

Hanna Nedreberg Burud

# Proteinhydrolysater fra ulike restråstoff

Sammenheng mellom peptidstruktur og  
antioksidantkapasitet

Masteroppgave i Industriell kjemi og bioteknologi  
Veileder: Turid Rustad

Mars 2021



Hanna Nedreberg Burud

# **Proteinhydrolysater fra ulike restråstoff**

Sammenheng mellom peptidstruktur og  
antioksidantkapasitet

Masteroppgave i Industriell kjemi og bioteknologi  
Veileder: Turid Rustad  
Mars 2021

Norges teknisk-naturvitenskapelige universitet  
Fakultet for naturvitenskap  
Institutt for bioteknologi og matvitenskap



Kunnskap for en bedre verden





## Forord

Denne masteroppgaven ble gjennomført som en avsluttende del av det femårige studieprogrammet Industriell kjemi og bioteknologi ved NTNU. Oppgaven ble utført ved Institutt for bioteknologi og matvitenskap ved NTNU Trondheim. Arbeidet er en fortsettelse av prosjektoppgaven som ble utført våren 2020. Prosjektoppgaven var en rent teoretisk oppgave som resulterte i en kvalitativ og kvantitativ litteraturstudie om sammenhenger mellom peptidstruktur og bioaktivitet.

Jeg vil først og fremst rette en stor takk til min veileder, professor Turid Rustad, for god ekspertise, veiledning og oppmuntring gjennom hele semesteret. Dette året har vært annerledes, med mye hjemmekontor. Jeg setter stor pris på den gode oppfølgingen, og at du alltid har tatt deg tid til møter, spørsmål og faglige diskusjoner når det har vært behov for det. Jeg vil også takke professor i datavitenskap, Kristin Tøndel ved NMBU, for råd og svar om multivariat dataanalyse, og spesielt da om PLSR-analysen. I tillegg vil jeg takke lærling Peter Benjamin Kelley som gjorde målinger på noen av prøvene på lab, som så kunne bli en del av dataanalysen i denne masteroppgaven.

Til slutt vil jeg rette en stor takk til mine foreldre Tove og Jens Petter, og min samboer Magnus. Tusen takk for god støtte, oppmuntrende ord og motivasjon gjennom masteroppgaven, men også gjennom hele studiet.

---

Trondheim, mars 2021

Hanna Nedreberg Burud

## Abstract

The Norwegian fishing and aquaculture industry generates several hundreds of thousands of tonnes of rest raw material on an annual basis. The utilization of rest raw material has increased the last couple of years. In the whitefish sector there is still a substantial amount of fish heads and guts that is not being utilized. The development is heading towards increased conservation and utilization of the rest raw material. The market for animal feed is currently the largest market for marine rest raw material, and only a small share of the rest raw material is used for human consumption and in areas like cosmetics, pharmacy and supplements. There is an increasing interest for the production of high-value products. Among others, there is an increased interest for the use of hydrolyzed proteins for human consumption. It has been shown that peptides from fish proteins may possess important bioactive properties, like antioxidant capacity. Oxidative reactions in food cause severe loss of food quality, and distinct changes in both flavour, aroma and texture. Today there is an increased demand for antioxidants of natural origin. More knowledge is needed to understand relationships between molecular structure/composition of peptides and antioxidative properties.

Quantitative Structure-Activity-Relationship (QSAR) models can potentially reveal important relationships between peptide structure and properties in food. If these properties can be predicted from known peptide structures, then this knowledge could be used for the development and isolation of peptides with desirable properties. Statistical regression methods are used to determine the mathematical relationship between X and Y. The aim of this thesis was to use data analysis as a tool to study relationships between antioxidant capacity and different measures of peptide structure. Several master's theses have been completed at the Department of Biotechnology and Food Science at NTNU studying rest raw material from different types of fish and fish fractions. In this thesis several data matrices were constructed based on data from other completed master's theses. Partial least squares regression (PLSR) with Leave-One-Out Cross-Validation was used to examine each matrix. Different measures of peptide structure were used as predictor variables, including variables like % amino acid from total amino acid composition, % amino acid from the composition of free amino acids (FAA), the thiol content, and the degree of hydrolysis. ABTS antioxidant capacity was used as the response variable. The results of the modelling were visualized through different plots: validation plot, prediction plot, score plot, correlation loadings plot, coefficient plot, X and Y score plot and Variable importance for the projection (VIP) plot.

The results of the modelling showed that the two models with the largest data foundation described the datasets well, with relatively small differences between measured and predicted ABTS antioxidant capacity. Validation on multiple independent samples is necessary to confirm if the dataset represents relevant variation in ABTS antioxidant capacity. This is especially the case for the matrices that contained the least number of samples. Three of these

models showed limited ability to predict and thereby describe ABTS antioxidant capacity for different samples. The regression coefficients of the models with the most samples showed that some of the X-variables had greater impact on the modelling of ABTS antioxidant capacity. Among these were: Asp/Asn, Glu/Gln, Gly/Arg, Lys, Asp FAA and His FAA for matrix 1 and polar R, positively charged R, negatively charged R and aromatic R FAA for matrix 5. The main conclusion is that PLSR is a method that can be used to describe variation in these types of data, and therefore can be used to model and potentially also describe antioxidant capacity from variables as such as amino acid composition.

## Sammendrag

Fiskeri- og havbruksnæringen i Norge genererer flere hundre tusen tonn restråstoff årlig. Den samlede utnyttelsesgraden av restråstoff har økt de siste årene, men i hvitfisksektoren er det fortsatt en vesentlig andel hoder og slo som ikke blir utnyttet. Utviklingen går mot økt bevaring og utnyttelse av restråstoffet. I dag er førmarkedet det største markedet for marint restråstoff, og kun en liten andel går til humant konsum og bransjer som kosmetikk, farmasi og kosttilskudd. Det er likevel en økende interesse for produksjon av produkter med høyere verdi. Det er blant annet økt interesse for bruk av hydrolyserte proteiner til humant konsum. Det er vist at peptider fra fiskeproteiner kan ha viktige bioaktive egenskaper, blant annet antioksidantkapasitet. Oksidative reaksjoner i mat fører til vesentlig tap av matkvalitet, og store endringer i både smak, aroma og tekstur. Det er i dag større etterspørsel etter naturlige antioksidanter. Mer kunnskap er nødvendig for å forstå sammenhenger mellom henholdsvis molekylstruktur/sammensetning av peptider og antioksidative egenskaper.

*Quantitative Structure-Activity-Relationship* (QSAR) modeller kan potensielt avdekke relevante sammenhenger mellom peptidstruktur og egenskaper i mat. Hvis egenskapene kan predikeres for kjente peptidstrukturer, så kan slik kunnskap brukes til utvikling og isolering av peptider med spesielle egenskaper. Statistiske regresjonsmetoder brukes for å finne den matematiske sammenhengen mellom X og Y. Målet med masteroppgaven var å bruke dataanalyse som et verktøy for å studere sammenhenger mellom antioksidantkapasitet og ulike mål på peptidstruktur. Ved Institutt for bioteknologi og matvitenskap ved NTNU har det tidligere blitt gjennomført flere masteroppgaver der restråstoff fra ulike typer fisk og fiskefraksjoner har blitt studert. Med disse oppgavene som datagrunnlag ble det lagd flere datamatriser som inngikk i en dataanalyse. Det ble gjort en *Partial least squares regression* (PLSR) analyse med kryssvalidering (*Leave-One-Out Cross-Validation*) for hver matrise. Prediktorvariablene besto av ulike mål på peptidstruktur, og inkluderte variabler som % aminosyre fra total aminosyresammensetning, % aminosyre fra sammensetning av frie aminosyrer (FAA), mengde tioler og hydrolysegrad. Responsvariabelen var ABTS antioksidantkapasitet. Resultatene av modelleringen ble visualisert gjennom en rekke ulike plott: valideringsplott, prediksjonsplott, *scores*-plott, korrelasjon-*loadings*-plott, koeffisientplott, X og Y-*score*-plott og *Variable importance for the projection* (VIP)-plott.

Resultatene av modelleringen viser at de to modellene med størst datagrunnlag beskriver datasettene godt med relativt små avvik mellom målt og predikert ABTS-antioksidantkapasitet. Validering på flere uavhengige prøver er nødvendig for å svare på om datasettet representerer relevant variasjon i ABTS-antioksidantkapasitet. Dette gjelder spesielt for de matrixene som hadde færrest prøver. Tre av disse modellene viste begrenset evne til å predikere og dermed beskrive ABTS-antioksidantkapasitet på tvers av ulike prøver. De to modellene med størst datagrunnlag ga regresjonskoeffisienter som viste at noen X-variabler hadde vesentlig større betydning for modelleringen av ABTS

antioksidantkapasitet. Blant disse var Asp/Asn, Glu/Gln, Gly/Arg, Lys, Asp FAA og His FAA for matrise 1 og polare R, positivt ladete R, negativt ladete R og aromatiske R FAA for matrise 5. Hovedkonklusjonen er at PLSR er en metode som kan brukes til å beskrive variasjon i denne typen data, og som derfor kan modellere og dermed også potensielt beskrive antioksidantkapasitet fra variabler som aminosyresammensetning.

# Innhold

|   |          |
|---|----------|
| Figurer . . . . .   | ix       |
| Tabeller . . . . .  | xii      |
| Forkortelser . . . . .  | xv       |
| <b>1 Introduksjon</b>   | <b>1</b> |
| 1.1 Restråstoff fra marine ressurser . . . . .  | 1        |
| 1.1.1 Tilgang og bruk av restråstoff fra fiske og havbruk . . . . .                             | 3        |
| 1.1.2 Fisk som kilde til høyverdi komponenter . . . . .   | 4        |
| 1.1.3 Peptider som bioaktive komponenter . . . . .  | 5        |
| 1.1.4 Enzymatisk hydrolyse av proteiner . . . . .   | 7        |
| 1.2 Proteinstruktur og bioaktivitet . . . . .   | 9        |
| 1.2.1 Proteinstruktur . . . . .   | 9        |
| 1.2.2 Struktur-bioaktivitet sammenhenger . . . . .  | 11       |
| 1.3 Antioksidative peptider . . . . .   | 11       |
| 1.3.1 Oksidasjon av biologiske makromolekyler . . . . .   | 12       |
| 1.3.2 Antioksidative mekanismer . . . . .   | 13       |
| 1.3.3 <i>In vitro</i> analyser for måling av antioksidantkapasitet . . . . .                    | 15       |
| 1.4 Kjemometriske metoder i analyse av proteiner og peptider . . . . .                          | 18       |
| 1.4.1 Bruk av kjemometriske metoder i analyse . . . . .   | 18       |
| 1.4.2 Prinsippal komponent analyse og regresjon . . . . .                                       | 19       |
| 1.4.3 Delvis minste-kvadratets regresjon . . . . .  | 20       |
| 1.4.4 Validering av statistiske modeller . . . . .  | 22       |
| 1.4.5 Utfordringer ved dagens bruk av multivariat dataanalyse for analyse av bioaktive peptider | 24       |
| 1.5 Målet med masteroppgaven . . . . .  | 25       |
| 1.5.1 Relevante begreper i statistisk analyse . . . . .   | 26       |

|          |   |            |
|----------|---|------------|
| <b>2</b> | <b>Materialer og metoder</b>  | <b>27</b>  |
| 2.1      | Fremgangsplan . . . . .   | 28         |
| 2.2      | Innsamling og forhånds-behandling av data . . . . .                           | 28         |
| 2.2.1    | Datagrunnlaget . . . . .  | 29         |
| 2.2.2    | Utvelging og sortering av variabler og observasjoner . . . . .                | 30         |
| 2.2.3    | Beregninger . . . . .   | 33         |
| 2.2.4    | Sortering av dataene i matriser . . . . .                                     | 34         |
| 2.2.5    | Generell analyse av dataene . . . . .   | 36         |
| 2.3      | Delvis minste-kvadraters regresjon (PLSR) . . . . .                           | 37         |
| 2.3.1    | Standardisering av variabler . . . . .  | 37         |
| 2.3.2    | PLSR-modellen . . . . .   | 38         |
| <b>3</b> | <b>Resultater og Diskusjon</b>  | <b>40</b>  |
| 3.1      | Forhånds-behandling av variablene . . . . .                                   | 41         |
| 3.1.1    | ABTS antioksidantkapasitet . . . . .  | 41         |
| 3.2      | Matrise 1: TAA, FAA, ABTS antioksidantkapasitet . . . . .                     | 41         |
| 3.2.1    | Oversikt over dataene: spredning og korrelasjon . . . . .                     | 41         |
| 3.2.2    | Delvis minste-kvadraters regresjon (PLSR) . . . . .                           | 44         |
| 3.3      | Matrise 5: Aminosyredeskriptorer, ABTS antioksidantkapasitet . . . . .        | 53         |
| 3.3.1    | Oversikt over dataene: Spredning og korrelasjon . . . . .                     | 53         |
| 3.3.2    | Delvis minste-kvadraters regresjon (PLSR) . . . . .                           | 55         |
| 3.4      | Matrise 2 og 6: TAA, FAA, tioler, ABTS antioksidantkapasitet . . . . .        | 59         |
| 3.5      | Matrise 4 og 8: TAA, FAA, hydrolysegrad, ABTS antioksidantkapasitet . . . . . | 64         |
| 3.6      | Vurdering av litteratur som kilde til data . . . . .                          | 69         |
| <b>4</b> | <b>Konklusjon</b>   | <b>71</b>  |
| <b>5</b> | <b>Fremtidig perspektiv</b>   | <b>73</b>  |
|          | <b>Bibliografi</b>  | <b>i</b>   |
| <b>A</b> | <b>Rådata og beregninger</b>  | <b>vii</b> |
| A.1      | Oppdrettslaks . . . . .   | vii        |
| A.2      | Regnbueørret . . . . .  | xiii       |
| A.3      | Sei . . . . .   | xvi        |
| A.4      | Torsk . . . . .   | xviii      |

|          |  |            |
|----------|--|------------|
| <b>B</b> | <b>Dataanalyse</b>   | <b>xxi</b> |
| B.1      | Pearson korrelasjonskoeffisient . . . . .                                      | xxi        |
| B.2      | Datascript . . . . .   | xxi        |
| B.2.1    | Pearson korrelasjonsmatrise . . . . .  | xxi        |
| B.2.2    | PLSR-analyse . . . . .   | xxii       |
| B.3      | Matrise 1: TAA, FAA, ABTS antioksidantkapasitet . . . . .                      | xxv        |
| B.4      | Matrise 5: Aminosyredeskriptorer, ABTS antioksidantkapasitet . . . . .         | xxx        |
| B.5      | Matrise 2: TAA, FAA, tioler, ABTS antioksidantkapasitet . . . . .              | xxxvii     |
| B.6      | Matrise 6: Aminosyredeskriptorer, tioler, ABTS antioksidantkapasitet . . . . . | xlvii      |
| B.7      | Matrise 4: TAA, FAA, hydrolysegrad, ABTS antioksidantkapasitet . . . . .       | lix        |
| B.8      | Matrise 8: TAA, FAA, hydrolysegrad, ABTS antioksidantkapasitet . . . . .       | lxix       |



# Figurer

|     |   |    |
|-----|---|----|
| 1.1 | ACE sin rolle i Renin-Angiotensin systemet [25]. Denne figuren ble også vist i prosjektoppgaven våren 2020 [20]. . . . .  | 6  |
| 1.2 | Oversikt over den enzymatiske hydrolyseprosessen for å danne fiskeproteinhydrolysater. Figuren er en modifisert utgave av figuren fra Kristinsson et al (2000). [11] . . . . .  | 7  |
| 1.3 | Trypsin som hydrolyserer peptidbindingen ved aminosyreresten Lys. Figuren er en modifisert utgave av figuren fra boken Food: The Chemistry of its Components. [17] . . . . .  | 8  |
| 1.4 | Inndeling av aminosyrer etter R-sidegrupper innen kategoriene (ved pH 7): aromatiske, positivt ladete, negativt ladete, polare uladete og upolare alifatiske R-grupper. Figuren er opprinnelig fra boken Principles of Biochemistry, men rammene har blitt modifisert. [18] . . . . .                               | 10 |
| 1.5 | Plott som viser dataobservasjoner for en variabel 1 og en variabel 2. Linjene indikerer de to første PC-ene, som er plassert i retning størst varians. Det er også markert at de er ortogonale. Figuren er ikke basert på et reelt datasett, men er laget for å illustrere prinsippet bak PCA. . . . .              | 20 |
| 1.6 | Parametere i PLSR modellen: x-variabler $x_{ik}$ , y-variabler $y_{im}$ , x-score T, y-score U, loadings P', X-vektorer W', Y-vektorer C'. Indekser er gitt for matrisene: N for antall objekter, K for antall X variabler og M for antall Y variabler. Figuren er modifisert fra Wold et al (2001). [38] . . . . . | 22 |
| 2.1 | Flytskjema som viser trinnene i dataanalysen. . . . .   | 27 |
| 3.1 | Boksplokk av matrise 1 som ble brukt i dataanalyse. . . . .   | 42 |
| 3.2 | Pearson korrelasjon i form av en korrelasjonsmatrise for matrise 1. . . . .   | 43 |
| 3.3 | Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 1. . . . .  | 45 |
| 3.4 | Scores-plott for de to første PLS komponentene fra den kryssvaliderte PLSR-modellen av matrise 1. . . . .   | 46 |
| 3.5 | Korrelasjons-loadings plott for de to første PLS komponentene fra den kryssvaliderte PLSR-modellen av matrise 1. . . . .  | 49 |
| 3.6 | Regresjonskoeffisienter for de ulike variablene i den kryssvaliderte modellen av matrise 1 for 3 komponenter . . . . .  | 50 |

|      |   |        |
|------|---|--------|
| 3.7  | VIP-verdiene for variablene i den kryssvaliderte modellen av matrise 1 for 3 komponenter. . . . .   | 53     |
| 3.8  | Boksplott over de originale variablene i matrise 5. . . . .   | 54     |
| 3.9  | Pearson korrelasjon i form av en korrelasjonsmatrise for matrise 5. . . . .   | 55     |
| 3.10 | <i>Scores</i> -plott for de to første komponentene i PLSR modellen for matrise 5. . . . .   | 56     |
| 3.11 | Korrelasjon- <i>loadings</i> for de to første komponentene i PLSR-modellen for matrise 5. . . . .   | 57     |
| 3.12 | Regresjonskoeffisientene for de tre valgte komponentene i PLSR-modellen til matrise 5. . . . .  | 58     |
| 3.13 | <i>Scores</i> for de to første komponentene i PLSR-modellen for matrise 2. . . . .  | 61     |
| 3.14 | <i>Scores</i> for de to første komponentene i PLSR-modellen for matrise 6. . . . .  | 62     |
| 3.15 | Korrelasjon- <i>loadings</i> for de to første komponentene i PLSR-modellen for matrise 2. . . . .   | 63     |
| 3.16 | Regresjonskoeffisientene for de fire komponentene i PLSR-modellen for matrise 2. . . . .  | 64     |
| 3.17 | <i>Scores</i> for de to første komponentene i PLSR-modellen for matrise 4. . . . .  | 66     |
| 3.18 | Korrelasjon- <i>loadings</i> for de to første komponentene i PLSR-modellen for matrise 4. . . . .   | 67     |
| 3.19 | Regresjonskoeffisientene for de tre komponentene i PLSR-modellen for matrise 4. . . . .   | 68     |
| A.1  | Standardkurve hvor absorbans ved 734 nm er plottet som en funksjon av konsentrasjon av propylgallat (uM eller umol/L). Lineær trendlinje viser: $y = -0,0069x + 0.6504$ og $R^2=0.9624$ . . . . . | xi     |
| A.2  | Standardkurve hvor absorbans ved 734 nm er plottet som en funksjon av konsentrasjon av propylgallat (uM eller umol/L). Lineær trendlinje viser: $y = -0,0074x + 0.6641$ og $R^2=0.9999$ . . . . . | xv     |
| B.1  | R datascript for beregning og visualisering av korrelasjonsmatrisen til matrise 1. . . . .  | xxii   |
| B.2  | R datascript for PLSR-analysen av matrise 1, del A. . . . .   | xxiii  |
| B.3  | R datascript for PLSR-analysen av matrise 1, del B. . . . .   | xxiv   |
| B.4  | R datascript for PLSR-analysen av matrise 1, del C. . . . .   | xxv    |
| B.5  | Boksplott av matrise 1 etter standardisering av alle variabler. . . . .   | xxv    |
| B.6  | Sammendrag av PLSR-analysen for matrise 1. . . . .  | xxvi   |
| B.7  | MSEP for den kryssvaliderte modellen med ulikt antall komponenter for matrise 1. . . . .  | xxvi   |
| B.8  | Plott av X og Y-scores for PLSR-modellen med tre komponenter for matrise 1. . . . .   | xxvii  |
| B.9  | <i>Scores</i> -plott for alle komponentene i PLSR-modellen fra matrise 1. . . . .   | xxviii |
| B.10 | Korrelasjon- <i>loadings</i> -plott for alle komponentene i PLSR-modellen fra matrise 1. . . . .  | xxix   |
| B.11 | Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 1, gjennom funksjonen jack.test. . . . .  | xxx    |
| B.12 | Boksplott over variablene i matrise 5 etter standardisering. . . . .  | xxxi   |
| B.13 | Sammendrag av PLSR-analysen av matrise 5 . . . . .  | xxxi   |
| B.14 | MSEP for den kryssvaliderte modellen med ulikt antall komponenter for matrise 5. . . . .  | xxxii  |

|      |  |         |
|------|--|---------|
| B.15 | Plott over X og Y-scores for alle komponentene i PLSR-modellen for matrise 5. . . . .  | xxxiii  |
| B.16 | Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 5. . . . .             | xxxiv   |
| B.17 | Scores-plott for alle komponentene i PLSR-modellen for matrise 5. . . . .  | xxxv    |
| B.18 | Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen for matrise 5. . . . .                                      | xxxvi   |
| B.19 | VIP-verdiene for variablene i den kryssvaliderte modellen av matrise 5 for alle komponentene. . .                            | xxxvii  |
| B.20 | Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 5, gjennom funksjonen jack.test. . . . . | xxxvii  |
| B.21 | Boksplott over de originale variablene i matrise 2. . . . .  | xxxviii |
| B.22 | Boksplott over de standardiserte variablene i matrise 2. . . . .   | xxxix   |
| B.23 | Korrelasjonsmatrisen for variablene i matrise 2. . . . .   | xl      |
| B.24 | Sammendrag av PLSR-analysen for matrise 2. . . . .   | xl      |
| B.25 | MSEP for den kryssvaliderte PLSR-modellen med ulikt antall komponenter for matrise 2. . . . .                                | xli     |
| B.26 | Plott over X og Y-scores for de fire komponentene i PLSR-modellen for matrise 2. . . . .                                     | xlii    |
| B.27 | Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 2. . . . .             | xliii   |
| B.28 | Scores-plott for alle komponentene i PLSR-modellen for matrise 2. . . . .  | xliv    |
| B.29 | Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen for matrise 2. . . . .                                      | xliv    |
| B.30 | VIP-verdiene for variablene i PLSR-modellen med alle fire komponentene for matrise 2. . . . .                                | xlvi    |
| B.31 | Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 2, gjennom funksjonen jack.test. . . . . | xlvii   |
| B.32 | Boksplott over de originale variablene i matrise 6. . . . .  | xlviii  |
| B.33 | Boksplott over de standardiserte variablene i matrise 6. . . . .   | xlix    |
| B.34 | Korrelasjonsmatrisen for variablene i matrise 6. . . . .   | l       |
| B.35 | Sammendrag av PLSR-analysen for matrise 6. . . . .   | l       |
| B.36 | MSEP for den kryssvaliderte modellen med ulikt antall komponenter for matrise 6. . . . .                                     | li      |
| B.37 | Plott over X og Y-scores til PLSR-modellen med alle fire komponentene for matrise 6. . . . .                                 | lii     |
| B.38 | Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 6. . . . .             | liii    |
| B.39 | Scoresplott for alle komponentene i PLSR-modellen for matrise 6. . . . .   | liv     |
| B.40 | Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen for matrise 6. . . . .                                      | lv      |
| B.41 | Korrelasjon-loadings-plott for de to første komponentene i PLSR-modellen for matrise 6. . . . .                              | lvi     |
| B.42 | Regresjonskoeffisientene for variablene i matrise 6 fra PLSR-modellen med fire komponenter. . .                              | lvii    |

|      |  |         |
|------|--|---------|
| B.43 | VIP-verdiene for variablene i den kryssvaliderte modellen av matrise 6 for fire komponenter . . . .                          | lviii   |
| B.44 | Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 6, gjennom funksjonen jack.test. . . . . | lix     |
| B.45 | Boksplokk over de originale variablene i matrise 4. . . . .  | lx      |
| B.46 | Boksplokk over de standardiserte variablene i matrise 4. . . . .   | lxi     |
| B.47 | Korrelasjonsmatrisen for variablene i matrise 4. . . . .   | lxii    |
| B.48 | Sammendrag av PLSR-analysen for matrise 4. . . . .   | lxiii   |
| B.49 | MSEP for den kryssvaliderte PLSR-modellen med ulikt antall komponenter for matrise 4. . . . .                                | lxiii   |
| B.50 | Plott over X og Y-scores for de tre komponentene i PLSR-modellen for matrise 4. . . . .                                      | lxiv    |
| B.51 | Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 4. . . . .             | lxv     |
| B.52 | Scores-plott for alle komponentene i PLSR-modellen for matrise 4. . . . .  | lxvi    |
| B.53 | Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen for matrise 4. . . . .                                      | lxvii   |
| B.54 | VIP-verdiene for variablene i den kryssvaliderte modellen for matrise 4 med tre komponenter. . . .                           | lxviii  |
| B.55 | Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 4, gjennom funksjonen jack.test. . . . . | lxix    |
| B.56 | Boksplokk over de originale variablene i matrise 8. . . . .  | lxx     |
| B.57 | Boksplokk over de standardiserte variablene i matrise 8. . . . .   | lxxi    |
| B.58 | Korrelasjonsmatrisen for variablene i matrise 8. . . . .   | lxxii   |
| B.59 | Sammendrag av PLSR-analysen for matrise 8. . . . .   | lxxiii  |
| B.60 | MSEP for den kryssvaliderte PLSR-modellen med ulikt antall komponenter for matrise 8. . . . .                                | lxxiii  |
| B.61 | Plott over X og Y-scores for PLSR-modellen med tre komponenter for matrise 8. . . . .  | lxxiv   |
| B.62 | Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 8. . . . .             | lxxv    |
| B.63 | Scores-plott for de to første komponentene i PLSR-modellen for matrise 8. . . . .  | lxxvi   |
| B.64 | Scores-plott for alle komponentene i PLSR-modellen for matrise 8. . . . .  | lxxvii  |
| B.65 | Korrelasjon-loadings-plott for de to første komponentene i PLSR-modellen for matrise 8. . . . .                              | lxxviii |
| B.66 | Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen for matrise 8. . . . .                                      | lxxix   |
| B.67 | Regresjonskoeffisientene for variablene i den kryssvaliderte modellen av matrise 8 for tre komponenter. . . . .              | lxxx    |
| B.68 | VIP-verdiene for variablene i den kryssvaliderte PLSR-modellen av matrise 8 for tre komponenter. . . .                       | lxxxi   |
| B.69 | Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 8, gjennom funksjonen jack.test. . . . . | lxxxii  |

# Tabeller

|     |  |     |
|-----|--|-----|
| 1.1 | Oppsummerende oversikt av en rekke <i>in vitro</i> analyser for å studere antioksidantkapasitet [32]. . . . .  | 17  |
| 1.2 | Oversikt over relevante begreper i statistisk analyse . . . . .  | 26  |
| 2.1 | Oversikt over prøvene (hydrolysatene) som ble inkludert i dataanalysen, inkludert informasjon om råstoffet, enzym til hydrolyse og betydningen av prøvenavnene. RT: romtemperatur. . . . .                     | 33  |
| 2.2 | Oversikt over matrisene som ble brukt videre i dataanalyse, inkludert innhold av prøver (rader), X og Y-variabler (kolonner) . . . . .   | 35  |
| 2.3 | Aminosyredeskriptorer med tilhørende aminosyrer (gjelder for pH 7) [18]. . . . .   | 36  |
| 3.1 | ABTS antioksidantkapasitet (umol PG/g hydrolysat) for laks, ørret og torske-hydrolysater. Resultatene er gitt som gjennomsnitt±standardavvik, n=6 for laksehydrolysatene og n=5 for ørrethydrolysatet. . . . . | 41  |
| 3.2 | RMSEP og forklart x-og y-variens for den kryssvaliderte PLSR-modellen med 3 komponenter for matrise 1. . . . .   | 44  |
| 3.3 | RMSEP, forklart x- og y-variens for den gjennomsnittlige, kryssvaliderte PLSR modellen med 3 komponenter for matrise 5. . . . .  | 55  |
| 3.4 | Modellparametere ved valgt antall PLS komponenter for de to kryssvaliderte PLSR-modellene tilhørende matrise 2 og 6. . . . .   | 60  |
| 3.5 | Modellparametere ved valgt antall PLS komponenter for de to kryssvaliderte PLSR-modellene tilhørende matrise 4 og 8. . . . .   | 65  |
| A.1 | Oversikt over innveid laksehydrolysat for ABTS-målinger. . . . .   | x   |
| A.2 | Rådata for standardkurven i form av målte absorbansverdier for kjente konsentrasjoner av PG. . . . .   | x   |
| A.3 | Målte absorbansverdier ved 734 nm for kontrollprøvene av laks (0-6 måneder lagring). . . . .   | xi  |
| A.4 | Resultatene av beregningene som ble gjort for alle parallellene av de ulike lakseprøvene . . . . .   | xii |

|      |   |      |
|------|---|------|
| A.5  | Oversikt over gjennomsnittsverdiene (umol PG/g hydrolysat) av parallellene for de ulike kontrollprøvene av laksehydrolysat (0-6 måneder lagring). Standardavvikene ble beregnet med Excel-funksjonen STDAV.S. . . . . . | xii  |
| A.6  | Rådata for standardkurven i form av målte absorbansverdier for kjente konsentrasjoner av PG. . .  | xiv  |
| A.7  | Målte absorbansverdier ved 734 nm for 6 måneders lagret kontrollprøve av ørret. . . . .   | xv   |
| A.8  | Resultatene av beregningene som ble gjort for alle parallellene av 6 måneders lagret kontrollprøve av ørrehydrolysat. . . . .   | xv   |
| A.9  | Gjennomsnittsverdien av umol PG/g hydrolysat for 6 måneders lagret kontrollprøve av ørrehydrolysat, inkludert standardavvik. . . . .  | xv   |
| A.10 | Oversikt over umol PG/g hydrolysat for seihydrolysatene BOI, BOII, BJI og BJII. . . . .   | xvii |
| A.11 | Oversikt over umol PG/g protein (fra Marte Vilde Monslaup) og resultatene av omberegningene til umol PG/g hydrolysat for torskehydrolysatene T1, T5 og T9. . . . .  | xx   |

# Forkortelser

FPC = Fiskeproteinkonsentrater

FPH = Fiskeproteinhydrolysater

ACE = Angiotensin-konverterende enzymer

PRY = Pro-Arg-Tyr

WG = Trp-Gly

Gly = Glycin

Ala = Alanin

Pro = Prolin

Val = Valin

Leu = Leucin

Ile = Isoleucin

Met = Metionin

Ser = Serin

Thr = Treonin

Cys = Cystein

Asn = Asparagin

Gln = Glutamin

Phe = Fenylalanin

Tyr = Tyrosin

Trp = Tryptofan

Lys = Lysin

Arg = Arginin

His = Histidin

Asp = Aspartat

Glu = Glutamat

ROS = Reaktive oksygenforbindelser

HAT = *Hydrogen atom transfer*

ET = *Electron transfer*

ORAC = *Oxygen radical absorbance capacity*

TRAP = *Total peroxyl radical-trapping antioxidant parameter*

FCR = *Folin-Ciocalteu reagent*

TEAC = *Trolox equivalence antioxidant capacity*  
FRAP = *Ferric ion reducing antioxidant power*  
DPPH = *2,2-di(4-tertoctylphenyl)-1-picrylhydrazyl*  
ABTS= *2,2'-azinobis(3-ethylbenzothiazoline-6-suslfonic acid*  
QSAR = *Quantitative structure-activity relationship*  
QSPR = *Quantitative structure-property relationship*  
MLR = *Multiple linear regression*  
PCR = *Principal components regression*  
PLSR = *Partial least squares regression*  
ANN = *Artificial neural networks*  
PCA = *Principal Components analysis*  
PC = *Principal Components*  
PLS = *Partial least squares*  
MSE = *Mean squared error*  
LOOCV = *Leave-One-Out Cross-Validation*  
RMSEP = *Root mean squared error of prediction*  
VIP = *Variable importance for the projection*  
MW = *Molecular weight*  
RT = *Romtemperatur*  
FAA = *Sammensetning av frie aminosyrer*  
TAA = *Total aminosyresammensetning*  
HPLC = *High performance liquid chromatography*



# Kapittel 1

## Introduksjon

### 1.1 Restråstoff fra marine ressurser

Fiskeri-og havbruksnæringen i Norge genererer mange millioner tonn råstoff fra fisk og skalldyr årlig. I 2019 tilsvarte dette 3.55 millioner tonn levende fisk og skalldyr.[1] På verdensbasis har konsumet av sjømat og dermed også den globale etterspørselen økt de siste 60 årene. Globalt konsum av fisk som sjømat har økt dobbelt så mye som den globale befolkningsveksten i perioden 1961-2017 med en årlig, gjennomsnittlig økning på 3.1% mens befolkningsveksten har vært 1.6% (FAO rapport: “The State of World Fisheries and Aquaculture 2020. Sustainability in action”). Konsumet av fisk har også økt mer enn for noen andre proteinkilder fra dyr i samme tidsperiode. [2]

Denne utviklingen legger press på havressursene, hvor blant annet overfiske allerede er en global utfordring. Samme rapport viser til en økning i globalt overfiske fra 10% til 34.2% tidsperioden 1974-2017. På 1970-1980-tallet var det et vesentlig fiskepress i det nordøstlige Atlanterhavet, som i etterkant har blitt regulert for å bevare bestandene. [2] Senest i 2019 ble den totale torskekvote i Norge redusert på grunn av et overfiske på 10 500 tonn mer enn tildelt kvote for åpen gruppe i kystflåten [3]. Løsningen for å øke utbyttet fra havet er derfor ikke å øke fangsten. Overfiske har en negativ effekt på både fiskebestander, biodiversitet og økosystemer [2]. For å øke verdiskapningen og utbyttet fra fiskerinæringen må andre tiltak iverksettes, som å utnytte hele fangsten gjennom alle trinn i verdikjeden på en bærekraftig måte. Slik tankegang vil også være viktig for havbruksnæringen for å sikre god utnyttelse og bærekraftige løsninger. Flere av FNs bærekraftsmål peker på bærekraftig ressursutnyttelse og matproduksjon, både direkte og indirekte. Dette er relevant for både mål nummer 2 (utrydde sult), mål nummer 14 (livet i havet) og mål nummer 12 (ansvarlig forbruk og produksjon). [4]

FNs bærekraftsmål er globale mål som skal sikre en bærekraftig fremtid. Det er 17 mål med en tidsramme på 15 år, fra 2015 til 2030. Bærekraftsmål nummer 12 handler om ansvarlig forbruk og produksjon, med delmål som

omhandler reduksjon av både matsvinn og svinn i innhøsting- og produksjonskjeder. Bedre ressursutnyttelse er også et delmål, hvor det også er et ønske om å redusere avfallsmengden. [4] Regjeringen kom med en egen strategi for bioøkonomi i 2016 kalt “Kjente ressurser-uante muligheter”. Bioøkonomi omhandler en lønnsom og bærekraftig produksjon, uttak og utnyttelse av fornybare biologiske ressurser. Bakgrunnen for denne strategien er todelt. Norge har stor tilgang på fornybare ressurser, i tillegg til kompetansemiljøer som kan utnytte dette potensialet. Målene for strategien er i likhet med bærekraftsmålene tett knyttet opp mot mer effektiv og bærekraftig ressursutnyttelse og økt verdiskaping. Noen av fokusområdene omhandler bedre utnyttelse av restråstoff og utvikling av mer lønnsomme produkter. [5] Utnyttelse av restråstoff fra fisk er et gjennomgående tema i denne oppgaven, og bruk og utvikling av produkter fra restråstoff i dagens fiskeri- og havbruksnæring er nærmere beskrevet i delkapittel 1.1.1. Tilgang og bruk av restråstoff fra fiske og havbruk.

Fiskeri- og havbruksnæringens forskningsfinansiering (FHF) sin definisjon på marint restråstoff er “Som restråstoff defineres det som ikke er primære hovedprodukt ved anvendelse av et råstoff. Primære råstoffer er fisk og skalldyr (krepsdyr og bløtdyr) som oppdrettes og fanges fra norske kvoter i norske farvann og/eller landes i Norge” [6]. Restråstoff kan inndeles i flere kategorier avhengig av håndtering. Restråstoff som er behandlet etter hygieneregelverket kan brukes til humant konsum eller i fôr. Om det derimot er håndtert etter biproduktregelverket så er restråstoffet definert som et biprodukt, og kan ikke brukes til humant konsum. Det fins to kategorier av biprodukter som defineres basert på håndtering men også opprinnelse. Kategoriene er som følger [6]:

- Kategori II materiale: Dette kan være syk eller død fisk fra ventemerd eller produksjonsmerd. Det inkluderer også diverse avfall fra gulv og avløp i fiskemottak og slakteri. Bruk av denne typen biprodukt er begrenset til produksjon av bioenergi, gjødsel og i fôrproduksjon til pelsdyr.
- Kategori III materiale: Dette er hovedsakelig rester etter fisk som er slaktet for humant konsum, og behandlet deretter. Denne typen materiale kan kun benyttes i fôr til produksjonsdyr.

I 2019 ble 84% av restråstoffet fra sjømatnæringen utnyttet. Det ble generert totalt 965 000 tonn restråstoff fra 3.55 millioner tonn råstoff (SINTEF Ocean og Kontali analyse rapport: Analyse marint restråstoff 2019). [1] Fra havbruksnæringen er det kun blod som ikke utnyttes, men for hvitfisksektoren er det en vesentlig høyere andel som ikke utnyttes i form av for eksempel hoder. De siste årene har det likevel vært en positiv økning i benyttet restråstoff fra hvitfisk, ettersom utnyttelsesgraden har økt med 11% fra 2017 til 2019.[1] Samtidig eksporteres mye fisk som hel, ubearbeidet fisk. Tall fra 2018 viser at dette er tilfellet for både makrell (95%), laks (84%) og torsk (42%). [7] Flere kilder refererer til ubearbeidet eller hel fisk som sløyd fisk [8], [9]. Den lave bearbeidingsgraden medfører at restråstoffet og dermed også verdier går tapt for norsk industri. Norges sjømatråd har i den sammenheng uttalt at det er et betydelig verdiskapingspotensiale i form av arbeidsplasser, bevaring av restråstoff og høyere verdi på produkter. [7]

### 1.1.1 Tilgang og bruk av restråstoff fra fiske og havbruk

Marint restråstoff oppstår både i fiskeri- og havbruksnæringen. Det er mange typer restråstoff som oppstår under sløying og videreforedling av hvitfisk. Noen eksempler på kjente typer hvitfisk er torsk, sei, blåkveite, hyse, uer, lange og steinbit. Fra hvitfisk er det hoder som utgjør den største andelen tilgjengelig restråstoff. Andre tilgjengelige restråstoff er lever, tunger, rogn, melke, avskjær, skinn, rygger, mager, bein, svømmeblære og tarmer. I pelagisk sektor oppstår det restråstoff hovedsakelig fra sild og makrell. En stor andel sild fileteres, og det er derfor her mesteparten av restråstoffet i denne sektoren oppstår. For de andre pelagiske artene så utnyttes enten 100% av råstoffet til fiskemel/oljer eller så selges fisken hovedsakelig som rundfrossen. Det tilgjengelige restråstoffet fra havbrukssektoren oppstår hovedsakelig på slakteriene, men også under videreforedling. Dette er restråstoff som blant annet avskjær, slo, hoder, blod, spol, rygger, skinn og buklist fra laks og ørret. Ved matfiskanleggene oppstår restråstoff i form av dødfisk. I hvitfisksektoren er det i hovedsak hoder og slo som ikke blir utnyttet. Som nevnt tidligere er det kun blod som ikke utnyttes fra havbrukssektoren. Den samlede utnyttelsesgraden av restråstoff har de siste årene likevel økt. [1] Utviklingen går mot økt bevaring og utnytting av restråstoff, samtidig som det fortsatt er et potensiale for økt ressursutnytting.

Restråstoffet går gjennom flere prosesser før det ender opp som produkter innenfor ulike markeder. En av disse prosessene er separering av proteiner og oljer. En stor andel av restråstoffet går blant annet til produksjon av fiskemel. Under denne prosessen separeres ulike fraksjoner i form av fiskemel (tørstoff), olje og vann. Ensilering er en annen metode som muliggjør separasjon av oljer og proteiner. Mesteparten av restråstoffet prosesseres gjennom ensilering, det vil si ved tilsetning av maursyre og vann. [1] Endogene enzymer bryter så ned proteiner i restråstoffet og muliggjør utskilling av fiskeproteinkonsentrater, oljer og grakse [1], [10]. Fiskeproteinkonsentrater (FPC) tilsvarer den oppkonsentrerte proteinandelen. Hydrolyse er en annen prosess som muliggjør separasjon av proteiner og fett. [1] Biokjemisk hydrolyse er en hydrolyseprosess som tar i bruk enzymer for å hydrolysere peptidbindinger i proteiner og slik danne fiskeproteinhydrolysater (FPH) [11]. I denne masteroppgaven har fiskeproteinhydrolysater fra ulike fisketyper blitt studert, og derfor er enzymatisk hydrolyse nærmere beskrevet i delkapittel 1.1.4.

Ingredienser og produkter fra marint restråstoff anvendes til humant konsum, fôr og energi. Dette inkluderer blant annet produktkategorier som diverse sjømatprodukter og ekstrakter til humant konsum, i tillegg til andre produkter som fiskeolje, fiskemel, FPC, FPH og pelsdyrfôr. [1] Konsummarkedet består av produkter som frosne eller ferske torsketunger, tørkede hoder og rogn. [6] Andre mer bearbejdede konsumprodukter er tran, proteinekstrakter og oljer [1]. Kaviar er et annet kommersielt konsumprodukt som baserer seg på rogn fra torsk og skrei [12], [13]. Konsummarkedet inkluderer også bransjer som kosmetikk, farmasi og kosttilskudd. I disse bransjene er det lite restråstoff som utnyttes sammenlignet med de andre konsumproduktene, og dette er også bransjene med størst

betalingsvilje. Det har likevel skjedd en utvikling på dette området. [1] I Norge er det blant annet bedrifter og fagmiljøer som utvikler omega-3-produkter og kuldetilpassede enzymer fra marint restråstoff [5]. En del forskning vies også til å studere marine proteiner i forbindelse med human helse [1], [5].

Marine proteiner har også blitt et viktig fokus i andre markeder, spesielt i fôrmarkedet. Dette er det største markedet for marint restråstoff. Markedet omfatter både fiskefôr, pelsdyrfôr, husdyrfôr og fôr til kjæledyr, hvor fiskefôr opptar den største markedsandelen på 55%. Det er stigende interesse for bruk av marine proteiner og oljer som viktige ingredienser i fiskefôr. En viktig komponent i laksefôr er blant annet FPC fra hvitfisk og pelagisk restråstoff. Det er også økt interesse for hydrolyserte proteiner i både forbrukermarkedet og fôrmarkedet.[1] Dette viser at det har skjedd en utvikling mot mer forskning og bruk av marine proteiner i ulike markeder. Hydrolyserte proteiner fra torskehoder har blitt studert som et alternativ til tørkede hoder. Tidligere har torskehoder blitt tørket og eksportert, men her har lønnsomheten gått vesentlig ned på grunn av en ressurskrevende prosess og ustabil etterspørsel. SINTEF Ocean ledet forskningsprosjektet HEADS UP I i 2016-2017 for å se på nye muligheter for utnyttelse av torskehoder til humant konsum. Prosjektet viste at enzymatisk hydrolyse egner seg for utvinning av proteiner av høyere kvalitet enn fiskemel. [14] HEADS UP II er en videreføring av HEADS UP I, med samme forskningstema og flere forsøk i industriell skala. [15] Prosjektene viser til et potensiale tilknyttet proteiner fra marint restråstoff som høyverdi-produkter.

Selv om mye av restråstoffet fra fisk og skalldyr blir utnyttet i dag så er det fortsatt et verdiskapningspotensiale. I regjeringens bioøkonomistrategi defineres et av satsningsområdene som effektiv og lønnsom utnyttelse av fornybare ressurser, hvor et av delmålene er: “Økt foredling mot produkter med høy avkastning”. Det andre delmålet er: “Økt ressursutnyttelse og -gjenvinning”. [5] Disse delmålene er relevante for fiskeri- og havbruksnæringen i utvikling mot enda mer ressursutnyttelse, men også flere høyverdi-produkter. De neste delkapitlene vil omhandle fisk som kilde til slike verdifulle produkter, og et videre fokus vil være marine peptider ettersom det er fiskeproteinhydrolysater som har blitt studert i denne masteroppgaven.

### **1.1.2 Fisk som kilde til høyverdi komponenter**

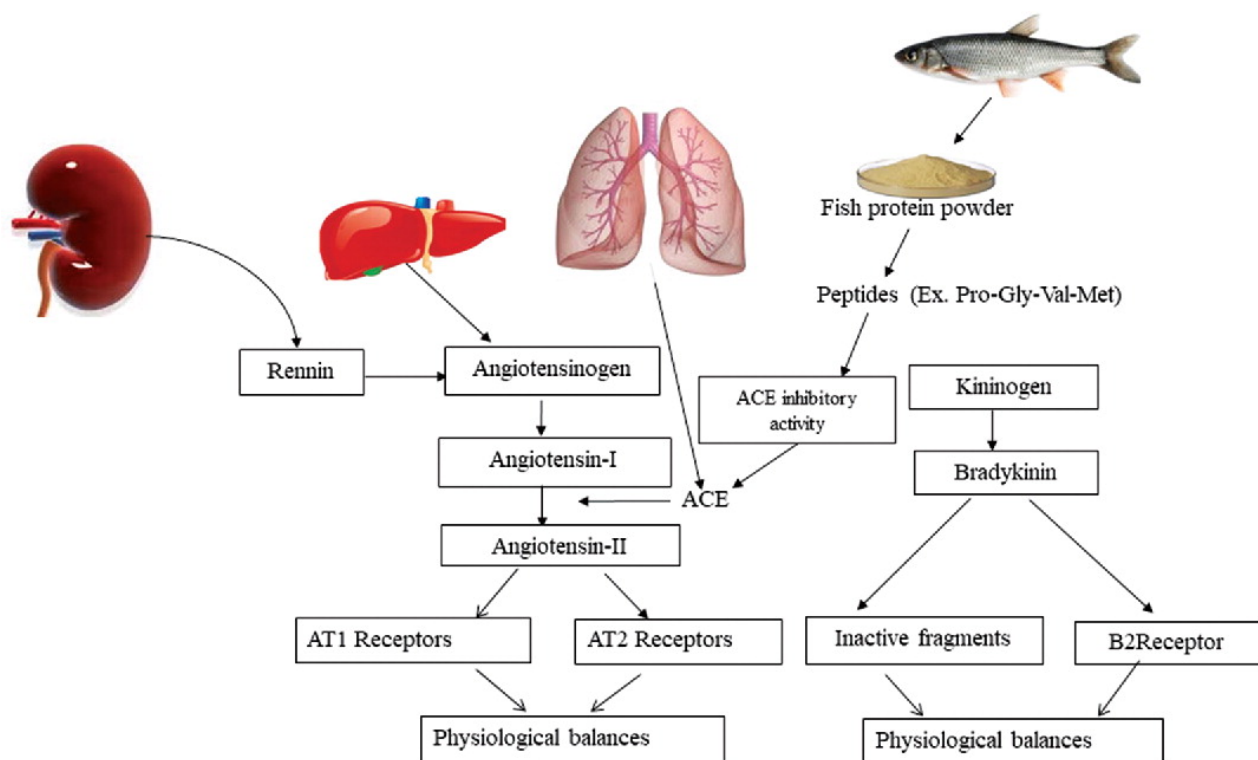
Fisk er i dag spesielt kjent som en god kilde til omega-3-fettsyrer (EPA, DPH og DHA), men det er også en god kilde til proteiner, vitaminer, jod og selen [16]. Fet fisk som makrell, laks, sild og ørret er gode kilder til omega-3-fettsyrer [17]. Tran er et av flere veletablerte produkter som inneholder marine fettsyrer, og utvinnes fra torskelever. Fiskeproteiner er en spesielt god kilde til essensielle aminosyrer, som er aminosyrer som må tilføres gjennom dietten ettersom mennesker ikke kan syntetisere disse selv [16], [18] Dette er grunnene til at fisk allerede er en viktig del av kostholdet.

Restråstoff fra fisk er en potensiell kilde til flere verdifulle produkter. Fiskeenzymer er av interesse for enzymindustrien, hvor eksempelvis fisk fra kaldere områder har enzymer som fungerer ved lave temperaturer. Dette kan være en fordel om prosessbetingelser helst skal skje ved lave temperaturer og fordi det er lettere å inaktivere enzymaktiviteten med varmebehandling. Det kan også være en ulempe ettersom enzymene er mer varmesensitive. Det er også vist at peptider fra fiskeproteiner har viktige bioaktive egenskaper, slik som: antihypertensive, anti-koagulerende, antimikrobielle og antioksidative egenskaper. [19] Litteraturstudien fra prosjektoppgaven som ble gjennomført våren 2020 gir en oversikt over flere studier som har vist at peptider fra proteinhydrolysater kan ha bioaktive egenskaper i form av ACE-inhibering eller antioksidantkapasitet [20]. Proteinkvalitet og tilhørende egenskaper vil også avhenge av ulike former for prosessering og råstoff, selv om proteininnholdet i fisk er relativt stabilt sammenlignet med både fett-og vanninnhold [16], [19]. Peptider som bioaktive komponenter er nærmere beskrevet i neste delkapittel. Denne masteroppgaven går ut på å studere marine proteinhydrolysater i sammenheng med målt antioksidantkapasitet, og se etter potensielle relasjoner mellom molekylstruktur og antioksidantkapasitet.

### **1.1.3 Peptider som bioaktive komponenter**

I regjeringens bioøkonomistrategi nevnes bioaktive forbindelser som en av flere produktkategorier som kan bidra til utvikling av ny industri [5]. Litteraturstudien fra prosjektoppgaven viste til flere studier som har blitt gjort de siste årene på bioaktive peptider, hvor flere av studiene var gjort på marine arter som hundhai, makrell, blekksprut, torsk, laks og tunfisk. De fleste funnene var enten direkte fra hydrolysater eller fraksjoner av hydrolysater. [20]

Bioaktive peptider kan grupperes etter type bioaktivitet, og en av disse er antihypertensiv aktivitet. Høyt blodtrykk (hypertensjon) kan gi kardiovaskulær sykdom, som defineres som sykdom relatert til hjerte og blodårer. Dette kan være hjerteinfarkt, blodpropp eller slag [19], [21]. Kardiovaskulær sykdom er et globalt helseproblem og er hovedårsaken til de fleste dødsfall i verden [22]. Antihypertensive peptider kan virke ved å inhibere angiotensin-konverterende enzymer (ACE), dette er såkalte ACE-inhiberende peptider. Det er funnet ACE-hemmende peptider i mange typer fisk, blant annet i tunfisk, sardiner og laks. [19] ACE har en viktig rolle i renin-angiotensin systemet (RAS), som er et system som regulerer blodtrykk. Systemet aktiveres av renin som kløyver angiotensinogen til angiotensin I. ACE katalyserer omgjøringen av decapeptidet angiotensin I til oktapeptidet angiotensin II. Sistnevnte kan bidra til sammentrekning av blodårer gjennom interaksjoner med reseptorer, som igjen bidrar til økt blodtrykk. [23] ACE kan også degradere og dermed inaktivere bradykinin, et molekyl som bidrar til utvidelse av blodårer [23], [24]. ACE sin rolle i renin-angiotensin systemet er vist mer detaljert i figur 1.1 [25].



Figur 1.1: ACE sin rolle i Renin-Angiotensin systemet [25]. Denne figuren ble også vist i prosjektoppgaven våren 2020 [20].

ACE-hemmende peptider kan vekselvirke med ACE gjennom ulike typer inhibering. Kompetitiv inhibering viser til at peptidet binder enzymets aktive sete, i motsetning til ikke-kompetitiv binding som viser til at peptidet kan binde enzymet uavhengig av om substratet opptar det aktive setet. Kinetikkstudier har for eksempel vist at det syntetiserte peptidet PRY (Pro-Arg-Tyr) er en kompetitiv ACE inhibitor, i motsetning til peptidet WG (Trp-Gly) som er en ikke-kompetitiv inhibitor. [26] Legemiddelet Captopril er et eksempel på en kompetitiv ACE-inhibitor [27]. I prosjektoppgaven viste flere av studiene i litteraturstudien til ulike typer interaksjoner mellom peptider og ACE: hydrogenbindinger, elektrostatiske, van der Waalske, polare og hydrofobe interaksjoner [20].

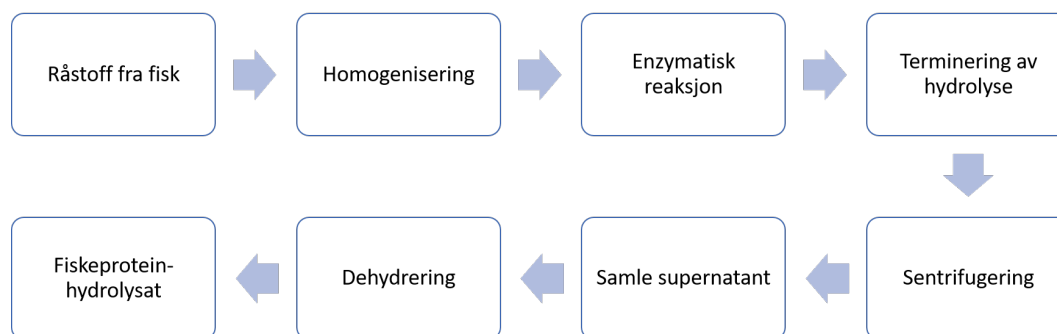
Antioksidantkapasitet er en annen type bioaktivitet, og det er også denne typen bioaktivitet som er studert i denne masteroppgaven i sammenheng med molekylstruktur i proteinhydrolysater. Tap av matkvalitet kan skje i form av oksidative reaksjoner i proteiner og lipider. Oksidasjon fører til vesentlige endringer i smak og aroma, men også i tekstur og farge. [28] Autooksidasjon av lipider i mat er en kjent prosess som fører til utvikling av harsk smak. Antioksidanter brukes generelt til å forsinke denne prosessen, og slik forlenge matens holdbarhet. [17] Det oksidative stresset øker ved prosessering og behandling av mat, gjennom prosesser som blant annet introduserer prooksidanter eller oksygen. En viktig bidragsyter til oksidative reaksjoner er frie radikaler som både kan reagere med lipider og proteiner. Proteiner har derfor et potensiale som antioksidanter om disse reaksjonene bidrar til inhibering av lipidoksidasjon. [28] Flere studier har vist at syntetiske antioksidanter kan ha negative helseeffekter,

og det er blant annet vist indikasjoner på karsinogenitet. Det er derfor en økt etterspørsel etter naturlige antioksidanter som mer helsefremmende alternativer. [17] Antioksidative peptider er nærmere beskrevet i kapittel 1.3 Antioksidative peptider, med fokus på antioksidative mekanismer og tilhørende måleanalyser.

#### 1.1.4 Enzymatisk hydrolyse av proteiner

Antioksidative mekanismer avhenger av aminosyresammensetning, og kan derfor også begrenses av proteiners tertiærstruktur. Tertiærstruktur kan gjøre flere aminosyrer utilgjengelige, og enzymatisk hydrolyse er en metode som kan brukes for å bryte opp strukturen og slik eksponere flere aminosyrer for reaksjon. [28] De fleste studiene i litteraturstudien fra prosjektoppgaven hadde brukt enzymatisk hydrolyse av råstoffet før videre analyse av bioaktivitet i peptidblandinger og peptidfraksjoner [20].

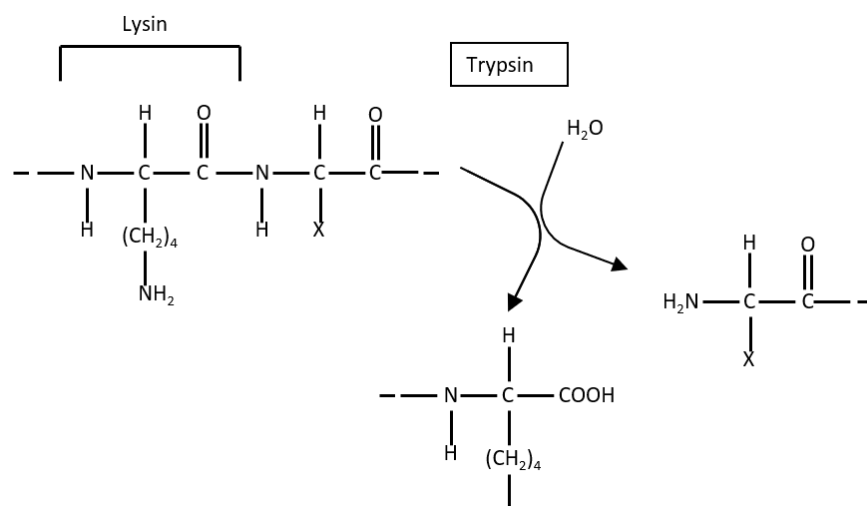
Et proteinhydrolysat kan defineres som “proteiner som enten kjemisk eller enzymatisk brytes ned til peptider av ulike størrelser” [11]. Kjemisk hydrolyse innebærer som regel bruk av enten base eller syre. Selv om dette er en rimelig prosess, så er det også en prosess som innebærer prosessbetingelser som i stor grad kan påvirke produktet negativt. Bruk av betingelser som ekstrem pH og temperatur kombinert med løsemidler og kjemikalier kan i stor grad føre til tap av næringsverdien i råstoffet. Dette kan være tap av aminosyrer og/eller dannelse av nye, giftige produkter gjennom  $\beta$ -elimineringssreaksjoner, eller rasemisering av L-aminosyrer. Kjemiske hydrolyseprosesser gir også produkter med stor variasjon i sammensetning og egenskaper, og er derfor vanskelig å kontrollere. En annen hydrolysemetode er enzymatisk hydrolyse, som er en prosess som innebærer bruk av enzymer. Formålet med denne prosessen er å endre proteinenes egenskaper og samtidig bevare næringsverdien. Figur 1.2 viser en generell oversikt over trinnene i en enzymatisk hydrolyseprosess. Resultatet er et fiskeproteinhydrolysat med ulike mengder frie aminosyrer, dipeptider, tripeptider og oligopeptider.[11]



Figur 1.2: Oversikt over den enzymatiske hydrolyseprosessen for å danne fiskeproteinhydrolysater. Figuren er en modifisert utgave av figuren fra Kristinsson et al (2000). [11]

I den enzymatiske reaksjonen i hydrolyseprosessen (figur 1.2) inngår bruk av enzymer. De fleste enzymer består av proteiner, og kjennetegnes av en unik primær, sekundær og tertiærstruktur. Dette betyr at både aminosyresekvensen og den tredimensjonale foldingen vil ha betydning for enzymets funksjon. Enzymets tredimensjonale struktur har et aktivt sete som kan binde spesifikke substrater basert på de funksjonelle gruppene i dette setet (aminosyrer, koenzymer og metallioner). Flere svake interaksjoner mellom substratet og enzymet vil da kunne oppstå, og danne et enzym-substrat kompleks. [17], [18]

Proteolytiske enzymer hydrolyserer peptidbindinger, og brukes mye til enzymatisk hydrolyse i matindustrien. Proteaser kan generelt grupperes etter mekanisme, hvor eksoproteinaser er proteaser som kløyver aminosyrer fra N eller C-terminal og endoproteinaser kløyver bindinger inne i selve proteinet. Flere proteaser kløyver kun bindinger ved siden av spesifikke aminosyrer og er derfor relativt substratspesifikke. Eksempelvis så vil trypsin gjenkjenne Lys og Arg og bare kløyve bindingen mellom disse to. [11], [18] Dette er vist i figur 1.3 hvor trypsin har kløyvet peptidbindingen ved aminosyreresten Lys og resultatet er to peptider [17].



Figur 1.3: Trypsin som hydrolyserer peptidbindingen ved aminosyreresten Lys. Figuren er en modifisert utgave av figuren fra boken Food: The Chemistry of its Components. [17]

Enzymatisk hydrolyse kan enten gjøres ved å utnytte de endogene enzymene som er i fiskeråstoffet, eller ved å tilsette enzymer. Autolyse er en type enzymatisk hydrolyse som tar i bruk endogene enzymer som allerede fins i råstoffet. De endogene enzymene er en blanding av flere enzymer, og kan eksempelvis være enzymer som trypsin, chymotrypsin eller pepsin som er typiske enzymer fra fiskeinnvoller og fordøyelsessystem. Innholdet av endogene enzymer varierer betraktelig mellom ulike råstoff, uavhengig av om det kommer fra samme eller ulike typer fisk. Noen av faktorene som påvirker denne sammensetningen er kjønn, alder og sesong. Prosessen er derfor vanskeligere å kontrollere sammenlignet med enzymatisk hydrolyse der enzymer tilsettes. Ulike enzymer har også ulik optimal pH og temperatur, noe som vil påvirke enzymkinetikken. Ved tilsetning av enzymer vil produsenten kunne



ha bedre kontroll og kan velge prosessbetingelser basert på substratspesifisitet og reaksjonskinetikk. [11] Slik kan prosessen tilpasses målet med hydrolyseprosessen, om det er å produsere spesifikke peptider eller å øke proteinutbyttet. Substratet vil også ha betydning for hydrolyseprosessen, hvor både enzym-substrat-forhold og type substrat er viktige faktorer som påvirker kjemisk struktur i sluttresultatet [11]. I litteraturstudien fra prosjektoppgaven brukte flere av studiene ulike enzymatiske hydrolysebetingelser, slik som type enzym, tid, enzym-substrat forhold og temperatur. Noen av enzymene som ble brukt var Alkalase, Protamex, esperase, pepsin, trypsin og chymotrypsin. [20] Hydrolysatene som er studert i denne masteroppgaven er basert på enzymatisk hydrolyse med enzymene papain og bromelain.

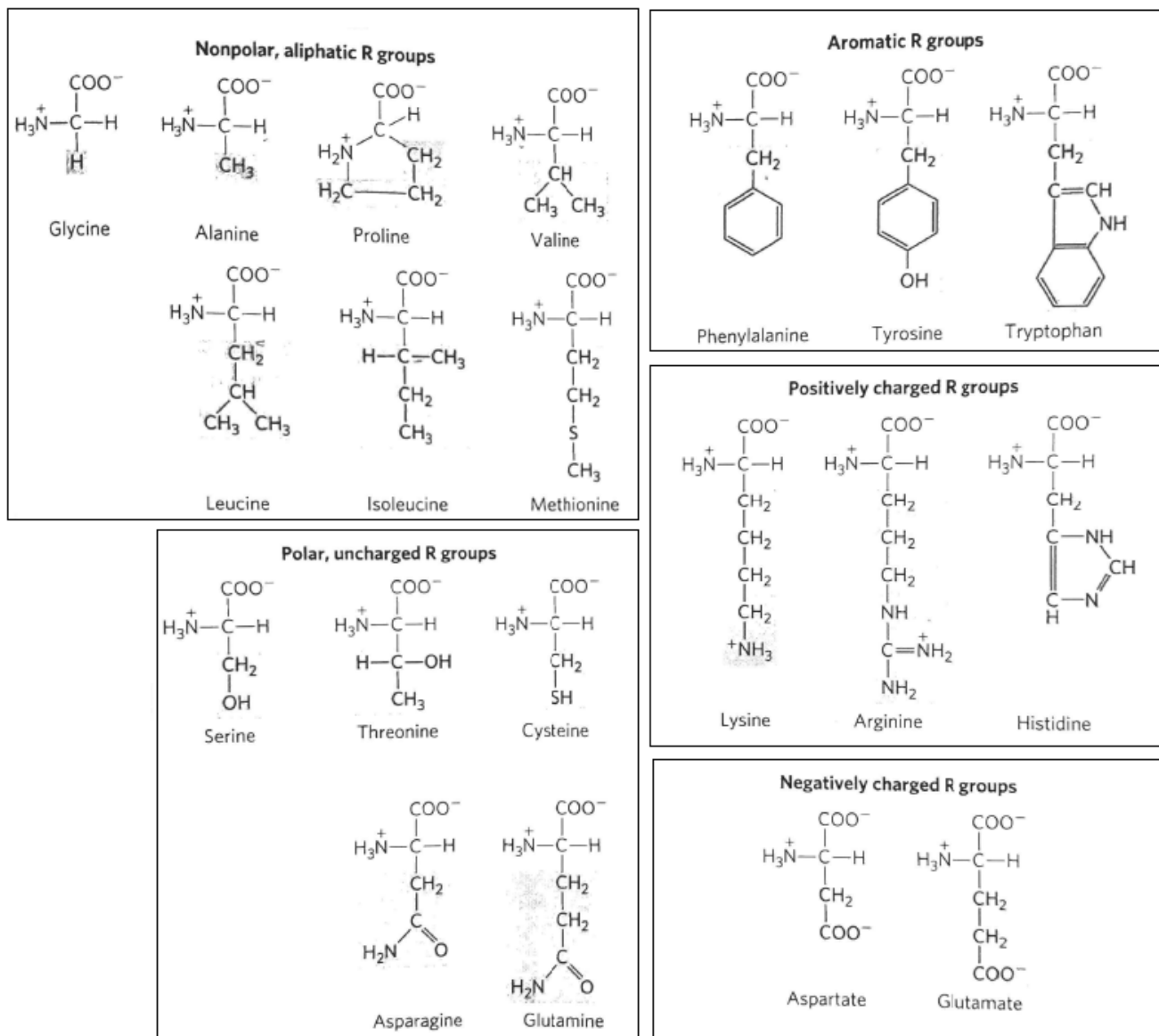
## 1.2 Proteinstruktur og bioaktivitet

Målet med denne masteroppgaven er å studere struktur-bioaktivitet sammenhenger for peptider, og derfor vil neste delkapittel definere proteinstruktur og beskrive hvordan aminosyrer og peptidkjeder kan interagere og foldes.

