Molecular Informatics: A finger in every Pie
Informatics...a ‘new’ word ...

- [informat(ion) + -ics.]
  - 23 million hits on Google
  - The central notion is the transformation of information

- For Molecular Informatics
  - I would add the collection, transformation and visualisation of chemical data to extract a deeper knowledge of the underlying properties of the data.
They knew about 'molecular informatics' before the word was coined.

Max Perutz and John Kendrew admire the structure of haemoglobin, and Watson and Crick with DNA.
The fundamental idea is that the use of models, built from experimental data and theory, can profoundly influence our philosophy of science - this is what we spend most of our time doing with chemical data, and it's easier now with computers - especially as data is available as never before...
A hundred BILLION (Chemoinformatics) challenges and opportunities

Biological data
- 91,170,934,635 nucleotide bases
- over 800 organisms
- 510792 protein x-ray, nmr crystal structures
- 23 million citations
- 40 main stream databases from EBI
- Ensemble : 24 million gene predictions

Chemical data
- 95 million chemical substances
- 3,700,000 chemical reactions
- 513,000 available reagents
- Bielstein has 600,000 reaction abstracts
- >270,000 organic x-ray structures

Patents
- European patents – 150,000,000 pages
- 150,000 applications/year
- 450,000 chemical patent hyperstructures
- Over 100 countries in patent cooperation treaty (PCT)

This is just some of the biological, chemistry and patent space – it’s very big! The role of molecular informatics is the collection, analysis and interpretation of these data.
Methods for the storage of molecular data

- Experimental Measurements
- Computation of properties
- Simulation

The Internet
- Data harvesting
- Data processing
- E-science

The other informatics sciences
- Bioinformatics
- ‘Omics, materials, Polymers.....

Data Pipelining

Statistical methods
- Signal processing

Machine Learning
- Chemometrics
- Pattern recognition

Visualisation

Compound selection
- And design

Structure Activity Relationships
- Structure Property Relationships
- Design rules...
If you are searching the literature, SciFinder Scholar results:-

Chemoinformatics and Cheminformatics (664 entries) from 1996-2007

Bioinformatics (35,962 entries)
1999-2007

Chemistry AND informatics (1965-2007)
4,795 entries
Molecular Informatics is, of course, pervasive, and refers to all aspects of data in molecules, and how it is stored and analysed, so we could potentially be here for the next three hours (and miss out on the pie)…

But let’s look at two aspects:

- **The Semantic Web and data**

- **Models from data**
- Potentially, we can now store nearly all of our data in computers so....
- This is central: particularly when we turn data into models, how do we know the models are real or even useful?
- They need
  - the availability of all the necessary data
  - and a relevant validated analysis

---

Henri Poincaré (mathematician)

“Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house”
Which brings us back to informatics...

- This leads to the first problem – what chemistry data is stored in the databases?

- What we actually store is often a very reduced data set. Examples are Smiles, SD files, pDB files, Inchi, etc.. (and also, there is an industry in converting one file format to the other, see OpenBabel, [http://openbabel.org/wiki/Main_Page](http://openbabel.org/wiki/Main_Page))

- Here’s a couple of examples
Some examples of Smiles

- **ortho-fluorophenol**
  \[ \text{H} - \left( \begin{array}{c} \text{F} \\ \text{H} \end{array} \right) \]

- **pyridine-N-oxide**
  \[ \text{[O-][n+]} - \left( \begin{array}{c} \text{H} \\ \text{N} \end{array} \right) \]

- **cis-resorcinol**
  \[ \text{H} - \left( \begin{array}{c} \text{H} \\ \text{OH} \\ \text{OH} \\ \text{OH} \end{array} \right) \]

- **Propranolol** – a beta-blocker
  \[ \text{C} - \left( \begin{array}{c} \text{C} \\ \text{N} \\ \text{C} \end{array} \right) - \left( \begin{array}{c} \text{H} \\ \text{O} \\ \text{C} \end{array} \right) - \left( \begin{array}{c} \text{C} \\ \text{C} \end{array} \right) - \text{H} - \text{C} \]

Very compact, clever, Covers most organics, Easy to interpret, widely used
An SD file of Alanine: a common standard format; contains additional 3Dimensional and property information

L-Alanine (13C)
GSMACCS-1110169115362D 1 0.00000 0.00000 0

6 5 0 0 1 0 3 V2000
-0.6622 0.5342 0.0000 C 0 0 2 0 0 0
0.6220 -0.3900 0.0000 C 0 0 0 0 0 0
-0.7207 2.0917 0.0000 C 1 0 0 0 0 0
-1.8622 -0.3995 0.0000 N 0 3 0 0 0 0
0.6220 -1.8037 0.0000 C 0 0 0 0 0 0
1.9454 0.4244 0.0000 C 0 5 0 0 0 0
1 2 1 0 0 0
1 3 1 0 0 0
1 4 1 0 0 0
2 5 2 0 0 0
2 6 1 0 0 0
M CHG 2 4 1 6 -1
M ISO 1 3 13
M END

x,y,z symbol, mass diff, charge, stereo, h-count....
But, the future of chemical data formats lies elsewhere...

