NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Mini School

Lecture 01
An introduction to computer aided drug design

Dr Krishna Govender
Senior Research Scientist
Centre for High Performance Computing
Senior Research Associate
Department of Chemical Sciences
University of Johannesburg

A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
Department: Science and Innovation
REPUBLIC OF SOUTH AFRICA
CSIR

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Overview

- Why?
- Drug discovery process
- What is a drug?
- What is a protein?
- Computer aided drug design (CADD)
- Molecular Docking
- Computational Methods
- Molecular Dynamics
- Software
- Why High-Performance Computing?

NITheCS
National Institute for
Theoretical and Computational Sciences

nicis
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Why?

Old

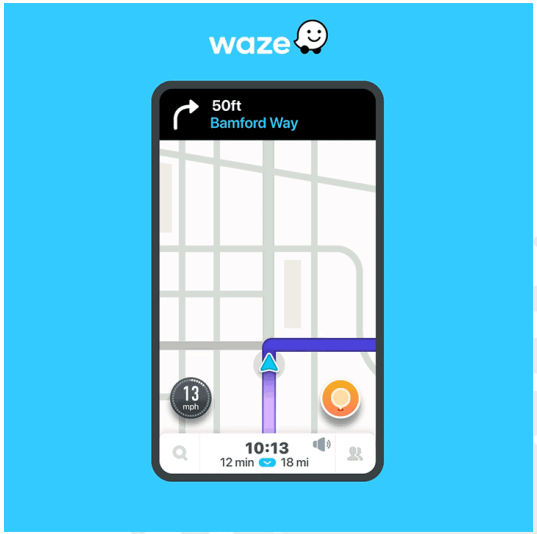
I have not failed.
I've just found 10,000 ways that won't work.
-Thomas Edison

A national initiative of the Department of Science and Innovation and implemented by the CSIR

science & innovation
Department: Science and Innovation
REPUBLIC OF SOUTH AFRICA

CSIR

Modern



NITheCS
National Institute for Theoretical and Computational Sciences

NICIS NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM CHPC

Drug discovery process

	Target Discovery	Target Validation	Lead Compound Identification	Lead Compound Optimization	Preclinical Development	Clinical Trials
Average Length	1 – 3 years				1,5 years	6 – 7 years
Average Cost	$196 million				$122 million	$1 – 2.5 billion
Goal(s)	Identification of molecule involved in a disease ➢ Identify the target: a molecule integral to gene regulation or intracellular signalling ➢ Ensure the target is 'druggable' and its activity can be modulated by another compound	➢ Validate initial hypothesis through gene knockdowns ➢ Test antibody interactions ➢ Modulate the drugs affinity to target by changing molecular structure	Generation of molecules(s) that can interact with the target previously identified ➢ Test drug mechanism of action ➢ Initial safety tests conducted in cell culture ➢ Test pharmacokinetics and pharmacodynamics	Compound modifications for increased effectiveness and safety ➢ Alter design of molecule to prevent off-target effects ➢ Optimize dosage and introduction route (oral, injection) ➢ Conduct tests for drugs uptake by 3D cell culture systems	Drug testing *in vivo* for side effects and safety ➢ Test drug in alternative cell lines and *in vivo*: most commonly mouse and rat research models ➢ Plan for either small- or large-scale production if approved	New drug approval by the FDA ➢ File Investigational New Drug Investigation to begin trials ➢ Includes three phases of human testing ➢ FDA conducts reviews and approvals after phase III ➢ Continued monitoring for dosage and safety

A national initiative of the Department of Science and Innovation and implemented by the CSIR

science & innovation Department: Science and Innovation REPUBLIC OF SOUTH AFRICA

CSIR

[1] https://www.taconic.com/taconic-insights/quality/drug-development-process.html, accessed 1 April 2022

What is a drug?

- A drug is a small molecule (key) that binds to a target such as a large protein or enzyme (lock) and as a result it turns on or off specific biochemical/physiological process in the body.

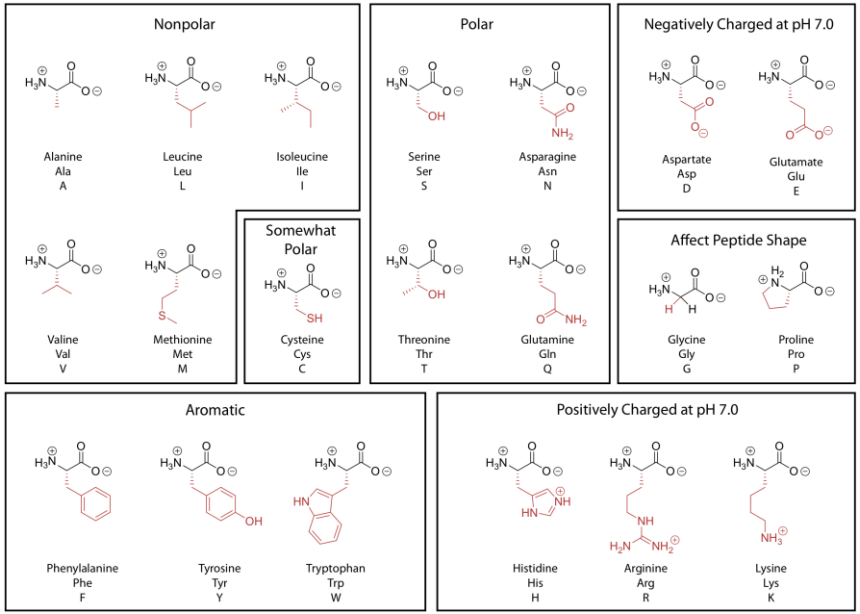
What is a protein?

- The building blocks of proteins are amino acids, which are small organic molecules that consist of alpha (central) carbon atom linked to an amino group, a carboxyl group, a hydrogen atom and a variable component called a side chain

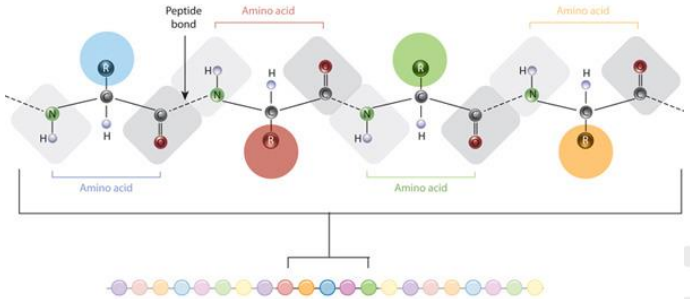
NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

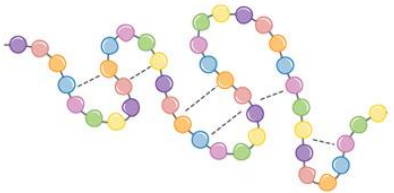
What is a protein?