### 1.2.1 Proteinstruktur

Aminosyrer er bundet med kovalente bindinger til en polypeptidkjede, hvor proteinets primærstruktur er rekkefølgen av disse aminosyrene. Oppbygningen av de ulike aminosyrene er like, med et karbonatom som sentrum koblet til en karboksylgruppe, en aminogruppe, et hydrogenatom og en R-sidegruppe. [18] I en peptidkjede er det en kovalent binding mellom aminogruppen i en aminosyre og karboksylgruppen til en annen aminosyre. Denne bindingen bidrar til tap av et vannmolekyl. [17]

Det er 20 ulike aminosyrer, med ulike R-sidegrupper som varierer i struktur. De ulike aminosyrene kan inndeles i grupper basert på R-gruppens struktur. Figur 1.4 viser en oversikt over de ulike aminosyrene med inndeling basert på kjemiske karaktertrekk ved R-sidegruppene (ved pH 7): aromatiske, positivt ladede, negativt ladede, polare uladede og upolare alifatiske R-sidegrupper. Det som ikke kommer frem fra figuren er at de ladede sidegruppene også er hydrofile. De aromatiske aminosyrene derimot er relativt upolare, men både tryptofan og tyrosin har kjemiske grupper som bidrar til polare egenskaper også (OH og N). Histidin regnes her som en positivt ladet aminosyre, men kan også være uladet ved nøytral pH på grunn av en  $pK_a$  på samme nivå. Som vist i figur 1.4 har også histidin en aromatisk sidegruppe i form av en imidazol-gruppe. [18] Flere av aminosyrene har derfor flere egenskaper som gjør at de kan passe inn under flere kategorier. De kjemiske egenskapene vil senere i oppgaven brukes som deskriptorer for aminosyresammensetning i form av totale og frie aminosyrer, og brukes i dataanalysen. Videre i oppgaven vil trebokstavs-forkortelsene for aminosyrene brukes.



Figur 1.4: Inndeling av aminosyrer etter R-sidegrupper innen kategoriene (ved pH 7): aromatiske, positivt ladete, negativt ladete, polare uladete og upolare alifatiske R-grupper. Figuren er opprinnelig fra boken Principles of Biochemistry, men rammene har blitt modifisert. [18]

Sekundærstruktur av proteiner beskriver arrangement av aminosyrer til stabile mønster, som i  $\alpha$ -helikser eller  $\beta$ -plater. Tertiærstruktur er videre tredimensjonal folding av polypeptidkjeden. Proteinkonformasjon begrenses av de kovalente bindingene mellom aminosyrene, hvor noen av bindingene kan rotere og andre ikke. [18] Aminosyresekvensen vil i stor grad styre foldingen av polypeptidkjeder [17]. Svake interaksjoner er essensielle for folding, slik som hydrofob effekt, hydrogen-bindinger, ioniske interaksjoner og van der Waalske krefter. [18]

## 1.2.2 Struktur-bioaktivitet sammenhenger

I prosjektoppgaven ble det gjort en kvalitativ vurdering av forskningsartikler for å studere sammenhenger mellom kjemisk struktur og bioaktivitet, med fokus på antioksidative og ACE-hemmende peptider. Det ble gjort en del funn som kunne relateres til konkrete struktur kategorier for peptider og aminosyrer, slik som molekylvekt, type aminosyrer, rekkefølge og sekvens. Funnene viste at molekylvekt kan være en viktig faktor i både antioksidantkapasitet og ACE-hemming. Flere av studiene viste til bedre antioksidantkapasitet i fraksjoner med en molekylvekt innenfor 0-4 kDa. Det var likevel ikke slik for alle studiene, noe som kan indikere at dette ikke er en avgjørende faktor for antioksidantkapasitet. Dette gjelder spesielt jernbinding. Funnene viste også ulike resultater for jernbinding som indikerte at spesielt denne mekanismen ikke kan relateres til en struktur kategori. [20] Metallbinding som mål på antioksidantkapasitet er nærmere beskrevet i delkapittel 1.3.2 Antioksidative mekanismer.

Litteraturstudien viste også at type aminosyre kan ha innflytelse på antioksidantkapasitet. Studiene viste eksempelvis til positiv korrelasjon mellom aromatiske eller nukleofile aminosyrer og antioksidantkapasitet, både i analyser som måler evne til å fjerne radikaler og reduserende evne. Et par funn indikerte også at ladede aminosyrer kan bedre antioksidantkapasiteten. I den sammenheng viste to studier til en vesentlig andel Glu i fraksjonen med best jernbinding. En annen studie viste til positiv korrelasjon mellom Arg og ABTS radikal fjerningsevne. Noen av studiene viste også til funn på kjente peptidsekvenser som i andre studier ble funnet i potente antioksidative fraksjoner. I litteraturstudien ble det også diskutert at peptider kan vekselvirke med hverandre, og at dette kan bidra positivt eller negativt til antioksidantkapasiteten. Ingen av studiene fra litteraturstudien hadde sett på interaksjoner mellom peptider, så det kunne hverken bekreftes eller utelukkes. [20] Antioksidative mekanismer er nærmere beskrevet i delkapittel 1.3 Antioksidative peptider. Resultatene fra dataanalysen i denne masteroppgaven vil sammenlignes med funnene som ble gjort på struktur-aktivitet sammenhenger i litteraturstudien i prosjektoppgaven. Resultatene av denne sammenligningen er beskrevet i kapittel 3 Resultater og Diskusjon.

## 1.3 Antioksidative peptider

I en oversiktsartikkel skrevet av R. J. Elias (2008) nevnes flere studier som viser til at aminosyrer kan oksideres av frie radikaler, og det foreslås også i denne artikkelen at selv om dette kan være negativt så kan det også indikere potensielle antioksidantmekanismer [28]. De siste årene har det blitt gjort mye forskning på antioksidative peptider. Et søk i litteratordatabasen Oria viser over 51 000 treff relatert til de to emneordene “antioxidant” og “peptides” kun de siste fem årene. Dette er etter at søkeresultatet har blitt avgrenset til å kun vise artikler fra fagfelle-vurderte tidsskrifter. Litteraturstudien i prosjektoppgaven gir en oversikt over flere studier som har målt antioksidantkapasitet for ulike proteinhydrolysater og peptidfraksjoner [20].

### 1.3.1 Oksidasjon av biologiske makromolekyler

Oksidasjon av biologiske makromolekyler som proteiner og lipider har stor negativ effekt på matkvalitet i form av endringer av sensoriske egenskaper som farge, aroma og smak [28]. De fleste lipider i mat består av uforgrenede fettsyrer med kun en karboksylgruppe. Mettede fettsyrer er fettsyrer uten dobbeltbindinger, i motsetning til umettede fettsyrer som kan ha en eller flere dobbeltbindinger. Autooksidasjon av umettede fettsyrer resulterer i harskning av fett, som gir en ubehagelig lukt og smak. Dette kan skje i mange ulike matvarer, blant annet i ulike oljer, fisk og kjøtt. [17] Autooksidasjon av fett kan beskrives gjennom tre faser: Initierting, propagering og terminering. Reaksjonene er gitt i følgende reaksjonsligninger, som også ble presentert i prosjektoppgaven 2020 [17], [20]:



Den første fasen kjennetegnes av dannelsen av frie radikaler  $\text{R}^\cdot$ . I denne fasen er det en initierende fri radikal som reagerer med umettede fettsyrer. Frie radikaler er molekyler med et uparet elektron, og vil derfor være relativt reaktive. Det er foreslått at singlet oksygen kan være en viktig initiator i initieringsprosessen og vil kunne reagere med dobbeltbindingene i umettede fettsyrer og danne hydroperoksider. Hydroperoksider kan videre brytes ned til reaktive alkoksy og hydroperoksy-radikaler. Disse radikalene vil kunne fjerne et hydrogenatom fra fettsyrer i mat, og starte autooksidasjonsprosessen. [17]

I propageringsfasen dannes flere ulike radikaler gjennom flere steg. Det første steget innebærer at  $\text{R}^\cdot$  reagerer med atmosfærisk oksygen, noe som resulterer i peroksy radikaler  $\text{ROO}^\cdot$ . Disse radikalene kan så reagere med andre umettede fettsyrer, og generere frie radikaler  $\text{R}^\cdot$  og hydroperoksider  $\text{ROOH}$ . Disse to bidrar videre til produksjon av flere frie radikaler. Som vist av ligningene kan hydroperoksider brytes ned til de frie radikalene  $\text{RO}^\cdot$  og  $\text{ROO}^\cdot$ , som også kan reagere videre med umettede fettsyrer.  $\text{R}^\cdot$  kan inngå i samme reaksjon som beskrevet gjennom det første steget i propageringsfasen, og slik repeterer denne prosessen. Dette gir en kjedereaksjon og resulterer i en akkumulering av frie radikaler. Termineringsfasen viser til at mengden frie radikaler når et metningspunkt, hvor de frie radikalene begynner å reagere med andre frie radikaler. Resultatet er stabile sluttprodukter. Underveis i autooksidasjonen dannes det også en rekke sideprodukter. Alkoksyradikaler inngår i reaksjoner som gir både alkoholer, ketoner og aldehyder, hvorav sistnevnte gir den karakteristiske lukten av harsknet fett. [17]

Et samlebegrep for singlet oksygen og frie radikaler er reaktive oksygenforbindelser (ROS). Som beskrevet over har disse forbindelsene evnen til å løsrive hydrogenatomer fra lipider. De ulike radikalene har ulike standard reduksjonspotensialer, noe som vil si at de har varierende reaktivitet. Det er gitt at høyere standard reduksjonspotensiale tilsvarer god oksideringsevne. Hydroksey radikaler er veldig reaktive, og har høyere standard reduksjonspotensiale sammenlignet med andre radikaler. [29]. Hydrogenatomene som er mest utsatte for frie radikaler er de som er tilknyttet CH<sub>2</sub>-gruppene ved siden av en eller mellom to dobbeltbindinger i umettede fettsyrer [17].

ROS kan også reagere med proteiner og aminosyrer. Det er spesielt R-sidegruppene til aminosyrer som angir hvor utsatt den er for frie radikaler. Dette er fordi de er strukturelt forskjellige, og derfor har ulik reaktivitet. De mest utsatte aminosyrene er gjerne aminosyrer hvor hydrogenatomer er lett tilgjengelige. Dette avhenger også av strukturen til nærliggende aminosyrer. En annen faktor er reaktiviteten til ROS, som ble beskrevet ovenfor. Hydroksey radikaler vil kunne reagere med alle aminosyrerester ettersom det er en svært reaktiv radikal. Andre radikaler er mindre reaktive og vil være mer selektive i reaksjon med aminosyrer. [28] Dette medfører også en lavere reaksjonsrate, eksempelvis med radikaler som superoksid anion og hydroperoksid radikaler [29] Evnen til å reagere med frie radikaler kan være en antioksidativ mekanisme hvis det også forhindrer oksidasjon av lipider, dette er nærmere beskrevet i neste delkapittel.

### 1.3.2 Antioksidative mekanismer

Elias et al. (2008) beskriver fem ulike antioksidantmekanismer for proteiner. Proteiner og peptider kan derfor potensielt hindre fettoksidasjon gjennom flere reaksjonsspor. [28] Dette viser til at proteiner og peptider har flere ulike funksjoner som bidrar til et potensiale som antioksidanter. Følgende mekanismer ble også beskrevet i prosjektoppgaven våren 2020, med unntak av den siste mekanismen [20], [28]:

- Å nøytralisere frie radikaler
- Å kompleksbinde overgangsmetaller
- Å fysisk separere reaktanter som er involvert i lipidoksidasjon.
- Å redusere hydroperoksider til stabile produkter gjennom redoksreaksjoner.
- Å interagere med aldehyder, som er sekundære produkter av lipid oksidasjon og slik motvirke harskning.

Antioksidanter som reagerer med frie radikaler kan hindre propageringsfasen i autooksidasjon av fett, ettersom det er en fase hvor mange frie radikaler produseres. En antioksidant kan nøytralisere frie radikaler ved å donere

hydrogenatomer til radikalene[17]. For at en såkalt *radical scavenger* skal være en potent antioksidant må den i tillegg oppfylle følgende tre krav [28]:

- Den må kunne nøytralisere frie radikaler
- Den må kunne oksideres fremfor lipider
- Proteinradikaler som dannes må ikke kunne bidra til lipid oksidasjon.

ROS kan reagere med proteiner og aminosyrer i større eller mindre grad avhengig av reaktiviteten til et radikal og hvor labil en aminosyre er for oksidasjon. For radikaler som er mindre reaktive og derfor mer selektive så vil interaksjonene i stor grad bestemmes av aminosyrenes oksidative potensial. Noen aminosyrer har et større oksidativt potensial enn andre, og det gjelder aminosyrer som har mer tilgjengelige hydrogenatomer. Dette er tilfellet for de nukleofile aminosyrene som har en SH-gruppe i sin R-sidekjede, som er tilfellet for aminosyrene Cys og Met. Aromatiske aminosyrer har også et generelt større oksidativt potensiale, dette inkluderer aminosyrer som Tyr, Trp og Phe. Den aromatiske ringen kan oksideres og slik donere ett elektron til en radikal. [28] Dette inkluderer også aminosyren His som har en imidazolring, som er en kjent nukleofil struktur.[18], [28]

En potent *radical scavenger* antioksidant må ha et større oksidativt potensiale enn lipider [28]. Den oksidative stabiliteten til aminosyrene Trp, Met og Cys ble undersøkt i en studie av  $\beta$ -laktoglobulin i en olje-i-vann emulsjon, og viste at kun Trp og Cys oksiderte før lipidene [30]. Dette indikerer at disse aminosyrene ikke bare har et større oksidativt potensiale enn lipider, men også sammenlignet med andre aminosyrer. Studien viser også til at Met-restene er utilgjengelige ettersom de befinner seg i kjernen av  $\beta$ -laktoglobulin, og at dette kan forklare den manglende oksidasjonen av Met [30]. For å kunne ha en netto antioksidativ effekt kan heller ikke proteinradikalene som dannes bidra til fettoksidasjon [28]. I noen tilfeller vil antioksidantradikalene være stabile på grunn av resonans, og derfor ikke reagere videre. Dette er tilfellet for kjemiske strukturer som aromatiske ringer. Antioksidantradikaler kan også reagere med andre radikaler og slik produsere stabile produkter.[17] Dette innebærer også at proteinradikaler kan reagere med andre proteinradikaler, og danne krysslinkede produkter [29]. Slik kan altså både mekanismen til den potente antioksidanten og resulterende antioksidantradikal være av interesse i utviklingen av nye antioksidanter.

En annen antioksidant-mekanisme er metallbinding av overgangsmetaller, blant annet jern og kobber [28]. Overgangsmetaller er viktige prooksidanter som gjennom redoksreaksjoner bidrar til produksjon av flere radikaler. Overgangsmetaller har en evne til å oksideres og/eller reduseres, og kan slik interagere med og bryte ned hydroperoksider. [17], [28] Proteiner kan potensielt hindre overgangsmetaller i å bidra til fettoksidasjon på flere måter: kompleksbinde overgangsmetaller, endre metallenes redokstilstand, sterisk forhindre interaksjon og påvirke metal-

lenes plassering i systemet. Endringer i pH kan bidra til mer eller mindre metallbinding, på grunn av elektrostatiske krefter. Proteiner med negative ladninger vil kunne tiltrekke seg metallkationer. [28] Alle aminosyrer kan på grunn av karboksylgruppen og aminogruppen ha en ionisert form med enten negative eller positive ladninger, men noen aminosyrer har også R-grupper som bidrar til dette. Asp og Glu er eksempler på aminosyrer med en R-gruppe som bidrar til en negativ ladning ved pH 7. [18] Som nevnt over kan denne typen antioksidanter også påvirke metallenes redokstilstand. Antioksidative proteiner kan potensielt binde en mindre reaktiv form av metallkationene. Det må likevel nevnes at den prooksidative evnen til overgangsmetallene i flere tilfeller kan bevares selv i ett protein-metallkompleks. [28] Dette gjør mekanismen bak denne typen velfungerende antioksidanter mer kompleks og uforutsigbar.

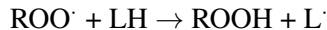
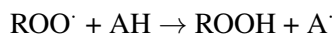
Blant andre antioksidant-mekanismer er evnen til å fysisk separere reaktanter som er involvert i lipidoksidasjon [28] I mange næringsmidler er fett emulgert, og det kan være i form av vann-i-olje emulsjoner (smør) eller olje-i-vann emulsjoner (melk, majones) [17]. Overgangsmetaller i vandig fase kan reagere med lipid hydroperoksider i mellomfasen, og denne kontakten kan en antioksidant potensielt forhindre. Dette kan være gjennom elektrostatisk repulsjon mellom positive proteiner/peptider og overgangsmetaller, eller ved å generelt danne et tykt lag mellom metallkationer og lipider. [28] Emulsjoner er eksempler på komplekse systemer hvor også plasseringen til en antioksidant påvirker antioksidant-aktiviteten. Flere studier viser at upolare antioksidanter er mer effektive ettersom de er i den lipide fasen der lipidoksidasjon skjer i størst grad, men det er også studier som tilsier at dette ikke stemmer i alle situasjoner. [31] Dette indikerer at potensialet som antioksidant også avhenger av type næringsmiddel og tilgjengelighet til en viss grad.

Den siste antioksidantmekanismen involverer reaksjoner med enten aldehyder eller hydroperoksider. Den ene mekanismen innebærer å redusere lipid hydroperoksider til produkter som ikke reagerer videre. Aminosyren Met har en tioeter gruppe som kan donere to elektroner og dermed redusere lipid hydroperoksider. Den andre mekanismen innebærer at aminosyrer kan reagere med sekundære lipidoksidasjonsprodukter. Dette er ikke en mekanisme som i seg selv forhindrer lipidoksidasjon, men den kan likevel forsinke harskningsprosessen. [28] Denne mekanismen nevnes for å vise at proteiner og peptider også kan reagere med oksidasjonsprodukter, som vil kunne påvirke matkvaliteten.

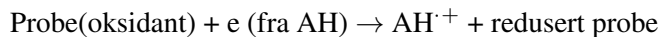
### **1.3.3 *In vitro* analyser for måling av antioksidantkapasitet**

Som beskrevet i forrige delkapittel så kan peptider virke gjennom flere ulike antioksidantmekanismer, noe som tilsier at antioksidantkapasitet er komplisert. Det er også flere ulike typer *in vitro* analyser som kan måle antioksidantkapasitet. Disse kan grovt inndeles etter type reaksjonsmekanisme [32]:

- Hydrogenatom-overføring (*hydrogen atom transfer* HAT)



- Elektron-overføring (*electron transfer* ET)



AH= antioksidant, LH=substrat og ROO $\cdot$ = peroksyldradikaler [32]. Selv om det kan gjøres en grov inndeling mellom HAT og ET-baserte analyser, så skjer de to mekanismene stort sett samtidig. Hvilken mekanisme som dominerer er i stor grad bestemt av antioksidantens struktur. [33] HAT-mekanismen viser til antioksidantens evne til å donere hydrogenatomer. Dette betyr at reaktiviteten til en antioksidant i stor grad er avhengig av dissosiasjonsenergien. [33] De fleste HAT-analysene består generelt av komponentene: substrater/prober, antioksidanter og radikaler. Antioksidantene vil konkurrere med probene om radikalene, og slik inhibere oksidasjon av probene. Denne prosessen kan beskrives gjennom en kinetisk kurve med ulike faser, som ofte inkluderer en lagfase hvor det ikke skjer noen målbare endringer av proben. Etterhvert som en potent antioksidant brukes opp, vil oksidasjon av proben skje i større grad. Generelt så vil fasene avhenge av både type antioksidant og probe, i tillegg til konsentrasjoner. Følgende analyser anses som HAT-analyser: *oxygen radical absorbance capacity assay* (ORAC) og *total peroxy radical-trapping antioxidant parameter assay* (TRAP). Felles for disse er at en azo forbindelse brukes til å generere en fluks av peroksyldradikaler gjennom termisk nedbrytning. Proben detekteres gjennom målinger av fluorescens over tid. [32]

ET-analyser måler den reduserende evnen til en potensiell antioksidant. Som vist i reaksjonsligningen over, så innebærer analysen en redoksreaksjon mellom en probe og en potensiell antioksidant. Proben vil i dette tilfellet fungere som oksidant. Slik vil målt antioksidantkapasitet reflektere evnen til å redusere oksidanter. [32] Dette betyr at reaktiviteten til en slik antioksidant i stor grad er avhengig av ioniseringspotensialet og deprotonering. [33] Følgende analyser anses som ET-analyser: *total phenols assay* (FCR), *trolox equivalence antioxidant capacity assay* (TEAC), *ferric ion reducing antioxidant power assay* (FRAP) and *2,2-di(4-tert-octylphenyl)-1-picrylhydrazyl assay* (DPPH). [32]

Tabell 1.1 gir en oversikt over disse analysene, inkludert en beskrivelse av hovedkomponenter, prinsipp og deteksjon. Det fins en rekke andre analyser, som blant annet baserer seg på inhibert oksygenopptak eller inhibering av induert lipid autooksidasjon. I tillegg er det andre analyser som måler nøytralisering av utvalgte ROS og oksidanter, eksempelvis for hydroksyl radikaler, O $_2^{\cdot-}$  og H $_2$ O $_2$ . [32] Det er altså relativt mange analyser som kan brukes til å måle antioksidantkapasitet. Mangel på en standardmetode gjør det vanskelig å kunne sammenligne resultater på antioksidantkapasitet, ettersom de ulike metodene har ulike betingelser og baserer seg på ulike anti-



oksidantmekanismer. ET-baserte analyser vil ikke kunne beskrive antioksidantmekanismer som nøytralisering av ROS og metallbinding. Mekanismen bak HAT-baserte analyser kan minne om lipid peroksidasjon, og involverer også peroksyldradikaler som er en stor pådriver i fettoksidasjon i mat. [32] Slik vil HAT-baserte analyser kunne beskrive mekanismer relatert til kjedereaksjonen med radikaler som skjer under fettoksidasjon i mat.

Tabell 1.1: Oppsummerende oversikt av en rekke *in vitro* analyser for å studere antioksidantkapasitet [32].

| Metode | Hovedkomponenter                                       | Måle evnen til å:   | Deteksjon   |
|--------|--|---|---|
| ORAC   | Fluorescein (probe),<br>Antioksidant,<br>AAPH          | Inhibere radikalindusert oksidasjon av probe. Oksidasjon av probe: tap av intensitet. | Fluorescens:<br>eks 485 nm/ em 525 nm,<br>pH 7.4, 37°C. |
| TRAP   | R-PE (probe),<br>Antioksidant,<br>AAPH                 | Se ORAC.  | Fluorescens:<br>eks 495 nm/em 575 nm.                   |
| FCR    | Folin-Ciocalteu reagent (oksidant),<br>Antioksidant    | Redusere FC-reagenten (gult) til produkt (blått).                                     | Absorbans.<br>Basisk miljø.                             |
| TEAC   | ABTS <sup>•-</sup> (oksidant),<br>Antioksidant         | Redusere ABTS <sup>•-</sup> (mørkeblått) til ABTS <sup>2-</sup> (fargeløst).          | Absorbans.<br>734 nm, 30°C.<br>Nøytralt miljø.          |
| FRAP   | Fe(III)(TPTZ) <sub>2</sub> (oksidant),<br>Antioksidant | Redusere Fe(III) til Fe(II).  | Absorbans.<br>593 nm, 37°C.<br>Surt miljø.              |
| DPPH   | DPPH (oksidant),<br>Antioksidant                       | Redusere DPPH radikal (fargetap).   | Absorbans.<br>515 nm.                                   |

Det er flere utfordringer knyttet til bruk av *in vitro* analyser for å måle antioksidantkapasitet, både i forhold til ulike antioksidantmekanismer, men også betingelser i reelle systemer. I motsetning til disse *in vitro* analysene er det som regel en betydelig større konsentrasjon substrat/lipider sammenlignet med antioksidanter i reelle systemer som mat. I tillegg så varierer betingelsene for de ulike *in vitro* metodene, blant annet i type oksidant, probe og pH. Sistnevnte er en parameter som i stor grad påvirker reduserende evne, på grunn av protonering og protodissosiasjon i miljøer med ulik pH. [32] Pérez-Jiménez et al. (2006) gjorde ulike *in vitro* målinger av antioksidantkapasitet (ORAC, ABTS, FRAP, DPPH) med løsemidler som vann, metanol og aceton i ulike ratioer, og resultatene indikerte at type løsemiddel også har betydning for ulike analyser. Spesielt for ORAC og ABTS var forskjellene mellom de ulike løsemidlene store. [34]

I tillegg så skjer de *in vitro* analysene i kontrollerte homogene miljøer. Mat er ofte heterogene blandinger, hvor ulike faktorer som metallioner og temperatur kan initiere oksidasjon. [32] Jónsdóttir et al. (2016) studerte torske-

hydrolysater og tangekstrakt sin potensielle antioksidative rolle ved å gjøre *in vitro* antioksidantanalyser i tillegg til studier av oksidasjon i torskemulsjon og kvernet torskemuskel. I likhet med resultatene av DPPH, ORAC og reduserende evne indikerte studiene av kvernet torskemuskel at tangekstrakt er en mer potent antioksidant. Dette ble ikke bekreftet i emulsjonen hvor torskehydrolysatet ga en større antioksidativ effekt. Jónsdóttir et al. (2016) konkluderte med at verdien av *in vitro* analyser avhenger av kunnskap om ulike matsystemer, og spesielt hvilke prooksidanter som er tilstede. [35] I matemulsjoner vil antioksidantkapasitet også avhenge av antioksidantens distribusjon i ulike faser som olje, vann og interfase, og påvirkes av hvilke andre matkomponenter som er tilstede [31]. Potensielle antioksidanter kan interagere med hverandre, i tillegg til resulterende aminosyreradikaler som også kan reagere med andre aminosyrer [29], [31] Metodene sier heller ikke noe om bioaktivitet, ettersom de ikke inkluderer viktige *in vivo* parametere som stabilitet, biotilgjengelighet, reaktivitet og ulike vev. [32]

## 1.4 Kjemometriske metoder i analyse av proteiner og peptider

### 1.4.1 Bruk av kjemometriske metoder i analyse

Iwaniak et al. (2015) gir en oversikt over en rekke definisjoner på kjemometri fra ulike kilder. En av disse beskriver kjemometri som en metode for å håndtere data ved å kombinere statistiske og matematiske metoder, og henviser spesielt til multivariate metoder. En mer detaljert definisjon beskriver kjemometri som en kombinasjon av matematikk og statistikk, men også logikk. I denne definisjonen er formålet til kjemometri tredelt: for å designe og/eller velge optimale eksperimentelle metoder, for å få kunnskap om kjemiske systemer og for å få relevant kjemisk informasjon fra analyse av kjemiske data. [36] Data regnes som multivariate når flere variabler er målt for samme prøve. Det er en utfordring å visualisere multivariate data grafisk. For å kunne studere mønster og sammenhenger i slike datasett må derfor andre dataanalyser benyttes. [37]

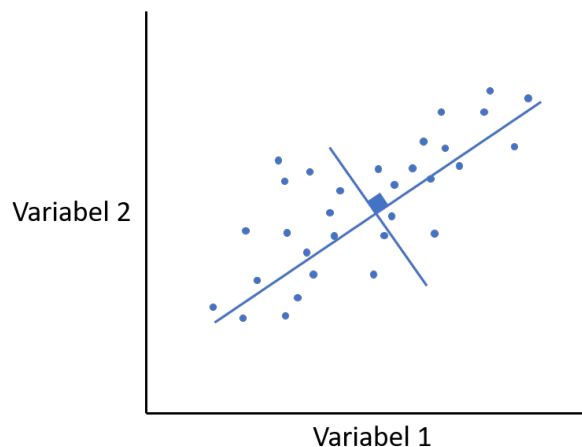
I mange dataanalyser er det ønskelig å relatere X-variabler til Y-variabler. I en kjemisk og/eller biologisk sammenheng henvises det ofte til QSAR (*quantitative structure-activity relationship*) eller QSPR (*quantitative structure-property relationship*). Dette er modeller som brukes for å relatere X-variabler som beskriver kjemisk struktur og sammensetning til Y-variabler i form av kjemiske egenskaper eller biologisk aktivitet. [38] Dette er også modeller som brukes innen forskning på peptider. Pripp et al (2007) brukte QSAR modeller til å studere ACE-hemming og bitterhet av kjente peptidsekvenser, hvor sekvensene ble beskrevet gjennom aminosyredeskriptorer. I denne studien ble predikert og observert ACE-hemming sammenlignet med predikert og observert bitterhet. [39] Det har også blitt gjort noe forskning på struktur-antioksidant sammenhenger i peptider ved bruk av QSAR modeller. Uno et al. (2020) baserte en QSAR modell på 55 tripeptider, målt antioksidantaktivitet i form av TEAC og utvalgte QSAR parametere (hydrofobisitet  $\Pi_{\alpha}$ , elektroniske og steriske parametere). Funnene i denne studien viser til ulike

struktur-aktivitet sammenhenger, blant annet at aromatiske aminosyrer på C-terminus og høyere hydrofobisitet på N-terminus kan bidra til en signifikant økning i antioksidantaktivitet. [40] Iwaniak et al. (2015) gir en oversikt over flere studier som viser til forskning på peptider ved bruk av QSAR modeller for å studere antioksidantaktivitet, ACE-hemming, antimikrobiell bioaktivitet, renin-hemming og bitterhet [36].

Statistiske regresjonsmetoder brukes til å finne den matematiske sammenhengen mellom X og Y. Dette inkluderer metoder som MLR (*multiple linear regression*), PCR (*principal components regression*), PLSR (*partial least squares regression*) og ANN (*artificial neural networks*). [41] MLR er en regresjonsmetode som fungerer godt når antall observasjoner er større enn antall prediktorer [37]. Dette betyr at metoden kun fungerer godt om det er få X-variabler [38]. En annen begrensning for denne metoden er at den ikke er tilpasset datasett med korrelerende X-variabler [37], [38]. Disse begrensningene gjør MLR uegnet for bruk i denne oppgaven hvor datasettene har flere variabler enn observasjoner. ANN gjør ingen approksimasjon av type sammenheng mellom X og Y på forhånd, og er derfor en fleksibel metode som egner seg for prediksjon. Samtidig så krever dette enda større datasett for å lage en god modell. ANN er også en *black box*-metode som gjør det vanskelig å tolke resultatene, og se eventuelle sammenhenger mellom variablene. [36], [37] Målet med dataanalysen i denne oppgaven var å i stor grad se etter sammenhenger mellom variabler i et datasett med relativt få prøver. ANN ble derfor vurdert som uegnet i denne sammenhengen. PCR og PLSR ble i utgangspunktet vurdert i denne oppgaven, ettersom begge metodene håndterer data med mange X-variabler og som tar høyde for at disse variablene korrelerer [37], [38]. PCR og PLSR er nærmere beskrevet i de neste delkapitlene.

## 1.4.2 Prinsipal komponent analyse og regresjon

*Principal Components Analysis* (PCA) er en metode som kan brukes til datavisualisering og regresjon. PCA fungerer godt på større datasett hvor flere av variablene korrelerer. PCA innebærer å generere et fåtall nye komponenter, *Principal Components* (PC) fra et originalt datasett med X-variabler. Prinsippet er dimensjonsreduksjon og at de nye PC-ene skal inneha så mye informasjon om X-variablene som mulig. Dette er informasjon i form av variabilitet mellom observasjoner i datasettet. Dette har også en geometrisk tolkning. Figur 1.5 viser en rekke observasjoner for variabel 1 og 2, men også et nytt koordinatsystem definert av de blå linjene. De blå linjene angir de to første PC-ene som er plassert i retning størst variabilitet blant observasjonene. PC1 og PC2 er ortogonale, som også innebærer at de ikke korrelerer med hverandre. PC1 angir retningen med størst variabilitet blant observasjonene. Dette gjør også PC2, men under betingelsen at den ikke korrelerer med PC1. Samme prinsipp gjelder for alle PC-ene som genereres i PCA. [42]



Figur 1.5: Plott som viser dataobservasjoner for en variabel 1 og en variabel 2. Linjene indikerer de to første PC-ene, som er plassert i retning størst varians. Det er også markert at de er ortogonale. Figuren er ikke basert på et reelt datasett, men er laget for å illustrere prinsippet bak PCA.

Alle PC-ene som dannes fra PCA er lineære kombinasjoner av de originale variablene. For den første PC-en vil følgende ligning beskrive denne kombinasjonen:

$$z_{i1} = \phi_{11}x_{i1} + \phi_{21}x_{i2} + \dots + \phi_{p1}x_{ip} \quad (1.6)$$

$z_{i1}$  og  $\phi_{p1}$  er definert som henholdsvis *scores* og *loadings* for den første prinsipale komponenten.  $x_{ip}$  henviser til de originale variablene. [42] *Scores* kan defineres som koordinatene til observasjonene i det nye koordinatsystemet definert av PC-er (figur 1.5). *Loadings* viser derimot til vektene som trengs for å definere retningen til PC-aksene i det originale koordinatsystemet. *Scores* og tilhørende *score* plott kan brukes til å studere grupperingsmønster blant objektene og observasjonenes bidrag til PC-ene. *Loadings* og tilhørende *loadings* plott kan brukes til å få oversikt over hvilke variabler som bidrar mest til komponentene, og om noen av variablene korrelerer. [43]

I en regresjonssammenheng brukes de prinsipale komponentene som prediktorvariabler i en lineær regresjonsmodell som tar i bruk minste kvadraters metode, dette er referert til som *Principal Components Regression* (PCR). Lineær regresjon brukes til prediksjon av en kvantitativ respons. Hovedprinsippet bak PCR er at et fåtall komponenter er nok til å beskrive variasjonen i det originale datasettet, og til å forklare responsvariabelen  $Y$ . [42]

### 1.4.3 Delvis minste-kvadraters regresjon

*Partial Least Squares Regression* (PLSR) er en type multivariat dataanalyse som først ble brukt i 1975. I ettertid har det blitt gjort justeringer av metoden for å kunne bruke den på mer komplekse datasett innen teknologi og vitenskap. Hovedidéen bak metoden er å modellere sammenhengen mellom prediktorvariabler  $X$  og responsvariabler  $Y$ . Dette kan være relevant i mange sammenhenger, hvor en av disse er å relatere bioaktivitet eller andre

kjemiske egenskaper til kjemisk struktur. Denne tilnærmingen er kjent som *Quantitative Structure–Activity Relationship* (QSAR) eller *Quantitative Structure–Property Relationship* (QSPR) modeller. [38] Iwaniak et al (2014) nevner PLSR som en av flere kjemometriske metoder som brukes i analyse av bioaktive peptider fra mat, og viser til eksempler på studier som bruker PLSR for å studere ACE-hemmende aktivitet og antioksidantkapasitet [36].

PLSR danner i likhet med metoden PCR nye komponenter basert på de originale prediktorvariablene X, men det er en essensiell forskjell. PLSR tar også hensyn til responsvariabelen Y. De nye komponentene er lineære kombinasjoner av de originale prediktorvariablene, men de er også relatert til responsen som undersøkes i systemet. Komponentene brukes videre i en lineær modell som tar i bruk minste kvadraters metode for å relatere X og Y. [42] Notasjonen for de nydannede komponentene er *X-score* eller  $t_a$  (a: indeks for komponentene), og de er ortogonale til hverandre [38]. Dette tilsier at de ikke korrelerer med hverandre [42].  $t_a$  er definert matematisk gjennom følgende lineære sammenheng [38]:

$$t_{ia} = \sum_k W_{ka} X_{ik} \quad (1.7)$$

Hvor  $X_{ik}$  tilsvarer de originale variablene, og  $W_{ka}$  tilsvarer tildelt vekt av disse. Følgende ligninger beskriver modellen videre [38]:

$$X_{ik} = \sum_a t_{ia} p_{ak} + e_{ik} \quad (1.8)$$

$$y_{im} = \sum_a u_{ia} c_{am} + g_{im} \quad (1.9)$$

Residualer ( $e_{ik}$  og  $g_{im}$ ) angir forskjellene mellom predikerte og observerte variabler, dette er altså data som ikke beskrives av modellen. En PLSR modell har både X og Y-residualer. Gjennom residualene kan modellen håndtere støy i form av X-variabler som ikke kan relateres til Y, i motsetning til MLR-metoden. Ligning 1.8 og 1.9 viser til at *X-score* ( $t_{ia}$ ) og *Y-score* ( $u_{ia}$ ) multiplisert med *loadings* ( $p_{ak}$ ) eller vekter ( $c_{am}$ ) skal beskrive de observerte X ( $x_{ik}$ ) og Y-verdiene ( $y_{im}$ ) slik at tilhørende residualer blir små. Ettersom de nye komponentene ( $t_a$ ) også er gode prediktorer for Y så gjelder følgende [38]:

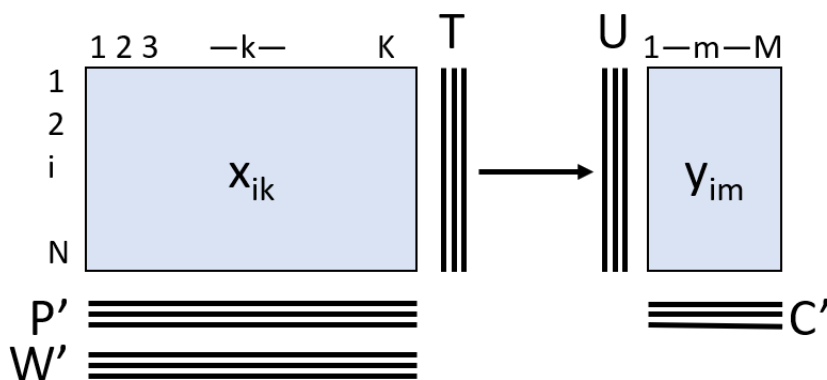
$$y_{im} = \sum_a c_{ma} t_{ia} + f_{im} \quad (1.10)$$

Hvor  $f_{im}$  tilsvarer Y-residualene. Sammenslåing av ligning 1.7 og 1.10 gir følgende uttrykk for regresjonsmodellen [38]:

$$y_{im} = \sum_a c_{ma} \sum_k w_{ka} x_{ik} + f_{im} = \sum_k b_{mk} x_{ik} + f_{im} \quad (1.11)$$

Hvor  $b_{mk}$  er definert som PLS-regresjonskoeffisienter. Oppsummert, så gir modellen *X-scores*, *Y-scores*, vekter, *loadings* og regresjonskoeffisienter. Figur 1.6 viser en oversikt over modellen og tilhørende parametere. PLSR

genererer *score*-matrisene T og U, hvor førstnevnte både skal beskrive X og modellere Y. Modellen har også en geometrisk forklaring i det at den projiserer de originale variablene X på et hyperplan som predikerer Y. Her er de nye koordinatene *X-scores*, T. Hyperplanet defineres av linjer bestemt av *loadings*-koeffisientene P' for hver komponent. [38] Slik kan *X-scores* T og *loadings* P' relateres som vist i ligning 1.8. W' og C' er vektorer som beskriver variablenes bidrag til relasjonen mellom X og Y, og derfor modellen (se ligning 1.9 og 1.7). Store verdier indikerer større bidrag. [38].



Figur 1.6: Parametere i PLSR modellen: x-variabler  $x_{ik}$ , y-variabler  $y_{im}$ , *x-score* T, *y-score* U, *loadings* P', X-vektorer W', Y-vektorer C'. Indekser er gitt for matrisene: N for antall objekter, K for antall X variabler og M for antall Y variabler. Figuren er modifisert fra Wold et al (2001). [38]

Fra et analyseperspektiv gir de ulike parametrene ulik informasjon. X og *Y-score* beskriver objektene i modellen, og kan derfor gi informasjon om grupperinger, uteliggere og fordelingen av objektene. Vektene beskriver hvordan variablene kombineres for å relatere X og Y. Dette betyr også at vektene inneholder informasjon om hvilke variabler som er viktige i denne relasjonen. PLSR koeffisientene er også et mål på dette. Loading  $p_{ak}$  gir også informasjon om betydningen av gitte X-variabler for modelleringen, men i form av et mål på hvor mye en variabel bidrar til modelleringen av X. [38]

#### 1.4.4 Validering av statistiske modeller

*Mean squared error* (MSE) er et mål på hvor god en statistisk metode er til å beskrive dataene. MSE gjør dette ved å måle forskjellen mellom predikerte responsverdier fra en modell og de faktiske responsverdiene. [42] Den matematiske definisjonen på MSE er:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (1.12)$$

hvor  $y_i$  er den faktiske responsverdien for prøve  $i$  og  $\hat{f}(x_i)$  er den predikerte responsverdien for tilsvarende prøve fra modellen  $\hat{f}$ . Det er forskjell på *training* MSE og *test* MSE. *Training* MSE refererer til MSE for datasettet som er brukt til å bygge modellen, i motsetning til *test* MSE som refererer til MSE som baserer seg på et helt nytt sett observasjoner. Det er *test* MSE som vil kunne si noe om hvor god modellen er til å predikere nye verdier. Det er ønskelig å velge en modell med minst mulig *test* MSE. I tilfeller hvor det ikke er nok data til å kunne lage et *test*-sett og beregne *test* MSE, så må andre metoder brukes til å estimere denne. Kryssvalidering er en metode som estimerer *test* MSE, og brukes dermed til å evaluere antall komponenter som bør brukes i en PCR eller PLSR modell. [42]

Det fins ulike former for kryssvalidering, men felles for metodene er at de innebærer en inndeling av datasettet der en del brukes til å lage modellen og den andre delen brukes til å estimere *test* MSE. En vanlig form for kryssvalidering er ved bruk av valideringssett. Dette går ut på å dele datasettet inn i to deler, denne inndelingen er helt tilfeldig. Det ene settet brukes til å lage regresjonsmodellen, og det andre settet (valideringssettet) brukes i prediksjon. Det er derfor valideringssettet som blir grunnlaget for estimering av *test* MSE. En annen form for kryssvalidering er *k-Fold Cross-Validation*. I denne typen deles datasettet tilfeldig inn i  $k$  like store grupper. En av disse gruppene (*fold*) brukes som valideringssett, samtidig som resten av gruppene brukes til å lage modellen. *Test* MSE estimeres for prediksjon av dataene i valideringssettet. Dette repeteres for alle gruppene, slik at det alltid er en gruppe data som holdes utenfor selve modelleringen og som senere predikeres. Estimaten for MSE blir gjennomsnittet av  $MSE_1, \dots, MSE_k$ . Det er vanlig å bruke enten 5 eller 10 grupper. Den siste strategien for kryssvalidering kan minne om sistnevnte om  $k =$  antall observasjoner. Dette er *Leave-One-Out Cross-Validation* (LOOCV), som derfor innebærer at en observasjon i datasettet brukes som valideringssett og resten brukes til å lage regresjonsmodellen. Dette er en prosess som repeteres for alle observasjonene, hvor gjennomsnittet av  $MSE_1, \dots, MSE_n$  er estimaten for *test* MSE til modellen. [42]

Det er noen fordeler og ulemper med de ulike typene kryssvalidering. I den førstnevnte strategien deles datasettet tilfeldig i to deler, noe som i mange tilfeller kan være en ulempe. Dette er fordi *test* MSE estimaten vil variere betraktelig avhengig av hvilke observasjoner som er i de ulike settene. Samtidig vil det begrensede antallet observasjoner medføre at modellen blir dårligere. [42] I denne masteroppgaven var antall observasjoner relativt få, og denne typen kryssvalidering ble derfor ikke brukt her. Når man skal vurdere hvilken variant av en modell som er best, så ser man etter hvilken variant som gir lavest *test* MSE [42]. Både LOOCV og  $k$ -fold kryssvalidering har vist evne til å gjøre dette. På et generelt grunnlag vil LOOCV være fordelaktig fordi nesten hele datasettet brukes til å lage modellen. Dette er også tilfellet for  $k$ -fold kryssvalidering, men kun til en viss grad. På den andre siden så vil  $k$ -fold kryssvalidering gi bedre estimater for *test* MSE enn LOOCV, ettersom det er mindre overlappende data mellom de ulike gruppene som brukes i modelleringen. [42] Med dette utgangspunktet er  $k$ -fold den beste

metoden, men samtidig så må det gjøres et valg basert på antall observasjoner tilgjengelig i datasettet.

Variablene som er brukt i modellen har stor betydning for regresjonsmodellen. Kjeldahl et al (2010) oppfordrer til å velge relevante variabler, men samtidig ikke begrense det til de viktigste variablene[44]. Om man bruker datasettet i en dataanalyse for å gjøre en filtrering av betydningsfulle variabler før regresjonsmodellen, så er det viktig å bruke kryssvalidering riktig. Kryssvalideringen blir feilaktig om det da kun gjøres en kryssvalidering av regresjonsmodellen, ettersom forhåndsseleksjonen allerede har “sett” og “brukt” dataverdiene i en form for dataanalyse (videoforelesning med professorene Trevor Hastie og Robert Tibshirani). [45] Om den typen seleksjon skal gjøres på forhånd må dette inkluderes i valideringsprosessen.

#### **1.4.5 utfordringer ved dagens bruk av multivariat dataanalyse for analyse av bioaktive peptider**

QSAR modellering har potensialet til å kunne avdekke relevante sammenhenger mellom peptidstruktur og aktivitet. Dette kan spesielt ha nytteverdi i utviklingen av peptider i mat med spesielle egenskaper, hvis disse egenskapene kan predikeres for kjente peptidstrukturer. [41] Målet med dataanalysen var å se etter og vurdere sammenhenger mellom peptidstruktur-kategorier (X) og antioksidantkapasitet (Y). Regresjon kan brukes til å avdekke sammenhenger som kan belyse bidraget av en X-variabel til en Y-variabel [36]. Regresjonsteknikkene PCR og PLSR ble i utgangspunktet vurdert til denne oppgaven. PCR ble etterhvert utelukket, ettersom metoden kun bruker korrelasjon mellom X-variablene [37]. Det gjøres dermed en antagelse om at de komponentene som beskriver mest variasjon i X også er relatert til Y [42]. I denne oppgaven er det viktig at modellen vektlegger sterke X-Y relasjoner. I PLSR blir X-variabler som korrelerer med Y spesielt vektlagt i analysen [37].

PLSR ble brukt i dataanalysen av fiskeproteinhydrolysater i denne masteroppgaven av flere grunner. Det er en metode som kan behandle datasett med sterke korrelasjoner, men også mye støy. Sistnevnte håndteres gjennom residualer, og på den måten “forstår” PLSR at deler av X ikke har betydning for Y. [38] I denne oppgaven er også antall variabler større enn antall prøver. PLSR blir spesielt mye brukt i akkurat slike tilfeller [36]. Det er også en metode som kan modellere flere Y-responser i en modell [38]. Dette gjør metoden fleksibel, og slik kan flere mål på antioksidantkapasitet også inkluderes i en og samme modell.

Det er likevel flere utfordringer som begrenser nytteverdien av multivariate analyser. Dette er spesielt relatert til input data som grunnlaget for modellen. For å kunne belyse sammenhenger mellom peptidstruktur og biologisk aktivitet, er det viktig at en QSAR modell dekker så mye strukturell variasjon som mulig. Input data bør derfor inkludere prøver som er representative, der alle relevante variasjoner i struktur er representert. [41] Dette gjør design av prosessen før og av modelleringen enda viktigere. En modell vil også avhenge av faktorer som antall prøver og



variabler. En modell kan tilsynelatende beskrive et datasett godt, men samtidig ha begrenset prediksjonsevne for nye observasjoner. [36] Modellen bør videre valideres for datasett med nye, uavhengige prøver, men der dette ikke er tilgjengelig så kan kryssvalidering gjøres [38], [42].

Det er også vanlig å bruke litteratur og databaser til QSAR modellering [41]. Generelt så har antall databaser på bioaktive peptider økt, men det er fortsatt begrenset hvor mange peptider som er fra mat. Funn på bioaktive peptider er som regel heller ikke i form av ett rent peptid, men heller peptidblandinger. [46]. Dette begrenser mulighetene for modellering av databasefunn for bioaktive peptider fra mat, og spesielt hvilke strukturkategorier som kan brukes i modelleringen. Kvaliteten på input data vil påvirke modelleringen i stor grad, og dette krever derfor en kritisk vurdering av litteratur og databaser. Pripp et al. (2005) stiller to krav til bruk av litteratur og databaser i denne typen modellering av peptider. Dette er krav til både likt eksperimentelt oppsett og betingelser, i tillegg til estimer på måleusikkerhet.[41] Dette stiller krav til sammenlignbarhet mellom prøver. Litteratur og databaser ble til en viss grad også vurdert til denne oppgaven. Dette ble etterhvert utelukket, men vurderingene som ble gjort av litteratur og databaser er presentert og diskutert i delkapittel 3.6. Vurdering av litteratur som kilde til data.

## 1.5 Målet med masteroppgaven

Hovedmålet med masteroppgaven var å bruke dataanalyse som et verktøy for å studere sammenhenger mellom antioksidantkapasitet og ulike mål på peptidstruktur som blant annet aminosyresammensetning og peptidstørrelse. Fokuset i oppgaven var på analyse av hydrolysater fra ulike typer fisk, og derfor også på sammenligning av disse.

I prosjektoppgaven fra våren 2020 ble det gjennomført en litteraturstudie og en mindre dataanalyse med samme formål. Dataanalysen var da basert på målinger av fire permeater fra ulike hydrolysater av torskehoder. [20] Grunnlaget for analysen var derfor begrenset til få prøver og prøver fra samme restråstoff. Det ble derfor anbefalt at datagrunnlaget bør være større i fremtidige studier for å generere mer robuste og troverdige resultater. Det ble også anbefalt å bruke andre *in silico* analyser enn PCA-analyse. [20] Et delmål med masteroppgaven var derfor å gjøre dataanalyse på et større datagrunnlag med flere ulike marine restråstoff. Et annet delmål var også å vurdere andre statistiske modeller for å ekstrahere mer informasjon fra datasettene.

Input data ble samlet inn fra flere typer fisk fra ulike master- og prosjektoppgaver som har blitt gjort på Institutt for bioteknologi og matvitenskap ved NTNU. Det ble gjort en kartlegging av prøver og variabler, samtidig som ulike metoder for statistisk modellering ble vurdert. Variabler ble valgt basert på ønsket om å relatere peptidstruktur til antioksidantkapasitet. Det ble gjort beregninger for å sikre sammenlignbarhet mellom de ulike prøvene som opprinnelig stammet fra ulike master- og prosjektoppgaver, både med tanke på enheter og antagelser. Da-

taene ble sortert i flere matriser som ble brukt i dataanalysen. Hovedanalysen bestod av en PLSR-analyse med peptidstruktur-variabler (x-variabler) og ABTS antioksidantkapasitet (y-variabel) for å studere mønster, korrelasjoner og sammenhenger variablene i mellom.

### 1.5.1 Relevante begreper i statistisk analyse

Tabell 1.2 oppsummerer og beskriver en rekke statistiske begreper som vil bli brukt gjennom oppgaven.

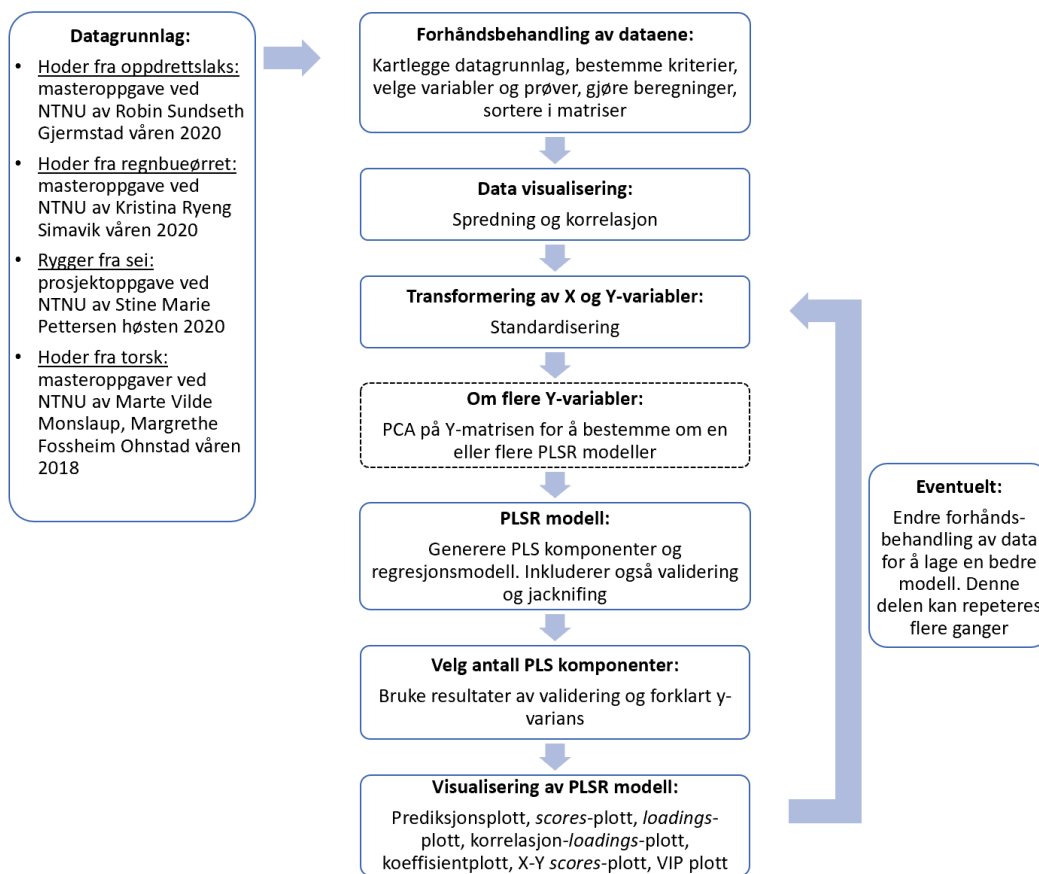
Tabell 1.2: Oversikt over relevante begreper i statistisk analyse

| Begreper  | Beskrivelse   |
|---|---|
| Prediktorvariabler                              | X-variabler som brukes for å predikere en respons Y. Kalles også input variabler. [42]  |
| Responsvariabler                                | Se over. Kalles også Y-variabler eller output variabler. [42]   |
| Deskriptorer                                    | Variabler som beskriver en forbindelse.   |
| <i>Test MSE</i>                                 | MSE er et mål på forskjellen mellom predikerte og faktiske responsverdier. <i>Test MSE</i> viser til modellens evne til å predikere nye verdier, og reflekterer derfor hvor god en modell er. [42]  |
| RMSEP   | $RMSEP = \sqrt{MSEP}$ [47].   |
| Kryssvalidering                                 | Metode som brukes til å estimere <i>test MSE</i> for å evaluere hvilken modell som best beskriver dataene [42].   |
| PLS komponenter                                 | Gjennom delvis minste kvadraters metode (PLS) dannes komponenter som både skal beskrive prediktorene og responsen. De brukes i PLS-regresjonen. [42]  |
| <i>Scores</i>                                   | PLSR modellen kan beskrives gjennom et hyperplan som både skal beskrive X og Y. Koordinatene til prøvene i dette hyperplanet tilsvarer <i>scores</i> . [38]   |
| <i>Loadings</i>                                 | Se forklaring om hyperplan over. Hyperplanet er definert av linjer i form av PLS komponenter. Retningen til linjene beskrives gjennom <i>loadings</i> [38].   |
| Korrelasjons <i>loadings</i>                    | Korrelasjonskoeffisient mellom en gitt prediktorvariabel X og en PLS komponent [47].  |
| Regresjonskoeffisienter                         | Koeffisientene i en regresjonsmodell, rent matematisk: $B$ i $Y = XB + F$ [38].   |
| <i>Variable importance for projection (VIP)</i> | Verdien av en gitt prediktorvariabel X for modelleringen av X og Y [38]. Den matematiske formelen er [48]:<br><br>$VIP = \sqrt{p \sum_{a=1}^A [SS_a(w_{aj} / \ w_a\ ^2)] / \sum_{a=1}^A (SS_a)}$ (1.13)<br>$SS_a$ : "sum of squares", a: indeks komponent, w: vektorer eller loading vektorer [38], [48]. |

## Kapittel 2

# Materialer og metoder

Målet med oppgaven var å bruke dataanalyse som et verktøy for å studere sammenhenger mellom peptidstruktur og bioaktivitet. Sentralt i masteroppgaven var derfor metodeutvikling innen dataanalyse for å kunne studere slike sammenhenger. Flere metoder og ressurser ble vurdert, slik som tidligere beskrevet i kapittel 1.4 Kjemometriske metoder i analyse av proteiner og peptider. Flytskjema i figur 2.1 viser trinnene i dataanalysen, inkludert datagrunnlaget. Trinnene er beskrevet mer detaljert senere i dette kapittelet.



Figur 2.1: Flytskjema som viser trinnene i dataanalysen.

Datagrunnlaget ble styrt av hva som tidligere hadde blitt målt på instituttet, både før og under tidsperioden for denne masteroppgaven. Underveis i oppgaven ble det etterhvert klart at kun en Y-variabel kunne inkluderes i modellen. Et av trinnene i flytskjemaet inkluderer en PCA-analyse av Y-variablene for å kunne se på mønster dem i mellom, og slik bestemme om PLSR-analysen skulle bestå av en eller flere modeller. I tilfeller hvor Y-variablene måler ulike mekanismer og ikke korrelerer vil det være mer fordelaktig å generere flere PLSR modeller [38]. Det var kun tilgang på en Y-variabel i denne oppgaven, og derfor ble en slik PCA-analyse ikke gjennomført. Det er likevel valgt å inkludere det som et trinn i flytskjemaet for å vise resultatet av metodeutviklingen, og at det er et trinn som ville vært naturlig å inkludere for utviklingen av en eller flere PLSR-modeller ved flere Y-variabler. De andre trinnene i flytskjemaet ble gjennomført som planlagt.

## 2.1 Fremgangsplan

Målet med oppgaven ble definert i begynnelsen av arbeidet. Basert på dette ble også en fremgangsplan utarbeidet tidlig. Denne ble etterhvert tilpasset i samsvar med kompetanseheving innenfor statistisk modellering. Ettersom metodeutvikling var en sentral del av denne oppgaven så inkluderes en beskrivelse av fremgangsplanen for å beskrive denne prosessen:

1. Kompetanseheving innen statistisk modellering: teori og programmeringsverktøy
2. Kartlegge mulige datavariabler og observasjoner
3. Valg av dataanalyser og programmeringsverktøy
4. Forberede datascript
5. Utvelging og sortering av variabler og observasjoner i datasett
6. Tilpasse og fullføre datascript
7. Analysere datatabeller for generell oversikt
8. Forhånds-behandling av data
9. Dataanalyse
10. Statistisk validering
11. Tilpasse forhånds-behandling av data for å forbedre modellen

## 2.2 Innsamling og forhånds-behandling av data

Datagrunnlaget, forhåndsbehandling og visualisering av dataene er beskrevet i dette delkapittelet. Dette er de første trinnene som er vist i flytskjema i figur 2.1.

## 2.2.1 Datagrunnlaget

Mulige datavariabler og observasjoner var avhengig av hva som tidligere hadde blitt målt ved Institutt for bioteknologi og matvitenskap. Grunnlaget for data baserte seg på arbeid gjort i forbindelse med tidligere prosjekt- og masteroppgaver. Det ble påpekt i prosjektoppgaven våren 2020 at kvaliteten på en dataanalyse i stor grad avhenger av tilgjengelige data. For å kunne generere mer robuste og troverdige resultater er man avhengig av et større datagrunnlag enn det som ble presentert i prosjektoppgaven. [20] Dette ble derfor et viktig utgangspunkt for denne oppgaven. Datagrunnlaget for denne oppgaven består av hydrolysater fra flere typer fisk: oppdrettslaks, regnbueørret, sei og torsk. Det inkluderer data fra masteroppgavene til Robin Sundseth Gjermstad, Kristina Ryeng Simavik, Marte Vilde Monslaup og Margrethe Fossheim Ohnstad, i tillegg til data fra prosjektoppgaven til Stine Marie Pettersen [49]–[53]. Datagrunnlaget fra de ulike oppgavene inkludert både hydrolyse og analyse-betingelser er beskrevet i mer detalj nedenfor:

- Hoder fra oppdrettslaks (Robin masteroppgave 2020)

Råstoffet ble først kvernet før det ble fryst ved  $-80^{\circ}$ . Råstoffet ble så tint før det gjennomgikk en enzymatisk hydrolyseprosess med enzymene papain og bromelain. Råstoff og destillert vann (1:1) ble tilsatt til en reaktor, før enzymatisk hydrolyse ved  $50^{\circ}\text{C}$  i 60 minutter. Enzymene ble deaktivert ved varmebehandling. Etter sentrifugering ble proteinfasen isolert. Prøvene som ble analysert videre var frysetørket proteinfraksjon. Prøvene ble lagret i romtemperatur i 0, 1, 2, 4 og 6 måneder før analyse.

Variablene som ble målt for mesteparten av de overnevnte laksehydrolysaterne var: tørrstoff, aske, hydrolysegrad, proteininnhold, nitrogeninnhold, mengde vannløselige proteiner, syreløselige proteiner, total aminosyresammensetning, frie aminosyrer, SH-grupper, molvektfordeling og fargemåling. [49] I etterkant av Gjermstad sin masteroppgave ble antioksidantkapasitet målt gjennom en ABTS-analyse på samme hydrolysater av Peter Benjamin Kelley (lærling på instituttet).

- Hoder fra regnbueørret (Kristina masteroppgave 2020)

Hydrolyseprosessen var lik for ørret- og laksehydrolysaterne med samme enzymer og hydrolysebetingelser. Prøvene ble derimot lagret i romtemperatur i 0, 2, 4 og 6 måneder.

Variablene som ble målt for mesteparten av de overnevnte ørrethydrolysaterne var: tørrstoff, aske, vannløselige proteiner, syreløselige proteiner, SH-grupper, nitrogeninnhold, proteininnhold, total aminosyresammensetning, frie aminosyrer, molvektfordeling, hydrolysegrad og fargemåling. [50] I etterkant av Simavik sin masteroppgave ble antioksidantkapasitet målt gjennom en ABTS-analyse på samme hydrolysater av Peter Benjamin Kelley (lærling på instituttet).

- Ryggbein fra sei (Stine prosjektoppgave 2020)

Råstoffet ble kvernet og fryst ned. Råstoff og destillert vann (1:1) ble tilsatt til en bioreaktor. Enzymene som

ble brukt i hydrolyseprosessen var papain og bromelain. Den enzymatiske hydrolysen varte i 60 minutter ved 50°C. Enzymene ble deaktivert ved varmebehandling. Etter sentrifugering ble proteinfasen isolert. Hydrolysatene ble så frysetørket. Prøvene ble så fryst før analyse. Prøvene som ble analysert var BOI, BOII, BJI og BJII. Forkortelsene viser til ryggbein fra seifangst i oktober (parallell I og II) og fra seifangst i januar (parallell I og II). [54]

Variablene som ble målt for de overnevnte seihydrolysaterne var: nitrogeninnhold, proteininnhold, lipidinnhold, frie aminosyrer, total aminosyresammensetning og ABTS-målinger [51].

- Hoder fra torsk (Marte Vilde og Margrethe masteroppgaver 2018)

Hydrolysaterne ble produsert av SINTEF Ocean i prosjektet HEADS UP. Råstoffet (400 kg) ble kvernet og tilsatt til en hydrolysetank, og videre ble ulike mengder vann tilsatt. Blandingen ble varmet opp til 50°C, og enzymene papain og bromelain ble tilsatt. Den enzymatiske hydrolysen varte i 60 minutter ved 50°C, før enzymene ble deaktivert ved °C i 10 minutter. De isolerte proteinhydrolysaterne ble fryst en periode før frysetørking. Det ble gjennomført 13 ulike hydrolyseprosesser med ulike hydrolysebetingelser. T1 (fangst i februar, 400 kg vann til hydrolyse), T5 (fangst i mars, 200 kg vann), T9 (fangst i mars, 200 kg limvann) og T13 (fangst i mai, 300 kg vann) ble brukt i analyser i masteroppgavene til Margrethe Fossheim Ohnstad og Marte Vilde Monslaup. [52], [53]

Variablene som ble målt for disse hydrolysaterne var: proteininnhold, hydrolysegrad, molvektfordeling, total aminosyresammensetning, frie aminosyrer, ACE-hemming, ABTS-målinger, jernbinding, Folin-Ciocalteu reduserende kapasitet og antioksidantkapasitet i liposom modellsystem. Hydrolysegrad ble presentert i både Marte Vilde og Margrethe sine oppgaver, men det ble henvist til Ayat Asfour (masteroppgave våren 2018) for målinger og beregninger av denne variabelen. [52], [53], [55]

## 2.2.2 Utvelging og sortering av variabler og observasjoner

### Valg av variabler og observasjoner

Det ble definert en rekke kriterier for valg av prediksjons-og responsvariabler og prøver fra de overnevnte masteroppgavene. Felles kriterier for prøvene og variablene var at det måtte være kvantitative parametere, uavhengig av enheter. Enheter vil ikke ha noen betydning ettersom både X og Y-variabler skal skaleres før PLSR-analysen, som beskrevet i delkapittel 2.3.1 Standardisering av variabler. De overnevnte dataene tilhører samme institutt, og det antas derfor at laboratorieanalysene ble utført med samme metoder og type utstyr.

Kriteriene som ble bestemt for prediktorvariablene i forkant av dataanalysen var følgende:

- Prediktorvariabler

Variablene må være relatert til peptidstruktur. Dette kan være type aminosyrer og størrelse, men begrenses ikke kun til dette.

Valg av prediktorvariabler ble gjort i tråd med prinsippet for QSAR-modeller som går ut på å relatere X= kjemisk struktur til Y= kjemiske egenskaper, biologisk aktivitet eller reaktivitet [38]. I litteraturstudien fra prosjektoppgaven våren 2020 ble det gjort flere funn på struktur-aktivitet sammenhenger, som i stor grad var relatert til størrelse på peptider (MW og kjedelengde) og type aminosyrer. Litteraturstudien ga derfor indikasjoner på hvilke struktur kategorier som kan være relatert til bioaktivitet. Dette betyr at disse kategoriene bør inkluderes som prediktorvariabler i dataanalysen. Av flere grunner ble også andre struktur kategorier inkludert i kriteriene. En av grunnene til dette var å ikke utelukke andre viktige sammenhenger, ettersom litteraturstudien kun baserte seg på et fåtall faglige artikler. Litteraturstudien viste også til funn som tilsier at aktiviteten er kompleks, mekanismene er mange og vil ikke nødvendigvis være de samme i ulike *in vitro* og *in vivo* miljøer. [20] Formålet med analysen var å studere mønster i dataene, og det var derfor et mål å unngå begrensninger som kan føre til tap av viktig informasjon.

Basert på kriteriene og overnevnte datagrunnlag ble følgende prediktorvariabler valgt til dataanalysen: hydrolysegrad, mengde SH-grupper, total aminosyresammensetning og mengde frie aminosyrer. Disse ble vurdert til å være ulike mål på peptidstruktur. De to sistnevnte sier noe om type aminosyrer. Hydrolysegrad reflekterer grad av brutte bindinger og derfor indirekte også peptidstørrelse. Mengde SH-grupper er også et direkte mål på peptidstruktur. Molekvektfordeling ble også vurdert til å være et mål på peptidstruktur, men den ble utelukket ettersom det ikke er en kvantitativ parameter som kan oversettes til en enkelt verdi for hver prøve. Det er en parameter som vises i form av en fordeling, og kan derfor ikke gi en konkret verdi for hver prøve.