- Old formats are pre-defined and fixed

- They are not *extensible*: This means they cannot be added to when new information appears, and they do not adhere to web standards, so important today for interoperability of software.

- This has led to the development of Chemical Markup Language, CML, which is written and defined in the Web standard language XML. (Peter Murray-Rust and Henry Rzepa)
Here are a couple of examples (available on the internet)

- **Project Prospect** at the Royal Society of Chemistry uses CML and XML to add information to documents – the documents are ‘Marked up’

  http://www.rsc.org/Publishing/Journals/ProjectProspect/Examples.asp

- **Crystaleyeye** reads X-ray data from documents and uses XML to create a new document with alerts of new structures (you can try this out)

  http://wwmm.ch.cam.ac.uk/crystaleyeye/
Project Prospect

See science come alive with a new initiative that allows papers published online in RSC journals to become machine-readable.

Features
How to access the enhanced features of our articles

Examples
View sample enhanced articles

Project Prospect News

Project Prospect's a big hit with authors and readers
27 February 2007
Since launch at the beginning of February, RSC has been delighted at the praise received for Project Prospect from the scientific community.

Project Prospect featured in webzine interview
27 February 2007
Read an interview with Robert Parker, Acting Managing Director of RSC Publishing, in popular webzine, Reactive Reports.

RSC Publishing pioneers next generation of enriched articles
01 February 2007
A new project allows data in RSC journal papers to be read, indexed and intelligently searched by machine: a first step towards the “semantic web”.

http://www.rsc.org/Publishing/Journals/ProjectProspect/index.asp
Beauty, Staker, O'Connor, Fargnoli, and Stephen L. Roser

Structure and Stability of DPPE Planar Blayers

SoMater. 2007, 3, 214-222; DOI: 10.1039/b712362a

Soft Matter
Introduction

To exploit certain types of surface and interfacial techniques in biological cell membrane studies it is necessary to use planar lipid bilayers located on or near to a substrate. These planar systems have allowed the application of AFM,3 impedance analysis,2 surface plasmon resonance,4 neutron reflectivity,5 and ellipsometry.6 Examples of types of planar systems range from single bilayers adsorbed or deposited onto substrates6,9 to multilamellar stacked bilayers10 to elaborate polymer supported bilayer11,12 and hybrid bilayers with one leaflet of lipids and the other of alkaneetholyl.13

One of the main advantages of single bilayer samples compared to multilamellar samples is that the information obtained is bilayer specific, it is not the average of hundreds or thousands of bilayers. Another advantage is that very low concentrations are needed to fabricate these samples. This is obviously interesting when expensive components are being studied. Often the main disadvantage of single bilayer systems is that the substrate can exert a restraining force upon the bilayer, inhibiting its phase behavior. Another key disadvantage is that only a very thin water layer separates the bilayer from the substrate (5–10 Å). This can restrict the inclusion of transmembrane proteins that protrude either side of the bilayer.14

To overcome some of the problems associated with single bilayer systems, a new type of planar membrane system has been developed. It consists of a bilayer floating above another bilayer that is in close proximity to the substrate.4 Fabrication is achieved by a combination of Langmuir-Blodgett and Langmuir-Schafer depositions. These techniques enable the fabrication of asymmetric bilayers where the composition of each leaflet can be selected to model the asymmetric nature of real membranes. Compared to single bilayer samples the upper bilayer is less constrained and is open to a large reservoir of water, making it ideally suited to study transmembrane phenomena and translocation. When the double bilayers are fabricated with phosphatidylcholines, the upper bilayer is separated from the lower bilayer by a water layer of 20–30 Å, whilst the lower bilayer is separated from the substrate by a water layer of 5–10 Å.15 The upper bilayer exhibits comparable gel, transition and fluid phase-behaviour to vesicles in solution.16 During phase-behavior studies, the main water layer was found to swell around the main transition temperature. This was interpreted in terms of competition between the inter-bilayer potential and membrane fluctuations and used to estimate the bending rigidity of the bilayer.17 Off-specular neutrons radiation measurements have allowed the measurement of the bending modulus and tension of the floating bilayer.18 Incorporation of low concentrations of cholesterol (1–5 mol%) was found to progressively decrease the swelling,19 and at a concentration of 10 mol% swelling was completely removed.20

Together with phosphatidylcholines, phosphatidylethanolamines are one of the most abundant components of the lipid bilayer in membranes and can even be found to account for up to a third of the total percentage of lipids present, as in the case of human and the rat erythrocyte plasma membranes. They are often found asymmetrically distributed in membranes, being predominantly located in the inner cytoplasm-facing leaflet. As well as being one of the major building blocks of membranes, they also have specific tasks such as supporting active transport by the lactose permease. They also act as a chaperone during the assembly of membrane proteins, guiding the folding path and aiding in the transition from the cytoplasmic to the membrane environment. In comparison to phosphatidylcholines, the smaller head-group of phosphatidyllethanolamines enables stronger hydrogen bonds between the phosphate oxygen and the primary amine parts of the lipid.21 Despite their predominance, phosphatidylethanolamines have not received as much attention as phosphatidylcholines in the literature. The phase-behavior of phosphatidylethanolamines vesicles has been well characterized, whilst literature on the behaviour of stacked multilamellar bilayers is rather limited. With a view to the application of phosphatidylethanolamine double bilayers as planar biomembrane mimics, the phase-behaviour of DEPE single and double bilayers was investigated by neutron reflectivity.22 Double bilayers with a ratio of 9:1 DEPE/cholesterol were also studied with the dual purpose of assessing the stabilizing effect of cholesterol on the upper bilayer and to increase the realism of the mimic by increased number of components. We have already shown that asymmetric double bilayers containing DEPE can be prepared and are stable in both the gel and fluid phases.23
Introduction