Primary structure



Secondary structure



A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

[2] https://www.nature.com/scitable/topicpage/protein-structure-14122136/#:~:text=The%20building%20blocks%20of%20proteins,side%20chain%20(see%20below),accessed 1 April 2022

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

What is a protein?

Protein

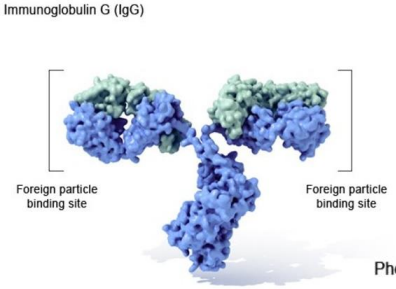
Antibody

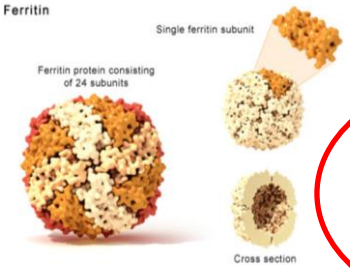
Transport/
Storage

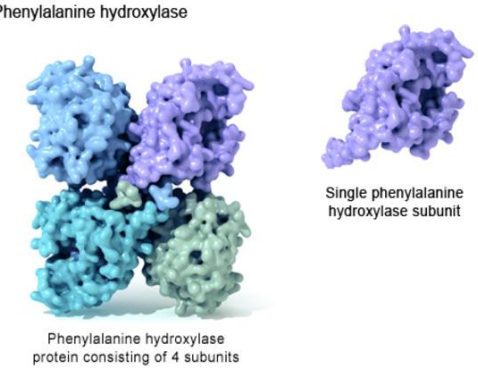
Enzyme

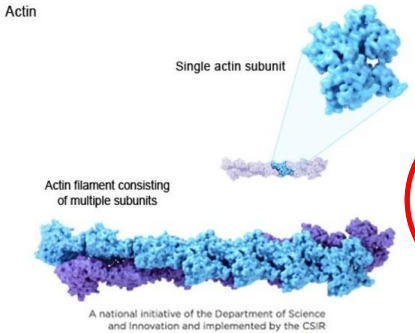
Structural
component

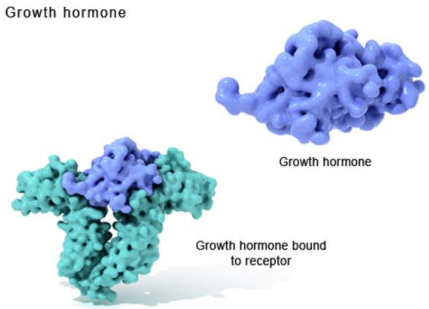
Messenger











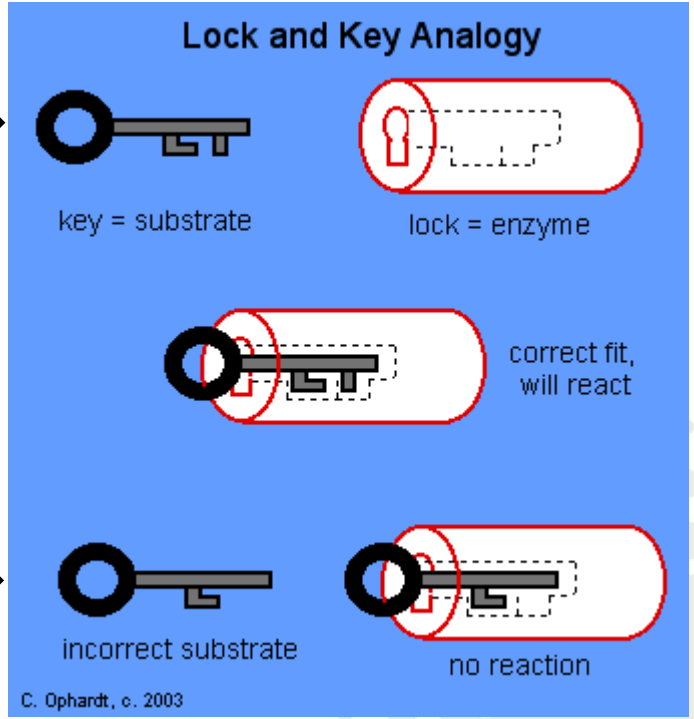
A national initiative of the Department of Science and Innovation and implemented by the CSIR

science & innovation
Department: Science and Innovation
REPUBLIC OF SOUTH AFRICA

CSIR

[3] https://medlineplus.gov/genetics/understanding/howgeneswork/protein/, accessed 1 April 2022

What is a drug?



[4] http://chemistry.elmhurst.edu/vchembook/571lockkey.html, accessed 1 April 2022

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

How do we find the key to our lock?

High-Throughput Screening (HTS)

This is a method for scientific experimentation especially used in drug discovery and is relevant in biology and chemistry. It combines robotics, data processing and control software, liquid handling devices and sensitive detectors allowing researchers to conduct numerous chemical, genetic or pharmacological tests.

Virtual Screening (VS)

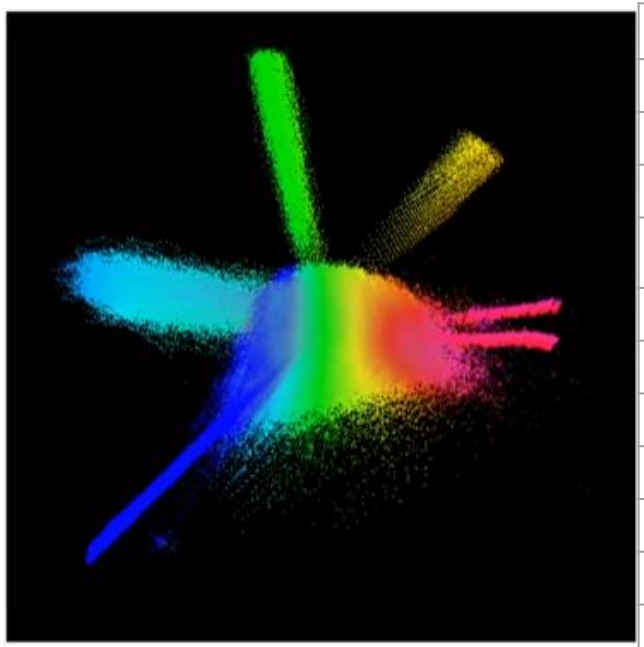
Is a computational technique used to search libraries of small molecules in order to identify those structures which are most likely to bind to a drug target, such as a large protein.

