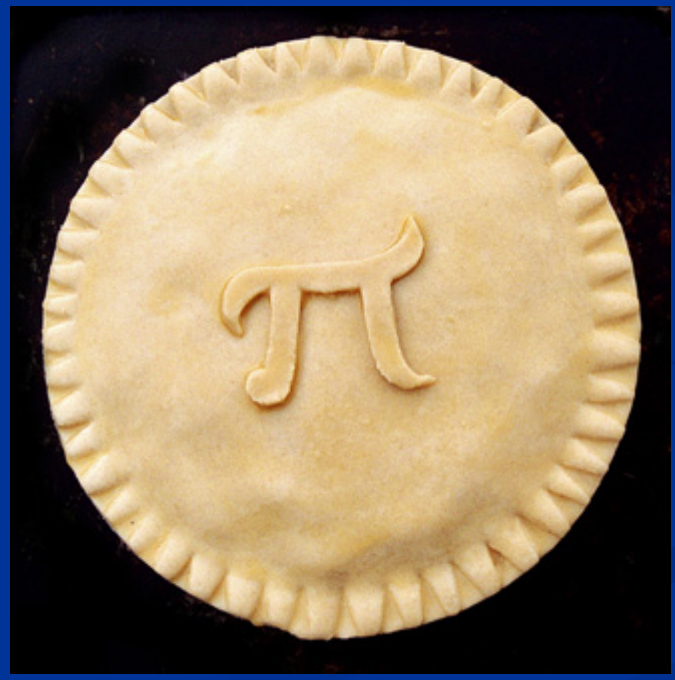
Molecular Informatics: A finger in every Pie



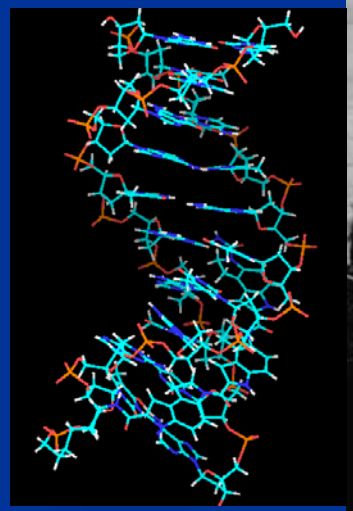
Informatics…a 'new' word …

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

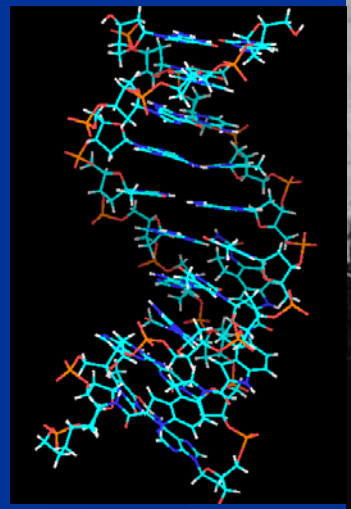
- For Molecular Informatics
  - I would add the collection, transformation and visualisation of **chemical data** to extract a deeper knowledge of the underlying properties of the data.

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

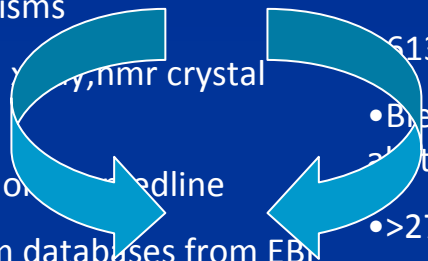


– this is what we spend most of our time doing with chemical data, and it's easier now with computers – especially as data is available as never before…

A hundred BILLION (Chemoinformatics) challenges and opportunities

Biological data

- 91,170,934,6353 **nucleotide bases**
- over 800 organisms
- 510792 protein x-ray,nmr crystal structures
- 23 million citations in Medline
- 40 main stream databases from EBI
- Ensemble : 24 million gene predictions



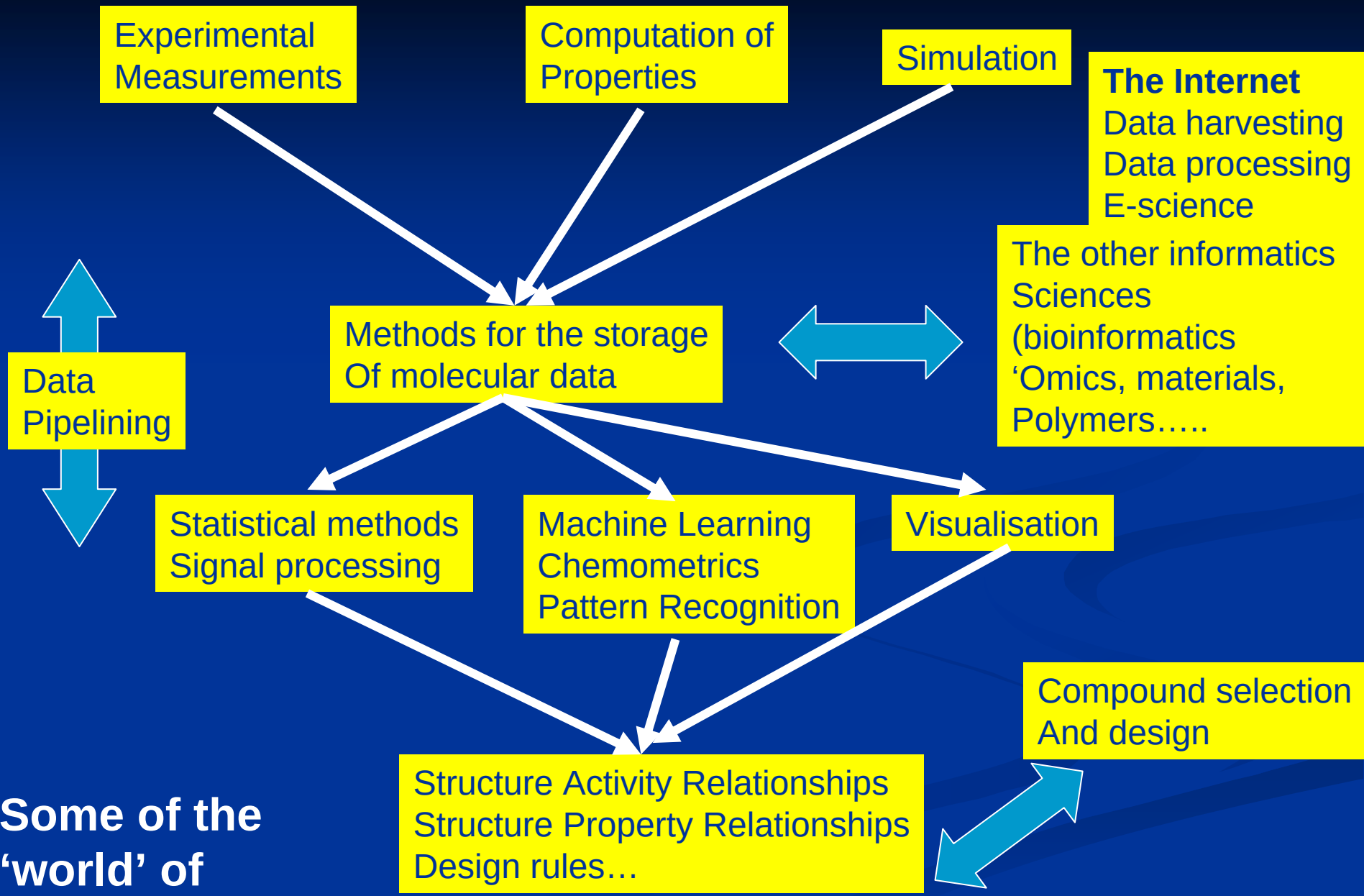
Chemical data

- **95 million chemical substances**
- 3,700,000 chemical reactions
- 613,000 available reagents
- Beilstein has 600,000 reaction abstracts
- >270,000 organic x-ray structures
- Combichem libraries of **Billions** of compounds

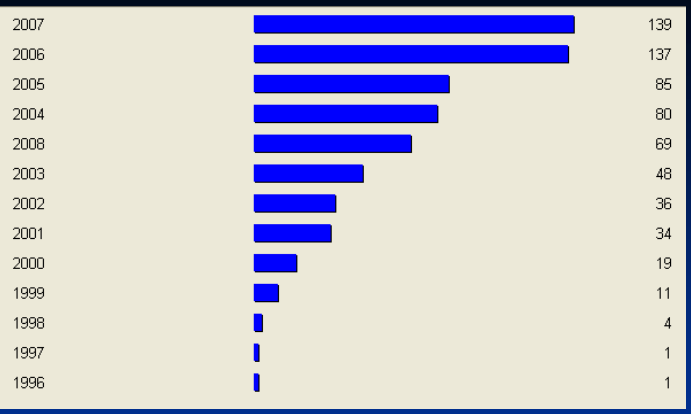
Patents

- **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 – its very big ! The role of molecular informatics is the collection, analysis and interpretation of these data.

