

## CRTL-ALT-SHIFT

**NMR er en yderst følsom og alsidig analytisk teknik. Metoden kan ligesom kromatografi give data, hvor toppenes positioner skifter fra prøve til prøve. På trods af at problemet ligner det for kromatografiske data, så kan det ikke effektivt løses med de samme værktøjer. Det kræver en skræddersyet algoritme, der kan håndtere de specielle udfordringer ved NMR-data**

**Af Francesco Savorani, Søren Balling Engelsen, Lars Nørgaard og Rasmus Bro, Institut for Fødevidenskab, Det Biovidenskabelige Fakultet, Københavns Universitet**

Fra at være en eksklusiv analytisk teknik for strukturanalyse bliver proton ( $^1\text{H}$ ) NMR nu ofte anvendt til forskellige områder lige fra fødevarer og autenticitet til ernæring, sundhed og medicin. En af styrkerne ved NMR er, at signalernes position er følsom over for det lokale kemiske miljø (kemisk skift), men NMR-signaler kan også være ganske følsomme over for en række eksterne faktorer som pH, temperatur og ionstyrke. Disse faktorer kan påvirke signalernes absolutte og relative position i NMR-spektret og i værste tilfælde give anledning til fejlfortolkninger.

Der er en stigende kvantitativ brug af NMR-spektre, hvor mange ensartede spektre analyseres vha. multivariate kemometriske metoder. For at en sådan kemometrisk analyse kan fungere optimalt, skal signalet for den samme kemiske komponent fremtræde ved de samme spektrale variable for alle prøver.

Vi har for nyligt introduceret en metode kaldet *interval Correlation Optimized SHIFTing algorithm (iCOSHIFT)*, der er specielt designet til at korrigere NMR-datasæt [1]. Metoden er baseret på lokale rigide skift af spektrale segmenter og benytter en speciel strategi ved segmentsamlingerne. For at reducere beregningstiden er metoden bygget på en Fast Fourier Transform (FFT)-beregningsmodel, som tillader, at selv store NMR-datasæt med mange tusinde variable kan beregnes inden for nogle få sekunder på en standard-pc.