Kriteriene som ble bestemt for responsvariablene i forkant av dataanalysen var følgende:

- Responsvariabler

Variablene må være mål på bioaktivitet. Dette begrenses av hva som måles på instituttet.

I prosjektoppgaven ble to typer bioaktivitet studert, både ACE-hemming og antioksidant-aktivitet. Kriteriene ovenfor ble definert før antioksidantkapasitet ble valgt som videre fokus. Dette valget var basert på hvilke analyser som ble gjort på instituttet. Litteraturstudien fra prosjektoppgaven viste til at det er flere analysemetoder som kan brukes til å måle antioksidantkapasitet, og at disse innebærer ulike analysebetingelser og baseres på ulike reaksjoner og mekanismer [20]. Av den grunn inkluderer kriteriene for responsvariabler alle mål på antioksidantkapasitet. På instituttet ble det gjort ABTS-målinger på alle prøvene, og derfor ble disse målingene brukt som responsvariabel i dataanalysen. De andre antioksidant-analysene som hadde blitt gjort på torsk ble utelukket fra dataanalysen, ettersom datagrunnlaget var begrenset til tre prøver.

Kriteriene som ble bestemt for prøvene i forkant av dataanalysen var følgende:

- Prøver

Det må være variasjon mellom prøvene. Variasjon kan bety ulike typer fisk, ulik lagringstid eller at prøvene har blitt analysert av ulike personer. Dette begrenses av hva som er karakterisert på instituttet.

Kriteriene for prøvene ble bestemt for å sikre et mer representativt utvalg av prøver før en dataanalyse. Basert på hva som hadde blitt karakterisert på instituttet (delkapittel 2.2.1 Datagrunnlaget) ble det inkludert prøver av oppdrettslaks, regnbueørret, sei og torsk i analysen. En sentral del av masteroppgavene til Kristina og Robin gikk ut på å studere forskjeller mellom prøver som har blitt tilsatt diverse antioksidanter under hydrolyseprosessen [49], [50]. Disse prøvene ble ikke inkludert i dataanalysen, ettersom det ikke er mulig å skille antioksidantkapasiteten til henholdsvis peptidene og tilsatte antioksidanter. Kontrollprøvene ble ikke tilsatt antioksidanter, og derfor ble disse prøvene inkludert i dataanalysen [49], [50]. Dette var også tilfellet for en av torskehdrolysatene i masteroppgavene til Margrethe og Marte Vilde, hvor prøve T13 hadde blitt tilsatt antioksidant under hydrolysen. Prøvene T1, T5 og T9 ble inkludert i analysen ettersom de ikke ble tilsatt antioksidanter under hydrolysen. [52], [53] Blant Stine sine prøver ble både BOI, BOII, BJI og BJII inkludert, altså prøver fra flere seifangster og flere parallelle hydrolyseprosesser.

På grunn av forsinkelser i laboratoriearbeidet på instituttet høsten og våren 2020/21 ble ikke antioksidantkapasitet målt for flere av ørrehydrolysatene: 0, 2 og 4 måneder lagring. Det ble gjort forberedelser på disse prøvene i form av beregninger og sortering av de andre variablene i matriser. Det ble ikke kjørt dataanalyse med disse prøvene ettersom dataanalysen er avhengig av antioksidantkapasitet som responsvariabel. Oppsummert ble følgende prøver inkludert i dataanalysen fra de ulike masteroppgavene [49]–[53]:



Tabell 2.1: Oversikt over prøvene (hydrolysatene) som ble inkludert i dataanalysen, inkludert informasjon om råstoffet, enzym til hydrolyse og betydningen av prøvenavnene. RT: romtemperatur.

| Prøvenavn | Råstoff       | Del av fisk | Enzymer            | Notasjoner  | Fra oppgave: |
|-----------|---------------|-------------|--------------------|---|--------------|
| R 0 mnd   | Oppdrettslaks | Hoder       | Papain + Bromelain | Lagringsperiode i RT: 0 mnd   | [49]         |
| R 1 mnd   | Oppdrettslaks | Hoder       | Papain + Bromelain | Lagringsperiode i RT: 1 mnd   | [49]         |
| R 2 mnd   | Oppdrettslaks | Hoder       | Papain + Bromelain | Lagringsperiode i RT: 2 mnd   | [49]         |
| R 4 mnd   | Oppdrettslaks | Hoder       | Papain + Bromelain | Lagringsperiode i RT: 4 mnd   | [49]         |
| R 6 mnd   | Oppdrettslaks | Hoder       | Papain + Bromelain | Lagringsperiode i RT: 6 mnd   | [49]         |
| K 6 mnd   | Regnbueørret  | Hoder       | Papain + Bromelain | Lagringsperiode i RT: 6 mnd   | [50]         |
| BOI       | Sei           | Ryggbein    | Papain + Bromelain | Fangst i oktober<br>I: nr 1 av parallelle hydrolyser                    | [51]         |
| BOII      | Sei           | Ryggbein    | Papain + Bromelain | Fangst i oktober<br>II: nr 2 av parallelle hydrolyser                   | [51]         |
| BJI       | Sei           | Ryggbein    | Papain + Bromelain | Fangst i januar<br>I: nr 1 av parallelle hydrolyser                     | [51]         |
| BJII      | Sei           | Ryggbein    | Papain + Bromelain | Fangst i januar<br>II: nr 2 av parallelle hydrolyser                    | [51]         |
| T1        | Torsk         | Hoder       | Papain + Bromelain | Prøvenavn viser til prøvenummer.<br>Se delkapittel 2.2.1 Datagrunnlaget | [52], [53]   |
| T5        | Torsk         | Hoder       | Papain + Bromelain | Prøvenavn viser til prøvenummer.<br>Se delkapittel 2.2.1 Datagrunnlaget | [52], [53]   |
| T9        | Torsk         | Hoder       | Papain + Bromelain | Prøvenavn viser til prøvenummer.<br>Se delkapittel 2.2.1 Datagrunnlaget | [52], [53]   |

Det ble bestemt egne kriterier for variabler og prøver fra litteratur og databaser. Artikkene fra litteraturstudien i prosjektoppgaven ble så gjennomgått basert på kriteriene i tillegg til relevante databaser. Av de databasene som ble vurdert var kun BIOPEP relevant for antioksidantkapasitet og peptider fra mat [36]. Data fra gjennomgått litteratur og database ble ikke inkludert i analysen av flere grunner. Kriteriene og vurderingene som ble gjort er nærmere beskrevet i kapittel 3 Resultater og diskusjon.

### 2.2.3 Beregninger

Dataene for total aminosyresammensetning og frie aminosyrer ble oppgitt i regneark (Microsoft Excel) fra de ulike master-og prosjektoppgavene. Regnearkene bestod av rådata og diverse beregninger som var relevante for de ulike oppgavene. De andre variablene var oppgitt i master-og prosjektoppgavene. Ettersom dataene var basert på ulike oppgaver fra ulike personer så varierte datainnholdet i både beregninger, tilgang på rådata og oppsett av regneark. Det ble derfor viktig å tilpasse dataene så de var sammenlignbare. Regnearkene ble derfor gjennomgått, de relevante dataene ble samlet i et regneark og nye beregninger ble gjort for å sørge for at dataene var direkte sammenlignbare med tanke på enheter, antagelser og beregninger. Det var ulike utgangspunkt for de ulike oppgavene og dermed også prøvene, som beskrevet i vedlegg A i appendiks: Rådata og beregninger. Blant annet så var det ulike antagelser om molekylvekt i de ulike oppgavene, hvor enten molekylvekten for frie aminosyrer eller for aminosyrer

bundet i peptidkjede ble brukt. Som beskrevet tidligere bidrar bindingen i en peptidkjede til tap av et vannmolekyl [17]. Derfor ble det bestemt at molekylvekten for frie aminosyrer skulle være inkludert molekylvekten for et vannmolekyl, i motsetning til molekylvekten til aminosyrene i total aminosyresammensetning. Molekylvekt ble endret og beregninger gjort på nytt for å sikre sammenlignbarhet mellom prøvene. Både datagrunnlag og hvilke endringer og beregninger som ble gjort i denne masteroppgaven er beskrevet for de ulike prøvene i vedlegg A i appendiks.

Generelt for alle dataene ble gjennomsnittsverdiene av parallellene for hver variabel brukt, istedenfor de individuelle verdiene. I de tidligere oppgavene ble det lagd og analysert flere paralleller for de ulike prøvene, men ettersom parallellene ikke nødvendigvis er de samme gjennom alle laboratorieanalysene ble heller gjennomsnittsverdiene brukt. Som mål på sammensetning av totale og frie aminosyrer ble % aminosyre brukt videre i matrisene.

#### **2.2.4 Sortering av dataene i matriser**

Basert på kriteriene beskrevet i delkapittel 2.2.2 ble det gjort en utvelging av variabler som i hovedsak baserte seg på om variablene beskrev enten peptidstruktur eller antioksidantkapasitet. Følgende prediktorvariabler X ble derfor brukt i dataanalysen: total aminosyresammensetning, frie aminosyrer, SH-grupper og hydrolysegrad. Aminosyren Aba ble ekskludert fra analysen, ettersom det ikke er en av de vanlige aminosyrene. Responsvariabelen Y ble antioksidantkapasitet i form av ABTS-målinger. Ved bruk av transponering ble variablene samlet i matriser med prøver (rader) og variabler (kolonner). Ettersom dataene var basert på ulike oppgaver fra ulike personer så varierte datainnholdet. Innholdet i matrisene ble basert på hvilke variabler som hadde blitt målt for hver prøve. Det ble i utgangspunktet lagd 8 matriser som skulle brukes i dataanalysen, men ettersom flere ørrethydrolysater ikke ble målt for ABTS antioksidantkapasitet ble det ikke gjort dataanalyse for matrise 3 og 7 (FAA, TAA, SH-grupper, hydrolysegrad, ABTS antioksidantkapasitet) som hadde for få prøver.

Tabell 2.2 viser en oversikt over hvilke prøver og variabler som ble inkludert i de ulike matrisene som ble brukt videre i dataanalyse som beskrevet under delkapittel 2.2.5 Generell analyse av dataene og i kapittel 2.3 Delvis minste-kvadraters regresjon (PLSR).

Tabell 2.2: Oversikt over matrisene som ble brukt videre i dataanalyse, inkludert innhold av prøver (rader), X og Y-variabler (kolonner)

| <b>Matrise:</b> | <b>Prøver</b>  | <b>X-variabler</b>   | <b>Y-variabler</b>                                   |
|-----------------|--|--|--|
| 1               | Alle   | FAA (%)<br>TAA (%)   | ABTS antioksidantkapasitet<br>(umol PG/g hydrolysat) |
| 2               | Oppdrettslaks (0, 1, 2, 4, 6 mnd)<br>Regnbueørret (6 mnd)              | FAA (%)<br>TAA (%)<br>SH-grupper (nmol/mg protein)                           | ABTS antioksidantkapasitet<br>(umol PG/g hydrolysat) |
| 4               | Oppdrettslaks (0, 6 mnd)<br>Regnbueørret (6 mnd)<br>Torsk (T1, T5, T9) | FAA (%)<br>TAA (%)<br>Hydrolysegrad (%)                                      | ABTS antioksidantkapasitet<br>(umol PG/g hydrolysat) |
| 5               | Alle   | FAA deskriptorer (%)<br>TAA deskriptorer (%)                                 | ABTS antioksidantkapasitet<br>(umol PG/g hydrolysat) |
| 6               | Oppdrettslaks (0, 1, 2, 4, 6 mnd)<br>Regnbueørret (6 mnd)              | FAA deskriptorer (%)<br>TAA deskriptorer (%)<br>SH-grupper (nmol/mg protein) | ABTS antioksidantkapasitet<br>(umol PG/g hydrolysat) |
| 8               | Oppdrettslaks (0, 6 mnd)<br>Regnbueørret (6 mnd)<br>Torsk (T1, T5, T9) | FAA deskriptorer (%)<br>TAA deskriptorer (%)<br>Hydrolysegrad (%)            | ABTS antioksidantkapasitet<br>(umol PG/g hydrolysat) |

Syrehydrolyse brukes for å bestemme total aminosyresammensetning før deteksjon gjennom HPLC [49], [50], [52], [53]. Hydrolyseprosessen gjør at Asn og Gln omdannes til Asp og Glu [56]. På den måten vil total aminosyresammensetning måle summen av Asp og Asn i hydrolysatene som Asp, og tilsvarende for Glu og Gln som Glu. Derfor ble variablene Asn, Gln, Asp, Glu (fra HPLC) endret til Asp/Asn og Glu/Gln i total aminosyresammensetning (%): % Asp/Asn= % Asp + % Asn og % Glu/Gln= % Glu + % Gln.

De fleste prediktorvariablene er relatert til aminosyrer, som kan beskrives gjennom ulike egenskaper og kategorier, som vist i delkapittel 1.2.1 Proteinstruktur. Det ble derfor også lagd nye variabler som kunne beskrive aminosyresammensetning gjennom noen av de kjente egenskapene til de ulike aminosyrene. De nye variablene ble kalt aminosyredeskriptorer, og inkluderer følgende strukturkategorier: upolare R, aromatiske R, polare R, positivt og negativt ladete R [18]. Tabell 2.3 viser en oversikt over hvilke aminosyrer som ble inkludert i de ulike kategoriene [18]. Kategoriene gjelder for nøytral pH, noe som også er tilfellet for antioksidantkapasitet målt med TEAC-metoden basert på ABTS [18], [32].

Tabell 2.3: Aminosyredeskriptorer med tilhørende aminosyrer (gjelder for pH 7) [18].

| Upolare R | Aromatiske R | Polare R | Positivt ladede R | Negativt ladede R |
|-----------|--------------|----------|-------------------|-------------------|
| Ala       | Phe          | Ser      | Lys               | Asp               |
| Val       | Tyr          | Thr      | His               | Glu               |
| Leu       | His          | Asn      |                   |                   |
| Ile       |              | Gln      |                   |                   |
| Met       |              | Lys      |                   |                   |
| Phe       |              | His      |                   |                   |
| Tyr       |              | Asp      |                   |                   |
|           |              | Glu      |                   |                   |

Flere aminosyrer ble ikke detektert gjennom den brukte eksperimentelle metoden og ble derfor ikke inkludert i inndelingen: Trp, Pro og Cys [49]–[53]. Aminosyrene Gly og Arg kan kun detekteres samlet fra HPLC, og ettersom Arg er en polar, ladet aminosyre og Gly er en upolar, nøytral aminosyre så har Gly/Arg blitt ekskludert fra inndelingen [18]. Som beskrevet over, så tilsvarende Asp og Glu fra HPLC-analyse henholdsvis Asn/Asp og Gln/Glu i total aminosyresammensetning i hydrolysatene. Glu og Asp er de eneste aminosyrene med en negativt ladet sidekjede ved nøytral pH [18]. For å inkludere denne aminosyredeskriptoren for total aminosyresammensetning ble Asp og Glu (fra HPLC) brukt i inndelingen, selv om dette også inkluderer de nøytrale aminosyrene Asn og Gln. Aminosyredeskriptorene ble brukt som variabler for sammensetningen av totale og frie aminosyrer ved å summere aminosyrene i hver kategori, og slik få sammensetning basert på kjemiske karaktertrekk. Aminosyredeskriptorene ble så inkludert i noen av matrisene som vist i tabell 2.2.

### 2.2.5 Generell analyse av dataene

Analyseverktøyet R er et verktøy som er tilpasset statistiske analyser, og som ble brukt gjennom hele dataanalysen i denne oppgaven [57]. I forkant av PLSR analysen ble det gjort en generell analyse for å få oversikt over spredning og korrelasjon i datamatrixene. Det ble generert et boksplokk gjennom den innebygde funksjonen “boxplot” i R for å studere innholdet i datamatrixene, inkludert spredning i dataene. Gjennomsnitt, varians og standardavvik ble også beregnet for hver variabel for å få oversikt over spredning i dataene.

PLSR er en metode som fungerer godt på data med mange korrelerende X-variabler [38]. Det ble derfor også gjort beregninger på korrelasjon ved hjelp av flere innebygde funksjoner i analyseverktøyet R. Resultatene av dette ble kun brukt til å få oversikt over korrelasjoner i matrisene før de samme matrisene ble brukt i en PLSR-analyse. Pearson korrelasjons-koeffisient, også kalt produkt-moment korrelasjonskoeffisient  $r$  er et mål på lineær sammenheng mellom to variabler [37], [42]. Rent matematisk tilsvarende koeffisienten  $r$  kovariansen av to variabler dividert på produktet av de to standardavvikene. Koeffisienten gir verdier mellom -1 og 1, hvor ytterpunktene viser perfekt negativ/positiv korrelasjon. Om  $r=0$  betyr dette at det ikke er en lineær sammenheng mellom de to variablene.

Andre korrelasjonskoeffisienter som Spearman og Kendall brukes i tilfeller hvor verdiene for variablene viser til en type rangering av egenskaper istedenfor egne kvantitative enheter. [37] Dette er ikke relevant i denne analysen. Den innebygde funksjonen “cor” ble brukt for å beregne Pearson korrelasjonskoeffisient mellom en og en variabel. Den innebygde funksjonen “corrplot” i pakken “corrplot” ble brukt for lage en grafisk fremstilling av korrelasjonsmatrisen.

Datascriptene fra analyseverktøyet for både den generelle analysen og videre PLSR-analyse er vist i vedlegg B i appendiks.

## 2.3 Delvis minste-kvadraters regresjon (PLSR)

For å gjøre PLSR-analysen ble “pls” pakken i analyseverktøyet R benyttet [57], [58]. Flytskjema i figur 2.1 viser trinnene i dataanalysen. Datagrunnlag, forhåndsbehandling og visualisering av data ble beskrevet i kapittel 2.2. Resten av dataanalysen skal beskrives i dette delkapittelet. Importeringen av datasettet fra Microsoft Excel til R ble gjort gjennom funksjonene “read\_excel” og “attach”. Funksjonene “rownames” og “as.matrix” ble brukt til å justere importert datasett, hvor førstnevnte ble brukt til å definere første kolonne som radnavn og sistnevnte ble brukt til å omgjøre format fra datatabell til matrise.

### 2.3.1 Standardisering av variabler

Standardisering av både X og Y-variabler gjøres normalt før PLS [42]. PLSR er en metode som i stor grad påvirkes av skalering av variabler [38]. Både X og Y-variablene har ulike enheter. Av den grunn ble både X og Y-variablene standardisert før PLSR-analyse ved bruk av funksjonen “stdize” i analyseverktøyet R. Dette er en vanlig type datatransformasjon, som innebærer å skalere og sentrere hver variabel. Rent matematisk gjøres dette ved å dividere hver enkelt variabel med tilhørende standardavvik og subtrahere fra gjennomsnittsverdien. [38], [58] Slik blir alle variablene likt vektlagt i analysen [38].

Skaleringen gjør at modellen ikke vil avhenge av de ulike måleenhetene, ettersom de standardiserte variablene har et standardavvik på 1 [42]. Regresjonskoeffisientene fra PLSR-modellen er viktige i tolkningen av viktige variabler fra datasettene, og de har også tilsvarende enheter som både X og Y. Derfor er det spesielt viktig at X-variablene er på samme skala for lettere tolkning. [44] Oppsummert, så bidrar standardiseringen til at variablene er sammenlignbare og enklere å tolke. Standardisering ble gjort for både X og Y i alle matrisene, selv om det strengt tatt ikke var nødvendig for de matrisene med X-variabler med like enheter.

### 2.3.2 PLSR-modellen

Scriptet for PLSR-analysen er vist i appendiks i delkapittel B.2.2, og ble kopiert, tilpasset etter de individuelle matrisene og lagret slik at det var ett PLSR-script per matrise. Tilpasningene innebærte spesifikasjoner for ny input i form av ny matrise, eksempelvis nye matrisedimensjoner. Flere innebygde funksjoner med tilpassede spesifikasjoner i analyseverktøyet R ble brukt i dataanalysen, som vist i datascriptet i vedlegg B i appendiks. Den innebygde funksjonen “`plsr`” ble brukt til å gjøre selve PLS regresjonen. Spesifikasjoner ble lagt til funksjonen for blant annet type kryssvalidering og PLSR algoritme. [58] De standardiserte X og Y-variablene i matriseform var input til PLSR modellen, og det ble utviklet henholdsvis en kryssvalidert modell per matrise. PLSR algoritmen NIPALS ble brukt til analysen [58]. Funksjonen “`set.seed`” ble brukt i forkant av funksjonen “`plsr`” for å sikre reproduserbarhet. For andre funksjoner som innebærer en form for generering av tilfeldige tall, så vil “`set.seed`” sørge for at resultatet kan reproduseres [42]. Derfor ble ett unikt tall for “`set.seed`” brukt for hver matrise, i hvert script.

Valget av PLS komponenter som ble inkludert i modellen ble basert på resultatene av forklart y-varians og kryssvalidering for hver komponent. Forklart y-varians er et mål på mengde informasjon om y-variabelen som er i en modell med n komponenter, dette er tilsvarende for forklart x-varians [42]. Resultatene for både forklart x- og y-varians angis ved bruk av funksjonen “`summary`” av `plsr`-analysen og vises for 1, 2...n komponenter. Denne funksjonen angir også resultatene av kryssvalideringen i form av *root mean squared error of prediction* RMSEP for 1, 2...n komponenter. I tillegg ble funksjonen “`validationplot`” brukt til å visualisere resultatene av kryssvalideringen i form av *mean squared error of prediction* MSEP og basert på disse velge antall PLS komponenter som bør brukes i modellen. [58] Modellen som ble valgt hadde derfor et antall komponenter som ga en relativt lav MSEP og i tillegg beskrev en betydelig andel y-varians.

Kryssvalideringsmetoden LOOCV ble brukt til validering og valg av antall PLS komponenter, ettersom data-grunnlaget var begrenset til relativt få observasjoner. Dette innebærte at k antall PLSR modeller ble lagd (k= antall prøver), og for hver modell ble en prøve utelatt fra treningen av modellen og ble heller brukt til prediksjon for å få et mål på MSEP. Gjennomsnittet av MSEP for en global og gjennomsnittlig PLSR modell med 1, 2...,n komponenter vises deretter gjennom de overnevnte funksjonene “`summary`” og “`validationplot`”. Den globale, gjennomsnittlige PLSR modellen med valgte n komponenter ble brukt til videre tolkning av datasettene gjennom blant annet *scores*, *loadings*, regresjonskoeffisienter, dette er nærmere beskrevet i slutten av dette delkapittelet. Den globale PLSR modellen kan anses som gjennomsnittet av alle modellene i kryssvalideringen, men ettersom kun en observasjon er utelukket i hver modell så vil modellen være tilnærmet lik en ny PLSR modell med alle prøvene. Dette ble også testet for en av matrisene ved å kjøre en PLSR modell uten kryssvalidering med valgte n komponenter, som viste at forklart x og y varians, *scores* og *loadings* så likt ut.

LOOCV ble nærmere beskrevet i delkapittel 1.4.4 Validering av statistiske modeller, i tillegg til hvordan filtrering av variabler bør behandles i en valideringsprosess. Datasettene i denne oppgaven ble ikke benyttet i selve filtreringsprosessen av variabler, ettersom utvelgelsen heller baserte seg på relevans for problemstillingen. Dette innebærte at X-variablene skulle reflektere peptidstruktur og Y-variablene skulle reflektere antioksidantkapasitet. Av den grunn ble valideringen gjort kun av PLSR-modellene. *Jack-knifing* ble brukt på regresjonskoeffisientene for å finne tilhørende standardavvik [38], [58]. Standardavvikene baseres på variasjon mellom komponentene i de ulike modellene som testes under kryssvalideringen [38].

Funksjonen “*pls*” genererer flere output [58]. En rekke andre funksjoner ble brukt til å visualisere resultatene av den globale og kryssvaliderede PLSR-modellen gjennom følgende plott som også ble brukt videre i tolkningen av datasettene:

- Kryssvalidert predikert y-verdi mot målt y-verdi i en modell med valgte n komponenter.
- *Scores*-plott over prøvene hvor aksene defineres av komponentene.
- Korrelasjon-*loadings*-plott og/eller *loadings*-plott over prediktorvariablene hvor aksene defineres av komponentene.
- Koeffisienter for de ulike prediktorvariablene i en modell med valgte n komponenter, inkludert standardavvik generert gjennom *jack-knifing* som beskrevet ovenfor.
- *X-scores* mot *Y-scores* i en modell med valgte n komponenter.
- VIP plott for prediktorvariablene. Som nevnt i delkapittel 1.4.6. Relevante begreper i statistisk analyse er VIP (*variable importance for the projection*) et matematisk mål på verdien av en prediktorvariabel for modelleringen av både X og Y [38].

Funksjonen “*jack.test*” ble brukt til å gjøre “*grove*” t-tester for regresjonskoeffisientene basert på variansestimater fra *jack-knifing* [47], [58].

## Kapittel 3

# Resultater og Diskusjon

Dataanalysen i denne masteroppgaven er i stor grad basert på målte prøver og variabler fra tidligere prosjekt- og masteroppgaver. I denne oppgaven ble det gjort nye beregninger av flere av variablene for å oppnå sammenlignbarhet. Beregningene er vist i appendiks: Tillegg A Rådata og beregninger. Resultatene for ABTS for laks- og ørrethydrolysatene var fra nye rådata, og derfor er de presentert i dette kapittelet under delkapittel 3.1. Forhåndsbehandling av variablene. Resultatene for ABTS for torskehydrolysatene, etter omberegning av enheter, er også presentert her. Videre er resultatene av dataanalysen presentert og diskutert i delkapitler inndelt etter matriser.

Først presenteres en generell oversikt over dataene i form av spredning og korrelasjon i de ulike matrisene. Deretter presenteres PLSR-modellene. For den første modellen er modellkvalitet og bakgrunnen for valg av antall PLS komponenter presentert og diskutert i større grad enn for de resterende modellene. Dette er for å redegjøre for viktige modellparametere tidlig, og dette vil også gjelde for resterende modeller. I lys av dette presenteres og diskuteres så funnene på proteinstruktur-antioksidantkapasitet sammenhenger. I resultater og diskusjon av resterende modeller er dette fokuset, og modellene vil kun presenteres i korte trekk. PLSR-modellene for henholdsvis matrise 1 og 5 vil bli gjennomgått i større grad på grunn av bedre modellrobusthet og kvalitet basert på et større datagrunnlag. De andre modellene består også i stor grad av samme variabler som matrise 1 og 5 med unntak av enkeltvariabler.

Til slutt presenteres vurderingene som ble gjort av litteratur og databaser som mulig kilde til data i en multivariat dataanalyse.



## 3.1 Forhånds-behandling av variablene

### 3.1.1 ABTS antioksidantkapasitet

Peter Benjamin Kelley gjorde ABTS-målinger av de samme ørret og lakse-hydrolysatene som ble studert i Gjermstad og Simavik sine masteroppgaver [49], [50]. I denne masteroppgaven ble det gjort beregninger av ABTS antioksidantkapasitet basert på rådataene fra Kelley, og resultatene er vist i tabell 3.1. ABTS antioksidantkapasitet for torskhydrolysatene er også gitt i denne tabellen som referansepunkt. ABTS-målinger og beregninger hadde blitt gjort for torsk i Monslaup sin masteroppgave. Resultatet ble presentert av Monslaup som gjennomsnittlig  $\mu\text{mol PG/g}$  løselig protein. [52] Rådata var ikke tilgjengelig, så i denne oppgaven ble det gjort omberegninger fra gjennomsnittlig  $\mu\text{mol PG/g}$  løselig protein til  $\mu\text{mol PG/g}$  hydrolysat for å sikre sammenlignbarhet mellom prøvene. Beregningene er beskrevet i appendiks for både laks, ørret og torsk (A.1, A.2 og A.4).

Tabell 3.1: ABTS antioksidantkapasitet ( $\mu\text{mol PG/g}$  hydrolysat) for laks, ørret og torsk-hydrolysater. Resultatene er gitt som gjennomsnitt  $\pm$  standardavvik,  $n=6$  for laksehydrolysatene og  $n=5$  for ørrethydrolysatet.

| <b>Prøver:</b>                 | <b><math>\mu\text{mol PG/g}</math> hydrolysat</b> |
|--------------------------------|---|
| Laks 0 mnd                     | $16,2 \pm 2.3$                                    |
| Laks 1 mnd                     | $17,9 \pm 0.5$                                    |
| Laks 2 mnd                     | $15,8 \pm 0.5$                                    |
| Laks 4 mnd                     | $16,3 \pm 1.9$                                    |
| Laks 6 mnd                     | $19,6 \pm 2.4$                                    |
| Ørret 6 mnd                    | $12,0 \pm 2.4$                                    |
| Torsk feb 400H <sub>2</sub> O  | 20,0  |
| Torsk mar 200H <sub>2</sub> O  | 21,7  |
| Torsk mar 200sH <sub>2</sub> O | 18,6  |

Ørrethydrolysatet viste lavest ABTS antioksidantkapasitet mens det ene torskhydrolysatet (torsk mar 200H<sub>2</sub>O) viste høyest ABTS-antioksidantkapasitet. Generelt så viser torskhydrolysatene høyere ABTS antioksidantkapasitet enn de andre hydrolysatene, med unntak av laksehydrolysatet (6 måneder lagring i romtemperatur). ABTS-resultatene for seihydrolysatene i Pettersen sin prosjektoppgave ble brukt i dataanalysen i denne masteroppgaven. Pettersen sine resultater viser en vesentlig høyere ABTS-antioksidantkapasitet for sei sammenlignet med de andre fisketyperne, med resultater mellom 60.4-65.8  $\mu\text{mol PG/g}$  hydrolysat.[51]

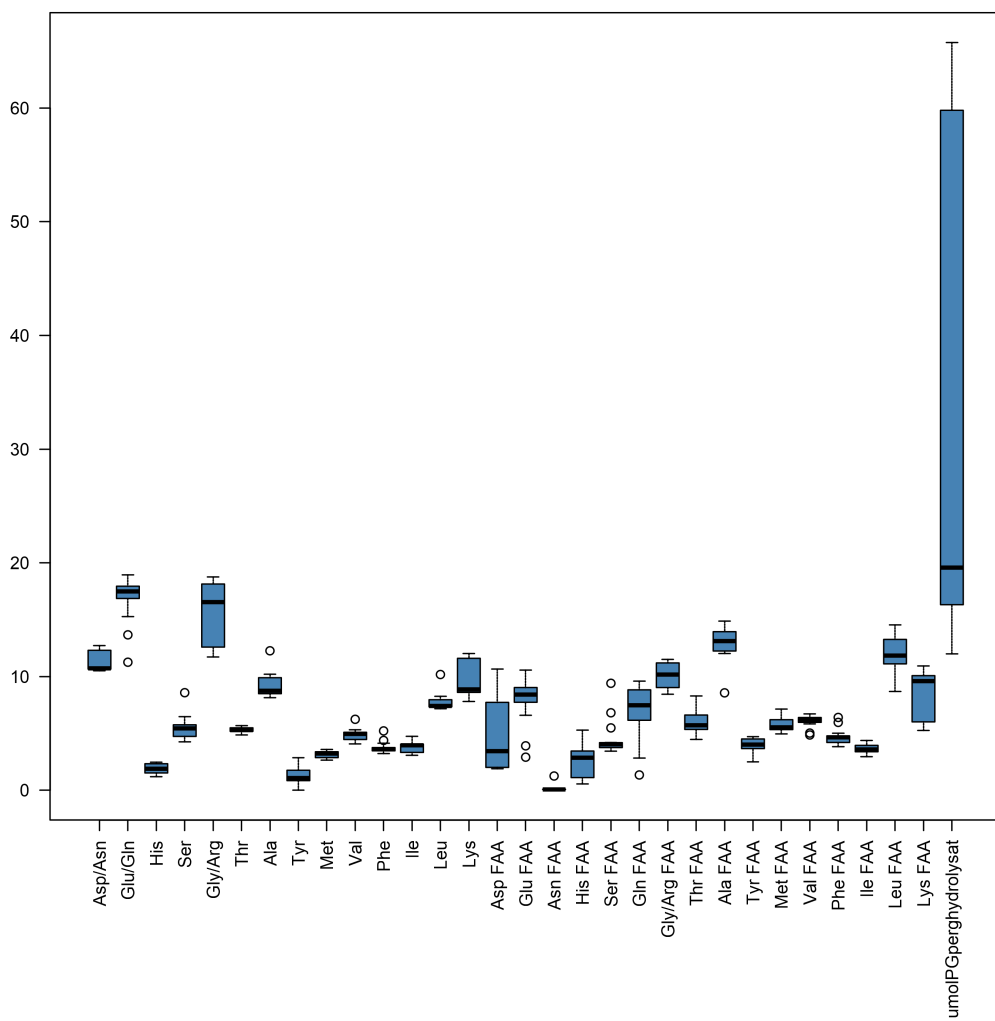
## 3.2 Matrise 1: TAA, FAA, ABTS antioksidantkapasitet

### 3.2.1 Oversikt over dataene: spredning og korrelasjon

Matrise 1 som ble brukt i dataanalysen består av variablene: total aminosyresammensetning (%), mengde frie aminosyrer (% FAA) og ABTS antioksidantkapasitet ( $\mu\text{mol PG/g}$  hydrolysat). Matrisen inkluderer følgende fiskehydrolysater: laks 0/1/2/4/6 mnd, ørret 6 mnd, sei okt/jan 1/2 og torsk feb/mar 400H<sub>2</sub>O/200H<sub>2</sub>O/200sH<sub>2</sub>O.

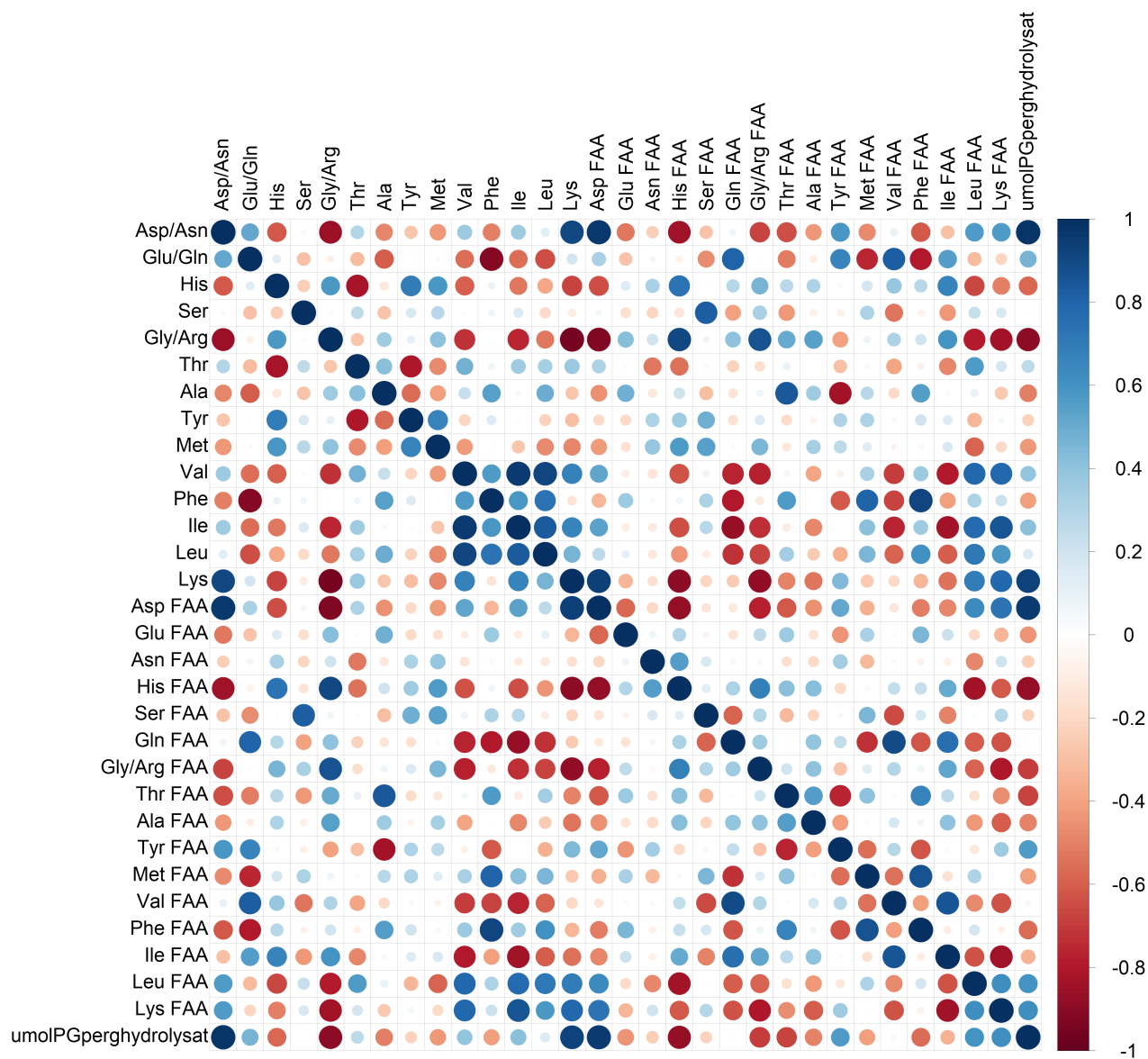
Prøvenavnene for laks og ørret viser til antall måneder lagringsperiode i romtemperatur. Prøvenavnene for sei viser til ulike fangstmåneder og to parallelle hydrolyseprosesser. For torsk viser prøvenavnene også til ulike fangstmåneder, men også til mengde vann (kg) brukt i hydrolyseprosessene. For et av torskehydrolysatene ble limvann fra produksjon av fiskeolje brukt istedenfor rent vann, og dette er indikert med “sH2O” [52], [53].

Figur 3.1 viser et boksplokk av variablene i matrise 1 før standardisering. Fra figuren er det tydelig at X-variablene er på samme skala, og at det er y-variabelen umol PG/g hydrolysat (ABTS antioksidantkapasitet) som skiller seg ut. Boksplokket viser også en asymmetrisk fordeling for ABTS antioksidantkapasitet, der minst 50% av observasjonene har verdier under 20 umol PG/g hydrolysat. Blant X-variablene varierer observasjonene i ulik grad, men flere uteliggere er indikert som punkter i figuren. Boksplokket for de standardiserte variablene er vist i appendiks i figur B.5. Dette viser at både Y og X-variabler var på samme skala etter standardisering, og varierte i området [-2,2] med unntak av noen uteliggere.



Figur 3.1: Boksplokk av matrise 1 som ble brukt i dataanalyse.

Pearson korrelasjonskoeffisient ble beregnet for hvert par av variabler. Figur 3.2 viser resulterende korrelasjonsmatrise for matrise 1, hvor fargestyrken er et mål på grad av korrelasjon. Korrelasjonsmatrisen viser at det er flere sterke korrelasjoner i datasettet, indikert ved sterk blå eller rød farge. Dette er tilfellet både mellom noen aminosyrer og ABTS antioksidantkapasitet, og mellom aminosyrer. Dette bekrefter at PLSR er en godt egnet metode for dette datasettet, ettersom det er en metode som kan håndtere både mange og korrelerende variabler [38].



Figur 3.2: Pearson korrelasjon i form av en korrelasjonsmatrise for matrise 1.

Korrelasjonsmatrisen viser en relativt sterk, positiv korrelasjon mellom umol PG/g hydrolysat og flere x-variabler: Asp/Asn, Asp FAA og Lys. Den viser også en relativt sterk, negativ korrelasjon mellom umol PG/g hydrolysat og x-variablene Gly/Arg og His FAA. Dette betyr at det er gjort funn av vesentlige sammenhenger. Korrelasjon betyr ikke nødvendigvis at det er en årsakssammenheng mellom variabler. Spesielt i datasett med

mange variabler er risikoen større for tilfeldige korrelasjoner. [44] Dette kan være tilfellet for denne matrisen. Korrelasjonene kan også være misvisende fordi det kan være andre underliggende faktorer som direkte påvirker antioksidantkapasitet. Korrelasjon alene er derfor ikke nok til å kunne forklare struktur-aktivitet sammenhenger. Funnene av sterke korrelasjoner er videre diskutert i kapittel 3.2.2. Delvis minste-kvadraters regresjon (PLSR) i lys av funnene i PLSR-analysen.

### 3.2.2 Delvis minste-kvadraters regresjon (PLSR)

Valget av antall PLS komponenter var basert på både resultatet av kryssvalideringen og forklart y-varians for den gjennomsnittlige, kryssvaliderte PLSR-modellen. Dette var for å velge den modellen som best beskrev dataene og dermed hadde den beste prediktive evnen, samtidig som at modellen også skulle beskrive en vesentlig del av variansen i antioksidantkapasitet (umol PG/g hydrolysat). % forklart y-varians er i stor grad også avhengig av uteliggere, og vil derfor alene ikke være nok til å bestemme antall komponenter [44]. Resultatet av kryssvalideringen er vist gjennom et valideringsplott av MSEP mot antall komponenter, se figur B.7 i appendiks. En oversikt over PLSR-modellen er også vist i figur B.6, her er både RMSEP og % forklart y-varians inkludert. Det er en betydelig nedgang i MSEP allerede etter to komponenter før den stabiliseres etter tre komponenter. % forklart y-varians er ganske høy kun etter en komponent, men øker jevnt før den også stabiliseres ved omtrent tre komponenter. Tre PLS komponenter ble derfor valgt til modellen. Den resulterende PLSR-modellen for matrise 1 er derfor den gjennomsnittlige, kryssvaliderte modellen med tre komponenter.

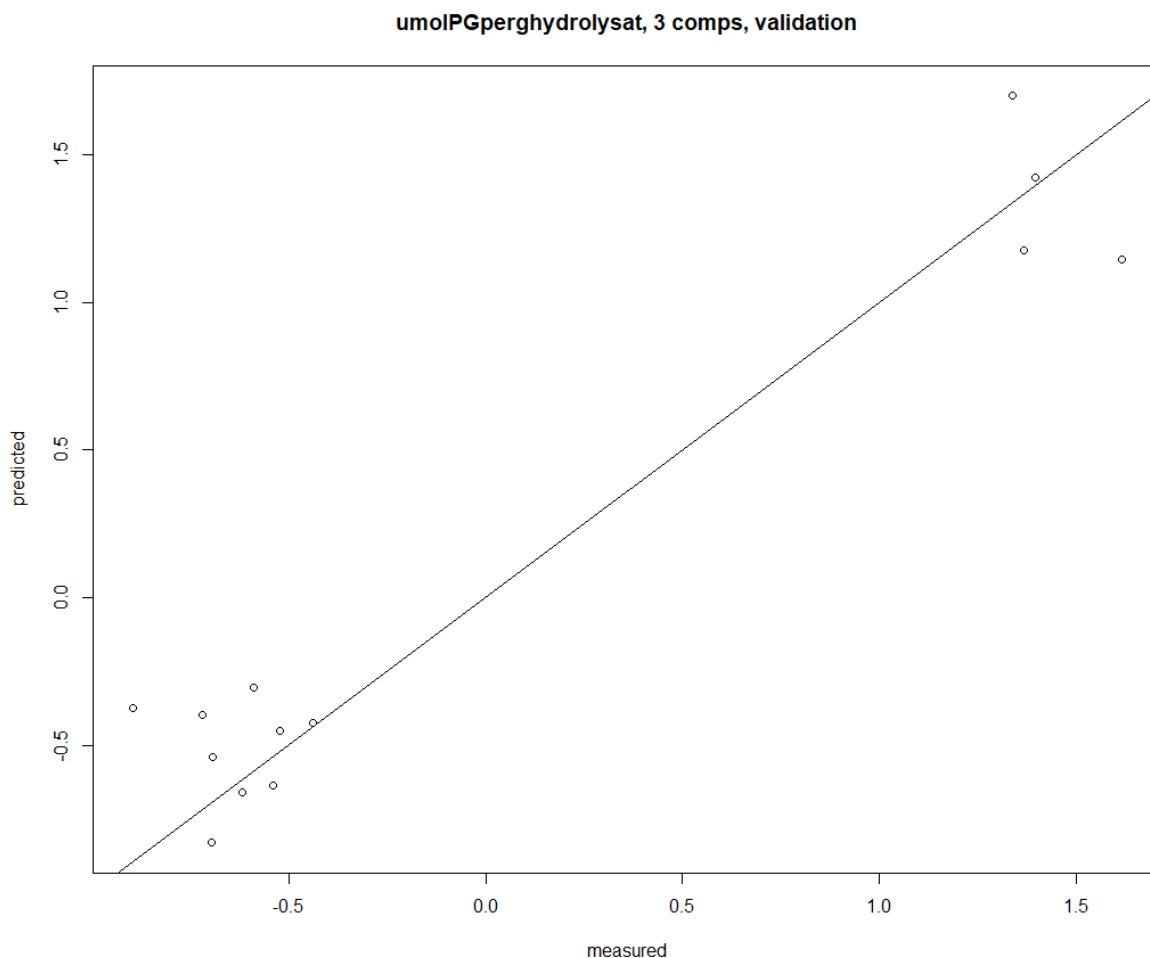
Det ble også generert et plott over X-og Y-scores for å se om det burde gjøres noen endringer for å få en bedre modell, i form av enten logaritmetransformering av datainput eller tillegging av kvadratiske ledd [38]. Plottet for den kryssvaliderte modellen med tre komponenter er vist i appendiks i figur B.8. Det viste ingen klar ulinearitet, og heller ingen klare uteliggere. Av den grunn ble det ikke gjort noen endringer, og det er den gjennomsnittlige, kryssvaliderte modellen som ble beskrevet i forrige avsnitt som ble brukt videre i analysen. Tabell 3.2 viser forklart x og y-varians inkludert RMSEP for den kryssvaliderte PLSR-modellen med tre komponenter.

Tabell 3.2: RMSEP og forklart x-og y-varians for den kryssvaliderte PLSR-modellen med 3 komponenter for matrise 1.

| Komponenter | RMSEP  | Forklart y-varians (%) | Forklart x-varians (%) |
|-------------|--------|------------------------|------------------------|
| 3           | 0.2639 | 98.23                  | 69.39                  |

Figur 3.3 viser de kryssvaliderte prediksjonene for ABTS antioksidantkapasitet for prøvene i en modell med tre komponenter. Prediksjonene viser ABTS antioksidantkapasitet på en standardisert form ettersom responsvariabelen også ble standardisert før PLSR-analysen. Figuren viser at det ikke er store avvik mellom målte og predikerte

verdier, noe som også er reflektert av en relativt god MSEP for modellen. Dette betyr at modellen har en god evne til å beskrive dataene. Figuren viser også to distinkte klynger, og dette er som forventet ettersom ABTS antioksidantkapasitet for prøvene er på to ulike nivåer. Modellen vil i stor grad prøve å beskrive variasjonene i y, og det er tydelig at dette i stor grad er relatert til forskjellen mellom disse to klyngene. Det er dermed også fornuftig at forklart y-varians var relativt høy (92.96%) allerede etter en PLS komponent.



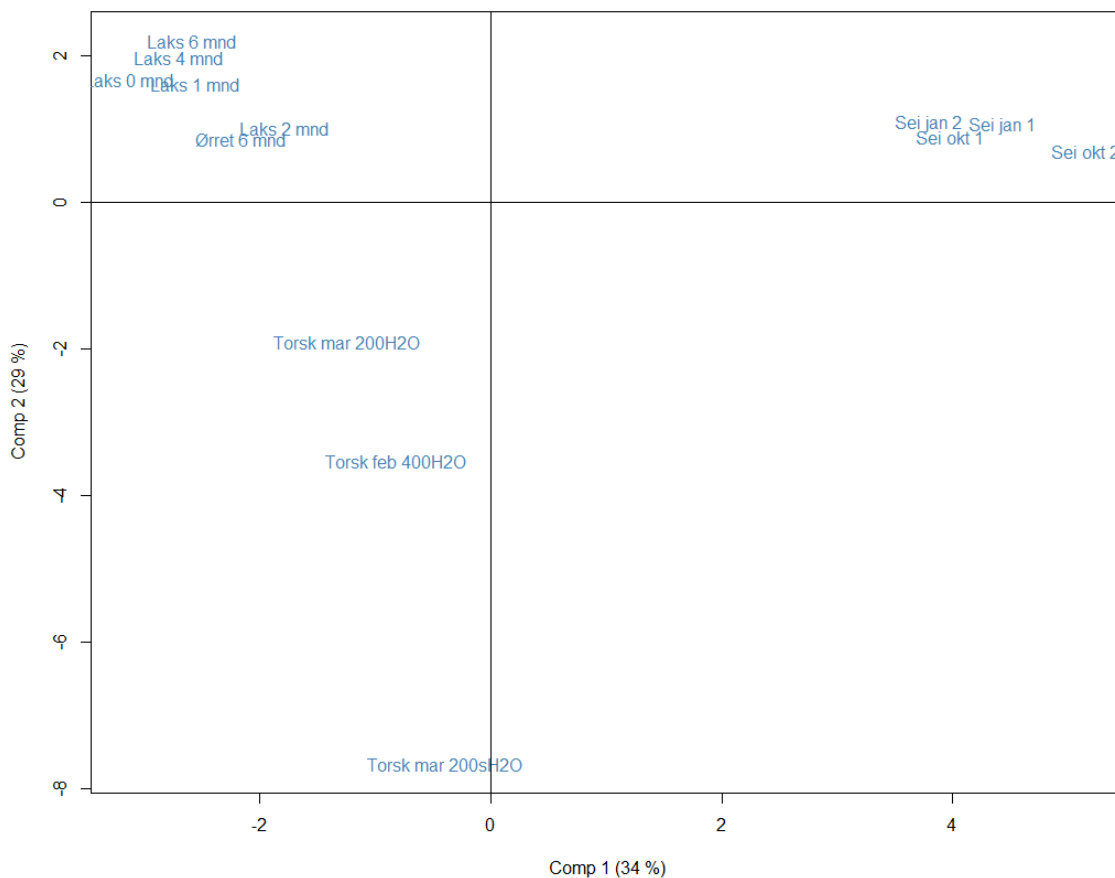
Figur 3.3: Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 1.

Modellen er likevel begrenset til relativt få prøver. Det kan være usikkerhet knyttet til om de karakteristiske klyngene oppstår på grunn av et begrenset datagrunnlag, eller om det faktisk er forskjeller i ulike typer fisk som gir definerte klynger. Flere av prøvene i matrisen er også fra samme hydrolysat, eller samme prøve fisk/fraksjon. Generelt så vil et større datagrunnlag gi en mer robust modell, og gi svar på om prøvene er representative for variasjon blant fisketyper og/eller fraksjoner. Med flere uavhengige prøver vil MSEP også i større grad kunne reflektere X sin reproduserbare evne til å predikere Y for helt nye prøver [47]. Om klyngene er tilfeldige så kan det være usikkerhet knyttet til hvor godt modellen beskriver annen varians i ABTS-antioksidantkapasitet, og om den da har

evnen til å predikere ABTS-antioksidantkapasitet for prøver hvor verdiene faller mellom de to nivåene.

Selv om resultatene av kryssvalideringen viser en modell med relativt god prediksjonsevne, så er MSEP også kun et estimat på test MSE [42]. Den beste valideringen vil alltid være å bruke modellen på et uavhengig valideringssett. Kryssvalidering brukes når store, uavhengige testsett ikke er tilgjengelig for ytterligere validering. [38] Oppsummert så ser det ut som om modellen beskriver dataene godt, men at det er noen begrensninger relatert til størrelsen på datagrunnlaget. Enda flere, uavhengige prøver kan brukes til å videre validere, avkrefte eller bekrefte funnene i modellen.

Figur 3.4 viser hydrolysatenes *scores* for de to første komponentene. Resultatene av PLSR modellen viste at % forklart y-varians var 92.96 % etter kun en komponent, og 96.48% etter to komponenter. Dette betyr at en betydelig andel av variansen i ABTS antioksidantkapasitet (umol PG/g hydrolysat) er forklart av de første to komponentene. Et *scores*-plott med de første komponentene vil derfor være nok til å forklare hovedtrendene i matrise 1. En fullstendig oversikt over *scores* for alle tre komponentene er vist i appendiks.



Figur 3.4: *Scores*-plott for de to første PLS komponentene fra den kryssvaliderte PLSR-modellen av matrise 1.

*Scores*-plott viser en oversikt over prøvene i PLSR-modellen, og vil derfor vise mønster og ulikheter i datasett. *Scores* tilsvarer også de nye koordinatene for X som predikerer Y, i et hyperplan definert av PLS komponentene. [38] Figur 3.4 viser at prøvene grupperer seg etter type fisk, som er forventet ettersom flere av prøvene er fra samme fisk og fraksjon. Det viser også at med tanke på aminosyresammensetning og ABTS-antioksidantkapasitet så er det små forskjeller mellom prøver av samme type fisk. Selv om det også er en form for gruppering av torskhydrolysatene, så er det større spredning blant disse. Dette kan være fordi de tre torskprøvene er fra tre ulike hydrolyseprosesser med variasjoner i hydrolyseparametere som reflekteres i X (aminosyresammensetning) og Y (ABTS antioksidantkapasitet) [52], [53]. Det kan se ut som torskhydrolysatet fra hydrolyseprosessen med limvann (Torsk mar200sH2O) skiller seg noe ut. Limvann vil tilføre blant annet nye proteiner og peptider i tillegg til råstoffet i en hydrolyseprosess, og det er derfor ikke uventet at dette påvirker total aminosyresammensetning og mengde frie aminosyrer. Monslaup viste blant annet at mengde vannløselige proteiner var vesentlig mindre i denne prøven enn i de andre [52]. Laksehydrolysatene er en av de klare grupperingene, noe som indikerer at variasjonene i lagringsperioder i romtemperatur hverken utgjør vesentlige forskjeller i aminosyresammensetningen eller i modelleringen av ABTS antioksidantkapasitet fra X.

Figuren viser også at laks og ørret grupperer seg. Generelt så har laks og ørret et større fettinnhold enn torsk og sei [59]. Dette vil ikke kunne forklare denne grupperingen av flere grunner. Den ene grunnen er at prøvene er proteinhydrolysater, og lipider har i stor grad blitt fjernet etter sentrifugering [11], [49], [50]. Dette er likevel ikke alltid tilfellet, og for fetere fisker kan dette kreve flere runder med sentrifugering [11]. Den andre grunnen er at modellen beskriver aminosyresammensetning (TAA og FAA) og basert på dette predikerer ABTS-antioksidantkapasitet. Proteininnholdet i rå laks, ørret, sei og torsk varierer mellom 16.5-20 g/100 g fisk, og er derfor relativt stabilt [59]. Mengde proteiner skal derfor ha liten betydning på variasjon mellom hydrolysatene. Det kan være andre underliggende faktorer som har betydning for aminosyreprofilene til de ulike fisketypene og/eller fiskefraksjonene. Dette kan for eksempel være andel ulike typer proteiner og dermed også peptider og aminosyrer. Både mengde av strukturelle muskelproteiner og bindevevsproteiner varierer mellom ulike typer fisk [24]. Grupperingen kan også skyldes at både ørret og laksehydrolysatene er fra hoder, i motsetning til sei som er fra ryggbein. Torskhydrolysatene er også fra hoder, men som nevnt over kan de ulike hydrolysebetingelsene være en grunn til at torsk skiller seg ut fra resten av fisketypene der samme ratio vann/råstoff ble brukt i hydrolyseprosessene [49], [50]. Torskhydrolysatene ble også laget i semiindustriell skala i motsetning til laks og ørret. [49], [50], [52], [53]

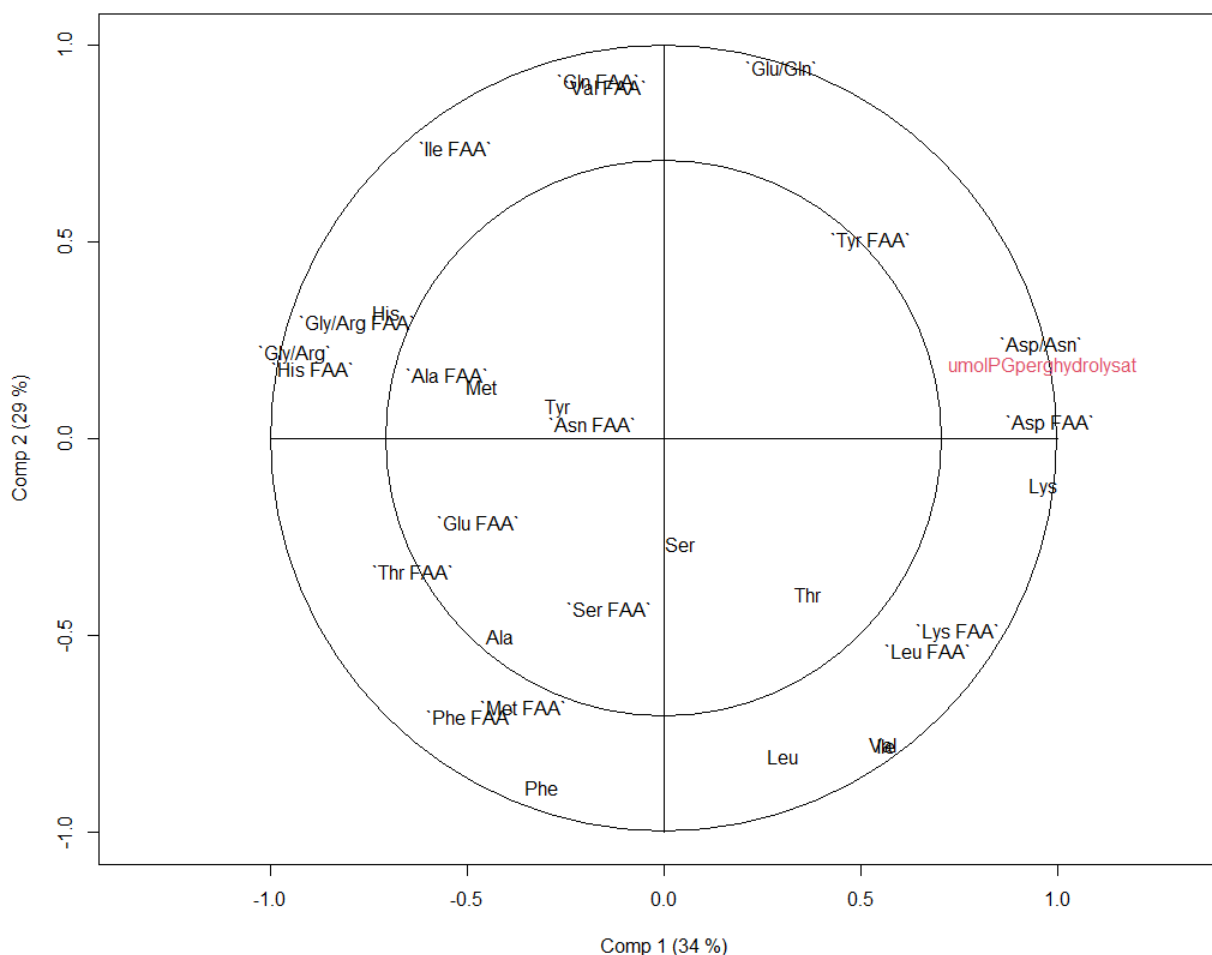
*Scores*-plottene for alle de tre komponentene er vist i appendiks i figur B.9. Plottet for komponent 1:3 og 2:3 viser ikke like tydelige grupperinger, men det kan se ut som ørret skiller seg ut i retning komponent tre. Komponent tre beskriver kun en liten del av variansen i ABTS antioksidantkapasitet, men det kan likevel være viktig

informasjon i denne variansen.

Figur 3.5 viser korrelasjons-*loadings* for de standardiserte variablene og de to første komponentene. En fullstendig oversikt over korrelasjon-*loadings* for alle tre komponentene er vist i appendiks. Korrelasjon-*loadings* er relatert til *loadings*, og reflekterer korrelasjonskoeffisienten mellom en variabel og en komponent [47]. Både korrelasjon-*loadings* og *loadings* viser hvilke variabler som korrelerer, men korrelasjon-*loadings* vil også indikere hvor godt forklart variablene er av komponentene i modellen [44]. Sirklene i figuren indikerer 50% og 100% forklart varians [47]. Dette vil si at variablene med størst betydning for hver komponent er plassert nær den ytterste sirkelen og samtidig lengst unna origo langs aksene for hver komponent. I denne masteroppgaven ble både korrelasjon-*loadings* og *loadings*-plott generert. Resultatene er angitt for korrelasjon-*loadings*, ettersom det kun er en skalert versjon av *loadings* og vil vise samme informasjon.

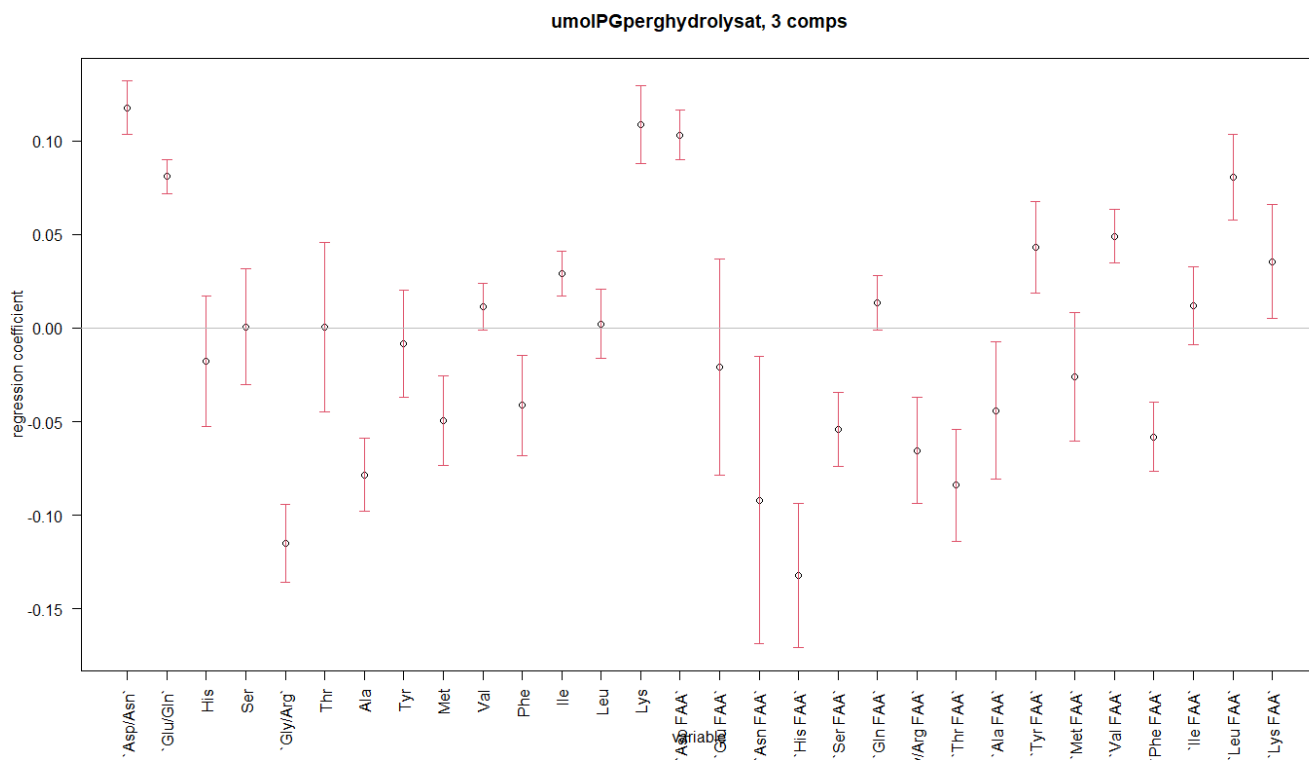
Figuren viser at det er en betydelig andel variabler som er opp mot 100% forklart av de første to komponentene med unntak av variabler som Tyr, Asn FAA, Ser, Thr, Ser FAA og Glu FAA. Figuren viser også at variablene som har størst betydning for den første komponenten er Gly/Arg, His, Asp/Asn, Lys og Gly/Arg FAA, His FAA, Thr FAA, Leu FAA, Lys FAA og Asp FAA. Den viser også at ABTS antioksidantkapasitet (umol PG/g hydrolysat) er godt modellert av den første komponenten. Figur 3.5 og B.10 (med alle tre komponentene) viser at Y ikke er spesielt godt modellert av hverken komponent 2 eller 3. Dette er som forventet ettersom forklart y-variens var spesielt høy kun etter en komponent.





Figur 3.5: Korrelasjons-loadings plott for de to første PLS komponentene fra den kryssvaliderte PLSR-modellen av matrise 1.

For at korrelasjonene i korrelasjons-loadings-plottet skal ha betydning må variablene være godt forklart av modellen [44]. Figur 3.5 indikerer en sterk, positiv korrelasjon mellom ABTS antioksidantkapasitet og henholdsvis Asp/Asn, Asp FAA og Lys. Figuren viser også en sterk, negativ korrelasjon mellom ABTS antioksidantkapasitet og henholdsvis Gly/Arg, Gly/Arg FAA, His og His FAA. Alle disse variablene er nært 100% forklart av de første to komponentene, og har spesielt stor betydning for den første komponenten som også er den komponenten som best modellerer ABTS antioksidantkapasitet. Dette indikerer at det er X-variabler som er viktige i modelleringen av Y. Regresjonskoeffisientene er også et mål på dette, der X-variablenes innflytelse på Y er reflektert gjennom store PLS-regresjonskoeffisienter [38]. Regresjonskoeffisientene for den kryssvaliderte PLSR-modellen med tre komponenter er vist i figur 3.6.



Figur 3.6: Regresjonskoeffisienter for de ulike variablene i den kryssvaliderte modellen av matrise 1 for 3 komponenter

Figur 3.6 viser regresjonskoeffisientene for alle X-variablene inkludert estimert standardavvik fra *jack-knifing* av koeffisientene. Estimert standardavvik er et mål på variasjon mellom regresjonskoeffisientene i de ulike delmodellene som ble generert under kryssvalideringen for å beregne MSEF [38]. For noen av variablene er de estimerte standardavvikene relativt store, og av den grunn er det knyttet stor usikkerhet til hvilken effekt variablene har på ABTS antioksidantkapasitet. Store standardavvik for en spesifikk variabel indikerer at det er uteliggende verdier av den variabelen for en eller flere av hydrolysatene. Plottet viser derimot mye av det samme som korrelasjon-loadings-plottet, der følgende variabler er spesielt viktige for modelleringen av ABTS antioksidantkapasitet: Asp/Asn, Glu/Gln, Gly/Arg, Lys, Asp FAA og His FAA. Disse variablene har relativt små standardavvik med unntak av His FAA. Koeffisienten for His FAA inkludert standardavvik er likevel av vesentlig størrelse.

