

Miljöriskanalys, kemometri och "virtual screening"

Tomas Öberg
Högskolan i Kalmar



Presentation vid Kemometrisektionens Vintermöte
i Tännålen, 13-16 april 2005

Risker och riskanalys

- Risk
 - Möjligheten att något oönskat ska inträffa
 - En kombination av sannolikhet och konsekvens
- $Risk = f(exponering, effekter)$

"Risk analysis is broadly defined to include risk assessment, risk characterization, risk communication, risk management, and policy relating to risk, in the context of risks of concern to individuals, to public and private sector organizations, and to society at a local, regional, national, or global level."

Society for Risk Analysis

Delmoment

- Riskbedömning
 - Faroidentifiering (hazard identification)
 - Dos-responsanalys (dose-response assessment)
 - Exponeringsanalys (exposure assessment)
 - Riskkaraktisering (risk characterization)
- Riskkommunikation
- Riskhantering



Miljöriskanalys i Kalmar

Forskning

- Faroidentifiering, dos-responsanalys
 - Virtual screening for environmental pollutants
- Exponeringsanalys, riskkarakterisering
 - Probabilistisk riskbedömning av förorenad mark
- Riskkommunikation
 - Conjoint analysis as a decision tool for evaluation of environmental performance

Utbildning

- Miljöriskanalys 10 p C-nivå
- Miljöriskanalys 5 p D-nivå
- Miljöriskkommunikation 5 p D-nivå

Var kommer kemometri in?

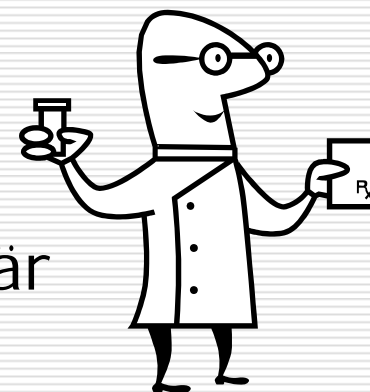
- Forskningsprojekten handlar om kemiska miljörisker och är fokuserade på tillämpning av statistisk metodik
 - Mönsterigenkänning och multivariat kalibrering
 - Statistisk försöksplanering
 - Återsampling och Monte Carlo-simulering
- "Virtual screening"-projektet är intressant att diskutera mer i detalj
 - Likheter med jakten på nya läkemedelskandidater
 - Påtaglig skillnad i hur det kan tillämpas

Virtuell screening

"Virtual screening is the computational or in silico analogue of biological screening. The aim is to score, rank and/or filter a set of structures using one or more computational procedures."

Leach & Gillet (2003): An introduction to chemoinformatics

- ❑ Läkemedelsforskningen har länge arbetat med "virtuell sållning"
- ❑ Är det något även för miljövetenskapen?
- ❑ Strukturaktivitetssamband (QSAR) är inget nytt inom miljökemi, men hur tillämpas de?



QSAR-modeller på miljöområdet

- Begränsad räckvidd
 - Små och homogena grupper av ämnen
- Ingen/dålig validering
 - Sällan korsvalidering, än mindre oberoende test-set
- Tillämpningsområdet är inte definierat
 - Ibland kvalitativa avgränsningar, men sällan mer

OECDs principer (november 2004):

1) *a defined endpoint*

2) *an unambiguous algorithm*

3) *a defined domain of applicability*

4) *appropriate measures of goodness-of-fit, robustness and predictivity*

5) *a mechanistic interpretation, if possible*

Begränsad tillämpning

- De flesta studier omfattar ett begränsat antal ämnen
- Tillämpningen inskränker sig ofta till kalibreringsdata
 - Ett undantag är beräkningsverktygen EPIWIN och ECOSAR (US EPA)
 - Begränsad validering
 - Saknar definition av tillämpningsområde
- Ytterligare en begränsning är att modeller för exponering och toxicitet behandlas separat