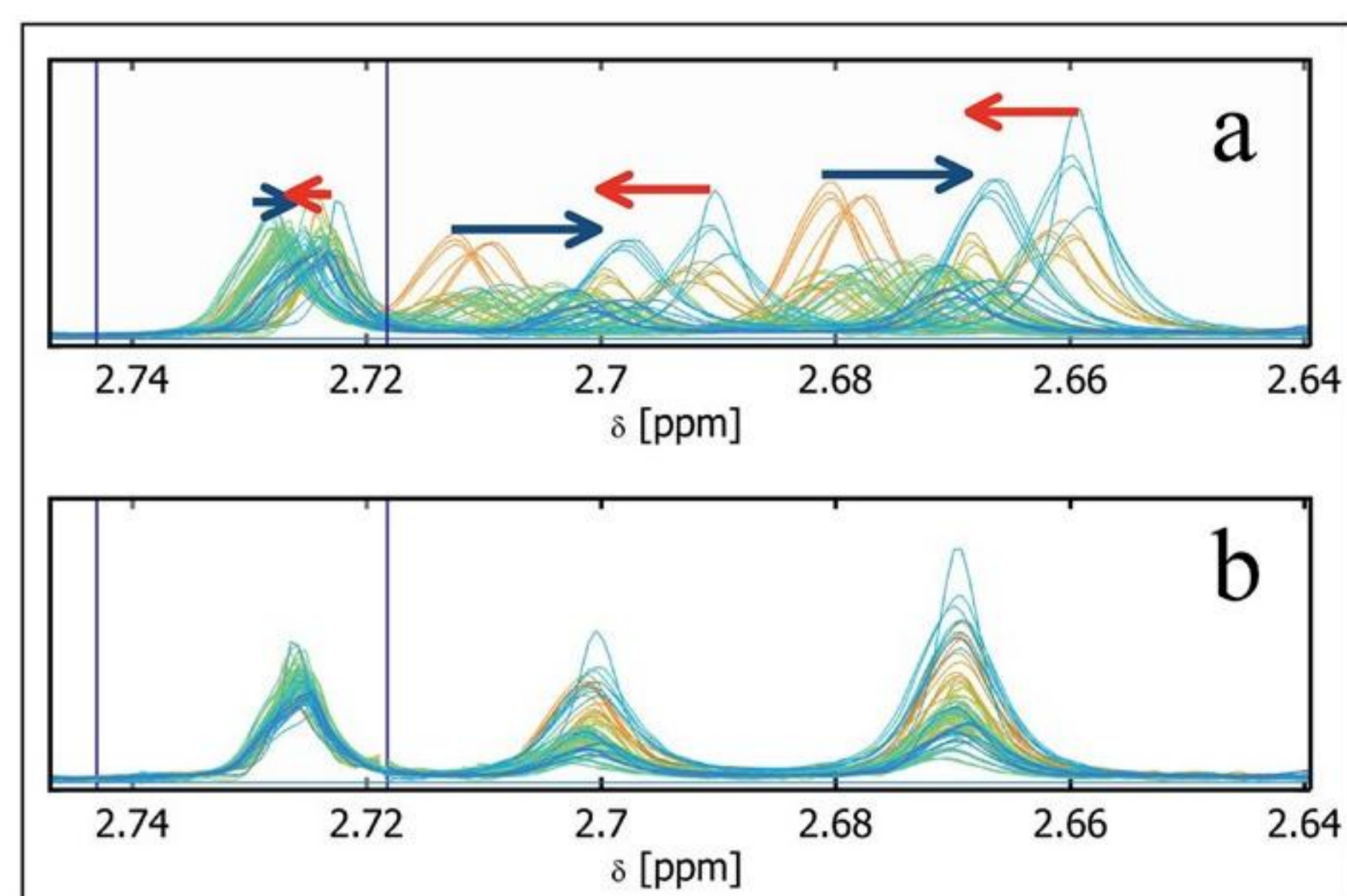
### Problem

I modsætning til kromatografiske data er det ikke kun position og højden af et NMR-signal, der indeholder den kemiske information, men også dets kurveform og multiplicitet. Det er derfor nødvendigt at en korrektionsmetode kan bevare så meget som muligt af de originale spektrale egenskaber. Vi har tidligere præsenteret algoritmen COW (Dansk Kemi nr. 9, 2010) [2], som er et effektivt redskab til at korrigere tidsaksen af kromatografiske toppe. COW er baseret på en strategi, der implicit forsøger at ændre kurveformen for at tilpasse den så nøjagtigt som muligt til en referencetop. Det virker fint for kromatografiske toppe, men er ikke hensigtsmæssigt til korrektion af NMR-data, da kurveformen her er informativ og derfor ikke bør ændres.

### Løsning

I *iCOSHIFT* rykkes et NMR-spektrum i små bidder, indtil hver bid har maks. korrelation med et tilsvarende stykke i et referencspektrum. Det udføres i praksis ved at flytte hvert spektrum