Funksjonen “jack.test” gjør en “grov” signifikanstest (t-test) av regresjonskoeffisientene ved valgt antall komponenter basert på usikkerhetsvarians i koeffisientene. Usikkerhetsvariansen er basert på differansen mellom koeffisientestimatene for en modell med alle prøvene og koeffisientestimatene for delmodellene i kryssvalideringen, der en og en prøve har blitt utelatt fra modellen. På grunn av begrensningene med denne t-testen så er den kun en indikator på *non-significance*. [47], [58] Testen indikerer at de overnevnte variablene med vesentlige regresjonskoeffisienter er signifikante, se vedlagt resultat i appendiks. Om ikke annet er beskrevet for de neste modellene, så

betyr det at de vesentlige koeffisientene i beskrevet modell ser ut til å være signifikante.

ABTS antioksidantkapasitet (umol PG/g hydrolysat) er et mål på evnen til å nøytralisere en ABTS radikal. Dette kan skje enten ved ET-basert reduksjon av radikalen, eller ved en HAT-basert mekanisme hvor radikalen mottar et hydrogenatom.[33] Reduksjon av ABTS radikalen er vist gjennom en fargeendring, der ekvivalent reduksjon av standardmolekylet PG er et mål på reduksjon av ABTS radikalen. [32], [33] Funnene over kan derfor indikere hvilke X-variabler som kan ha betydning for antioksidative peptider i form av evne til å nøytralisere ABTS radikaler. Både % Asp/Asn, %Gly/Arg, %Glu/Gln og %Lys ser ut til å ha relativt stor betydning for modelleringen av umol PG/g hydrolysat. ABTS-analysen skjer i et nøytralt miljø, og derfor er det også naturlig å se på aminosyrekarakteristikk i samme miljø [33]. Det er ikke mulig å skille Asp og Asn på grunn av bruk av syrehydrolyse i analysemetoden, som gjør at Asn omdannes til Asp. Dette er også tilfellet for Glu og Gln, som også ser ut til å ha en viss innflytelse på modellen. [56] De aminosyrene som er detektert sammen deler ikke samme aminosyrekarakteristikk, noe som vil skape en usikkerhet rundt hvilke aminosyrer som potensielt bidrar til ABTS antioksidantkapasitet. Felles for Glu, Asp, Arg og Lys er at dette er hydrofile aminosyrer med en ladet sidekjede ved nøytral pH. Dette gjør også disse aminosyrene til potente protondonorer eller akseptorer. [18] Hydrogendonorer er potente antioksidanter gjennom HAT-mekanismen [33]. Det er også vist at ET-mekanismen er avhengig av deprotonering og ioniseringspotensiale. Ioniseringsenergien reduseres med økt pH og deprotonering. [33] Dette kan indikere at de negativt ladede sidekjedene til Asp og Glu ved nøytral pH gjør disse aminosyrene til mer potente elektrondonorer.

Litteraturstudien fra prosjektoppgaven viste flere funn relatert til ladede aminosyrer [20]. Blant funnene var positiv korrelasjon mellom Arg og ABTS radikal fjerningsevne, i tillegg til en positiv korrelasjon mellom inhibering av lipidoksidasjon og henholdsvis Glu, His og Glu+His. [60] Et annet funn viste også til en betydelig andel Arg i fraksjonen med størst reduserende evne [61]. I motsetning til disse funnene er Gly/Arg negativt korrelert med antioksidantkapasitet i denne datamodellen. Det betyr at selv om Gly/Arg har stor innflytelse på modelleringen av Y, så reflekterer det ikke et positivt bidrag til ABTS antioksidantkapasitet. Lys derimot som også har en positiv sidekjede, ser ut til å bidra positivt til ABTS antioksidantkapasitet i denne modellen [18]. Damgaard et al. (2015) begrunner den positive korrelasjonen med Glu med aminosyrens evne til å binde metaller, noe negativt ladede aminosyrer kan gjøre gjennom elektrostatiske interaksjoner[28], [60]. Negativt ladede aminosyrer kan derfor også bidra til antioksidantkapasitet gjennom metallbinding, men denne mekanismen kan ikke måles gjennom ABTS-analysen og vil derfor ikke ha betydning for resultatene fra datamodellen. Likevel så kan funnene indikere at ladede aminosyrene potensielt kan reagere med ABTS-radikalen som proton-eller elektrondonorer, og bidra til bedre ABTS antioksidantkapasitet. Dette er videre studert i en modell med matrise 5 der variablene består av

aminosyredeskriptorer, der blant annet ladede aminosyrer er inkludert. Dette er nærmere beskrevet i delkapittel 3.3.

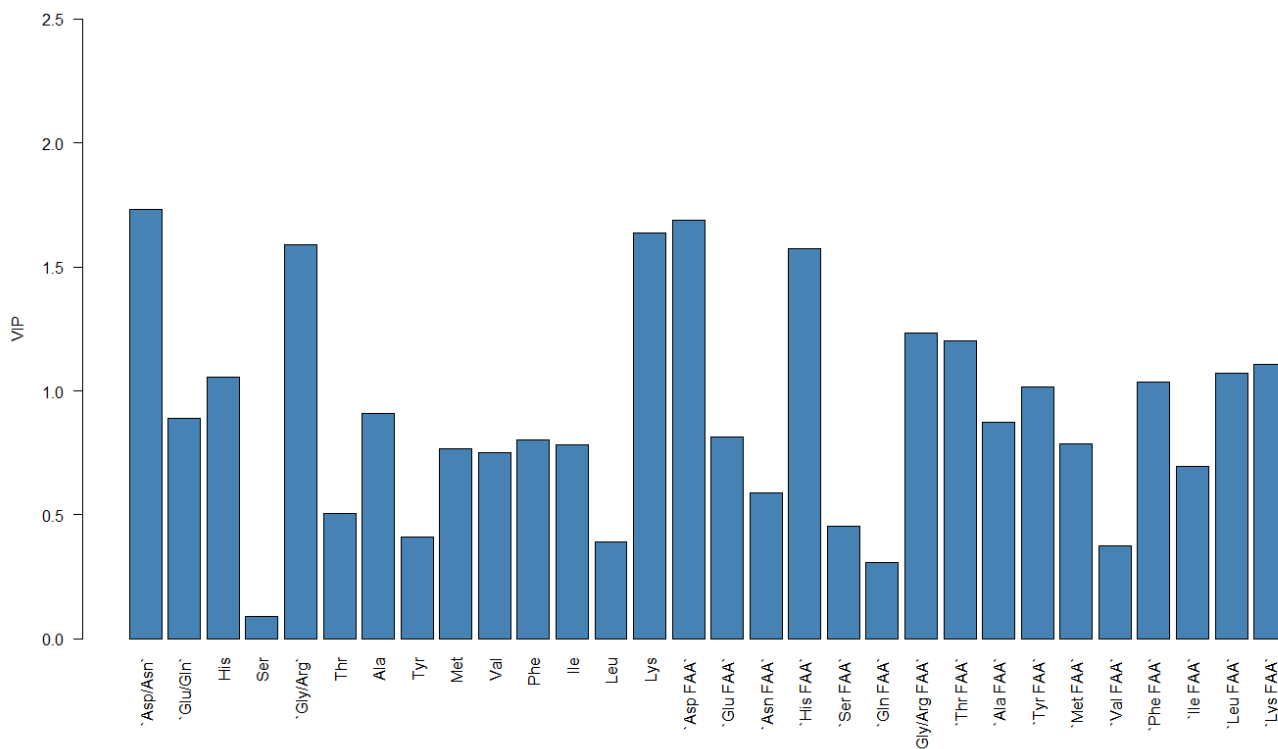
Fra regresjonskoeffisientene ser det også ut til at His FAA har en god innflytelse på modelleringen av ABTS antioksidantkapasitet. His har en aromatisk ring som lett kan oksideres og reagere med frie radikaler [28]. Damgaard et al. fant en positiv korrelasjon mellom %His og inhibering av lipidoksidasjon, og viste til His sin evne til å binde metaller og donere hydrogenatomer [20], [60]. Datamodellen viser derimot at det er en negativ korrelasjon mellom His FAA og ABTS antioksidantkapasitet. Dette indikerer at His har en negativ effekt på ABTS antioksidantkapasitet i motsetning til resultatene fra Damgaard et al. Dette trenger ikke å bety at His ikke har antioksidative egenskaper og kan nøytralisere radikaler. Hydrolysater er komplekse blandinger med mange peptider av ulike lengde og frie aminosyrer. Underliggende faktorer som interaksjoner mellom ulike peptider og aminosyrer kan gjøre noen aminosyrer utilgjengelige i reaksjon med frie radikaler. Blant annet kan proteinradikaler videre oksidere andre aminosyrer eller reagere med andre proteinradikaler [29]. Det er derfor også interessant at His FAA og His har så ulike regresjonskoeffisienter, der His FAA er vesentlig viktigere for modelleringen av ABTS antioksidantkapasitet.

Aromatiske og nukleofile aminosyrer har større oksidativt potensiale enn andre aminosyrer. Met og Cys har en SH-gruppe som fungerer som en god hydrogendonor, og Phe, Tyr, Trp og His har aromatiske ringer som kan oksideres gjennom elektrononasjon. Det er også vist at Met kan donere elektroner gjennom sulfidgruppen [28] Dette betyr at disse aminosyrene er spesielt reaktive og kan potensielt nøytralisere radikaler. Cys og Trp ble ikke detektert gjennom måleanalysen, og kunne derfor ikke brukes i modellen. Fra regresjonskoeffisientene er det tydelig at hverken His, Tyr, Phe eller Met hadde stor effekt på modelleringen av ABTS antioksidantkapasitet. Dette motstrider flere av funnene fra litteraturstudien i prosjektoppgaven som indikerte at aromatiske og nukleofile aminosyrer har betydning for radikal fjerningsevne [20] Damgaard et al. viste blant annet til positiv korrelasjon mellom Met og radikal fjerningsevne, men også mellom Tyr og ABTS radikal fjerningsevne [60]. Dette kan indikere at antioksidantkapasitet er mer kompleks enn hva som har blitt beskrevet i litteratur. Det vil også være noe måleusikkerhet på grunn av at ulike personer har gjort de ulike *in vitro*-analysene som er bakgrunn for alle variablene.

Det er en del usikkerhet knyttet til underliggende faktorer som interaksjoner mellom aminosyrer og peptider. Utfordringer knyttet til å beskrive peptidstruktur og interaksjoner i peptidblandinger er diskutert etter at alle modellene er presentert. Det er også andre faktorer som kan påvirke antioksidantkapasiteten til en peptidblanding. Ulike radikaler har ulike reaktivitet og spesifisitet, og i tilfeller med mer reaktive radikaler vil plassering av aminosyrer og peptider ha større betydning enn reaktiviteten til den enkelte aminosyren [28]. ABTS-analysen vil kun reflektere evnen til å nøytralisere ABTS-radikalen, som er en radikal som ikke eksisterer i biologiske systemer [33]. Usikkerhetene tilknyttet ABTS-analyse som mål på antioksidantkapasitet er også diskutert nærmere etter at alle modellene

er presentert.

VIP (*variable importance for the projection*) er også et mål på betydningen av en variabel for modelleringen, men denne parameteren viser til betydningen av X for både X og Y [38]. VIP vil i stor grad vise det samme som nevnt over, og vil derfor ikke gi mer informasjon om X-Y sammenhenger. VIP er vist for X-variablene i figur 3.7.



Figur 3.7: VIP-verdiene for variablene i den kryssvaliderte modellen av matrise 1 for 3 komponenter.

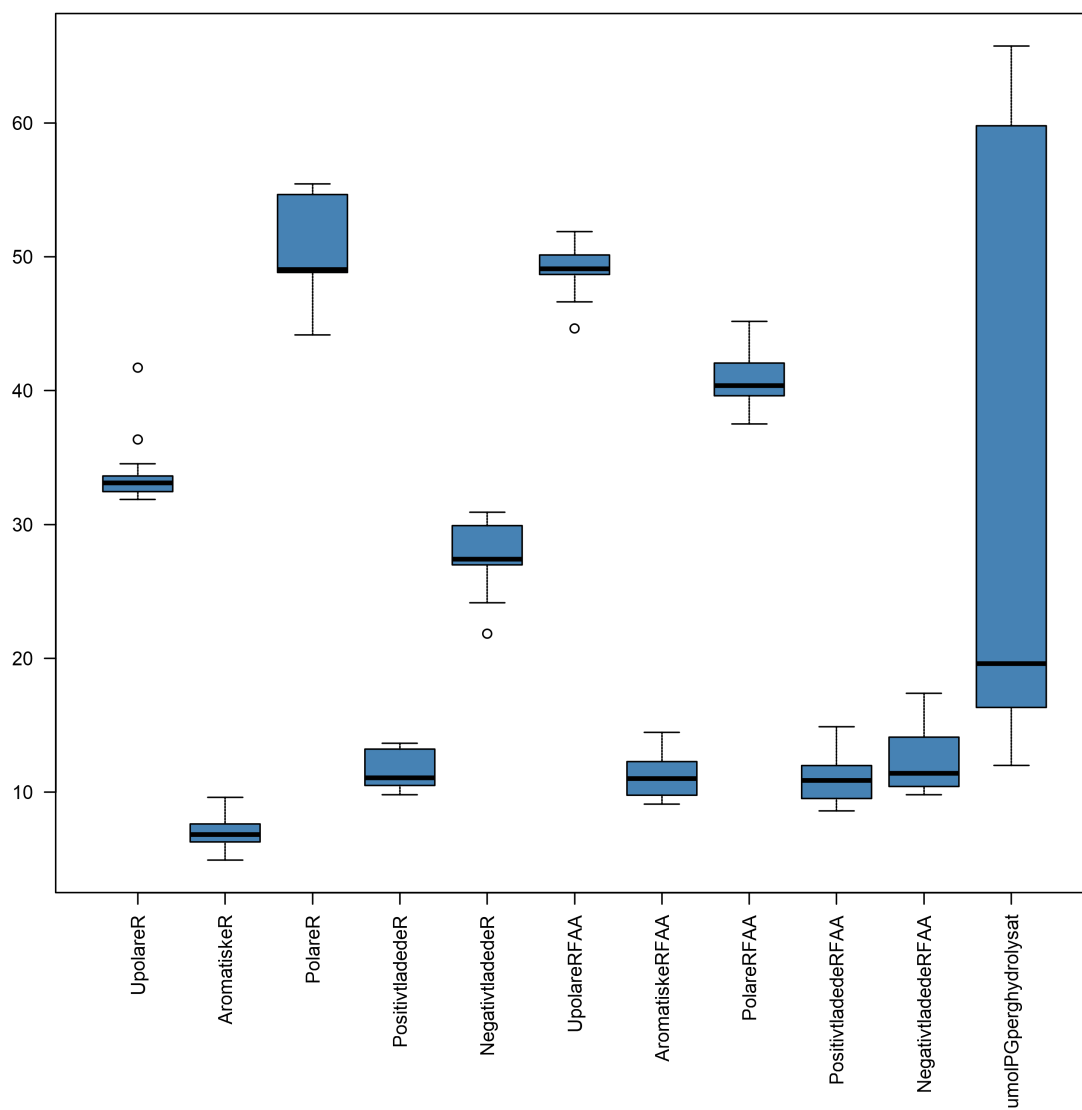
Resultatene fra modellene av de andre matrisene vil i de neste delkapitlene beskrives i korte trekk, før flere funn på proteinstruktur-antioksidantkapasitet skal presenteres og diskuteres.

### 3.3 Matrise 5: Aminosyredeskriptorer, ABTS antioksidantkapasitet

#### 3.3.1 Oversikt over dataene: Spredning og korrelasjon

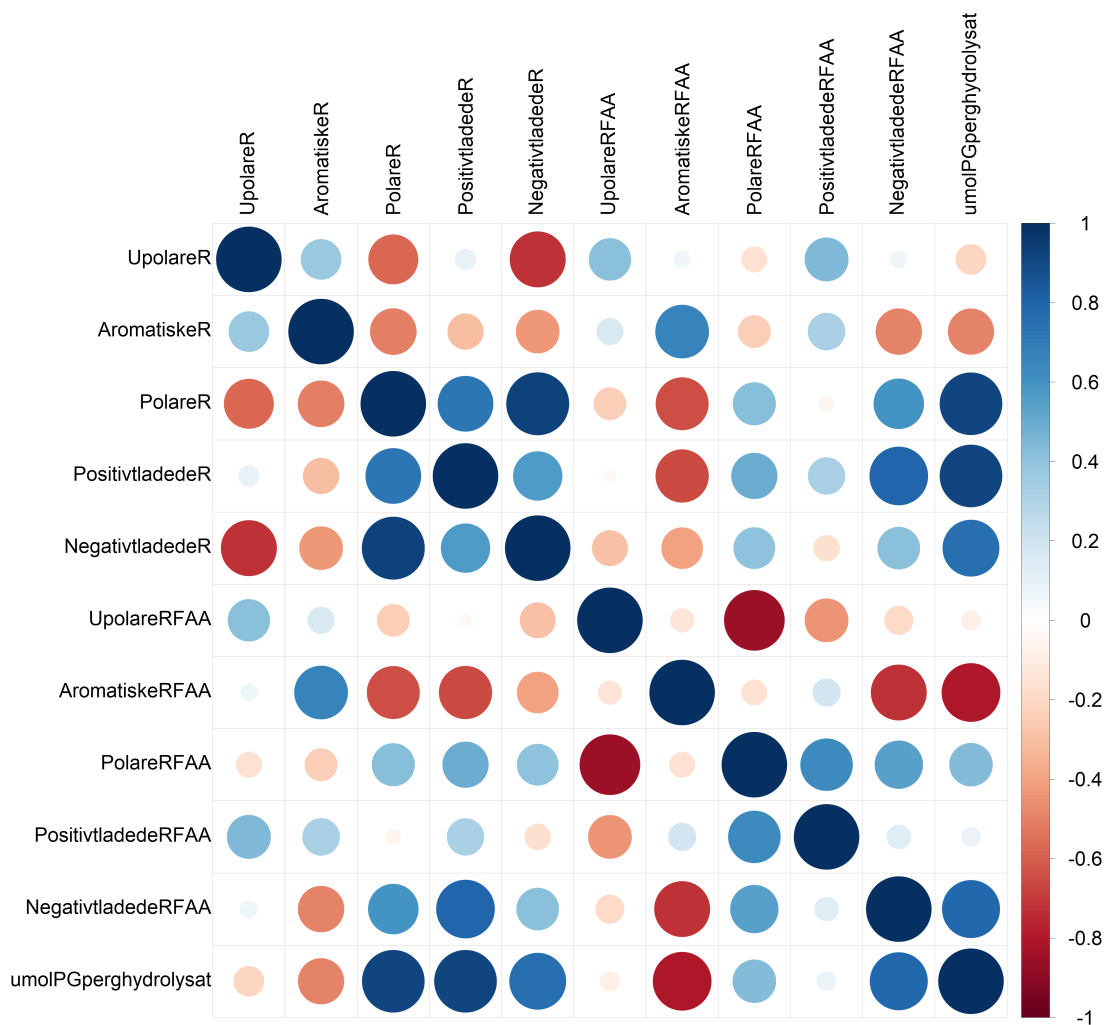
Matrise 5 består av variablene: Upolare R, Aromatiske R, Polare R, Positivt ladede R, Negativt ladede R, Upolare R FAA, Aromatiske R FAA, Polare R FAA, Positivt ladede R FAA, Negativt ladede R FAA og ABTS antioksidantkapasitet (umol PG/g hydrolysat). I denne matrisen så er aminosyrene fra total aminosyresammensetning (%) og mengde frie aminosyrer (%) i matrise 1 summert og inndelt etter kategorier basert på aminosyrekarakteristikk ved nøytral pH. Hydrolysatene i matrise 5 er derfor de samme som i matrise 1. Boksplottet i figur 3.8 viser at % av de ulike kategoriene varierer betraktelig, og at det kun er fire uteliggende observasjoner totalt fordelt på tre

aminosyrekategorier. Sammensetning av totale aminosyrer og frie aminosyrer varierer også for noen av kategoriene. Eksempelvis så er % negativt ladete R generelt vesentlig større enn % negativt ladete R FAA. Et boksplokk over de standardiserte variablene er vist i appendiks i figur B.12.



Figur 3.8: Boksplokk over de originale variablene i matrise 5.

Figur 3.9 viser korrelasjonsmatrisen for matrise 5. Den viser sterkt, positive korrelasjoner mellom umol PG/g hydrolysat og følgende x-variabler: polare R og positivt ladete R. Det er også en relativt sterk, negativ korrelasjon mellom ABTS antioksidantkapasitet og aromatiske R FAA. Det var sterke korrelasjoner også for matrise 1. I denne matrisen overlapper også aminosyrekategoriene til en viss grad, og på grunn av dette er det forventet at det er en del korrelasjoner i matrisen.



Figur 3.9: Pearson korrelasjon i form av en korrelasjonsmatrise for matrise 5.

### 3.3.2 Delvis minste-kvadraters regresjon (PLSR)

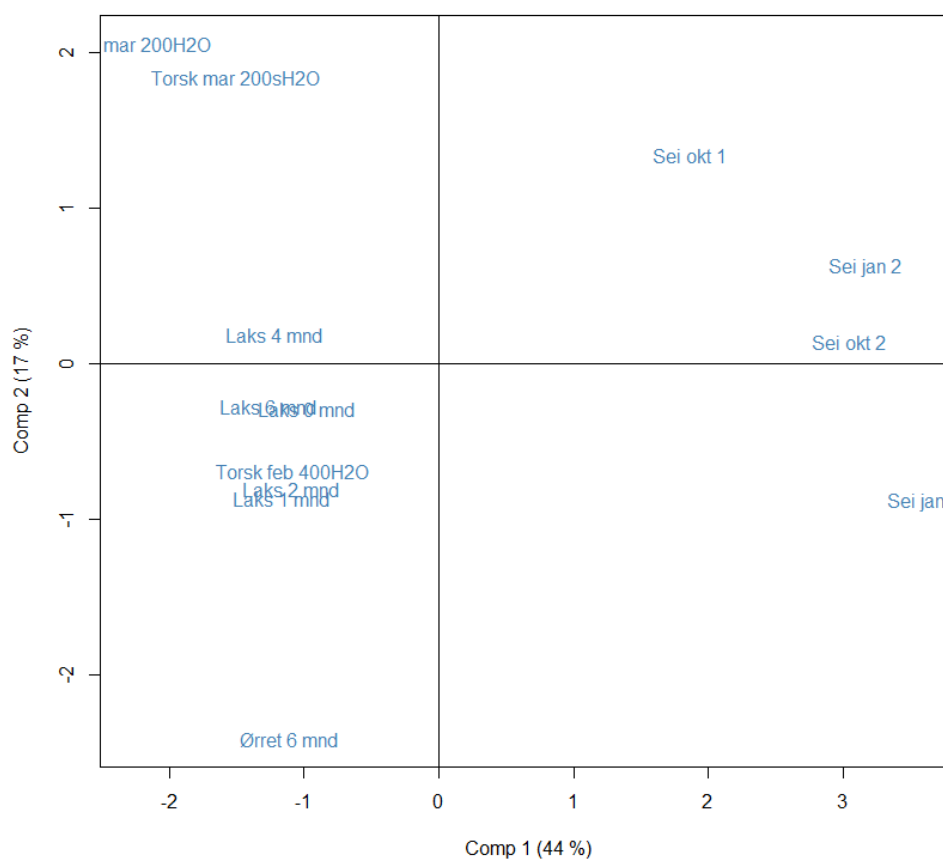
En gjennomsnittlig, kryssvalidert PLSR modell ble generert, og antall PLS komponenter ble så valgt på samme måte som for matrise 1. Et sammendrag av PLSR-modellen og tilhørende valideringsplott er vist i appendiks i figur B.13 og B.14. Det ble valgt å inkludere tre komponenter i modellen for å oppnå en lav MSEP, selv om forklart y-varians var høy allerede etter to komponenter. Dette var også tilfellet for matrise 1. Det var ingen klar ulinearitet eller uteliggere i plottet av X- og Y-scores, så derfor ble det heller ikke gjort noen endringer i modellen. Den kryssvaliderte PLSR-modellen med tre komponenter ble brukt videre i dataanalysen. Tabell 3.3 viser en oversikt over RMSEP og forklart x og y-varians for PLSR-modellen med tre komponenter.

Tabell 3.3: RMSEP, forklart x- og y-varians for den gjennomsnittlige, kryssvaliderte PLSR modellen med 3 komponenter for matrise 5.

| Komponenter | RMSEP  | Forklart y-varians (%) | Forklart x-varians (%) |
|-------------|--------|------------------------|------------------------|
| 3           | 0.2069 | 98.78                  | 72.51                  |

De kryssvaliderte prediksjonene for ABTS antioksidantkapasitet for prøvene i en modell med tre komponenter er vist i appendiks i figur B.16. Dette plottet viser mye av det samme som modellen for matrise 1. Det er relativt små avvik mellom predikert og målt umol PG/g hydrolysat (standardisert verdi), og det er også her to distinkte klynger med datapunkter. X-variablene i matrise 1 og matrise 5 er to varianter av de samme variablene, og av den grunn er det ikke uventet at modellene vise tilnærmet lik prediksjonsevne. Begrensninger i prøvegrunnlag og MSEF ble diskutert for matrise 1, men vil også gjelde her.

De to første komponentene beskriver 97.57% av variansen i umol PG/g hydrolysat. Derfor er både *scores*-og korrelasjon-loadings plottene vist for de to første komponentene i figur 3.10 og 3.11. Tilsvarende for alle tre komponentene i modellen er vist i appendiks i figur B.17 og B.18.



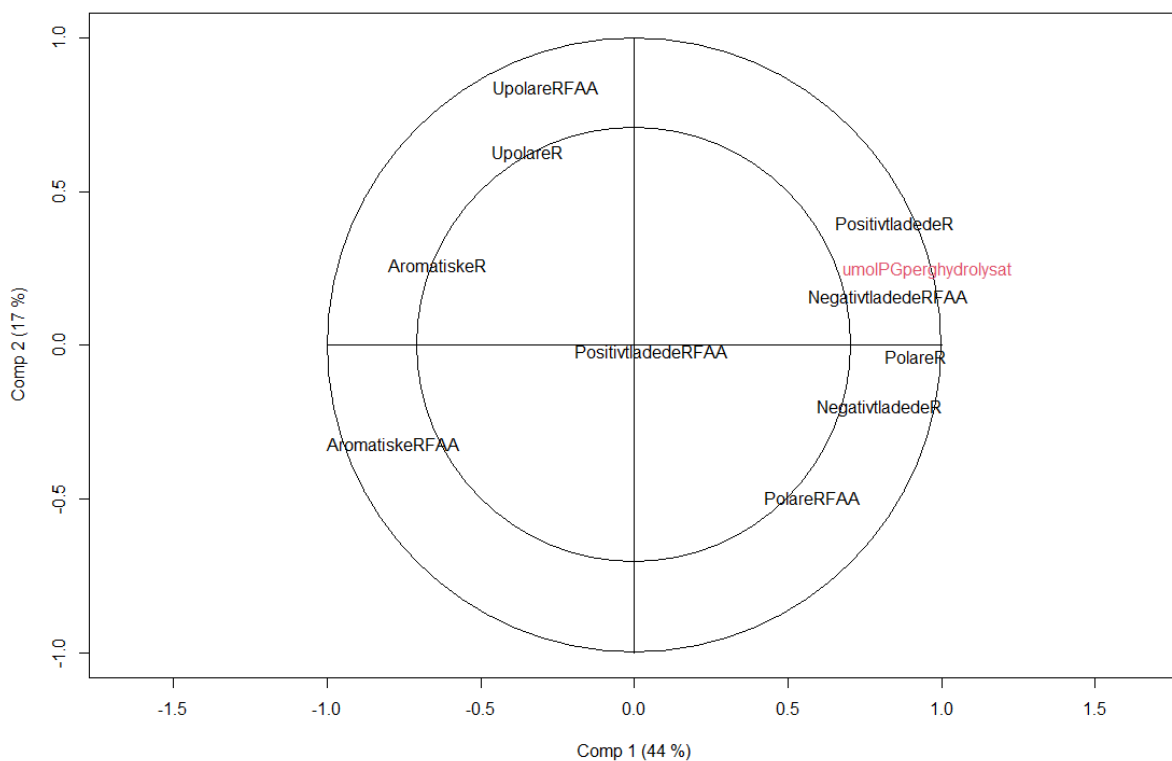
Figur 3.10: *Scores*-plott for de to første komponentene i PLSR modellen for matrise 5.

*Scores*-plottet for de to første komponentene viser at seihydrolysatene utgjør en klynge vekk fra de andre hydrolysatene. Det er ikke like tydelige grupperinger etter fisketype som for matrise 1. Laksehdrolysatene utgjør også en klynge, men sammen med det ene torskhydrolysatet. Ørrehydrolysatet er en uteligger i dette tilfellet.

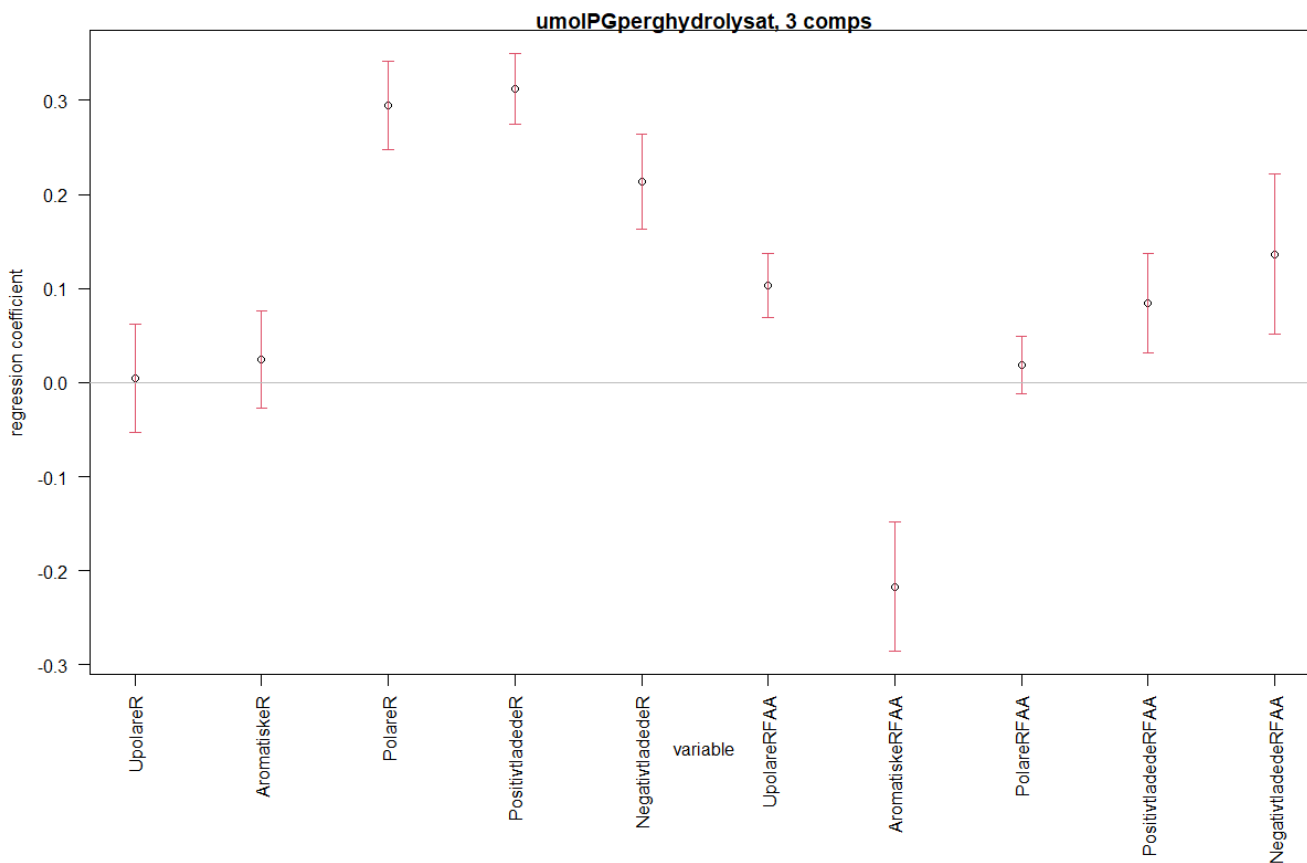


Forskjellen fra matrise 1 er at variablene viser sammensetning av ulike kategorier av aminosyrer. Plottet indikerer at fisketype alene ikke er nok til å gi utslag i sammensetning av ulike aminosyre-kategorier og i prediksjon av ABTS antioksidantkapasitet. Plottet kan til en viss grad antyde at type fiskefraksjon gir utslag på sammensetning av ulike kategorier av aminosyrer, hvor seihydrolysat fra ryggbein skiller seg betraktelig fra de andre som er fra hoder. Dette var også tilfellet for matrise 1, så dette ble også diskutert tidligere. Det hadde vært interessant å se om de samme mønstrene hadde vært i en modell med enda flere uavhengige prøver fra sei, torsk og laks, men fra både like og ulike fraksjoner.

Korrelasjon-loadings og regresjonskoeffisientene viser som sagt mye av det samme, se figur 3.11 og 3.12. Korrelasjon-loadings for de to første komponentene viser at de fleste variablene er godt forklart av de to første komponentene, med unntak av positivt ladete R FAA. Figuren viser også at følgende x-variabler er positivt korrelert med ABTS antioksidantkapasitet: positivt ladete R, negativt ladete R, polare R og negativt ladete R FAA. ABTS antioksidantkapasitet er også negativt korrelert med aromatiske R og aromatiske R FAA. Det er også tydelig at ABTS antioksidantkapasitet er best modellert av den første komponenten. Det er polare R, positivt ladete R, negativt ladete R og aromatiske R FAA som har størst regresjonskoeffisienter, og dermed har størst innflytelse på modelleringen av Y [38].



Figur 3.11: Korrelasjon-loadings for de to første komponentene i PLSR-modellen for matrise 5.



Figur 3.12: Regresjonskoeffisientene for de tre valgte komponentene i PLSR-modellen til matrise 5.

Bidraget av ladede aminosyrer til antioksidantkapasitet ble diskutert for matrise 1. I denne modellen ser det ut til at % negativt ladede og % positivt ladede aminosyrer gir et positivt bidrag til ABTS antioksidantkapasitet. Dette var ikke tilfellet i modellen av matrise 1 der aminosyrer som Arg og His så ut til å bidra negativt til antioksidantkapasitet. Det kan være at enkeltaminosyrer alene ikke er nok til å beskrive og/eller predikere ABTS-antioksidantkapasitet. Ved å inkludere denne matrisen var målet å se om antioksidantkapasitet heller kunne relateres til ulike aminosyrekarakteristika på tvers av enkeltaminosyrer. Denne modellen antyder at ladede aminosyrer er viktige for å beskrive ABTS antioksidantkapasitet, i hvertfall med dette datagrunnlaget. Som nevnt tidligere så ble det gjort noen funn i litteraturstudien i prosjektoppgaven som indikerte det samme, både relatert til bedre metallbinding, ABTS radikal fjerningsevne og reduserende evne [20]. Heller ikke i denne modellen ser aromatiske aminosyrer ut til å ha vesentlig påvirkning på ABTS radikal fjerningsevne, selv om dette er aminosyrer med betydelig oksidasjonspotensiale [28].

His er klassifisert som en positivt ladet aminosyre, men er også en aromatisk aminosyre. Det er derfor noe overlapp mellom aminosyrekategoriene. Selv om ladede aminosyrer ser ut til å ha en effekt på ABTS radikal fjerningsevne, så kan det være at noen av de ladede aminosyrene også bidrar til antioksidantkapasitet av andre underliggende

grunner. Asp og Glu vil som nevnt tidligere også inneholde aminosyrene Asn og Gln, og heller ikke her kan det derfor gjøres et tydelig skille mellom aminosyrene.

Peptidstruktur er kompleks, og det er derfor også utfordrende å beskrive peptidstruktur i peptidblandinger. Iwaniak et al (2015) nevner flere ulike kategorier av deskriptorer som kan brukes til å beskrive et molekyl, blant annet elektrostatiske, geometriske, kvantekjemiske og termodynamiske parametere. Z-skala og HESH (*Hydrophobic, Electronic, Steric and Hydrogen*) er kjente aminosyredeskriptorer, hvor den sistnevnte ble generert fra en PCA på 171 aminosyreegenskaper. [36] En vanlig tilnærming til QSAR-analyse av peptider er også å oversette kjente peptidsekvenser til et datasett som beskriver peptider i form av aminosyredeskriptorer og posisjon i sekvensen [41]. Dette kan i større grad beskrive noen av peptidene tilstede i hydrolysatene, men samtidig så består hydrolysater både av frie aminosyrer og flere peptider av ulik størrelse. Enkeltpeptider alene vil ikke kunne beskrive interaksjoner mellom peptider, hverken synergiske, additive eller antagonistiske interaksjoner [46]. Dette kan også være faktorer som direkte påvirker antioksidantkapasitet, men som er vanskelige å måle. I litteraturstudien i prosjektoppgaven ble det gjort flere struktur-aktivitet funn relatert til proteinstrukturkategorier som sekvens og molekylvekt [20]. For at en *radical scavenger* skal være en god antioksidant kan heller ikke proteinradikalene som dannes bidra til lipid oksidasjon. Når en aminosyre i en peptidkjede oksideres så kan det oppstå forskjellige reaksjoner videre i peptidsekvensen. [28] Det er derfor sannsynlig at peptidsekvens og størrelse også vil ha betydning for reaktiviteten til et proteinradikal og derfor også antioksidantkapasiteten. I de neste matrisene er det derfor inkludert andre mål på proteinstruktur basert på de tilgjengelige variablene som var målt for prøvene: mengde tioler og hydrolysegrad.

### 3.4 Matrise 2 og 6: TAA, FAA, tioler, ABTS antioksidantkapasitet

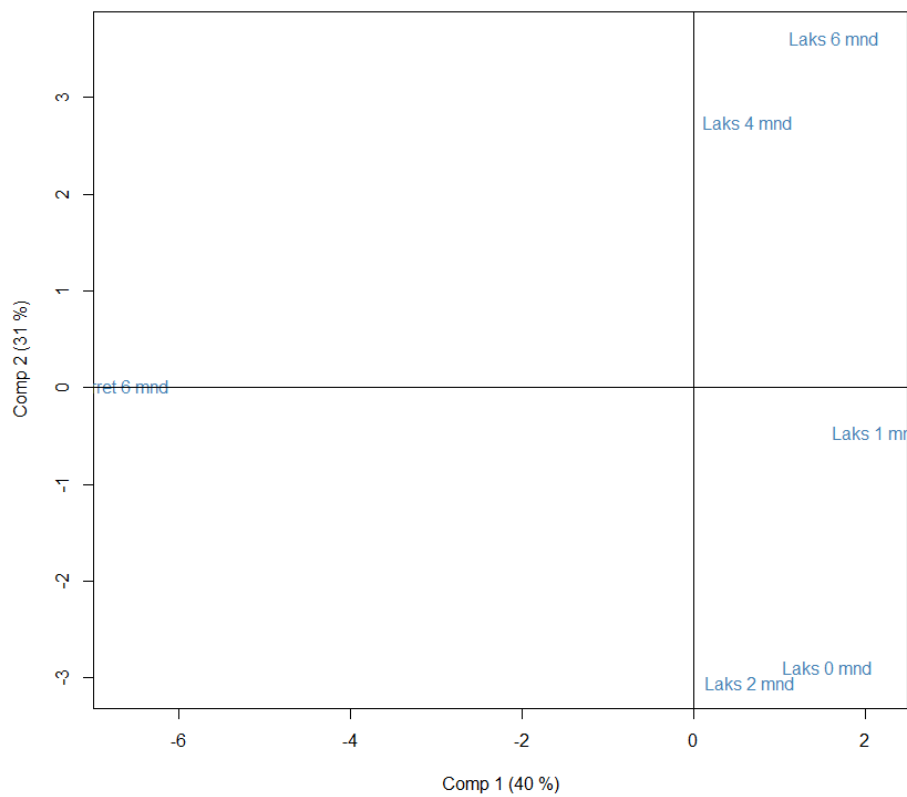
Matrise 2 består av variablene: TAA (%), FAA (%), mengde tioler (nmol/mg protein) og ABTS antioksidantkapasitet (umol PG/g hydrolysat). Matrise 6 har de samme variablene, men TAA og FAA var inndelt etter aminosyredeskriptorer på samme måte som i matrise 5. Matrise 2 og 6 består av lakse- og ørrehydrolysatene. Boksplottet er vist i appendiks og viser at det er stor variasjon i mengde tioler mellom prøvene. Selv om datagrunnlaget var begrenset til seks prøver, så ble PLSR likevel kjørt for begge matrisene for å se om modellene viste spesielle mønster i datasettene. Resultatene av kryssvalideringen er vist gjennom valideringsplott i appendiks i figur B.25 og B.36. For de andre to matrisene så flatet MSEP ut med økt antall komponenter, men for matrise 2 og 6 så viste valideringsplottene en vesentlig nedgang i MSEP for hver komponent til maksimalt antall komponenter var nådd. På grunn av de vesentlige forskjellene i MSEP mellom hver komponent ble maksimalt antall komponenter valgt, 4, for de to ulike PLSR modellene, den ene for matrise 2 og den andre for matrise 6. Tabell 3.4 viser en oversikt over forklart y-varians, forklart x-varians og RMSEP ved valgt antall PLS komponenter for de to ulike PLSR-modellene tilhørende matrise 2 og 6.

Tabell 3.4: Modellparametere ved valgt antall PLS komponenter for de to kryssvaliderte PLSR-modellene tilhørende matrise 2 og 6.

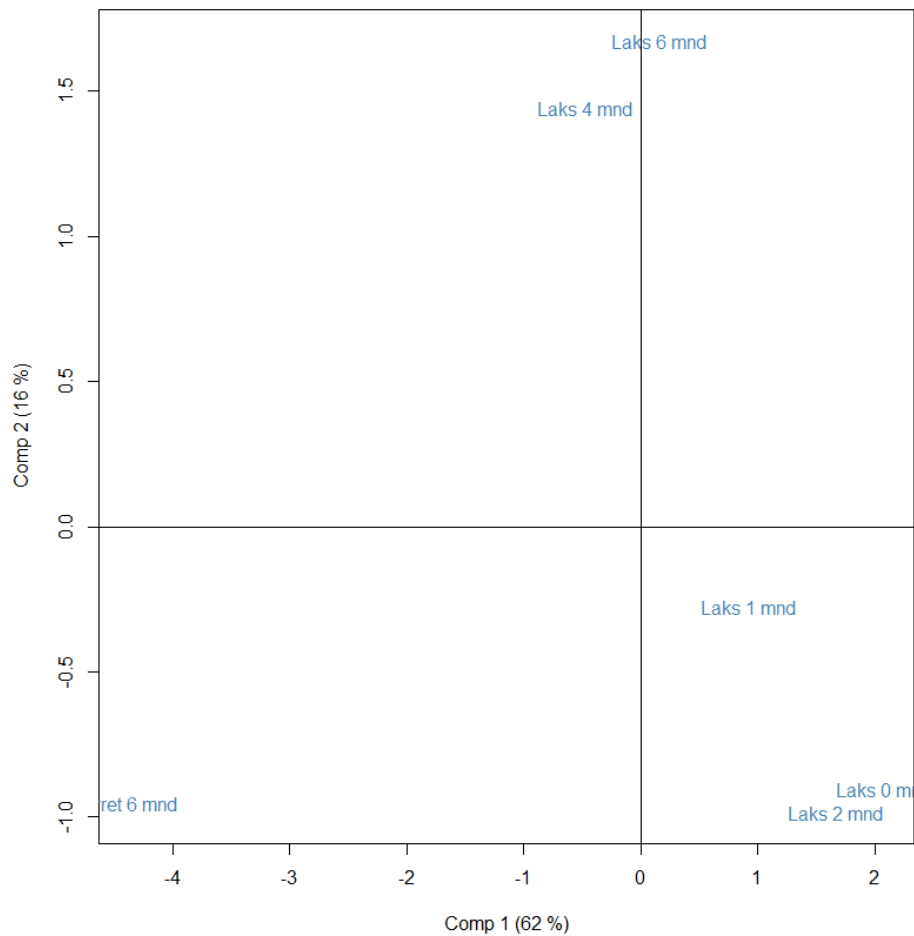
| <b>Matrise</b> | <b>Komponenter</b> | <b>RMSEP</b> | <b>Forklart y-varians (%)</b> | <b>Forklart x-varians (%)</b> |
|----------------|--------------------|--------------|-------------------------------|-------------------------------|
| 2              | 4                  | 0.08964      | 99.96                         | 92.15                         |
| 6              | 4                  | 0.4186       | 99.98                         | 93.25                         |

MSE er et mål på hvor godt modellen beskriver dataene, og kryssvalideringen skal si noe om modellens evne til å predikere nye data [38], [42]. Relativt til de standardiserte verdiene i datasettene [-2,2] så var RMSEP etter tre komponenter relativt store for begge modellene, 0.4185 for matrise 2 og 0.74 for matrise 6. Etter fire komponenter var RMSEP som vist i tabell 3.4 mindre, men relativt stor fortsatt for matrise 6 selv om forklart y-varians var tilnærmet 100%. Dette indikerer at modellen for matrise 6 har begrenset prediktiv evne, og dermed ikke beskriver alle prøvene godt. Dette er også reflektert i prediksjonsplottet som viser de kryssvaliderte prediksjonene av ABTS antioksidantkapasitet mot de faktiske verdiene, se figur B.38 i appendiks.

Det er generelt stor usikkerhet knyttet til både matrise 2 og 6 ettersom modellene er basert på så få prøver, og vil dermed naturlig nok være sensitive til uteliggere. *Scores*-plottene for de to første komponentene i modellen med henholdsvis matrise 2 er vist i figur 3.13 og matrise 6 i figur 3.14. Begge plottene viser at ørret er en tydelig uteligger fra de andre prøvene, som også er naturlig ettersom det er en annen fisketype enn laks. Samtidig så er laksehydrolysatene fra samme hydrolysat bare med ulik lagringstid i romtemperatur. Dette betyr at modellen i stor grad begrenses til å kun beskrive samme hydrolysat med små variasjoner i lagring. Det er derfor mulig at MSEP i stor grad reflekterer modellens repeterbarhet for samme type prøve istedenfor den prediktive evnen til å beskrive Y fra X i nye prøver [47]. Det er derfor usikkert om prøvene beskriver representabel variasjon på tvers av type fisk og fraksjoner for både X og Y, selv om matrise 2 viste en relativt god RMSEP. Datagrunnlaget i denne oppgaven var begrenset til hva som var tilgjengelig på instituttet.

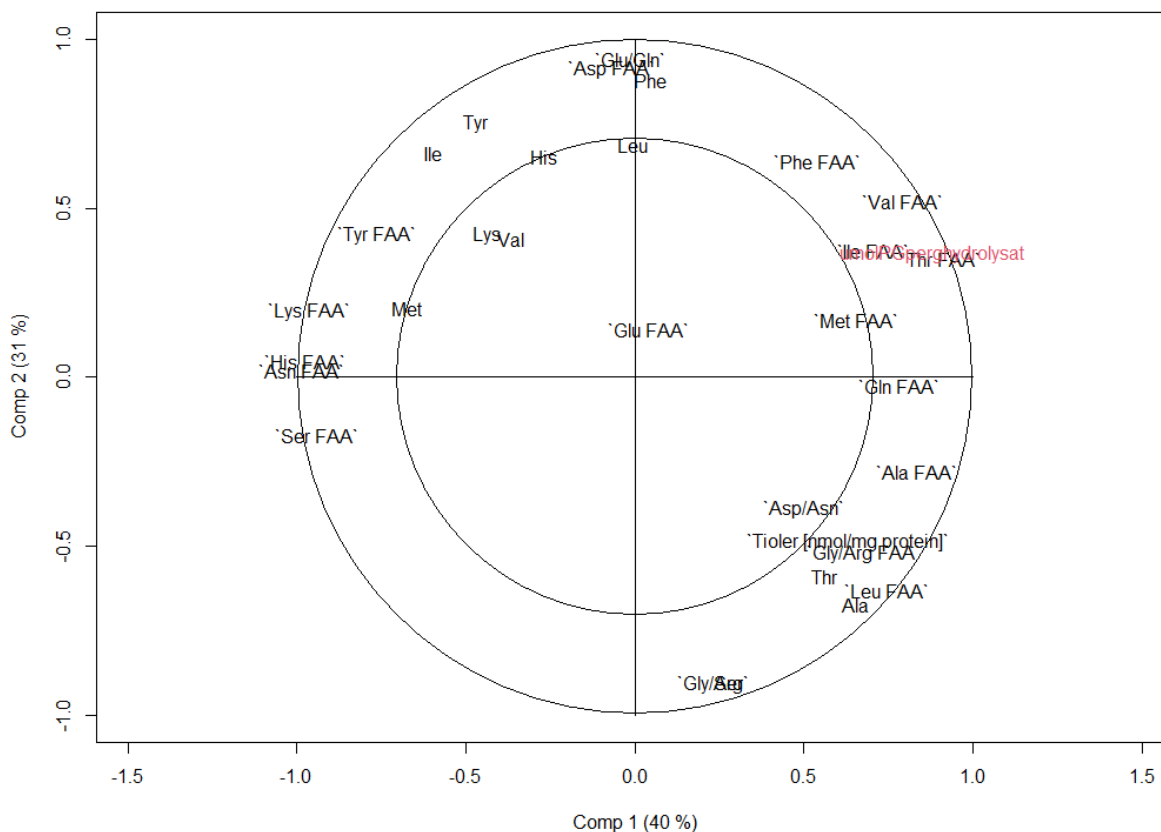


Figur 3.13: Scores for de to første komponentene i PLSR-modellen for matrise 2.



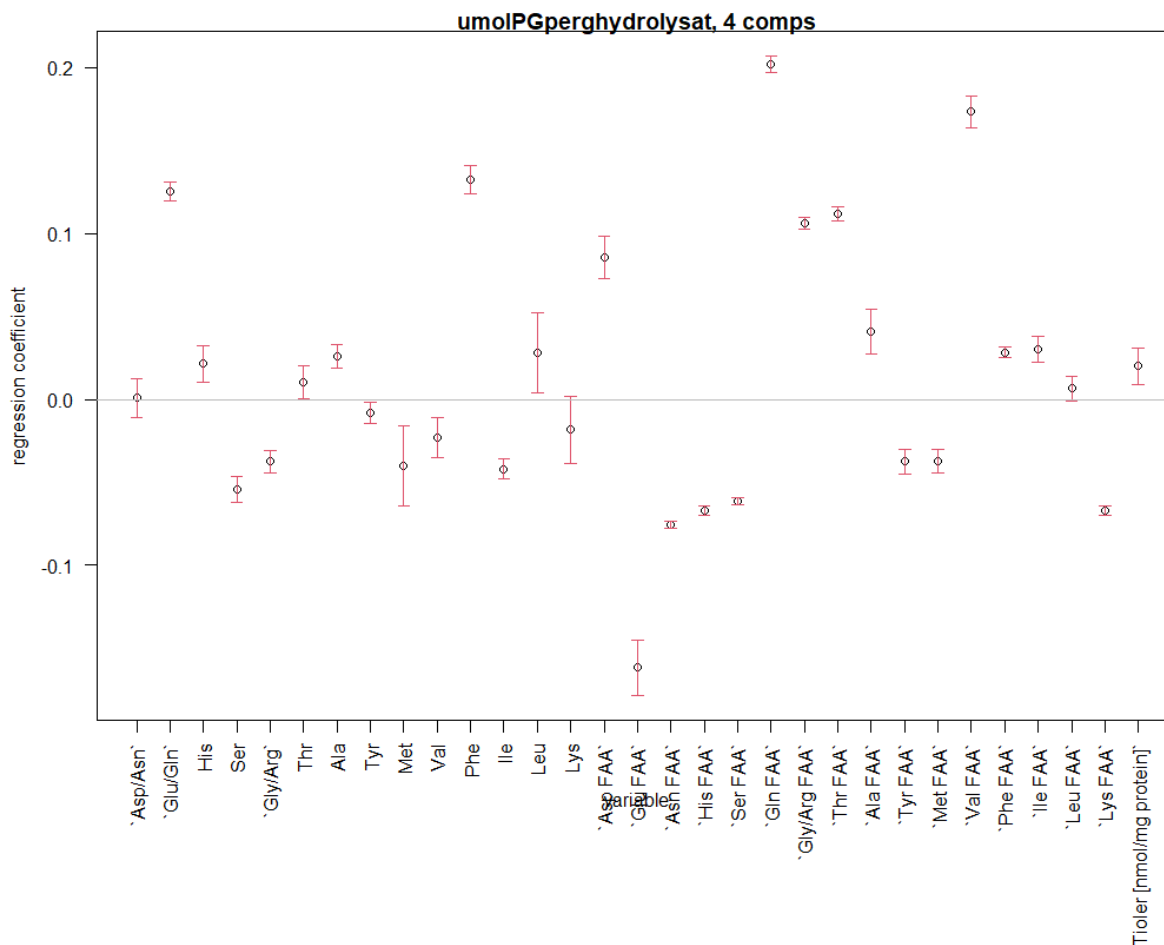
Figur 3.14: Scores for de to første komponentene i PLSR-modellen for matrise 6.

Regresjonskoeffisientene for PLSR modellen av matrise 6 bekrefter det som er beskrevet over, med relativt store standardavvik som gjør det vanskelig å konkludere hvilken betydning de ulike X-variablene har for ABTS-antioksidantkapasitet. Av den grunn er ikke resultatene av denne modellen presentert og diskutert her, men heller vist i appendiks. Resultatene av modellen for matrise 2 er vist i de neste figurene. Figur 3.15 viser korrelasjonsloadings for de to første komponentene i PLSR modellen til matrise 2. Figur 3.16 viser regresjonskoeffisientene til PLSR-modellen med alle fire komponentene.



Figur 3.15: Korrelasjon-loadings for de to første komponentene i PLSR-modellen for matrise 2.

Fra korrelasjon-loadings ser det ut til at stort sett alle variablene er godt forklart av de to første komponentene, inkludert  $\mu\text{mol PG/g}$  hydrolysat. ABTS antioksidantkapasitet korrelerer positivt og negativt med flere variabler. Regresjonskoeffisientene viser at Gln FAA, Glu/Gln, Phe og Val FAA er spesielt viktige for modelleringen av ABTS antioksidantkapasitet. Det er også Glu FAA, selv om den er svakere forklart av de to første komponentene. Som vist i korrelasjon-loadings for alle fire komponentene i appendiks, så er Glu FAA godt forklart av komponent tre. Phe, Val FAA, Gln FAA og Glu FAA hadde ikke spesielt stor betydning for Y i modellen for matrise 1. Prøvegrunnet kan i stor grad forklare forskjellene mellom de to modellene. Modellen for matrise 1 vil sannsynligvis bedre beskrive varians i ABTS antioksidantkapasitet på tvers av ulike typer fisk og fraksjoner. Dette kan valideres ytterligere ved tilgang på et testsett med uavhengige prøver.



Figur 3.16: Regresjonskoeffisientene for de fire komponentene i PLSR-modellen for matrise 2.

I modellen for matrise 2 kan det se ut som noen typer frie aminosyrer har betydning for ABTS antioksidantkapasitet, selv om frie aminosyrer generelt ikke er kjent for å være gode antioksidanter [28]. Mengde tioler for laksehdrolysatene sank med antall måneder lagret i romtemperatur, og indikerte derfor oksidasjon av Cys [49]. Fra denne modellen ser det derimot ikke ut som om mengde tioler og dermed Cys har noen spesiell påvirkning på ABTS antioksidantkapasitet, selv om Cys er en nukleofil aminosyre med et stort oksidativt potensiale [28]. Resultatene fra Robin indikerer at det har skjedd oksidasjon av proteinhydrolysatene, og dermed også endringer i peptidstruktur deriblant oksidasjon av Cys [49]. Det er usikkert om oksidasjon kan ha gitt andre produkter som ikke kan måles i hydrolysatene, men som har en direkte eller indirekte effekt på antioksidantkapasiteten.

### 3.5 Matrise 4 og 8: TAA, FAA, hydrolysegrad, ABTS antioksidantkapasitet

Matrise 4 består av variablene: TAA (%), FAA (%), hydrolysegrad (%) og ABTS antioksidantkapasitet (umol PG/g hydrolysat). Matrise 8 består av de samme variablene, men TAA og FAA var inndelt etter aminosyredeksriptorer. Følgende prøver var en del av de to matrisene: Laks 0/6 mnd, Ørret 6 mnd, Torsk feb/mar 400H<sub>2</sub>O/200H<sub>2</sub>O/200sH<sub>2</sub>O.



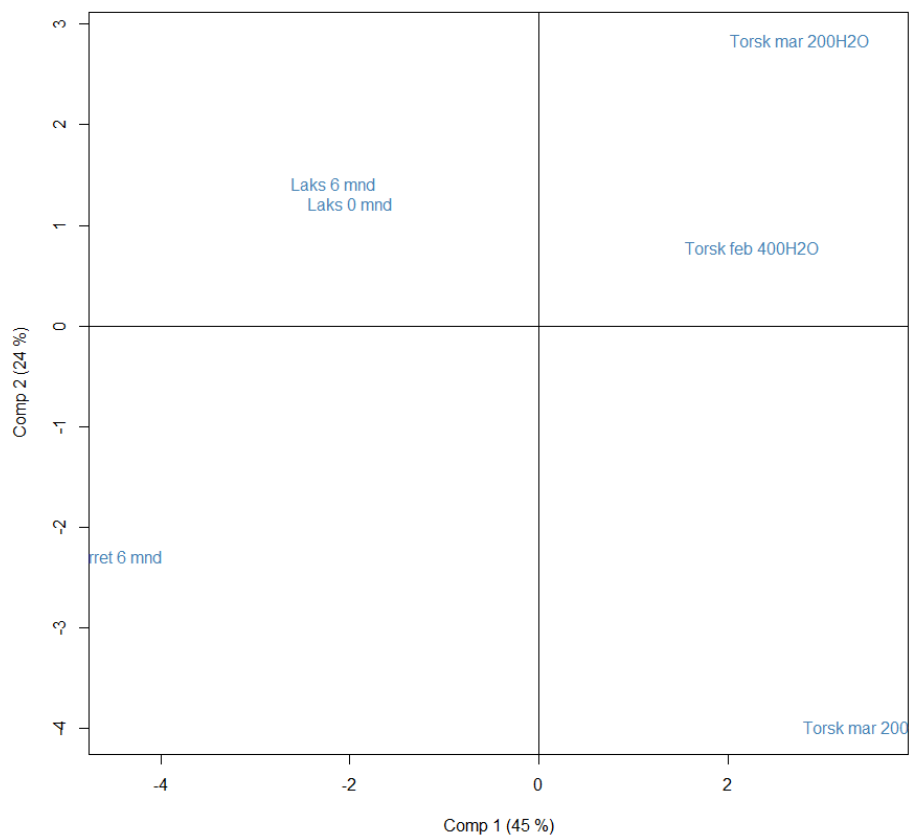
Datagrunnlaget var her også begrenset til kun seks prøver. Koden genererte maksimalt fire komponenter. Resultatet av kryssvalideringen er vist gjennom valideringsplott for hver av matrisene i figur B.49 og B.60 i appendiks. For matrise 4 flatet MSEP ut etter tre komponenter. For matrise 8 var det et minimumspunkt ved tre komponenter før MSEP steg ved fire komponenter. Forklart y-varians var relativt stor ved tre komponenter for begge matrisene. Tre komponenter ble derfor inkludert i de to gjennomsnittlige PLSR-modellene henholdsvis for matrise 4 og 8. Tabell 3.5 viser en oversikt over modellparameterne i de to modellene.

Tabell 3.5: Modellparametere ved valgt antall PLS komponenter for de to kryssvaliderte PLSR-modellene tilhørende matrise 4 og 8.

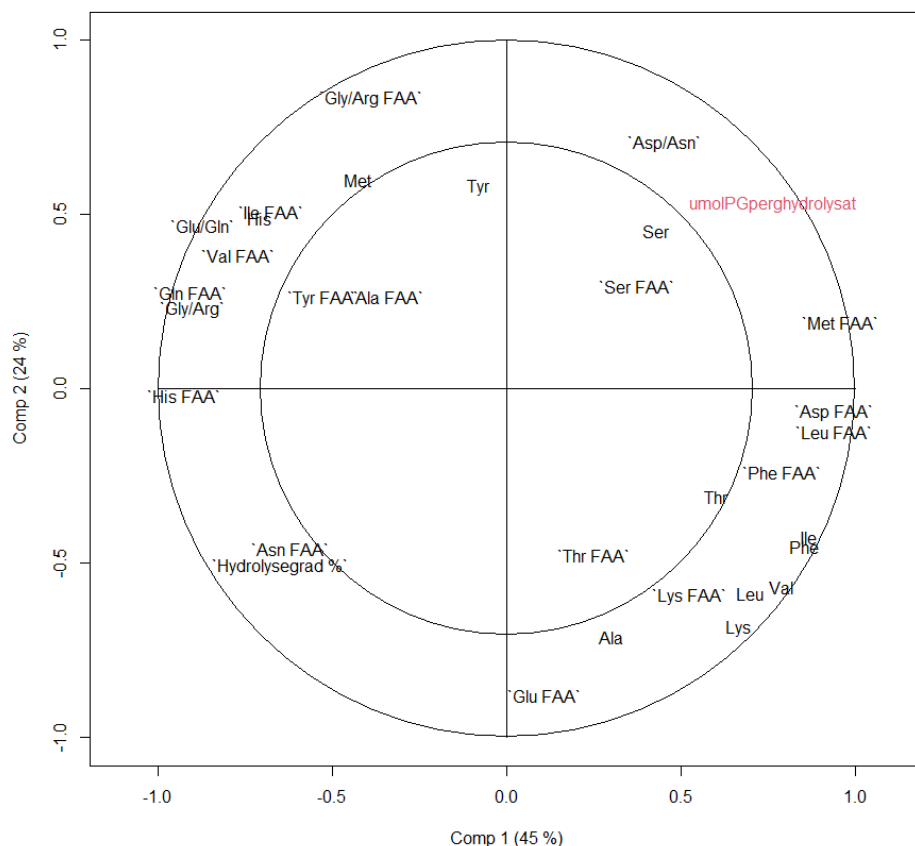
| Matrise | Komponenter | RMSEP  | Forklart y-varians (%) | Forklart x-varians (%) |
|---------|-------------|--------|------------------------|------------------------|
| 4       | 3           | 0.6323 | 91.19                  | 88.55                  |
| 8       | 3           | 0.7071 | 85.63                  | 86.09                  |

Relativt til de standardiserte verdiene i datasettet [-2,2] så er RMSEP for både matrise 4 og 8 relativt store, og enda større enn for matrise 2 og 6. Dette indikerer også at evnen til å predikere ABTS antioksidantkapasitet ikke er like god som for eksempel i modell 1 og 5. Dette bekreftes ved å studere de kryssvaliderte prediksjonene for begge modellene, der det er betydelige avvik mellom predikert og målt ABTS antioksidantkapasitet (se prediksjonsplott i appendiks). Dette kan også bety at evnen til å predikere og dermed også beskrive ABTS-antioksidantkapasitet gjennom x-variablene ikke er spesielt god for de tilgjengelige prøvene. Regresjonskoeffisientene for begge matrisene har også relativt store standardavvik, som gjør det vanskelig å tolke betydningen av variablene for ABTS antioksidantkapasitet i modellene. På grunn av disse begrensningene vil funnene kun presenteres i korte trekk for modellen med minst RMSEP, altså matrise 4. Fokuset er på funn knyttet til den nye variabelen: hydrolysegrad.

De to første komponentene beskriver 87.04% av variansen i ABTS antioksidantkapasitet, og vil derfor vise hovedtrendene for hydrolysatene og variablene i både *scores*-plottet og korrelasjon-*loadings*-plottet. *Scores*-plottet for PLSR-modellen av matrise 4 er vist i figur 3.17 for de to første komponentene, og tilsvarende plott for alle komponentene er vist i figur B.52 i appendiks. Figuren viser stor spredning blant de ulike fiskehydrolysatene. Figur 3.18 viser korrelasjon-*loadings* for de to første komponentene i modellen. Korrelasjon-*loadings* viser at både hydrolysegrad og ABTS antioksidantkapasitet er godt forklart av modellen, og ser ut til å korrelere negativt.



Figur 3.17: Scores for de to første komponentene i PLSR-modellen for matrise 4.

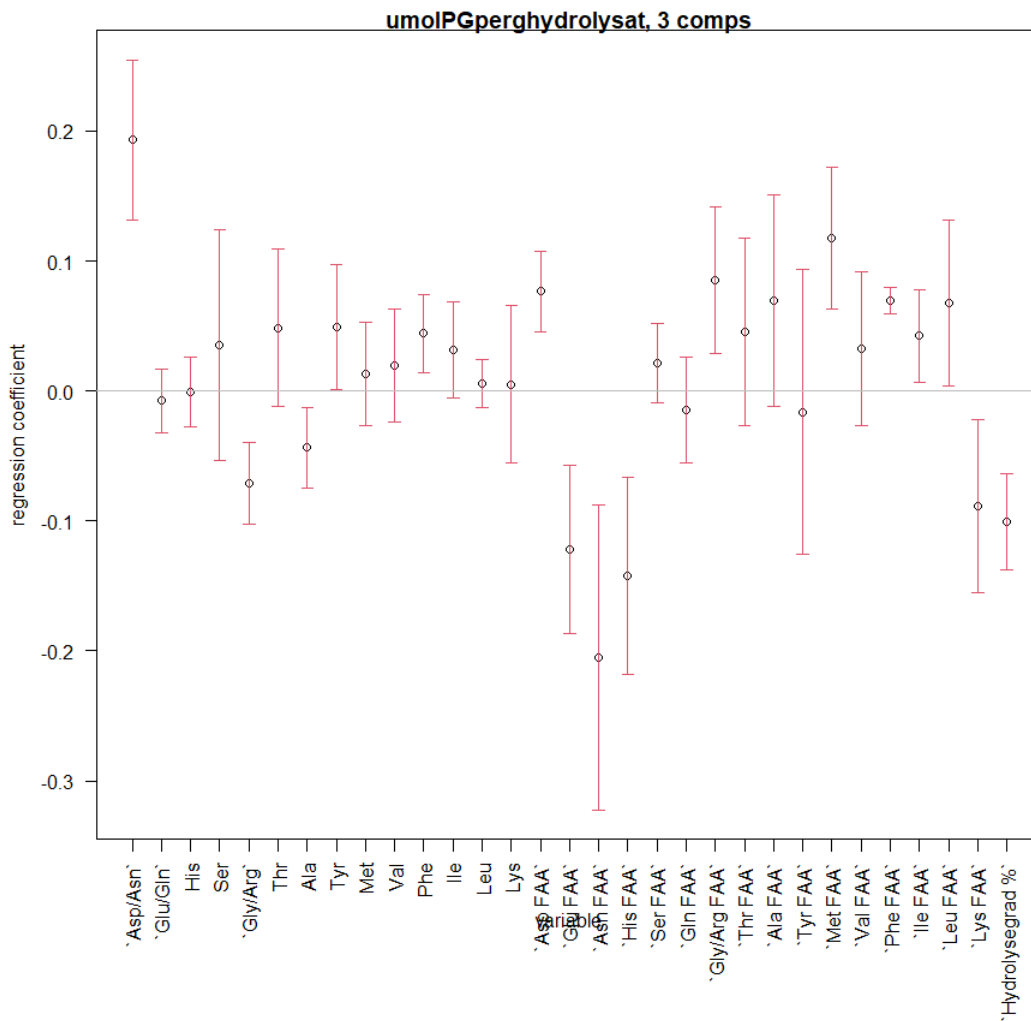


Figur 3.18: Korrelasjon-loadings for de to første komponentene i PLSR-modellen for matrise 4.