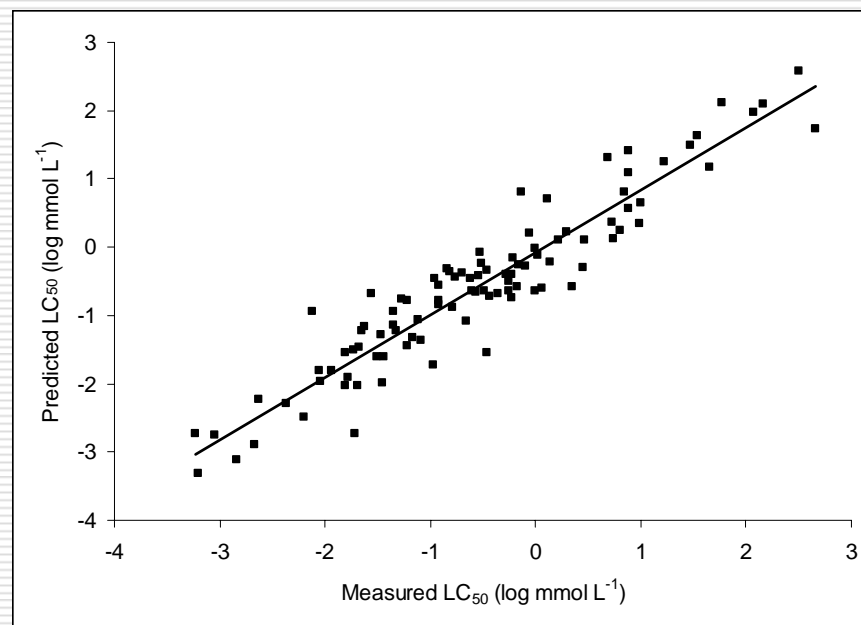
Två QSARs med en bredare ansats

- Icke-specifik toxicitet (narkos)
 - Den minimala giftigheten hos organiskt ämne
 - Icke-specifik toxicitet är additiv
 - I komplexa blandningar tenderar den narkotiska effekten att dominera

- Kemisk nedbrytning i atmosfären (reaktion med OH-radikaler)
 - Dominerande nedbrytningsväg för många organiska miljöföroreningar
 - Ett generellt mått på persistens

QSAR för icke-specifik toxicitet

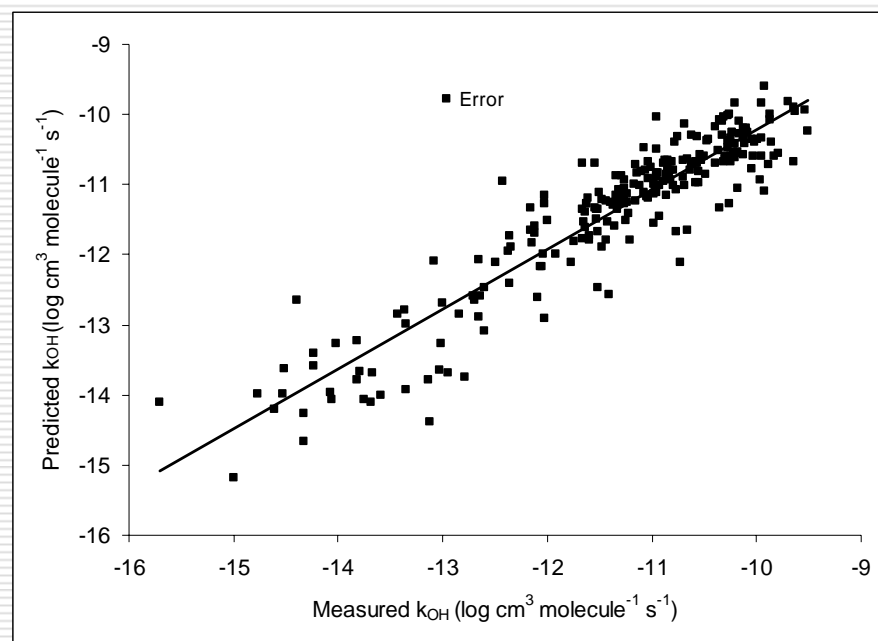
- Akvatisk toxicitet (*Pimephales promelas*)
 - 311 ämnen i vikt-intervallet 32-285 amu
 - 218 deskriptorer från 2D strukturer med *Dragon v. 5.0*, PLSR1-modell med *Unscrambler v. 9.0*
 - $R^2_{\text{Cal}}=0.909$ och $Q^2_{\text{Ext}}=0.888$