To exploit certain types of surface and interfacial techniques in biological cell membrane studies it is necessary to use planar lipid bilayers because in planar systems have allowed the application of AFM, impedance analysis, surface plasmon resonance, neutron reflectivity, and ellipsometry. Single bilayers range from single bilayers adsorbed or deposited onto substrates to multilamellar stacked bilayers to elaborate polymer systems with one leaflet of lipid and the other of alkane thiol.

One of the main advantages of single bilayer samples compared to multilamellar samples is that the information obtained is bilayer specific. Another advantage is that very low concentrations are needed to fabricate these samples. This is obviously interesting studied. Often the main disadvantage of single bilayer systems is that the substrate can exert a restraining force upon the bilayer, inhibiting its disadvantage is that only a very thin water layer separates the bilayer from the substrate (5–10 Å). This can restrict the inclusion of transmembrane proteins or the bilayer can be formed on either side of the bilayer.

To overcome some of the problems associated with single bilayer samples, a new type of planar membrane system has been developed. In the arrangement above another bilayer that is in close proximity to the substrate. Fabrication is achieved by a combination of Langmuir–Blodgett and Langmuir–Schaefer techniques, the bilayer is made to span a liquid-crystalline phase that is separated from the bilayer by a water layer of 20–30 Å, which forms from the substrate by a water layer of 5–10 Å. The upper bilayer exhibits comparable gel, transition, and fluid phase behaviour to vesicles in solution studies. The main water layer was found to swell around the main transition temperature. This was interpreted in terms of the vesicles in the liquid crystalline phase being more disordered and the liquid crystalline phase being more disordered and the disordered liquid crystalline phase that is more disordered (1–6 mol%) was found to progressively decrease the swelling was completely removed. Together with phosphatidylcholines, phosphatidylethanolamines are one of the most abundant components of the lipid bilayer in membranes up to a third of the total percentage of lipids present, in the case of human and the outer erythrocyte plasma membranes. They are often found in membranes being predominantly located in the inner cytoplasmic-facing leaflet. As well as being one of the major building blocks of membranes supporting active transport by the lactate permease. They also act as a chaperone during the assembly of membrane proteins, guiding the folding process and terminating the transition from the cytoplasmic to the membrane environment. In comparison to phosphatidylcholines, the smaller head-group of phosphatidylethanolamines enables stronger hydrogen bonds between the phosphate oxygen and the primary amine parts of the lipids. Despite their predominance, phosphatidylethanolamines have not received as much attention as phosphatidylcholines in the literature. The phase-behaviour of phosphatidylethanolamines has been well characterised, whilst literature on the behaviour of stacked bilayers is rather limited. With a view to the application of phosphatidylethanolamine bilayers as planar biomembrane mimics, the phase-behaviour of DPPE single and double bilayers was investigated by neutron reflectivity. Double bilayers with a ratio of 9:1 DPPE/cholesterol were also studied with the dual purpose of assessing the stabilising effect of cholesterol on the upper bilayer and to increase the realism of the mimic by increased number of components. We have already shown that asymmetric double bilayers containing DPPE can be prepared and are stable in both the gel and fluid phases.
Compound Information for cholesterol

Synonyms:
- cholesterol
- Cholesteryl
- Cholesterol
- cholest-5-en-3beta-ol
- CHOLESTEROL

SMILES:
\[ H\{C@@1\}(CC\{C@@2\}3\{E\}\{C@@3\}3\{E\}CC=O\}C\{C@4\}C\{C@4\}C\{C@4\}C\{C@4\} \]

InChI:
InChI=1/C17H34O6/c1-18(2)7-6-8-19(3)23-11-12-24-22-10-9-20-17-21(28)13-15-26(19)

CML (Chemical Markup Language Representation)

2-D Image

Other articles referencing this compound:

Supporting active transport by facilitated transporters. They also act as a gatekeepers during the assembly of membrane proteins, guiding the formation of membranes and facilitating the transition from the cytosol to the membrane environment. In comparison to phosphatidylcholines, the smaller head-group of phosphatidylethanolamines enables stronger hydrogen bonds between the phosphate oxygen and the primary amine parts of the lipids. Despite their predominance, phosphatidylethanolamines have not received as much attention as phosphatidylcholines in literature. The phase-behaviour of phosphatidylethanolamines vesicles has been well characterised, whilst literature on the behaviour of stacked multilamellar bilayers is rather limited. With a view to the application of phosphatidylethanolamines, double bilayers as planar biomembrane mimics, the phase-behaviour of DPPE single- and double bilayers was investigated by neutron reflectivity. Double bilayers with a ratio of 9:1 DPPE:cholesterol were also studied with the dual purpose of assessing the stabilising effect of cholesterol on the upper bilayer and to increase the realism of the mimics by increased number of components. We have already shown that asymmetric double bilayers containing DPPE can be prepared and are stable in both the gel and fluid phases.
Introduction