Regresjonskoeffisientene for modellen med tre komponenter er vist i figur 3.19. Som nevnt over så er standardavvikene relativt store, og viser til ustabilitet i modellen basert på variasjoner mellom de individuelle prøvene [47]. Det kan likevel se ut som hydrolysegrad har noe betydning for ABTS antioksidantkapasitet i denne modellen, samtidig som det ikke er den viktigste variabelen. De kan likevel antyde at en økning i hydrolysegrad og dermed også i kortere peptider til en viss grad kan relateres til en dårligere ABTS antioksidantkapasitet.

Peptidstørrelse ble også undersøkt i litteraturstudien i prosjektoppgaven, der flere av funnene var motstridende [20]. Flere av studiene viste til bedre antioksidantkapasitet ved lavere molekylvekt [62], [63]. Damgaard et al. (2015) fant derimot ingen korrelasjon mellom andel peptider med lav molekylvekt og hverken ABTS antioksidantkapasitet eller inhibering av lipid oksidasjon [60]. Det kan derfor se ut til at peptidstørrelse i mange tilfeller ikke er avgjørende for antioksidantkapasitet. Antioksidantkapasitet til proteiner er begrenset av tertiærstruktur og folding som kan gjøre flere aminosyrerester utilgjengelige [28]. Kortere peptider vil ideelt sett bety flere terminalender og mindre folding, og dermed mer eksponerte aminosyrerester. Folding er i stor grad styrt av svake interaksjoner [18]. Slike interaksjoner vil fortsatt kunne skje mellom aminosyrer i kortere peptider, men det er usikkert i hvor stor

grad dette kan påvirke tilgjengeligheten av aminosyrerester for reaksjon.



Figur 3.19: Regresjonskoeffisientene for de tre komponentene i PLSR-modellen for matrise 4.

For hver modell har det blitt diskutert om funnene kan indikere noen peptidstruktur-antioksidantkapasitet sammenhenger. I denne masteroppgaven har ABTS antioksidantkapasitet blitt brukt som et mål på antioksidantkapasitet selv om det kun er en av mange mulige metoder for å måle antioksidantkapasitet. Frie radikaler er viktige pådrivere under autooksidasjon av fettsyrer, og på den måten kan det anses som relevant å studere ABTS antioksidantkapasitet for å si noe om evnen til å nøytralisere frie radikaler [17]. ABTS er derimot ikke en vanlig radikal i biologiske systemer [33]. Den har også et helt annet reduksjonspotensiale enn andre radikaler. Hydroperoksyldikaler er eksempler på naturlige radikaler i oksidert mat. Hydroperoksyldikaler har lavere reduksjonspotensiale (1060 mV) enn hydroksyldikaler (2310 mV), og er derfor også mer selektive i reaksjon med aminosyrer. [28], [29] ABTS radikalen har et enda lavere reduksjonspotensiale på 680 mV [33]. Dette indikerer at ABTS radikalen er mindre reaktiv, og forventes å være mer selektiv. På den måten er kanskje ikke ABTS antioksidantkapasitet alene nok til å kunne reflektere antioksidantkapasitet i matsystemer.

Det er også vist at resultater av ABTS-analysen ikke er direkte sammenlignbare med andre metoder. Kristinová (2008) målte antioksidantkapasitet til flere ulike molekyler gjennom flere *in vitro* analyser, og viste at ABTS-analysen ga helt andre resultater enn de andre analysene. FC, FRAP og DPPH-analysene derimot viste relativt like trender for de målte molekylene. Dette er på tross av at DPPH-metoden kan minne om ABTS-metoden. [64] ABTS-analysen gir i dette tilfellet helt annen informasjon enn de andre analysemetodene. Det er derfor også sannsynlig at det hadde vært tilfellet med flere *in vitro* analyser på hydrolysatene i denne masteroppgaven. utfordringer ved bruk av *in vitro*-metoder ble presentert i delkapittel 1.3.3. *In vitro* analyser for måling av antioksidantkapasitet, spesielt med tanke på varierende mekanismer, eksperimentelle betingelser og sammenligning med *in vivo* bioaktivitet. Det er mange faktorer som varierer og gjør metodene vanskelige å sammenligne, samtidig som en metode kun kan vise et av mange mål på antioksidantkapasitet. Enda en styrke med PLSR som metode er at den kan modellere flere Y-responser enten i en modell eller i flere. En PCA-analyse av flere Y-responser kan avdekke om de korrelerer eller ikke, og derfor om de bør modelleres separat i PLSR. [38] På grunn av forskjellene i mekanismer mellom de ulike *in vitro* analysene på antioksidantkapasitet er det sannsynlig at de må modelleres separat. Et forslag til hvordan PCA kan gjøres i R for flere mål på antioksidantkapasitet er presentert i datascriptet i appendiks (figur B.2-B.3). QSAR-modellering har potensiale til å kunne belyse viktige sammenhenger mellom struktur og aktivitet, og dermed også mulige mekanismer. [41] Det kunne derfor vært interessant å inkludere enda flere mål på antioksidantkapasitet, og se hva resultatet av modelleringen viser.

### 3.6 Vurdering av litteratur som kilde til data

Litteratur og databaser ble også vurdert for bruk i dataanalysen. Formålet med å bruke denne typen data ville vært å bruke et større datagrunnlag i analysen. Selv om datagrunnlaget er større, så betyr det derimot ikke at modellen er bedre. Bruk av litteratordata vil kun være nyttig om prøvene som er inkludert i en datamatrikse er sammenlignbare. Det ble bestemt egne kriterier for prøver fra litteratur og databaser:

- Likt eksperimentelt oppsett og betingelser [41].
- Oversikt og estimat på måleusikkerhet [41].
- Sammenlignbare måleenheter.

De to første kriteriene ble anbefalt av Pripp et al. (2005) for QSAR modellering fra litteratur og databaser [41]. Litteraturstudien i prosjektoppgaven viste bruk av flere metoder for å måle antioksidantkapasitet basert på ulike mekanismer og radikaler, med ulike analyseparametere. Litteraturstudien viste at det kan være variasjon mellom metoder i følgende parametere: substrater, løsemidler, deteksjon, blank-prøve og måleenheter. [20] For at litteratordata skal kunne tilføre noe ekstra til analysen må alle eksperimentelle betingelser og måleenheter være sam-

menlignbare.

Basert på de overnevnte kriteriene ble flere av oversiktsartiklene fra den kvantitative delen av litteraturstudien i prosjektoppgaven gjennomgått [20]. Dette var oversiktsartikler som hadde et stort fokus på identifiserte peptidsekvenser. Flere av oversiktsartiklene beskrev eksempler på peptidsekvenser og målt aktivitet. Resultater for mange ulike *in vitro*-analyser var presentert med varierende måleenheter. Felles for oversiktsartiklene var at det manglet informasjon i form av analysebetingelser og estimert usikkerhet. [19], [24], [65], [66] For at disse dataene skal være brukbare i en analysemodell, så må alle kildene gjennomgås for å sjekke sammenlignbarhet mellom eksperimentelle metoder og måleenheter.

Iwaniak et al. (2015) viser en oversikt over databaser med bioaktive peptider. Denne listen ble gjennomgått, og fra den var det kun BIOPEP som var relatert til antioksidantkapasitet.[36] BIOPEP nevnes også av Li-Chan (2015) som en av databasene som kun fokuserer på peptider fra mat [46]. BIOPEP ble derfor studert nærmere for antioksidative peptider. Tilgjengelig informasjon var sekvens, masse, type aktivitet, referanse og EC50. EC50 var ikke tilgjengelig for de peptidene som ble sett på, og heller ikke informasjon om hvilken *in vitro* antioksidantkapasitet analyse som hadde blitt brukt. [67] Dette gjør det spesielt vanskelig å bruke informasjonen, ettersom det fins så mange analyser som også måler ulike mekanismer og substrater [33]. Som for oversiktsartiklene blir dette for lite informasjon til å kunne brukes, med mindre hver kilde grundig gjennomgås for å sammenligne eksperimentelt oppsett, prøver, målte variabler og estimert usikkerhet.

Forskningsartiklene på antioksidative peptider fra den kvalitative litteraturstudien i prosjektoppgaven ble også vurdert basert på de overnevnte kriteriene [20]. Disse har naturlig nok mer informasjon om eksperimentelle metoder og usikkerhet. To av fem artikler hadde tabellerte verdier for målt antioksidantkapasitet inkludert standardavvik, men de hadde brukt ulike *in vitro*-analyser for å måle antioksidantkapasitet og var derfor ikke sammenlignbare [60], [62]. På grunn av begrensningene i forhold til kriteriene som ble satt, så ble ingen data fra hverken litteratur eller databaser brukt i dataanalysen i denne masteroppgaven. Variasjon i *in vitro*-analysemetoder for antioksidantkapasitet ble diskutert i etterkant av PLSR-modellene, og indikerer hvor vanskelig det er å skulle finne sammenlignbare resultater som derfor kan brukes i en felles dataanalyse.

## Kapittel 4

# Konklusjon

Målet med masteroppgaven var å bruke dataanalyse som et verktøy for å studere sammenhenger mellom antioksidantkapasitet og ulike mål på peptidstruktur. For å svare på dette ble det gjennomført *Partial least squares regression* (PLSR) analyser med kryssvalidering (*Leave-One-Out Cross-Validation*) på ulike matriser. I tråd med *Quantitative structure-activity relationship* (QSAR) modeller tilsvarte X ulike mål på peptidstruktur og Y tilsvarte bioaktivitet.

Resultatene av modelleringen viser at de to modellene med størst datagrunnlag (matrise 1 og 5) beskriver dataene i datasettene godt med relativt små avvik mellom målt og predikert ABTS-antioksidantkapasitet. Dette viser at PLSR-analyse er en metode som kan brukes til å beskrive variasjon i denne typen data, og som potensielt kan modellere og dermed også beskrive antioksidantkapasitet fra variabler som aminosyresammensetning. Blant prøvene i modellene med størst datagrunnlag var det to distinkte nivåer for ABTS-antioksidantkapasitet, og modellene beskriver derfor denne variasjonen. Validering på flere uavhengige prøver er nødvendig for å svare på om datasettet representerer relevant variasjon i ABTS-antioksidantkapasitet. Med flere uavhengige prøver fra ulike typer fisk og fraksjoner vil MSEP også i større grad kunne reflektere X sin reproduserbare evne til å predikere Y for helt nye prøver. Dette gjelder spesielt for de matrisene som hadde et begrenset datagrunnlag til seks prøver. Tre av disse modellene (matrise 6, 4 og 8) viste begrenset evne til å predikere og dermed beskrive ABTS-antioksidantkapasitet på tvers av ulike prøver.

Regresjonskoeffisientene fra PLSR-modellen til matrise 1 viste at følgende variabler hadde vesentlig betydning for modelleringen av ABTS-antioksidantkapasitet: Asp/Asn, Glu/Gln, Gly/Arg, Lys, Asp FAA og His FAA. For matrise 5 bidro følgende variabler vesentlig til modelleringen av ABTS-antioksidantkapasitet: polare R, positivt ladede R, negativt ladede R og aromatiske R FAA. Disse funnene kan derfor indikere hvilke X-variabler som har betydning for peptider sin evne til å nøytralisere ABTS radikaler. Funnene indikerer at ladede aminosyrer har be-

tydning for ABTS antioksidantkapasitet. Funn i litteraturen tilsier at ladede aminosyrer er potente hydrogen eller elektron-donorer. Nukleofile og aromatiske aminosyrer har blitt beskrevet i litteratur som aminosyrer med stort oksidativt potensiale, og som slik potensielt kan nøytralisere radikaler. % av disse typene aminosyrer hadde derimot ikke spesiell betydning i modellene for matrise 1 og 5. Dette kan blant annet indikere at det er andre underliggende faktorer som påvirker ABTS-antioksidantkapasitet i peptidblandinger, som er vanskelige å måle. Dette kan være interaksjoner og reaksjoner mellom ulike peptider, aminosyrer og proteinradikaler.

PLSR er også en metode som håndterer flere responsvariabler, og derfor potensielt kan modellere flere mål på antioksidantkapasitet enten i samme modell eller i flere modeller. ABTS antioksidantkapasitet er kun et mål på antioksidantkapasitet. Det kunne derfor vært relevant å inkludere flere mål på antioksidantkapasitet for å studere flere mekanismer. Litteratur og databaser ble også vurdert for bruk i dataanalysen. Oversiktsartikler og databaser har begrenset informasjon om eksperimentelle betingelser, måleenheter og estimater på måleusikkerhet. Om data fra litteratur skal brukes må prøvene være sammenlignbare, og på grunn av disse begrensningene ble ikke gjennomgått litteratur brukt i dataanalysen i denne masteroppgaven.



## Kapittel 5

# Fremtidig perspektiv

Denne masteroppgaven viste at PLSR-analyse kan brukes til å beskrive variasjon i ABTS-antioksidantkapasitet fra variabler som aminosyresammensetning, og potensielt også avdekke relevante sammenhenger mellom molekylstruktur/sammensetning av peptider og ABTS antioksidantkapasitet. Til det videre arbeidet gis dermed følgende anbefalinger:

- Ved å validere funnene av modellene på et datasett med nye, uavhengige prøver så kan det undersøkes om modellen beskriver relevant forklart varians i ABTS-antioksidantkapasitet på tvers av prøver. Generelt så vil også et større datagrunnlag i selve modellen gi en enda mer robust modell.
- Ved å inkludere flere mål på antioksidantkapasitet så kan også flere antioksidantmekanismer studeres. Det kan gjøres en PCA av Y-variablene for å se om noen av *in vitro*-analysene gir korrelerende resultater, og om de skal modelleres separat eller ikke.
- Peptidstruktur er kompleks, og spesielt i peptidblandinger. Det kunne vært interessant å også utforske andre x-variabler, blant annet andre mulige aminosyredeskriptorer.
- Antioksidantkapasitet er en av mange typer bioaktivitet og proteinegenskaper. Potensielt kan denne typen modeller beskrive og avdekke andre viktige sammenhenger.

# Bibliografi

- [1] R. Richardsen, M. Myhre (SINTEF Ocean), R. Nystøyl og G. Strandheim (Kontali Analyse AS), «Analyse marint restråstoff 2019», 2020.
- [2] FAO, *The State of World Fisheries and Aquaculture 2020. Sustainability in action. Rome*. adresse: <https://doi.org/10.4060/ca9229en>, (accessed: 26.11.2020).
- [3] Nærings- og fiskeridepartementet, *Torskereguleringen er fastsatt*. adresse: <https://www.regjeringen.no/no/aktuelt/torskereguleringen-er-fastsatt/id2623661/>, (accessed: 23.11.2020).
- [4] FN-sambandet, *FNs bærekraftsmål*. adresse: <https://www.fn.no/om-fn/fns-baerekraftsmaal>, (accessed: 18.09.2020).
- [5] Nærings- og fiskeridepartementet, «Kjente ressurser-uante muligheter. Regjeringens bioøkonomistrategi», 2016.
- [6] R. Richardsen, M. Myhre (SINTEF Ocean), R. Nystøyl, G. Strandheim og A. Marthinussen (Kontali Analyse AS), «Analyse marint restråstoff 2018», 2019.
- [7] Norges sjømatråd, *Sjømateksport for 99 milliarder i 2018*. adresse: <https://seafood.no/aktuelt/nyheter/sjomateksport-for-99-milliarder-i-2018-/>, (accessed: 18.09.2020).
- [8] B. I. Bendiksen (Nofima), «Referansepriser i førstehåndsmarkedet for hvitfisk», 2018.
- [9] U. Winther, M. G. Sandberg, R. Richardsen, T. Olafsen (SINTEF Fiskeri og havbruk), R. K. Brandvik og J.-H. Hauvik (Kontali Analyse), «Potensial for økt verdiskaping i lakse- og ørretoppdrettsnæringen», 2011.
- [10] Lovdata, *Animaliebiproduktforskriften. (2016). Forskrift om animalske biprodukter som ikke er beregnet på konsum. (FOR-2016-09-14-1064)*. adresse: <https://lovdata.no/forskrift/2016-09-14-1064>, (accessed: 11.02.2021).
- [11] H. G. Kristinsson og B. A. Rasco, «Fish protein hydrolysates: production, biochemical, and functional properties», *Critical reviews in food science and nutrition*, årg. 40, nr. 1, s. 43–81, 2000.
- [12] Mills, *Mills Original kaviar*. adresse: <https://mills.no/produkt/mills/original-kaviar/>, (accessed: 26.11.2020).

- [13] Kavli, *Kaviar 60% Kavli*. adresse: <https://www.kavli.no/vaare-produkter-og-varemerker/kavli/Kaviar/kavli-kaviar-60>, (accessed: 26.11.2020).
- [14] FHF, *Restråstoff: Anvendelser av hoder til humant konsum (HEADS UP)*. adresse: <https://www.fhf.no/prosjekter/prosjektbasen/901308/>, (accessed: 27.11.2020).
- [15] —, *Kvalitetsprotein fra torskehoder (HEADS UP II)*. adresse: <https://www.fhf.no/prosjekter/prosjektbasen/901499/>, (accessed: 27.11.2020).
- [16] Mattilsynet, *Næringsstoffer i fisk og annen sjømat*. adresse: [https://www.matportalen.no/matvaregrupper/tema/fisk\\_og\\_skalldyr/naeringsstoffer\\_i\\_fisk\\_og\\_annen\\_sjomat-1](https://www.matportalen.no/matvaregrupper/tema/fisk_og_skalldyr/naeringsstoffer_i_fisk_og_annen_sjomat-1), (accessed: 30.09.2020).
- [17] T. Coultate, *Food: The Chemistry of its Components*. The Royal Society of Chemistry, 2016, ISBN: 9781849738804.
- [18] D. L. Nelson og M. M. Cox, *PRINCIPLES of BIOCHEMISTRY*. Macmillan Higher Education, 2017, ISBN: 9781464126116.
- [19] A.-L. Välimaa, S. Mäkinen, P. Mattila, P. Marnila, A. Pihlanto, M. Mäki og J. Hiidenhovi, «Fish and fish side streams are valuable sources of high-value components», *Food Quality and Safety*, årg. 3, nr. 4, s. 209–226, 2019.
- [20] H. Nedreberg Burud, «Proteinhydrolysater fra ulike restråstoff- sammenheng mellom struktur og bioaktivitet (Prosjektoppgave)», 2020, NTNU: Institutt for bioteknologi og matvitenskap.
- [21] WHO, *Hypertension*. adresse: <https://www.who.int/news-room/fact-sheets/detail/hypertension>, (accessed: 06.11.2020).
- [22] —, *Cardiovascular diseases (CVDs)*. adresse: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)), (accessed: 06.11.2020).
- [23] K. Majumder og J. Wu, «Molecular targets of antihypertensive peptides: understanding the mechanisms of action based on the pathophysiology of hypertension», *International Journal of Molecular Sciences*, årg. 16, nr. 1, s. 256–283, 2015.
- [24] A. V. Le Gouic, P. A. Harnedy og R. J. FitzGerald, «Bioactive peptides from fish protein by-products», *Bioactive molecules in food*. Cham: Springer International Publishing, s. 1–35, 2018.
- [25] Y. UG, I. Bhat, I. Karunasagar og M. BS, «Antihypertensive activity of fish protein hydrolysates and its peptides», *Critical reviews in food science and nutrition*, årg. 59, nr. 15, s. 2363–2374, 2019.
- [26] Y. Fu, A. M. Alashi, J. F. Young, M. Therkildsen og R. E. Aluko, «Enzyme inhibition kinetics and molecular interactions of patatin peptides with angiotensin I-converting enzyme and renin», *International journal of biological macromolecules*, årg. 101, s. 207–213, 2017.

- [27] Felleskatalogen, *Captopril*. adresse: <https://www.felleskatalogen.no/medisin/captopril-mylan-547252#legemiddelfoto>, (accessed: 25.11.2020).
- [28] R. J. Elias, S. S. Kellerby og E. A. Decker, «Antioxidant activity of proteins and peptides», *Critical reviews in food science and nutrition*, årg. 48, nr. 5, s. 430–441, 2008.
- [29] E. Choe og D. B. Min, «Chemistry and reactions of reactive oxygen species in foods», *Critical reviews in food science and nutrition*, årg. 46, nr. 1, s. 1–22, 2006.
- [30] R. J. Elias, D. J. McClements og E. A. Decker, «Antioxidant activity of cysteine, tryptophan, and methionine residues in continuous phase  $\beta$ -lactoglobulin in oil-in-water emulsions», *Journal of agricultural and food chemistry*, årg. 53, nr. 26, s. 10 248–10 253, 2005.
- [31] T. Waraho, D. J. McClements og E. A. Decker, «Mechanisms of lipid oxidation in food dispersions», *Trends in Food Science & Technology*, årg. 22, nr. 1, s. 3–13, 2011.
- [32] D. Huang, B. Ou og R. L. Prior, «The chemistry behind antioxidant capacity assays», *Journal of agricultural and food chemistry*, årg. 53, nr. 6, s. 1841–1856, 2005.
- [33] R. L. Prior, X. Wu og K. Schaich, «Standardized methods for the determination of antioxidant capacity and phenolics in foods and dietary supplements», *Journal of agricultural and food chemistry*, årg. 53, nr. 10, s. 4290–4302, 2005.
- [34] J. Pérez-Jiménez og F. Saura-Calixto, «Effect of solvent and certain food constituents on different antioxidant capacity assays», *Food research international*, årg. 39, nr. 7, s. 791–800, 2006.
- [35] R. Jónsdóttir, M. Geirsdóttir, P. Y. Hamaguchi, P. Jamnik, H. G. Kristinsson og I. Undeland, «The ability of in vitro antioxidant assays to predict the efficiency of a cod protein hydrolysate and brown seaweed extract to prevent oxidation in marine food model systems», *Journal of the Science of Food and Agriculture*, årg. 96, nr. 6, s. 2125–2135, 2016.
- [36] A. Iwaniak, P. Minkiewicz, M. Darewicz, M. Protasiewicz og D. Mogut, «Chemometrics and cheminformatics in the analysis of biologically active peptides from food sources», *Journal of Functional Foods*, årg. 16, s. 334–351, 2015.
- [37] J. Miller og J. C. Miller, *Statistics and chemometrics for analytical chemistry*. Pearson education, 2018.
- [38] S. Wold, M. Sjöström og L. Eriksson, «PLS-regression: a basic tool of chemometrics», *Chemometrics and intelligent laboratory systems*, årg. 58, nr. 2, s. 109–130, 2001.
- [39] A. H. Pripp og Y. Ardö, «Modelling relationship between angiotensin-(I)-converting enzyme inhibition and the bitter taste of peptides», *Food Chemistry*, årg. 102, nr. 3, s. 880–888, 2007.

- [40] S. Uno, D. Kodama, H. Yukawa, H. Shidara og M. Akamatsu, «Quantitative analysis of the relationship between structure and antioxidant activity of tripeptides», *Journal of Peptide Science*, årg. 26, nr. 3, e3238, 2020.
- [41] A. H. Pripp, T. Isaksson, L. Stepaniak, T. Sørhaug og Y. Ardö, «Quantitative structure activity relationship modelling of peptides and proteins as a tool in food science», *Trends in Food Science & Technology*, årg. 16, nr. 11, s. 484–494, 2005.
- [42] G. James, D. Witten, T. Hastie og R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013, ISBN: 9781461471370.
- [43] E. Riccardi, *Introduction to Chemometrics*, Forelesningsnotater (NTNU), TKJ4175 - Kjemometri 2019.
- [44] K. Kjeldahl og R. Bro, «Some common misunderstandings in chemometrics», *Journal of Chemometrics*, årg. 24, nr. 7-8, s. 558–564, 2010.
- [45] T. Hastie og R. Tibshirani (Statistical Learning), *StatsLearning Lect7c 110613*. adresse: [https://www.youtube.com/watch?v=S06JpVoNaA0&list=PL5-da3qGB5IA6E6ZNXu7dp89\\_uv8yocmf](https://www.youtube.com/watch?v=S06JpVoNaA0&list=PL5-da3qGB5IA6E6ZNXu7dp89_uv8yocmf), Videoforelesning tilknyttet boken: An introduction to statistical learning: with applications in R (accessed: 02.12.20).
- [46] E. C. Li-Chan, «Bioactive peptides and protein hydrolysates: research trends and challenges for application as nutraceuticals and functional food ingredients», *Current Opinion in Food Science*, årg. 1, s. 28–37, 2015.
- [47] H. Martens og M. Martens, «Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR)», *Food quality and preference*, årg. 11, nr. 1-2, s. 5–16, 2000.
- [48] T. Mehmood, K. H. Liland, L. Snipen og S. Sæbø, «A review of variable selection methods in partial least squares regression», *Chemometrics and Intelligent Laboratory Systems*, årg. 118, s. 62–69, 2012.
- [49] R. Sundseth Gjermstad, «Hydrolysater av restråstoff fra laks (*Salmo salar*) - effekten av antioksidanter på kvaliteten under lagring (Masteroppgave)», 2020, NTNU: Institutt for bioteknologi og matvitenskap.
- [50] K. Ryeng Simavik, «Hydrolyse av restråstoff fra regnbueørret (*Oncorhynchus mykiss*)- Effekten av antioksidanter på kvalitet under lagring (Masteroppgave)», 2020, NTNU: Institutt for bioteknologi og matvitenskap.
- [51] S. M. Pettersen, «Antioxidant Activity of Saithe (*Pollachius virens*) Backbone Hydrolysates and the Effect of Ultrafiltration. (Prosjektoppgave)», 2020, NTNU: Institutt for bioteknologi og matvitenskap.
- [52] M. V. Monslaup, «*In Vitro* Antioxidant Properties of Cod (*Gadus morhua*) Head Hydrolysates and Ultrafiltration Fractions», 2018, NTNU: Institutt for bioteknologi og matvitenskap.

- [53] M. Fossheim Ohnstad, «Marine Cod (*Gadus morhua*) Head Hydrolysates *in vitro* ACE inhibitory activity and investigation of structure-activity relationship», 2018, NTNU: Institutt for bioteknologi og matvitenskap.
- [54] V. Hammer Hjellnes, T. Rustad og E. Falch, «Enzymatic hydrolysis of pre-spawned and spawning saithe (*Pollachius virens*) in bioreactors and its potential for implementation in the Norwegian whitefish industry», Manuskript innsendt, NTNU.
- [55] A. Asfour, «Fish protein powders - characterisation of composition and properties», 2018, NTNU: Institutt for bioteknologi og matvitenskap.
- [56] A. Lamp, M. Kaltschmitt og O. Lüdtke, «Improved HPLC-method for estimation and correction of amino acid losses during hydrolysis of unknown samples», *Analytical biochemistry*, årg. 543, s. 140–145, 2018.
- [57] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. adresse: <https://www.R-project.org/>.
- [58] B. Mevik, R. Wehrens, K. H. Liland og M. B. Mevik, *Package 'pls'*, 2020.
- [59] Mattilsynet, *Matvaretabellen 2020*. adresse: [www.matvaretabellen.no](http://www.matvaretabellen.no), (accessed: 16.02.2021).
- [60] T. Damgaard, R. Lametsch og J. Otte, «Antioxidant capacity of hydrolyzed animal by-products and relation to amino acid composition and peptide size distribution», *Journal of Food Science and Technology*, årg. 52, nr. 10, s. 6511–6519, 2015.
- [61] K. S. Farvin, L. L. Andersen, J. Otte, H. H. Nielsen, F. Jessen og C. Jacobsen, «Antioxidant activity of cod (*Gadus morhua*) protein hydrolysates: Fractionation and characterisation of peptide fractions», *Food Chemistry*, årg. 204, s. 409–419, 2016.
- [62] Z. Khiari, D. Rico, A. B. Martin-Diana og C. Barry-Ryan, «Structure elucidation of ACE-inhibitory and antithrombotic peptides isolated from mackerel skin gelatine hydrolysates», *Journal of the Science of Food and Agriculture*, årg. 94, nr. 8, s. 1663–1671, 2014.
- [63] S. Saidi, A. Deratani, M.-P. Belleville og R. B. Amar, «Antioxidant properties of peptide fractions from tuna dark muscle protein by-product hydrolysate produced by membrane fractionation process», *Food Research International*, årg. 65, s. 329–336, 2014.
- [64] V. KRISTINOVÁ, «Evaluation of antioxidant effect using different analytical methods», ph.d.-avh., Diploma thesis. Faculty of Chemistry Institute of Food Science and . . . , 2008.
- [65] A. Freitas, J. Andrade, F. Silva, T. Rocha-Santos, A. Duarte og A. Gomes, «Antioxidative peptides: trends and perspectives for future research», *Current medicinal chemistry*, årg. 20, nr. 36, s. 4575–4594, 2013.

- [66] K. Korczek, J. Tkaczewska og W. Migdał, «Antioxidant and antihypertensive protein hydrolysates in fish products—a review», *Czech Journal of Food Sciences*, årg. 36, nr. 3, s. 195–207, 2018.
- [67] P. Minkiewicz, A. Iwaniak og M. Darewicz, «BIOPEP-UWM database of bioactive peptides: Current opportunities», *International journal of molecular sciences*, årg. 20, nr. 23, s. 5978, 2019.

# Tillegg A

## Rådata og beregninger

Datagrunnlaget for denne oppgaven baserte seg på relevante resultater for målte prøver (oppdrettslaks, regnbueørret, sei, torsk) fra master-og prosjektoppgaver som tidligere har blitt gjennomført ved Institutt for bioteknologi og matvitenskap ved NTNU. Felles for datagrunnlaget var tilgangen på ferdige master-og prosjektoppgaver i form av pdf-filer. For noen av variablene var også excelfiler med ulike rådata og beregninger tilgjengelig. Det ble også gjort målinger av ABTS antioksidantkapasitet for oppdrettslaks og regnbueørret av Peter Benjamin Kelley underveis i tidsperioden for denne oppgaven. Både tilgjengelig datagrunnlag og beregninger blir presentert videre i appendiks A. Datagrunnlaget varierte (målte variabler, type prøver, beregninger) mellom oppgavene, ettersom det er unike oppgaver som har blitt gjort med ulike mål. En stor del av denne oppgaven var derfor å samle data fra flere prøver/oppgaver, og samtidig sørge for at dataene var sammenlignbare på tvers av de ulike prøvene/oppgavene. Denne prosessen vil beskrives mer i kommende delkapitler, inkludert hvilke beregninger som ble gjort i andre oppgaver, og hvilke beregninger som ble gjort i denne oppgaven for å oppnå sammenlignbarhet. Mengde tioler (nmol/mg protein) og hydrolysegrad (%) var variabler som ble brukt direkte fra de relevante masteroppgavene i datamatrixene i denne oppgaven (uten endringer/beregninger), ettersom enhetene og beregningene som hadde blitt gjort i de relevante masteroppgavene var sammenlignbare.

### A.1 Oppdrettslaks

Målte variabler for oppdrettslaks ble inkludert i dataanalysen i denne oppgaven. Variablene ble målt gjennom ulike *in vitro*-analyser av Robin Sundseth Gjermstad i sin masteroppgave ved Institutt for bioteknologi og matvitenskap ved NTNU våren 2020 [49]. ABTS-målingene på samme prøver ble gjennomført av Peter Benjamin Kelley høsten 2020.

### Total aminosyresammensetning



Gjermstad gjorde beregninger på total aminosyresammensetning i sin oppgave i et excel regneark, fra umol aminosyre/l til mg aminosyre/g hydrolysat for tre paralleller. Fra dette regnearket ble gjennomsnittsverdien av mg aminosyre/g hydrolysat av parallellene brukt som utgangspunkt for denne oppgaven. Utgangspunktet var derfor gjennomsnittsverdien mg aminosyre/g hydrolysat for alle kontrollprøvene med ulik lagringsperiode (0, 1, 2, 4, 6 måneder).

I denne oppgaven ble total aminosyresammensetning omgjort til prosent, med utgangspunkt i gjennomsnittlig mg aminosyre/g hydrolysat og sum av gjennomsnittlig mg aminosyrer/g hydrolysat:

$$\% \text{ aminosyre} = \frac{100 \cdot \text{mg aminosyre/g hydrolysat}}{\text{sum mg aminosyrer/g hydrolysat}} \quad (\text{A.1})$$

En eksempelberregning av % Asp for den ene kontrollprøven (lagring 0 måneder) er vist her:

$$\% \text{ Asp} = \frac{100 \cdot 49.99 \text{ mg Asp/g hydrolysat}}{468.44 \text{ mg aminosyrer/g hydrolysat}} = 10.67\% \quad (\text{A.2})$$

Datascriptet håndterer kolonner som variabler og rader som prøver. Ettersom dette ikke var tilfellet for dataene ble det gjort en transponering av dataene i excel for å oppnå rett form før analyse i datascriptet. Etter transponering ble total aminosyresammensetning (%) for oppdrettslaks limt inn i datamatriser med de andre prøvene og variablene.

## Sammensetning av frie aminosyrer

Gjermstad gjorde beregninger for frie aminosyrer i sin oppgave i et excelark, fra umol aminosyre/l til mg/g hydrolysat. I denne oppgaven ble beregningene gjort på nytt av flere grunner. For det første så måtte molekylvekten oppdateres til å inkludere vann (frie aminosyrer). For det andre så var ikke mg/g hydrolysat beregnet for alle parallellene i tildelt regneark. Fra rådata umol/l ble molekylvekt brukt til å beregne frie aminosyrer (ug/ml) for to paralleller per ekstrakt (to ekstrakter per kontrollprøve):

$$\text{ug aa/mL} = \frac{M_w \text{ (g/mol)} \cdot M_{\text{HPLC}} \text{ (umol/l)}}{1000} \quad (\text{A.3})$$

Hvor  $M_w$  er molekylvekten til gitt aminosyre inkludert vann og  $M_{\text{HPLC}}$  er detektert mengde aminosyre fra HPLC (rådata). Videre ble mg frie aminosyrer/g hydrolysat beregnet:

$$\text{mg aa/g hydrolysat} = \frac{\text{ug aa/mL} \cdot 1.25 \cdot V \text{ (mL)} \cdot F}{1 \cdot 1000 \cdot m_s \text{ (g)}} \quad (\text{A.4})$$

Hvor  $A$  er mengden frie aminosyrer (ug/ml),  $V$  viser til volum av destillert vann (mL) for å lage proteinekstrakt,  $m_s$  viser til innveid mengde hydrolysat (g) og  $F$  viser til fortynningsfaktor for prøvene. En eksempelberregning er

vist her for en parallell for en av ekstraktene til kontrollprøve (lagring 0 måneder):

$$\text{ug Asp/mL} = \frac{133 \text{ g/mol} \cdot 6.54 \text{ umol/l}}{1000} = 0.87 \text{ ug Asp/mL} \quad (\text{A.5})$$

$$\text{mg Asp/g hydrolysat} = \frac{0.87 \text{ ug Asp/mL} \cdot 1.25 \cdot 10\text{mL} \cdot 50}{1 \cdot 1000 \cdot 0.5004\text{g}} = 1.09 \text{ mg Asp/g hydrolysat} \quad (\text{A.6})$$

Gjennomsnittsverdien mg aminosyre/g hydrolysat av alle parallellene ble så beregnet, med unntak av parallell 2 for ekstrakt 2 ettersom den avvek betraktelig fra de andre parallellene (kontrollprøve 0 måneder lagring). Standardavvik ble beregnet med excelfunksjonen STDAV.S.

Gjennomsnittlig frie aminosyrer (mg aminosyre/g hydrolysat) per kontrollprøve (0, 1, 2, 4, 6 måneder lagring) ble så omgjort til prosent med utgangspunkt i mg aminosyre/g hydrolysat og sum mg aminosyre/g hydrolysat for frie aminosyrer. Det ble så gjort en transponering av dataene så rader tilsvare prøvenavn og kolonner tilsvare frie aminosyrer (%), før dataene kunne limes inn i matriser med flere prøver og variabler.

## ABTS antioksidantkapasitet

Peter Benjamin Kelley gjorde ABTS-målingene av oppdrettslaks i samme tidsperiode som denne oppgaven, og bidro med rådata. Disse dataene ble brukt i denne masteroppgaven, ved å lage standardkurve og gjøre beregninger til umol PG/g hydrolysat. Kelley lagde standardløsninger av propylgallat (PG) med kjente konsentrasjoner (0, 10, 20, 30, 40, 50 umol/L) og målte tilhørende absorbans ved 734 nm. Absorbans (y) ble plottet som en funksjon av konsentrasjon propylgallate umol/L (x). Standardkurven er vist i figur A.1. Ved lineær regresjon ble en trendlinje tilpasset dataene og følgende ligning funnet:

$$y = -0.0069x + 0.6504 \quad (\text{A.7})$$

Ved å løse ligning A.7 for x ble dette følgende ligning:

$$x \text{ (umol/L)} = \frac{y - 0.6504}{-0.0069} \quad (\text{A.8})$$

Ligning A.8 kan brukes til å finne antioksidantkapasitet i form av ekvivalent konsentrasjon av PG fra målte absorbansverdier. For å finne umol PG/g hydrolysat ble følgende beregninger gjort:

$$\text{umol PG/mL} = \frac{x \text{ (umol PG/L)}}{1000\text{mL/L}} \quad (\text{A.9})$$

$$\text{umol PG/g hydrolysat} = \frac{\text{umol PG/mL} \cdot F}{\frac{m_s \text{ (g)}}{V \text{ (mL)}}} \quad (\text{A.10})$$

Hvor  $m_s$  er innveid hydrolysat (g) oppløst i V (mL) vann, og F er fortynningsfaktoren. For laks: V= 10 mL og F= 25. Innveid laksehydrolysat (g) er vist i tabell A.1. Overnevnte beregninger ble gjort for hver parallell, før gjennomsnittsverdien av parallellene for hver prøve (umol PG/g hydrolysat) ble beregnet.

Tabell A.1: Oversikt over innveid laksehydrolysat for ABTS-målinger.

| <b>Prøvenavn laks:</b> | <b>Innveid hydrolysat (g)</b> |
|------------------------|-------------------------------|
| 0A                     | 0,5032                        |
| 0B                     | 0,4809                        |
| 1A                     | 0,5285                        |
| 1B                     | 0,5434                        |
| 2A                     | 0,606                         |
| 2B                     | 0,5218                        |
| 4A                     | 0,5302                        |
| 4B                     | 0,613                         |
| 6A                     | 0,4848                        |
| 6B                     | 0,5186                        |

En eksempelberegning er vist for parallell 0A-a (kontrollprøve 0 måneder lagring):

$$x = \frac{0.440 - 0.6504}{-0.0069} = 30.5 \text{ umol PG/L} \quad (\text{A.11})$$

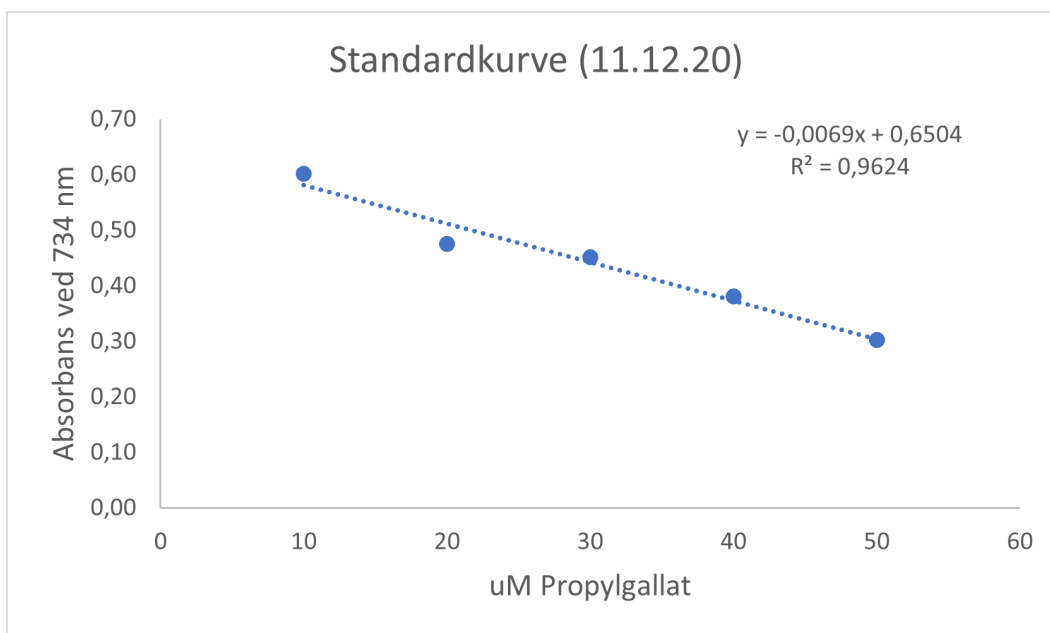
$$\text{umol PG/mL} = \frac{30.5 \text{ umol PG/L}}{1000\text{mL/L}} = 0.0305 \text{ umol PG/mL} \quad (\text{A.12})$$

$$\text{umol PG/g hydrolysat} = \frac{0.0305 \text{ umol PG/mL} \cdot 25}{\frac{0.5032 \text{ g}}{10 \text{ mL}}} = 15.1 \text{ umol PG/g hydrolysat} \quad (\text{A.13})$$

Tabell A.2 viser rådata for standardkurven, det vil si de kjente konsentrasjonene av PG og målte absorbansverdier ved 734 nm. Standardkurven er så vist i figur A.1.

Tabell A.2: Rådata for standardkurven i form av målte absorbansverdier for kjente konsentrasjoner av PG.

| <b>Absorbans ved 734 nm:</b> |          |          |          |                      |
|------------------------------|----------|----------|----------|----------------------|
| <b>uM PG</b>                 | <b>a</b> | <b>b</b> | <b>c</b> | <b>Gjennomsnitt:</b> |
| 0                            | 0,7      |          |          | 0,7000               |
| 10                           | 0,602    | 0,6      | 0,603    | 0,6017               |
| 20                           | 0,475    | 0,461    | 0,49     | 0,4753               |
| 30                           | 0,453    | 0,452    | 0,449    | 0,4513               |
| 40                           | 0,383    | 0,382    | 0,376    | 0,3803               |
| 50                           | 0,303    | 0,299    | 0,304    | 0,3020               |



Figur A.1: Standardkurve hvor absorbans ved 734 nm er plottet som en funksjon av konsentrasjon av propylgallat (uM eller umol/L). Lineær trendlinje viser:  $y = -0,0069x + 0,6504$  og  $R^2 = 0,9624$ .

De målte absorbansverdiene for kontrollprøvene av laks (0-6 måneder lagring) er vist i tabell A.3. Resultatet av beregningene er vist i tabell A.4. Gjennomsnittsverdiene (umol PG/g hydrolysat) av parallellene for de ulike prøvene er vist i tabell A.5.

Tabell A.3: Målte absorbansverdier ved 734 nm for kontrollprøvene av laks (0-6 måneder lagring).

| Prøver: | Absorbans ved 734 nm: |       |       |
|---------|-----------------------|-------|-------|
|         | a                     | b     | c     |
| 0A      | 0,440                 | 0,464 | 0,453 |
| 0B      | 0,400                 | 0,414 | 0,411 |
| 1A      | 0,387                 | 0,388 | 0,382 |
| 1B      | 0,376                 | 0,396 | 0,384 |
| 2A      | 0,376                 | 0,386 | 0,384 |
| 2B      | 0,430                 | 0,420 | 0,433 |
| 4A      | 0,386                 | 0,387 | 0,387 |
| 4B      | 0,405                 | 0,402 | 0,402 |
| 6A      | 0,363                 | 0,362 | 0,357 |
| 6B      | 0,413                 | 0,382 | 0,403 |

Tabell A.4: Resultatene av beregningene som ble gjort for alle parallellene av de ulike lakseprøvene

| Prøver: | Ekvivalent PG kons (umol/L): |      |      | umol PG/mL: |        |        | umol PG/g hydrolysat : |      |      |
|---------|------------------------------|------|------|-------------|--------|--------|------------------------|------|------|
|         | a                            | b    | c    | a           | b      | c      | a                      | b    | c    |
| 0A      | 30,5                         | 27,0 | 28,6 | 0,0305      | 0,0270 | 0,0286 | 15,1                   | 13,4 | 14,2 |
| 0B      | 36,3                         | 34,3 | 34,7 | 0,0363      | 0,0343 | 0,0347 | 18,9                   | 17,8 | 18,0 |
| 1A      | 38,2                         | 38,0 | 38,9 | 0,0382      | 0,0380 | 0,0389 | 18,1                   | 18,0 | 18,4 |
| 1B      | 39,8                         | 36,9 | 38,6 | 0,0398      | 0,0369 | 0,0386 | 18,3                   | 17,0 | 17,8 |
| 2A      | 39,8                         | 38,3 | 38,6 | 0,0398      | 0,0383 | 0,0386 | 16,4                   | 15,8 | 15,9 |
| 2B      | 31,9                         | 33,4 | 31,5 | 0,0319      | 0,0334 | 0,0315 | 15,3                   | 16,0 | 15,1 |
| 4A      | 38,3                         | 38,2 | 38,2 | 0,0383      | 0,0382 | 0,0382 | 18,1                   | 18,0 | 18,0 |
| 4B      | 35,6                         | 36,0 | 36,0 | 0,0356      | 0,0360 | 0,0360 | 14,5                   | 14,7 | 14,7 |
| 6A      | 41,7                         | 41,8 | 42,5 | 0,0417      | 0,0418 | 0,0425 | 21,5                   | 21,6 | 21,9 |
| 6B      | 34,4                         | 38,9 | 35,9 | 0,0344      | 0,0389 | 0,0359 | 16,6                   | 18,8 | 17,3 |

Tabell A.5: Oversikt over gjennomsnittsverdiene (umol PG/g hydrolysat) av parallellene for de ulike kontrollprøvene av laksehydrolysat (0-6 måneder lagring). Standardavvikene ble beregnet med Excel-funksjonen STDAV.S.

| Laks  | Snitt umol PG/g hydrolysat | Stdavvik |
|-------|----------------------------|----------|
| 0 mnd | 16,2                       | 2,3      |
| 1 mnd | 17,9                       | 0,5      |
| 2 mnd | 15,8                       | 0,5      |
| 4 mnd | 16,3                       | 1,9      |
| 6 mnd | 19,6                       | 2,4      |

## A.2 Regnbueørret

Målte variabler for regnbueørret ble inkludert i dataanalysen i denne oppgaven. Variablene ble målt gjennom ulike *in vitro*-analyser av Kristina Ryeng Simavik i sin masteroppgave ved institutt for bioteknologi og matvitenskap ved NTNU våren 2020 [50]. ABTS-målingene på samme prøver ble gjennomført av Peter Benjamin Kelley høsten 2020.

### Total aminosyresammensetning

Simavik gjorde også beregninger på total aminosyresammensetning i sin oppgave i et excel regneark, fra nmol aminosyre/ml til mg aminosyre/g hydrolysat for tre paralleller. I denne oppgaven ble det gjort noen endringer i dette regnearket. Fra før ble kun en innveid mengde hydrolysat (g) brukt i beregningene for alle parallellene av kontrollprøve (lagring 0 måneder), selv om det var oppgitt ulik innveid mengde hydrolysat per parallell. Dette var ikke tilfellet for de andre prøvene, hvor innveid mengde hydrolysat (g) for hver parallell ble brukt. Dette ble korrigert slik at innveid mengde hydrolysat (g) for hver parallell ble brukt i beregningene av mg aminosyre/g hydrolysat for gitt parallell. Fra regnearket ble så gjennomsnittsverdien av total aminosyresammensetning (mg aminosyre/g hydrolysat) for kontrollprøvene med ulik lagringstid (0, 2, 4, 6 måneder) brukt videre. Parallell 2 for kontrollprøve (lagring 2 måneder) avvek betraktelig fra de andre parallellene, og ble derfor ikke inkludert i gjennomsnittsverdien.

Total aminosyresammensetning ble så omgjort fra mg aminosyre/g hydrolysat til prosent på samme måte som for oppdrettslaks (se ligning A.1). Det ble så gjort en transponering av dataene slik at prøvenavnene tilsvarte rader og total aminosyresammensetning tilsvarte kolonner. Total aminosyresammensetning (%) for ørret ble så limt inn i datamatriser med de andre prøvene og variablene.

### Sammensetning av frie aminosyrer

Simavik gjorde beregninger for frie aminosyrer i sin oppgave i et excelark fra  $\mu\text{mol}$  aminosyre/l til mg aminosyre/g hydrolysat. For denne oppgaven ble det gjort noen justeringer i regnearket. Innveid hydrolysat (g) brukes i beregningene, og det var fra før angitt vekt for hvert ekstrakt. Først ble beregningene oppdatert til å bruke angitt vekt for paralleller av tilhørende ekstrakt. Molekylvekten ble oppdatert til å inkludere vann (frie aminosyrer). Med disse endringene ble beregningene gjort automatisk og regnearket oppdatert (slik som vist i ligning A.3 og A.4).

Fra mg aminosyre/g hydrolysat ble gjennomsnittet av alle fire parallellene (to paralleller per ekstrakt) beregnet for alle kontrollprøvene (0 måneder, 2 måneder, 4 måneder, 6 måneder). Deretter ble sammensetning av frie aminosyrer omgjort til prosent med utgangspunkt i mg aminosyre/g hydrolysat og sum mg aminosyre/g hydrolysat for frie aminosyrer. Det ble så gjort en transponering av dataene til rett format for datascriptet, som beskrevet tidligere.

## ABTS antioksidantkapasitet

Peter Benjamin Kelley gjorde ABTS-målingene av ørret i samme tidsperiode som denne oppgaven, og bidro med rådata. Disse dataene ble brukt i denne oppgaven på samme måte som beskrevet for laks, først ved å lage en standardkurve for så å gjøre beregninger frem til gjennomsnittsverdien  $\mu\text{mol PG/g}$  hydrolysat av alle parallellene for kontrollprøve (6 måneder lagring). I beregningene ble følgende verdier brukt:  $m_s = 0.5\text{g}$ ,  $V = 10\text{ mL}$  og  $F = 25$ . Standardkurven som viser absorbans ( $y$ ) som funksjon av kjent konsentrasjon propylgallat  $\mu\text{mol/L}$  ( $x$ ) er vist i figur A.2. Følgende ligning ble funnet ved lineær regresjon:

$$y = -0.0074x + 0.6641 \quad (\text{A.14})$$

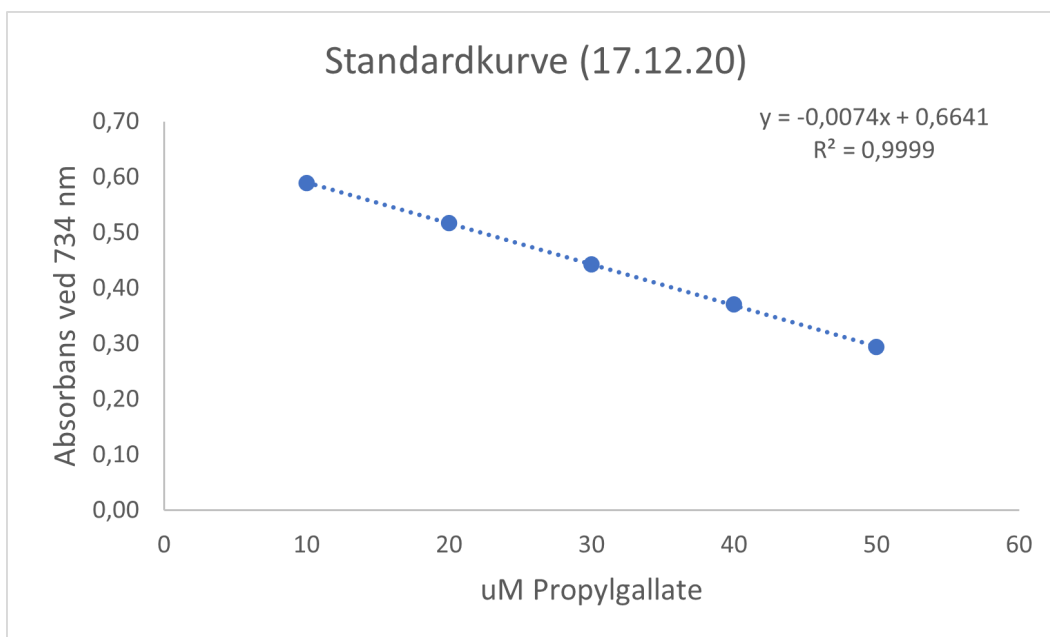
Som også tilsvareer ligningen:

$$x (\mu\text{mol/L}) = \frac{y - 0.6641}{-0.0074} \quad (\text{A.15})$$

En oversikt over rådata for standardkurven er vist i tabell A.6, med selve standardkurven vist i påfølgende figur A.2.

Tabell A.6: Rådata for standardkurven i form av målte absorbansverdier for kjente konsentrasjoner av PG.

| <b>Absorbans ved 734 nm:</b>       |          |          |          |                      |
|------------------------------------|----------|----------|----------|----------------------|
| <b><math>\mu\text{M PG}</math></b> | <b>a</b> | <b>b</b> | <b>c</b> | <b>Gjennomsnitt:</b> |
| 0                                  | 0,673    |          |          | 0,6730               |
| 10                                 | 0,593    | 0,588    | 0,588    | 0,5897               |
| 20                                 | 0,518    | 0,516    | 0,516    | 0,5167               |
| 30                                 | 0,441    | 0,449    | 0,440    | 0,4433               |
| 40                                 | 0,367    | 0,377    | 0,367    | 0,3703               |
| 50                                 | 0,290    | 0,289    | 0,303    | 0,2940               |



Figur A.2: Standardkurve hvor absorbans ved 734 nm er plottet som en funksjon av konsentrasjon av propylgallat (uM eller umol/L). Lineær trendlinje viser:  $y = -0,0074x + 0,6641$  og  $R^2 = 0,9999$ .

De målte absorbansverdiene for ørrethydrolysatet er vist i tabell A.7, og resultatet av beregningene er vist i tabell A.8 og A.9. Parallell 6B-a avviker betydelig fra b og c, og derfor ble denne parallellen utelatt fra beregningen av gjennomsnittsverdien av umol PG/g hydrolysat for 6 måneders lagret kontrollprøve av ørrethydrolysat.

Tabell A.7: Målte absorbansverdier ved 734 nm for 6 måneders lagret kontrollprøve av ørret.

| Absorbans ved 734 nm: |       |       |       |
|-----------------------|-------|-------|-------|
| Prøver:               | a     | b     | c     |
| 6A                    | 0,500 | 0,470 | 0,433 |
| 6B                    | 0,608 | 0,516 | 0,514 |

Tabell A.8: Resultatene av beregningene som ble gjort for alle parallellene av 6 måneders lagret kontrollprøve av ørrethydrolysat.

| Prøver: | Ekvivalent PG kons (umol/L): |      |      | umol PG/mL : |        |        | umol PG/g hydrolysat: |      |      |
|---------|------------------------------|------|------|--------------|--------|--------|-----------------------|------|------|
|         | a                            | b    | c    | a            | b      | c      | a                     | b    | c    |
| 6A      | 22,2                         | 26,2 | 31,2 | 0,0222       | 0,0262 | 0,0312 | 11,1                  | 13,1 | 15,6 |
| 6B      | 7,58                         | 20,0 | 20,3 | 0,00758      | 0,0200 | 0,0203 | 3,79                  | 10,0 | 10,1 |

Tabell A.9: Gjennomsnittsverdien av umol PG/g hydrolysat for 6 måneders lagret kontrollprøve av ørrethydrolysat, inkludert standardavvik.

| Ørret | Snitt umol PG/g hydrolysat | Stdavvik: |
|-------|----------------------------|-----------|
| 6 mnd | 12,0                       | 2,38      |



## A.3 Sei

Målte variabler for sei ble inkludert i dataanalysen i denne oppgaven. Variablene ble målt gjennom ulike *in vitro*-analyser av Stine Marie Pettersen i sin prosjektoppgave ved institutt for bioteknologi og matvitenskap ved NTNU høsten 2020 [51].

### Total aminosyresammensetning

Pettersen gjorde beregninger av total aminosyresammensetning i sin oppgave i et eget excel regneark, fra  $\mu\text{mol/l}$  til  $\text{mg/g}$  hydrolysat for tre paralleller. Pettersen hadde valgt å bruke molekylvekt for aminosyrer som ikke er bundet i en peptidkjede (med vann). For å sørge for sammenlignbarhet mellom de ulike typene fisk, ble det korrigert til molekylvekt for aminosyrer som er bundet i en peptidkjede (uten vann). Fra regnearket ble så oppdatert gjennomsnitt av total aminosyresammensetning ( $\text{mg aminosyre/g hydrolysat}$ ) for prøvene BOI, BOII, BJI og BJII brukt videre.

Total aminosyresammensetning ble så omgjort fra  $\text{mg aminosyre/g hydrolysat}$  til prosent på samme måte som for oppdrettslaks og ørret (se ligning A.1). Det ble så gjort en transponering av dataene slik at prøvenavnene tilsvarte rader og total aminosyresammensetning tilsvarte kolonner. Total aminosyresammensetning (%) for ørret ble så limt inn i datamatriser med de andre prøvene og variablene.

### Sammensetning av frie aminosyrer

Pettersen gjorde beregninger av sammensetning frie aminosyrer i eget excelark, fra  $\mu\text{mol aminosyre/L}$  til  $\text{mg aminosyre/g hydrolysat}$ . For denne oppgaven ble molekylvekten oppdatert til å inkludere vann (frie aminosyrer), og beregningene ble oppdatert automatisk i regnearket deretter. Dette inkluderte gjennomsnittet av parallellene for hver prøve (BOI, BOII, BJI, BJII). Gjennomsnittsverdiene ble så omgjort til prosent med utgangspunkt i  $\text{mg aminosyre/g hydrolysat}$  og sum  $\text{mg aminosyre/g hydrolysat}$  for frie aminosyrer. I etterkant av dette ble det gjort en transponering av dataene til rett format for datascriptet, som beskrevet tidligere.

### ABTS antioksidantkapasitet

Pettersen gjennomførte ABTS-målinger i sin oppgave, og gjorde beregninger på ABTS antioksidantkapasitet fra rådata (absorbansmålinger) til  $\mu\text{mol PG/g hydrolysat}$  i et eget excelark. De originale resultatene fra Stine er vist i tabell A.10. De originale resultatene ble transponert og limt inn i datamatriser før dataanalyse.

Tabell A.10: Oversikt over umol PG/g hydrolysat for seihydrolysatene BOI, BOII, BJI og BJII.

| <b>Prøver:</b> | <b>umol PG/g hydrolysat</b> |
|----------------|-----------------------------|
| BOI            | 60.4                        |
| BOII           | 59.8                        |
| BJI            | 61.0                        |
| BJII           | 65.8                        |

## A.4 Torsk

Målte variabler for torsk ble inkludert i dataanalysen i denne oppgaven. Variablene ble målt gjennom ulike *in vitro*-analyser av Marte Vilde Monslaup, Margrethe Fossheim Ohnstad og Ayat Asfour i sine masteroppgaver ved institutt for bioteknologi og matvitenskap ved NTNU våren 2018 [52], [53], [55].

### Total aminosyresammensetning

Total aminosyresammensetning for torskehydrolysater var tilgjengelig kun som rådata i form av umol aminosyre/l fra HPLC-analyser som ble gjennomført av Marte Vilde Monslaup og Margrethe Fossheim Ohnstad i sine oppgaver. I denne oppgaven ble mg aminosyre/g hydrolysat beregnet fra disse rådataene. Molekylvekt for aminosyrer i en peptidkjede (uten vann) ble brukt i beregningene, slik som for de andre fisketyperne. Først ble ug aminosyre/mL beregnet på følgende måte:

$$\text{ug aa/mL} = \frac{M_w \text{ (g/mol)} \cdot M_{\text{HPLC}} \text{ (umol/l)}}{1000} \quad (\text{A.16})$$

Hvor  $M_w$  er molekylvekten (g/mol) til en gitt aminosyre bundet i en peptidkjede (uten vann) og  $M_{\text{HPLC}}$  er total aminosyresammensetning (umol/l) målt fra HPLC. Videre ble mg aminosyre/g hydrolysat beregnet:

$$\text{mg aa/g hydrolysat} = \frac{\text{ug aa/mL} \cdot V \text{ (mL)} \cdot F}{1000 \cdot m_s \text{ (g)}} \quad (\text{A.17})$$

Hvor  $V$  viser til volum (mL) etter at vann ble tilført filtratet,  $F$  viser til fortynningsfaktoren og  $m_s$  er innveid hydrolysat (g). Marte Vilde Monslaup og Margrethe Fossheim Ohnstad gjorde HPLC-analyser på samme torskhydrolysater. Rådata for T1 og T9 så ut til å ha blitt målt av Margrethe Fossheim Ohnstad, og T5 av Marte Vilde Monslaup (fra tilgjengelig rådata). Informasjon om volum, fortynningsfaktor og innveid hydrolysat til beregningene over ble funnet i masteroppgavene til Ohnstad og Monslaup. For Ohnstad ble følgende tall brukt:  $V = 10$  mL,  $F = 500$  og  $m_s = 0,05$  g. For Monslaup ble følgende tall brukt:  $V = 20$  mL,  $F = 1000$  og  $m_s = 0,1$  g. Uten tilgang på nøyaktig innveid vekt vil det kunne være noen små variasjoner i resultatet. En eksempelberegning er vist for parallell 1 av hydrolysat T1:

$$\text{ug Asp/mL} = \frac{115 \text{ g/mol} \cdot 8.6490 \text{ umol/l}}{1000} = 0.9946 \text{ ug Asp/mL} \quad (\text{A.18})$$

$$\text{mg Asp/g hydrolysat} = \frac{0.9946 \text{ ug Asp/mL} \cdot 10 \text{ mL} \cdot 500}{1000 \cdot 0,05 \text{ g}} = 99,4640 \text{ mg Asp/g hydrolysat} \quad (\text{A.19})$$

I etterkant av disse beregningene ble gjennomsnittlig total aminosyresammensetning beregnet for alle parallellene til hver prøve (T1, T5, T9). For hydrolysat T9 ble første parallell av totalt tre paralleller utelatt fra gjennomsnittet, ettersom den avvek betydelig fra de andre parallellene. Deretter ble total aminosyresammensetning omgjort til

prosent med utgangspunkt i mg aminosyre/g hydrolysat og sum mg aminosyrer/g hydrolysat for total aminosyre-sammensetning. Dataene ble så transponert til rett format for datascripnet, slik som beskrevet tidligere.

## Sammensetning av frie aminosyrer

Det var Marte Vilde Monslaup som gjennomførte HPLC-analyser av frie aminosyrer på torskhydrolysater gjennom sin oppgave. Monslaup hadde gjort beregninger på frie aminosyrer i et eget excelark i egen oppgave, men det var noen forskjeller i beregningene for torsk og for de andre fisketypene/oppgavene. Derfor måtte beregningene gjøres på nytt for å sikre sammenlignbarhet mellom prøvene. I denne oppgaven ble derfor rådata for frie aminosyrer umol aminosyre/l beregnet til mg aminosyre/g hydrolysat på nytt basert på informasjon om volum, fortynning og innveid hydrolysat i oppgaven til Marte Vilde Monslaup og Margrethe Fossheim Ohnstad. Rådata ble hentet fra excelarket til Marte Vilde Monslaup, og så ble sammensetning av frie aminosyrer og aminosyre/mL beregnet på samme måte som vist i ligning A.3, og mg aminosyre/g hydrolysat som i ligning A.4. Molekylvekt for frie aminosyrer (inkludert vann) ble brukt i beregningene slik som for de andre fisketypene. Tallene som ble brukt i beregningene ble hentet fra Monslaup og Ohnstad sine oppgaver:  $V = 10$  mL,  $F = 25$  og  $m_s = 0.02$ g. Slik som for TAA var det heller ikke her tilgang på nøyaktig innveid vekt, så det vil kunne være noen små variasjoner i resultatet.

Etter at mg aminosyre/g hydrolysat ble beregnet for alle parallellene, så ble gjennomsnittlig sammensetning av frie aminosyrer beregnet for alle prøvene (T1, T5 og T9). Sammensetning av frie aminosyrer ble så omgjort til prosent med utgangspunkt i mg aminosyre/g hydrolysat og sum mg aminosyrer/g hydrolysat for frie aminosyrer. Dataene ble så transponert til rett format for datascripnet, slik som beskrevet for de andre prøvene.

## ABTS antioksidantkapasitet

Marte Vilde Monslaup hadde gjort ABTS-målinger i sin oppgave. Resultatene viste umol PG/g løselig protein. For å oppnå sammenlignbarhet mellom prøvene ble resultatet beregnet om til umol PG/g hydrolysat på følgende måte:

$$\text{umol PG/g hydrolysat} = \text{umol PG/g protein} \cdot \frac{\text{g protein/mL}}{\frac{m_s (\text{g})}{V (\text{mL})}} \quad (\text{A.20})$$

Hvor umol PG/g protein viser til resultatet av ABTS-målingene og beregningene til Monslaup og g protein/mL viser til proteinkonsentrasjonen funnet ved Lowry metoden.  $m_s$  viser til innveid torskhydrolysat (g) i  $V$  (mL) vann. Tallene som ble brukt ble hentet fra Monslaup sin masteroppgave:  $V = 10$  mL og  $m_s = 0.2$ g. Tabell A.11 viser tallene fra Marte Vilde Monslaup sin oppgave (umol PG/g protein) og resultatene av omberegningene som ble gjort i denne oppgaven til umol PG/g hydrolysat for torskhydrolysatene T1, T5 og T9.

Tabell A.11: Oversikt over umol PG/g protein (fra Marte Vilde Monslaup) og resultatene av omberegningene til umol PG/g hydrolysat for torskehydrolysatene T1, T5 og T9.

| <b>Prøver:</b> | <b>umol av PG/g protein</b> | <b>g protein/mL</b> | <b>umol av PG/mL</b> | <b>umol PG/g hydrolysat</b> |
|----------------|-----------------------------|---------------------|----------------------|-----------------------------|
| T1             | 37,5                        | 0,0107              | 0,400                | 20,0                        |
| T5             | 52,4                        | 0,00830             | 0,435                | 21,7                        |
| T9             | 58,3                        | 0,00637             | 0,371                | 18,6                        |

## Tillegg B

# Dataanalyse

### B.1 Pearson korrelasjonskoeffisient

Pearson korrelasjonskoeffisient er definert gjennom følgende ligning:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{B.1})$$

Som vist gjennom ligningen så er  $r$  enhetsløs, og vil være den samme uavhengig av enhetene til variablene  $x$  og  $y$ .