Öberg, T. (2004) *Chemical Research in Toxicology* **17**, 1630-1637

QSAR för reaktion med OH-radikalen

- Hastighetskonstant för gasfasreaktion
 - 743 ämnen i viktintervallet 30-300 amu
 - 333 deskriptorer från 2D strukturer med *Dragon v. 5.0*, PLSR1-modell med *Unscrambler v. 9.0*
 - $R^2_{Cal}=0.906$ och $Q^2_{Ext}=0.840$



Öberg, T. (2005) *Atmospheric Environment*, in press

Virtuell screening med QSAR-modellerna

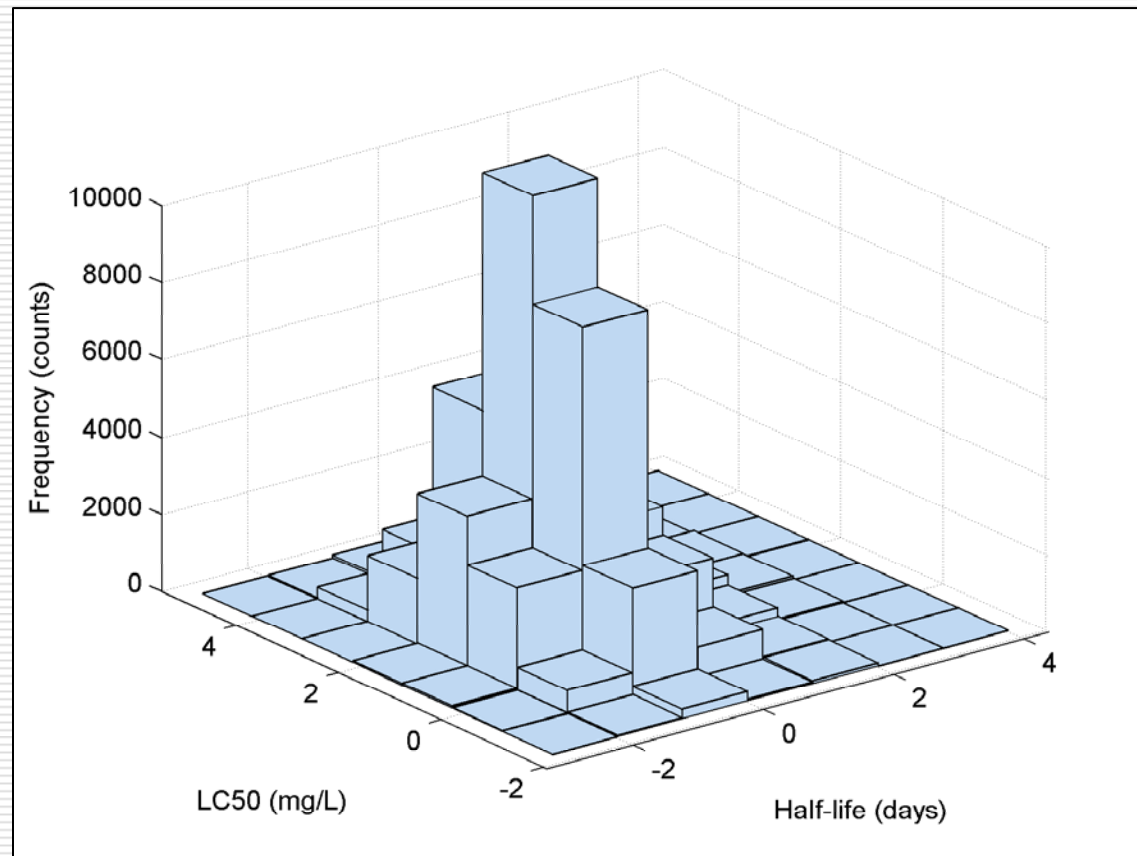
- The SMILESCAS Database (Syracuse Research Corporation)
 - God täckning av industrikemikalier
 - 109 611 ämnen med CAS-nr och 2D struktur
 - Deskriptorer genererade för 98 083 ämnen med Dragon 5.0 (exklusive metall- och kiselorganiska föreningar)