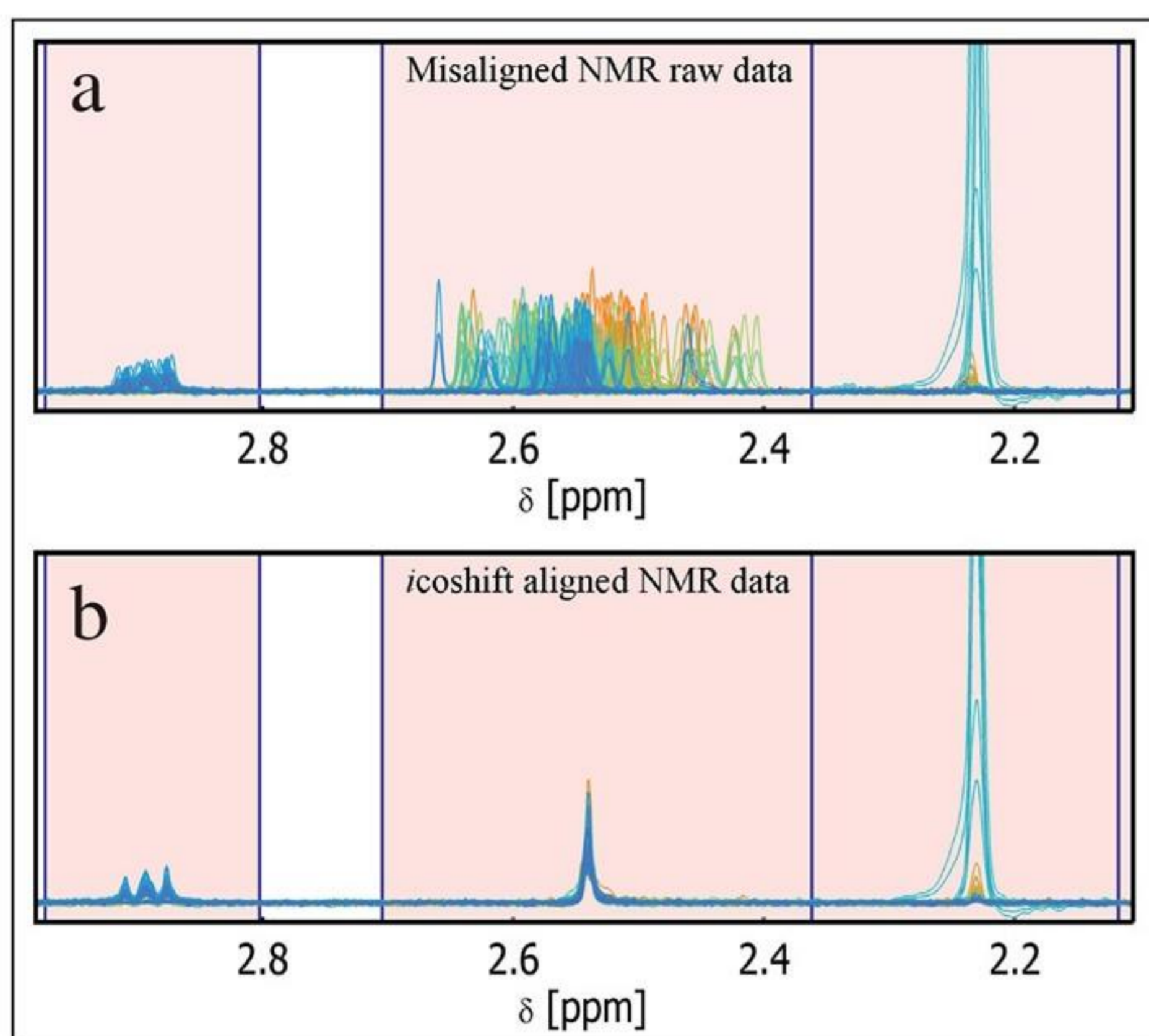
til venstre/højre indtil den bedste korrelation til referencspekret er opnået. Ethvert interval fra alle analyserede spektre kan flyttes i begge retninger, indtil en endelig bedste korrelation er opnået for hele datasættet. Metoden kræver, at det på alle intervalgrænser er nødvendigt enten at slette nogle punkter (som forhåbentlig kun indeholder støj/baseline) eller indsætte nogle nye punkter til at udfylde hullet forårsaget af det valgte skift, alt efter hvilken retning intervallet er blevet flyttet. Dette koncept er illustreret i figur 1, hvor tre NMR-toppe med forskydningsfejl er blevet korrigeret med *iCOSHIFT*.



Figur 1. Illustration af NMR-datakorrektion med *iCOSHIFT*. (A) En typisk NMR-situation hvor tre toppe er næsten fuldstændigt sammenblandet. Pilene angiver det nødvendige skift for nogle toppe for at flytte dem til en referenceposition (venstre-rød, højre-blå). (B) Den samme spektrale region efter at der er korrigeret med *iCOSHIFT*. De tre toppe er nu skarpt adskilt. De blå lodrette linjer indikerer intervalgrænser.

Men hvad forstås egentlig ved begrebet "interval"? Svaret på dette spørgsmål er kernen *iCOSHIFT*'s strategi. Det første skridt er netop at opsplutte NMR-datasættet i segmenter, kaldet intervaller. Disse intervaller kan defineres på mange forskellige måder. Intervallerne kan alle have samme størrelse, ligge i forlængelse af hinanden (tilstødende) eller de kan være af variabel størrelse og ikke-tilstødende. Intervallerne kan dække hele det spektrale område eller være fokuseret på én specifik og interessant region, der skal tilpasses. Når intervallerne er defineret, vil ethvert interval blive behandlet separat (men ud fra en overordnet strategi) i *iCOSHIFT* ved at lede efter den bedst mulige kor-





