

Bestemmelse af dioxin i fiskemel ud fra fedtsyresammensætningen

I forrige kemometriske klumme udforskede vi, hvordan de forskellige fedtsyrer i fiskemel indbyrdes korrelerer, og hvordan de er kvalitativt relaterede til dioxinindholdet. Her vises en kvantitativ model, der ud fra fedtsyresammensætningen forudsiger mængden af dioxin i fiskemelet

Af Karin Kjeldahl, Thomas Skov, Søren Balling Engelsen, Rasmus Bro og Lars Nørgaard, Københavns Universitet

Dioxiner er giftige og derfor særdeles uønskede i fødevarer. Især fiskeprodukter er i fokus, da dioxiner fra (naturlig og menneskeskabt) havforurening ophobes gennem fødekæden i fiskenes fedtvæv, og en hurtig og effektiv metode til kvantificering af dioxiner er derfor højst ønskværdig. I forrige klumme fandt vi ud af, at der er korrelationer mellem dioxinindholdet og fedtsyresammensætningen, som måske skyldes, at fedtsyresammensætningen afspejler fiskens føde, art, niveau i fødekæden og fangststed. Vi vil her forsøge at opbygge en kvantitativ model, der kan give et bud på dioxinindholdet i en given fiskemelsprøve ud fra dens fedtsyresammensætning. Værktøjet til det er *Partial Least Squares* (PLS) regression, som tidligere er beskrevet i Det kemometriske Rum (Dansk Kemi 11, 2008).

Data

Som beskrevet i Dansk Kemi 5, 2009 består datasættet af analyser af 64 fiskemelsprøver fra forskellige fede fiskearter fanget i en række farvande omkring Danmark over en periode på næsten to år. Fedtsyreprofiler med 32 fedtsyrer er bestemt vha. GC med flammeioniseringsdetektion (GC-FID). Dioxin er bestemt med HRGC-HRMS efter WHO standard (se mere i [1]) og varierer mellem 1,1 og 47,1 ng toksiske ækvivalenter (TEQ).

Til PLS-modellerne opdeles data i en X-blok med GC-målingerne $[64 \times 32]$ og en y-blok med den tilhørende dioxinværdi $[64 \times 1]$. Da det forventes, at fedtsyrer i små koncentrationer kan være lige så væsentlige som fedtsyrer i høje koncentrationer, autoskaleres X-blokken. Med kun én variabel (dioxin) i y-blokken har skalering ingen betydning, så y-blokken centrerer kun.