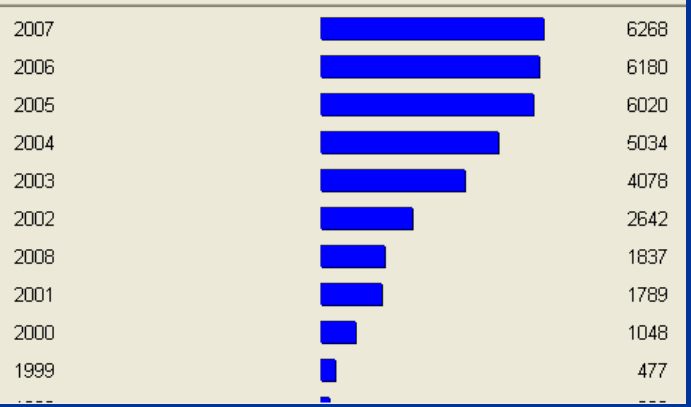


Some of the 'world' of Molecular Informatics

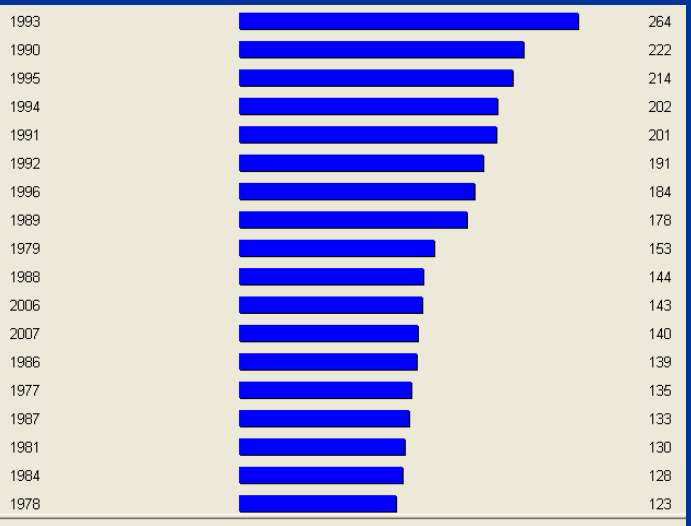


If you are searching the literature, SciFinder Scholar results:-

Chemoinformatics and Cheminformatics (664 entries) from 1996-2007

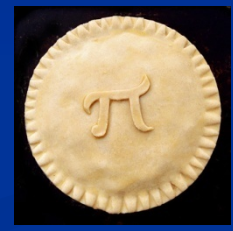


Bioinformatics (35,962 entries) 1999-2007



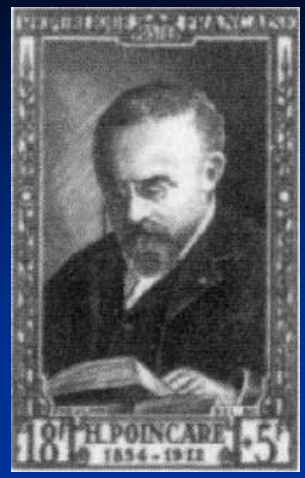
Chemistry AND informatics (1965-2007) 4,795 entries

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 lets look at two aspects



- The Semantic Web and data
- Models from data

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"

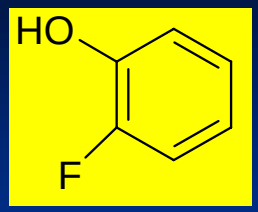
- Potentially, we can now store nearly all of our data in computers so…..
- This is central : particularly when we turn data into models, how do we know the models are real or even useful ?
- They need
  - the availability of *all* the *necessary* data
  - and a *relevant validated* analysis

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)
- Here's a couple of examples

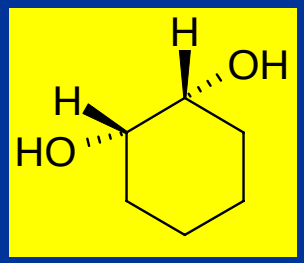
Some examples of Smiles



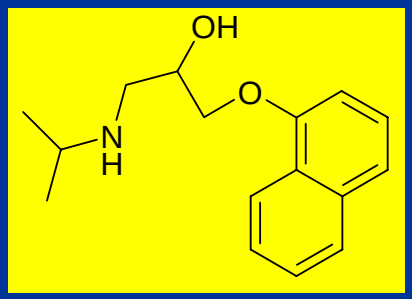
Fc1ccccc1O
orthofluorophenol

Very compact, clever,
Covers most organics,
Easy to interpret, widely used

[O-][n+]1ccccc1
pyridine-N-oxide

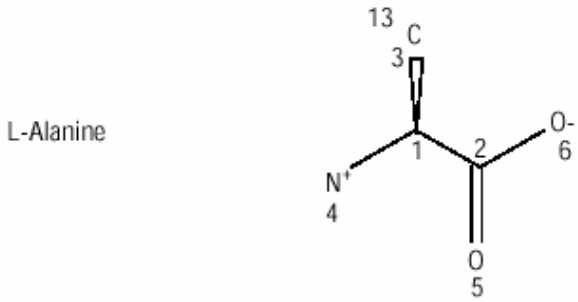


O[C@H]1CCCC[C@H]1O
cis-resorcinol



CC(C)NCC(O)COc1cccc2ccccc12
Propranolol – a beta-blocker

An SD file of Alanine: a common standard format; contains additional 3Dimensional and property information




```
L-Alanine (13C)
GSMACCS-II10169115362D 1   0.00366     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   -0.3000    0.0000 C   0  0  0  0  0  0
   -0.7207    2.0817    0.0000 C   1  0  0  0  0  0
   -1.8622   -0.3695    0.0000 N   0  3  0  0  0  0
    0.6220   -1.8037    0.0000 O   0  0  0  0  0  0
    1.9464    0.4244    0.0000 O   0  5  0  0  0  0
  1  2  1  0  0  0
  1  3  1  1  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
```

Header block

Counts line

Atom block

Bond block

Blocks not used in this Ctab

List block

Stext block

Properties block

Connection table (Ctab)

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

- Crystaleye 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/crystaleye/

RSC Publishing

Online Shop | Contact us | Advanced search

Publishing | Other RSC site areas... GO | Site Search GO

Home › Publishing › Journals › Project Prospect

Project Prospect

- Home
- Features
- Examples
- FAQ
- Contact
- Feedback
- News archive

Alerting Services

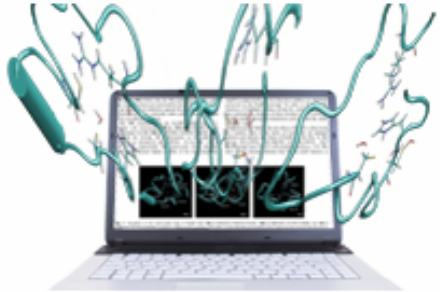
- RSS Feeds
- E-Alerts Service

Tools

- Email this to a friend
- Email your librarian

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

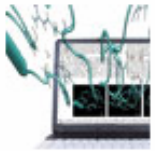
Examples
View sample enhanced articles

Project Prospect News

Latest news items relating to Project Prospect

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".

RSC quicklinks

I want information on:
Select a subject GO

Information for:
Select a role GO

I am interested in:
Select a product GO

Subscribe to Journal RSS Feeds

'...the best online submission system...' ReSourCe

RSC Open Science
more choice for journal authors

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

About this Journal
- Journal Home
- About this Journal
- Subscriptions
- People and Contacts
- Cover Gallery
- Copyright & Permissions

Reader Services
- Current Issue
- Search RSC Journals
- Article Finder
- Advance Articles
- Previous Issues
- Hot Articles
- Top 10

Alerting Services
- E-Alerts Service
- RSS feed

Author Services
- Submissions
- Guidelines
- Journals Grants for International Authors

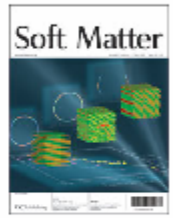
Referee Services
- Referee Report Forms

Customer Services
- Advertising
- Librarians
- Sample Content

Related Links
- Chemical Biology
- Chemical Science
- Chemical Technology
- Molecular BioSystems
- Soft Matter

Soft Matter

A new journal providing a forum for the communication of generic science underpinning the properties and applications of soft matter.