- Tillämpningsdomänen definierades av PLSR1-modellerna
 - 5 % signifikansgränsen för restvärdena
 - 3 ggr medelvärdet för "leverage" i kalibreringsdata

Prediktionsresultat

- Icke-specifik toxicitet
 - 66 884 ämnen inom tillämpningsdomänen
 - 26 % (25 316) hade ett skattat LC_{50} -värde under $10 \text{ mg}\cdot\text{L}^{-1}$, motsvarande kategori I och II i OECD Global Harmonized Classification System
- Reaktionshastighet med OH-radikalen
 - 51 894 ämnen inom tillämpningsdomänen
 - 15 % (7 557) hade en halveringstid över två dagar, d.v.s. persistenta
 - $t_{1/2} = \ln 2 / ([\text{OH}] \cdot k_{\text{OH}})$, antar $[\text{OH}] = 1 \cdot 10^6 \text{ molekyl}\cdot\text{cm}^{-3}$
- Båda modellerna: 50 074 ämnen inom tillämpningsdomänen

En bivariat lognormalfördelning

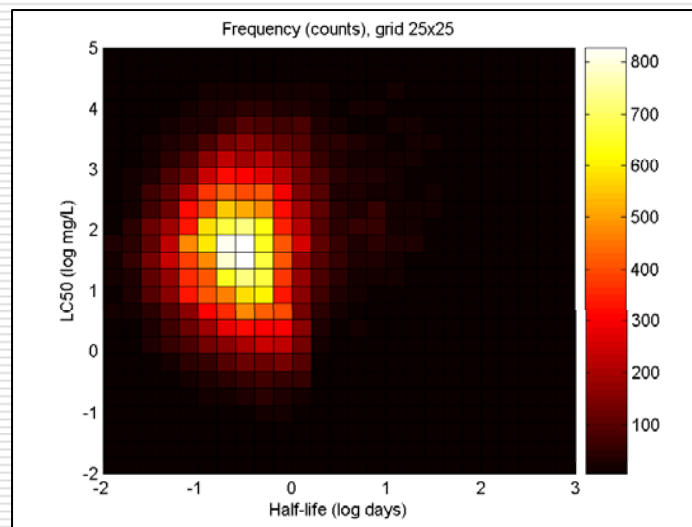


Går det att generalisera?

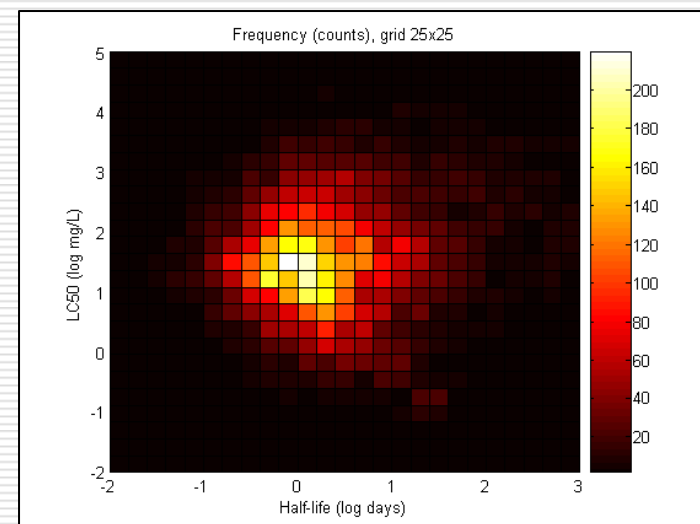
- Samma andel toxiska och persistenta ämnen erhöles från en mer begränsad databas (the PhysProp Database)
- Kan fördelningen av toxicitet och persistens bland små antropogena kemikalier beskrivas med statistiska fördelningar?
- För tidigt att säga med bestämdhet
 - Öppnar för nya ansatser för miljöriskanalys
 - Probabilistisk metoder
 - Policybeslut på generella grunder, snarare än ämnesvis