A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

NITheCS
National Institute for
Theoretical and Computational Sciences

nicis
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

How do we find the key to our lock



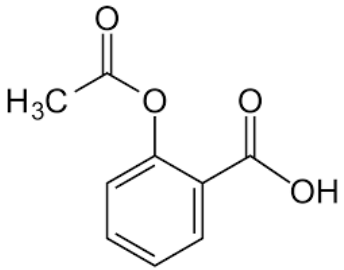
Database	Description	Size	Web address
PubChem	Known molecules from various public sources	32.5 M	http://pubchem.ncbi.nlm.nih.gov
Chemspider	Online resource from the Royal Society of Chemistry	26.0 M	http://www.chemspider.com
ZINC	Commercially available small molecules	21.0 M	http://zinc.docking.org
NCI Open	Anticancer and AIDS compounds with screening data	0.25 M	http://cactus.nci.nih.gov/ncidb2.1
ChemDB	Commercially available small molecules	4.1 M	http://cdb.ics.uci.edu
BindingDB	Bioactive molecules with binding affinity data	0.36 M	http://www.bindingdb.org
ChemBank	Small molecules annotated with screening data	1.2 M	http://chembank.broadinstitute.org
ChEMBL	Small molecules annotated with experimental data	1.1 M	https://www.ebi.ac.uk/chembldb
CTD	Comparative toxicogenomics database	0.17 M	http://ctdbase.org
HMDB	Human metabolome database	0.0085 M	http://www.hmdb.ca
SMPDB	Small molecule pathway database	0.001 M	http://www.smpdb.ca
DrugBank	Experimental and approved small molecule drugs	0.0065 M	http://www.drugbank.ca

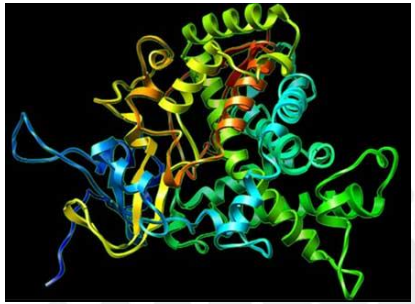
A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

[5] Reymond, J-L and Awale, M. ACS Chem. Neurosci. 2012, 3, 649 – 657.

Computer aided drug design (CADD)

- There are two general types of CADD namely structure-based drug design (SBDD) and ligand-based drug design (LBDD)



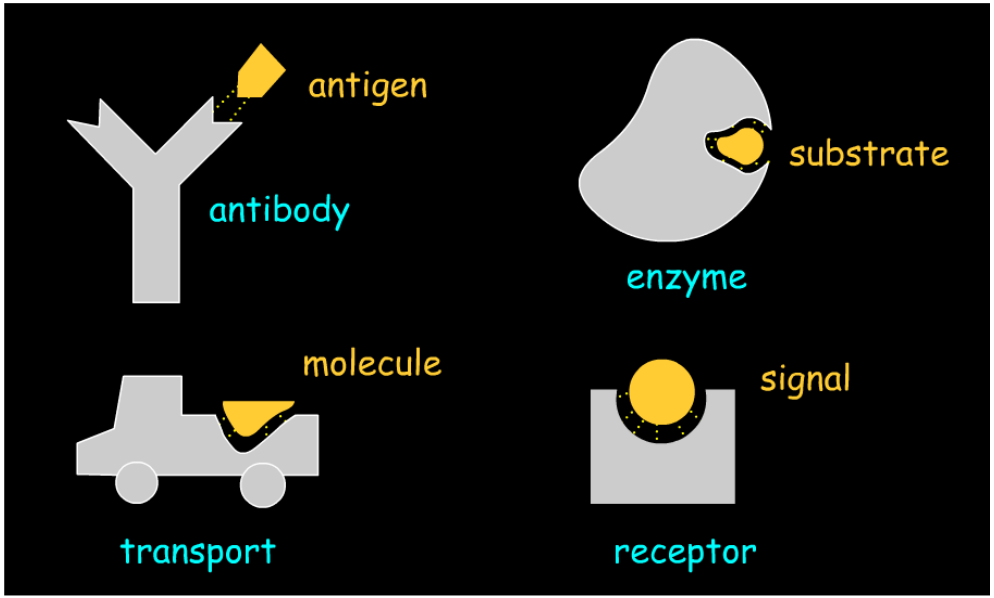


NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Molecular Docking

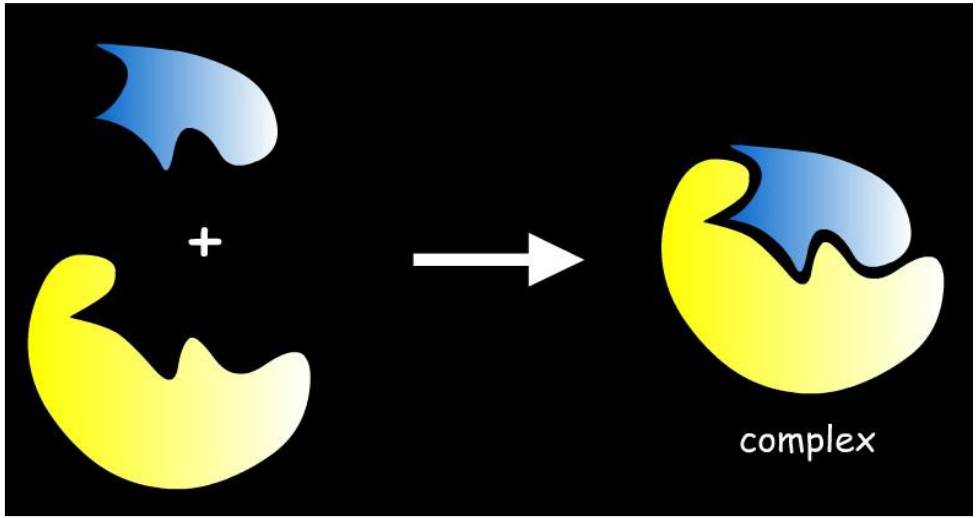
- Predicts the optimal orientation and conformation of interacting molecules in space and estimates the stability of the complex formed