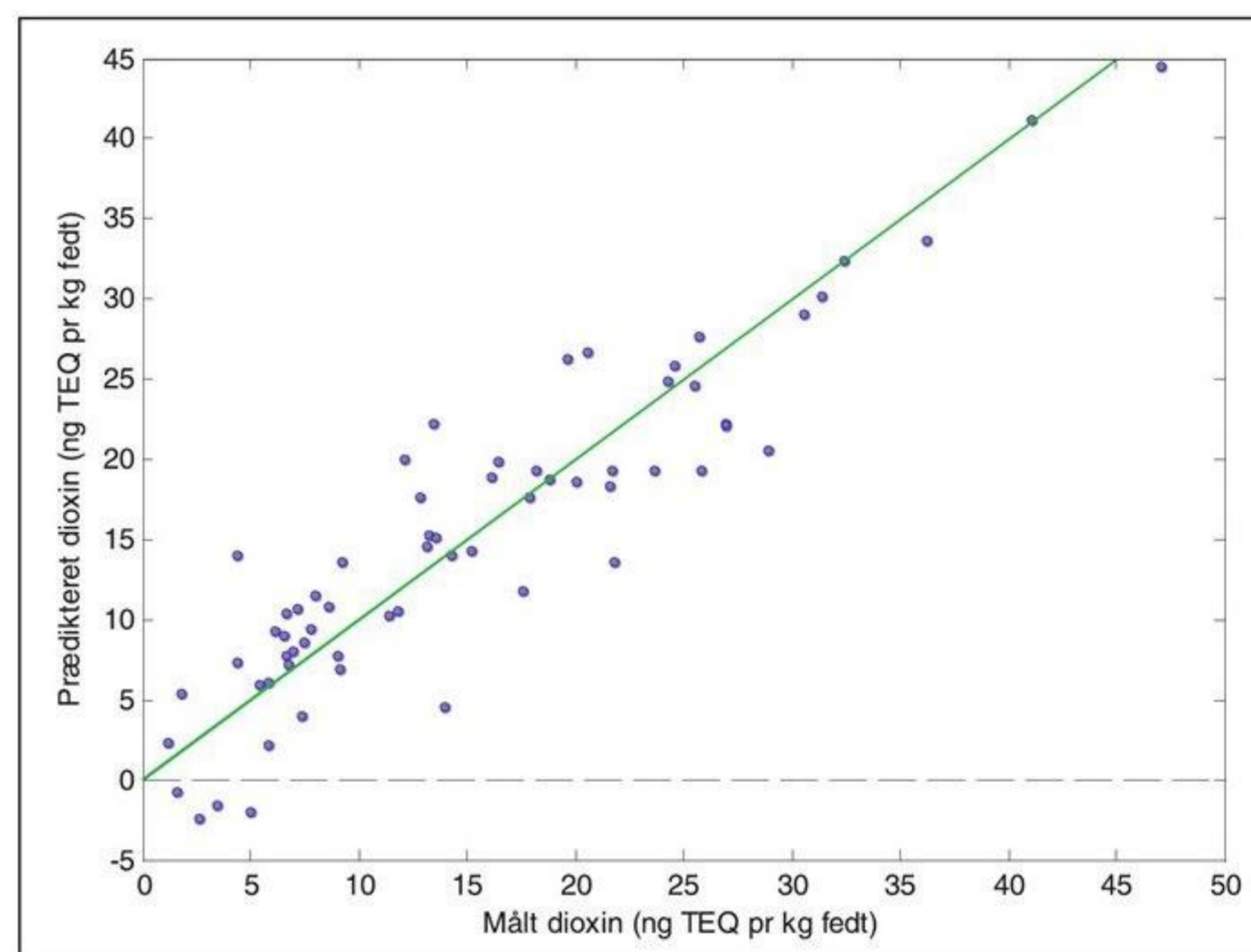
Kan vi overhovedet lave en model?

Først beregnes en PLS-model, der kvantitativt relaterer mønstret af samtlige 32 fedtsyrer til dioxinindholdet. Resultatet er bl.a. en regressionsvektor, der angiver en vægtning for hver fedtsyre (evt. negativ), sådan at det summerede produkt over alle fedtsyrer giver en modelleret dioxinværdi. Hvis regressionsvektoren \mathbf{b} 's elementer b_1, b_2, \dots, b_{32} angiver en vægt for hver af de 32 fedtsyrevariable i \mathbf{X} , så kan dioxinværdien y for den i 'te prøve beregnes således:

$$y_i = b_1 x_{i,1} + b_2 x_{i,2} + \dots + b_{32} x_{i,32} + \text{offset}$$

For en god model er der fin overensstemmelse mellem de prædikterede værdier og referenceværdierne, for hver prøve viser, at vi med vores PLS-model med 6 komponenter kan prædiktere dioxinindholdet med en gennemsnitlig krydsvalideringsfejl på

4.2 ng TEQ pr. kg fedt, hvilket må siges at være en fornuftig model set over hele koncentrationsområdet (1,1-47,1 ng TEQ pr. kg fedt). Krydsvalideringsfejlen betegnes RMSECV (Dansk Kemi 8, 2008) og er et bud på, hvor stor en prædiktionsfejl vi kan forvente på fremtidige prædiktioner. Prædiktionsfejlen skyldes en kombination af utilstrækkelighed i modellen og måleusikkerhed på referenceværdierne og fedtsyreværdierne.