Effekten av halogenering

- 1 653 ämnen uppfyller kriterierna för både toxicitet ($LC_{50} < 10 \text{ mg} \cdot \text{L}^{-1}$) och persistens ($t_{1/2} > 2$ dagar). 88 % av dessa var halogenerade
- Effekten av halogenering syns dock långt utanför den här gruppen



39 428 icke-halogenerade ämnen



10 646 halogenerade ämnen

Nuvarande riskbedömnings- filosofi

- Kemikalierisker bedöms ämnesvis och i miljöarbetet har det lett till:
 - Att vi bortser från kumulativa effekter av additivt verkande ämnen
 - Att vi fokuserar på att prioritera och hitta de mest toxiska och persistenta ämnen
- Är det rätt väg att gå?
 - Det finns ingen distinkt avgränsning mellan persistenta och icke-persistenta ämnen
 - Det finns inte hellre någon tydlig avgränsning mellan toxiska och icke-toxiska ämnen

En riskindikator

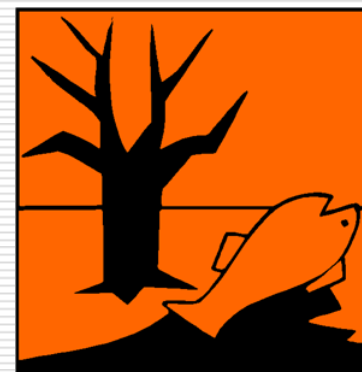
- De reciproka LC_{50} -värdena är ett mått på toxisk potens, som kan användas för att jämföra olika ämnen
 - Jfr. TEF-faktorerna för dioxinlika ämnen
- Persistens i miljön kan tas med genom att multiplicera med halveringstiden i atmosfären
- "Toxic persistence rating"
 - $TPR = t_{1/2} \text{ (dagar)} / LC_{50} \text{ (mg} \cdot \text{L}^{-1}\text{)}$
 - 2 dagar och $10 \text{ mg} \cdot \text{L}^{-1}$ kan då uttryckas som $TPR = 2 / 10 = 0.2$
 - 6 447 ämnen har en TPR-ranking över 0.2

Vilka ämnen har en hög ranking?

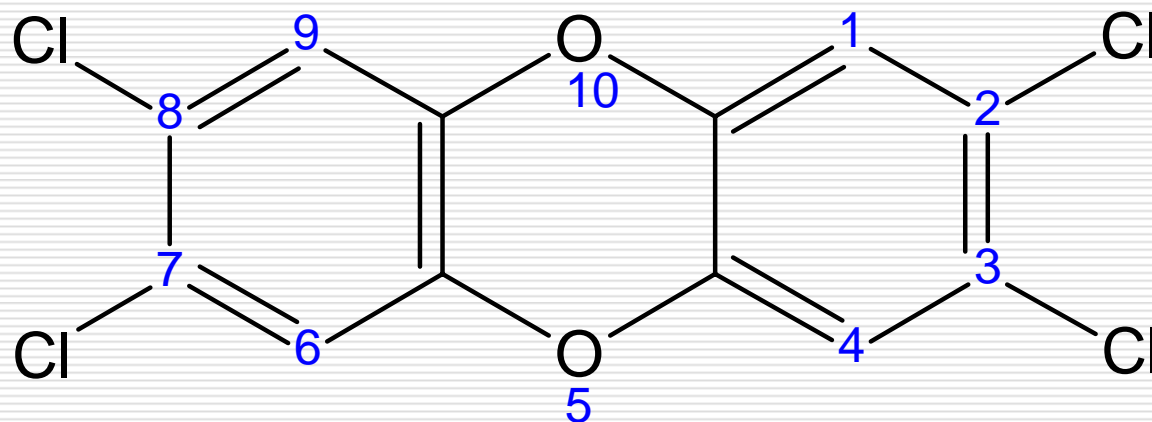
- I hela databasen
 - 62 % aromater, 23 % halogenerade och 16 % halogenerade aromater
- De 6 447 ämnena med TPR > 0.2
 - 69 % aromater, 55 % halogenerade och 40 % halogenerade aromater
- Nuvarande fokus på halogenerade aromater förefaller alltså välmotiverat, men det är många ämnen som tillhör andra grupper

"Topp 100"

- ❑ Alla är halogenerade
- ❑ Hög molekylvikt (median 292 amu)
- ❑ Välbekanta miljögifter
 - DDT metaboliter och derivat: 9 st
 - Tetraklorerade bifenyler: 42 st
 - Tetraklorerade difenyletrar: 19 st
 - Tetraklorerade dibensofuraner: 6 st
 - Klorparaffiner: 2 st
 - Klorerade bensener & derivat: 12 st
 - Ofullständigt halogenerade klorfluorkarboner (HCFC): 3 st
 - Övriga: 7 st



Vilka saknas?