A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

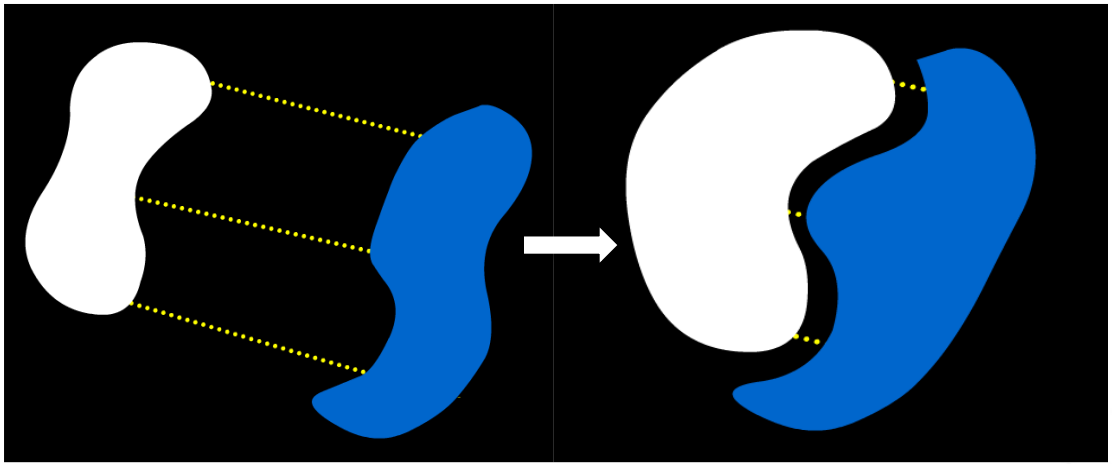
Molecular Docking

- **Lock-and-key**
  - A drug/ligand fits into the active site of a macromolecule, just like the key fits into a lock



Molecular Docking

- **Induced fit**
  - Both drug/ligand and target protein mutually adapt to each other through small conformational changes until optimal fit is achieved

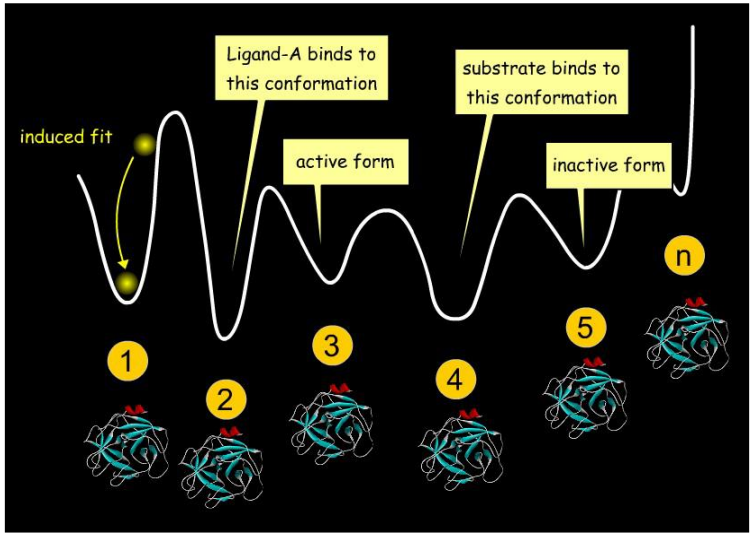


NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Molecular Docking

- **Conformational Ensemble**
  - Just like drugs/ligands can undergo conformational changes there are proteins that undergo large conformational changes.