### B.2 Datascript

#### B.2.1 Pearson korrelasjonsmatrise

Figur B.1 viser R-scriptet som ble brukt for å beregne Pearson korrelasjon mellom variablene og visualisere det gjennom korrelasjonsmatrisen. Scriptet er for matrise 1, men samme oppsett ble brukt for de andre matrisene også. Kommentarer i scriptet er angitt ved ‡.

```

library(pls)
library(plsVarSel)
library(corrplot)
library(Cairo)
library(cairoDevice)
set.seed(8010)

#Import,view and adjust data set for use in R
library(readxl)
MatriselUo <- read_excel("C:\\\\Users\\\\\\Hanna\\\\\\Documents\\\\\\
\Masteroppgave2020\\\\\\Dataanalyse\\\\\\Data\\\\\\Data som skal brukes\\\\\\
\MatriserUorret024mnd.xlsx",
                        range = "B8:AG21")

View(MatriselUo)
attach(MatriselUo)
fix(MatriselUo) #view spreadsheet
MatriselUoA=MatriselUo[,2:32]
rownames(MatriselUoA)=MatriselUo$Prøvenavn #Making a new matrix where the
first column is the row names.
View(MatriselUoA)
fix(MatriselUoA)
Ferdig.MatriselUo=as.matrix(MatriselUoA) #converting data.table into a
matrix
View(Ferdig.MatriselUo)

Correlation=cor(Ferdig.MatriselUo, method = "pearson") #Scale free, will
not depend on units

setwd("C:/Users/Hanna/Documents/R/Masteroppgave/020221-EndeligeResultater/
Matrisel-uorret024mnd")
CairoPNG(filename = "Korrplott-MatriselUorret.png", width = 4500, height =
4500, pointsize = 12)
corrplot(Correlation, method = "circle", tl.col = "black", tl.cex =
6,cl.cex = 6)
dev.off()

```

Figur B.1: R datascrip for beregning og visualisering av korrelasjonsmatrisen til matrise 1.

## B.2.2 PLSR-analyse

Figur B.2, B.3 og B.4 viser R-scriptet for PLSR-analysen for matrise 1. Samme oppsett ble brukt for de andre matrisene. Koden inkluderer et forslag til PCA om flere Y-responser hadde blitt målt, det er angitt med ‡ for kommentar.

```

library(pls)
library(plsVarSel)
library(Cairo)

#Import,view and adjust data set for use in R
library(readxl)
MatriselUo <- read_excel("C:\\\\Users\\\\\\Hanna\\\\\\Documents\\\\\\
\\Masteroppgave2020\\\\\\Dataanalyse\\\\\\Data\\\\\\Data som skal brukes\\\\\\
\\MatriserUorret024mnd.xlsx",
                      range = "B8:AG21")

View(MatriselUo)
attach(MatriselUo)
fix(MatriselUo) #view spreadsheet
MatriselUoA=MatriselUo[,2:32]
rownames(MatriselUoA)=MatriselUo$Prøvenavn #Making a new matrix where the
first column is the row names.
View(MatriselUoA)
fix(MatriselUoA)
Ferdig.MatriselUo=as.matrix(MatriselUoA) #converting data.table into a
matrix
View(Ferdig.MatriselUo)

#Define X and Y in dataset
HydrolysatesX=Ferdig.MatriselUo[,1:30, drop=FALSE]
HydrolysatesY=Ferdig.MatriselUo[,31, drop=FALSE]
View(HydrolysatesX)
View(HydrolysatesY)

#View and learn about data set.
#study mean and variance of columns, see the variance of data
apply(Ferdig.MatriselUo,2,mean) #calculating mean of the columns in x, 2
indicates columns
apply(Ferdig.MatriselUo,2,var) #calculating variance of the columns
apply(Ferdig.MatriselUo,2,sd) # standard deviation

#visualize data
setwd("C:/Users/Hanna/Documents/R/Masteroppgave/020221-EndeligeResultater/
Matrisel-uorret024mnd")
CairoPNG(filename="Boksplott-orgvar.png", width = 10, height = 10, units =
'in', res = 300)
par(mar=c(10,3,1,1)) #set margins so full labels will show
boksplottX=boxplot(Ferdig.MatriselUo, col="steelblue", las=2) #visualize
the data
dev.off()

std.Hydrolysates=stdize(Ferdig.MatriselUo, center=TRUE, scale=TRUE)
std.HydrolysateX=std.Hydrolysates[,1:30, drop=FALSE]
std.HydrolysateY=std.Hydrolysates[,31, drop=FALSE]
View(std.HydrolysateX)
View(std.HydrolysateY)
View(std.Hydrolysates)

CairoPNG(filename="Boksplott-stdvar.png", width = 10, height = 10, units =
'in', res = 300)
par(mar=c(10,3,1,1))
boksplott.std.Hydrolysates=boxplot(std.Hydrolysates,col="steelblue",
las=2) #visualize the standardized data.
dev.off()
#PCA on Y variables (adjust if several models for Y)

```

Figur B.2: R datascrip for PLSR-analysen av matrise 1, del A.



```

#PCA.Y=prcomp(std.HydrolysateY, scale=FALSE, center=FALSE) #singular value
decomposition.
#summary(PCA.Y)
#plot(pc) #scree plot, variance explained
#PCA.Y #rotations and dard deviations
#scoreplot(PCA.Y,comps = 1:2, identify=TRUE,col="light blue") #get an
overview of the samples and Y
#corrplot(PCA.Y, comps=1:2, identify=TRUE, col="light blue") #look for
clustering/correlations and decide if separate regression models
#loadingplot(PCA.Y, comps=1:2, identify=TRUE, col="light blue",scatter =
TRUE)

#PLSR modell
set.seed(9050)
plsr.modell=plsr(umolPGperghydrolysate~., data=std.Hydrolysates,
scale=FALSE, center=FALSE, validation="LOO", jackknife=
TRUE,method="oscorespls")
summary(plsr.modell)#CV RMSEP given and explained variance components for
fitting
#explained variance by X and Y is given in the summary

#Validation and choosing number of PLS directions
validationplot(plsr.modell, val.type="MSEP", legendpos="top") #look for
the lowest cross validation error, an estimate for test MSE
selectNcomp(plsr.modell,"onesigma", plot = TRUE) #Choosing the number of
components for the first simplest model that is within one std from
optimal CV error

#OPPDATER ANTALL KOMPONENTER UNDER: VALGTE HER 3
#Resulting plots for interpretation
predplot(plsr.modell,ncomp=3,line=TRUE ) #predicted versus measured
values, the crossvalidated predictions, insert # of PLS directions

scoreplot(plsr.modell,comps=1:2,labels = "names", type="p", col="steel
blue")
abline(h=0,v=0) #adding vertical and horizontal line at x=0, y=0 in
scoreplot
scoreplot(plsr.modell,comps=1:3,labels = "names", type="p", col="steel
blue")

loadingplot(plsr.modell, comps=1:2,labels="names",pretty.xlabels =
FALSE,scatter=TRUE, col="steel blue")
abline(h=0,v=0)
loadingplot(plsr.modell, comps=1:3,labels="names",pretty.xlabels =
FALSE,scatter=TRUE, col="steel blue")

corrplot(plsr.modell, comps=1:2,
labels="names",plotx=TRUE,ploty=TRUE,type="p")
corrplot(plsr.modell, comps=1:3,
labels="names",plotx=TRUE,ploty=TRUE,type="p")

coefplot(plsr.modell,ncomp = 3,legendpos = "bottomright", se.whiskers =
TRUE,separate=TRUE,labels="names", type="p", pretty.xlabels = FALSE,
las=2) #Plotting the regression coefficients. se.whiskers: +/- one
estimated std.error, if cv with jackknifing. labels=names, x axis labels
are the variable names.
#coef gives the B in y=BX

#X,Y score plot
Xscores=scores(plsr.modell)

```

Figur B.3: R datascrip for PLSR-analysen av matrise 1, del B.

```

Yscores=Yscores(plsr.modell)
plot(Xscores[,3],Yscores[,3])

#VIP (variable importance for the projection)
vip=VIP(plsr.modell, 3) #Measure of importance of an X variable for X and
Y modelling
par(mar=c(8,5,1,1))
barplot(vip,ylab="VIP",col="steelblue",ylim=c(0,2.5),las=2) #barplot of
VIP

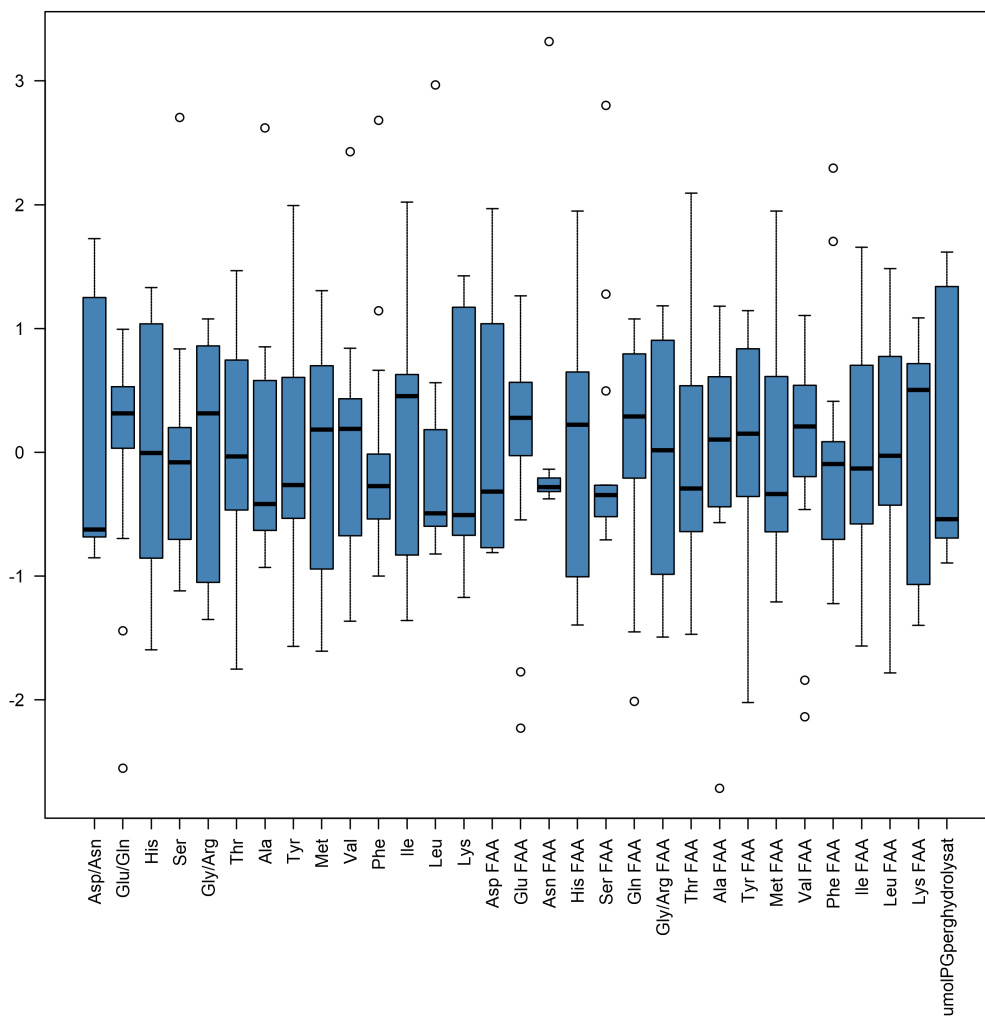
#Jack.test: gjør t-tester av reg.koeff. basert på jackknife variance
estimates.
jack.test(plsr.modell,ncomp=3, use.mean=FALSE)

```

Figur B.4: R datascrip for PLSR-analysen av matrise 1, del C.

### B.3 Matrise 1: TAA, FAA, ABTS antioksidantkapasitet

Figur B.5-B.11 viser diverse resultater av dataanalysen av matrise 1.



Figur B.5: Boksplott av matrise 1 etter standardisering av alle variabler.

```

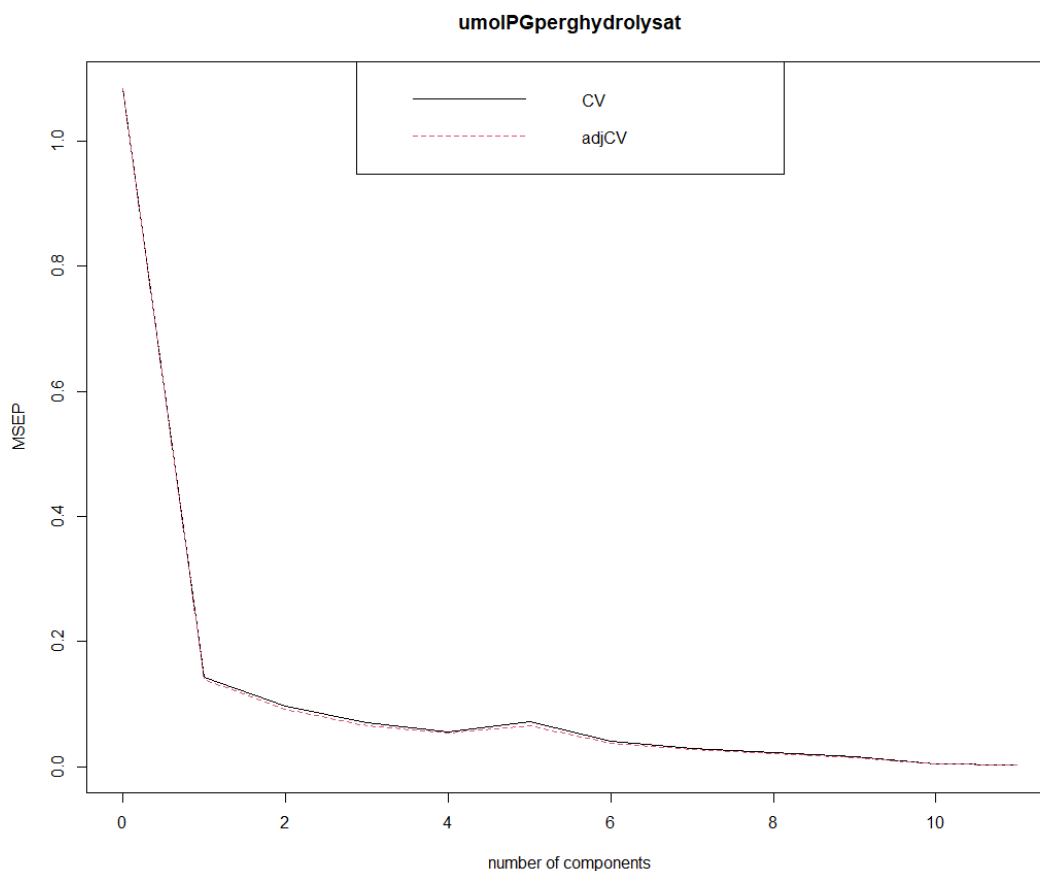
> set.seed(9050)
> plsr.modell=plsr(umolPGperghydrolysat~., data=std.Hydrolysat, scale=FALSE, center=FALSE, validation="LOO", jackknife= TRUE,method="oscorespls")
> summary(plsr.modell)#CV RMSEP given and explained variance components for fitting
Data: X dimension: 13 30
      Y dimension: 13 1
Fit method: oscorespls
Number of components considered: 11

VALIDATION: RMSEP
Cross-validated using 13 leave-one-out segments.
      (Intercept) 1 comps 2 comps 3 comps 4 comps 5 comps 6 comps 7 comps 8 comps 9 comps 10 comps 11 comps
CV          1.041  0.3775  0.3106  0.2639  0.2357  0.2686  0.2017  0.1701  0.1487  0.1271  0.06598  0.04687
adjCV       1.041  0.3731  0.3040  0.2558  0.2332  0.2563  0.1944  0.1648  0.1424  0.1217  0.06308  0.04480

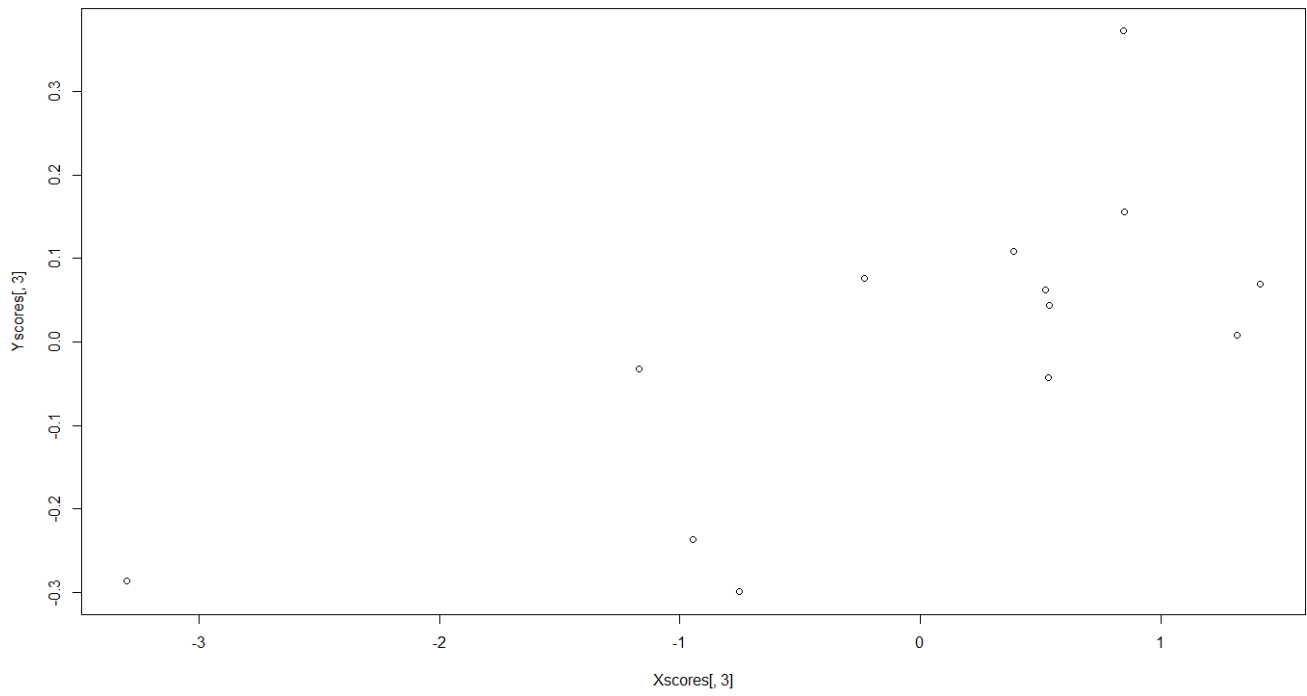
TRAINING: % variance explained
      1 comps 2 comps 3 comps 4 comps 5 comps 6 comps 7 comps 8 comps 9 comps 10 comps 11 comps
X          33.52  62.44  69.39  82.37  84.39  90.73  95.94  97.60  98.70  99.52  99.61
umolPGperghydrolysat 92.96  96.48  98.23  98.55  99.60  99.71  99.81  99.97  99.99  100.00  100.00
> |

```

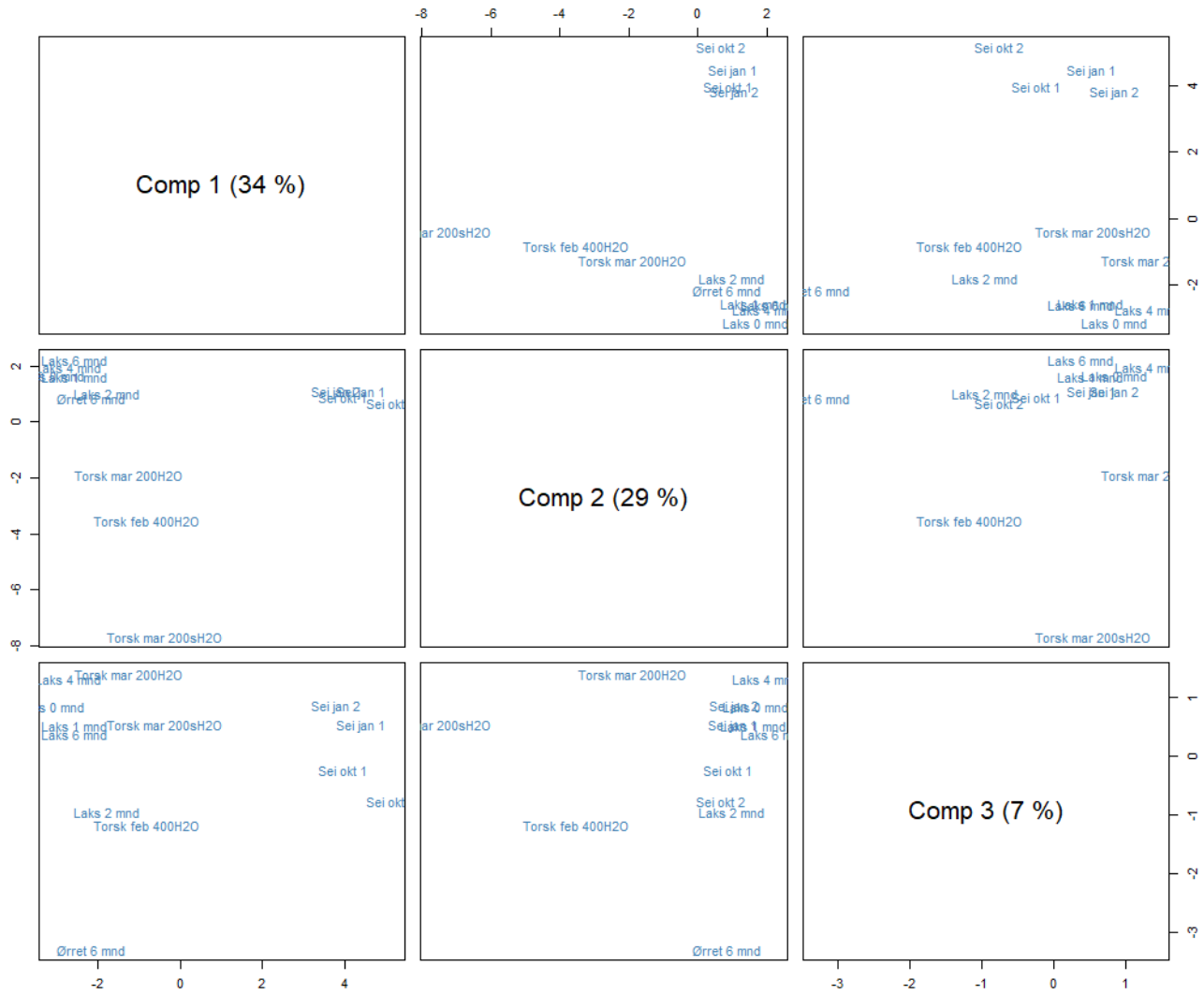
Figur B.6: Sammendrag av PLSR-analysen for matrise 1.



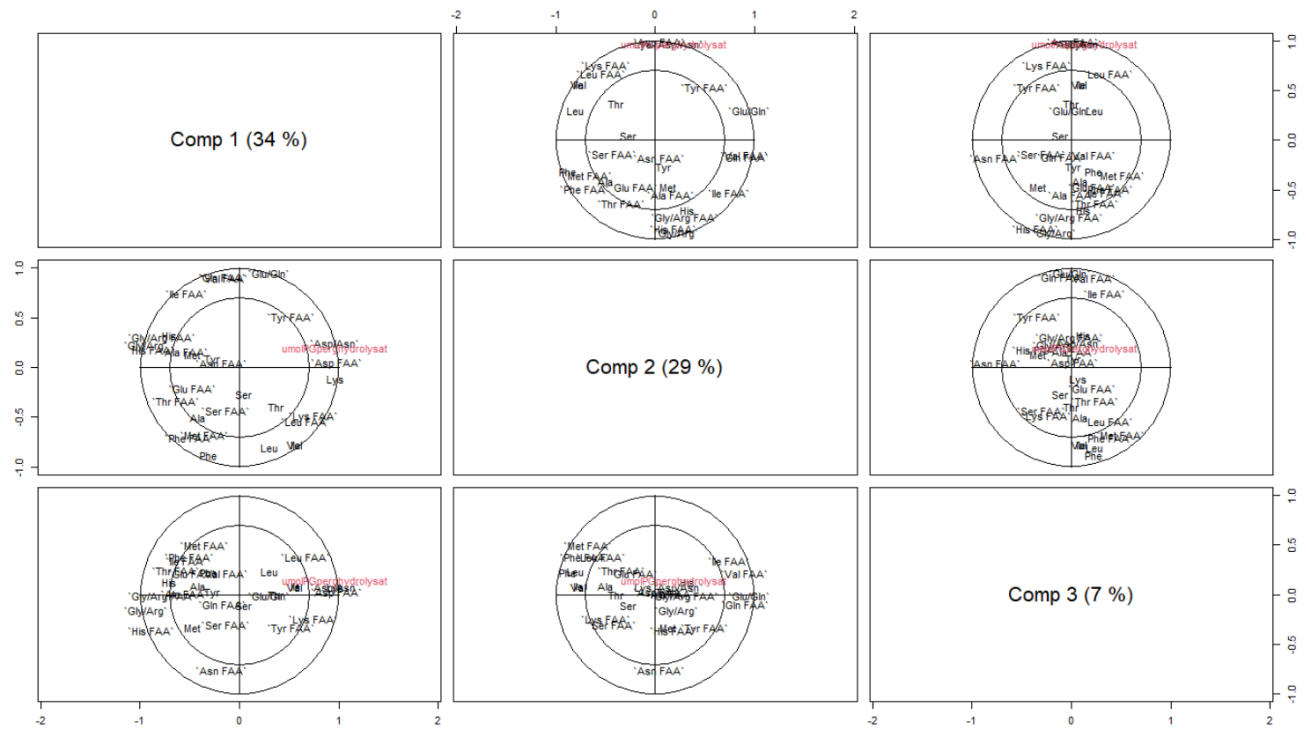
Figur B.7: MSEP for den kryssvaliderte modellen med ulikt antall komponenter for matrise 1.



Figur B.8: Plott av X og Y-scores for PLSR-modellen med tre komponenter for matrise 1.



Figur B.9: Scores-plott for alle komponentene i PLSR-modellen fra matrise 1.



Figur B.10: Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen fra matrise 1.

```

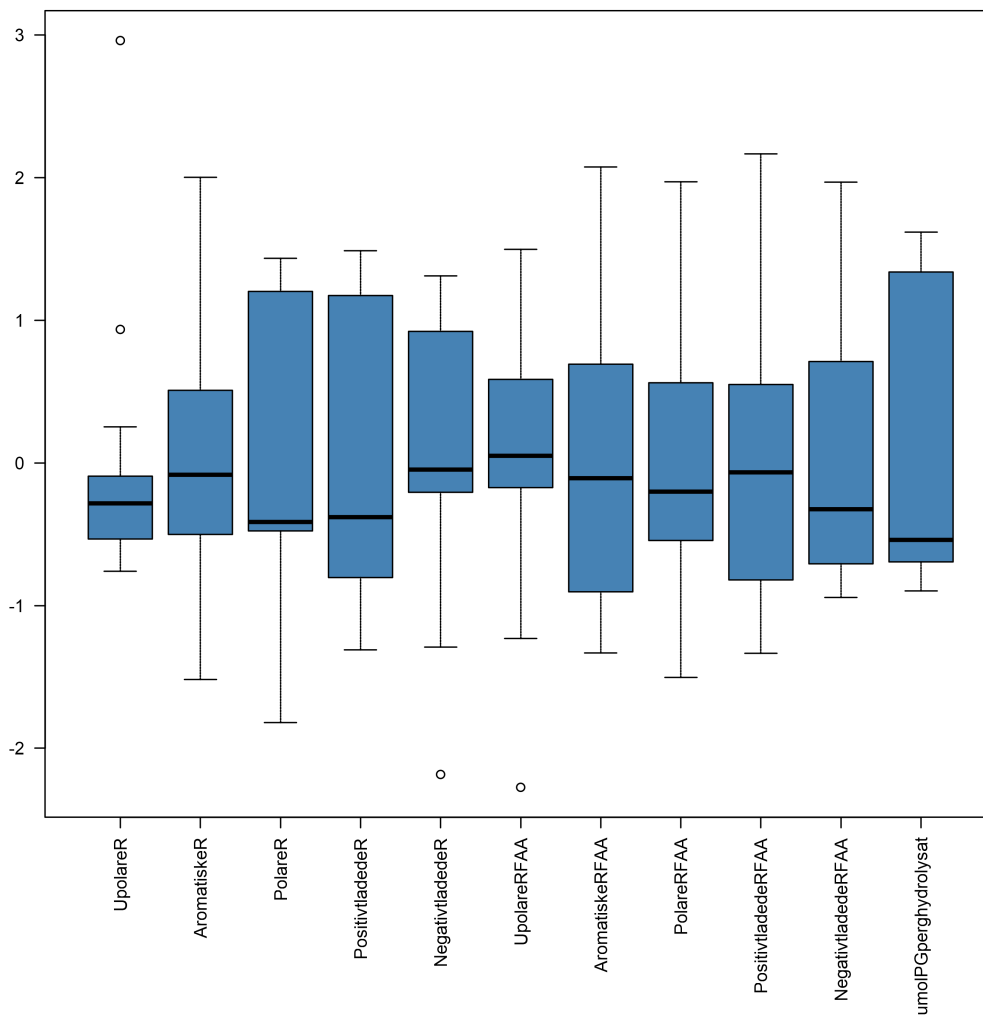
> jack.test(plsr.modell,ncomp=3, use.mean=FALSE) #replace n with #comp
Response umolPGperhydrolysat (3 comps):
      Estimate Std. Error Df t value Pr(>|t|)
`Asp/Asn`      0.11778509 0.01563827 12  7.5319 6.932e-06 ***
`Glu/Gln`      0.08106880 0.01145007 12  7.0802 1.282e-05 ***
His            -0.01771270 0.03529311 12 -0.5019 0.6248373
Ser            0.00066813 0.03118700 12  0.0214 0.9832599
`Gly/Arg`     -0.11499633 0.02214265 12 -5.1934 0.0002243 ***
Thr            0.00032133 0.04526540 12  0.0071 0.9944527
Ala           -0.07845176 0.02022347 12 -3.8792 0.0021909 **
Tyr           -0.00820528 0.02856380 12 -0.2873 0.7788123
Met           -0.04961701 0.02497400 12 -1.9867 0.0702687 .
Val            0.01143761 0.01337901 12  0.8549 0.4093550
Phe           -0.04135513 0.02670982 12 -1.5483 0.1475050
Ile            0.02922595 0.01199473 12  2.4366 0.0313540 *
Leu            0.00224993 0.01858730 12  0.1210 0.9056574
Lys            0.10894641 0.02098284 12  5.1922 0.0002248 ***
`Asp FAA`     0.10326407 0.01359632 12  7.5950 6.374e-06 ***
`Glu FAA`    -0.02094732 0.05791302 12 -0.3617 0.7238627
`Asn FAA`    -0.09190216 0.08222859 12 -1.1176 0.2855965
`His FAA`    -0.13236706 0.04163503 12 -3.1792 0.0079326 **
`Ser FAA`    -0.05407671 0.02121125 12 -2.5494 0.0254922 *
`Gln FAA`    0.01354876 0.01491510 12  0.9084 0.3815525
`Gly/Arg FAA` -0.06535486 0.02866761 12 -2.2797 0.0416974 *
`Thr FAA`    -0.08406446 0.02973170 12 -2.8274 0.0152481 *
`Ala FAA`    -0.04402886 0.03670150 12 -1.1996 0.2534266
`Tyr FAA`    0.04332440 0.02575843 12  1.6820 0.1183959
`Met FAA`    -0.02615988 0.03609471 12 -0.7248 0.4824945
`Val FAA`    0.04915664 0.01804782 12  2.7237 0.0184812 *
`Phe FAA`    -0.05806403 0.01881592 12 -3.0859 0.0094339 **
`Ile FAA`    0.01214547 0.02458809 12  0.4940 0.6302537
`Leu FAA`    0.08064523 0.02580117 12  3.1256 0.0087625 **
`Lys FAA`    0.03558259 0.03191284 12  1.1150 0.2866857
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```

Figur B.11: Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 1, gjennom funksjonen jack.test.

## B.4 Matrise 5: Aminosyredeskriptorer, ABTS antioksidantkapasitet

Figur B.12-B.20 viser diverse resultater av dataanalysen av matrise 5.



Figur B.12: BoksploTT over variablene i matrise 5 etter standardisering.

```

> set.seed(9051)
> pls.model=pls(umolPGperghydrolysat~., data=std.Hydrolysat, scale=FALSE, center=FALSE, validation="LOO", jackknife= TRUE,method="oscorespls")
> summary(pls.model)#CV RMSEP given and explained variance components for fitting
Data: X dimension: 13 10
      Y dimension: 13 1
Fit method: oscorespls
Number of components considered: 10

VALIDATION: RMSEP
Cross-validated using 13 leave-one-out segments.
      (Intercept) 1 comps 2 comps 3 comps 4 comps 5 comps 6 comps 7 comps 8 comps 9 comps 10 comps
CV      1.041  0.3601  0.2667  0.2069  0.1579  0.1421  0.1398  0.1698  0.1748  0.1748  0.2258
adjCV   1.041  0.3572  0.2588  0.2012  0.1558  0.1402  0.1408  0.1653  0.1700  0.1701  0.2188

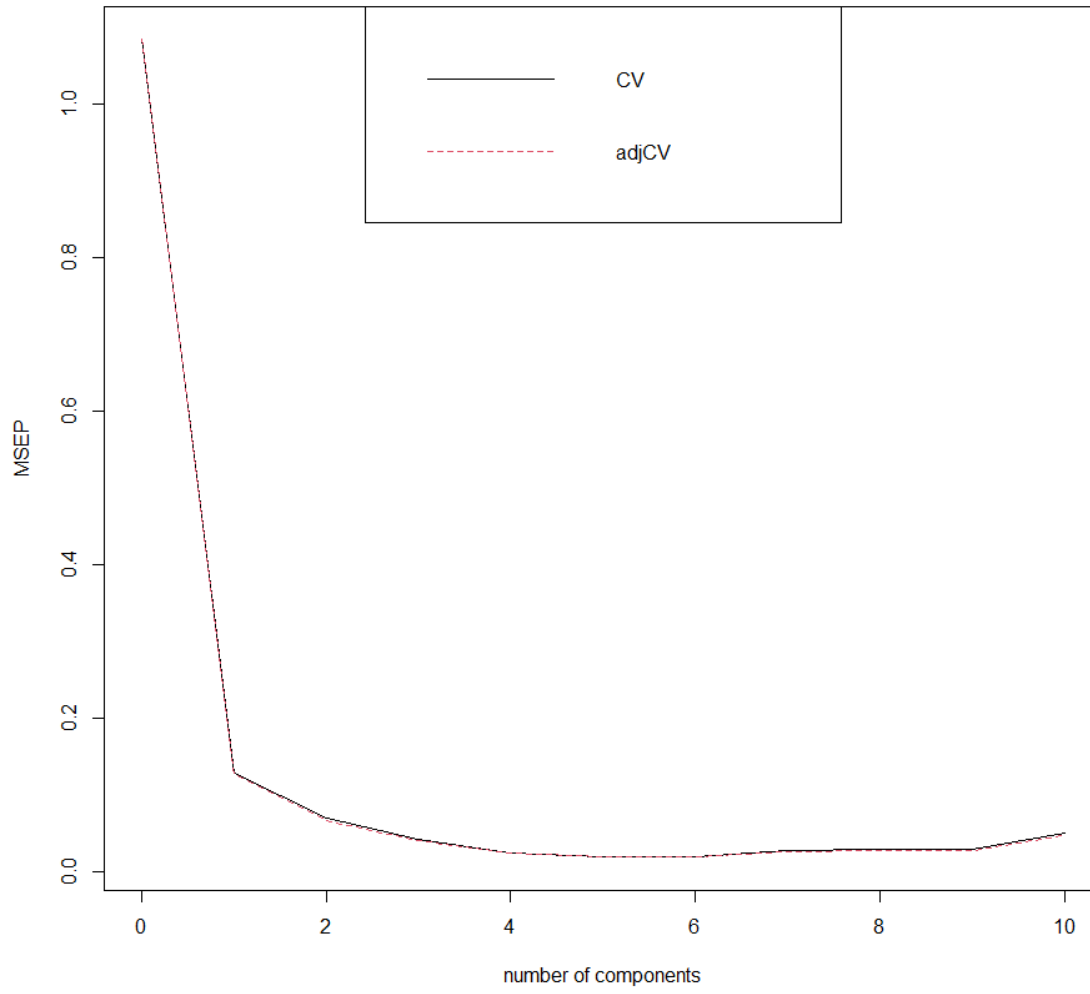
TRAINING: % variance explained
      1 comps 2 comps 3 comps 4 comps 5 comps 6 comps 7 comps 8 comps 9 comps 10 comps
X      44.46  61.78  72.51  90.80  97.10  98.89  99.40  99.79  99.98  100.00
umolPGperghydrolysat  91.55  97.57  98.78  98.91  99.13  99.26  99.58  99.64  99.65  99.65
> |

```

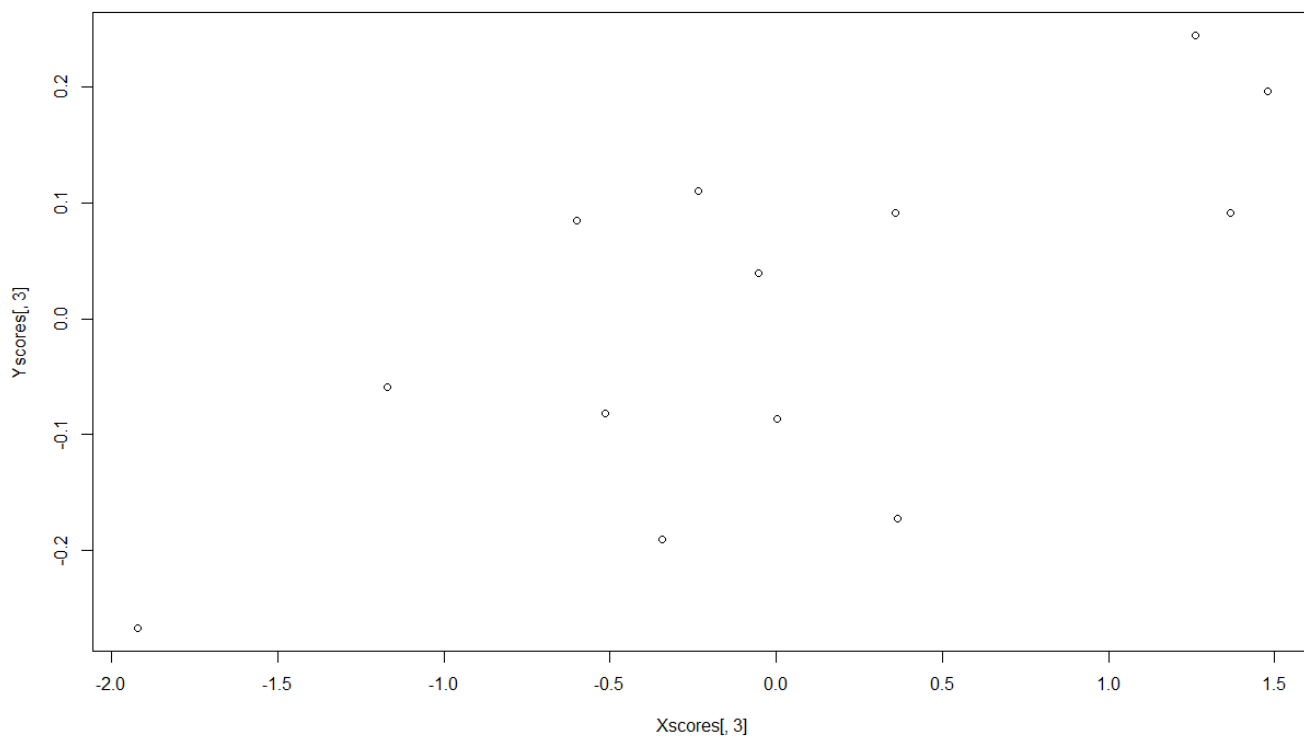
Figur B.13: Sammendrag av PLSR-analysen av matrise 5



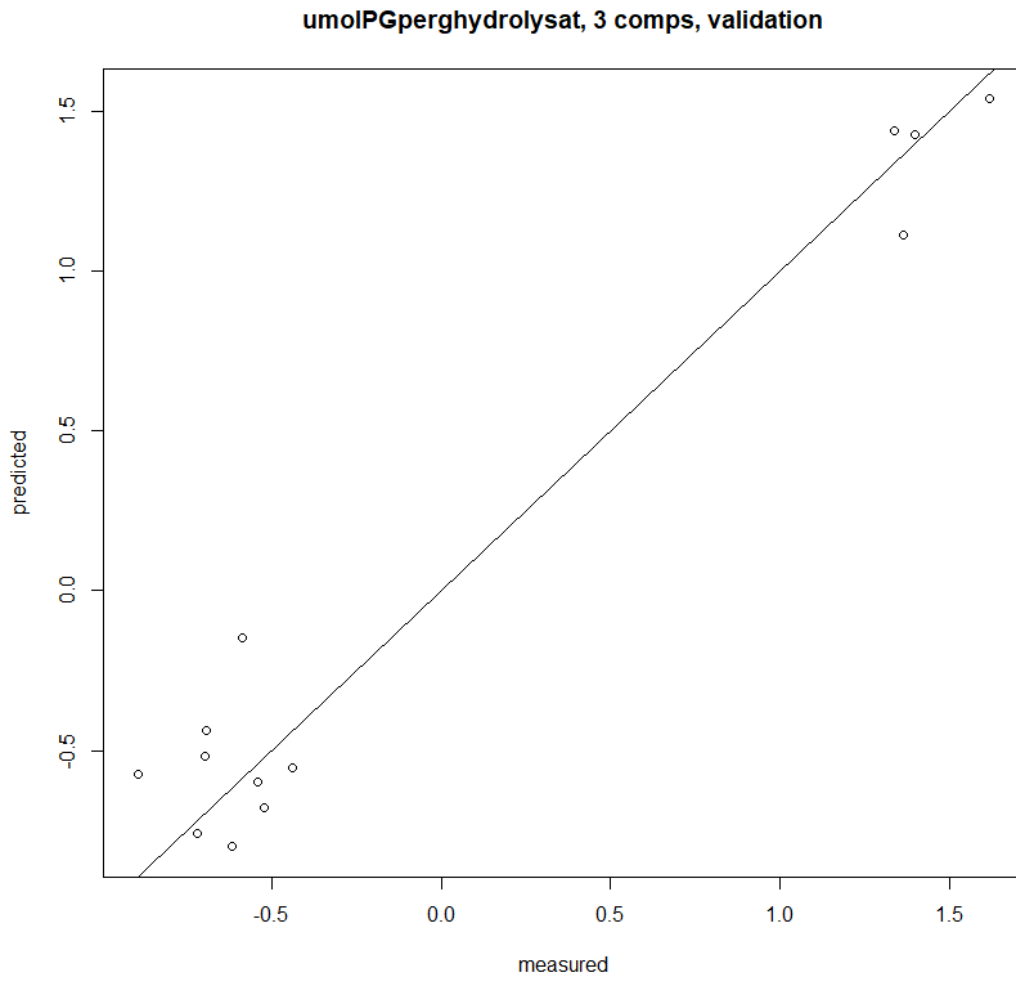
umolPGperhydrolysat



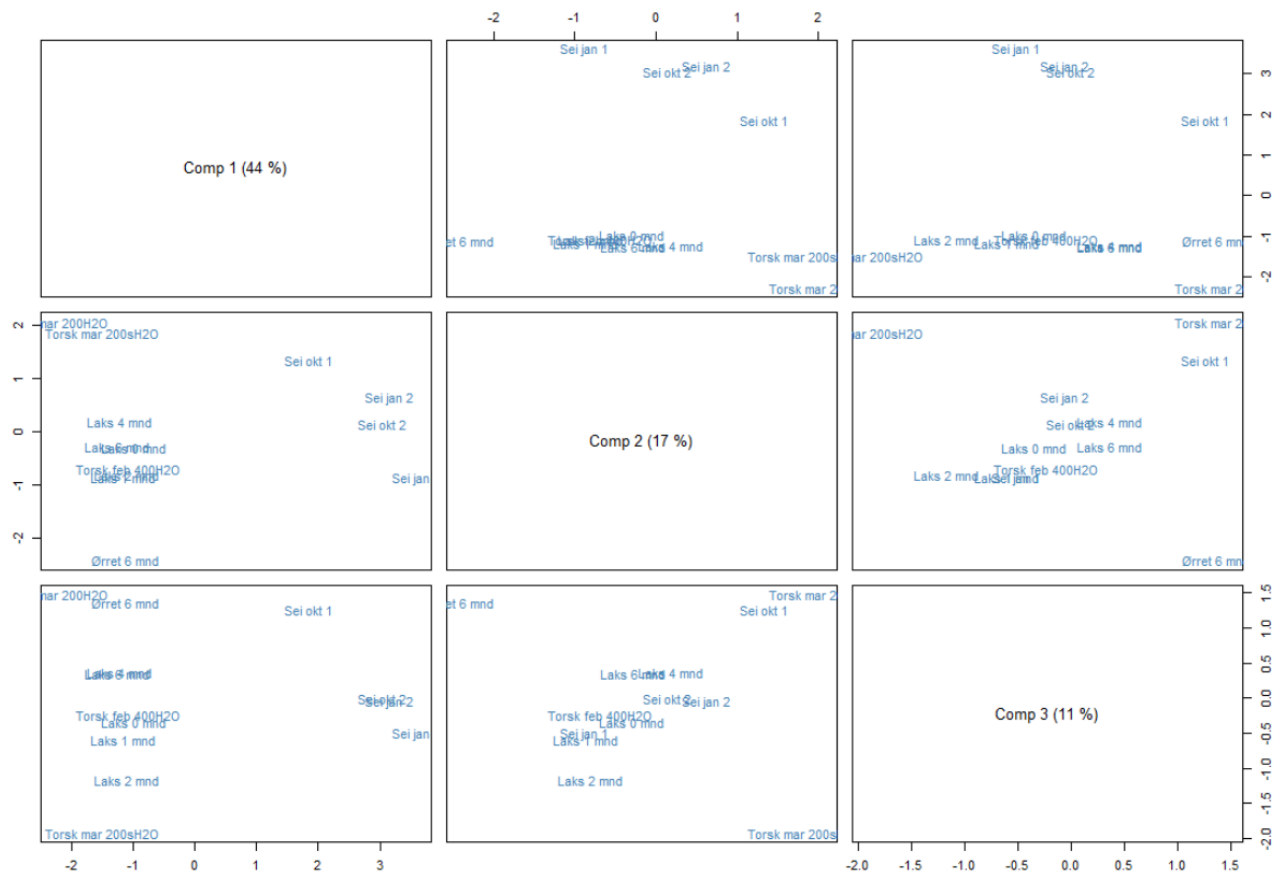
Figur B.14: MSE for den kryssvaliderte modellen med ulikt antall komponenter for matrise 5.



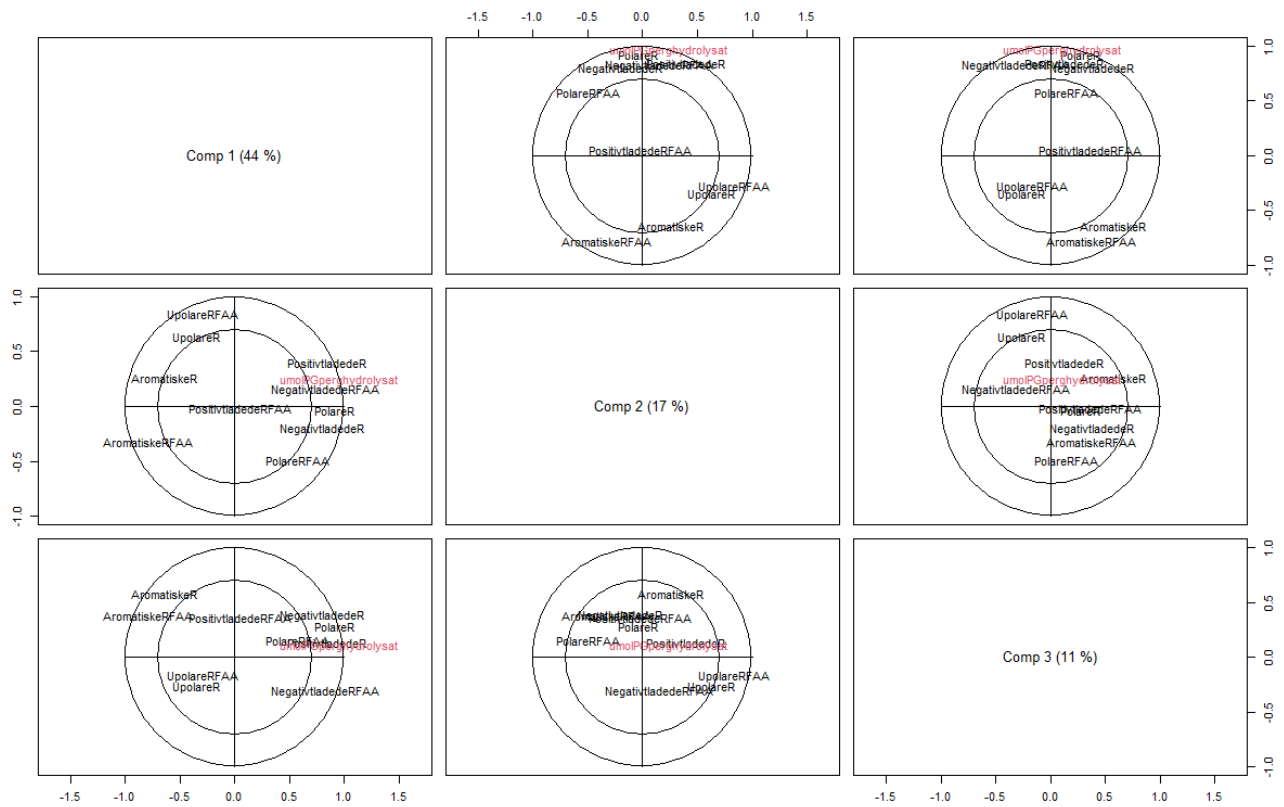
Figur B.15: Plott over X og Y-scores for alle komponentene i PLSR-modellen for matrise 5.



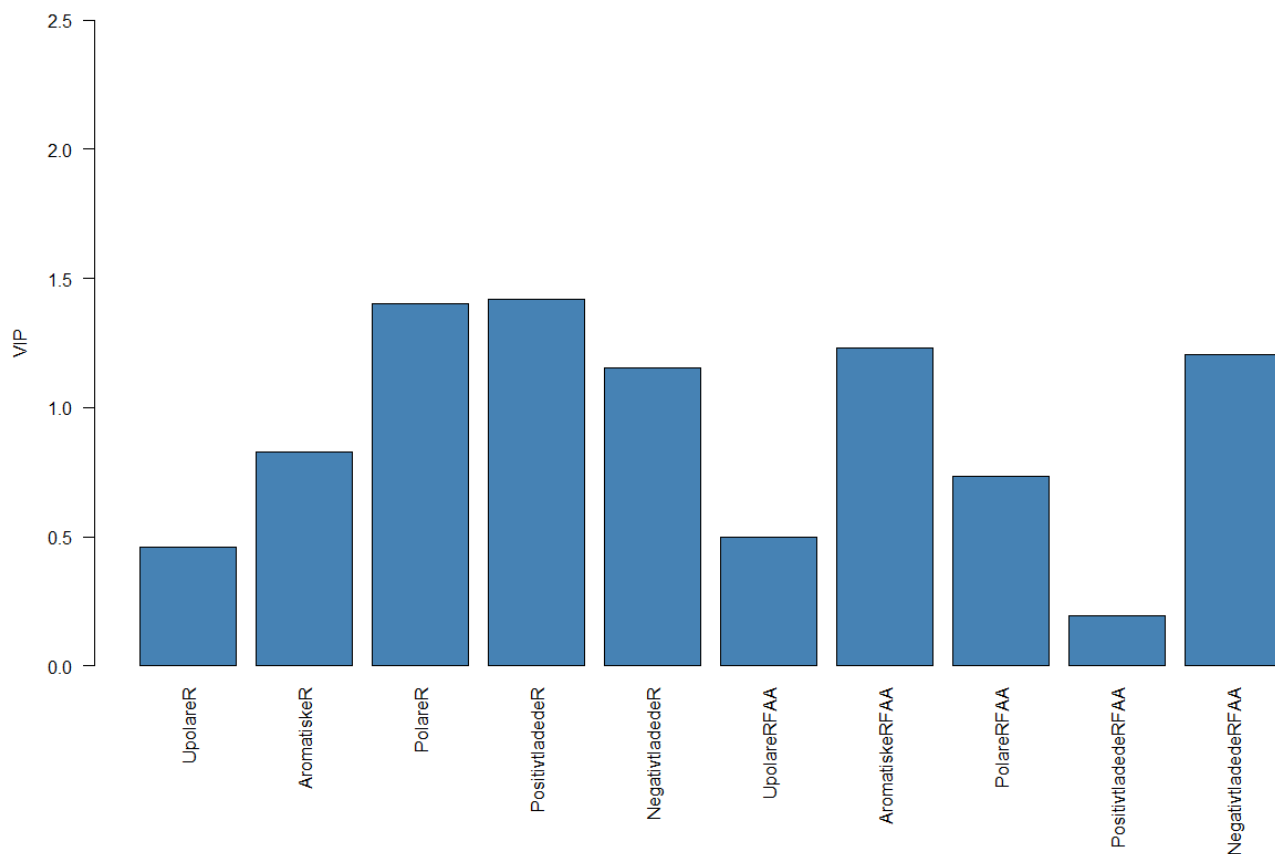
Figur B.16: Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 5.



Figur B.17: Scores-plott for alle komponentene i PLSR-modellen for matrise 5.



Figur B.18: Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen for matrise 5.



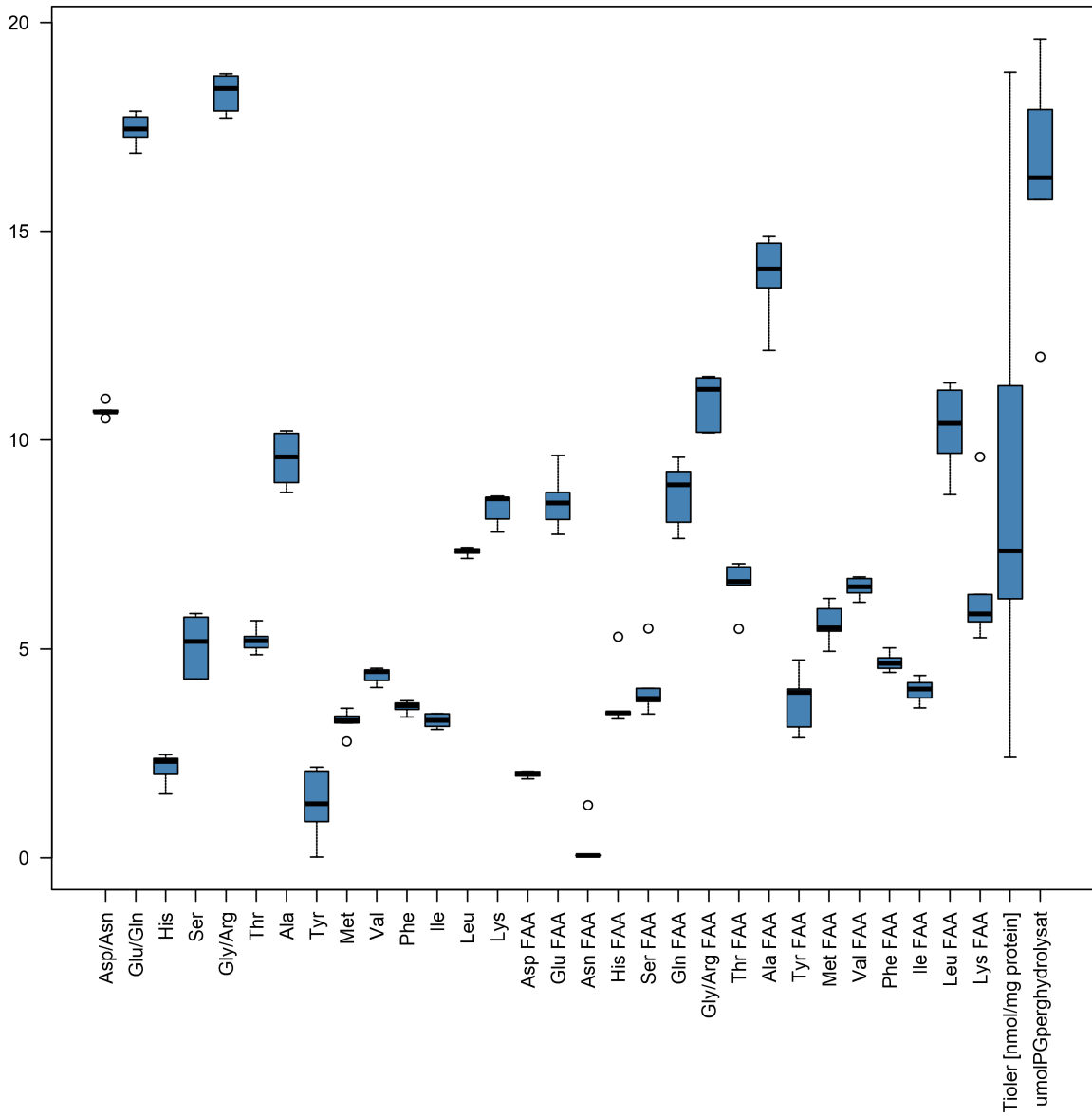
Figur B.19: VIP-verdiene for variablene i den kryssvaliderte modellen av matrise 5 for alle komponentene.

```
> jack.test(plsr.modell, ncomp=3, use.mean=FALSE) #replace n with #comp
Response umolPGperhydrolysat (3 comps):
      Estimate Std. Error Df t value Pr(>|t|)
UpolareR      0.0041623  0.0588703 12  0.0707  0.944798
AromatiskeR   0.0240302  0.0548120 12  0.4384  0.668877
PolareR       0.2942828  0.0511956 12  5.7482 9.189e-05 ***
PositivtladedeR 0.3122321  0.0371517 12  8.4042 2.262e-06 ***
NegativtladedeR 0.2138680  0.0518880 12  4.1217  0.001416 **
UpolareRFAA   0.1028229  0.0343528 12  2.9931  0.011209 *
AromatiskeRFAA -0.2168918  0.0686855 12 -3.1578  0.008255 **
PolareRFAA    0.0186406  0.0330907 12  0.5633  0.583591
PositivtladedeRFAA 0.0840048  0.0529217 12  1.5873  0.138420
NegativtladedeRFAA 0.1363337  0.0857366 12  1.5901  0.137787
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

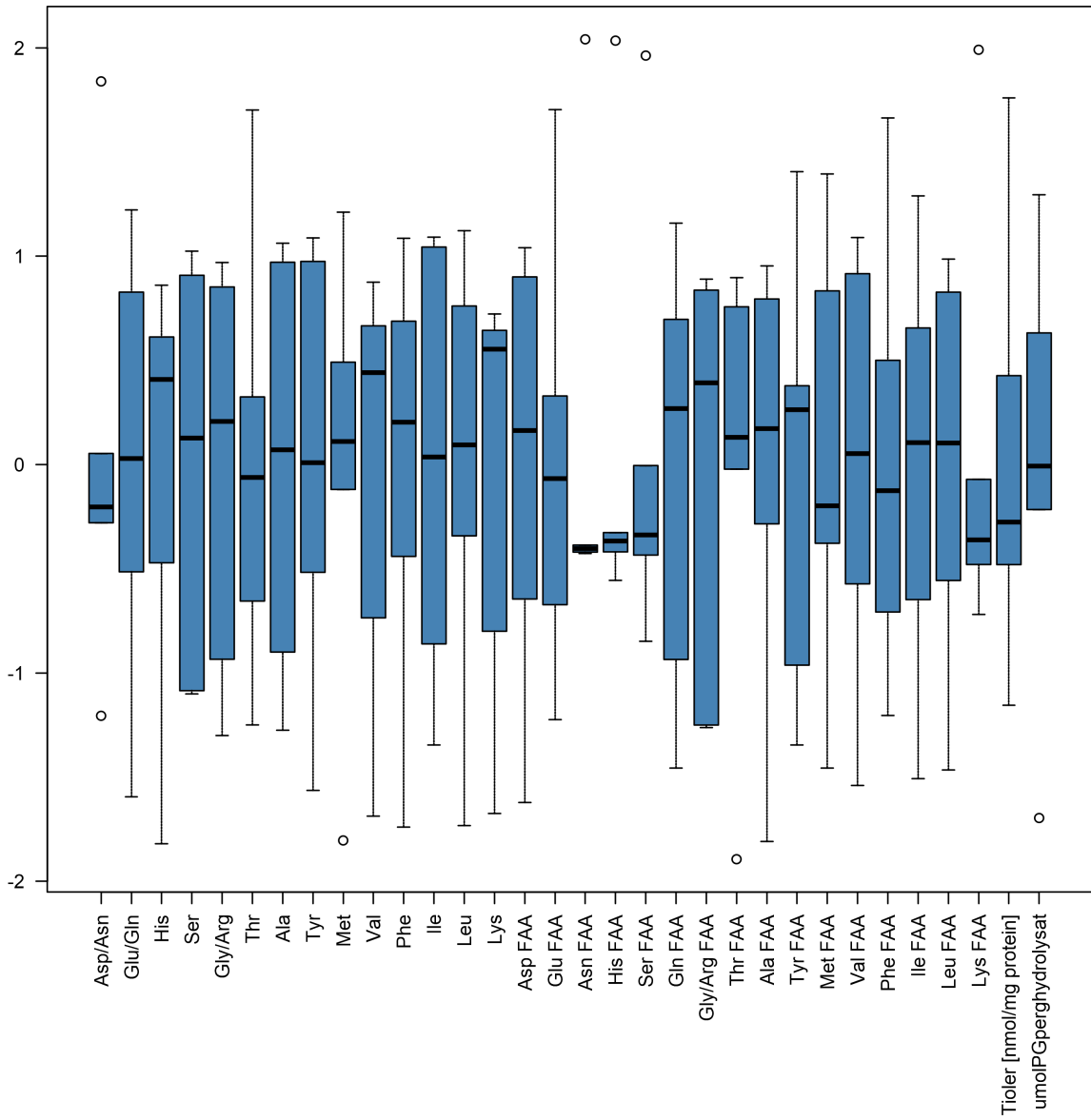
Figur B.20: Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 5, gjennom funksjonen jack.test.

## B.5 Matrise 2: TAA, FAA, tioler, ABTS antioksidantkapasitet

Figur B.21-B.31 viser diverse resultater av dataanalysen av matrise 2.

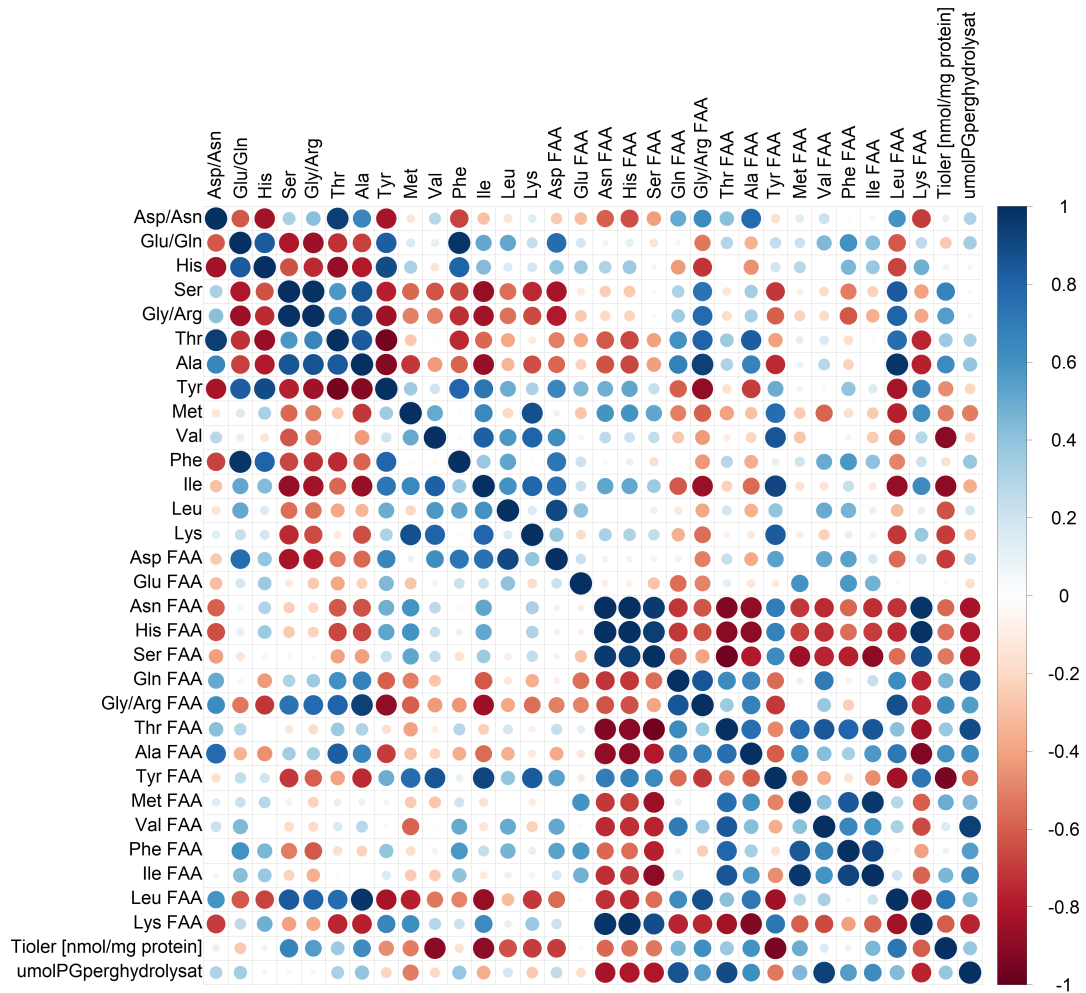


Figur B.21: Bokplott over de originale variablene i matrise 2.



Figur B.22: Boksploott over de standardiserte variablene i matrise 2.





