Flexible Molecular Alignment

Sofia Oliveira

HASLab/INESC TEC, Universidade do Minho
with collaboration of BIAL

April 10, 2019
Q Days 2019
QuantaLab Workshop in Quantum Computation

Outline

Motivation

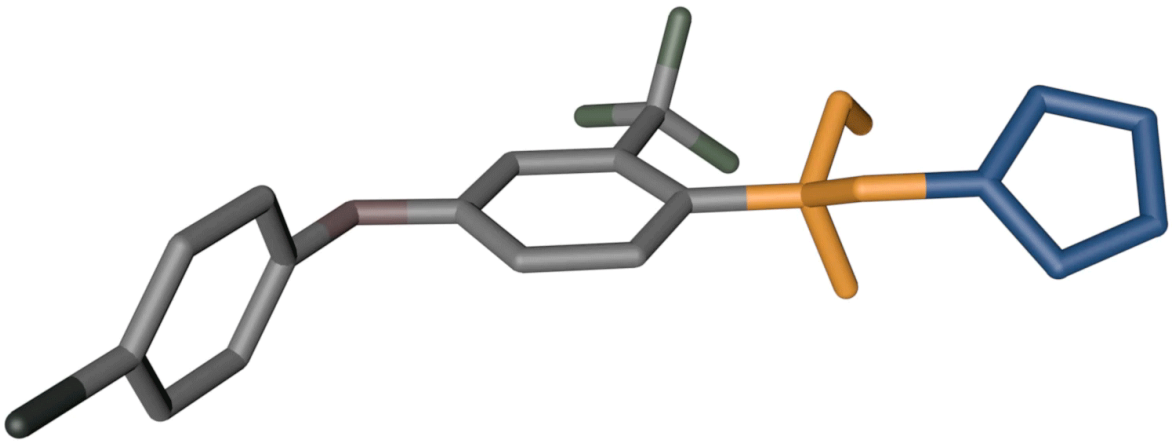
Flexible molecular alignment is a complex problem in the area of Medicinal Chemistry where, even nowadays, the approach to find the solution is, due to the tremendous computational power required, not a certainty but instead a probability.
Quantum computation, due to his natural parallelism, may improve algorithmic solutions for this kind of problems because it may test and/or simulate all possible solutions in an execution cycle.

Why a Computational Challenge?

- What is a flexible molecule?
- What is a conformation of a flexible molecule?
- How to align such molecules?