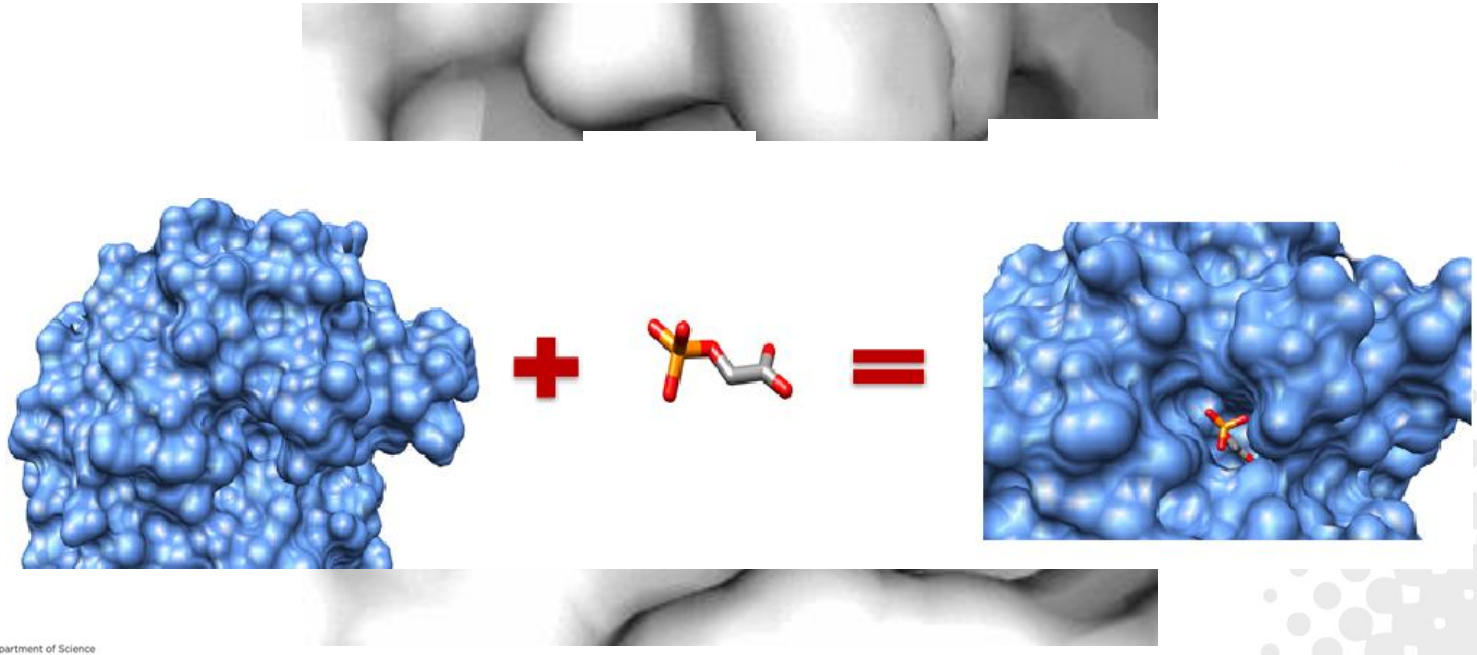


A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

NITheCS
National Institute for
Theoretical and Computational Sciences

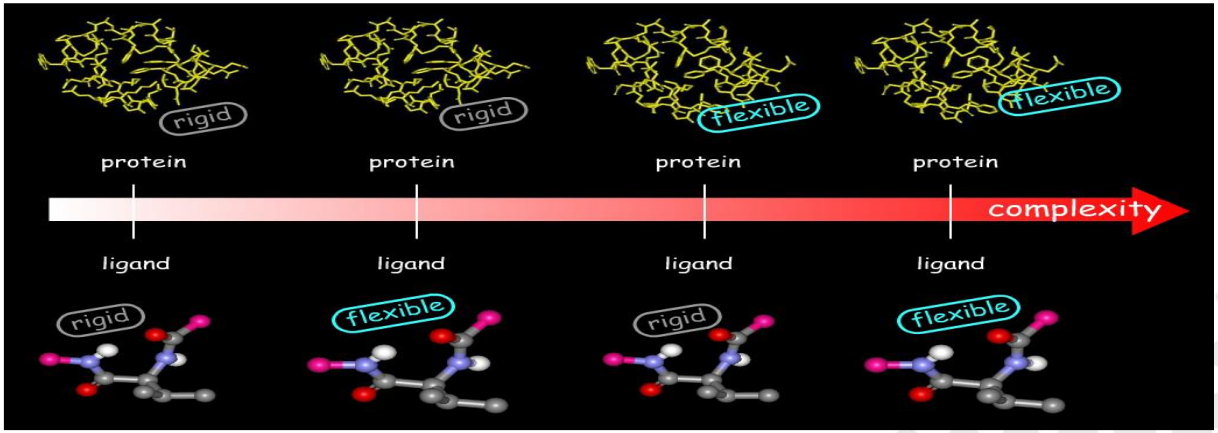
NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Molecular Docking



Molecular Docking

- **Different docking categories:**
  - Protein-ligand
  - Protein-protein
  - Protein-nucleic acid
  - Enzyme-substrate
  - Ligand-nucleic acid

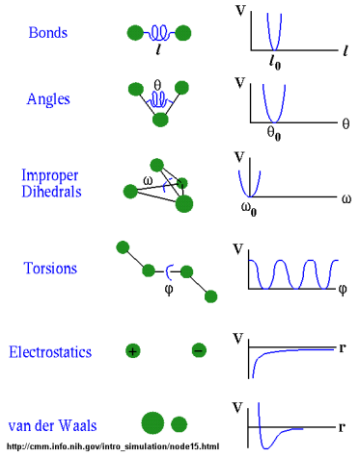


- The interactions between protein and ligand are by far much better understood compared to those between protein-protein or protein-nucleic acid

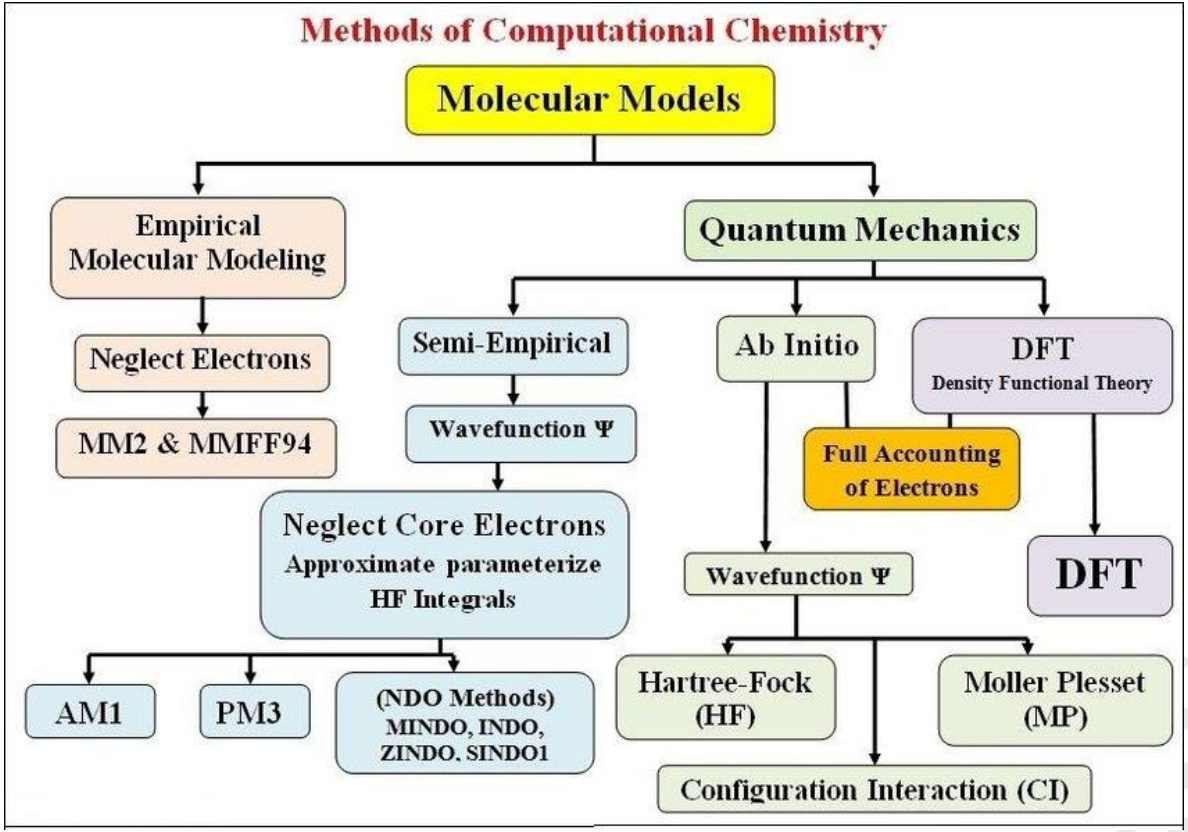
NITheCS
National Institute for Theoretical and Computational Sciences

NICIS NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM CHPC

Empirical Potential Energy Function



Methods of Computational Chemistry



$$\hat{H}\,\Psi = E\,\Psi$$

Hamiltonian Operator (Energy operator) — Energy eigenvalue

$$\frac{-\hbar^2}{2m}\nabla^2\Psi(\mathrm{r}) + V(r)\Psi(\mathrm{r}) = E\Psi(\mathrm{r})$$

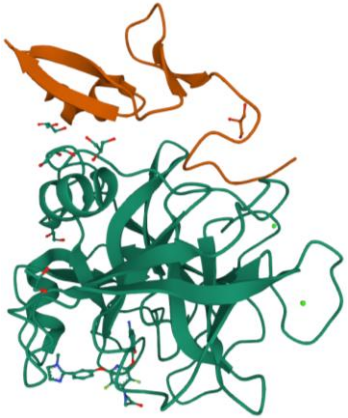
Kinetic Energy + *Potential Energy* = *Total Energy*

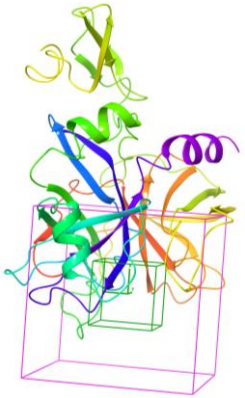
A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation Department: Science and Innovation REPUBLIC OF SOUTH AFRICA
CSIR

