The Properties of Known Drugs

Shuqiang Wang

Department of Applied Mathematics & Statistics

1

- The Properties of Known Drugs: 1. Molecular Frameworks, 1996, JMC
- Properties of Known Drugs: 2. Side Chains, 1999, JMC

by Guy W. Bemis and Mark A. Murcko

2

Problem

- The data is a database of known drugs

  the MDL® Comprehensive Medicinal Chemistry (CMC) database with more than 6700 compounds (over 8,400 pharmaceutical compounds now) .

- The problem is to try to classify the shapes of drug molecules and provide a “high-level overview of the gross structural features of these molecules.

3

Purpose

- The purpose of doing this is that there is a wealth of information implicitly encoded in the two-dimensional and three-dimensional structures of drug molecules and they might be applied to design of new drugs.

4

Methods

- The analysis of the structures has been carried out on two levels, using graph properties and atomic properties.
- Two classification:
  - G
  - G--->Atomic Properties

5

Methods

- Graph representation of a molecule is gotten by considering each atom to be a vertex and each bond to be an edge on a graph.
- Graph properties of molecules are the connectivity properties of atoms representing a molecule.
- Atomic properties include such information as element type, atomic hybridization, and atomic charge.

6

Methods

graph representation of a molecule (pyridine)

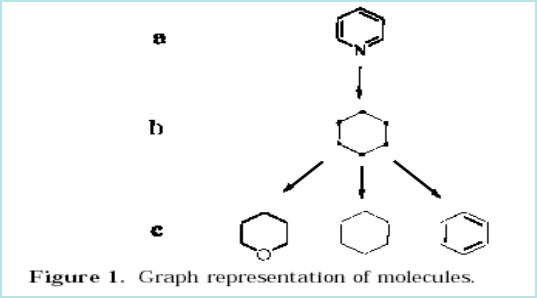


Figure 1. Graph representation of molecules.

Pyridine, Benzene, cyclohexane, and pyran

7

Methods

structural elements of a molecule graph

- Ring Systems: cycles within graph representation of molecules and cycles sharing an edge, benzene.
- Linker Atoms: atoms that are on the direct path connecting two ring systems.
- Side Chain Atoms:
- Framework: union of ring systems & linkers in a molecule.

8

Methods

structural elements of a molecule graph

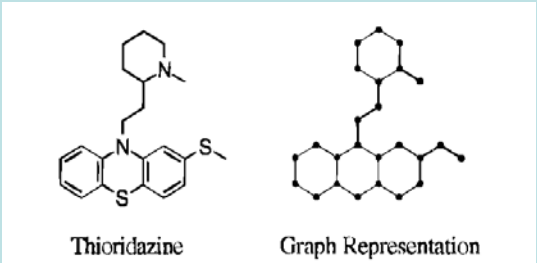


Figure 2. Graph representation of a typical drug molecule.

9

Methods

structural elements of a molecule graph

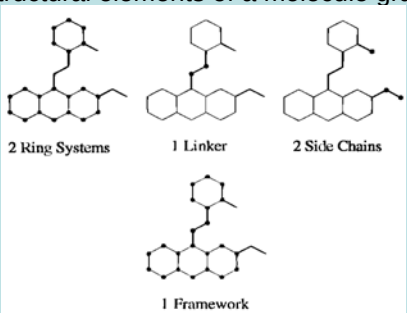


Figure 3. Distinguishing between ring systems, linkers, and side chains.

10

Methods

structural elements of a molecule graph

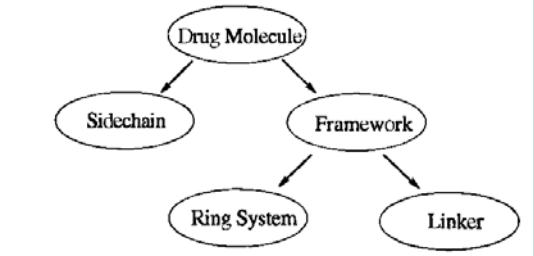


Figure 4. Hierarchical description of molecules.

11

Framework

Methods (Framework part)

algorithm for identification of linker & framework

- Identifying side chain atoms first: each atoms bonded to only one other atom is identified as a side chain atom and removed from the molecule; repeat. The remaining atoms are framework atoms.
- Ring systems are identified using depth-first search.
- Linker atoms.