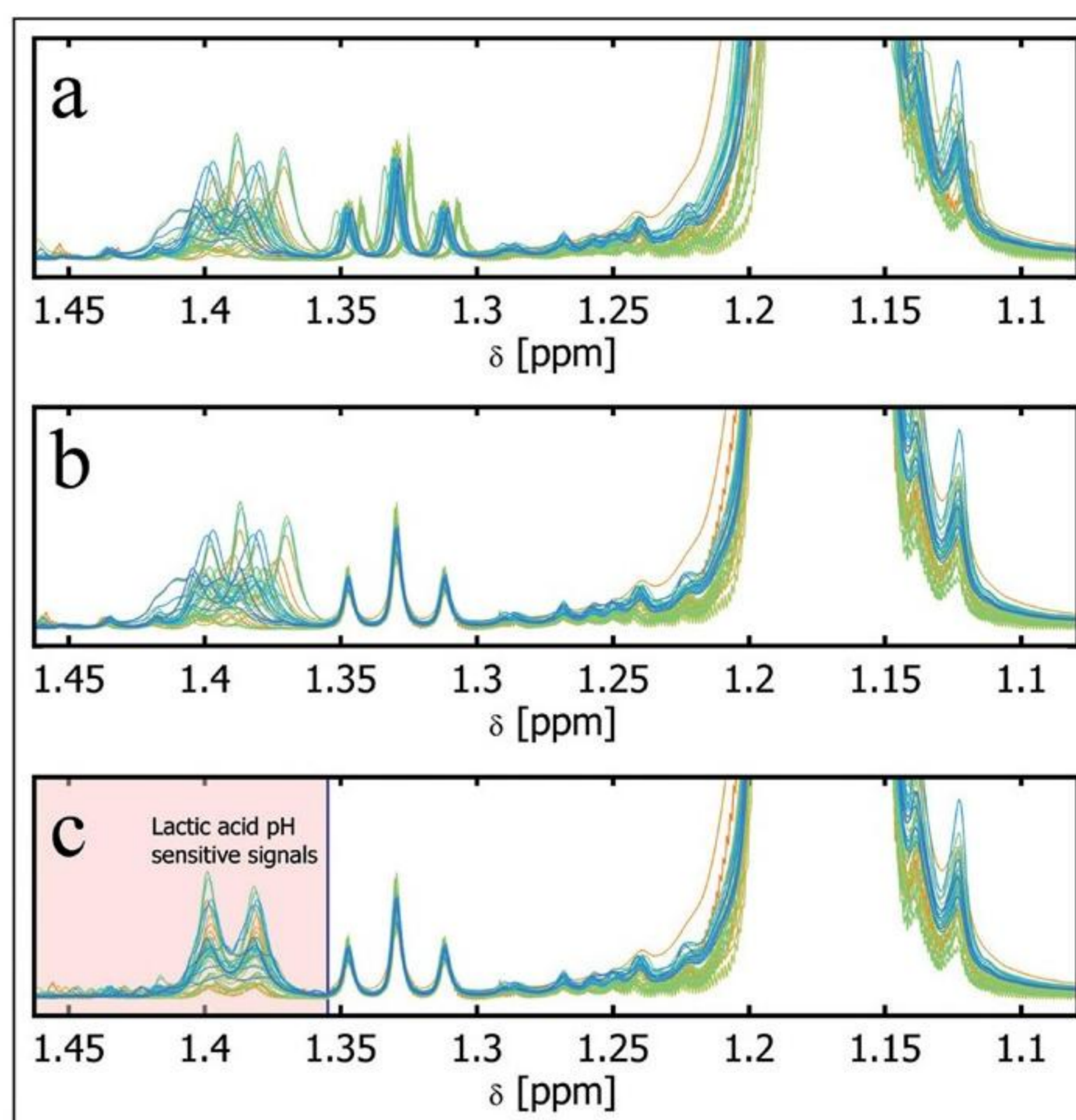
Figur 2. Illustration af intervalkorrektion. (A) Tre intervaller (lyserød baggrund) er defineret af de lodrette blå linjer. De tre intervaller repræsenterer tydeligvis forskellige grader af forskydningsfejl, som er et problem, der ikke kan løses vha. andre tilgængelige korrektionsværktøjer. (B) De tre intervaller efter automatisk iCOSHIFT-korrektion.

reaktion. Først ved afslutningen af optimeringsprocessen bliver de tilpassede intervaller forenet igen for at returnere et komplet korrigeret datasæt. Denne måde at arbejde selvstændigt på med flere intervaller er illustreret i figur 2.