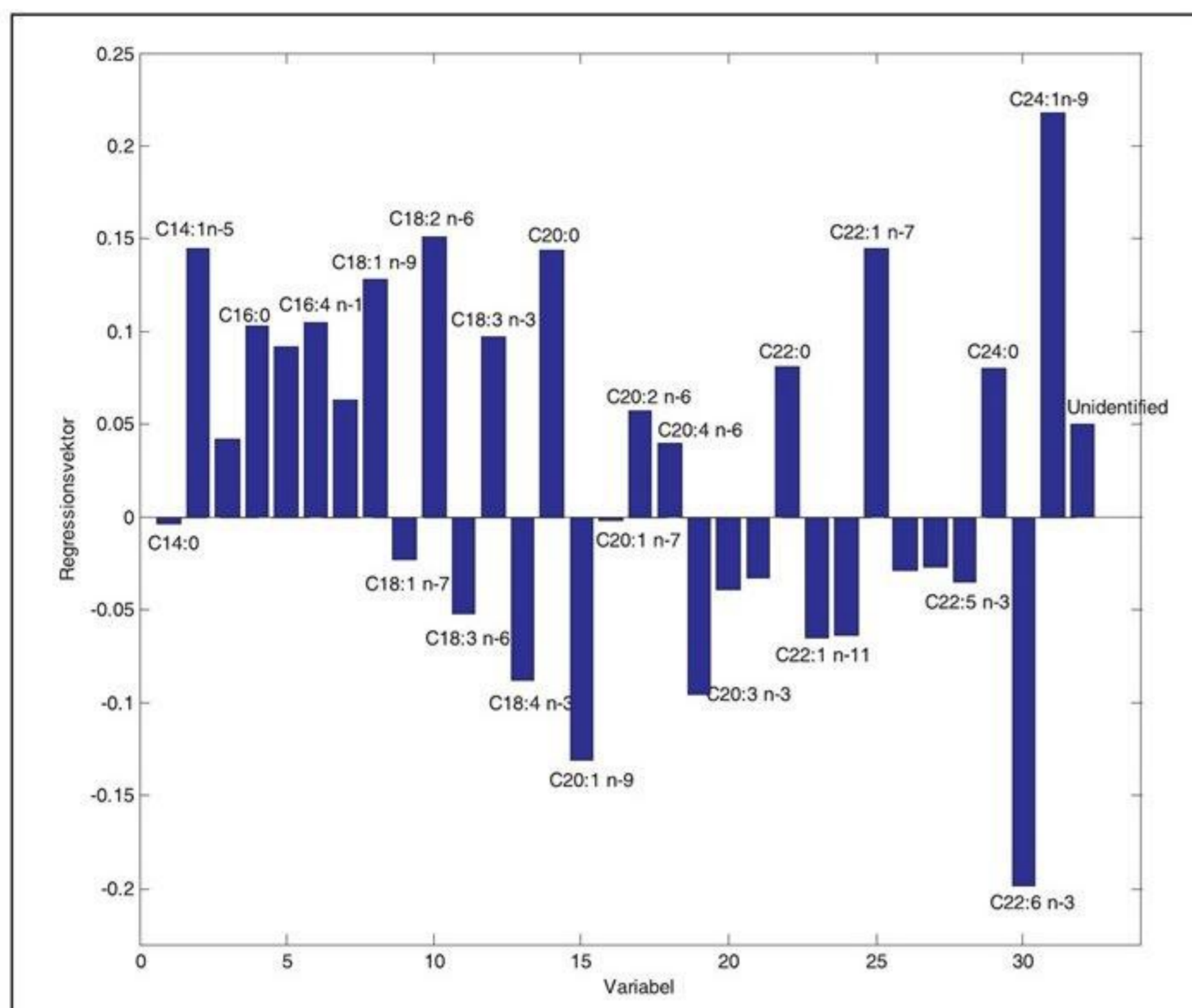


Figur 1. Prædikterede værdier vs. målte dioxinværdier for alle 64 prøver. Den gennemsnitlige fejl (RMSECV) er 4.2 ng TEQ pr. kg fedt. Under ideelle omstændigheder (dvs. perfekt model og målinger uden støj) skulle punkterne ligge på den grønne linje.

I PCA er score- og loadingplot nyttige i fortolkning, og noget tilsvarende findes også for PLS. Med loadings kan man i forskellige kombinationer af PLS-komponenter se de interne relationer mellem variablene, men altid kun for den del af variationen, der er forklaret af de viste PLS-komponenter. Det betyder, at to variable godt kan være positivt korrelerede for nogle kombinationer af loadings og negativt for andre. Helt tilsvarende gælder for PCA-loadings. Det samlede billede over alle komponenter er opsummeret i regressionsvektoren, der viser hvor meget og med hvilket fortegn hver enkelt variabel bidrager til at forudsige y-variablen.