To exploit certain types of surface and interfacial techniques in biological systems, planar substrates have allowed the application of AFM. Impedance spectroscopy is also a powerful technique that is able to characterize the electrical properties of the bilayer and the surrounding environment. The system is composed of two layers of phospholipids with one leaflet of lipids and the other of alkane tails. One of the main advantages of single bilayer samples compared to double layers is the ease with which they can be studied. Often the main disadvantage of single bilayer systems is the disadvantage that only a very thin water layer separates the bilayer from the substrate by a water layer of 5–10 Å. In studies with the main water layer was found to swell around the main transition temperature. This was interpreted in terms of membrane fluctuations and used to estimate the bending rigidity of the bilayer. Incorporation of low concentrations of cholesterol (1–5 mol%) was found to concentration of 10 mol% swelling was completely removed.

Together with phosphatidylcholines, phosphatidylethanolamines are one of the most abundant components of the lipids up to a third of the total percentage of lipids present, in the case of human and the rat erythrocyte plasma membranes. Membranes, being predominantly located in the inner cytoplasm-facing leaflet. As well as being one of the major building blocks, phosphatidylethanolamines are a common among the primary parts of the lipids. Despite their predominance, phosphatidylethanolamines have been well characterized. Multimullerian-like bilayers is rather limited. With a view to the application of phosphatidylethanolamine double bilayers as planar bilayer membrane, the phase behavior of DPPC in the occurrence of DPPC in single and double bilayers was investigated by neutron reflectivity. Double bilayers with a ratio of 9:1 DPPC:cholesterol were also studied with the dual purpose of assessing the stabilizing effect of cholesterol on the upper bilayer and to increase the realism of the mimic by increased number of components. We have already shown that asymmetric double bilayers containing DPPC can be prepared and are stable in both the gel and fluid phases.
An example with a patent (from Peter Corbett, developer of the software OSCAR3)
QUATERNARY AMMONIUM SALTS AS ANTI-AGING ACTIVES IN PERSONAL CARE COMPOSITIONS

A personal care product is provided which includes a package filled with a personal care composition and instructions printed on or associated with
the package indicating topical use of the composition on skin for purposes
of controlling the signs of aging. The composition includes a quaternary
ammonium compound selected from (a) salts of hydroxypropyltri(C1-C
alkyl)ammonium mono-substituted monosaccharides; (b) salts of
hydroxypropyltri(C1-C alkyl)ammonium mono-substituted polyols, the salts
having a cation of an average molecular weight no higher than 450 and the
salt having a Thy no higher than 10 deg. C; (e) dihydroxypropyltri(C1-C
alkyl)ammonium salts; (d) chlorhydroxypropyltri(C1-C alkyl)ammonium
salts; and (e) mixtures thereof. QUATERNARY AMMONIUM SALTS AS ANTI-AGING
ACTIVES IN PERSONAL CARE
COMPOSITIONS Technical Field of the Invention [0001] The invention
concerns quaternary ammonium salts, particularly such salts of polyols, in
personal care compositions for purposes of delaying onset and treating the
signs of aging. Background of the Invention [0002] Forever young,
adults as they age seek to preserve the indicia of youth. Through the ages
cosmetics have proved valuable for retarding the signs of the aging
process. Facial foundations, creams and lotions have all helped in the
cover up. Yet few really effective actives are available in the cosmetic
chemist's arsenal. [0003] Two classes of materials have been
clinically proven as providing some relief from the signs of aging.

Advanced: Enter an OscarFlow command here (or leave blank):

Submit  Reset
The chemistry is detected, annotated and converted to CML (note: the subsets of data types, experimental. Ontology term, reaction, etc.)