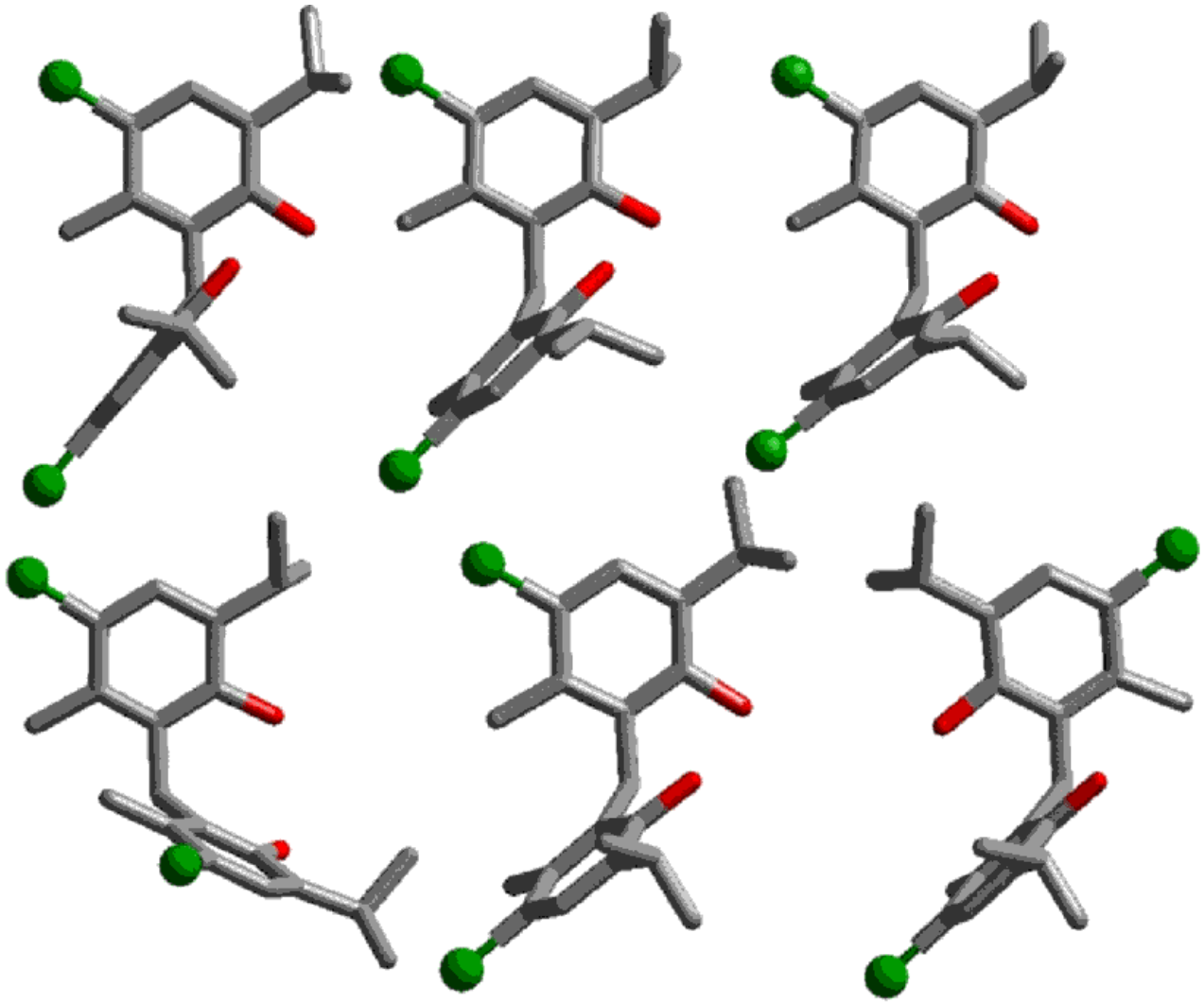
What is a Flexible Molecule?

Molecule that can have dozens of degrees of freedom i.e. connections between atoms that can suffer rotations and translations.
Because this degrees of freedom are continuous, there are infinite 'shapes' that this molecule can have.



What is a Conformation?

Each shape a flexible molecule can take is named conformation.

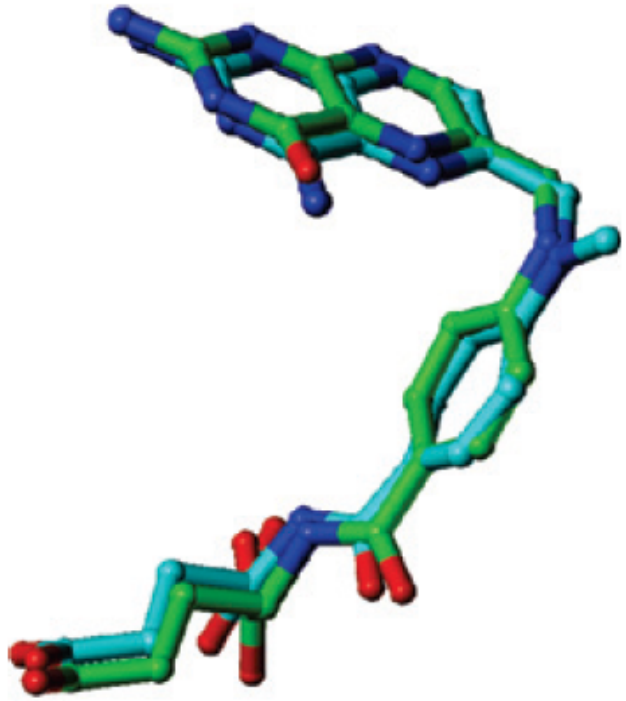


How to align such molecules?