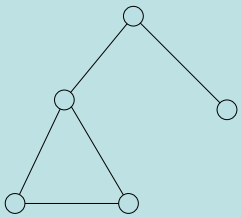
12

Depth-first Search of a Graph

Edges are explored out of the most recently discovered vertex v that still has unexplored edges leaving it. When all of v's edges have been explored, the search "backtracks" to explore edges leaving the vertex from which v was discovered.

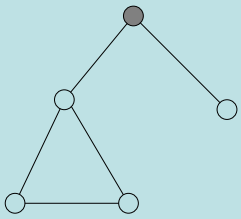
13

Depth-first Search of a Graph



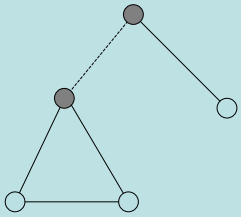
14

Depth-first Search of a Graph



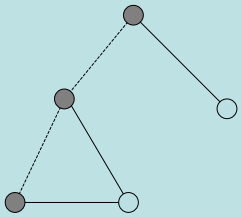
15

Depth-first Search of a Graph



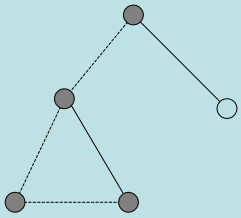
16

Depth-first Search of a Graph



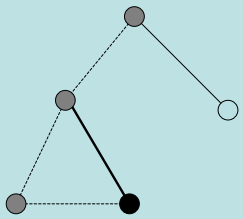
17

Depth-first Search of a Graph



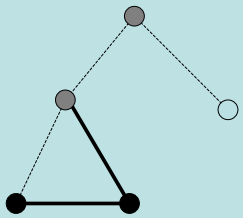
18

Depth-first Search of a Graph



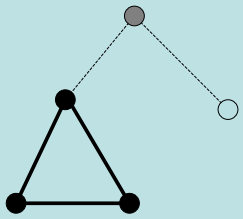
19

Depth-first Search of a Graph



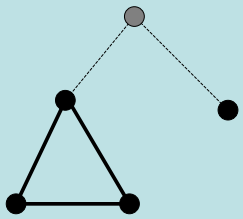
20

Depth-first Search of a Graph



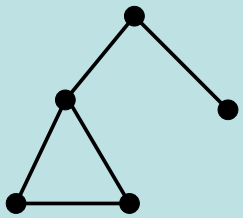
21

Depth-first Search of a Graph



22

Depth-first Search of a Graph



23

Data (Framework part)

- All compounds for which no therapeutic activity class was given, as well as such compounds: radiopaque agents, contrast agents, solvents, anesthetics, disinfectants, topicals, ….
- After the above process, the CMC database had 5120 remaining entries.

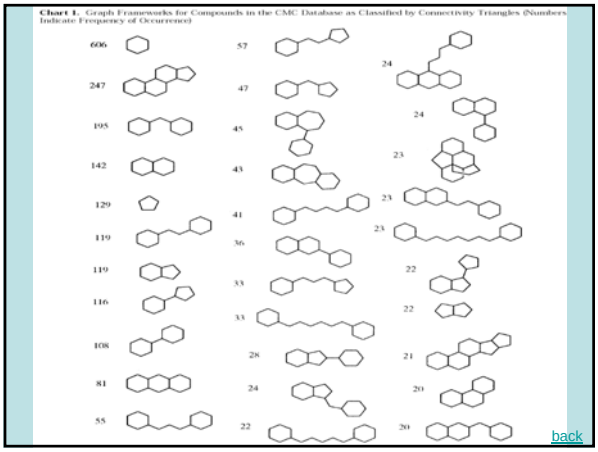
24

Results (Framework part)

the graph theory analysis

- There are 1179 different frameworks among the 5120 compounds analyzed.
- Of these framework, 783 (66%) are unique.
- There are 32 frameworks which exist in at least 20 drugs. This set of frameworks accounts for 50% of the total.
- Acyclic molecules account for 306 (6%) of the total.