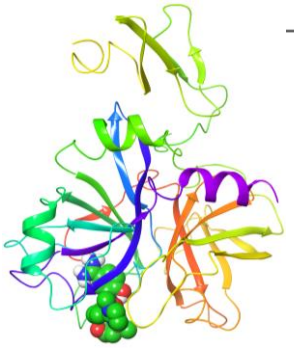
NITheCS
National Institute for
Theoretical and Computational Sciences

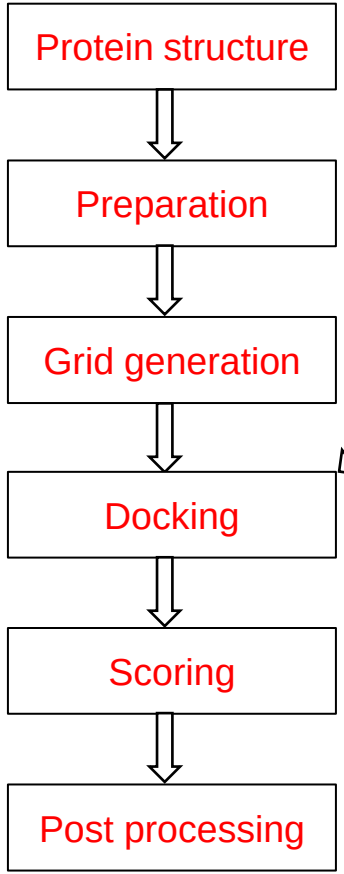
NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Molecular docking









A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

[6] PDB database, https://www.rcsb.org/, PDB ID: 1FJS, Accessed 4th April 2022.

Post docking

- Take the hits, synthesis and test experimentally
- Run further simulations to verify that the possible hits are stable when exposed to biological conditions
- This means explicitly taking the effects of solvent into account
- This is done with the aid of molecular dynamics

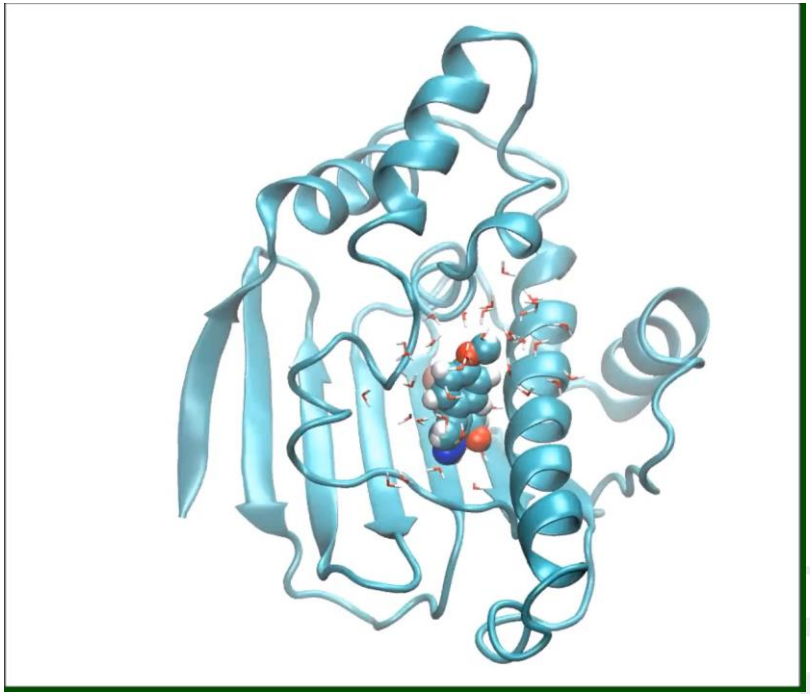
Molecular Dynamics

- It is often used to explore conformational space
- In this approach a single-point model is replaced by a dynamic model in which the nuclear system is forced into motion
- It assists in determining if a ligand stays bound to an active site or not

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Molecular Dynamics



A national initiative of the Department of Science and Innovation and implemented by the CSIR

science & innovation
Department
Science and Innovation
REPUBLIC OF SOUTH AFRICA

CSIR

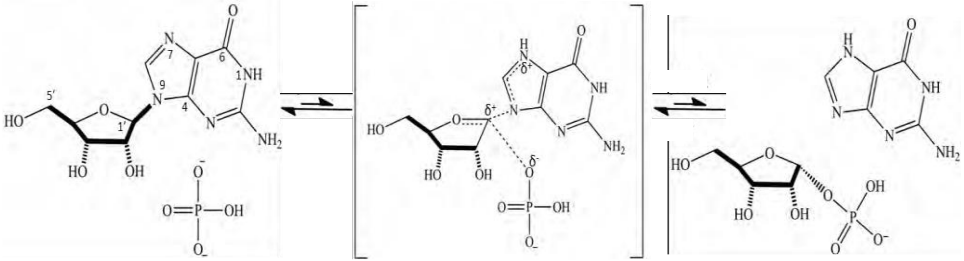
[7] Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A. and Rocchia, W. Scientific Reports, 2015, 5, 11539.
https://doi.org/10.1038/srep11539

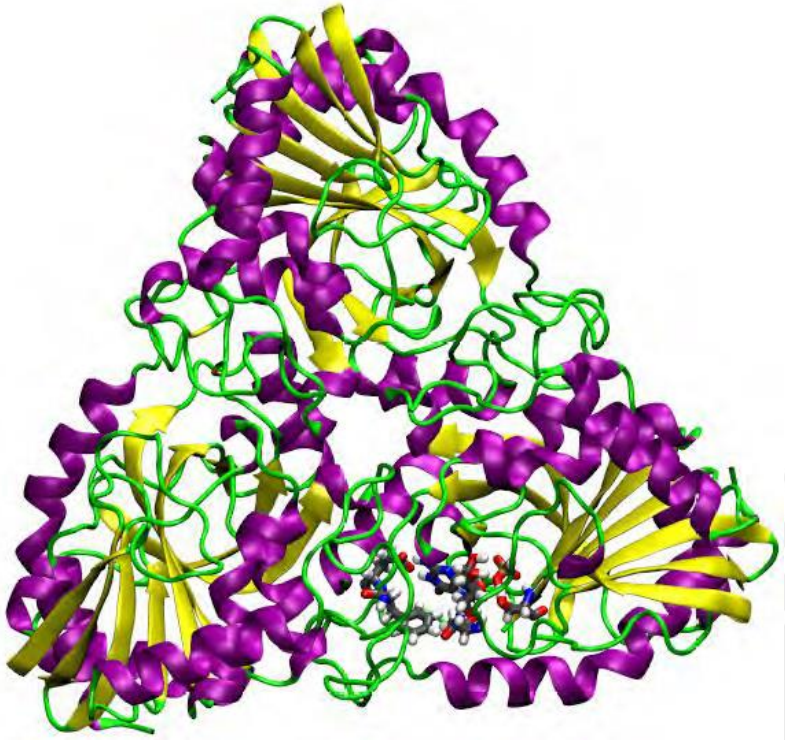
NITheCS
National Institute for
Theoretical and Computational Sciences

nicis
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Molecular Dynamics

- What if you have a reaction taking place in the active site





Secondary structure of Purine Nucleoside Phosphorylase (PNP) with α-helices in purple, β-sheets in yellow and random coil in green

A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

[8] Govender, K. K., The development of Hybrid Quantum Classical Computational Methods for Carbohydrate and Hypervalent Phosphoric systems, August 2014.