Paper

Soft Matter, 2007, **3**, 214 - 222, **DOI:** 10.1039/b612538g

Structure and stability of DPPE planar bilayers

Barry Stidder, Giovanna Fragneto and Stephen J. Roser

Biomembrane mimics in the form of supported planar bilayers allow the application of a wide range of surface and interface analytical techniques. The structure and phase-behavior of single and double bilayers of 1,2-dipalmitoylphosphoethanolamine (DPPE) were investigated by specular neutron reflectivity for their viability as biomembrane mimics. Whilst single bilayer samples were found to exhibit stable gel and fluid structures, double bilayers were found to be intrinsically unstable in the fluid phase as a planar structure. A Bragg peak was observed in the reflectivity data at just above the gel-to-fluid transition temperature, indicating the partial rearrangement of the upper bilayer into a repeat stacked structure. The lower bilayer was structurally stable. The structure and phase-behaviour of a double bilayer containing a ratio of 9 : 1 DPPE/cholesterol was also investigated to assess the stabilising effect of cholesterol on the upper bilayer. The presence of cholesterol completely destabilised the upper bilayer, causing it to detach 7 °C below the gel-to-fluid transition temperature of DPPE. It is possible that the cholesterol increases the overall conical shape of DPPE molecule by residing in

Introduction

(Show/Hide) Toolbox

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,[1] impedance analysis,[2,3] surface plasmon resonance,[4] neutron reflectivity,[5,6] and ellipsometry.[7] Examples of types of planar systems range from single bilayers adsorbed or deposited onto substrates[2,8–10] to multilamellar stacked bilayers[11,12] to elaborate polymer supported bilayers[13–15] and hybrid bilayers with one leaflet of lipids and the other of alkanethiols.[16]

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.[5] Fabrication is achieved by a combination of Langmuir–Blodgett and Langmuir–Schaefer 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.[17] The upper bilayer exhibits comparable gel, transition and fluid phase-behaviour to vesicles in solution.[18] 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 synchrotron radiation measurements have allowed the measurement of the bending modulus and tension of the floating bilayer.[19] Incorporation of low concentrations of cholesterol (1–6 mol%) was found to progressively decrease the swelling,[20] and at a concentration of 10 mol% swelling was completely removed.[21]

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 phosphatidylethanolamines enables stronger hydrogen bonds between the phosphate oxygen and the primary amine parts of the lipids.[22] Despite their predominance, phosphatidylethanolamines have not received as much attention as phosphatidylcholines in the 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 phosphatidylethanolamine 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 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.[20]

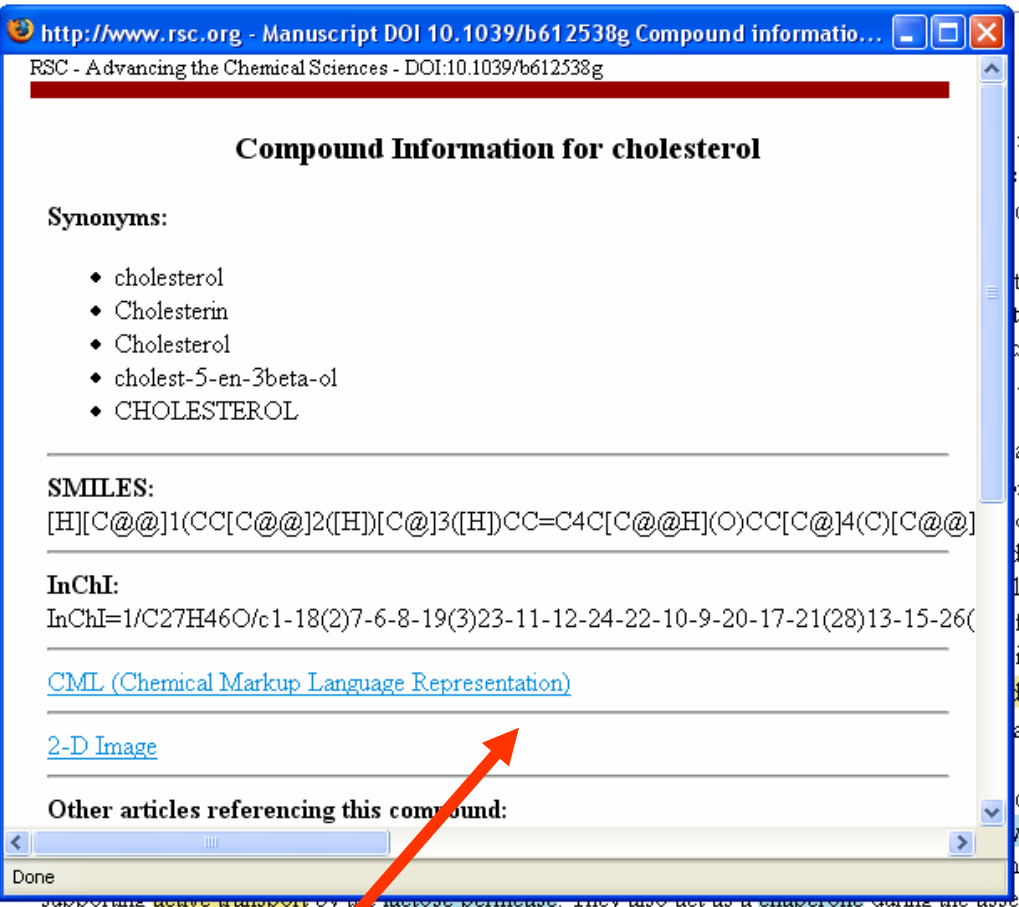
Introduction

To exploit certain types of surface and interfacial techniques in biological cell membrane studies it is necessary to use planar lipid bilayers loc…e. These planar systems have allowed the application of AFM,[1] impedance analysis,[2,3] surface plasmon resonance,[4] neutron reflectivity,[5,6] and ellipso… planar systems range from single bilayers adsorbed or deposited onto substrates[2,8–10] to multilamellar stacked bilayers[11,12] to elaborate polymer sup… rid bilayers with one leaflet of lipids and the other of alkanethiols.[16]

One of the main advantages of single bilayer samples compared to multilamellar samples is that the information obtained is bilayer specifi… dreds or thousands of bilayers. Another advantage is that very low concentrations are needed to fabricate these samples. This is obviously interesting … 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 … disadvantage is that only a very thin water layer separates the bilayer from the substrate (5–10 Å). This can restrict the inclusion of transmem… 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. … above another bilayer that is in close proximity to the substrate.[5] Fabrication is achieved by a combination of Langmuir–Blodgett and Langmuir–Sc… chniques enable the fabrication of asymmetric bilayers where the composition of each leaflet can be selected to model the asymmetric nature of real m… gle bilayer samples the upper bilayer is less constrained and is open to a large reservoir of water, making it ideally suited to study transmembrane phenc… en 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 t… l from the substrate by a water layer of 5–10 Å.[17] The upper bilayer exhibits comparable gel, transition and fluid phase-behaviour to vesicles in solutio… tudies, the main water layer was found to swell around the main transition temperature. This was interpreted in terms of competition between the inter-t… ne fluctuations and used to estimate the bending rigidity of the bilayer.[17] Off-specular synchrotron radiation measurements have allowed the mea… dulus and tension of the floating bilayer.[19] Incorporation of low concentrations of cholesterol (1–6 mol%) was found to progressively decrease the swe… n of 10 mol% swelling was completely removed.[21]

Together with phosphatidylcholines, phosphatidylethanolamines are one of the most abundant components of the lipid bilayer in membrar… 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 four… in membranes, being predominantly located in the inner cytoplasm-facing leaflet. As well as being one of the major building blocks of membrar… sks such as supporting active transport by the lactose permease. They also act as a chaperone during the assembly of membrane proteins, guiding the fo… ansition 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.[22] Despite their predominance, phosphatidylethanolamines have not received as much attention as phosphatidylcholines in the 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 phosphatidylethanolamine 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 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.[20]



is necessary to use planar lipid bilayers loc…e. These …onance,[4] neutron reflectivity,[5,6] and ellipso… planar …cked bilayers[11,12] to elaborate polymer sup… rid bilayers

…t the information obtained is bilayer specifi… dreds or …these samples. This is obviously interesting … are being …raining force upon the bilayer, inhibiting its … y …. This can restrict the inclusion of transmem… either side of

…ar membrane system has been developed. … above …n of Langmuir–Blodgett and Langmuir–Sc… chniques …d to model the asymmetric nature of real m… gle bilayer …deally suited to study transmembrane phen… en the double …layer by a water layer of 20–30 Å, whilst t… from the …fluid phase-behaviour to vesicles in solutio… tudies, the …in terms of competition between the inter-b… ne …diation measurements have allowed the mea… dulus and …as found to progressively decrease the swe… of 10 mol%