atomic 25



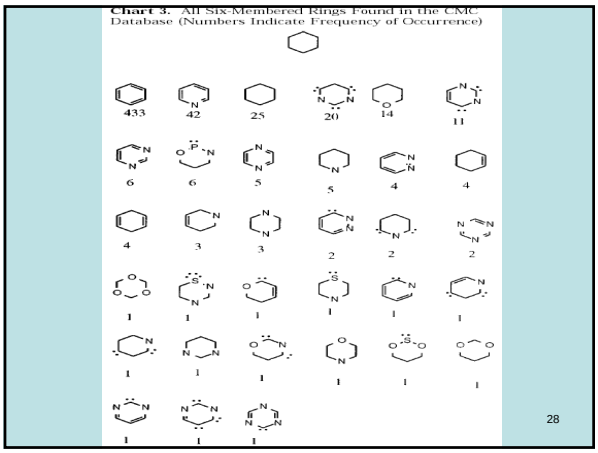
back

Results (Framework part)

relation between graph & atomic property

- the graph theory frameworks can be viewed as providing a generic classification scheme, and the atom property-based framework further subdivide classes of framework based on their chemical properties.

27



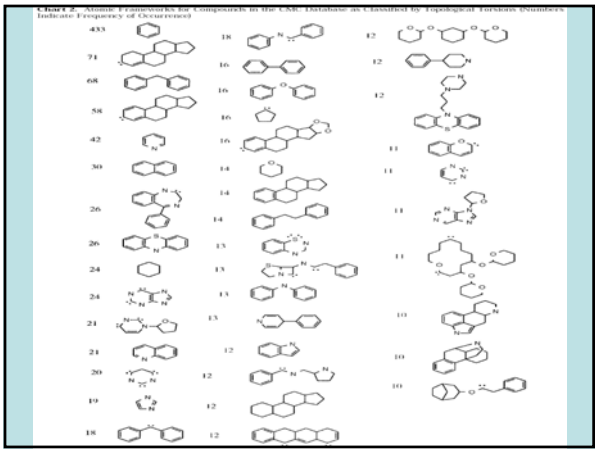
Results (Framework part)

the atomic property analysis

- Atom properties (atom type, hybridization, and bond order) are considered.
- There are 2506 different frameworks
- A large majority of these frameworks (1908, or 76%) are unique.
- The 41 frameworks that occur at least 10 times accounts for 1235 (24%) of the total.

graph

29



Discussion (Framework part)

- The preferences for certain shapes do not necessarily reveal some fundamental truth about drugs, receptors, metabolism, or toxicity.
- Instead, it may reflect the constraints due to synthetic or patent considerations, cost, or the tendency to make new compounds which are structurally similar to known compounds.

31

Discussion (Framework part)

- Half of the known drugs fall into only 32 shape categories (graph theory). The drugs are quite different in their chemical properties.
- The analysis result might be used to direct a de novo design effort.

32

Side Chains

- Properties of Known Drugs: 2. Side Chains, 1999, JMC

  by Guy W. Bemis and Mark A. Murcko

33

Methods (Side Chains)

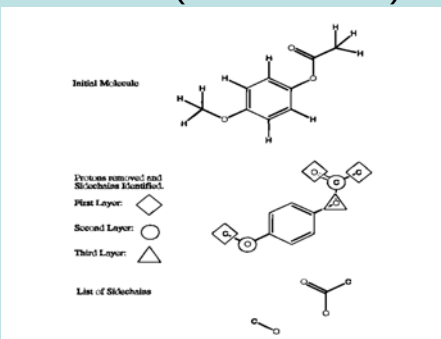


Figure 1. Example of the procedure for determination of side chains in a simple molecule.

34

Methods (Side Chains)

- But it is important to recognize both chain patterns and the way these patterns are connected to molecule frameworks
- Augmented side chains..

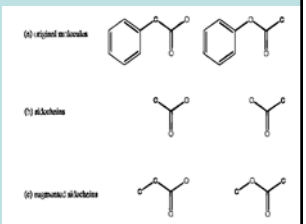


Figure 2. Example of side chain overlap problem. Side chains (b) determined from molecules (a) are identical. Augmented side chains (c) are necessary to distinguish these.

35

Methods (Side Chains)

- But it is important to recognize both chain patterns and the way these patterns are connected to molecule frameworks
- Augmented side chains..

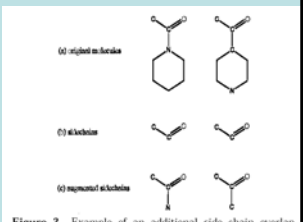


Figure 3. Example of an additional side chain overlap problem. Side chains (b) determined from molecules (a) are identical. Augmented side chains (c) are again necessary to distinguish these.

36

Methods (Side Chains)

- But it is important to recognize both chain patterns and the way these patterns are connected to molecule frameworks
- Augmented side chains..
- Labeled side chains.