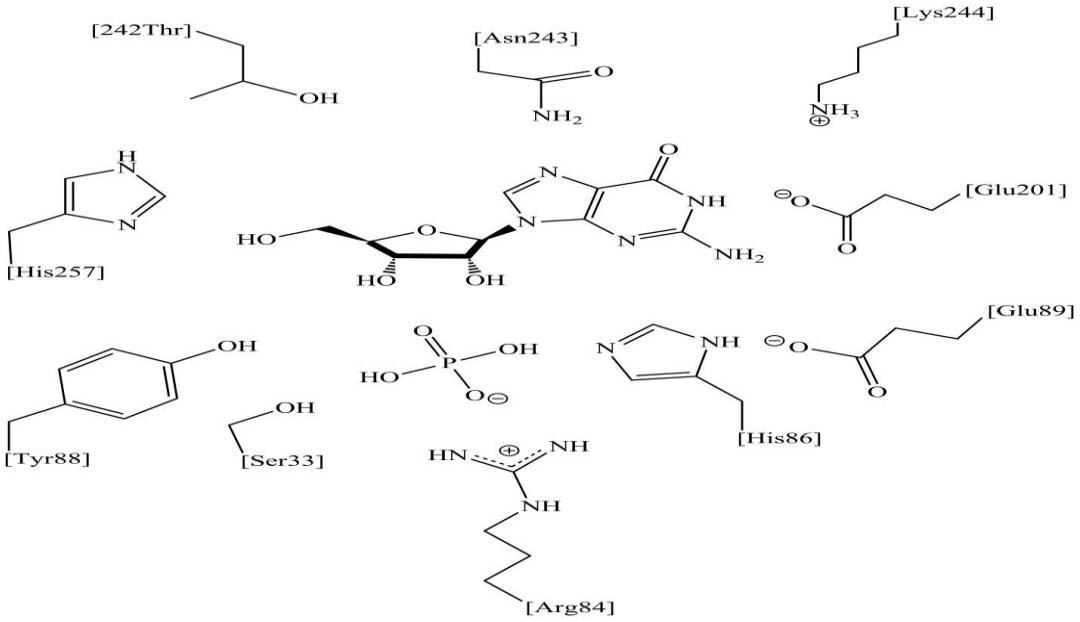
Molecular Dynamics

- One can use the ONIOM approach (Our own n-layered Integrated Molecular Orbital and Molecular Mechanics)
- This is not going to take the motion of the enzyme into consideration
- In order to take the motion of all atoms into consideration one needs to do QM/MM MD

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Molecular Dynamics

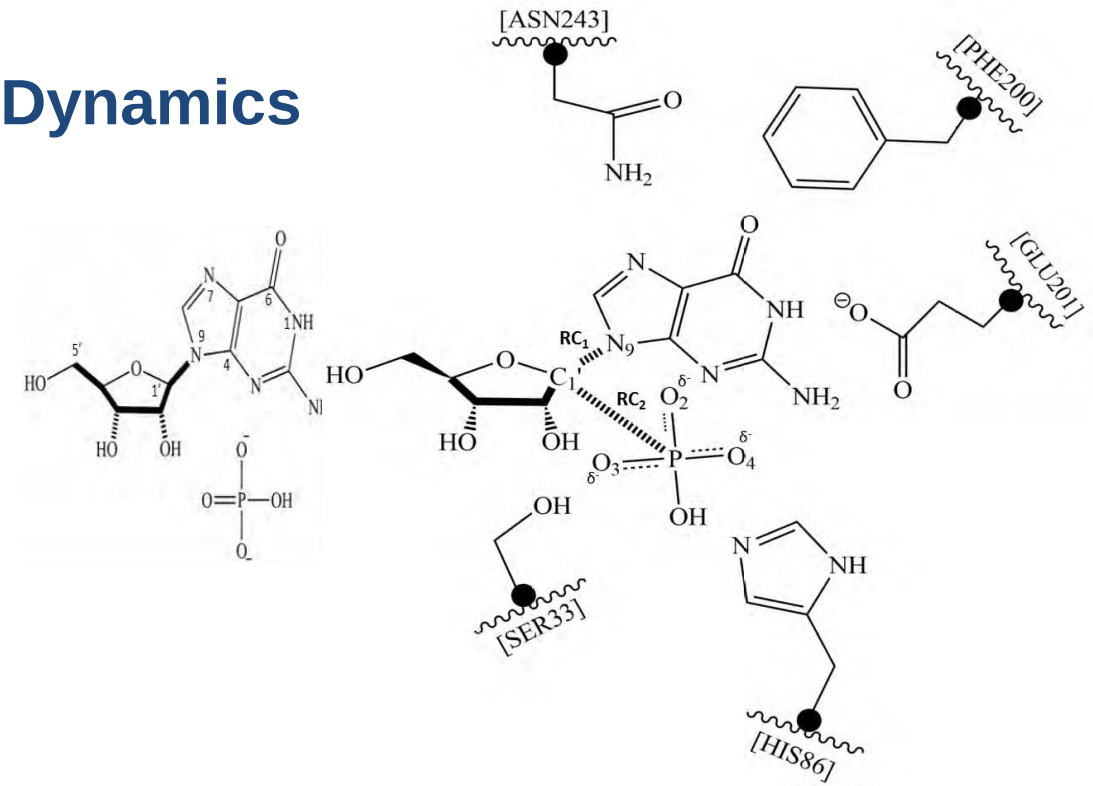


PNP active site model constructed from the atomic coordinates for the PNP-guanine complex

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

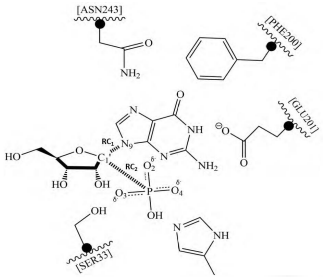
Molecular Dynamics



Chosen active site for PNP with QM region indicated. GHO atoms are represented with black spheres. RC_1 and RC_2 indicate the two reaction coordinates used for the reaction.