Obvious though:

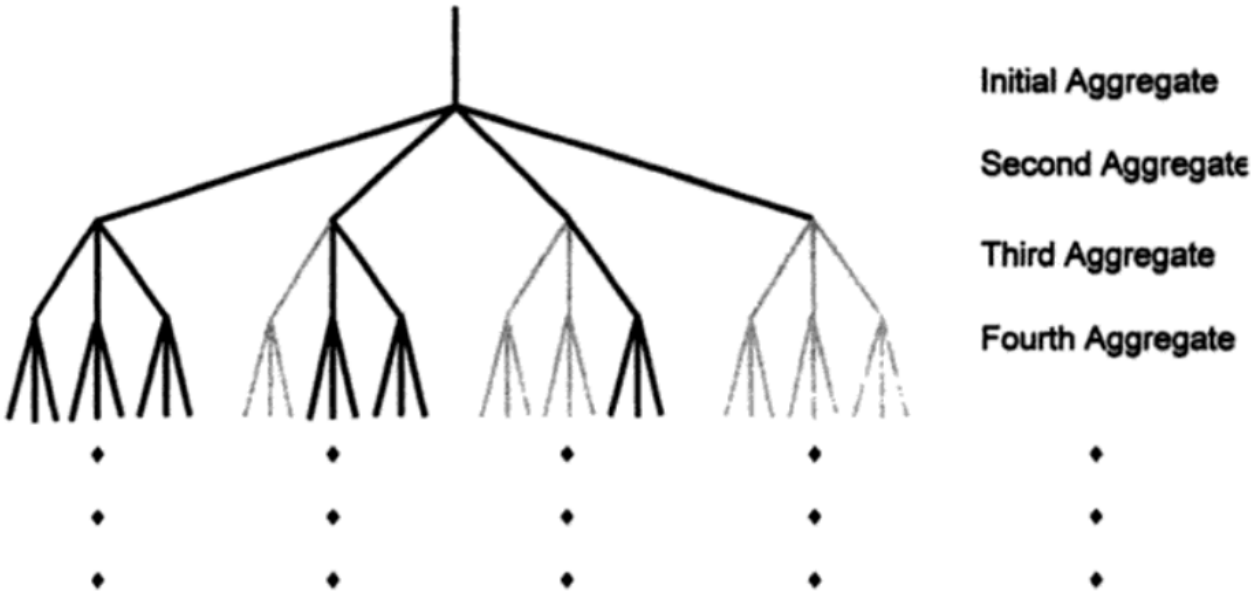
"Test all possible alignments."

When comes to complex flexible molecules, we are talking about hundreds of possible conformations for each molecule.
Testing all possible solution becomes unbearable for a classical computer to analyze in an acceptable length of time.



How to align such molecules?

Nowadays, the most used approach it based on a 'tree method'. [1]



Why a quantum approach can be better?

Due to it's natural parallelism, it could possibly solve:

- Time taken to analyze all possible solutions
- By doing this, one would be choosing, with certainty, the best alignment

The Big Picture - Abstraction

Because we want to see the actual behaviour of a quantum algorithm in this problem, one needs to take into accounts the available quantum computation power. Because of that, one cannot start to solve this problem as a whole, too much data.

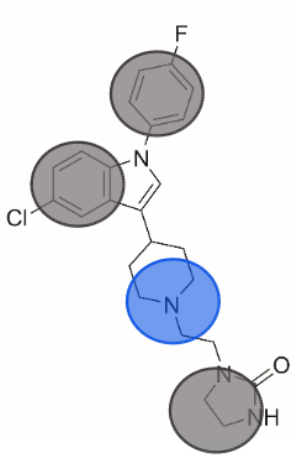
Abstraction is needed.

