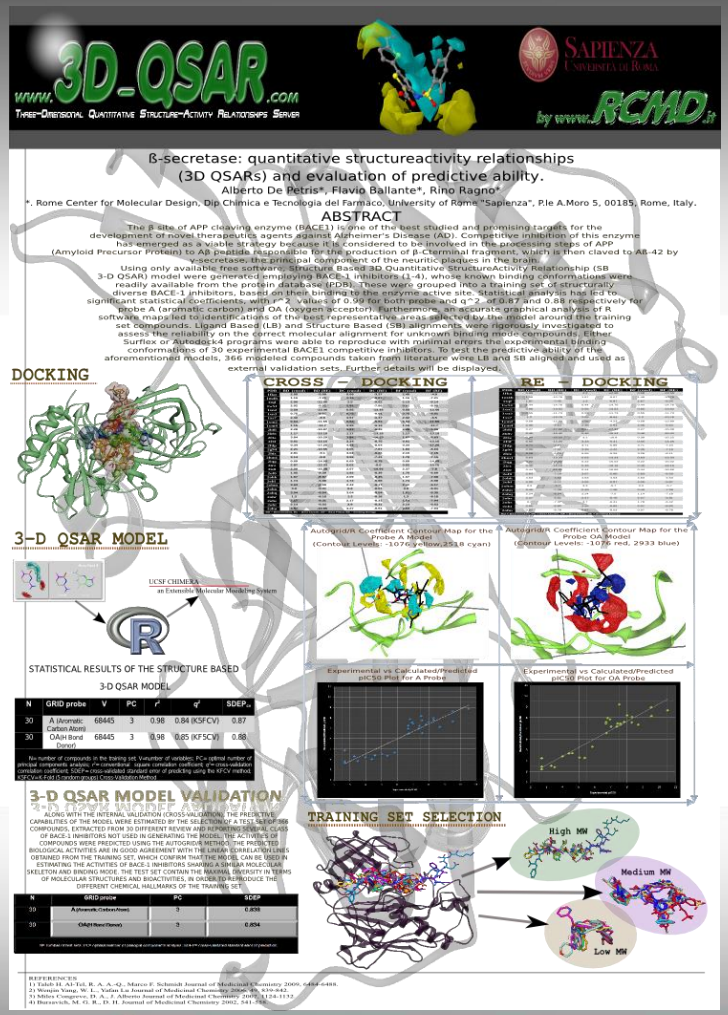
Most of these methods are computationally intensive and complex. 3-D QSAR methods are nowadays used widely in drug design, since they are computationally not demanding and afford fast generation of QSARs from which the biological activity of molecules can be predicted.

We have developed a software system for automatically score, rank and/or filter a set of structures using pre-built 3-D QSAR models.

A Web interface incorporating a molecular drawing interface enables the users to process their own molecules by drawing or uploading them to the server and selecting the target for the virtual screening and biological activity prediction.
Multiple targets screening are allowed, making the RCMD 3-D QSAR Server a very interesting tool for virtual high throughput screening.

The system consists of a library of pre-built 3-D QSAR models. This database is constantly updated with published or newly developed 3-D QSAR models.

The 3-D QSAR server:
http://www.3d-qsar.com/



University of Siena

VIII EWDD

Eighth European Workshop in Drug Design

May 22 - 28, 2011 - Certosa di Pontignano, Siena (Italy)

Density Functional Theory Based Ab Initio Molecular Dynamics Using the Car-Parrinello Approach