Ved at anvende denne strategi på hele datasættet er det muligt at udføre en skræddersyet og optimal global korrektion på meget kort tid. Ikke desto mindre kan man benytte nogle tricks for at fremskynde beregningen yderligere eller i det mindste at gøre processen lettere. Normalt er det få toppe, der udviser en uafhængig forskydning i NMR-spektre, mens flertallet kun behøver at blive korrigeret til en fælles intern standard. Der er tale om f.eks. pH-afhængige toppe, der drastisk kan variere deres position, selv med ganske små pH-variationer imellem prøverne. I sådanne tilfælde er det en god idé først at tilpasse hele spektret til én fælles reference og derefter lade iCOSHIFT arbejde på et par udvalgte intervaller, der indeholder de pH-afhængige signaler. Denne situation er illustreret i figur 3 for NMR-spektre af forskellige bordvine. I figuren har vi zoomet ind på et område, der indeholder både pH-uafhængige signaler (ethanol) og den meget pH-afhængige dublet fra mælkesyre. Ved første iCOSHIFT-operation korrigeres spektret ift. de dominerende ethanol-signaler, og derefter arbejder iCOSHIFT lokalt på mælkesyreintervallet for at opnå et optimalt globalt resultat.

## Faldgruber

Når man bruger korrektionsalgoritmer, er de korrigerede data muligvis ikke længere repræsentative for de prøver, de stammer fra, og det er derfor altid en god ide at prøve at forstå, hvorfor en korrektion er nødvendig, og om den giver mening. I tilfældet med spektral korrektion er det sandsynligt, at de observerede forskydninger indeholder kemiske eller fysiske informationer



Figur 3. Illustration af to trin udført af iCOSHIFT for at løse forskellige typer af forskydningsfejl.

(A) Rå  $^1\text{H}$  NMR-spektre af vin domineret af den forskudte ethanolmethyl-triplet (1,17 ppm) og dens  $^{13}\text{C}$  satellit (1,33 ppm), hvor også mælkesyre-dubletten (1,39 ppm) er tydeligt forskudt. (B) Den første globale korrektion (trin 1) vha. iCOSHIFT: kun ethanol-tripletten er nu perfekt afstemt, mens det pH-følsomme signal fra mælkesyre (lyserød baggrund) stadig er dårligt ukorrigeret. (C) Den efterfølgende intervalbaserede korrektion vha. iCOSHIFT: mælkesyresignalet er nu korrigeret separat for at opnå en god global korrektion.

om prøverne, som er relevante for beskrivelsen af prøverne. Hvis det er tilfældet, kan vi under bestræbelserne på at opnå et perfekt afstemt datasæt miste vigtige oplysninger, som vi netop var på udkig efter.

## Outro

Vi har i denne klumme vist, hvordan  $^1\text{H}$ -NMR-data påvirkes af forskydningsfejl, der kan vanskeliggøre datafortolkning og føre til, at vigtige resultater overses. Vi har præsenteret iCOSHIFT-værktøjet som en hurtig og effektiv løsning til korrektion af forskydningsfejl inden for spektroskopi. Programmet er lavet som open source og kildetekst i MatLab, dokumentation og demo kan downloades frit fra [www.models.life.ku.dk](http://www.models.life.ku.dk).

## E-mail-adresser

Francesco Savorani: [frsa@life.ku.dk](mailto:frsa@life.ku.dk)  
 Søren Balling Engelsen: [se@life.ku.dk](mailto:se@life.ku.dk)  
 Lars Nørgaard: [lan@life.ku.dk](mailto:lan@life.ku.dk)  
 Rasmus Bro: [rb@life.ku.dk](mailto:rb@life.ku.dk)

## Referencer

- Savorani, F., Tomasi, G., and Engelsen, S.B. 2010. iCOSHIFT: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance*, 202 (2): 190-202.
- Tomasi, G., van den Berg, F., and Andersson, C. 2004. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, 18 (5): 231-241.