Alignment based on Pharmacophore Features

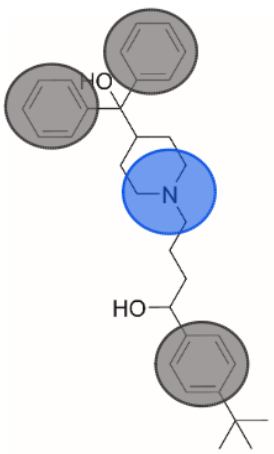
Definition of Pharmacophore: [2]

"A pharmacophore can be seen as an abstract description of molecular features which are necessary for molecular recognition of a ligand by a biological macromolecule."

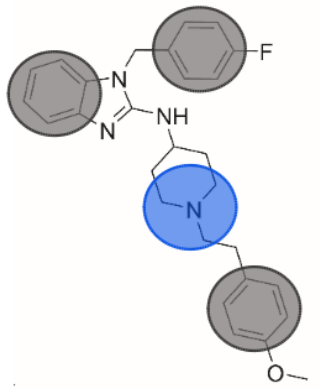
- Positive charge
- Hydrophobic



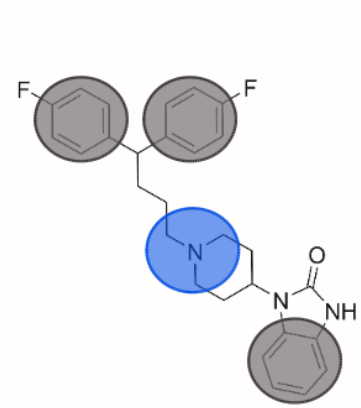
Sertindole



Terfenadine



Astemizole



Pimozide

Statement of the Abstracted Problem

There are N sets. For each $S \in N$, M subsets. For each $R \in M$, a variable number of points.
Choose a subset from each N that contributes to the best alignment.

Statement of the Abstracted Problem

There are N sets. For each $S \in N$, M subsets. For each $R \in M$, a variable number of points.
Choose a subset from each N that contributes to the best alignment.

- *This alignment depends on the distance between the 'connected' points of the same type, from each subset.*

Statement of the Abstracted Problem

There are N sets. For each $S \in N$, M subsets. For each $R \in M$, a variable number of points.
Choose a subset from each N that contributes to the best alignment.

- *This alignment depends on the distance between the 'connected' points of the same type, from each subset.*
- Each point can only be connected to a maximum of N-1 points, one from each set.

Statement of the Abstracted Problem

There are N sets. For each $S \in N$, M subsets. For each $R \in M$, a variable number of points.
Choose a subset from each N that contributes to the best alignment.

- *This alignment depends on the distance between the 'connected' points of the same type, from each subset.*
- Each point can only be connected to a maximum of N-1 points, one from each set.
- The calculation of the 'fitness' of a given N subsets takes into account all possibly connected points that are at a maximum of 2 Å distance from each other.

Statement of the Abstracted Problem

There are N sets. For each $S \in N$, M subsets. For each $R \in M$, a variable number of points.
Choose a subset from each N that contributes to the best alignment.

- *This alignment depends on the distance between the 'connected' points of the same type, from each subset.*
- Each point can only be connected to a maximum of N-1 points, one from each set.
- The calculation of the 'fitness' of a given N subsets takes into account all possibly connected points that are at a maximum of 2 Å distance from each other.

Each R is a rigid set of points that can undergo translation and rotation.

Fair Comparison

Classically, this problem is solved as a whole. Until now, what I have been doing, is a classic algorithm version of this abstraction to posteriorly make fair comparisons to the quantum algorithm that will also be made based on this abstractions.