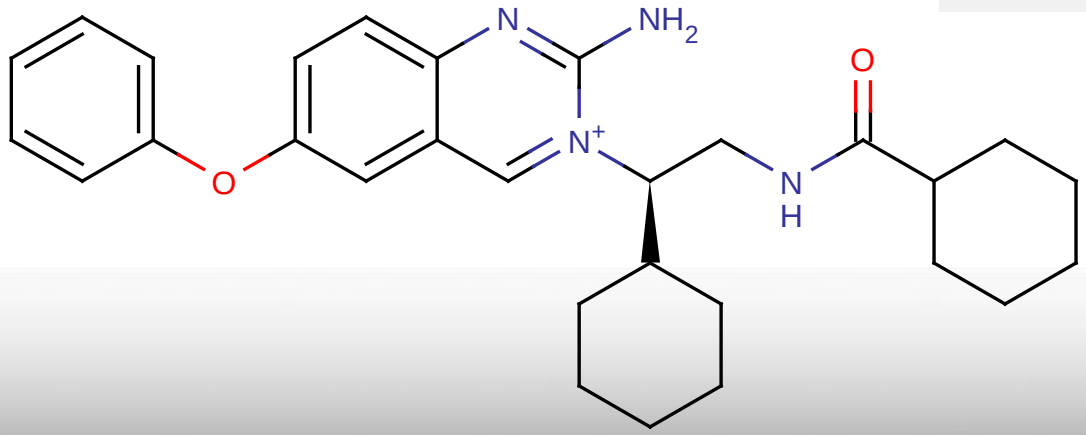
SAPIENZA
UNIVERSITÀ DI ROMA

Density Functional Theory Based Ab Initio MD Using the Car-Parrinello Approach

- IUPAC: N-[(2R)-2-(2-amino-6-phenoxyquinazolin-3-ium-3-yl)-2-cyclohexylethyl] cyclohexanecarboxamide

- Binding Bioassay: Inhibition of BACE-1

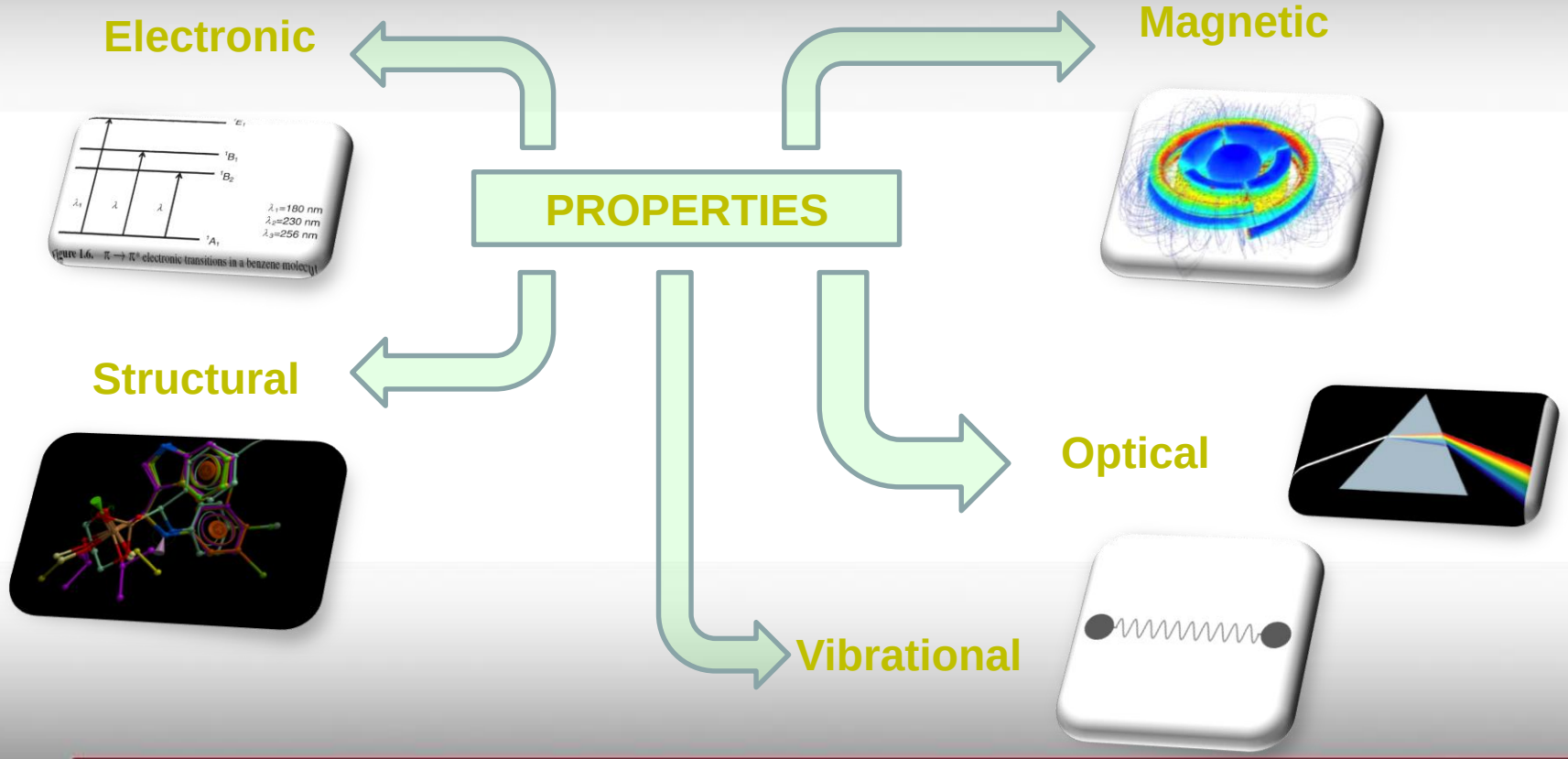
Compound ID	44631815
Molecular Weight	473.62968 [g/mol]
Molecular Formula	$C_{29}H_{37}N_4O_2^+$
XLogP3-AA	6.6
H-Bond Donor	2
H-Bond Acceptor	4