…components of the lipid bilayer in membra… account for …yte plasma membranes. They are often fou… in …e of the major building blocks of membra… sks such as …supporting active transport by the lactose permease. They also act as a chaperone during the assembly of membrane proteins, guiding the folding pathway and aiding in 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.[22] Despite their predominance, phosphatidylethanolamines have not received as much attention as phosphatidylcholines in the 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 phosphatidylethanolamine 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 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.[20]

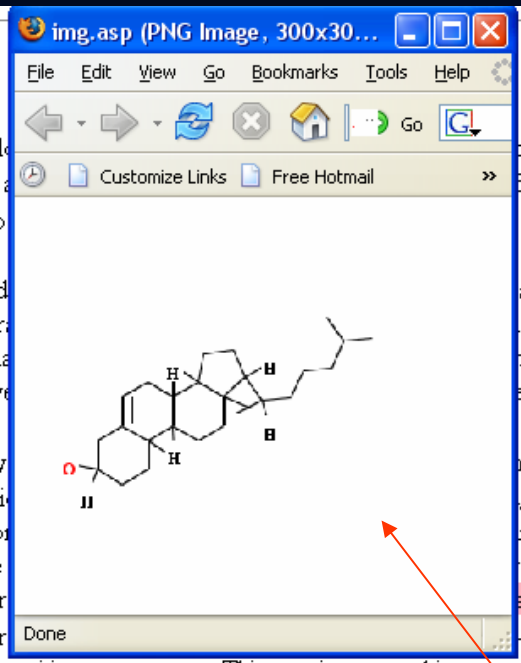
Introduction

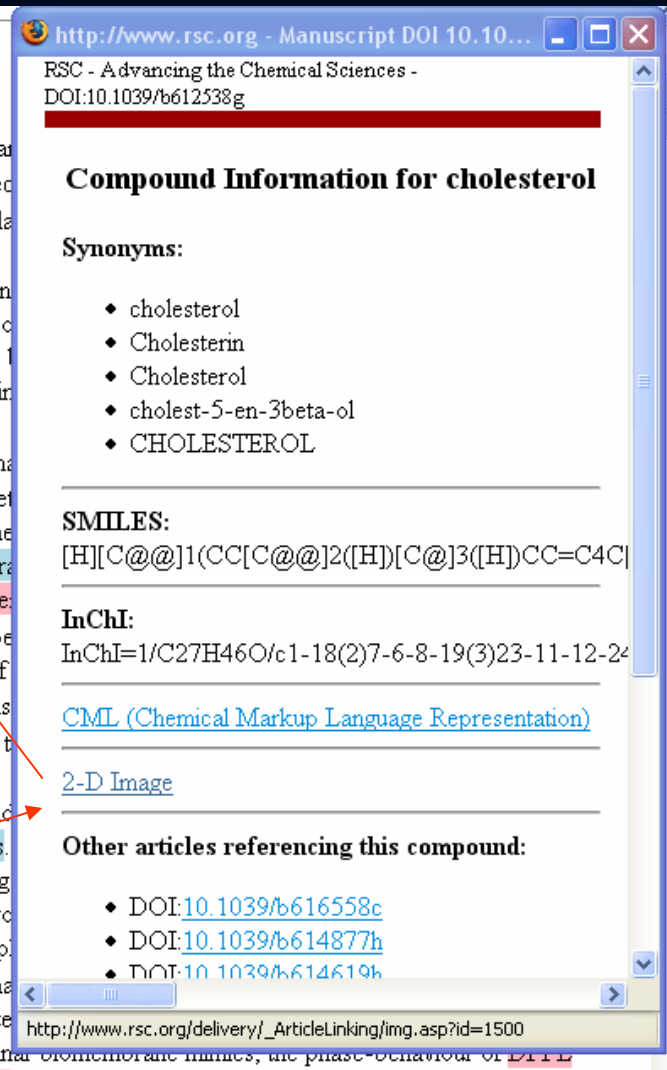
To exploit certain types of surface and interfacial techniques in biol… planar systems have allowed the application of AFM,[1] impedance … systems range from single bilayers adsorbed or deposited onto sub… with one leaflet of lipids and the other of alkanethiols.[16]

One of the main advantages of single bilayer samples compared … thousands of bilayers. Another advantage is that very low concentr… studied. Often the main disadvantage of single bilayer systems is tha… disadvantage is that only a very thin water layer separates the bilay… of the bilayer.[14]

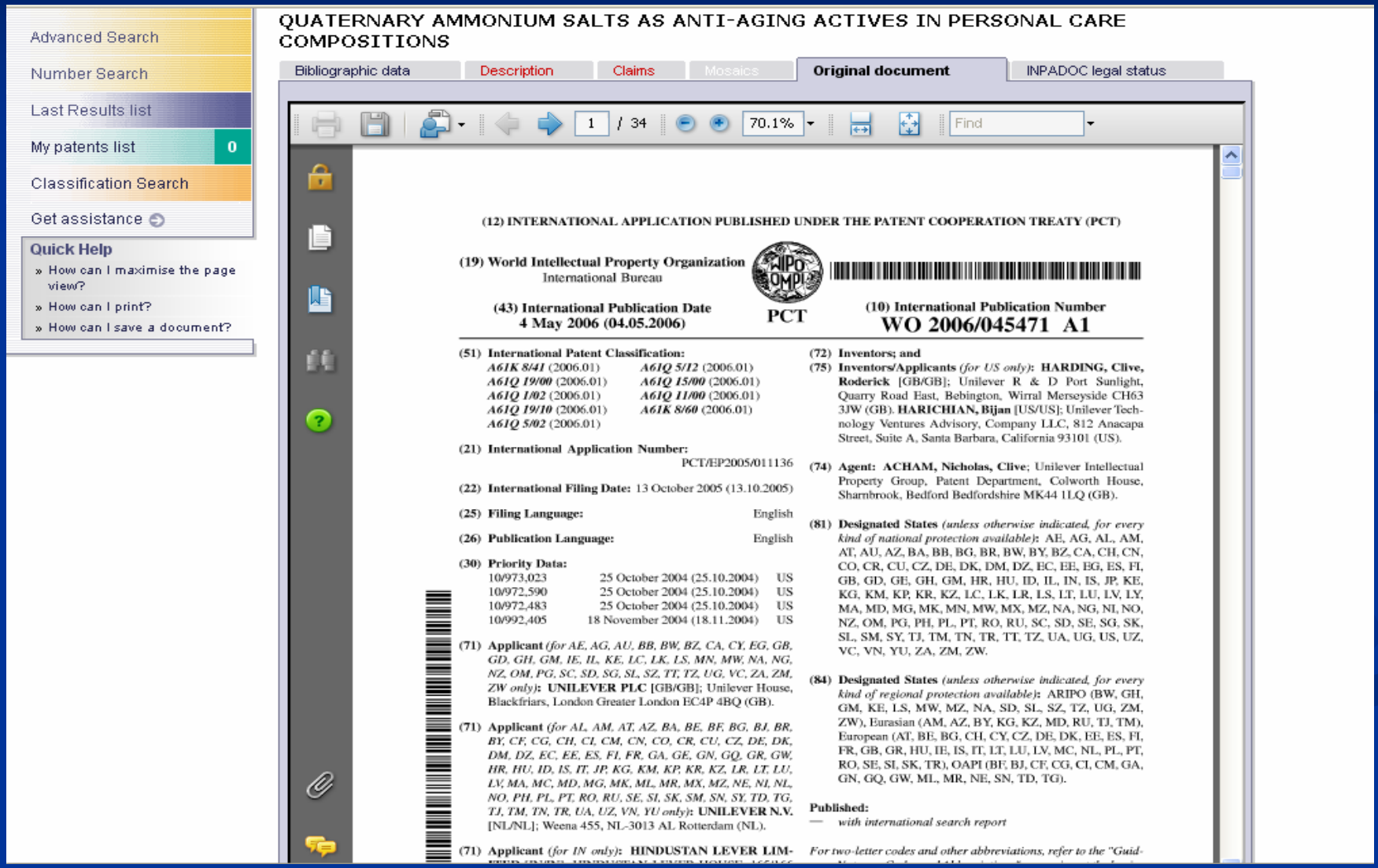
To overcome some of the problems associated with single bilay… another bilayer that is in close proximity to the substrate.[5] Fabricati… enable the fabrication of asymmetric bilayers where the compositio… samples the upper bilayer is less constrained and is open to a large … double bilayers are fabricated with phosphatidylcholines, the upper… from the substrate by a water layer of 5–10 Å.[17] The upper bilayer … studies, 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.[17] Off-specular synchrotron radiation meas… modulus and tension of the floating bilayer.[19] Incorporation of low concentrations of cholesterol (1–6 mol%) was found t… concentration of 10 mol% swelling was completely removed.[21]

Together with phosphatidylcholines, phosphatidylethanolamines are one of the most abundant components of the lipid… up to a third of the total percentage of lipids present, as 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 … supporting active transport by the lactose permease. They also act as a chaperone during the assembly of membrane pro… from the cytoplasmic to the membrane environment. In comparison to phosphatidylcholines, the smaller head-group of p… between the phosphate oxygen and the primary amine parts of the lipids.[22] Despite their predominance, phosphatidyletha… phosphatidylcholines in the literature. The phase-behaviour of phosphatidylethanolamines vesicles has been well characte… multilamellar bilayers is rather limited. With a view to the application of phosphatidylethanolamine double bilayers as pla… 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.[20]





An example with a patent (from Peter Corbett, developer of the software OSCAR3)



The paper is pasted into the parser,

Oscar3 Parse Form

Type or paste some text in here, and Oscar3 will parse it.