QUATERNARY AMMONIUM SALTS AS ANTI-AGING ACTIVES IN PERSONAL CARE COMPOSITIONS

A personal care product is provided which includes a package filled with a personal care composition and instructions printed on or associated with the package indicating topical use of the composition on skin for purposes of controlling the signs of aging. The composition includes a quaternary ammonium compound selected from (a) salts of hydroxypropyl(C1-C3) mono-substituted monoammonium, (b) salts of hydroxypropyl(C1-C3) alkylammonium mono-substituted polyelectrolytes, the salt having a ratio of an average molecular weight no higher than 450 and the salt having a Tg no higher than 10°C, (c) hydroxypropyl(C1-C3) alkylammonium salt, (d) chlorohydroxypropyl(C1-C3) alkylammonium salt, and (e) mixtures thereof. QUATERNARY AMMONIUM SALTS AS ANTI-AGING ACTIVES IN PERSONAL CARE COMPOSITIONS Technical Field of the Invention: [0001] The invention concerns quaternary ammonium salts, particularly such salts of polyelectrolytes, in personal care compositions for purposes of delaying onset and treating the signs of aging. Background of the Invention: [0002] Forever young. Adults seek and preserve the signs of youth. Throughout the ages, cosmetics have proved valuable for reducing the signs of aging. Foundation, cream, and lotion have all helped in this regard. Yet few really effective agents are available in the cosmetic community's arsenal. [0003] Two classes of materials have been clinically proven as providing some relief from the signs of aging. Alpha-hydroxy acids, or their derivates, are used widely in cosmetic commerce. Amino acids are US Patent 5,091,171 [Vü et al.], Retinol (Vitamin A) is an exogenous compound which occurs naturally in the human body. This material and its derivatives have been used extensively in the treatment of a variety of skin disorders and as repair or renewal agents [0004]. Both alpha-hydroxy acids and retinol as well as many of their derivatives tend to produce a skin rash and even reduce the skin when present at levels sufficient to be effective. Consumers would of course prefer performance without side effects [0005]. Accordingly, there remains a need for materials which can be effective against the signs of aging and that yet have no adverse side effects. Summary of the Invention: [0006] A personal care product is provided which includes: (A) a package filled with a personal care composition which includes: (c) from 0.1 to 30% by weight of a quaternary ammonium compound selected from the group consisting of: (a) salts of hydroxypropyl(C1-C3) mono-substituted monoammonium, (b) salts of hydroxypropyl(C1-C3) alkylammonium mono-substituted polyelectrolytes, the salt having a ratio of an average molecular weight no higher than 450 and the salt having a Tg no higher than 10°C, (c) hydroxypropyl(C1-C3) alkylammonium salt, (d) chlorohydroxypropyl(C1-C3) alkylammonium salt, and (e) mixtures thereof; (b) from 1 to 59.9% by weight of a cosmetically acceptable carrier, and (2) instructions printed on or associated with the package indicating topical use of the composition on skin for purposes of controlling the signs of aging. Detailed Description of the Invention: [0007] Now it has been found that certain types of quaternary ammonium salts, particularly certain such salts of polyelectrolytes, can control the signs of human skin agin by delaying onset and treating the signs of aging. A first category of useful quaternary ammonium compounds are salts of hydroxypropyl(C1-C3)monoammonium mono-substituted monoammonium. There can be prepared by a variety of procedures. Most preferred via reaction of 2-hydroxy-3-chloropropyltrimethylammonium chloride with a monoammonium in an approximately 1:1 molar ratio in an alkaline medium. By typical Wilkinson cyclization, sodium chloride is eliminated through forming an etherlike linkage between the hydroxypropyl end of the quaternary ammonium compound and the monoammonium [0009]. Monoammonium, particularly reducing and non-reducing cyclic monoaainoniums, are the smallest carbohydrate molecules incorporating the four-, five-, and six-membered sugars. Illustrative monoammoniums are those. Deoxyribose, thymidine, ribose, arabinose, rhamnose, xylulose, xylose, allose, talose, and mannose and their derivatives and isomers. Most preferred are sugars and inositol as the monoammonium moiety which is to be substituted with the hydroxypropyltrimethylammonium group. [0010] Ordinarily, the C3 to C5 alkyl constituent on the quaternary ammonium group will be methyl, ethyl, or propyl, isopropyl or hydroxymethyl and mixtures thereof. Particularly preferred is a trimethyl ammonium group known through INCI nomenclature as a "trimmonium" group. Any ratio can be used in the quaternary ammonium. The union may be organic or inorganic with the premise that the material is cosmetically acceptable. Typical inorganic anions are halides, sulphates, phosphates, tartrates, and borates. Most preferred
The highlighted chemical name is automatically detected, converted to CML, and the structure and associated data is now added.

A poster and demonstration is available from Peter Corbett.
Example: a search of PubMed for ‘grapefruit’ – fetched 200 abstracts - chemistry marked up

- Cecal parameters of rats fed diets containing grapefruit polyphenols and inulin as single supplements or in a combination.
- Interaction of grapefruit juice and calcium channel blockers.
- Identification of isomeric flavonoid glucuronides in urine and plasma by metal complexation and LC-ESI-MS/MS.
- UV-irradiated Grapefruit Juice Loses Pharmacokinetic Interaction with Nifedipine in Rats.
- Effect of surface waxes on the persistence of chlorpyrifos-methyl in apples, strawberries and grapefruits.
- Does gender, food or grapefruit juice alter the pharmacokinetics of primaquine in healthy subjects?
- Development and validation of an HPLC/UV/MS method for simultaneous determination of 18 preservatives in grapefruit seed extract.
- Delayed effect of grapefruit juice on pharmacokinetics and pharmacodynamics of tacrolimus in a living-donor liver transplant recipient.
- Therapeutic drug monitoring: A pharmacotherapeutic tool in psychiatry
  - A furanocoumarin-free grapefruit juice establishes furanocoumarins as the mediators of the grapefruit juice-felodipine interaction.
- Effect of grapefruit juice on the disposition of mandipine enantiomers in healthy subjects.
- Biological and physical approaches to improve induced resistance against green mold of stored citrus fruit.
- Narirutin does not alter caffeine pharmacokinetics, energy expenditure, or cardiovascular haemodynamics in humans following caffeine consumption.
- Design, synthesis and evaluation of furanocoumarin monomers as inhibitors of CYP3A4.
- The effects of grapefruit on weight and insulin resistance: relationship to the metabolic syndrome.
- Modelling intestinal absorption of salbutamol sulphate in rats.
- Grapefruit juice and potential drug interactions.
- Nonvolatiles of commercial lime and grapefruit oils separated by high-speed countercurrent chromatography.
- Pomelo juice, but not cranberry juice, affects the pharmacokinetics of cyclosporine in humans.
- Effect of extended exposure to grapefruit juice on cytochrome P450 3A activity in humans.
Search Form
(look for structures that contain piperidine)