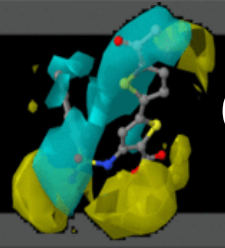
Density Functional Theory (CP2K)

The ground state electronic density ρ(r) determines uniquely all possible properties of an electronic system:

ρ(*r*) ⇒ Properties *P* (e.g. conductance), i.e. *P* ≡ *P*[ρ(*r*)]



22. Hohenberg-Kohn Theorem: Phys. Rev. 136, B864 (1964)



CP2K and QUICKSTEP METHOD

Calculation of properties:

GPW and GAPW methods

- GPW: Gaussian and Plane Wave basis sets
- Quickstep: DFT code within CP2K (cp2k.berlios.de)

Plane wave vs gaussian basis sets: plane waves pros and cons

Advantages
- independent of the nuclei position (good for forces)
- no BSSE
- one parameter controls the basis set size
- orthogonal
- numerical efficiency through use of FFT

Disadvantages
- large number of basis set elements needed
- **Necessary use of pseudo-potentials**
- loss of chemical insight

Border line
- fill the whole simulation box

Plane wave vs gaussian basis sets: gaussians pros and cons

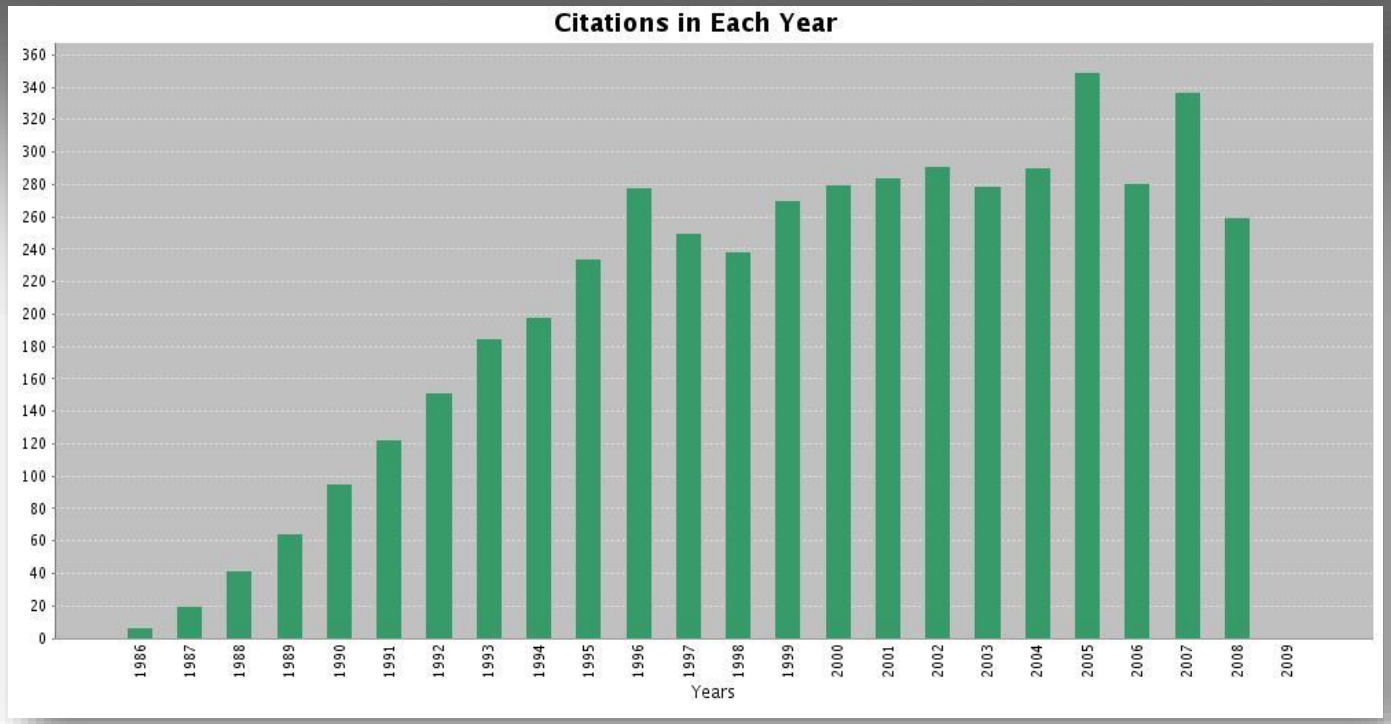
Advantages
- Good already for small basis set sizes
- correspond to chemical insight
- Computationally efficient (multi-centre integrals)
- Possibility to perform all-electrons calculations