Figur B.23: Korrelasjonsmatrisen for variablene i matrise 2.

```

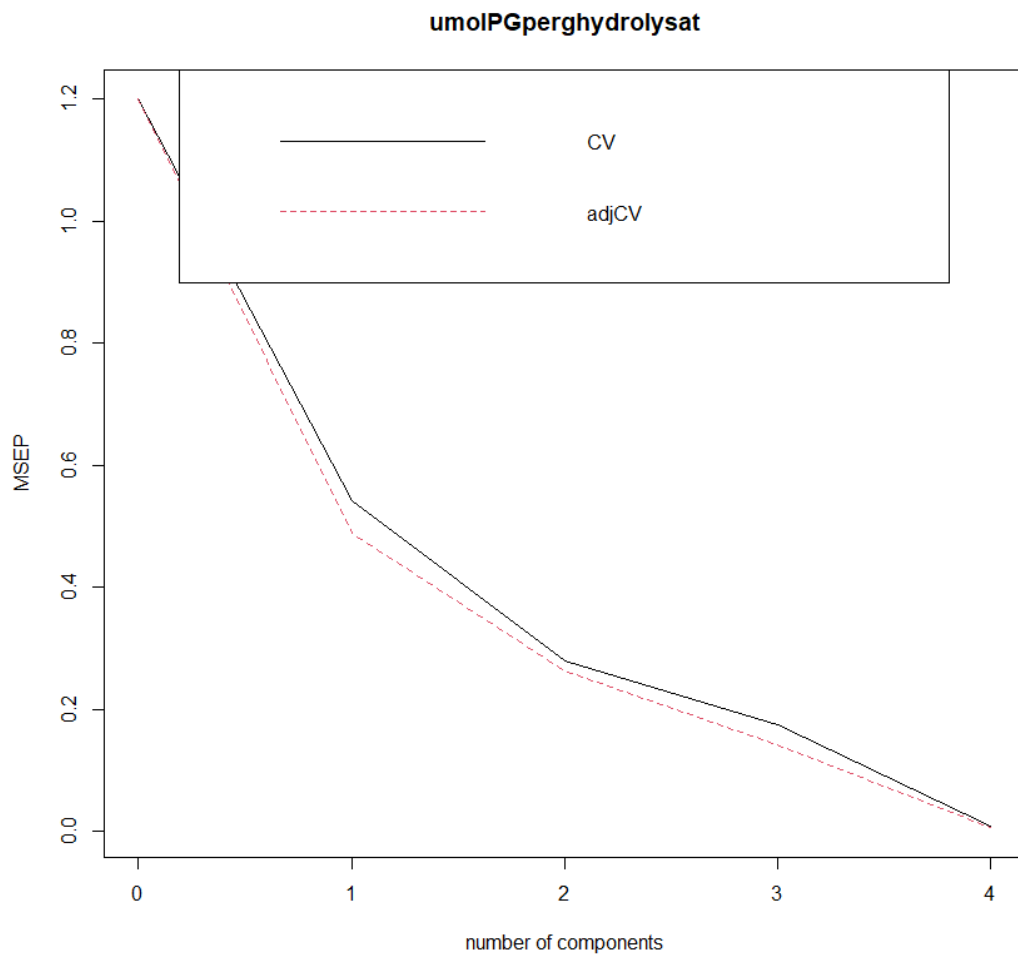
> set.seed(9052)
> pls.model=plsr(umolPGperghydrolysat~., data=std.Hydrolysat, scale=FALSE, center=FALSE, validation="LOO", jackknife= TRUE,method="oscorespls")
> summary(pls.model)#CV RMSEP given and explained variance components for fitting
Data: X dimension: 6 31
      Y dimension: 6 1
Fit method: oscorespls
Number of components considered: 4

VALIDATION: RMSEP
Cross-validated using 6 leave-one-out segments.
      (Intercept) 1 comps 2 comps 3 comps 4 comps
CV          1.095  0.7372  0.5299  0.4185  0.08964
adjCV       1.095  0.6990  0.5134  0.3763  0.07984

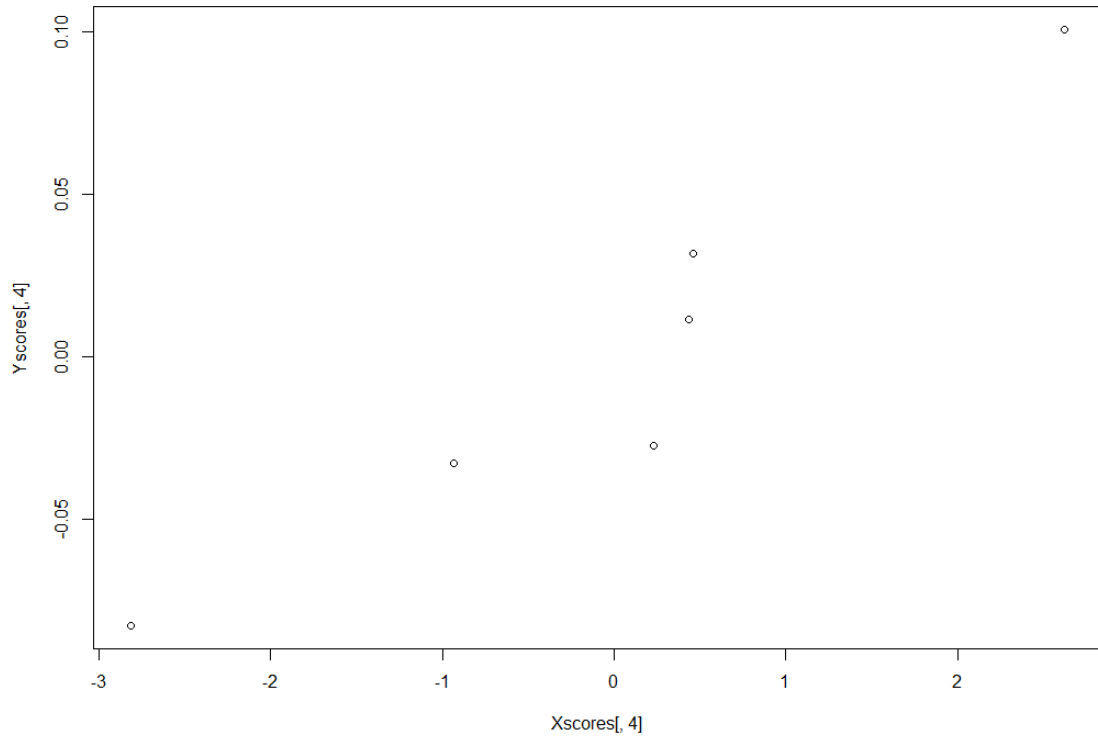
TRAINING: % variance explained
      1 comps 2 comps 3 comps 4 comps
X          39.57  70.61  80.96  92.15
umolPGperghydrolysat  77.64  91.13  99.60  99.96

```

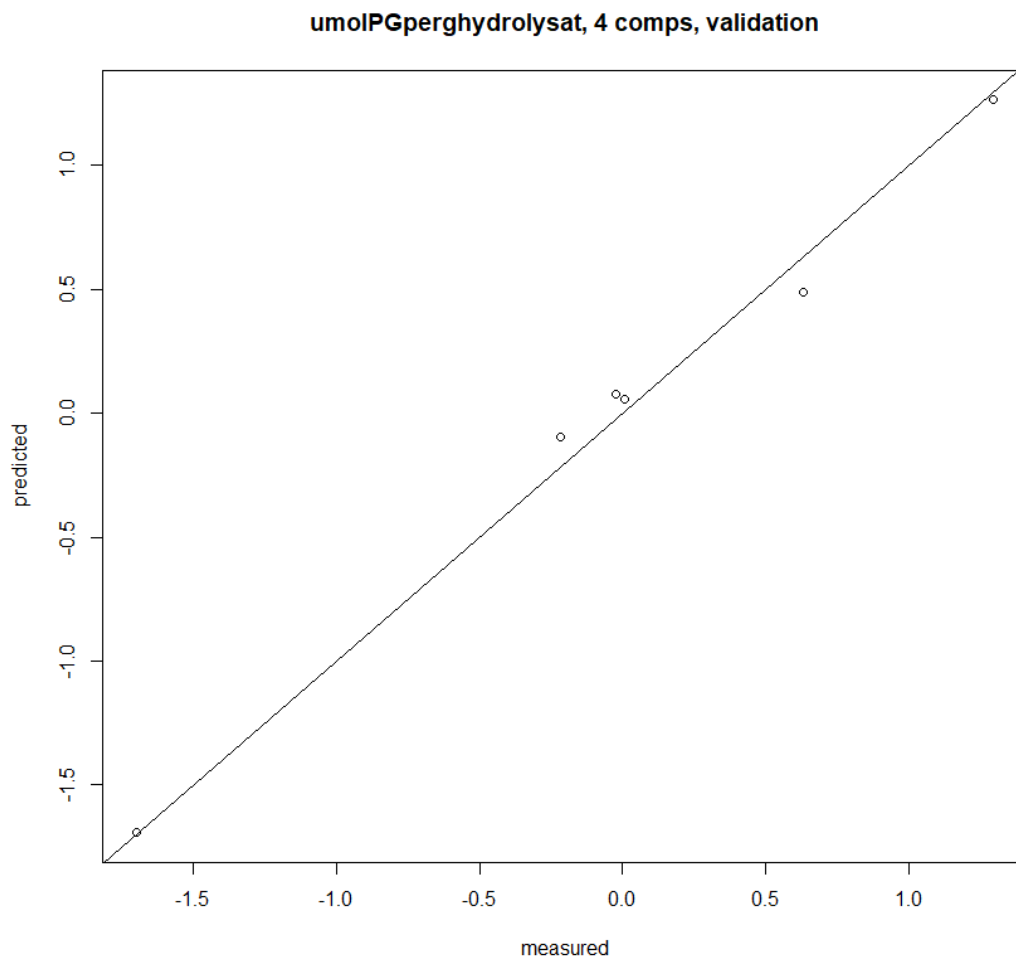
Figur B.24: Sammendrag av PLSR-analysen for matrise 2.



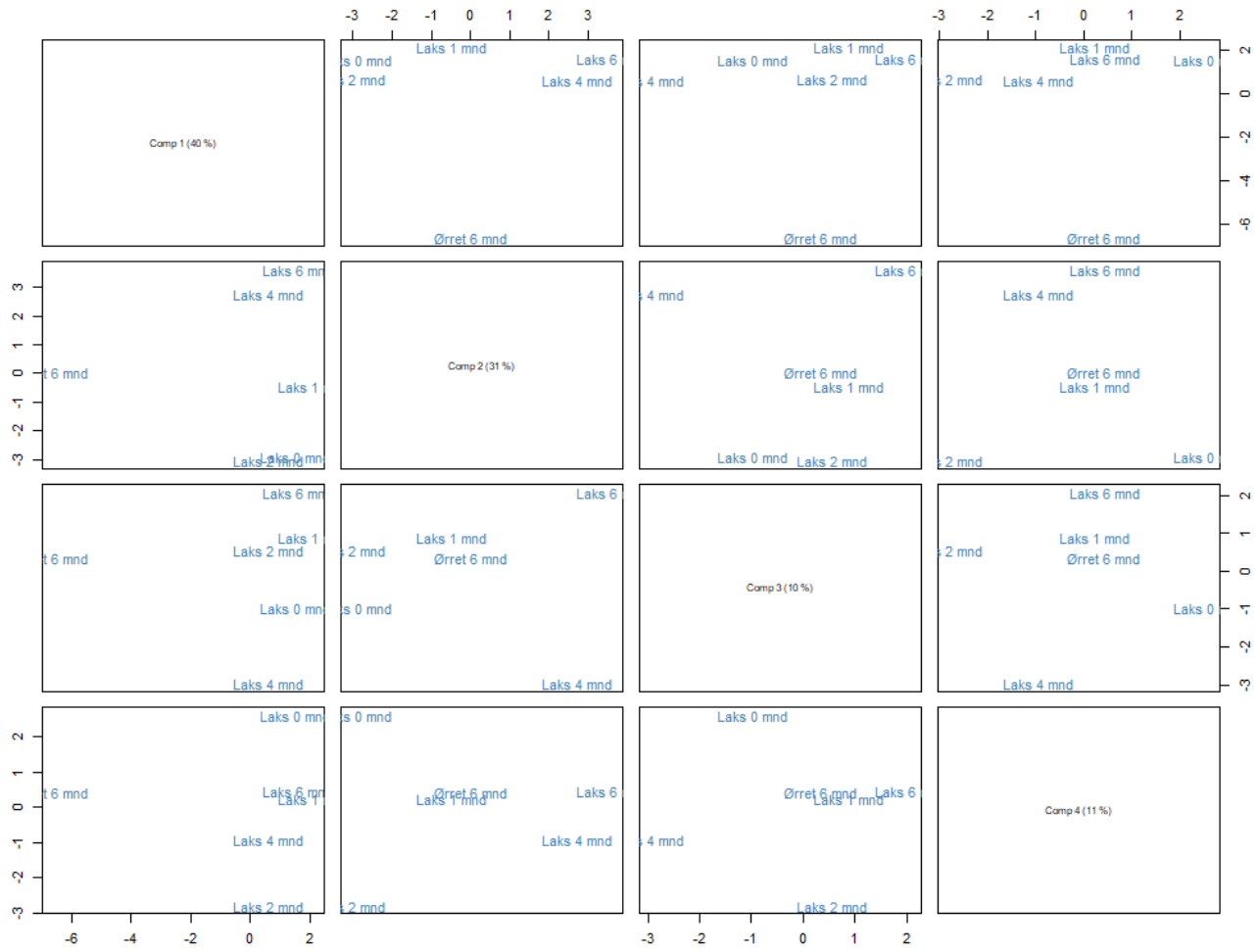
Figur B.25: MSEP for den kryssvaliderte PLSR-modellen med ulikt antall komponenter for matrise 2.



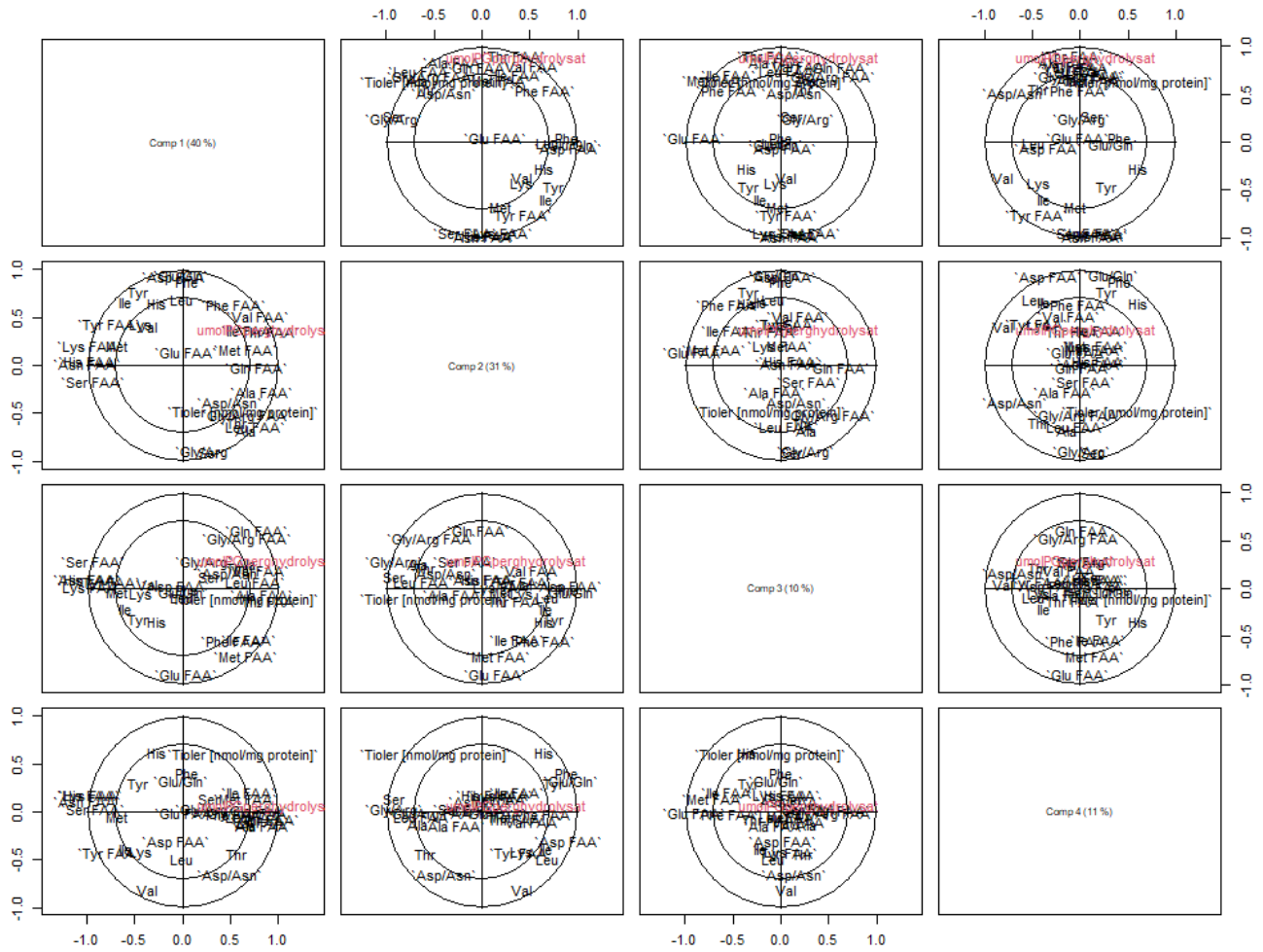
Figur B.26: Plott over X og Y-scores for de fire komponentene i PLSR-modellen for matrise 2.



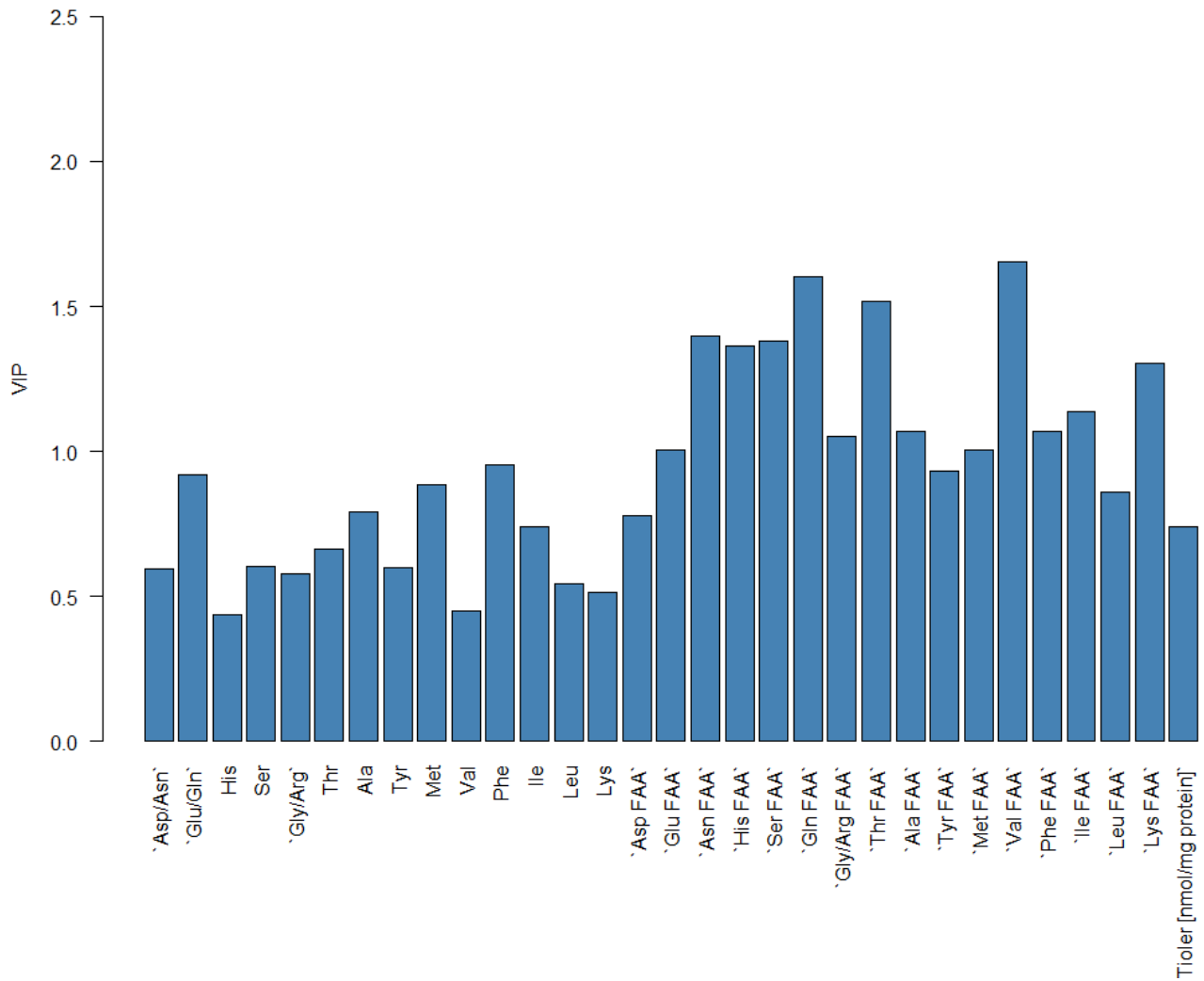
Figur B.27: Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 2.



Figur B.28: Scores-plott for alle komponentene i PLSR-modellen for matrise 2.



Figur B.29: Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen for matrise 2.



Figur B.30: VIP-verdiene for variablene i PLSR-modellen med alle fire komponentene for matrise 2.

```

> #Jack.test: gjør t-tester av reg.koeff. basert på jackknife variance estimates.
> jack.test(plsr.modell,ncomp=4, use.mean=FALSE) #replace n with #comp
Response umolPGperghydrolysat (4 comps):
      Estimate Std. Error Df  t value Pr(>|t|)
`Asp/Asn`      0.00084655 0.01394245 5    0.0607 0.9539362
`Glu/Gln`      0.12521586 0.00653628 5   19.1570 7.146e-06 ***
His            0.02149899 0.01227762 5    1.7511 0.1403285
Ser           -0.05405836 0.00935149 5   -5.7807 0.0021794 **
`Gly/Arg`     -0.03750385 0.00874477 5   -4.2887 0.0077986 **
Thr            0.01006837 0.01227841 5    0.8200 0.4495212
Ala            0.02582108 0.00953268 5    2.7087 0.0423423 *
Tyr           -0.00813551 0.00712500 5   -1.1418 0.3052417
Met           -0.04009660 0.03292887 5   -1.2177 0.2776733
Val           -0.02295121 0.01197562 5   -1.9165 0.1134374
Phe            0.13262652 0.01196579 5   11.0838 0.0001042 ***
Ile           -0.04190093 0.00629846 5   -6.6526 0.0011576 **
Leu            0.02780153 0.03260817 5    0.8526 0.4328166
Lys           -0.01841178 0.02686474 5   -0.6854 0.5235975
`Asp FAA`      0.08548773 0.01830109 5    4.6712 0.0054768 **
`Glu FAA`     -0.16160920 0.01840821 5   -8.7792 0.0003180 ***
`Asn FAA`     -0.07554676 0.00258611 5  -29.2125 8.811e-07 ***
`His FAA`     -0.06710102 0.00308570 5  -21.7458 3.816e-06 ***
`Ser FAA`     -0.06154288 0.00334843 5  -18.3796 8.769e-06 ***
`Gln FAA`      0.20161507 0.00570365 5   35.3484 3.410e-07 ***
`Gly/Arg FAA` 0.10591778 0.00493436 5   21.4654 4.070e-06 ***
`Thr FAA`      0.11197941 0.00558758 5   20.0408 5.718e-06 ***
`Ala FAA`      0.04098129 0.01921134 5    2.1332 0.0860567 .
`Tyr FAA`     -0.03747983 0.00797509 5   -4.6996 0.0053390 **
`Met FAA`     -0.03706154 0.01244715 5   -2.9775 0.0308884 *
`Val FAA`      0.17334820 0.01552026 5   11.1692 0.0001004 ***
`Phe FAA`      0.02837294 0.00382398 5    7.4197 0.0007004 ***
`Ile FAA`      0.03016339 0.01130555 5    2.6680 0.0444542 *
`Leu FAA`      0.00664805 0.00941040 5    0.7065 0.5114543
`Lys FAA`     -0.06718049 0.00314190 5  -21.3821 4.149e-06 ***
`Tioler [nmol/mg protein]` 0.01990112 0.01273265 5    1.5630 0.1788154
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |

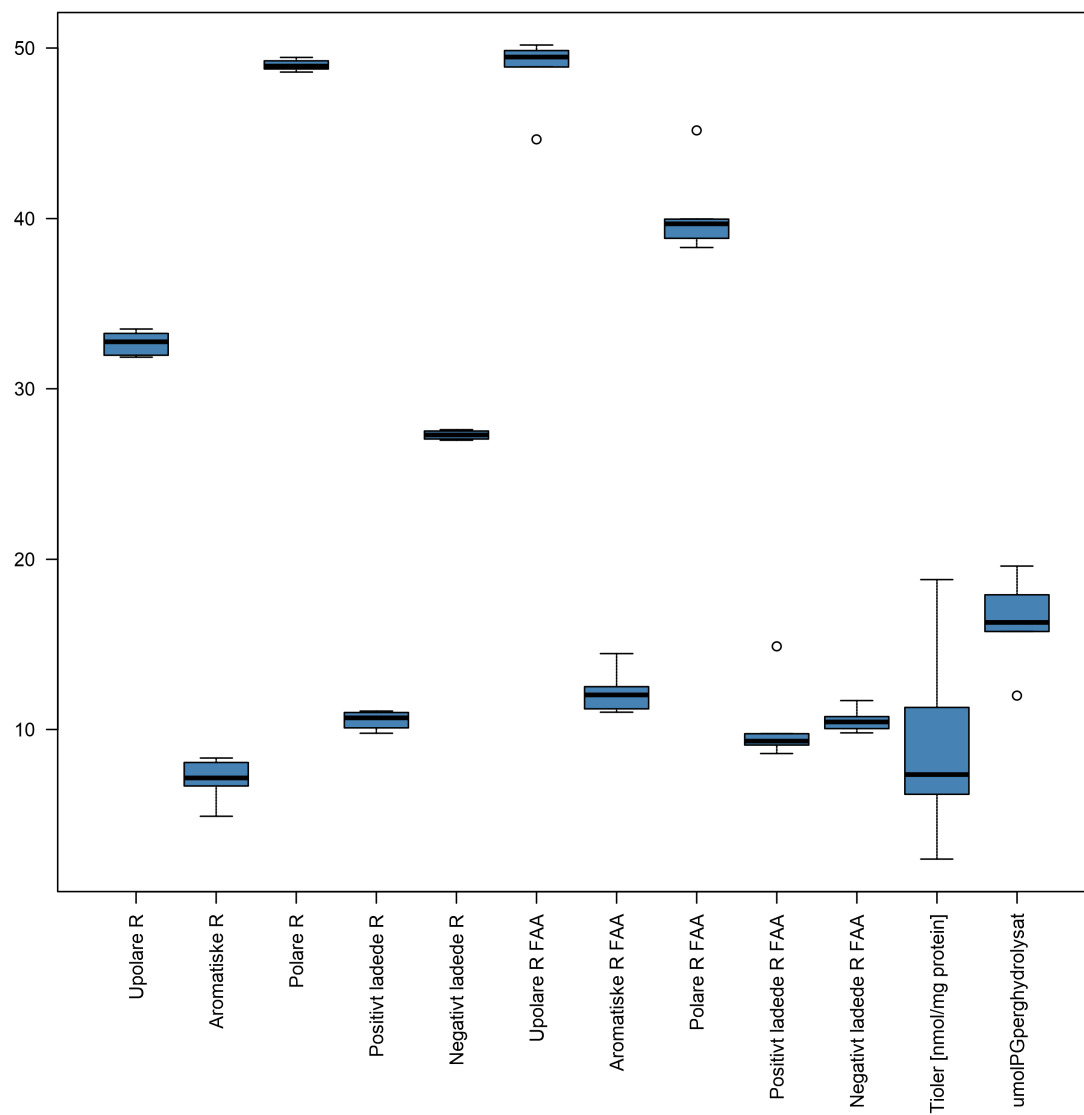
```

Figur B.31: Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 2, gjennom funksjonen jack.test.

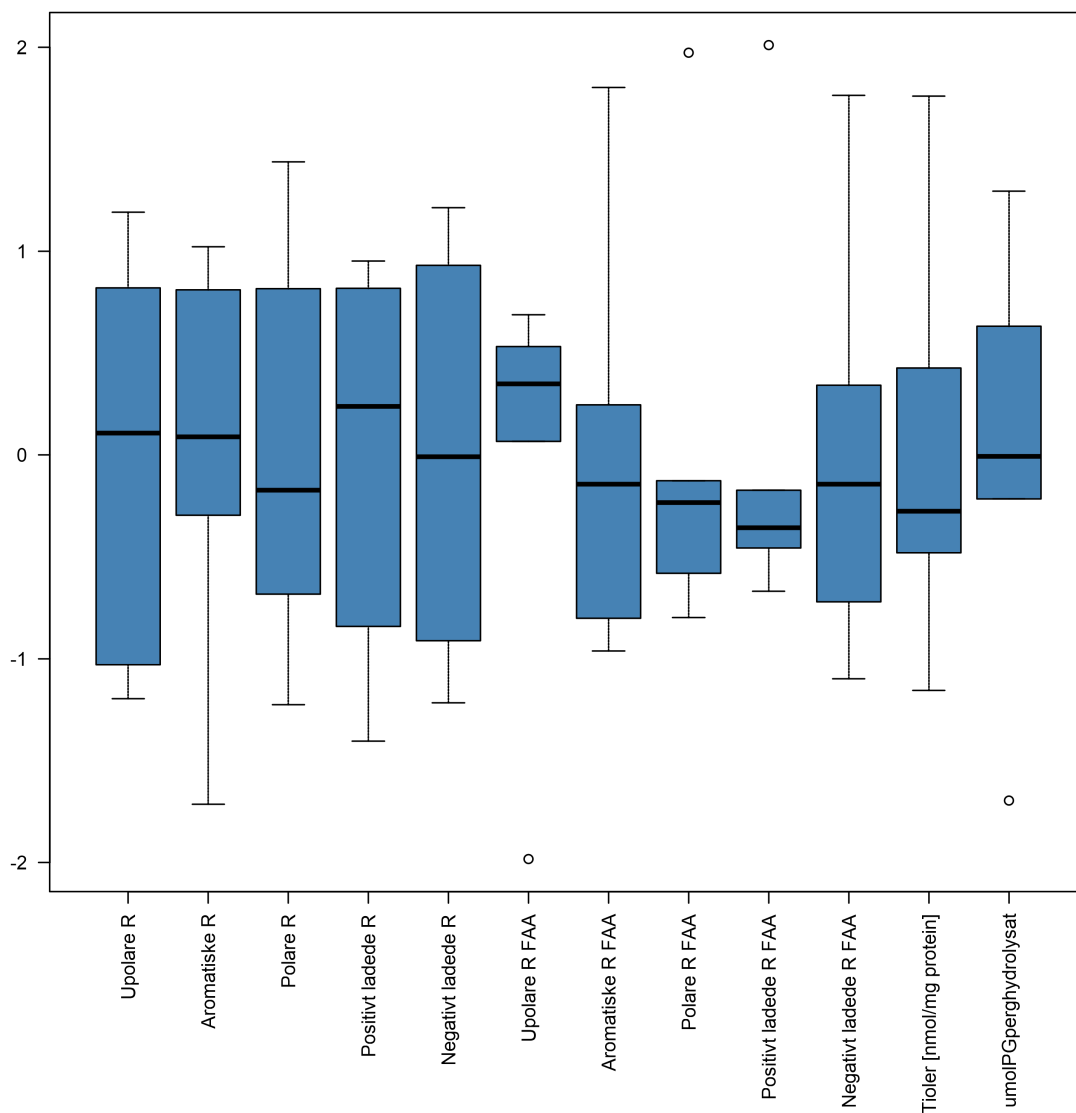
## B.6 Matrise 6: Aminosyredeskriptorer, tioler, ABTS antioksidantkapasitet

Figur B.32-B.44 viser diverse resultater av dataanalysen av matrise 6.

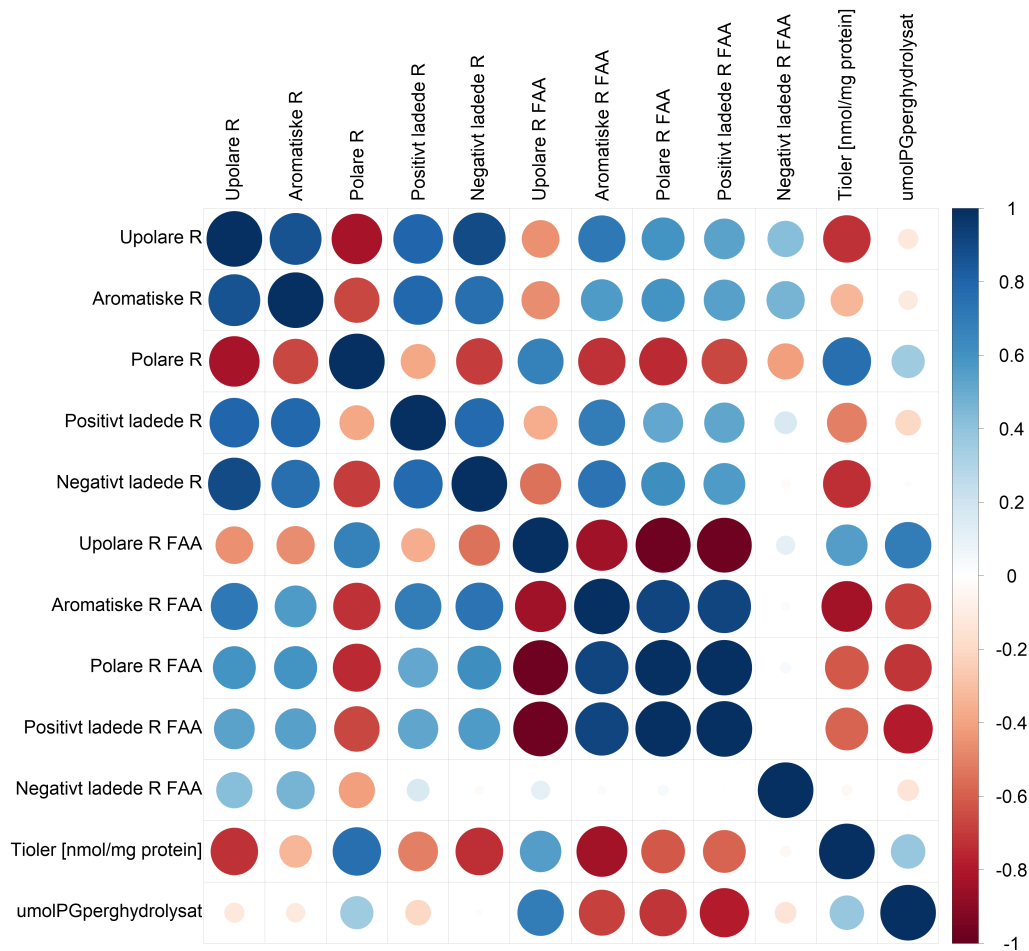




Figur B.32: Bokplott over de originale variablene i matrise 6.



Figur B.33: BoksploTT over de standardiserte variablene i matrise 6.



Figur B.34: Korrelasjonsmatrisen for variablene i matrise 6.

```

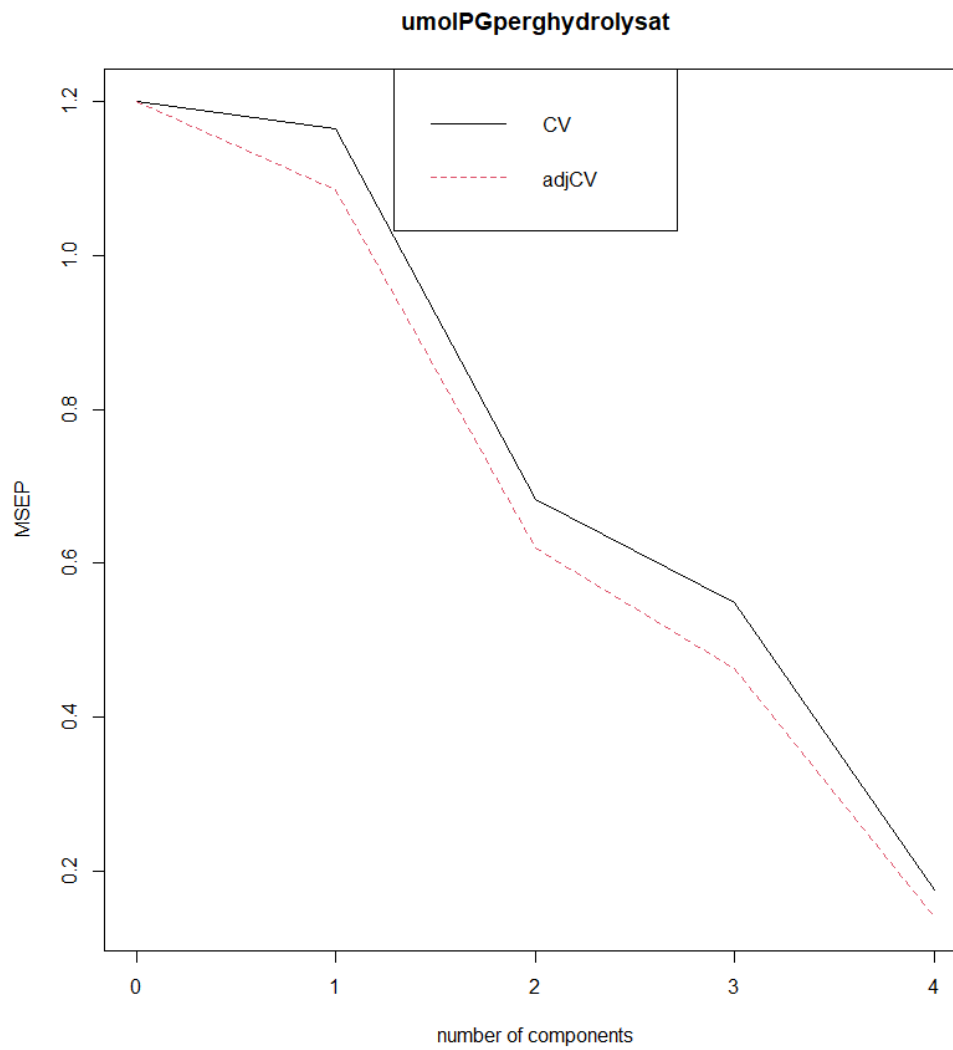
> set.seed(9053)
> pls.model=plsr(umolPGperhydrolysat~., data=std.Hydrolysates, scale=FALSE, center=FALSE, validation="LOO", jackknife= TRUE,method="oscorespls")
> summary(pls.model)#CV RMSEP given and explained variance components for fitting
Data: X dimension: 6 11
      Y dimension: 6 1
Fit method: oscorespls
Number of components considered: 4

VALIDATION: RMSEP
Cross-validated using 6 leave-one-out segments.
      (Intercept) 1 comps 2 comps 3 comps 4 comps
CV          1.095  1.079  0.8260  0.7407  0.4186
adjCV       1.095  1.042  0.7869  0.6807  0.3726

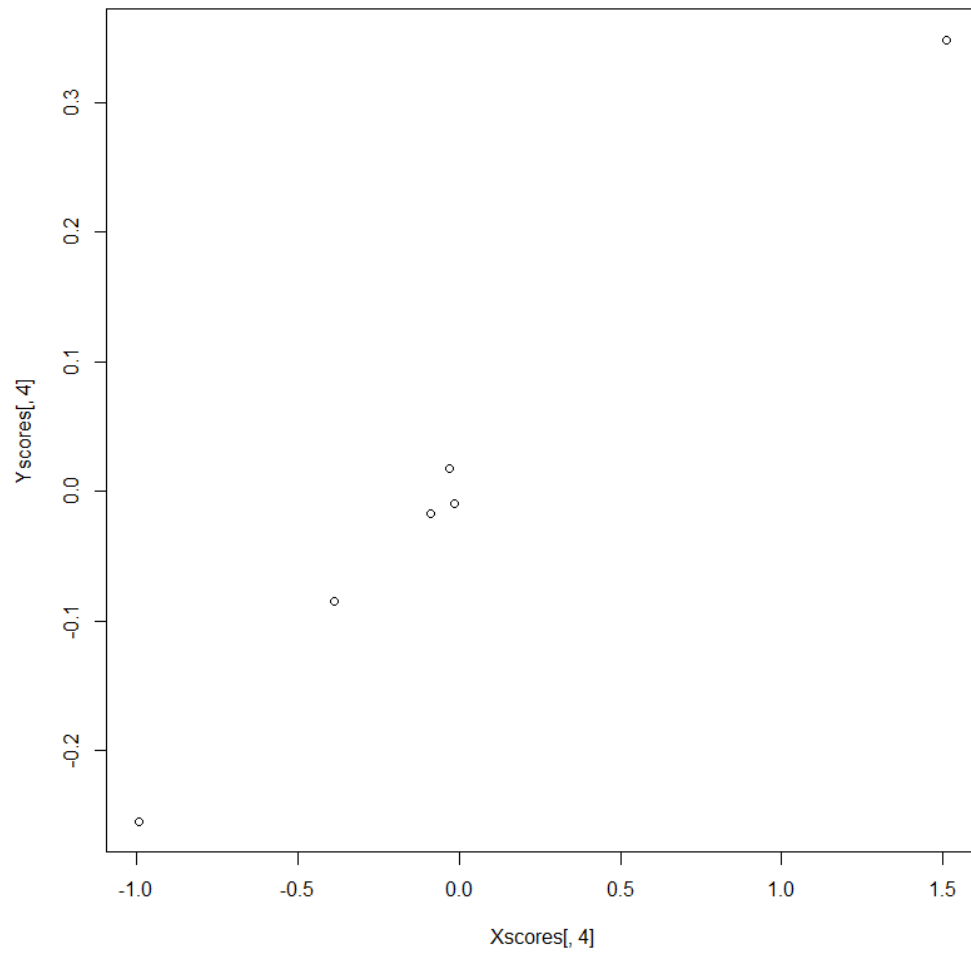
TRAINING: % variance explained
      1 comps 2 comps 3 comps 4 comps
X          61.73  77.94  86.97  93.25
umolPGperhydrolysat  44.62  83.47  96.11  99.98
> #Validation and choosing number of PLS directions
> validationplot(pls.model, val.type="MSEP", legendpos="top") #look for the lowest cross validation error, an estimate for test MSE
> |

```

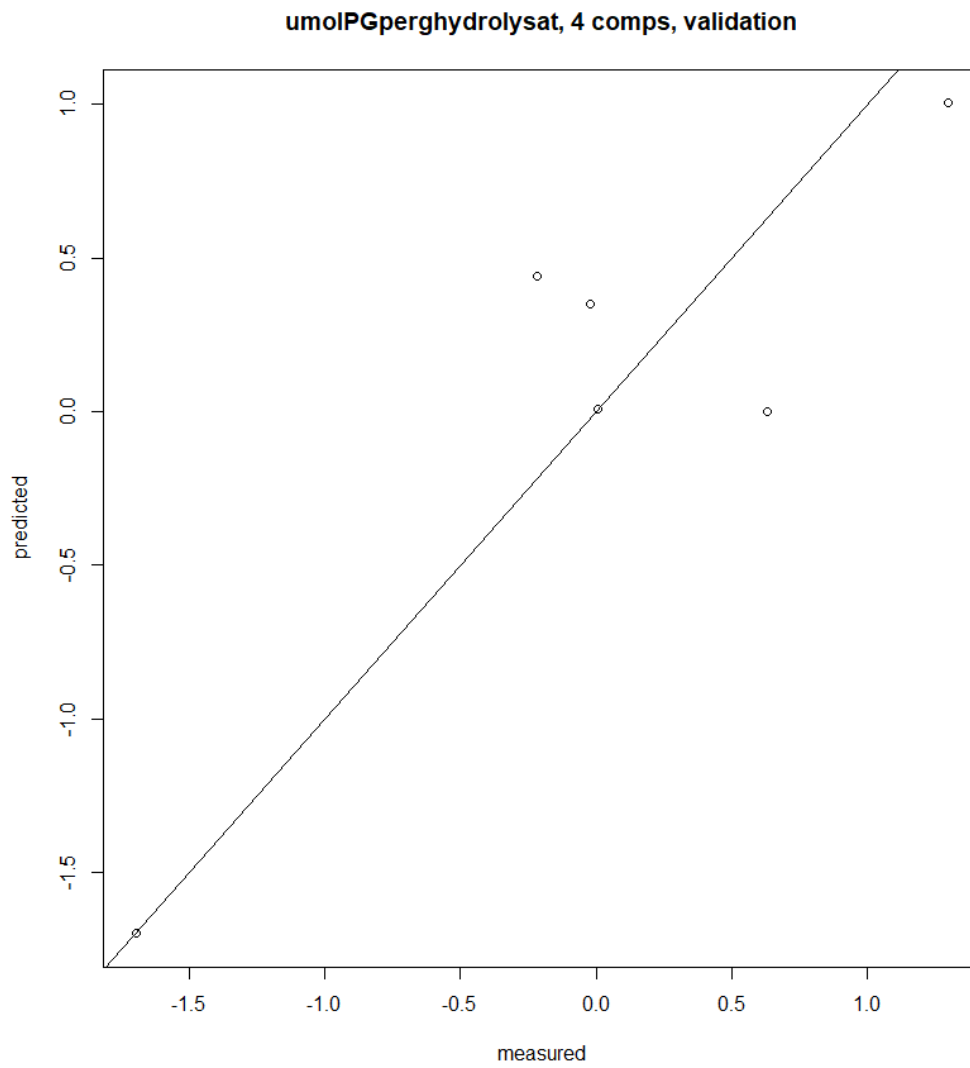
Figur B.35: Sammendrag av PLSR-analysen for matrise 6.



Figur B.36: MSEP for den kryssvaliderte modellen med ulikt antall komponenter for matrise 6.



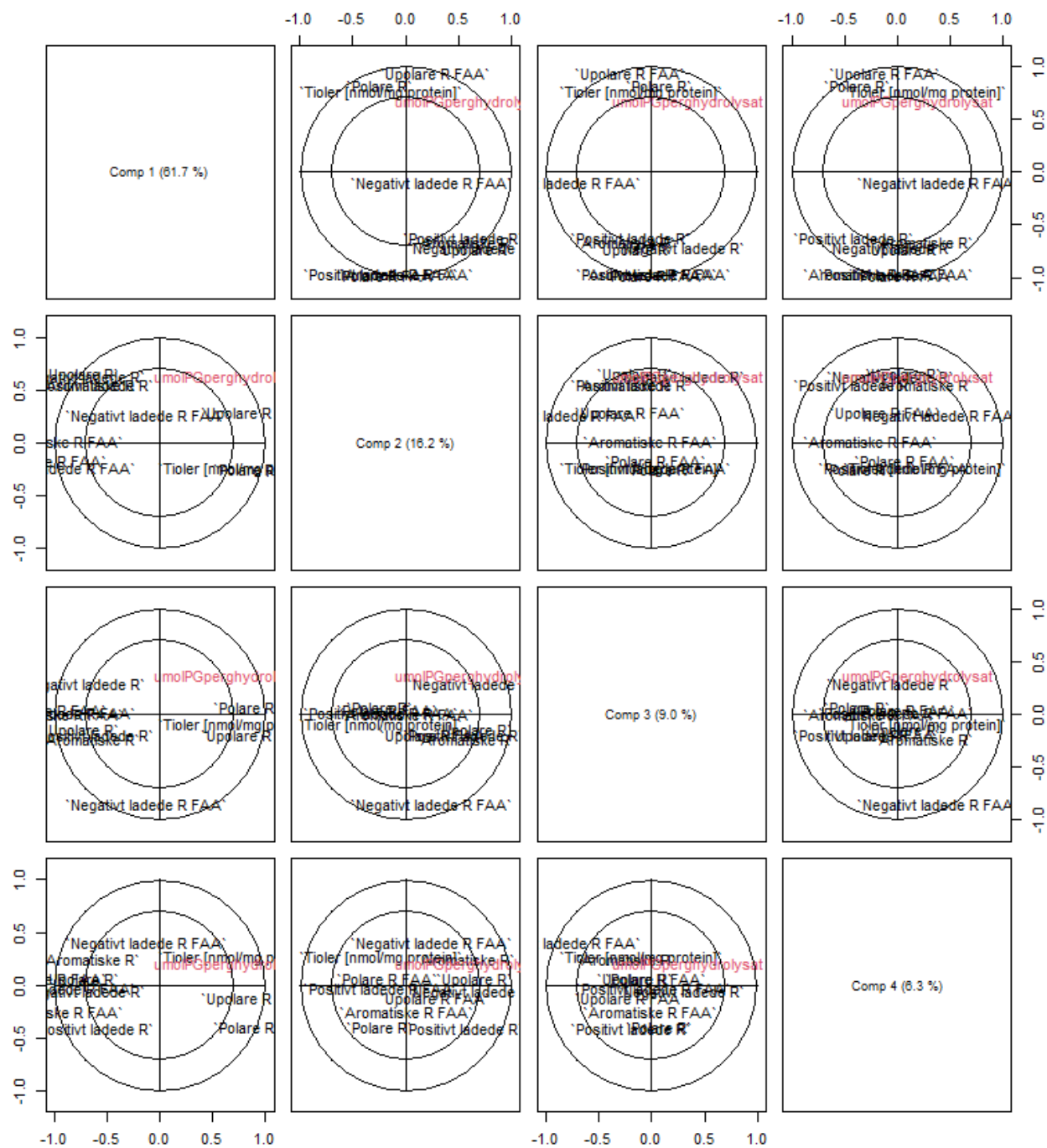
Figur B.37: Plott over X og Y-scores til PLSR-modellen med alle fire komponentene for matrise 6.



Figur B.38: Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 6.

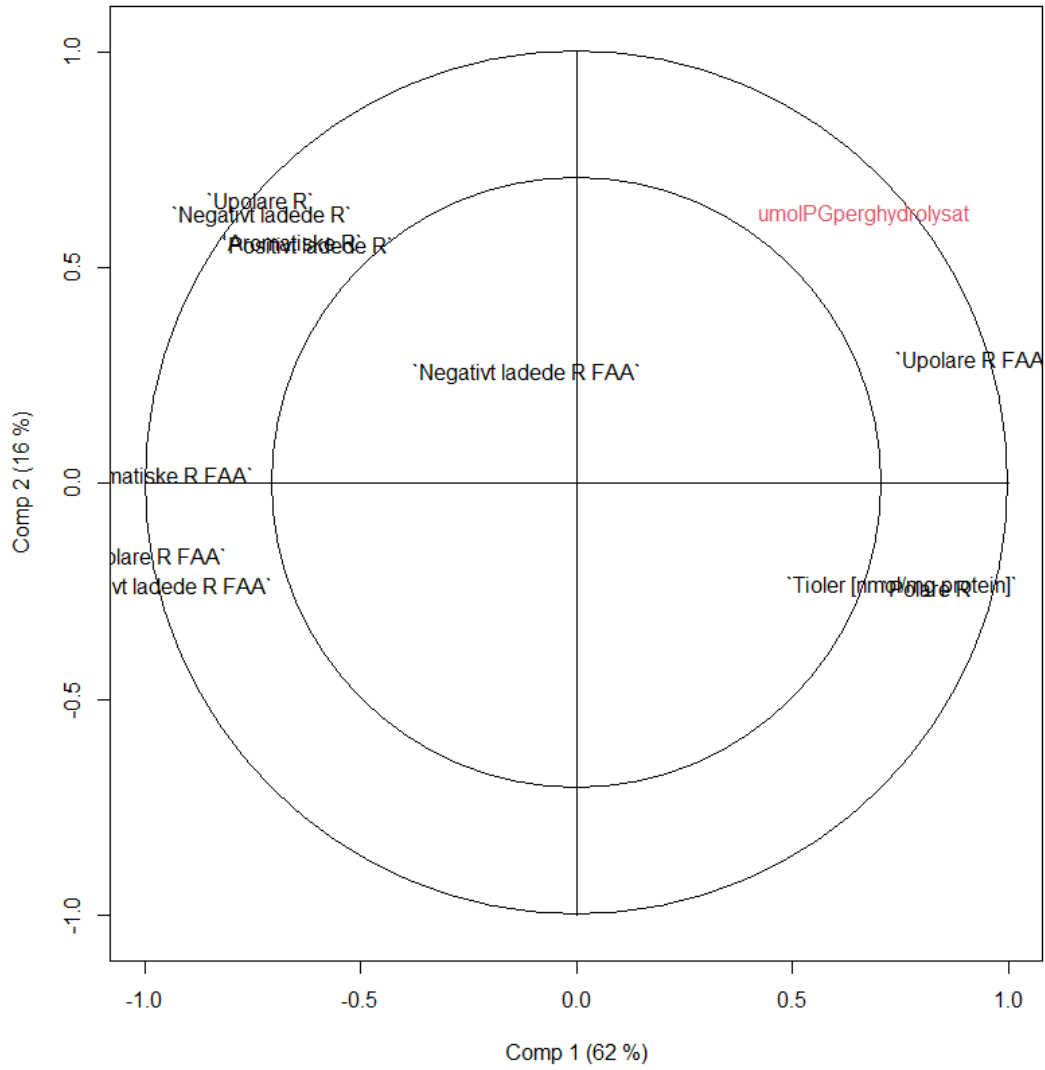


Figur B.39: Scoresplott for alle komponentene i PLSR-modellen for matrise 6.

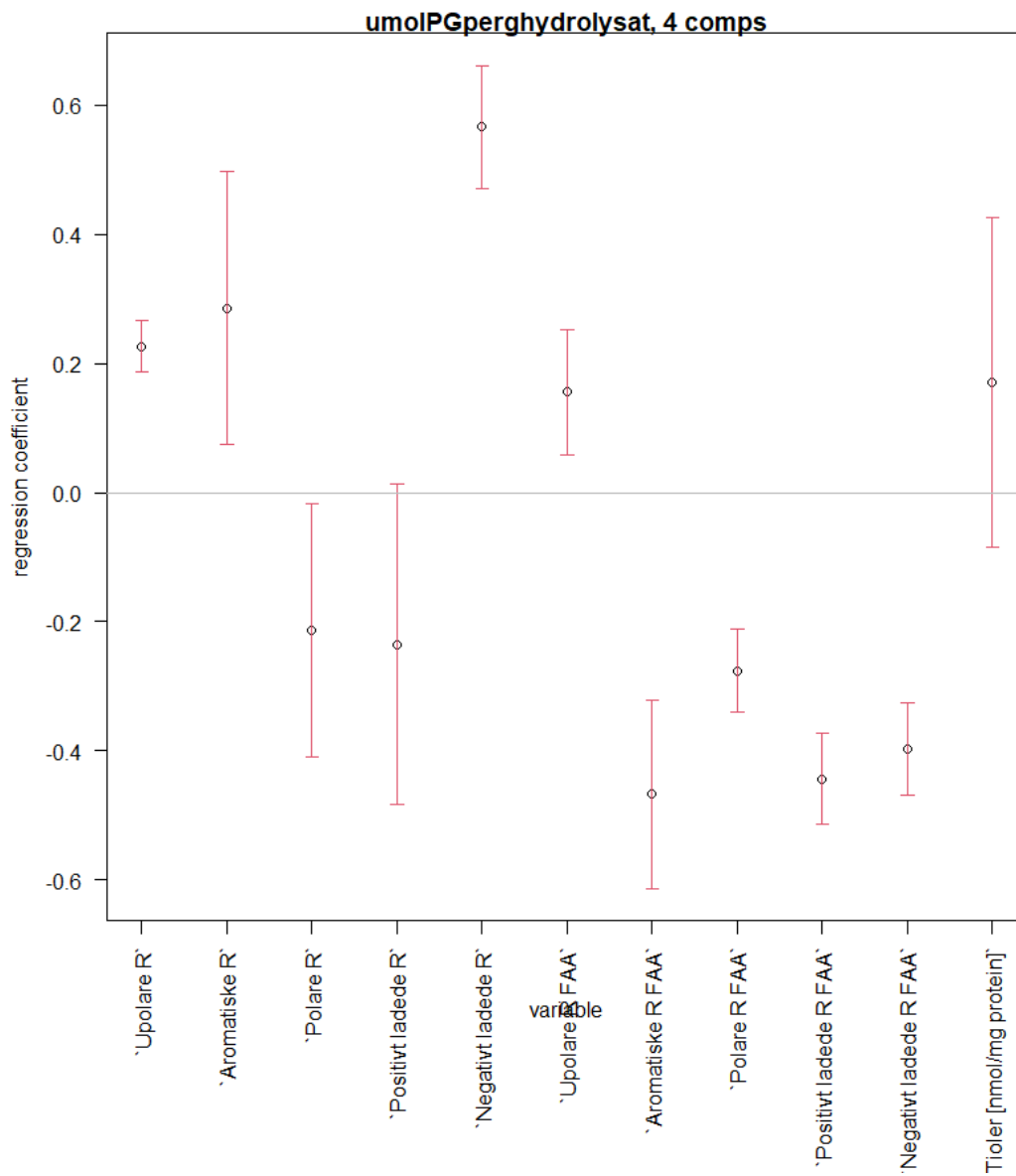


Figur B.40: Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen for matrise 6.

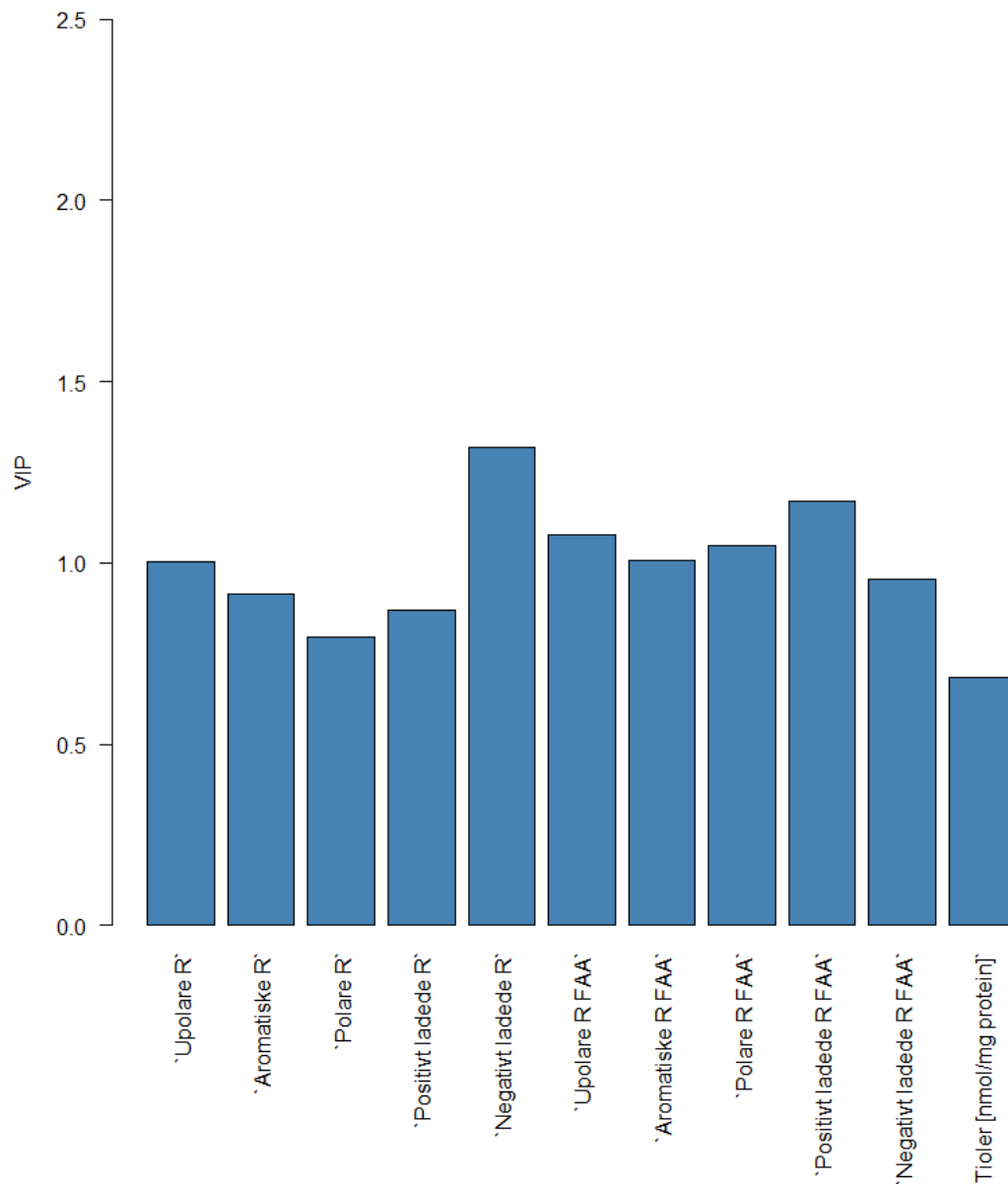




Figur B.41: Korrelasjon-loadings-plott for de to første komponentene i PLSR-modellen for matrise 6.



Figur B.42: Regresjonskoeffisientene for variablene i matrise 6 fra PLSR-modellen med fire komponenter.



