QSAR

Quantum Biology
An EMBnet introductory course

12-11-2008

José R. Valverde
EMBnet/CNB
www.es.embnet.org
jr@cnb.csic.es

Lecture Outline

- What is QSAR?
- 3D-QSAR
- Identifying active elements
- Ligand-receptor interactions
- Ligand databases

QSAR

- Quantitative Structure Activity Relationship
  - A body of techniques
  - Aim is activity prediction of small compounds
    - Active role
    - Dynamic transport
    - Side effects
    - Etc...
  - An alternative to S. James Black method
    - Modify substrates in search of analogues
    - Test experimentally
  - Not a substitute:
    - Experimental testing still needed
    - But for a smaller set
  - ADME-TOX
  - 3D-QSAR

ADME-TOX

- ADME
  - Absoprtion, distribution, metabolism and excretion
    - Administration method
    - Transport and barrier-crossing
    - Metabolism (liver redox by CytP450
    - Elimination (kidney, lungs, seat, stools..)
- TOX
  - Potential for toxicity
    - Of the compound
    - Of its metabolites
- Methods
  - QSPR, QSAR
- Side effects?
  - Usually by experiment

QSAR

- Similar molecules have similar functions and similar structures
  - Similar structures should have similar functions
  - Wouldn't it be nice if we could say...
    - why one analogue works and other doesn't?
    - if an analogue will work or block function?
    - which analogue will work in advance?
    - what will be the effects of a substance?
    - What is the specific binding activity of different substrates?
    - what is the actual activity / reactivity of different species?
    - etc...
- Look at structural *similarities* and try to predict.
  - *Any* similarity in *any property* is worth using for our guess

3D-QSAR

- Traditional QSAR models used any structural property
  - Mainly bulk properties
    - Volume
    - Dimensions
    - Charges, etc...
  - Generally simplifications to speed up computation
- With bigger computers we can now consider the 3D structure
  - We may start from simplifications
    - Extreme coordinates/dimensions
    - Distribution of charges in space
    - Etc...
  - And go all the way to full models
    - Rigid
    - Flexible...

We want to

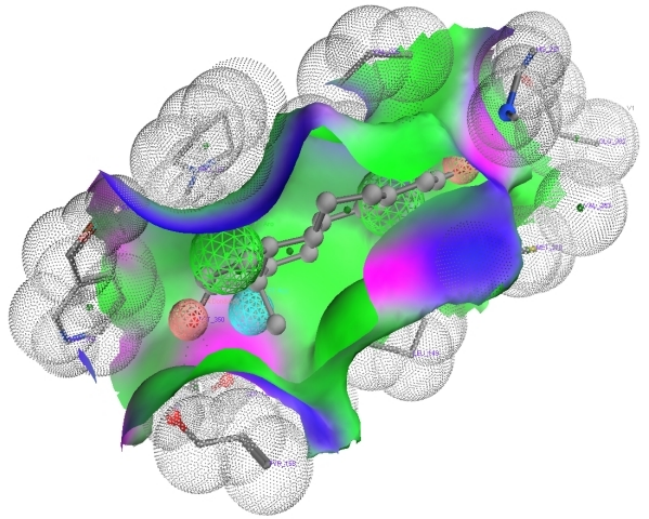
- Tell how structures interact
  - Identify possible interaction sites
  - Identify other molecules that might interact
  - Identify the 3D structure of the interaction
  - Analyze the 3D interactions
- Make inferences
  - Where is the active site?
    - What are the relevant residues for activity?
    - How can I interpret a polymorphism/mutation?
  - What are potential analogues
  - Will they bind more weakly or strongly?
  - Will the interaction be stable?
  - How will the reaction proceed?
    - Better or worst
    - Not at all (blocker)

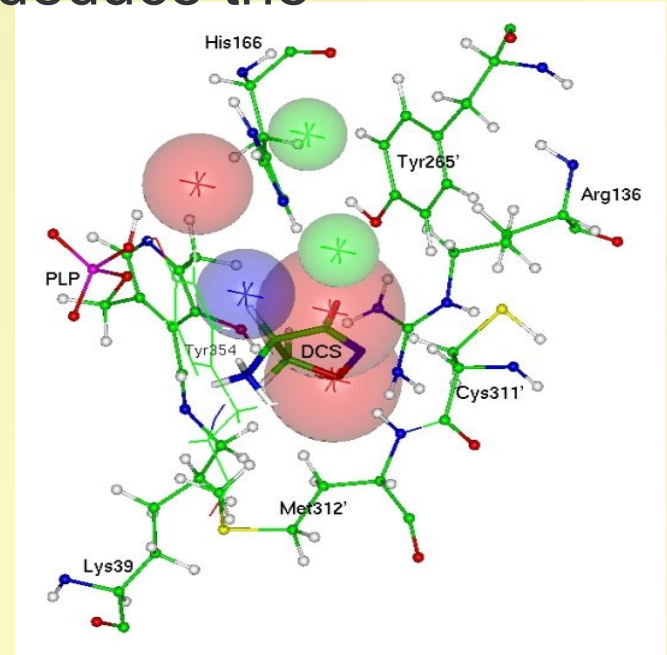
Identifying analogues

- Protein structure:
  - Experiment: X-ray, NMR, ME...
  - Modelling: homology, threading
  - Else...?
- Active analogue
  - Analyze fine structure of an active substrate
    - Groups most probably involved in the interaction
  - Refine fine structure
    - Most stable structure
  - Use several alternatives and look for possibly common features
  - Use several substrates and look for common features
  - If nothing is preserved use other properties (VdW, charge, etc...)

Analyzing analogues

- Pharmacophore (Paul Ehrlich 1909)
  - Molecular framework that carries the essential features for a drug's biological activity
  - The set of features recognized by the receptor and reponsible for the biological activity.
- Receptor volumes
  - From the superposition we may deduce the complementary cavity





Ligand-receptor binding

- Once the properties of ligands and binding site are known we can explore the binding process
  - Experiment (e. g. X-ray)
  - Elements
    - Receptor (experiment, model or receptor volume)
    - Ligand (one, pharmacophore or superposed group)
  - Grid maps
    - Define a grid around the protein
    - use functional sites as probes (e. g. H_2O for -OH)
    - Place probes on grid and look for putative interaction sites
  - Multiple copy minimization
    - Build a cloud of copies of the probe around the receptor
    - Minimize each of them independently
    - Model approximation of probe to receptor
    - Refine

Docking

- Grid maps and minimization identify putative interaction sites, but not the mechanism
- Docking
  - Simulate the interaction process
    - Start from many configurations and bind
    - MC
    - MM/MD
  - Compute interaction energies
  - Computatinally costly
    - Rigid models
    - Flexible model
    - Electrostatic interactions
    - Full MM/MD
  - Once the active site is known we can repeat/compare with other target drugs.

Searching

- Once the active site is known we can look for other target drugs
  - Match them in the active site
  - Match them in a receptor map
  - Match them with the pharmacophore
- Convert chemical formulas to structures
- Search databases
  - Commercial
  - CORINA (gives isolated structures and synthesis info)
  - HIC-UP (PDB ligands)
  - NCI (antitumour drugss)
  - ZINC (*the* database: millions of compounds)
  - Lead compounds: representatives of drug families
  - Dimensions → pKa, charge → aspect → structure

De novo design

- Good molecules may be known
- But *better* molecules *might* be designed
- Start from an interaction map
  - Match interaction points with interaction groups
  - Join groups with bridges that stabilize their position without damaging the interaction
- Very difficult process
  - Requires expert chemist
- Not limited to known existing molecules
- May potentially yield a much better drug