Disadvantages
- non-orthogonal
- atomic position dependent (Pulay forces)
- Basis set superposition error (BSSE)
- Systematic improvment not straightforward
- linear dependencies, over-completeness
- wrong asymptotic behaviour

Border line
- no implicit periodicity
- can be tuned for each application

19. Joost VandeVondele et al. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. Computer Physics Communications, 167:103 – 128, 2005.
20. A hybrid Gaussian and plane wave density functional scheme G. Lippert, J. Hutter, and M. Parrinello; Mol. Phys. 92, 477 (1997)
21. The Gaussian and Augmented-Plane-Wave Density Functional Method G. Lippert, J. Hutter,and M. Parrinello; Theor.Chem Accounts 103, 124 (1999)
22. M. Krack and M. Parrinello, Phys. Chem. Chem. Phys., 2, 2105 (2000)

CP2K



- **Car-Parrinello is a method for performing molecular dynamics with forces obtained from electronic structure calculations performed “on the fly” as the simulation proceeds. This is known as *ab initio* molecular dynamics (AIMD).**
- **As a result, AIMD calculations are considerably more expensive than force-field calculations, which only involve evaluation of simple functions of position.**

23. R. Car and M. Parrinello, *Phys. Rev. Lett*. **55**, 2471 (1985)

CP2K input file

Section path: CP2K_INPUT

Subsections

I. ATOM
II. DEBUG
III. EXT_RESTART
IV. FARMING
V. FORCE_EVAL
VI. GLOBAL
VII. MOTION
VIII. MULTIPLE_FORCE_EVALS
IX. OPTIMIZE_INPUT
X. TEST
XI. VIBRATIONAL_ANALYSIS

```
@SET CutOff    350
@SET dt        0.25

&GLOBAL
  PROJECT secretase_2wjo
  RUN_TYPE ENERGY
  PRINT_LEVEL HIGH
&END GLOBAL

&FORCE_EVAL
  METHOD Quickstep
  &DFT
    BASIS_SET_FILE_NAME /RCMD/opt/linux64/CP2K/cp2k/tests/QS/BASIS_MOLOPT
    POTENTIAL_FILE_NAME /RCMD/opt/linux64/CP2K/cp2k/tests/QS/POTENTIAL
    &MGRID
      CUTOFF ${CutOff}
    &END MGRID
    &QS
      EPS_DEFAULT 1.0E-12
    &END QS
    &SCF
      EPS_SCF 1.0E-6
      MAX_SCF 150
      SCF_GUESS atomic
    &END SCF
```

```
    &XC
      &XC_FUNCTIONAL Pade
      &END XC_FUNCTIONAL
    &END XC
    &POISSON
      PERIODIC NONE
      POISSON_SOLVER MT
    &END POISSON
    &PRINT
       &MOMENTS
           PERIODIC FALSE
       &END
    &END PRINT
  &END DFT
  &SUBSYS
    &CELL
      PERIODIC NONE
      ABC 30.000 30.000 30.000
    &END CELL
```

```
&MOTION
  &PRINT
    &VELOCITIES
      FORMAT ATOMIC
    &END
  &END PRINT
  &MD
    ANGVEL_ZERO
    COMVEL_TOL 0.000001
    &PRINT
      &CENTER_OF_MASS
      &END
    &END PRINT
    &AVERAGES
      &PRINT_AVERAGES
      &END
    &END AVERAGES
    ENSEMBLE NVE
    STEPS 10000
    TIMESTEP ${dt}
    TEMPERATURE 600.0
  &END MD
&END MOTION
```