OSCAR3 Search

Query: ClCCCCN1
Type:
- Plain text
- InChI
- SMILES
- SMILES Substructure
- SMILES Similarity

Similarity: Top 5 matches

- snippets - document titles and search terms in their context
- compoundsList - all compounds in the documents found
- hitsList - compounds found by the search only
Influence of hepatic and intestinal cytochrome P4503A activity on the acute disposition and effects of oral transmucosal fentanyl citrate.

Influence of hepatic and intestinal cytochrome P4503A activity on the acute disposition and effects of oral transmucosal fentanyl citrate. BACKGROUND: Oral transmucosal fentanyl citrate (OTF) was developed to provide rapid analgesia and is specifically approved for treating breakthrough cancer pain. Fentanyl in OTF is absorbed across the oral mucosa, but a considerable portion is swallowed and absorbed enterally. Fentanyl metabolism is catalyzed by cytochrome P4503A4 (CYP3A). The role of intestinal or hepatic first-pass metabolism and CYP3A activity in OTF disposition.

ampin, hepatic/intestinal CYP3A inhibition by troleandomycin, selective intestinal CYP3A inhibition by grapefruit juice, or nothing (control). Plasma fentanyl and norfentanyl concentrations were determined by mass spectrometry. Fentanyl effects were measured by dark-adapted pupil diameter and subjective self-assessments using visual analog scales. RESULTS: Peak plasma fentanyl concentrations, time to peak, and maximum pupil diameter change from baseline were unchanged after rifampin, troleandomycin, and grapefruit juice. Fe

significantly affected by CYP3A alterations. After control, rifampin, troleandomycin and grapefruit juice, respectively, area under the curve of plasma fentanyl versus time was 5.9 +/- 3.7, 2.2 +/- 0.8, 10.4 +/- 8.9, 10.4 +/- 3.3 h x ng/ml; norfentanyl/fentanyl plasma area under the curve ratios were 0.92 +/- 0.63, 3.2 +/- 1.8, 0.08 +/- 0.14, 0.08 +/- 0.33 (P < 0.05 versus control). DISCUSSION: Peak fentanyl concentrations and clinical effects after OTF were minimally affected by altering both intestinal and hepatic CYP3A activity, whereas fentanyl metabolism, elimination, and duration of effects were significantly affected: selective intestinal CYP3A inhibition had minimal effects. This suggest

More like this

Effects of sweetie juice on nifedipine pharmacokinetics in rats
A Similarity Search

OSCAR3 Search

Query: like limonene
Type:
- Plain text
- InChI
- SMILES
- SMILES Substructure
- SMILES Similarity

Similarity: Top 5 matches

- snippets - document titles and search terms in their context
- compoundsList - all compounds in the documents found
- hitsList - compounds found by the search only
A similarity search for compounds “Like Limonene”

Search Results

Results 1 to 5 of 7:  next

Bioactive compounds of grapefruit (Citrus paradisi Cv. Rio Red) respond differently to postharvest irradiation, storage, and freeze drying.

ced (P < or = 0.05) the lycopene content, but the reduction (P < or = 0.05) in beta-carotene content occurred only in the control fruit. Reduction in d-limonene and myrcene was observed in the irradiated fruits at 6 days after harvest and in the freeze-dried samples. These results warrant testing of the effect of posthar

More like this

Use of novel compounds for pest control: insecticidal and acaricidal activity of essential oil components from heartwood of Alaska yellow cedar.

Iaris Say nymphs, Xenopsylla cheopis (Rothchild), and Aedes aegypti (L.) adults. Four of the compounds from the essential oil have been identified as monoterpenes, five as eremophilane sesquiterpenes, five as eremophilane sesquiterpene derivatives from valencene and nootkatone, and one as a sesquiterpene outside the eremophilane parent group. Carvacrol was the only monoterpenes that demonstrated biocidal activ

efruit extract exhibited the greatest biocidal activity against fleas (LC50 = 0.0029%). Mosquitoes were most susceptible to one of the derivatives of valencene, valencene-13-aldehyde (LC50 = 0.0024%), after 24 h. Bioassays to determine residual activity of the most effective products were conducted at 1, 2,

More like this

Volatile constituents of redblush grapefruit (Citrus paradisi) and pummelo (Citrus grandis) peel essential oils from Kenya.
Moving on to Structure-
Activity/Property models- some observations

- The objective here is to relate measured or computed parameters to some new property, e.g. bioactivity at a target, absorption, melting point, solubility...