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-C3 alkyl)ammonium mono-substituted monosaccharide; (b) salts of hydroxypropyltri(C1-C3 alkyl)ammonium mono-substituted polyols, the salts having a cation of an average molecular weight no higher than 450 and the salt having a Tg no higher than 10 deg. C; (c) dihydroxypropyltri(C1-C3 alkyl)ammonium salts; (d) chlorohydroxypropyltri(C1-C3 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

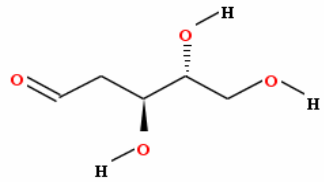
A personal care product is provided which includes a package filled with a personal care composition and instructions printed on or associated with the package indicating topical use of the composition on skin for purposes of controlling the signs of aging. The composition includes a quaternary ammonium compound selected from (a) salts of hydroxypropyltri(C1-C3 alkyl)ammonium mono-substituted monosaccharide; (b) salts of hydroxypropyltri(C1-C3 alkyl)ammonium mono-substituted polyols, the salts having a cation of an average molecular weight no higher than 450 and the salt having a Tg no higher than 10 deg. C; (c) dihydroxypropyltri(C1-C3 alkyl)ammonium salts; (d) chlorohydroxypropyltri(C1-C3 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. Alpha-hydroxycarboxylic acid derivatives are used widely in cosmetic commerce. Illustrative is US Patent 5,091,171 (Yu et al.). Retinol (Vitamin A) is an endogenous 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-hydroxycarboxylic acids and retinol as well as many of their derivatives tend to produce a stinging sensation and even redness on the skin when present at levels sufficient to be effective. Consumers would of course prefer performance without side effects. [0005]

Accordingly, there still 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: (i) from 0.1 to 30% by weight of a quaternary ammonium compound selected from the group consisting of: (a) salts of hydroxypropyltri(C1-C3 alkyl)ammonium mono-substituted monosaccharide; (b) salts of hydroxypropyltri(C1-C3 alkyl)ammonium mono-substituted polyols, the salt having a cation of an average molecular weight no higher than 450 and the salt having a Tg no higher than 10'C; (c) dihydroxypropyltri (Cl-C3 alkyl)ammonium salts; (d) chlorohydroxypropyltri (Cl-C3 alkyl) ammonium salts; and (e) mixtures thereof; (ii)from 1 to 99.9% by weight of a cosmetically acceptable carrier; and (B) 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 ammonium salts, particularly certain such salts of polyols, may control the signs of human skin aging. Three categories of quaternary ammonium (quat) salts are particularly useful to achieve the objectives of the present invention. These are outlined below. [0008] A first category of useful quaternary ammonium compounds are the salts of hydroxypropyltri(C1-C3alkyl) ammonium mono-substituted monosaccharides. These can be prepared by a variety of procedures. Most preferred is via reaction of 2-hydroxy-3-chloropropyl trimethylammonium chloride with a monosaccharide in an approximately 1:1 molar ratio in an alkaline medium. By typical Williamson synthesis, sodium chloride is eliminated thereby forming an ether linkage between the hydroxypropyl end of the quat group and the mono-saccharide. [0009] Monosaccharides, particularly reducing and non-reducing cyclic monosaccharides, are the smallest carbohydrate molecules encompassing the four-, five- and six- carbon sugars. Illustrative monosacchrides are ribose, deoxyribose, glucose, fructose, arabinose, xylose, lyxose, allose, altrose, gulose, mannose, idose, galactose and talose. Most preferred are glucose and fructose as the monosaccharide moiety which is to be substituted with the hydroxypropyltrimonium group. [0010] Ordinarily the C1-C3 alkyl constituent on the quaternized ammonium group will be methyl, ethyl, n-propyl, isopropyl or hydroxyethyl and mixtures thereof. Particularly preferred is a trimethyl ammonium group known through INCI nomenclature as a "trimonium" group. Any anion can be used in the quat salts of this invention. The anion may be organic or inorganic with the proviso that the material is cosmetically acceptable. Typical inorganic anions are halides, sulfates, phosphates, nitrates and borates. Most preferred

- Experimental data
- Ontology term
- Chemical (etc.) with structure
- Chemical (etc.), without structure
- Reaction
- Chemical adjective
- enzyme -ase word
- Chemical prefix

The highlighted chemical name is automatically detected, converted to CML, and the structure and associated data is now added.