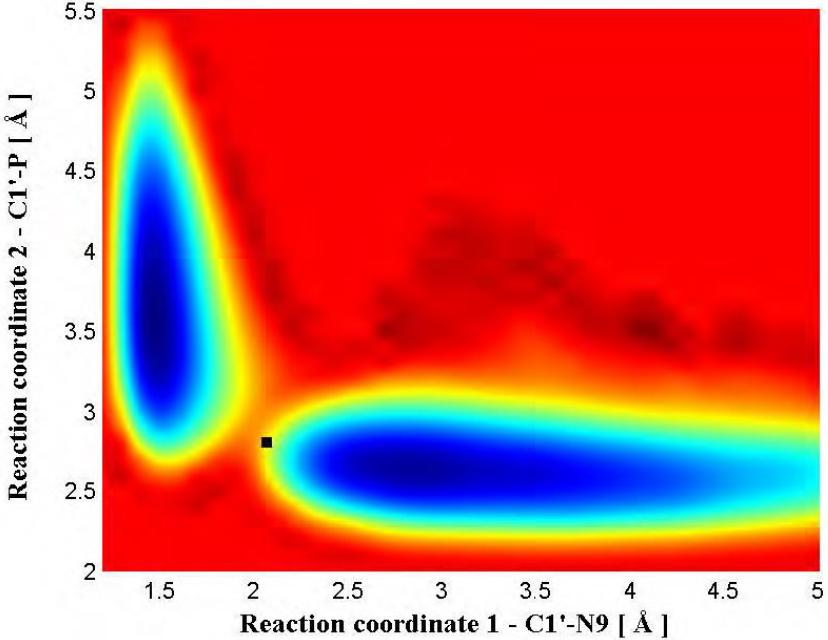
A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

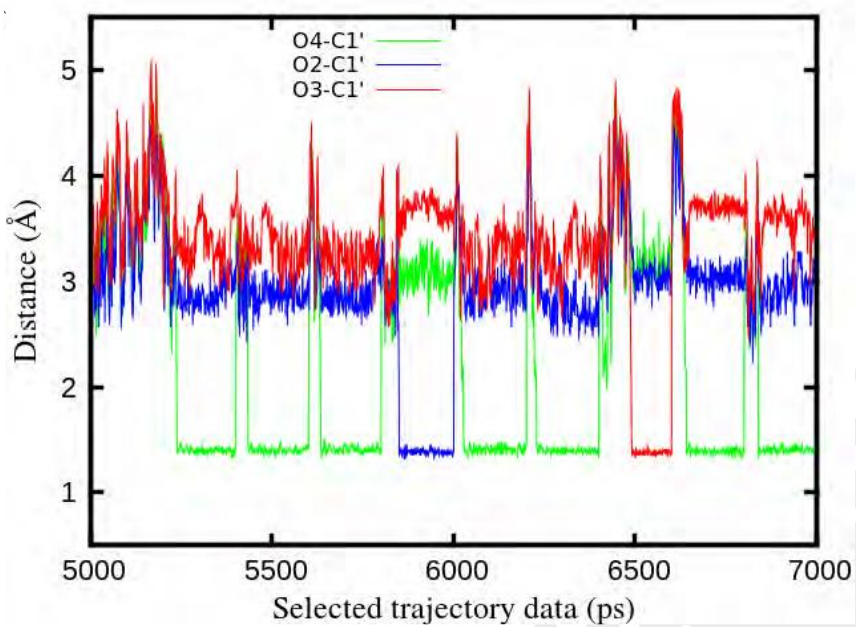
NITheCS
National Institute for
Theoretical and Computational Sciences



NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Molecular Dynamics





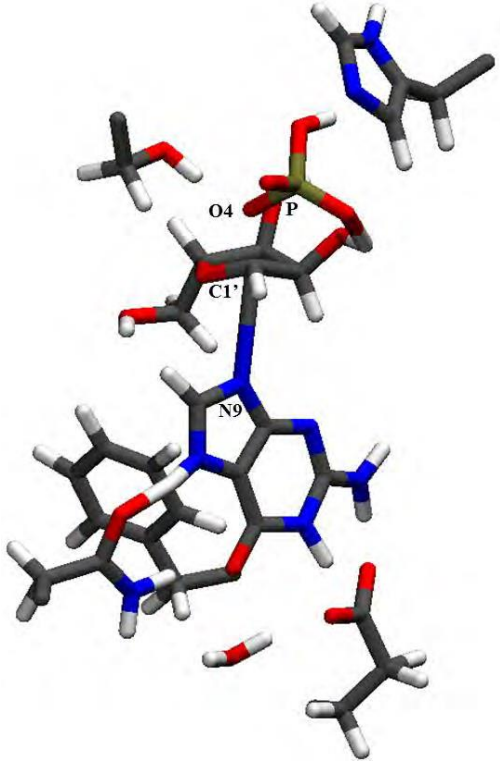
A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

Free energy surface viewed along the two reaction coordinates with the transition state indicated in black

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Molecular Dynamics



Transition state structure obtained from free energy simulations of PNP

A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
CSIR

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Software

Package	Application	Licensed
Schrödinger	Docking, MD and QM	Yes
AMBER	MD, QM/MM MD	Partially
GROMACS	MD, QM/MM MD	No
NWChem	MD, QM/MM MD, QM	No
Gaussian	QM, QM/MM	Yes
ORCA	QM	No
DMol3	QM, QM/MM	Yes
NAMD	MD	No
Autodock / Autodock VINA	Docking	No
CHARMM	MD, QM/MM MD	Yes
VMD	Visualization and post-processing	No
GaussView	Visualization, pre- and post-processing	Yes
Avogadro	Visualization, pre- and post-processing	No

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Why HPC?

- Docking simulations would typically involve using 100's to millions of possible drug candidates
- For each ligand different conformations need to be identified
- Each conformer needs to be docked into the target molecule and important properties related to this binding need to be computed
- Once you have possible targets and you wish to run MD you require a lot of computing resources
- MD can be sped up by making use of multiple CPU cores or graphical processing units (GPUs)

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

Access to the software

https://users.chpc.ac.za

The license for Schrödinger is only accessible to academics based at a South African tertiary institution

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED
CYBERINFRASTRUCTURE SYSTEM
CHPC

Acknowledgements

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED
CYBERINFRASTRUCTURE SYSTEM
CHPC

A national initiative of the Department of Science
and Innovation and implemented by the CSIR

science & innovation
Department:
Science and Innovation
REPUBLIC OF SOUTH AFRICA

CSIR

NITheCS
National Institute for
Theoretical and Computational Sciences

NICIS
NATIONAL INTEGRATED CYBERINFRASTRUCTURE SYSTEM
CHPC

THANK YOU

kgovender3@csir.co.za

A national initiative of the Department of Science and Innovation and implemented by the CSIR
science & innovation
Department: Science and Innovation
REPUBLIC OF SOUTH AFRICA
CSIR