Regressionsvektoren for vores model (figur 2) viser, at de fire vigtigste fedtsyrer er C24:1 n-9, C18:2 n-6, C22:6 n-3 og C20:1 n-9, og at disse to og to tilsyneladende er modsat korrelerede ift. dioxin – jo mere C24:1 n-9 eller C18:2 n-6, des højere dioxin-

indhold og modsat for C22:6n-3 og C20:1n-9. Derimod synes eksempelvis C14:0 og C20:1 n-7 at være tilnærmelsesvis uden betydning for forudsigelsen af dioxinindholdet. Nogle af disse konklusioner kan umiddelbart virke som lidt modstridende med, hvad der blev fundet med PCA (Dansk Kemi 5, 2009, men det skyldes netop, at kun de to første PCA-komponenter blev inspiceret, og at de "kun" forklarede cirka 59% af variationen.



Figur 2. Regressionsvektoren for en 6-komponent PLS-model med alle 32 fedtsyrer. Variable med stor positiv eller negativ vægt betyder mest for dioxinbestemmelsen.

Ved PLS-modellering er fokus ofte mindre på det eksplorative og mere på nøjagtige prædiktioner og en *robust* model, dvs. en model der er mindre følsom for, at kommende prøver måske adskiller sig en smule fra dem, modellen er bygget på. Det kan jo være, at der er små ændringer for de fisk, der fanges i fremtiden, ift. de fisk der er benyttet i dette datasæt, og det skulle modellen gerne kunne håndtere.

Kan vi gøre det lidt bedre?

En måde at forbedre modellen er ved at anvende *variabelselektion*. Man kan f.eks. vælge kun at medtage de fire vigtigste fedtsyrer fundet via figur 2. Det giver en simplere model med kun 3 PLS-komponenter, dvs. der er langt færre parametre i modellen, hvilket kan øge robustheden. Til gengæld falder nøjagtigheden på prædiktionerne; RMSECV bliver for denne model 4.6 ng TEQ pr. kg fedt. Sådan kan man på forskellig vis lege med forskellige kombinationer af variablerne for at finde den optimale løsning. Vi fandt her, at en kombination af 13 variable med 4 PLS-komponenter kunne give os en prædiktionsfejl på 3.3 ng TEQ pr. kg fedt, hvilket altså både forbedrer og forenkler vores prædiktionsmodel.

Man kunne også forestille sig, at man var særligt interesseret i et bestemt koncentrationsområde, f.eks. lige omkring grænseværdien, og at man optimerede modellen i dette interval, evt. med yderligere prøvetagning. Vil man have et mere retvisende bud på den forventede fejl i fremtiden, er den bedste vej at gentage forsøget og teste modellen på et sæt nye prøver - såkaldt testsæt-validering.

Outro

Vi har her lavet en model, der er i stand til at bestemme en

vanskeligt tilgængelig værdi, dioxin, ud fra en fedtsyreprofil, der er forholdsvis enkel at bestemme. Der er tale om en model baseret på *indirekte* sammenhænge, nemlig at dioxinindholdet er relateret til fedtsyreprofilen via den art og fangstområde som fedtsyresammensætningen afspejler. Det er fuldt tilstrækkeligt for f.eks. rutinekontrol af mange prøver, men nye omstændigheder, f.eks. andre arter, fangstområder m.v., vil kunne bryde den indirekte korrelation og ødelægge modellens validitet.

E-mail-adresser:

Karin Kjeldahl: kkd@life.ku.dk

Thomas Skov: thsk@life.ku.dk

Søren Balling Engelsen: se@life.ku.dk

Rasmus Bro: rb@life.ku.dk

Lars Nørgaard: lan@life.ku.dk

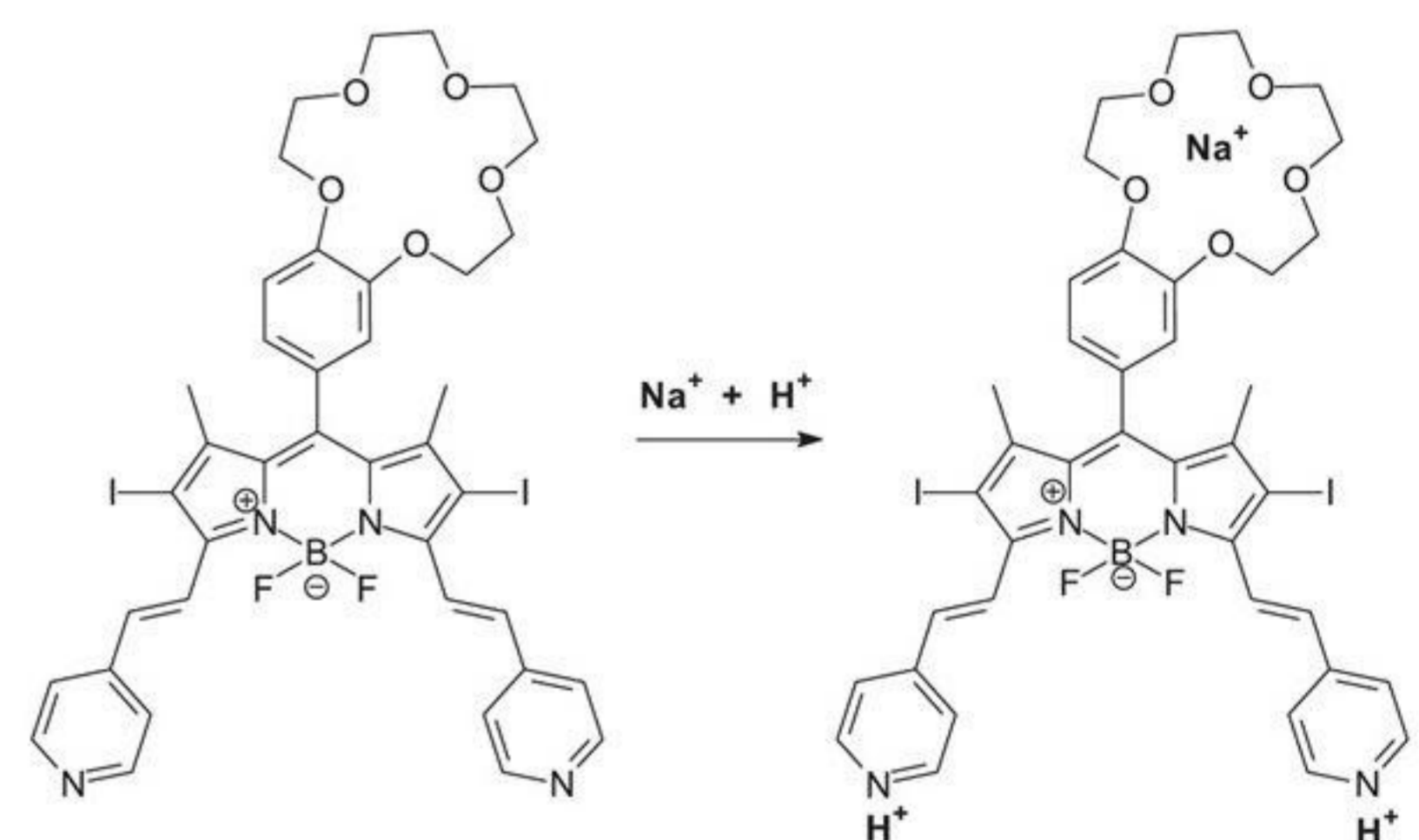
Referencer

1. Rapid dioxin assessment in fish products by fatty acid pattern recognition, Marc Bassompierre, Lars Munck, Rasmus Bro and Søren Balling Engelsen, *Analyst*, 129, 553-558, 2004.

Nyt om...

... Intelligente cancermidler

Tyrkiske forskere har fremstillet et cancermiddel, der skal genkende høje koncentrationer af både protoner og natriumioner, før det kan virke. Sådanne koncentrationer findes



inde i en cancercelle, hvor pH kan være så lavt som 4. Når ovenstående forbindelse har konstateret begge disse krav opfyldt, kan den fungere som sensitizer og ved belysning give anledning til dannelse af singlet oxygen, som dræber cellen.

Carl Th.

Molecular Logic Gates Meet Cancer Therapy, *Journal of the American Chemical Society* **131**, 2009, side 48