Geometry Optimization: Two ways of solving the Kohn-Sham equation:

Reach for self-consistency (Traditional Scheme)

- Obtain lowest occupied orbitals from diagonalization of the KS matrix for a test density from occupied orbitals.

Minimization of the energy functional

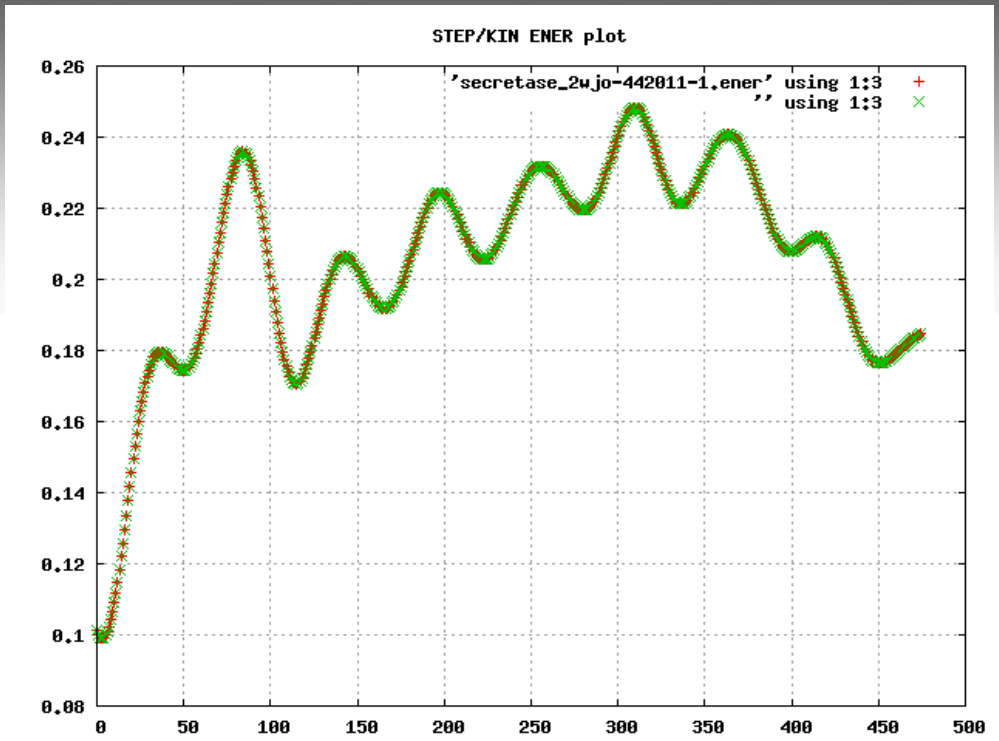
- Use a non-linear minimization scheme to minimize the KS energy

```
Core Hamiltonian energy:                                  192.6745297129
Hartree energy:                                           251.6513055795
Exchange-correlation energy:                              -80.2271045708
Coulomb (electron-electron) energy:                      2243.4298229402
      Maximum deviation from MO S-orthonormality              0.7105E-14
      Minimum/Maximum MO magnitude          0.3853E+00        0.1095E+01

      DIIS | Current SCF DIIS buffer size:           4
      DIIS | Maximum SCF DIIS error vector element:  8.401E-07
      DIIS | Current SCF convergence:                1.493E-06
      DIIS | Threshold value for a DIIS step:        1.000E-01
      DIIS | => The SCF DIIS buffer will be updated
      DIIS | => A SCF DIIS step will be performed

  29 B-13/Diag.  0.84E-06    3e-6      0.00000085      -257.9639771237 -2.16E-12

*** SCF run converged in    29 steps ***


Electronic density on regular grids:       -186.0000000195       -0.0000000195
Core density on regular grids:              185.9999999999       -0.0000000001
Total charge density on r-space grids:       -0.0000000196
Total charge density g-space grids:          -0.0000000196

Overlap energy of the core charge distribution:              0.00003527047067
Self energy of the core charge distribution:              -622.06274311582854
Core Hamiltonian energy:                                   192.67452971294904
Hartree energy:                                            251.65130557947546
Exchange-correlation energy:                               -80.22710457078813
Coulomb Electron-Electron Interaction Energy
- Already included in the total Hartree term              2243.42982294020931

Total energy:                                             -257.96397712372152

The electron density is written in cube file format to the file:

secretase_2wjo-442011-ELECTRON_DENSITY-1_0.cube
 Electronic kinetic energy:                               202.95295743270270
```