- Polyklorerade dibenso-*p*-dioxiner (PCDD)
 - Polybromerade difenyletrar (PBDE)
 - Varför? Modellernas tillämpningsdomän?
-

Är miljörankingen praktiskt användbar?

- Den globala exponeringen för antropogena ämnen går knappast att utvärdera utan QSAR-modeller
- Ytterligare information som behövs
 - Produktion och handel
 - Förluster till miljön i hela produktcykeln ("från vaggan till graven")
- Svagheter i TPR-rankingen
 - En typ av toxicitet, för en art
 - Beaktar inte andra nedbrytningsvägar
 - Tar inte hänsyn till fördelning i olika media

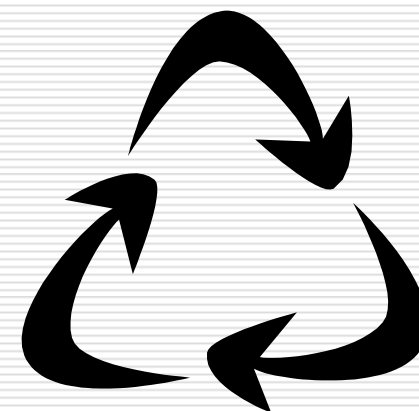


Kan fungera ändå

- ❑ Produktion och import kan följas i nationella produktregister, t.ex. det svenska
- ❑ Många undersökningar indikerar att additiva toxiska effekter dominerar vid exponering för blandningar i låga koncentrationer
- ❑ Artskillnader har sannolikt mindre inverkan vid icke-specifik toxicitet
- ❑ Reaktionen med OH-radikalen är sannolikt den dominerande nedbrytningsvägen för de flesta stabila organiska miljögifter

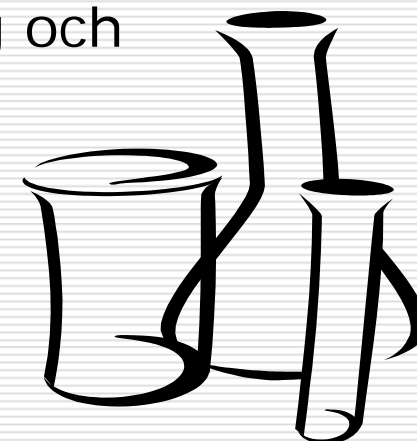
En möjlighet för "grön kemi"

- TPR-rankingen kan användas som ett verktyg
 - för att successivt välja mindre miljöskadliga ämnen (*substitutions- eller produktvalsprincipen*)
 - vid design av nya produkter (jfr. läkemedelsdesign)
- De 100 ämnena med lägst ranking
 - Dioler: 13 st
 - Sockerarter: 28 st
 - Sockerderivat: 15 st
 - Andra hydroxylerade: 37 st
 - Övriga: 7 st