- The first issue is data quality.

  - Biological data is always problematic as it is often not possible to reliably reproduce, isolate the variables, combine data. Physical data is easier to measure (in general) and there is a lot more of it.

  - Our experience with a common physical property, solubility (of course....)
How reliable are solubility data?

<table>
<thead>
<tr>
<th>Caffeine solubility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>25</td>
</tr>
</tbody>
</table>

‘Solubility’ in the literature

- Katritzky observed an average standard deviation of 0.58 log units.

- Jorgensen and Duffy suggested the average uncertainty of 0.6 log units - 1.5 log units.
  - data can have wide ranges in the literature: guanine has - 3.58 and 1.86 – take your pick.

- Recent study by Dearden, re-measured 113 organic drug-like compounds,
  - 22 differed by >0.5 log unit
  - 9 differed by >1.0 log unit
  - 1 differed by >2.0 log units
‘Solubility’ in the literature

- Thus, any computational method that gives estimates (usually based on SAR) better than 0.5 log units is over fitted – many are!

- The lit. data usually has no information on the experimental method, the material whose solubility is being studied, or the definition of the reported solubility - and commonly, many datasets are combined to build models.

In this case, we have decided to create our own data and not to combine it with other literature data.
Potentiometric cycling method for very accurate and controlled measurement of solubility

\[
\text{A} \rightarrow \text{A}^+ \rightarrow \text{Na}^+ \rightarrow \text{AH} \rightarrow \text{AH}
\]


We use a Sirius glpKa instrument With a DPAS detector
Solubility Challenge

- *JCIM* Solubility Challenge: Coming Soon! Please check *Journal of Chemical Information and Modeling* for more details.

- We deposit ca. 100 accurate measurements of intrinsic solubility of drug-like molecules.

- You predict 40 unknowns

- *JCIM* publish the ‘best’ attempts.
So, before creating a model (or using someone else’s)

- This is obvious, but....even if you use a pre-computed model, check the data sources
  - Are data compatible, and can they be combined
    - It is often the case that non-compatible data are merged to create a database 'large enough' to do statistics on
  - Is there sufficient background information to determine the model’s relevance
    - The 'ontology' of the information can be vital - what were the units of measurement? (in the solubility example, some have mixed up ug/ml and umol/ml
  - Do they cover the 'chemical property space' required
    - Are my compounds very different from those used in the model?
So, if we have accurate data, what’s in a model?

Molecular database → Calculate/measure molecular parameters → Analysis

This is the most common ‘paradigm for molecular analysis and prediction

\[ \log P = \sum_{i=1}^{N} a_i f_i + \sum_{j=1}^{M} b_j F_j \]

Prediction → Analysis
What’s in a model?

- The objective is usually to select a molecule (e.g. molecular similarity) or predict a property.

- ***All models rely on the variance of the data***
- ***All models are susceptible to database bias***

- That is, the range of data values and their distribution.
  - If the points all had the same value, they would be easy to look up, there would be no model and one prediction for everything.
  - The point is to extract a relationship between calculable parameters and the property of interest.
  - The design of the experiment to obtain the data is therefore very important (and often ignored) - experimental design (Chemometrics can help).
Methods to discover models

- Models are generated using statistical or machine learning methods
  - Statistical methods usually rely on a normal distribution of the data and provide a fit to the data while minimising the error in the fit.
  - Are either supervised (e.g. regression) or unsupervised (e.g. principal components)
  - Machine learning methods are usually heuristic based and nearly all depend on local clustering (classification) - There are lots of flavours...
Methods for Machine Learning…there are many…many…

Modeling conditional probability density functions: regression and classification
- Artificial neural networks
- Decision trees
- Gene expression programming
- Genetic algorithms
- Genetic programming
- Dynamic programming
- Gaussian process regression
- Linear discriminant analysis
- K-nearest neighbor
- Minimum message length
- Perceptron
- Quadratic classifier
- Radial basis function networks
- Support vector machines

Modeling probability density functions through generative models
- Expectation-maximization algorithm
- Graphical models including Bayesian networks and Markov Random Fields
- Generative Topographic Mapping

Approximate inference techniques
- Markov chain
- Monte Carlo method
- Variational Bayes
- Variable-order Markov models
- Variable-order Bayesian networks

Optimization
- Most of methods listed above either use optimization or are instances of optimization algorithms

Meta-learning (ensemble methods)
- Boosting
- Bootstrap aggregating aka bagging
- Random forest
- Weighted majority algorithm

Inductive transfer and learning to learn
- Inductive transfer
- Reinforcement learning
- Temporal difference
- Monte-Carlo method

They can be traced back to the ID3 method of Ross Quinlan – worth a look
Some comments about making models
(includes QSAR, SAR, QSPR…)

- The parameters used to predict a physical property (like solubility and logP) compared to e.g. a binding affinity must often behave in a fundamentally different way.

- Reason: a property like logP in octanol/water is consistent in that the medium doesn’t change.