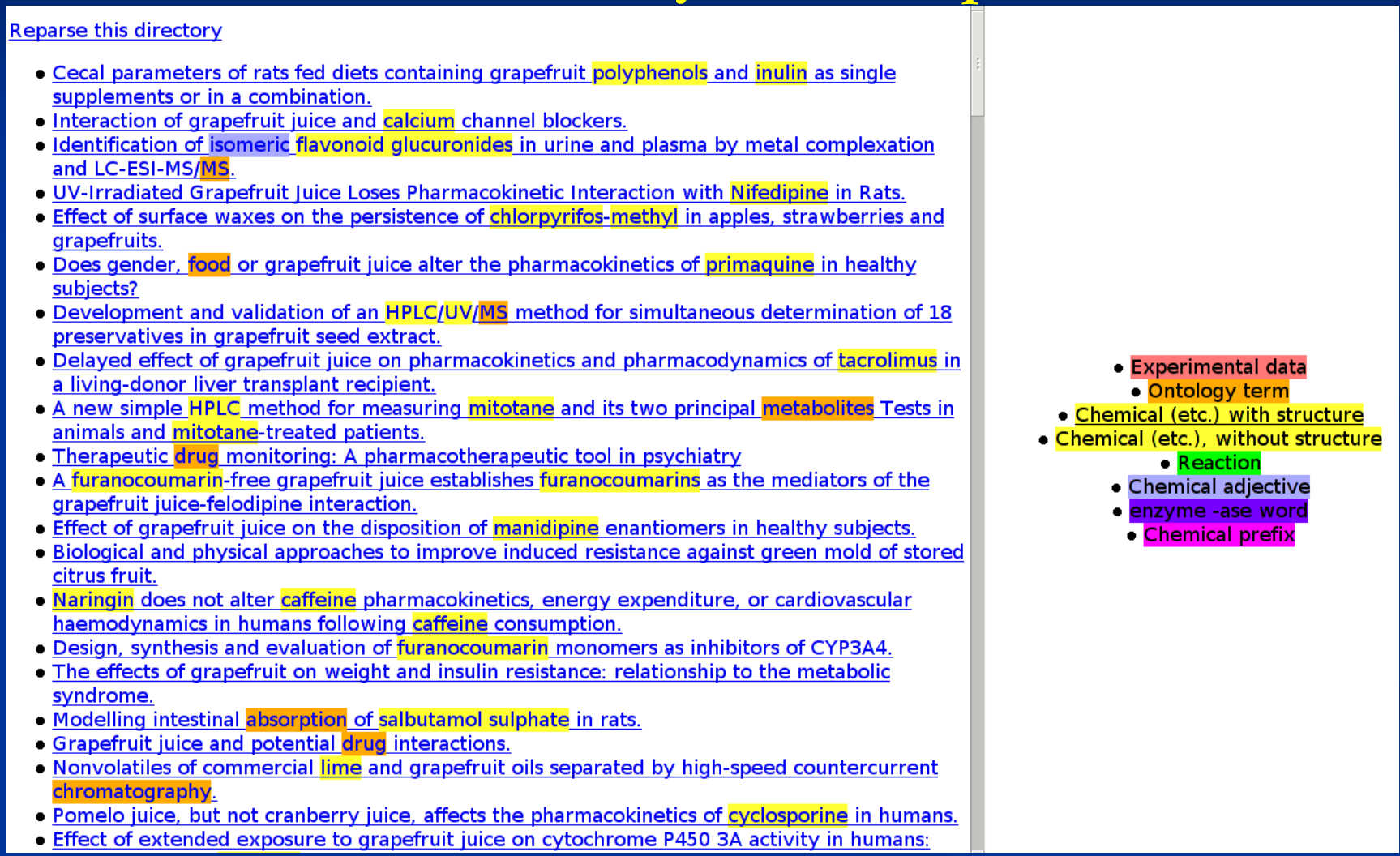
Summary of the Invention [0006] A personal care product is provided which includes: (A) a package filled with a personal care composition which includes: (i) from 0.1 to 30% by weight of a quaternary ammonium compound selected from the group consisting of: (a) salts of hydroxypropyltri(C1-C3 alkyl)ammonium mono-substituted monosaccharide; (b) salts of hydroxypropyltri(C1-C3 alkyl)ammonium mono-substituted polyols, the salt having a cation of an average molecular weight no higher than 450 and the salt having a Tg no higher than 10'C; (c) dihydroxypropyltri (Cl-C3 alkyl)ammonium salts; (d) chlorohydroxypropyltri (Cl-C3 alkyl) ammonium salts; and (e) mixtures thereof; (ii)from 1 to 99.9% by weight of a cosmetically acceptable carrier; and (B) 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 ammonium salts, particularly certain such salts of polyols, may control the signs of human skin aging. Three categories of quaternary ammonium (quat) salts are particularly useful to achieve the objectives of the present invention. These are outlined below. [0008] A first category of useful quaternary ammonium compounds are the salts of hydroxypropyltri(C1-C3alkyl) ammonium mono-substituted monosaccharides. These can be prepared by a variety of procedures. Most preferred is via reaction of 2-hydroxy-3-chloropropyl trimethylammonium chloride with a monosaccharide in an approximately 1:1 molar ratio in an alkaline medium. By typical Williamson synthesis, sodium chloride is eliminated thereby forming an ether linkage between the hydroxypropyl end of the quat group and the mono-saccharide. [0009] Monosaccharides, particularly reducing and non-reducing cyclic monosaccharides, are the smallest carbohydrate molecules encompassing the four-, five- and six- carbon sugars. Illustrative monosacchrides are ribose, deoxyribose, glucose, fructose, arabinose, xylose, lyxose, allose, altrose, gulose, mannose, idose, galactose and talose. Most preferred are glucose and fructose as the monosaccharide moiety which is to be substituted with the hydroxypropyltrimonium group. [0010] Ordinarily the C1-C3 al… surface = deoxyribose; type = CM; provenance = ChemNameDict; SMILES = [H]C([H])… methyl, ethyl, n-propyl, isopropyl or hydroxyethyl and mixtures thereof. Particularly preferred is a trimethyl ammonium group known through INCI nomenclature as a "trimonium" group. Any anion can be used in the quat salts of this invention. The anion may be organic or inorganic with the proviso that the material is cosmetically acceptable. Typical inorganic anions are halides, sulfates, phosphates, nitrates and borates. Most preferred are the halides, especially chloride. Organic anionic counter ions include methosulfate, toluoyl sulfate, acetate, citrate, tartrate, glycolate, lactate, gluconate, and benzenesulfonate. [0011] Particularly preferred quaternary ammonium salts of the first category are illustrated by structures I and II below, wherein X- is a halide. These formulas are intended as including all comformational isomers of the depicted structures. <EMI ID=1.1> <EMI ID=2.1> [0012] Advantageously, compositions of the present invention will be formulated with a quaternary ammonium salt where the monosaccharide is only mono-substituted with hydroxypropyltri(C1-C3 alkyl)ammonium groups. However, smaller amounts of di- and tri-substituted monosaccharide may also be present. These amounts normally may range from 0 to 20%, possibly from 2 to 10% by weight based on the weight of the quaternary ammonium compound present. More specifically, the multi-substituted monosaccharide may be di-[hydroxypropyltri(C1-C3 alkyl)ammonium] monosaccharide, tri-[hydroxypropyltri(C1-C3 alkyl)ammonium] monosaccharide and mixtures thereof. [0013] A second category of quaternary ammonium compound useful for this invention are the salts of hydroxypropyl tri(C1-C3 alkyl)ammonium monosubstituted polyols. These can be formed in a variety of procedures. Most preferred is via reaction of 2-hydroxy-3-chloropropyl trimethyl ammonium chloride with a polyol, particularly a linear polyol in an approximately 1:1 molar ratio in an alkaline medium. By typical Williamson synthesis, sodium chloride is eliminated thereby forming an ether linkage between the hydroxypropyl end of the quat group and the polyol. Typical polyols are sorbitol, pentaerythritol, neopentyl glycol, propylene glycol, dipropylene glycol and isoprene glycol. [0014] The second category will comprise salts with a cation having an average molecular weight no higher than 450, preferably no higher than 400, and optimally between 300 and 400. Further, the salt advantageously is liquid at 23 deg. C. Thus, the Tg preferably is no higher than 10 deg. C, more preferably no higher than 0 deg. C. The Tg can be measured in a Differential Scanning Calorimeter. [0015] A third category of suitable ammonium compounds are the hydroxypropyltri(C1-C3 alkyl)ammonium salts. These salts may be obtained in a variety of synthetic procedures, most particularly by hydrolysis of chlorohydroxypropyltri(C1-C3 alkyl)ammonium salts. A most preferred species is 1,2-dihydroxypropyltrimonium chloride, wherein the C1-C3 alkyl is a methyl group. Also useful may be 2-hydroxy-3-chloropropyl trimethylammonium chloride. [0016] Amounts of the quaternary ammonium compound may range from 0.1 to 30%, preferably from 0.5 to 20%, optimally from 1% to 12% by weight of the composition. [0017]



- Experimental data
- Ontology term
- Chemical (etc.) with structure
- Chemical (etc.), without structure
- Reaction
- Chemical adjective
- enzyme -ase word
- Chemical prefix

A poster and demonstration is available from Peter Corbett

Example: a search of PubMed for 'grapefruit' – fetched 200 abstracts - chemistry marked up



Reparse this directory

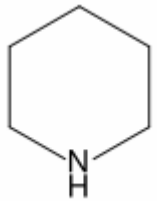
- Cecal parameters of rats fed diets containing grapefruit polyphenols and inulin as single supplements or in a combination.
- Interaction of grapefruit juice and calcium channel blockers.
- Identification of isomeric flavonoid glucuronides in urine and plasma by metal complexation and LC-ESI-MS/MS.
- 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.
- A new simple HPLC method for measuring mitotane and its two principal metabolites Tests in animals and mitotane-treated patients.
- 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 manidipine enantiomers in healthy subjects.
- Biological and physical approaches to improve induced resistance against green mold of stored citrus fruit.
- Naringin 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:

- Experimental data
- Ontology term
- Chemical (etc.) with structure
- Chemical (etc.), without structure
- Reaction
- Chemical adjective
- enzyme -ase word
- Chemical prefix

Search Form

(look for structures that contain piperidine

OSCAR3 Search



Query: C1CCCCN1

Type:

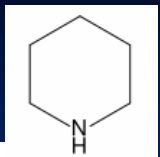
- Plain text
- InChI
- SMILES
- SMILES Substructure
- SMILES Similarity

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

Submit! Add Term

Search Results



Search Results

Results 1 to 5 of 21: next

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

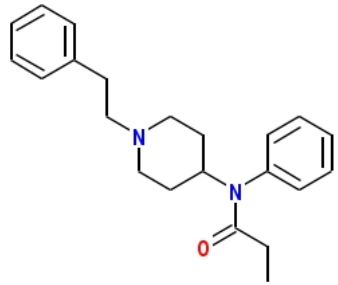
id = o5; surface = Fentanyl; type = CM; SMILES = CCC(=O)N(C1CCN(CC1)CCC2...