Classical Algorithm - Approach

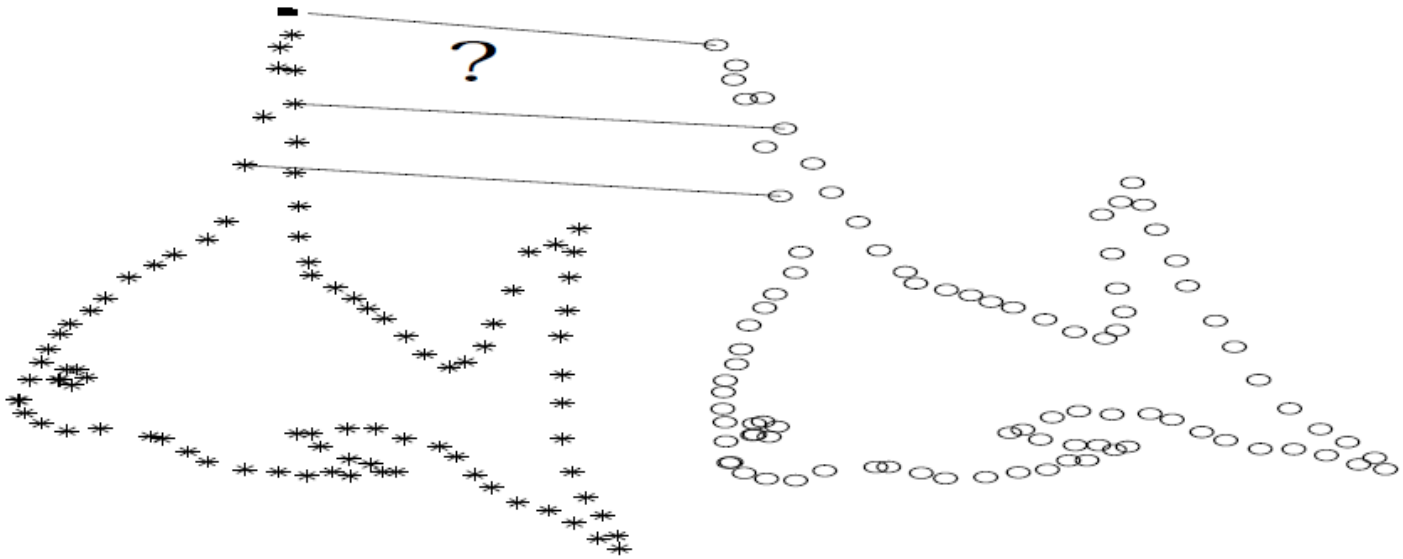
Each solution is a N-Dimensional Matrix, where each dimension is a conformation. Each position of the matrix is the distance between those points, if 'connected'.

With the purpose of analyze and find the best alignment, there are 3 steps:

- Pre-alignment of conformations - Coherent Point Drift Algorithm (PyCPD)
- Compliance with the imposed restrictions - SAT Model (Numberjack-SCIP)
- Choice of the best solution

Pre-alignment of conformations

Point Set Registration Problem: [3]
Given two sets of points, assign the correspondences and the transformation that maps one point set to the other.



Compliance with the imposed restrictions

It was used the solver SCIP of the numberjack library to impose the following restrictions to the generation of each possible solution:

- Points can only be 'connected' if they have the same feature (Hydrophobic or Positive Charge)
- Points can only be 'connected' if they are at a distance of equal or less than 2 Å
- One point can only be 'connected' to one other from each molecule

Choice of the best solution

Hand-made algorithm that abstracts from every matrix the number of non-zero elements and their sum and chooses the one that have the most non-zero elements and the lowest value as theirs sum.

Quantum Algorithm

- Pre-alignment of conformations
- Compliance with the imposed restrictions
- Choice of the best solution - Grover's Algorithm

Each of the above steps will have a different corresponding quantum algorithm.
The focus of this project, from now on, will be on finding or creating such algorithms.

References


Denise D. Beusen and Garland R. Marshall. "Pharmacophore Definition Using the Active Analog Approach". In: *Pharmacophore Perception, Development, and Use In Drug Design*. Ed. by Osman F. Güner. La Jolla, California: International University Line, 2000. Chap. 3, pp. 21–45.


Rafik Karaman. *What is Pharmacophore*. Acessed: 04-03-2019. 2016. URL: https://www.researchgate.net/post/What_is_Pharmacophore.


A. Myronenko and X. Song. "Point Set Registration: Coherent Point Drift". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.12 (Dec. 2010), pp. 2262–2275. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.46.