3D-Pharmacophore Based VIRTUAL SCREENING IN DRUG DISCOVERY

Chinmaya Chidananda Behera
Asst. Professor (Pharmaceutical Chemistry)
School of Pharmacy and Life Sciences
Centurion University, Bhubaneswar
3D based similarity

- Shape-based ROCS (Rapid Overlay of Chemical Structures) Silicos-it.com (Shape it)

- Computationally more expensive than 2D methods

- Requires consideration of conformational flexibility
  - Rigid search - based on a single conformer
  - Flexible search
    - Conformation explored at search time
    - Ensemble of conformers generated prior to search time with each conformer of each molecule considered in turn
    - How many conformers are required?
Multiple actives known: pharmacophore searching

- **IUPAC Definition**: “An ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response”

- In drug design, the term 'pharmacophore' refers to a set of features that is common to a series of active molecules.
  - Hydrogen-bond donors and acceptors, positively and negatively charged groups, and hydrophobic regions are typical features.

We will refer to such features as 'pharmacophoric groups'.
3D-Pharmacophores

- A three-dimensional pharmacophore specifies the spatial relationships between the groups
- Expressed as distance ranges, angles and planes
Workflow of Pharmacophore modeling

1. Collect a set of experimental compounds for a target.
2. Select a threshold for the activity values and mark compounds as Actives and Inactives. Create a training and test set.
3. Generate conformers and create pharmacophore model.
4. Identify the points essential for the activity of the molecules.
5. Perform statistical analysis on your model for model refinement.
6. VIRTUAL MOLECULES
   - Conformer generation and 3D database creation
6. 3D Search
5. Hits
4. Selected Pharmacophore model

Sensitivity Se = TP/TP + FN
Specificity Sp = TN/FP + FP
Fold of active VA/FP
Enrichment Factor EF = V(a/A/N)
Many Actives and Inactives known: Machine learning methods

- **SAR Modeling**
  - Use knowledge of known active and known inactive compounds to build a predictive model

- **Quantitative-Structure Activity Relationships (QSARs)**
  - Long established (Hansch analysis, Free-Wilson analysis)
  - Generally restricted to small, homogeneous datasets eg lead optimization.

- **Structure-Activity Relationships (SARs)**
  - “Activity” data is usually treated qualitatively
  - Can be used with data consisting of diverse structural classes and multiple binding modes
  - Some resistance to noisy data (HTS data)
  - Resulting models used to prioritize compounds for lead finding (not to identify candidates or drugs)
Protein Ligand Docking

Computational method which mimics the binding of a ligand to a protein.

It predicts ..
  a) the pose of the molecule in the binding site
  b) The binding affinity or score representing the strength of binding
Pose and Binding Site

• **Binding Site (or “active site”)**
  - the part of the protein where the ligand binds.
  - generally a cavity on the protein surface.
  - can be identified by looking at the crystal structure of the protein bound with a known inhibitor.

• **Pose ( “binding mode”)**
  - the geometry of the ligand in the binding site
  - Geometry- location, orientation and conformation of the molecule
Protein Ligand Docking

• How does a ligand (small molecule) bind into the active site of a protein?

• Docking algorithms are based on two key components
  – search algorithm
    • to generate “poses” (conformation, position and orientation) of the ligand within the active site
  – scoring function
    • to identify the most likely pose for an individual ligand
    • to assign a priority order to a set of diverse ligands docked to the same protein – estimate binding affinity
The search space

- The difficulty with protein–ligand docking is in part due to the fact that it involves many degrees of freedom
  - The translation and rotation of one molecule relative to another involves six degrees of freedom
  - These are in addition the conformational degrees of freedom of both the ligand and the protein
  - The solvent may also play a significant role in determining the protein–ligand geometry (often ignored though)
- The search algorithm generates poses, orientations of particular conformations of the molecule in the binding site
  - Tries to cover the search space, if not exhaustively, then as extensively as possible
  - There is a tradeoff between time and search space coverage
Dock Algorithms

• DOCK: first docking program by Kuntz et al. 1982
  – Based on shape complementarity and rigid ligands

• Current algorithms
  – Fragment-based methods: FlexX, DOCK (since version 4.0)
  – Monte Carlo/Simulated annealing: QXP(Flo), Autodock, Affinity & LigandFit (Accelrys)
  – Genetic algorithms: GOLD, AutoDock (since version 3.0)
  – Systematic search: FRED (OpenEye), Glide (Schrödinger)
DOCK (Kuntz et al. 1982)

- Rigid docking based on shape
  - A negative image of the cavity is constructed by filling it with spheres
- Spheres are of varying size
- Each touches the surface at two points
- The centres of the spheres become potential locations for ligand atoms.
DOCK

• Ligand atoms are matched to sphere centers so that distances between atoms equals distances between sphere centers.

• The matches are used to position the ligand within the active site.

• If there are no steric clashes the ligand is scored.

• Many different mappings (poses) are possible

• Each pose is scored based on goodness of fit

• Highest scoring pose is presented to the user
Energetics of protein-ligand binding

a) Ligand-receptor binding is driven by
   - electrostatics (including hydrogen bonding interactions)
   - dispersion or van der Waals forces
   - hydrophobic interactions
   - desolvation: surfaces buried between the protein and the ligand have to be desolvated
   - conformational changes to protein and ligand
   - ligand must be properly orientated and translated to interact and form a complex
   - loss of entropy of the ligand due to being fixed in one conformation.

b) Free energy of binding
\[ \Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib} \]
Conclusions

- Wide range of virtual screening techniques have been developed
- The performance of different methods varies on different datasets
- Increased complexity in descriptors and method does not necessarily lead to greater success.
- Combining different approaches can lead to improved results.
- Computational filters should be applied to remove undesirable compounds from further consideration.
References


• Hartenfeller, M.; Schneider, G. Enabling future drug discovery by de novo design. Wiley Interdisciplinary Reviews-Computational Molecular Science 2011, 1, 742-759.

• Open-source platform to benchmark fingerprints for ligand-based virtual screening Sereina Riniker, Gregory A Landrum Journal of Cheminformatics 2013, 5:26 (30 May 2013)
THANK YOU