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

gnificantly 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,* and 5.8 +/- 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,* and 0.67 +/- 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



- Experimental data
- Ontology term
- Chemical (etc.) with structure
- Chemical (etc.), without structure
- Reaction
- Chemical adjective
- enzyme -ase word
- Chemical prefix

A Similarity Search

OSCAR3 Search

(Like Limonene)

Query: like limonene

Type:

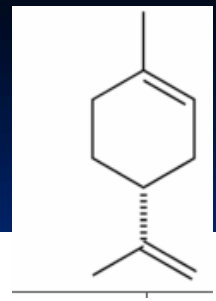
- Plain text
- InChI
- SMILES
- SMILES Substructure
- SMILES Similarity

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

Submit! Add Term

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

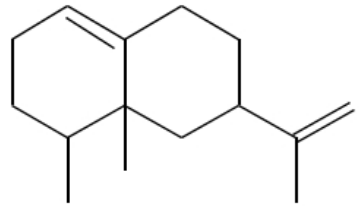
Use of novel compounds for pest control: insecticidal and acaricidal activity of essential oil components from heartwood of Alaska yellow cedar.

laris 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 monoterpene 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.



- Experimental data
- Ontology term
- Chemical (etc.) with structure
- Chemical (etc.), without structure
- Reaction
- Chemical adjective
- enzyme -ase word
- Chemical prefix

id = o7; surface = valencene; type = CM; SMILES = CC1CCC=C2C1(CC(CC2)C(...

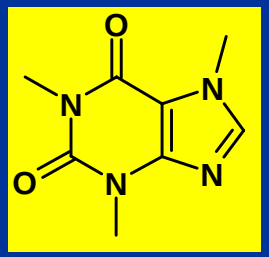
Moving on to Structure-Activity/Property models- some observations

- 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 ?

Caffeine solubility



Temperature	Solubility g/l	Year
25	2.132	1926 [1]
25	896.2	1985 [2]
25	21.0	2002 [3]
25	49.79	Merck Index
25	18.67	2005 [4]
25	21.6	SRC PhysProp Database

[1] Oliveri-Mandala, E. (1926), *Gazzetta Chimica Italiana 56*, 896-901

[2] Ochsner, A. B., Belloto, R. J., and Sokoloski, T. D. (1985), *Journal of Pharmaceutical Sciences 74*, 132-135

[3] Al-Maaieh, A., Flanagan, D. R. (2002), *Journal of Pharmaceutical Sciences 91*, 1000-1008

[4] Rytting, Erik, Lentz, Kimberley A., Chen, Xue-Qing, Qian, Feng, Venkatesh, Srini. *AAPS Journal* (2005), 7(1), E78-E105.

‘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

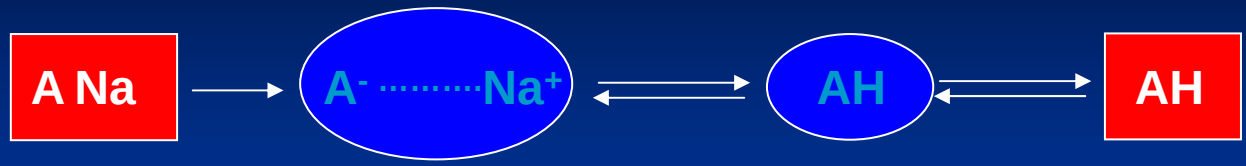
‘Solubility’ in the literature

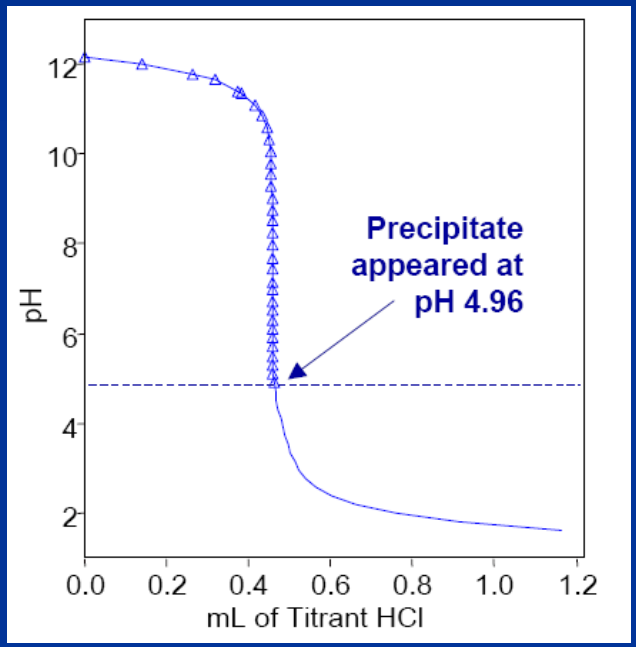
- Thus, any computational method that gives estimates (usually based on SAR) better than 0.5 log units is over fitted – many are!
- Dearden J.C. Expert Opin. Drug Discov. (2006), 1(1).

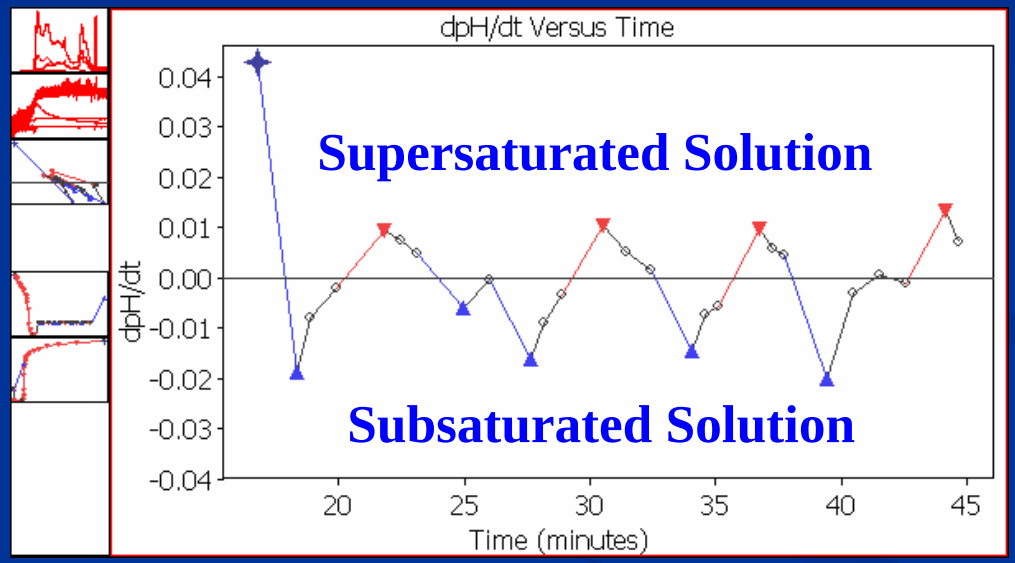
- The lit. data usually has no information on the experimental method, the material whose solubility is being studied, or the definition of the reported solubility – and commonly, many datasets are combined to build models.

In this case, we have decided to create our own data and not to combine it with other literature data.

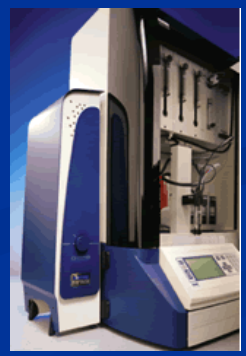
Potentiometric cycling method for very accurate and controlled measurement of solubility







Stuart, M., Box, K. Chasing equilibrium: measuring the intrinsic solubility of weak acids and bases. Anal. Chem. **2005**, 77(4), 983-990.



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)

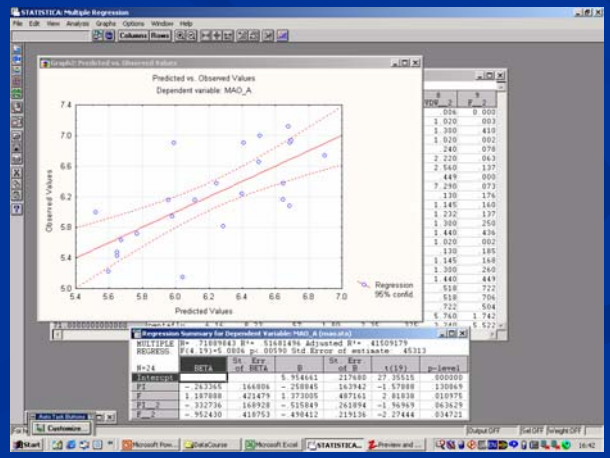
- This is obvious, but….even if you use a pre-computed model, check the data sources
  - Are data compatible, and can they be combined
    - It is often the case that non-compatible data are merged to create a database 'large enough' to do statistics on
  - Is there sufficient background information 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

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



Analysis

$$\mathrm{LogP} = \sum_{i=1,N} a_i f_i + \sum_{j=1,M} b_j F_j$$

Prediction

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!