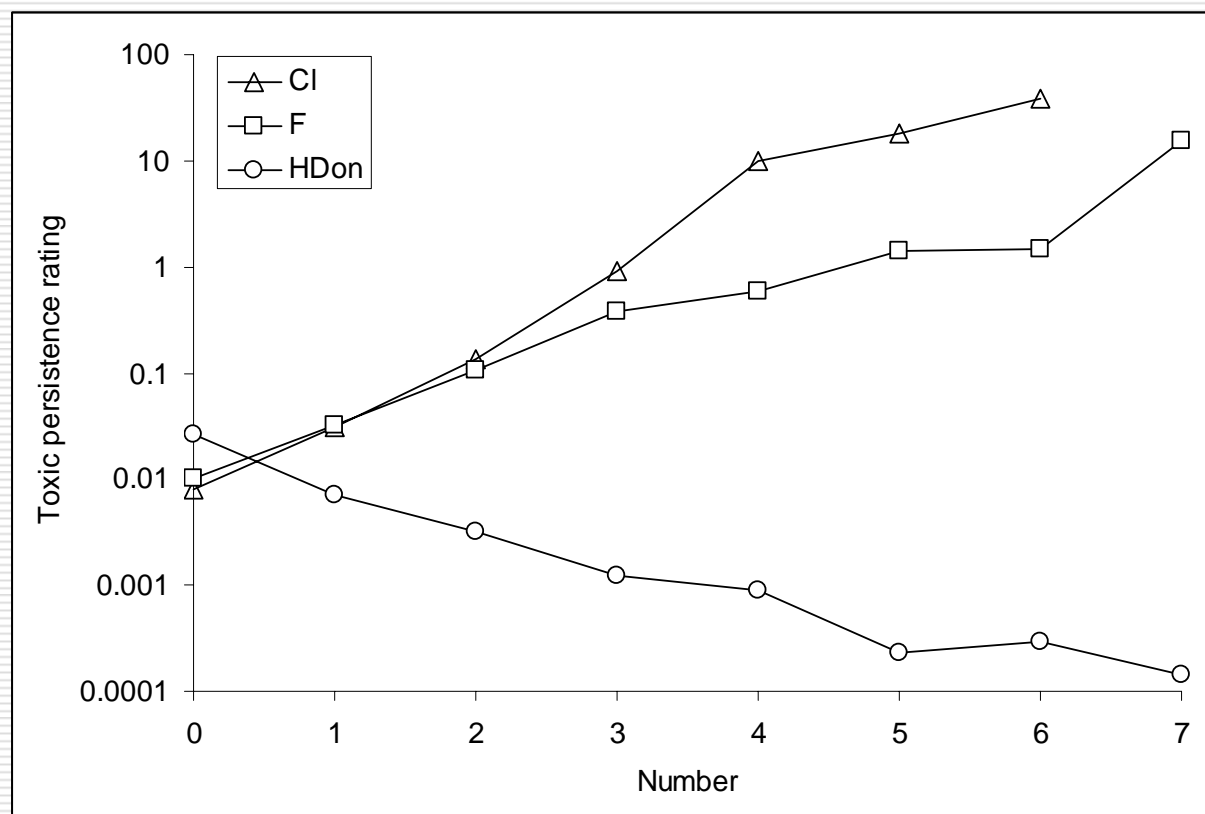


Kemin bakom

- Vilka faktorer som inverkar på QSAR-modellerna redovisas i artiklarna
 - Icke-specifik toxicitet: Storlek och polära interaktioner
 - Atmosfärisk persistens: Halogenering och reaktiva strukturelement
- TPR-rankingen styrs givetvis av samma faktorer



Monotona samband



TPR-ranking som en funktion av antalet klor, fluor resp. väteatomer bundna till syre eller kväve.

Slutsatser

- Kemoinformatik och beräkningskemi erbjuder rationella verktyg för utveckling av produkter som är lätt nedbrytbara och har låg toxicitet
- En än viktigare egenskap är att det kan stimulera till ett alternativt tänkande kring miljörisker
 - Tidigare ansatser har varit inriktade på att sälla fram ett fåtal riskfyllda ämnen
 - Mitt förslag är att istället använda skattningar av toxicitet och persistens för att beräkna den samlade miljöbelastningen

Nästa steg

- Modeller som kan användas för att skatta fördelning i miljön kan förbättra riskuppskattningen
 - Ångtryck och dess temperaturberoende är en nyckelparameter
- Att testa modellerna på dokumenterade materialflöden är nästa stora utmaning
 - Det svenska produktregistret är en tänkbar informationskälla för det

Hjälp behövs

- Virtuellt screening är ett nytt inslag på miljöområdet, särskilt vad gäller stora databaser
- Jag är mycket tacksam för alla synpunkter, tips och uppslag från Er, t.ex. avseende:
 - Metodval
 - Programvaror
 - Ämnesdatabaser