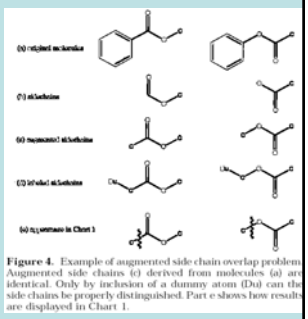


Figure 4. Example of augmented side chain overlap problem. Augmented side chains (c) derived from molecules (a) are identical. Only by inclusion of a dummy atom (Du) can the side chains be properly distinguished. Part e shows how results are displayed in Chart 1.

37

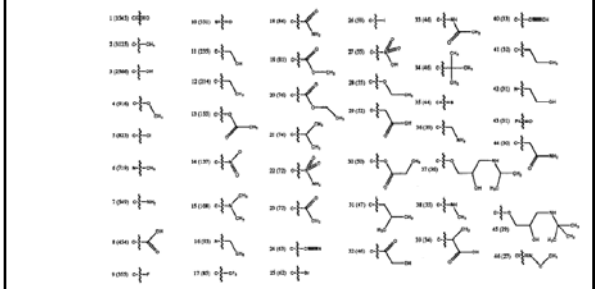
Data (side chains)

- Part of the undesirable compounds are removed from CMC database.
- 5090 entries remains.

38

Results (side chains)

Chart 1. Labeled Side Chains from CMC[a]



[a] First number indicates rank order among drug side chains; number in parentheses is frequency of occurrence.

The most frequently occurring drug side chains

39

Results (side chains)

- A total of 4689 contains labeled side chains.
- There are a total of 18664 labeled side chains (including C=O) attached to these scaffolds. The “average” drug scaffold contains 4 labeled side chains.

40

Most scaffolds have between one to five labeled side chains.

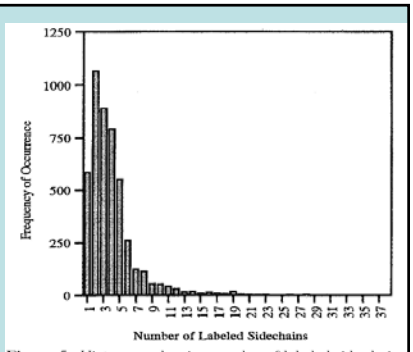


Figure 5. Histogram showing number of labeled side chains per drug framework.

41

Of these labeled side chains the vast majority contain one heavy atom in their actual side chain component.

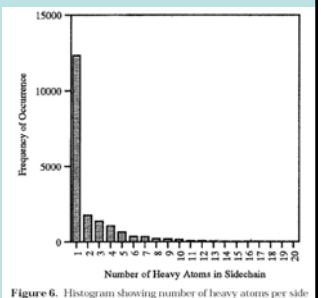


Figure 6. Histogram showing number of heavy atoms per side chain (number of atoms to the right of wavy line as shown in Chart 1).

42

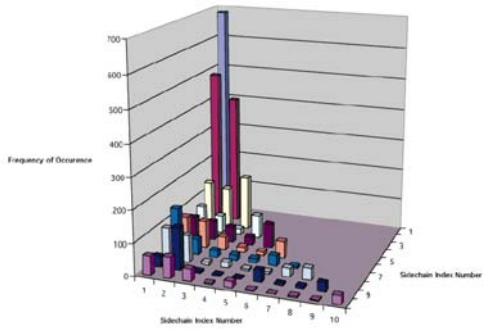


Figure 7. Labeled side chain pair distribution for the 10 most common labeled side chains. Side chain index number corresponds to rank order for labeled side chains shown in Chart 1.

43

Table 2. Most (top) and Least (bottom) Commonly Found Labeled Side Chain Pairs from Figure 7

side chain pair	frequency
Most Commonly Found	
C=O/C=O	686
C=O/C−CH_3	488
C−CH_3/C−CH_3	412
C−OH/C−OH	172
C=O/C−OH	139
C=O/C−NH_2	138
C−CH_3/C−F	131
C−CH_3/C−OH	127
C=O/C−CO_2H	98
C=O/N−CH_3	89
Least Commonly Found	
C−NH_2/S=O	1
C−Cl/C−F	3
N−CH_3/S=O	3
C−F/S=O	3
C−OCH_3/C−F	4
C−OCH_3/S=O	4
C−NH_2/C−F	4
C−CO_2H/S=O	4

44

thanks

45