- However, both the medium (the receptor) and the ligand change upon binding and different ligand/receptor combinations really require different models!
Property behaviour

- So, in property space, we should expect behaviour that was consistent in that it was: linear, exponential, parabolic – i.e. predictable

- However, in SAR space - it’s disjointed and, if we’re lucky, clustered e.g. depending on the mode of binding (if you look at SAR predicted/measured plots in the literature, many join clusters and not compounds)

- So, parameters must have the following ‘property’
  - Small changes in the parameter should produce small changes in the bio-activity (e.g. affinity)
  - Large changes in the parameter can produce large or small changes in the affinity
  - This is exactly how medicinal chemists optimise compounds
This is neatly summed up in this paper, which analysed diversity and similarity

So – does (Q)SAR work?

- Yes, for localised sets of compounds - often simple parameters, if spatially localised and linearly dependant, will e.g. provide a useful regression
- A mistake is often to use a dataset of molecules and their activities that actually requires multiple models
- Another is to rely on vast numbers of parameters and model selection such as cross validation. I'm not a great fan of 'let's use all the available parameters and cross-validation will save the day' - the variance of a large number of parameters will often match the variance of the data - just put in enough variables.
Overfitting and cross validation
- three papers to read by Douglas Hawkins

- The Problem of Overfitting
  Hawkins, D. M.
  J. Chem. Inf. Comput. Sci.; (Perspective); 2004; 44(1); 1-12. DOI: 10.1021/ci0342472

- Assessing Model Fit by Cross-Validation
  Hawkins, D. M.; Basak, S. C.; Mills, D.

- QSAR with Few Compounds and Many Features
  Hawkins, D. M.; Basak, S. C.; Shi, X.

Which leads on to another interesting aspect of molecular data database bias...
Database bias. 'Sophisticated models' are sometimes little better than simple models. The 'Database bias' in activity databases is simply that the active molecules are generally very similar classes and are memorised!

Put another way, the information content of many common structure-based descriptors for virtual screening purposes is, in some cases, not higher than the nonstructural information about the number of atoms per element in the structure.

Below, is an example using only atom counts compared to more complex similarity descriptors. Note the high performance of the 'dumb descriptors'.

**Figure 1.** Average hit rate using “dumb” atom count descriptors, compared to a variety of 2D and 3D similarity searching methods. Even atom count descriptors achieve an enrichment of about 4-fold, which is already superior to one of the virtual affinity fingerprint methods, DOCKSIM, and around half the enrichment achieved by other methods employed.

**Figure 2.** Fraction of active compounds found using simple atom counts, in comparison to Unity fingerprints and the MOLPRINT 2D method. Although Unity fingerprints outperform atom counts, overall, this margin is smaller than one might expect, given the fact that atom counts do not contain any structural information whatsoever, whereas Unity fingerprints have that information available.

*Bender et al. J. Chem. Inf. Model. 2005, 45, 1369-1375*
The failure to account for the influence of the solid state on solubility

The General Solubility Equation is a rare example that does.

\[ \log S = 0.8 \cdot \log P - 0.01(MP - 25) \]

A simple model example, again using solubility – putting in parameters that relate to the phenomenon

(from Wassvik, C. Uppsala Pharmaceutical Profiling Conference)
What works best?

- Using parameters that have a physical foundation in the phenomenon being studied.
- Selecting the correct method for model creation, functional (statistical) or clustering (machine learning).
- Always use a properly selected test set. The various cross validation and bootstrapping methods don’t take account of database bias (despite what they say).
Conclusions

- The data is king – comprehensive, in an extensible format is best
- Parameters in a model should relate to the phenomenon being studied. If not, smell a rat.
- Machine learning methods have the property of local models – best for discontinuous SAR data

- An example of a typical model.....
So, models of data…..

- A Biologist, a chemist and a mathematician are asked to explain reproduction
So, models of data…..

- A Biologist, a chemist and a mathematician are asked to explain reproduction
- The Biologist says “It’s very complex, involves lots of Bioinformatics and I’ll need ten years and a £10M grant to sort it out”
So, models of data…..

- A Biologist, a chemist and a mathematician are asked to explain reproduction

- The Biologist says “It’s very complex, involves lots of Bioinformatics and I’ll need ten years and a £10M grant to sort it out”

- The Chemist says “It’s even more difficult (and involves Hydrogen-bonds), but given 15 years and £30M I can sort it out”
A Biologist, a chemist and a mathematician are asked to explain reproduction.

The Biologist says “It’s very complex, involves lots of Bioinformatics and I’ll need ten years and a £10M grant to sort it out.”

The Chemist says “It’s even more difficult (and involves Hydrogen-bonds), but given 15 years and £30M I can sort it out.”

The mathematician says “We’ve already solved the problem. We take two spherical bodies connected by an N-dimensional attraction…….”
Sources of information on Chemoinformatics

- Chemoinformatics in Drug Discovery. Tudor I. Oprea (Editor), Raimund Mannhold (Series Editor), Hugo Kubinyi (Series Editor), Gerd Folkers (Series Editor). 2005. ISBN: 978-3-527-30753-1
- http://www.cheminformatics.org/
- http://www.raell.demon.co.uk/chem/cheminformatics/index.htm
- ….lots more