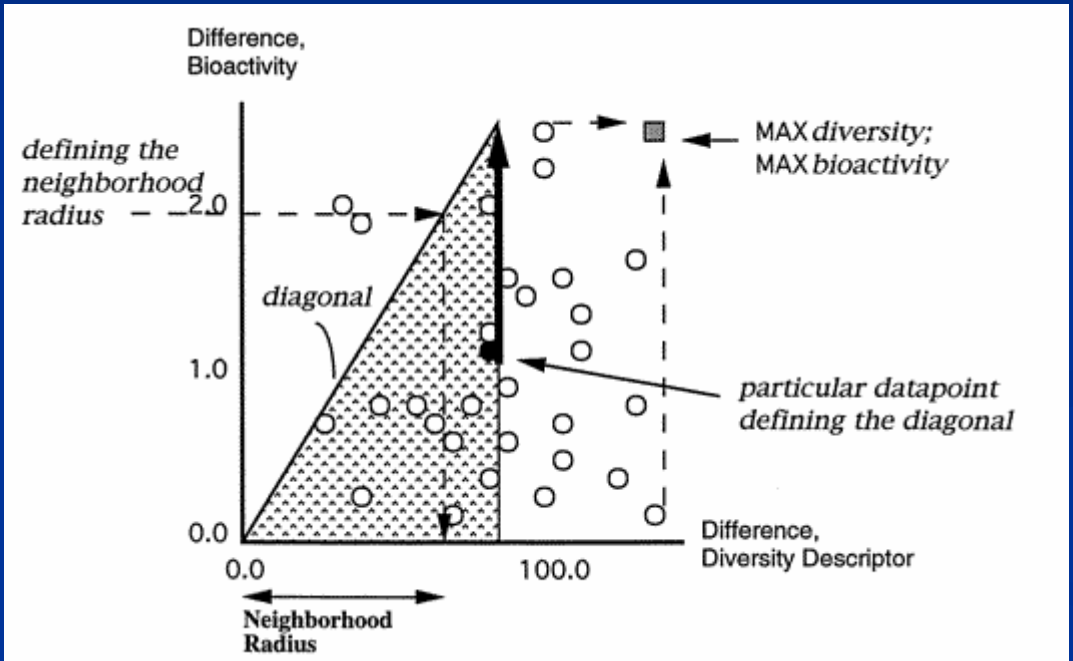
Property behaviour

- 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



- Neighborhood Behavior: A Useful Concept for Validation of "Molecular Diversity" Descriptors. Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. J. Med. Chem.; (Expedited Article); 1996; 39(16); 3049-3059. DOI: 10.1021/jm960290n

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 ' lets 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.
  J. Chem. Inf. Comput. Sci.; (Article); 2003; 43(2); 579-586. DOI: 10.1021/ci025626i
- QSAR with Few Compounds and Many Features
  Hawkins, D. M.; Basak, S. C.; Shi, X.
  J. Chem. Inf. Comput. Sci.; (Article); 2001; 41(3); 663-670. DOI: 10.1021/ci0001177

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'

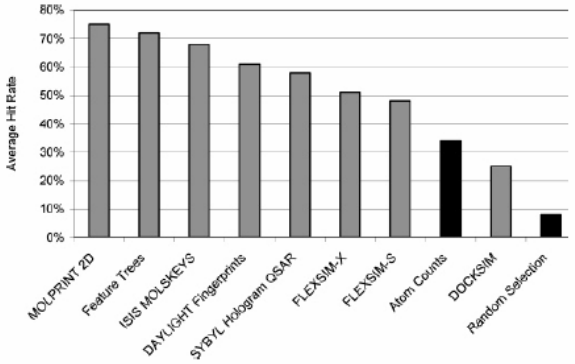


Figure 1. Average hit rate using "dumb" atom count descriptors, compared to a variety of 2D and 3D similarity searching methods. Even atom count descriptors achieve an enrichment of about 4-fold, which is already superior to one of the virtual affinity fingerprint methods, DOCKSIM, and around half the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

based methods, in principle, are able to exploit a wealth of

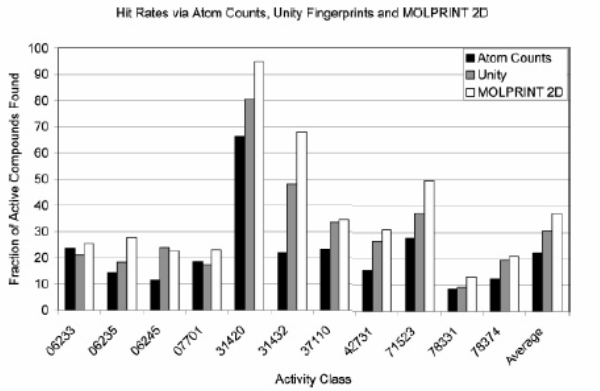


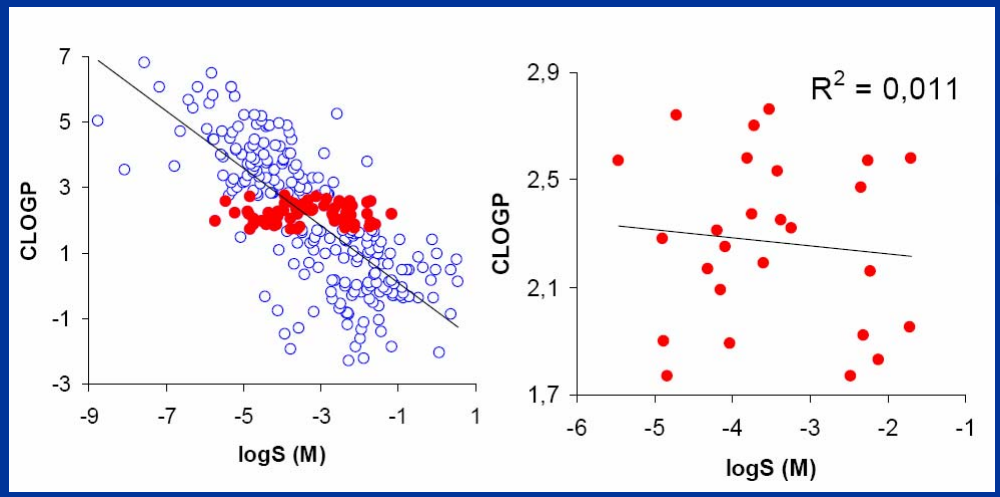
Figure 2. Fraction of active compounds found using simple atom counts, in comparison to Unity fingerprints and the MOLPRINT 2D method. Although Unity fingerprints outperform atom counts, overall, this margin is smaller than one might expect, given the fact that atom counts do not contain any structural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

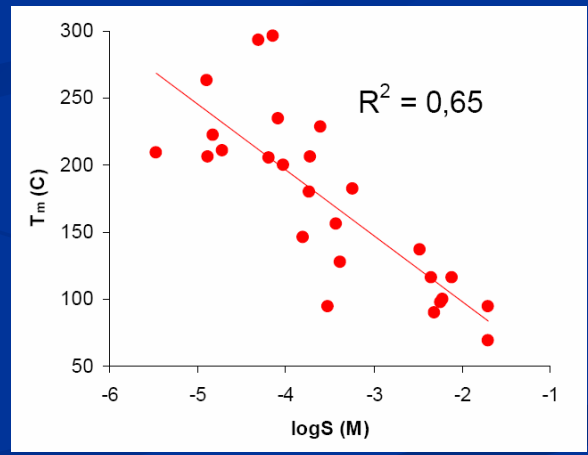
Bender et al. J. Chem. Inf. Model. ***2005, 45, 1369-1375 1369***

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

- The failure to account for the influence of the solid state on solubility
- The General Solubility Equation is a rare examples that does.

$$LogS = 0.8 - logP - 0.01(MP-25)$$





(from Wassvik, C. Uppsala Pharmaceutical Profiling Conference)

What works best ?

- Using parameters that have a physical foundation in the phenomenon being studies
- 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 bootsrapping 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 "Its even more difficult (and involves Hydrogen-bonds) , but given 15 years and £30M I can 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 "Its 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

- Gasteiger J.(Editor), Engel T.(Editor): Chemoinformatics : A Textbook. John Wiley & Sons, 2004, ISBN 3-527-30681-1
- A.R. Leach, V.J. Gillet: An Introduction to Chemoinformatics. Springer, 2003, ISBN 1-4020-1347-7
- Encyclopedia of Computational Chemistry, 5 volumes, ISBN: 0-471-96588-X
- 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
- Peter Ertl, Paul Selzer and Jörg Mühlbacher Web-based cheminformatics tools deployed via corporate Intranets. Drug Discovery Today: BIOSILICO, Volume 2, Issue 5, September 2004, Pages 201-207.
- http://www.cheminformatics.org/
- http://www.emolecules.com/doc/cheminformatics-101.htm
- http://www.raell.demon.co.uk/chem/cheminformatics/index.htm
- ….lots more