CP2K

Lagrangian:

The fictitious kinetic energy of the electrons and thus their temperature is a measure for the departure from the exact Born-Oppenheimer surface during Car-Parrinello dynamics.



Car and Parrinello postulated the following class of Lagrangians [108]

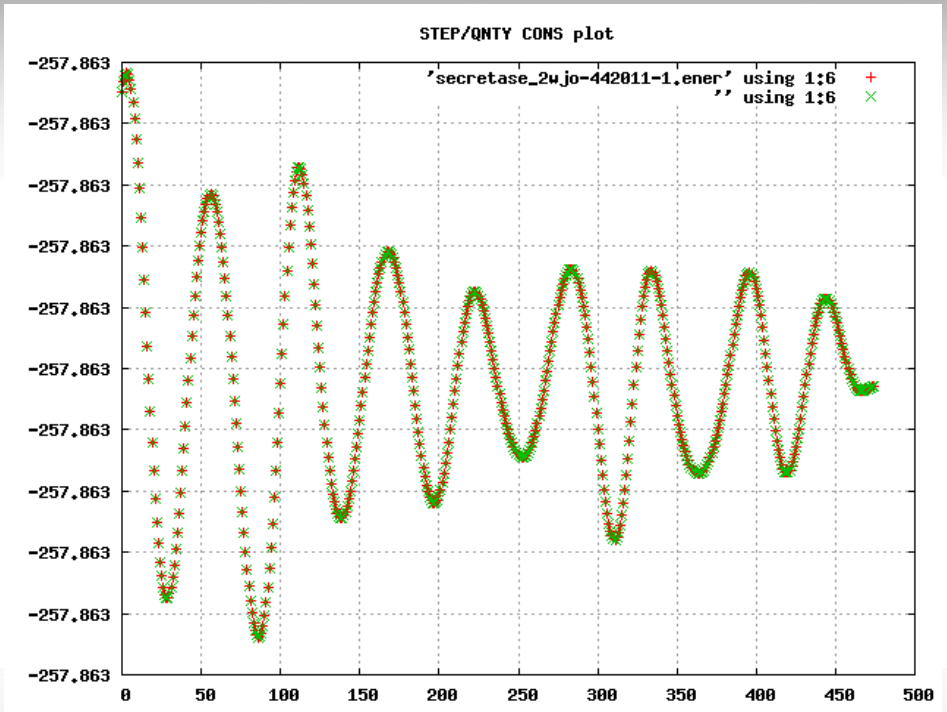
$$\mathcal{L}_{\mathrm{CP}} = \underbrace{\sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2} \mu_i \left\langle \dot{\psi}_i \middle| \dot{\psi}_i \right\rangle}_{\text{kinetic energy}} - \underbrace{\langle \Psi_0 | \mathcal{H}_{\mathrm{e}} | \Psi_0 \rangle}_{\text{potential energy}} + \underbrace{\textit{constraints}}_{\text{orthonormality}}$$

Energy Conservation:

•Conserving the total energy well is not enough if the system is very **heterogeneous** and the interesting part is small

•To improve the **energy conservation** one can either improve the basis-sets, increase the cutoff, and reduce the timestep.

•Normally having good forces and larger **timestep** is advantageous.



A structure based virtual screening against Ornithine carbamoyltransferase 2, phaseolotoxin-insensitive from Pseudomonas Syringae pv Actinidiae

SAPIENZA
UNIVERSITÀ DI ROMA

A structure based virtual screening against OTCase 2, phaseolotoxin-insensitive from *Pseudomonas Syringae pv Actinidiae*

Pseudomonas Syringae pv. actidiniae

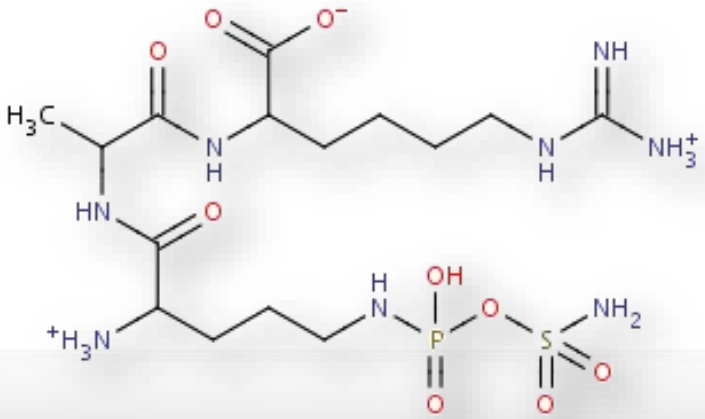
TOSSINE

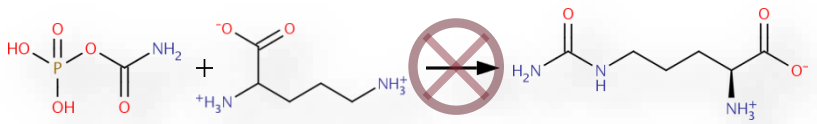
✓**Syringolin**

✓**Phaseolotoxin**

•**Citolisi per formazione di pori membranosi e fuoriuscita massiva di Ca^{2+} , H^{+} e K^{+}**

•**Blocco della sintesi della citrullina nell'ospite**





•Lai, J. R.; Koglin, A.; Walsh, C. T. *Biochemistry* **2006**, *45*, 14869-79.
•Tamura, K.; Imamura, M.; *Physiological and Molecular Plant Pathology* **2002**, *60*, 207-214.
•Bender, C. L.; Alarcon-Chaidez, F.; Gross, D. C. *Microbiology and Molecular Biology Reviews* **1999**, *63*, 266.
•Scortichini, M.; Ferrante, P. *Journal of Phytopathology* **2009**, *157*, 768-770.

The Target

STEP 1

OTCase 2: FASTA sequence

MKITSLKNRNLLTMNEFNQSELSHLIDRAIECKRLKKDRIFNLGLNHLNICGI
FLKPSGRTSTSFVVASYDEGAHFQFFPADNIRFGHKESIKDFARVVGRLF
DGIAFRGFEHEVAEELAKHSGIPVWNALTDTHHPTQVLADVMTVKEEFGR
IEGVTIAYVGDGRNNMVTSLAIGALKFGYNLRIIAPNALHPTDAVLAGIYEQ
TPERNGSIEIFTEVAAGVHQADVIYTDVWISMGESVSVEERIALLKPYKVTE
KMMALTGKADTIFMHCLPAFHDLDTEVARETPDLVEVEDSVFE
GPQSRVFDQGENRMHTIKALMLETVVP

STEP 2

Homology Modeling:

- Swiss model server
- M4T Server
- Uniprot server
- CPHmodels-3.0 Server

STEP 4

Structure Based Virtual Screening

STEP 3

- Geometric and structural optimization(Amber Suite)
- Protein Alignment(Chimera)

Conclusions

- ✓ Structure-Based 3-D QSAR model
- ✓ LB and SB Alignment: Assessment
- ✓ 3-D QSAR Model: External Validation
- ✓ AIMD simulation using DFT code
- ✓SB/VS of natural compounds vs. *Pseudomonas Syringae pv actinidiae*