Figur B.43: VIP-verdiene for variablene i den kryssvaliderte modellen av matrise 6 for fire komponenter

```

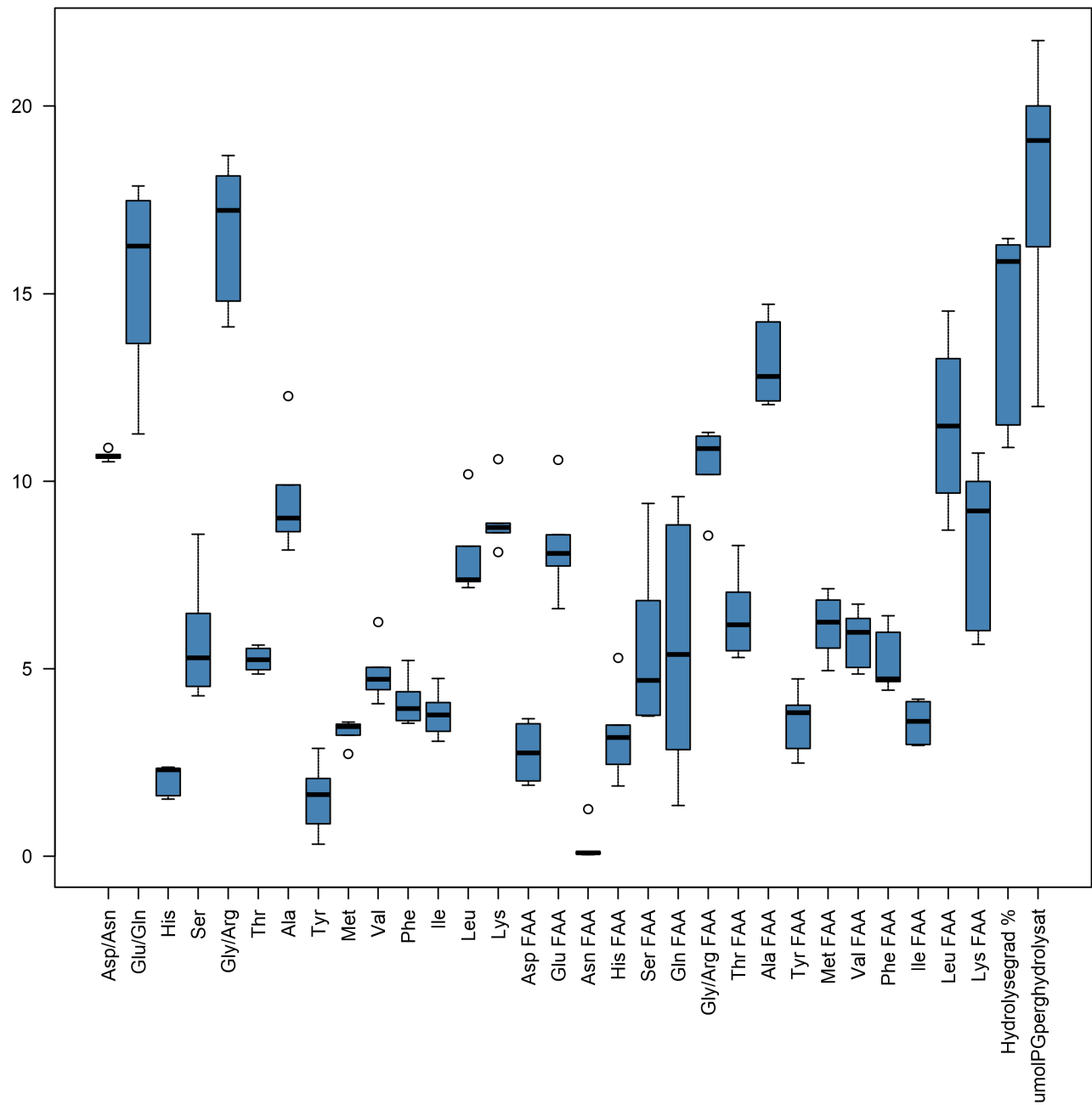
> #Jack.test: gjør t-tester av reg.koeff. basert på jackknife variance estimat
> jack.test(plsr.modell,ncomp=4, use.mean=FALSE) #replace n with #comp
Response umolPGperhydrolysat (4 comps):
      Estimate Std. Error Df t value Pr(>|t|)
`Upolare R`      0.226889  0.070100  5  3.2367 0.023034 *
`Aromatiske R`  0.285408  0.220817  5  1.2925 0.252686
`Polare R`      -0.214062  0.242477  5 -0.8828 0.417753
`Positivt ladede R` -0.235305  0.285722  5 -0.8235 0.447684
`Negativt ladede R`  0.566823  0.155375  5  3.6481 0.014777 *
`Upolare R FAA`  0.155959  0.100496  5  1.5519 0.181390
`Aromatiske R FAA` -0.466934  0.176747  5 -2.6418 0.045875 *
`Polare R FAA`   -0.275842  0.065254  5 -4.2272 0.008270 **
`Positivt ladede R FAA` -0.443364  0.096727  5 -4.5836 0.005928 **
`Negativt ladede R FAA` -0.397930  0.090596  5 -4.3923 0.007072 **
`Tioler [nmol/mg protein]` 0.170542  0.255304  5  0.6680 0.533727
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```

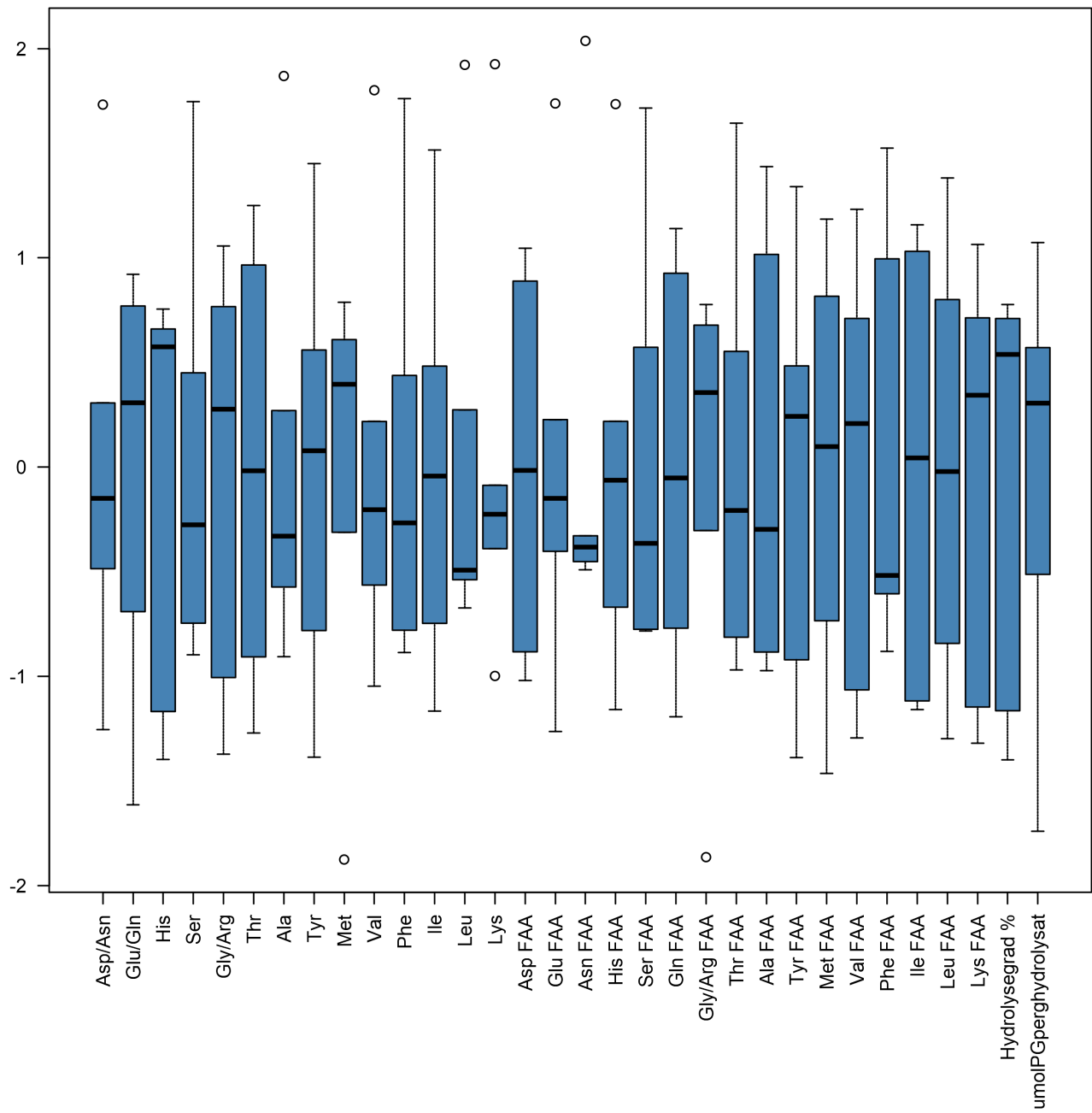
Figur B.44: Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 6, gjennom funksjonen jack.test.

## B.7 Matrise 4: TAA, FAA, hydrolysegrad, ABTS antioksidantkapasitet

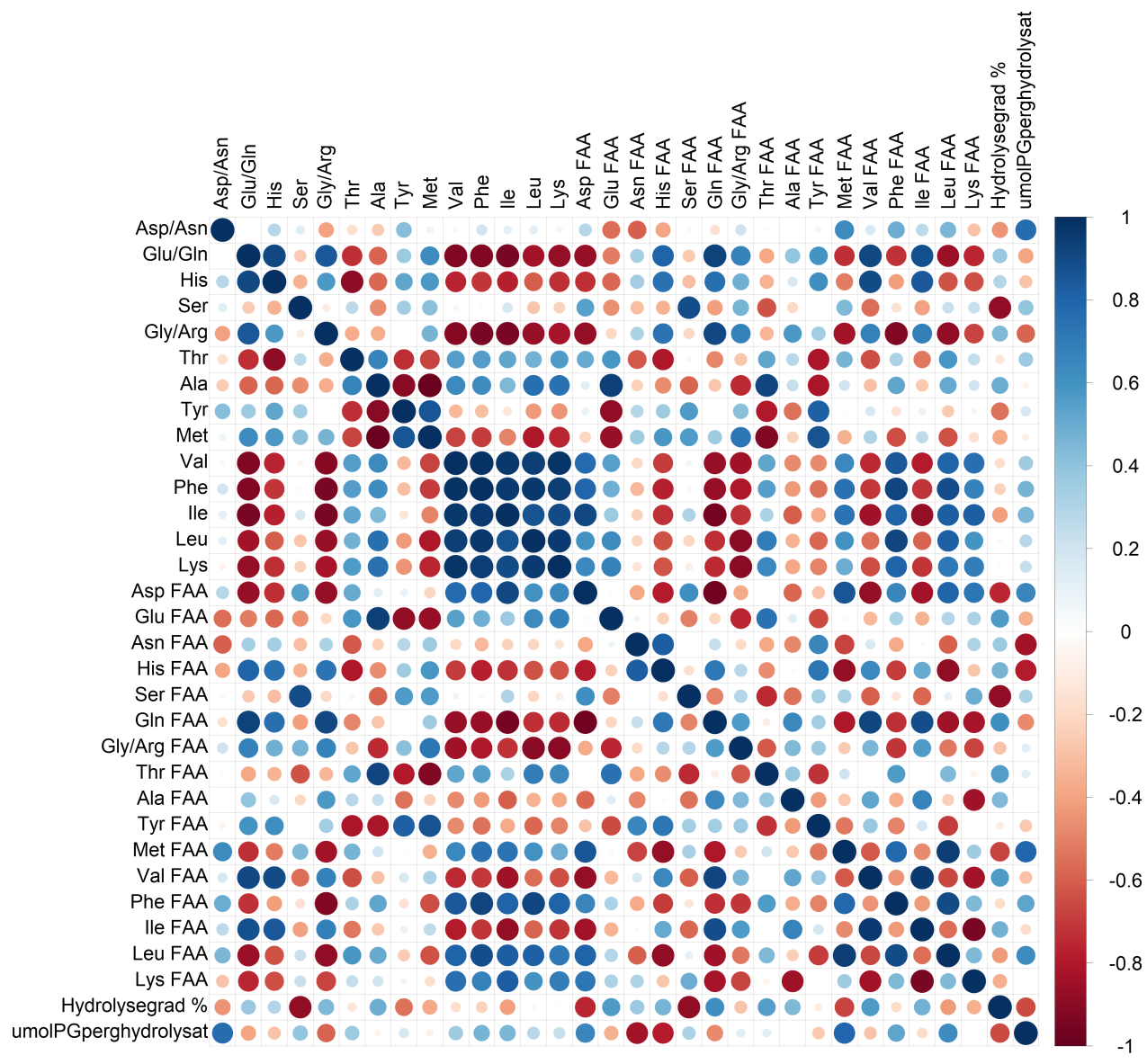
Figur B.45-B.55 viser diverse resultater av dataanalysen av matrise 4.



Figur B.45: Bokplott over de originale variablene i matrise 4.



Figur B.46: Boksplott over de standardiserte variablene i matrise 4.



Figur B.47: Korrelasjonsmatrisen for variablene i matrise 4.

```

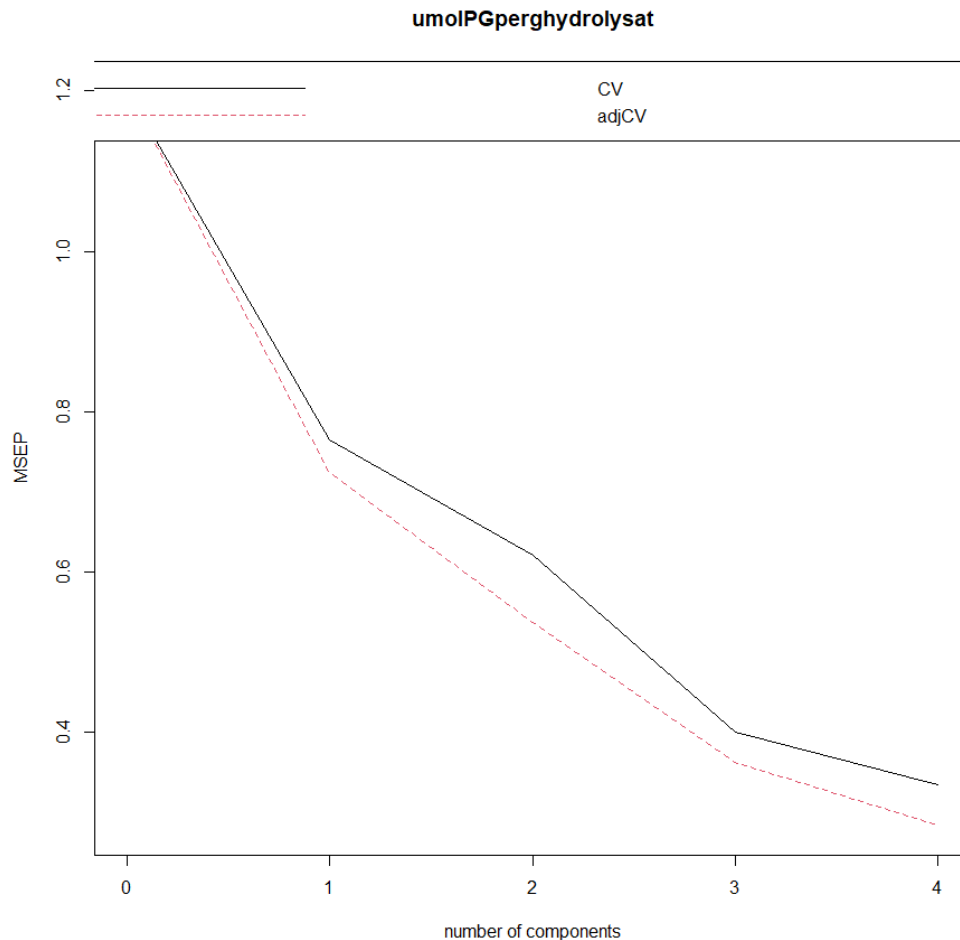
> #PLSR modell (6, 7, 8 in one)
> set.seed(9054)
> pls.model=plsr(umolPGperhydrolysat~., data=std.Hydrolysat, s
nife= TRUE,method="oscorespls")
> summary(pls.model)#CV RMSEP given and explained variance compor
Data:  X dimension: 6 31
      Y dimension: 6 1
Fit method: oscorespls
Number of components considered: 4

VALIDATION: RMSEP
Cross-validated using 6 leave-one-out segments.
      (Intercept) 1 comps 2 comps 3 comps 4 comps
CV          1.095  0.8747  0.7874  0.6323  0.5776
adjCV       1.095  0.8510  0.7323  0.6012  0.5322

TRAINING: % variance explained
      1 comps 2 comps 3 comps 4 comps
X          45.17 68.95 88.55 93.53
umolPGperhydrolysat 58.66 87.04 91.19 97.51
> #Validation and choosing number of PLS directions
> validationplot(pls.model, val.type="MSEP", legendpos="top") #l
mate for test MSE

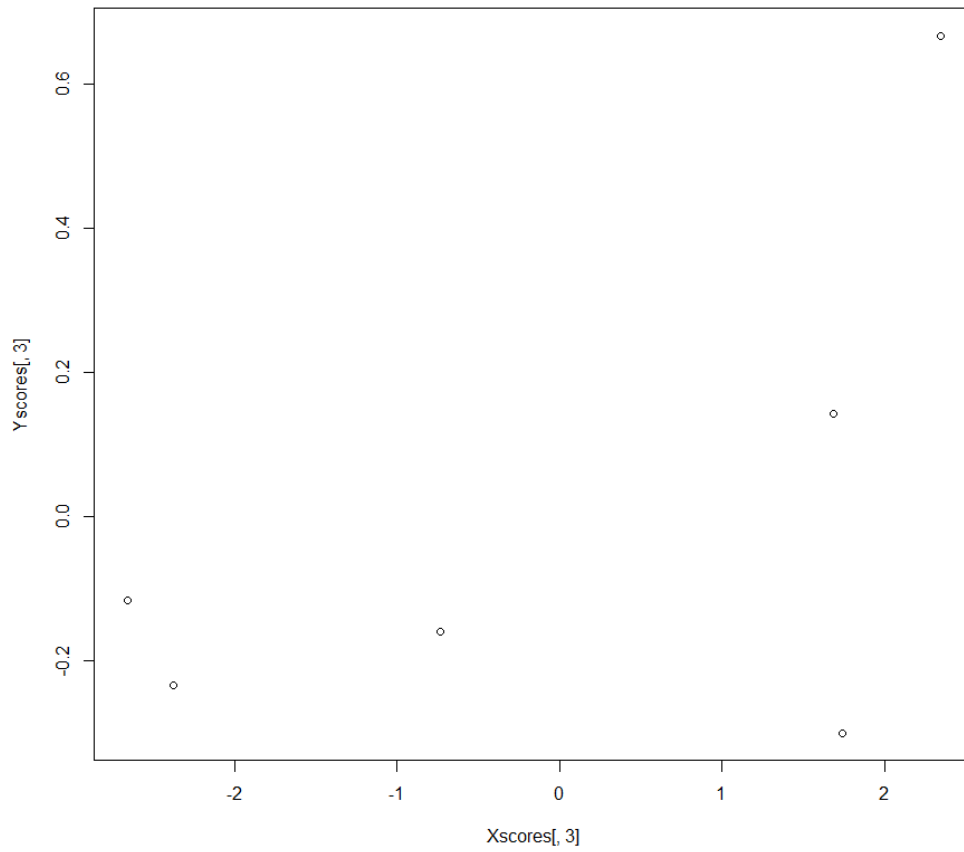
```

Figur B.48: Sammendrag av PLSR-analysen for matrise 4.

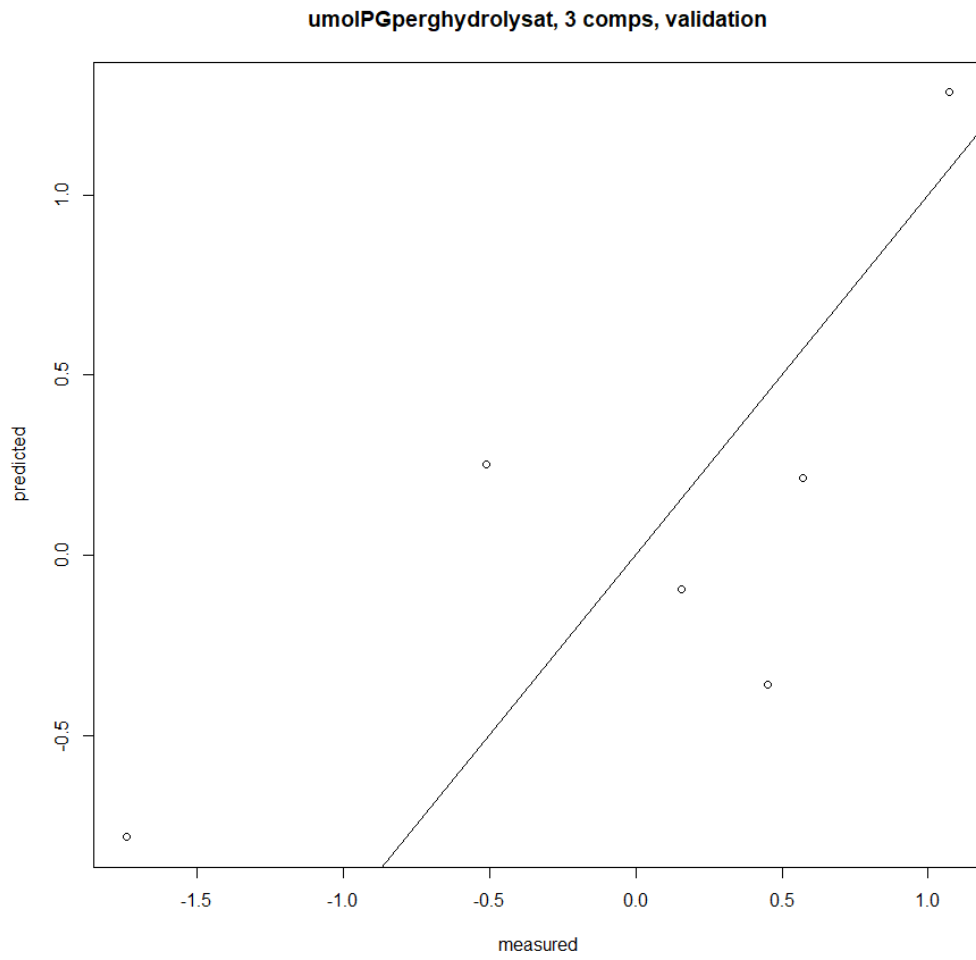


Figur B.49: MSEP for den kryssvaliderte PLSR-modellen med ulikt antall komponenter for matrise 4.

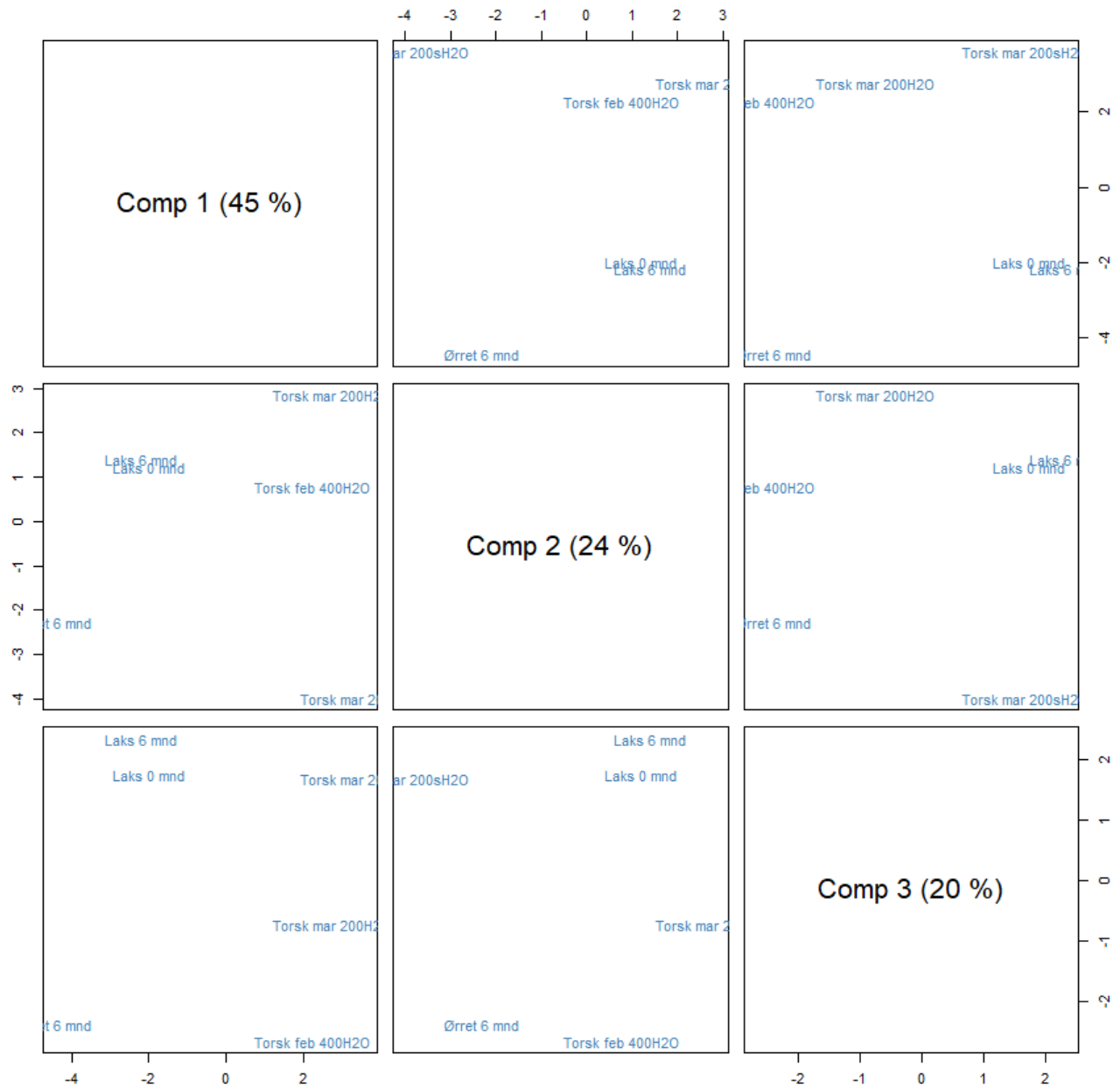




Figur B.50: Plott over X og Y-scores for de tre komponentene i PLSR-modellen for matrise 4.

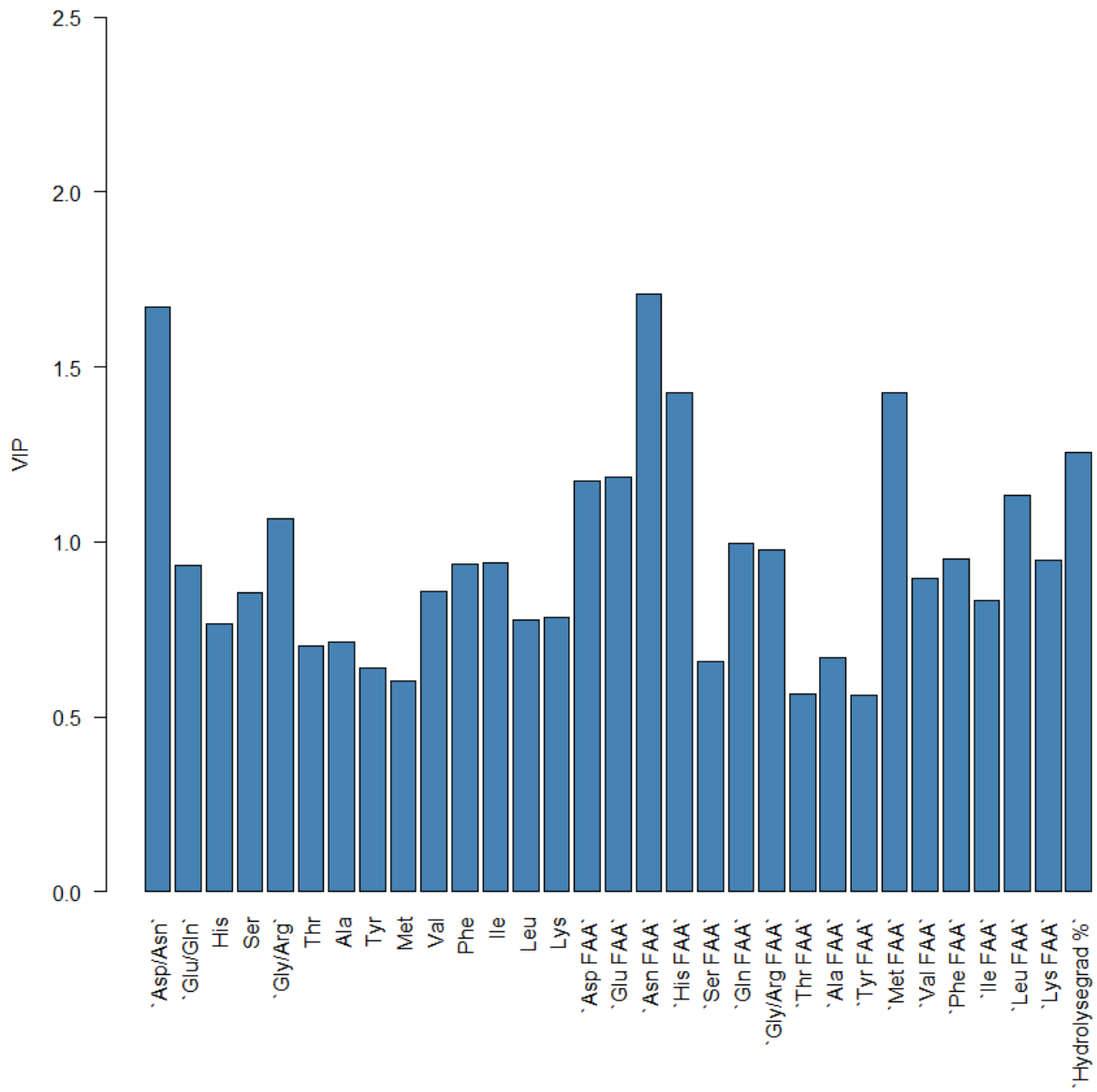


Figur B.51: Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 4.



Figur B.52: Scores-plott for alle komponentene i PLSR-modellen for matrise 4.





Figur B.54: VIP-verdiene for variablene i den kryssvaliderte modellen for matrise 4 med tre komponenter.

```

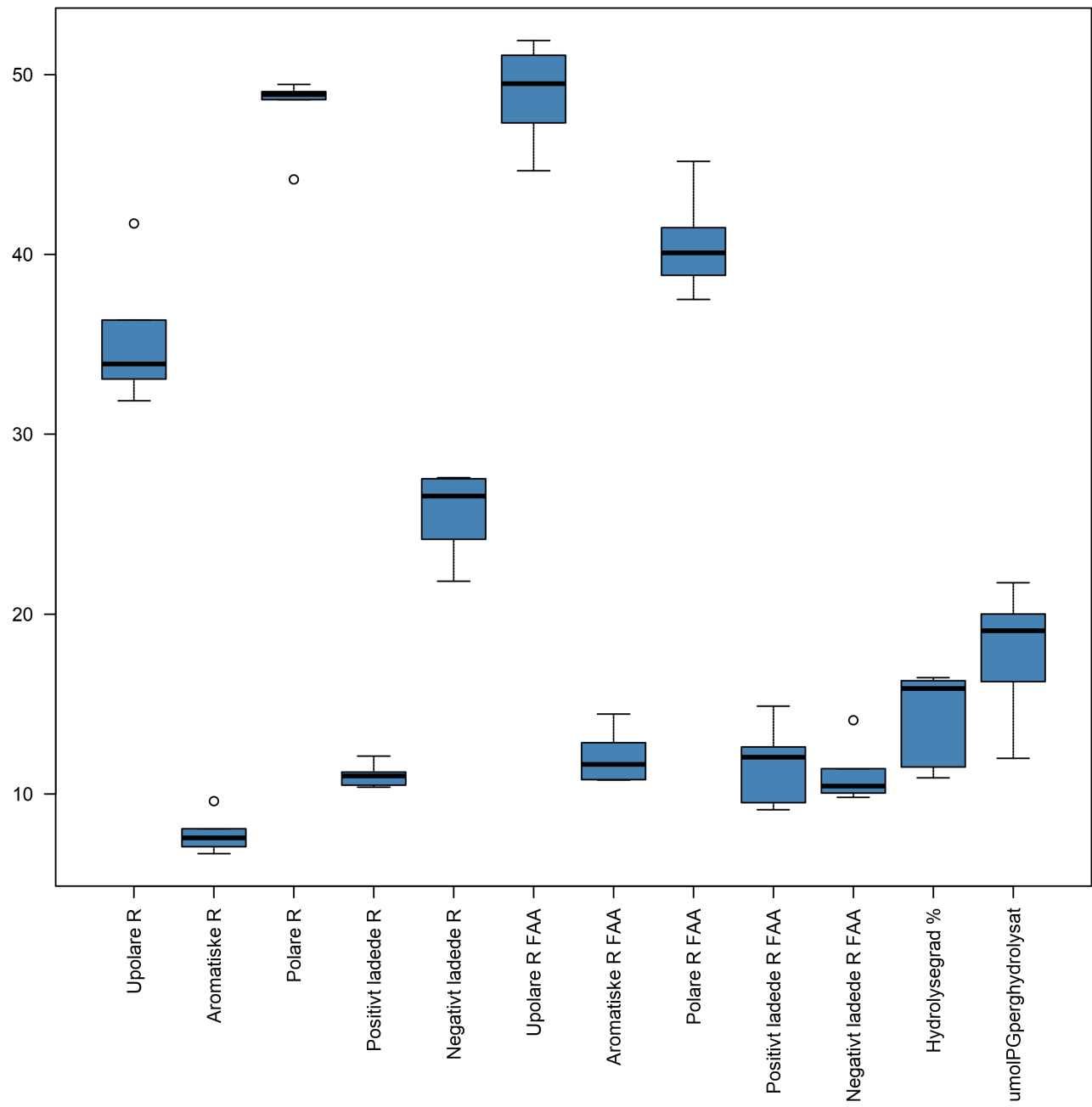
> #Jack.test: gjør t-tester av reg.koeff. basert på jackknife variance estimata
> jack.test(plsr.modell,ncomp=3, use.mean=FALSE) #replace n with #comp
Response umolPGperhydrolysat (3 comps):
      Estimate Std. Error Df t value Pr(>|t|)
`Asp/Asn`      0.19301274 0.06174135 5  3.1262 0.026070 *
`Glu/Gln`     -0.00755038 0.02810608 5 -0.2686 0.798939
His           -0.00067297 0.02852709 5 -0.0236 0.982092
Ser            0.03506718 0.10433721 5  0.3361 0.750445
`Gly/Arg`     -0.07104909 0.03492187 5 -2.0345 0.097548 .
Thr            0.04859330 0.06468594 5  0.7512 0.486352
Ala           -0.04335060 0.03488869 5 -1.2425 0.269132
Tyr            0.04910184 0.05698823 5  0.8616 0.428278
Met            0.01303947 0.04563817 5  0.2857 0.786555
Val            0.01953772 0.04912824 5  0.3977 0.707271
Phe            0.04437022 0.03277977 5  1.3536 0.233840
Ile            0.03148562 0.04099165 5  0.7681 0.477118
Leu            0.00574082 0.01997927 5  0.2873 0.785381
Lys            0.00518807 0.06880666 5  0.0754 0.942820
`Asp FAA`     0.07652757 0.03137862 5  2.4388 0.058737 .
`Glu FAA`    -0.12212361 0.06773197 5 -1.8030 0.131243
`Asn FAA`    -0.20480103 0.11875088 5 -1.7246 0.145193
`His FAA`    -0.14226525 0.07680799 5 -1.8522 0.123197
`Ser FAA`     0.02157387 0.03067464 5  0.7033 0.513251
`Gln FAA`    -0.01441969 0.04292263 5 -0.3359 0.750550
`Gly/Arg FAA` 0.08512954 0.05861030 5  1.4525 0.206096
`Thr FAA`     0.04556625 0.07424827 5  0.6137 0.566255
`Ala FAA`     0.06948750 0.08525105 5  0.8151 0.452082
`Tyr FAA`    -0.01635476 0.13004353 5 -0.1258 0.904819
`Met FAA`     0.11761271 0.06582336 5  1.7868 0.134017
`Val FAA`     0.03281055 0.06443826 5  0.5092 0.632295
`Phe FAA`     0.06945849 0.01062828 5  6.5353 0.001255 **
`Ile FAA`     0.04256345 0.03662684 5  1.1621 0.297651
`Leu FAA`     0.06775241 0.07546673 5  0.8978 0.410447
`Lys FAA`    -0.08867660 0.06838814 5 -1.2967 0.251359
`Hydrolysegrad %` -0.10043782 0.04143448 5 -2.4240 0.059820 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |

```

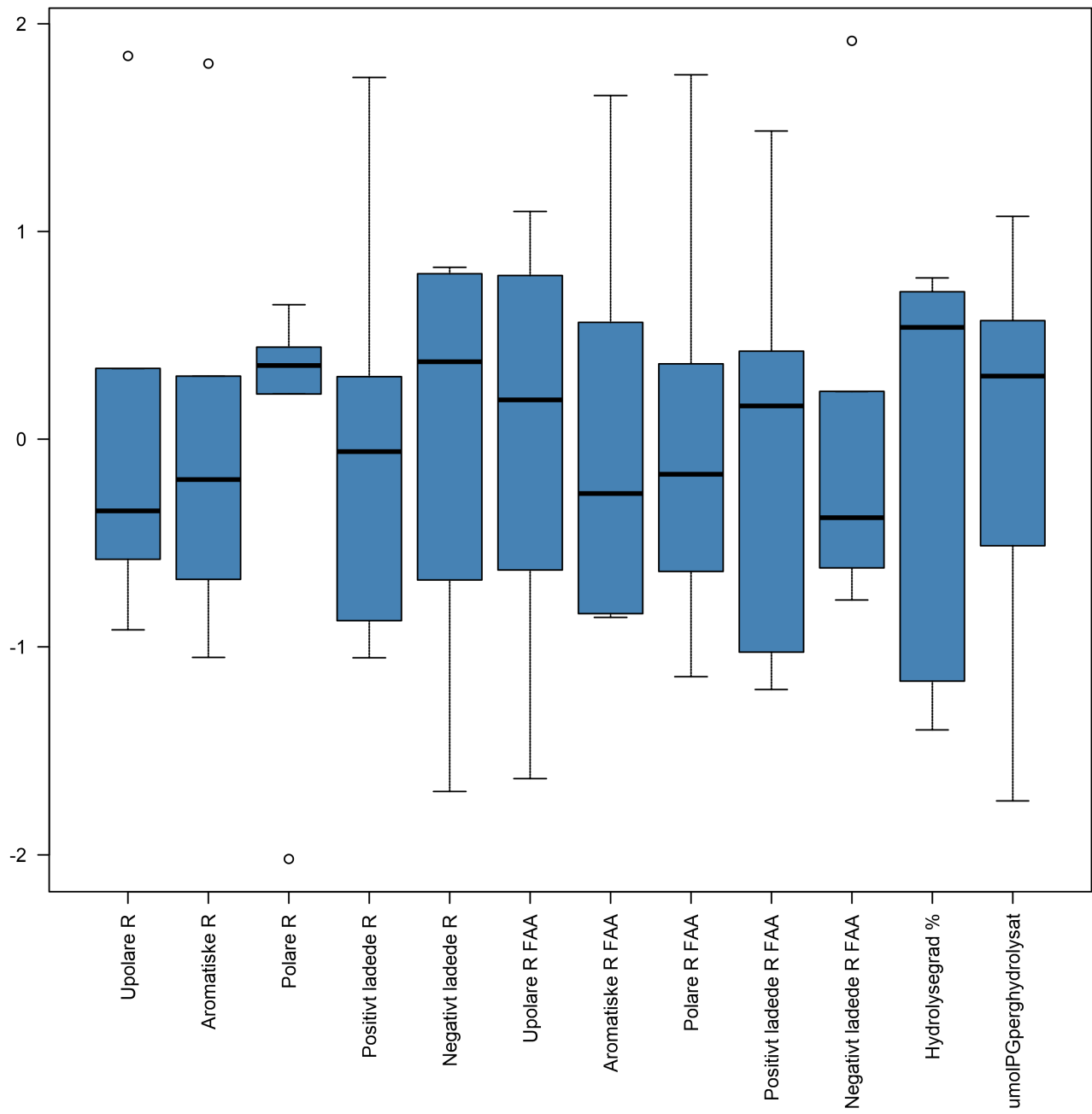
Figur B.55: Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 4, gjennom funksjonen jack.test.

## B.8 Matrise 8: TAA, FAA, hydrolysegrad, ABTS antioksidantkapasitet

Figur B.56-B.69 viser diverse resultater av dataanalysen av matrise 8.

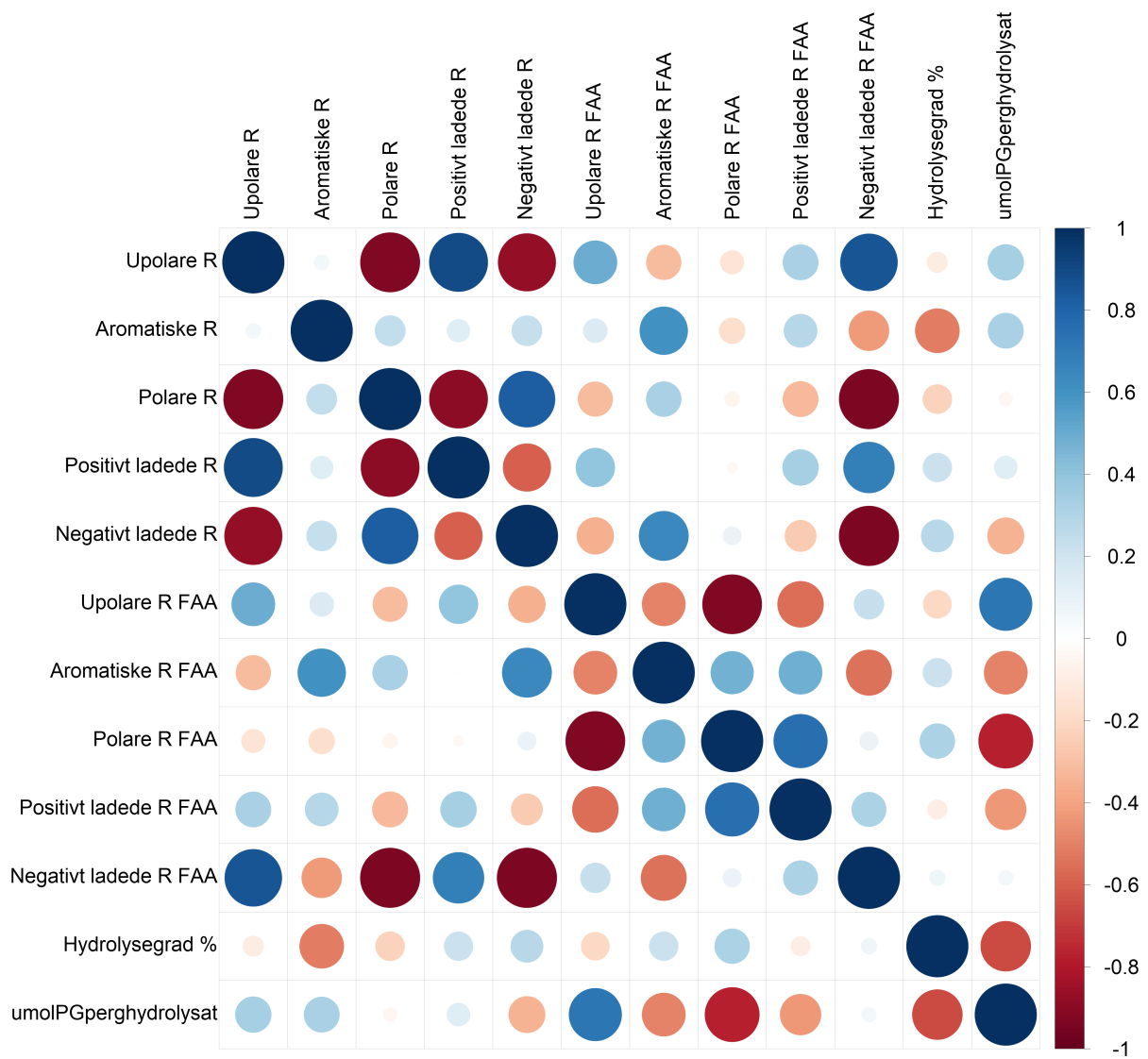


Figur B.56: Bokplott over de originale variablene i matrise 8.



Figur B.57: Boksplott over de standardiserte variablene i matrise 8.





Figur B.58: Korrelasjonsmatrisen for variablene i matrise 8.

```

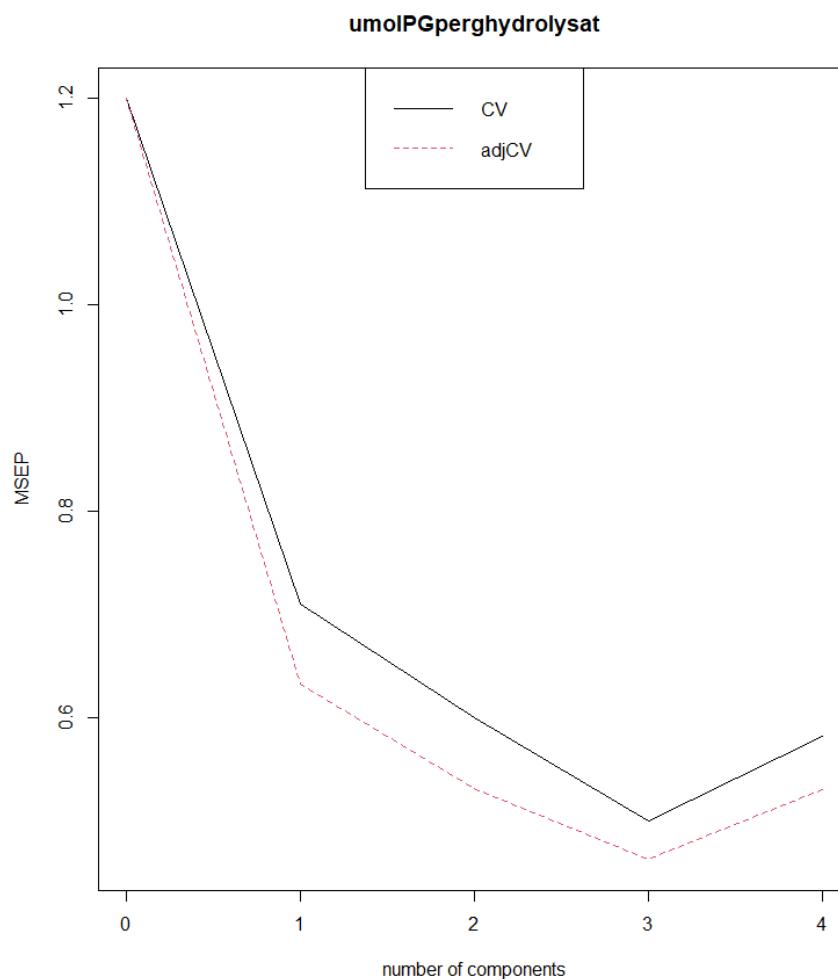
> #PLSR MODELL (0, 1, 0 III OIE)
> set.seed(9055)
> pls.model=plsr(umoIPGperghydrolysat~., data=std.Hydrolysates, scale=FALSE, cente
> summary(pls.model)#CV RMSEP given and explained variance components for fitting
Data: X dimension: 6 11
      Y dimension: 6 1
Fit method: oscorespls
Number of components considered: 4

VALIDATION: RMSEP
Cross-validated using 6 leave-one-out segments.
      (Intercept) 1 comps 2 comps 3 comps 4 comps
CV          1.095  0.8424  0.7743  0.7071  0.7631
adjCV       1.095  0.7951  0.7293  0.6800  0.7280

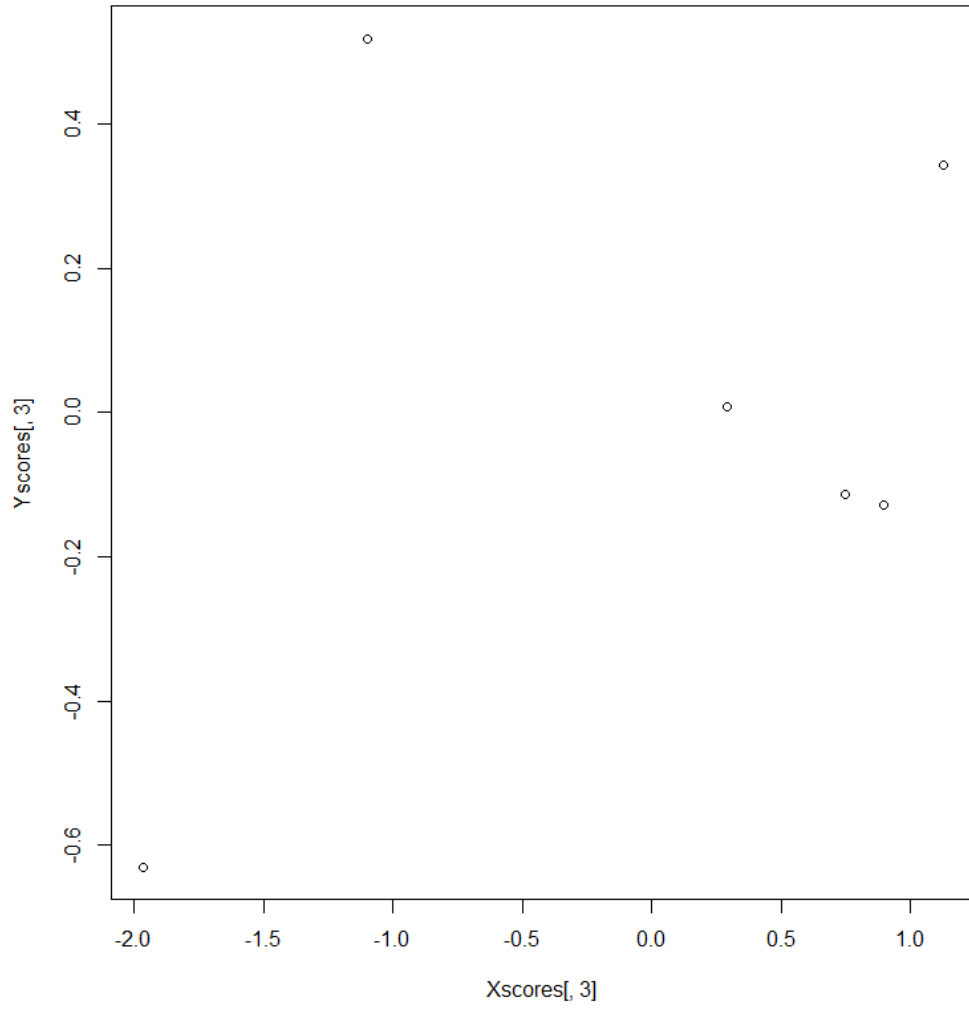
TRAINING: % variance explained
              1 comps 2 comps 3 comps 4 comps
X              30.75  66.33  86.09  96.89
umoIPGperghydrolysat 76.91  83.73  85.63  89.60
> #Validation and choosing number of PLS directions
> validationplot(pls.model, val.type="MSEP", legendpos="top") #look for the lowest
> |

```

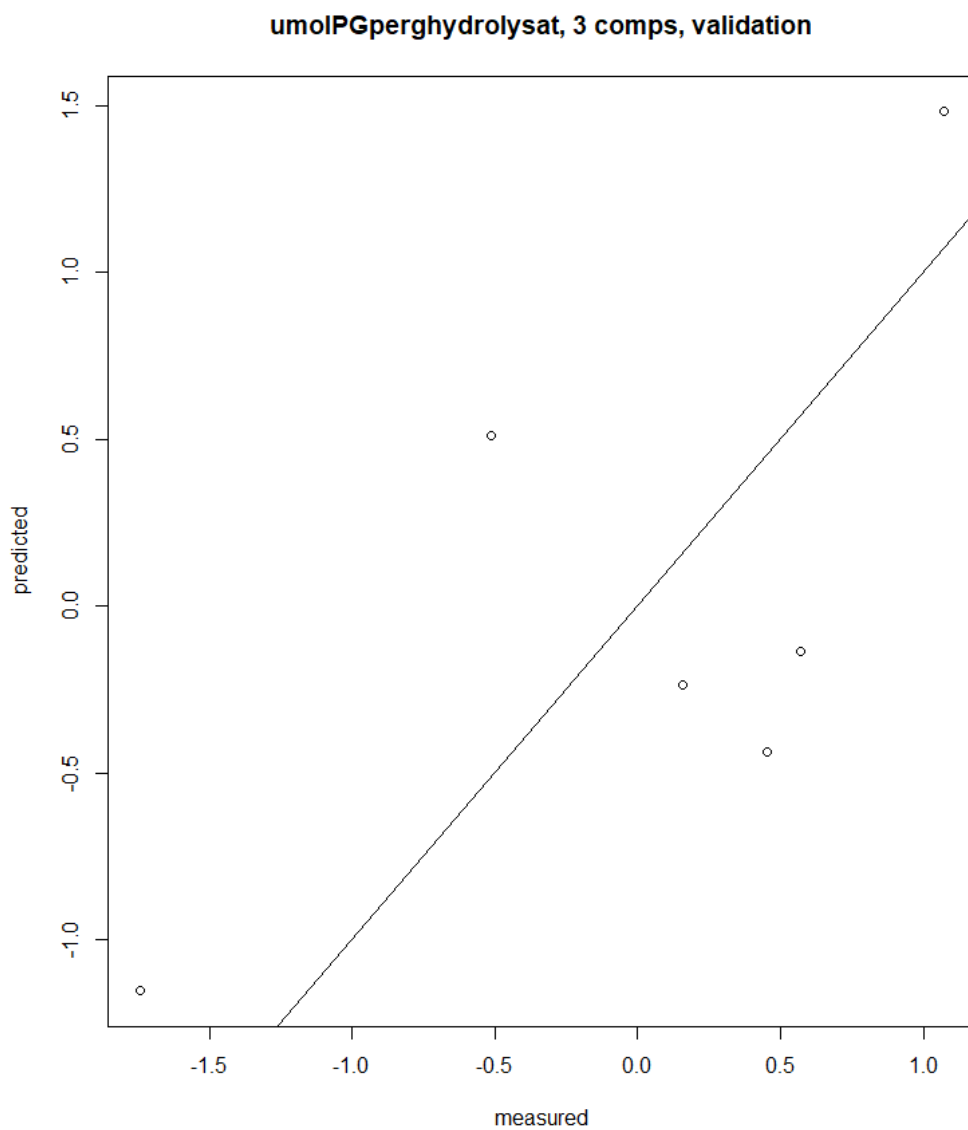
Figur B.59: Sammendrag av PLSR-analysen for matrise 8.



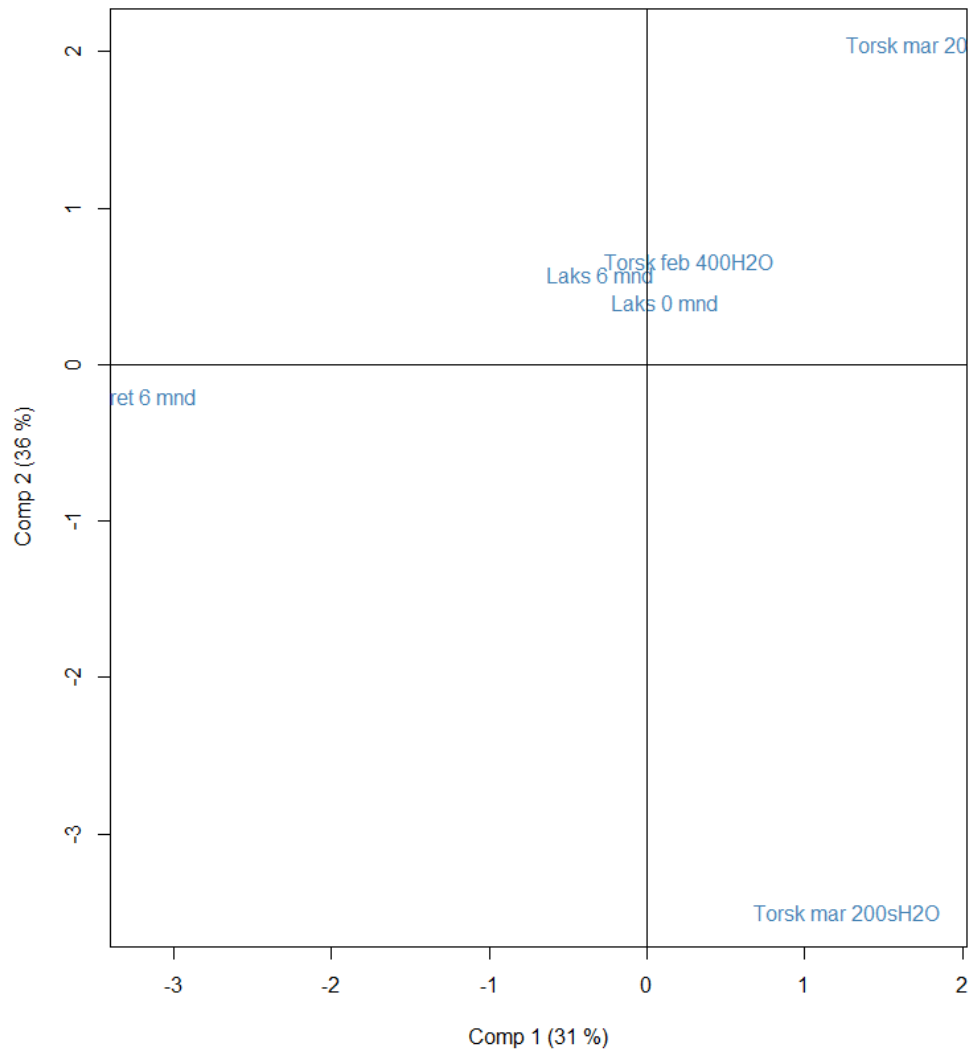
Figur B.60: MSEP for den kryssvaliderte PLSR-modellen med ulikt antall komponenter for matrise 8.



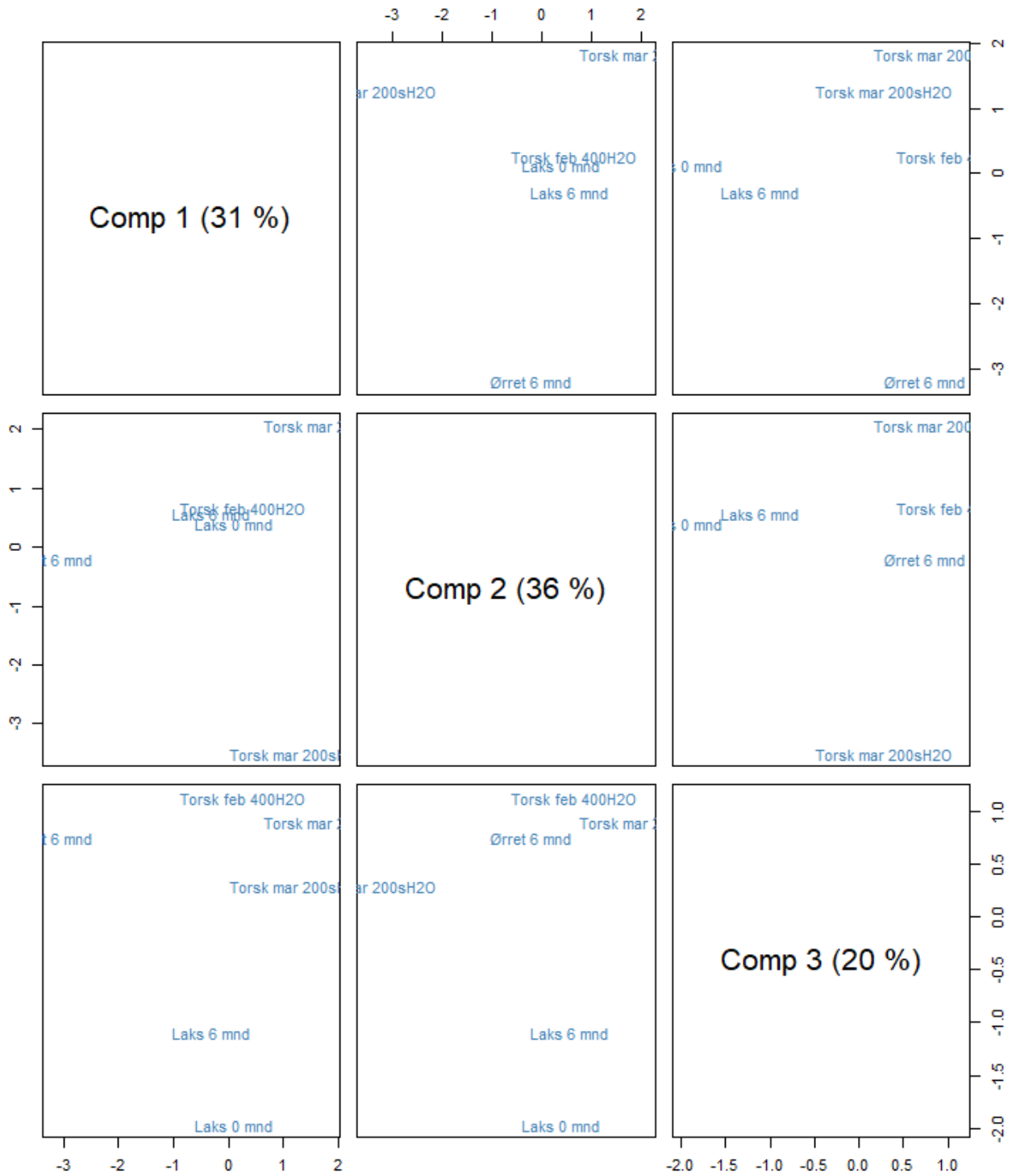
Figur B.61: Plott over X og Y-scores for PLSR-modellen med tre komponenter for matrise 8.



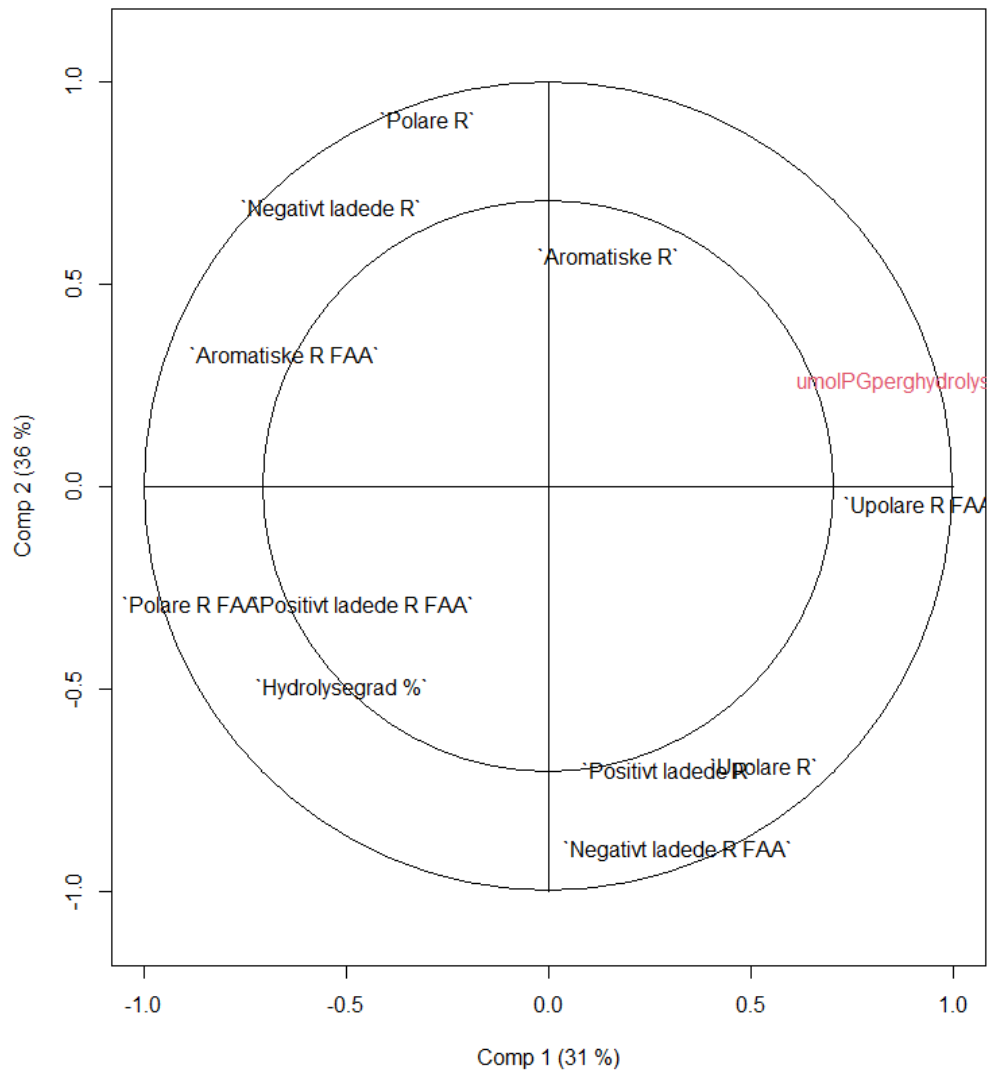
Figur B.62: Kryssvaliderte prediksjoner for ABTS-antioksidantkapasitet (standardisert form) for prøvene i matrise 8.



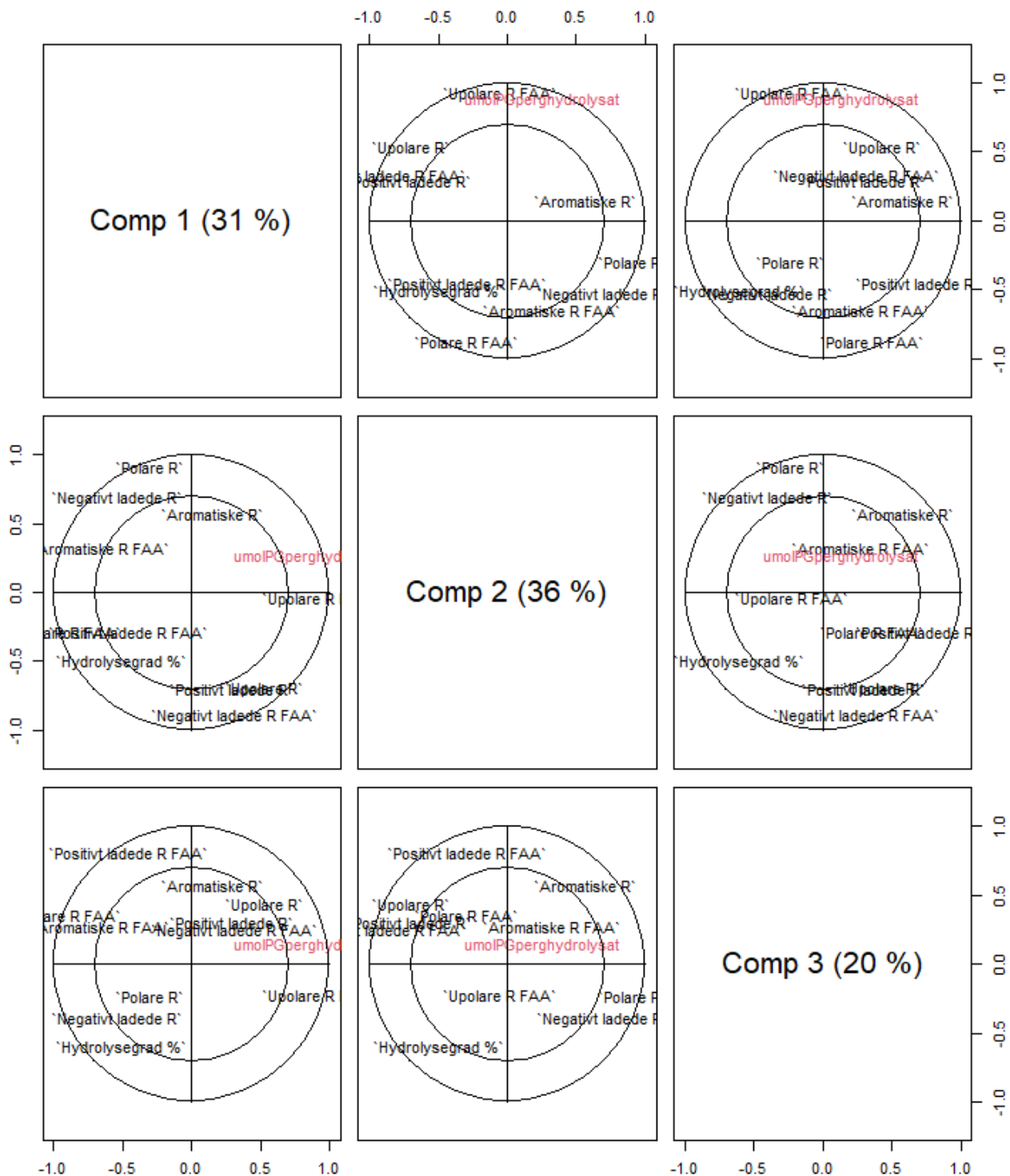
Figur B.63: Scores-plott for de to første komponentene i PLSR-modellen for matrise 8.



Figur B.64: Scores-plott for alle komponentene i PLSR-modellen for matrise 8.

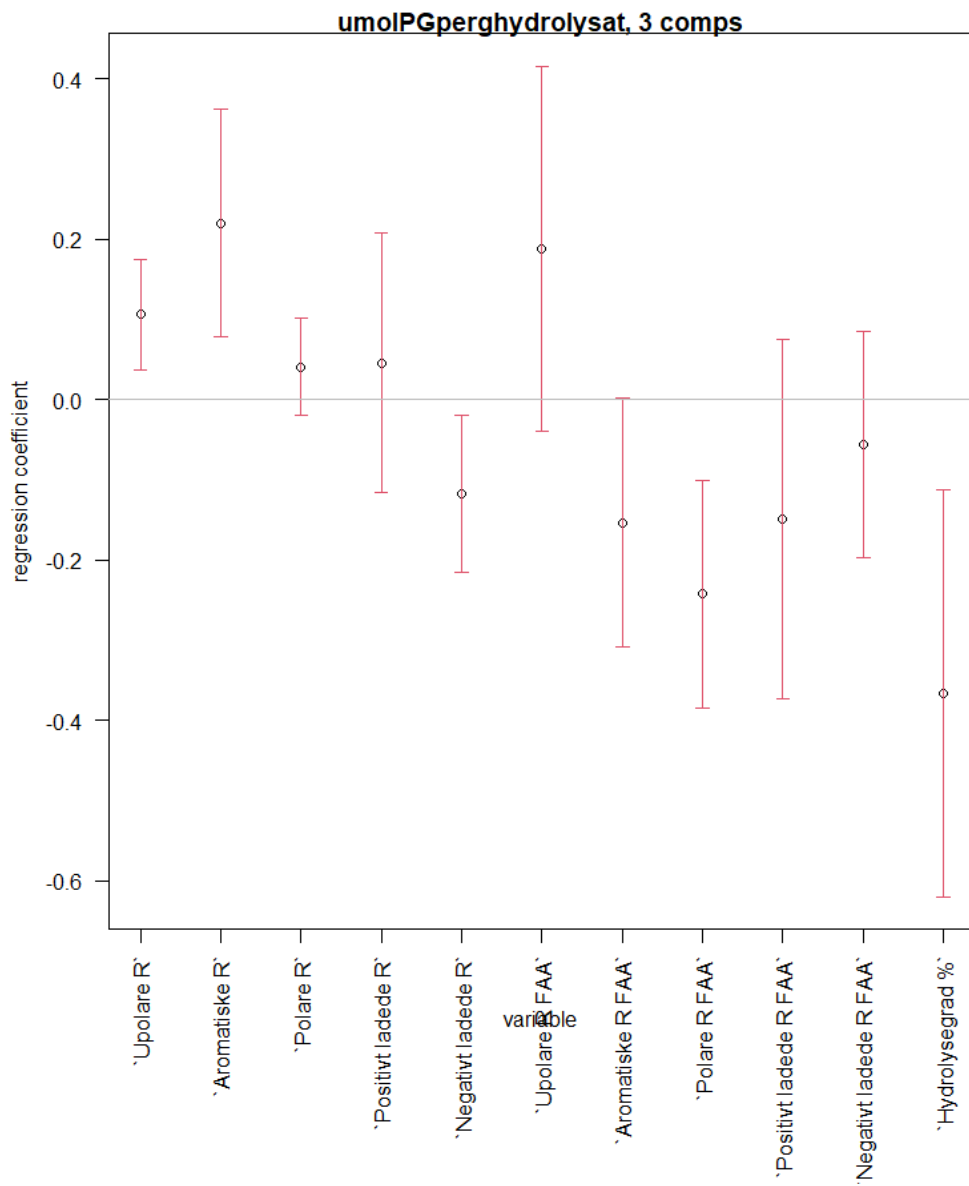


Figur B.65: Korrelasjon-loadings-plott for de to første komponentene i PLSR-modellen for matrise 8.

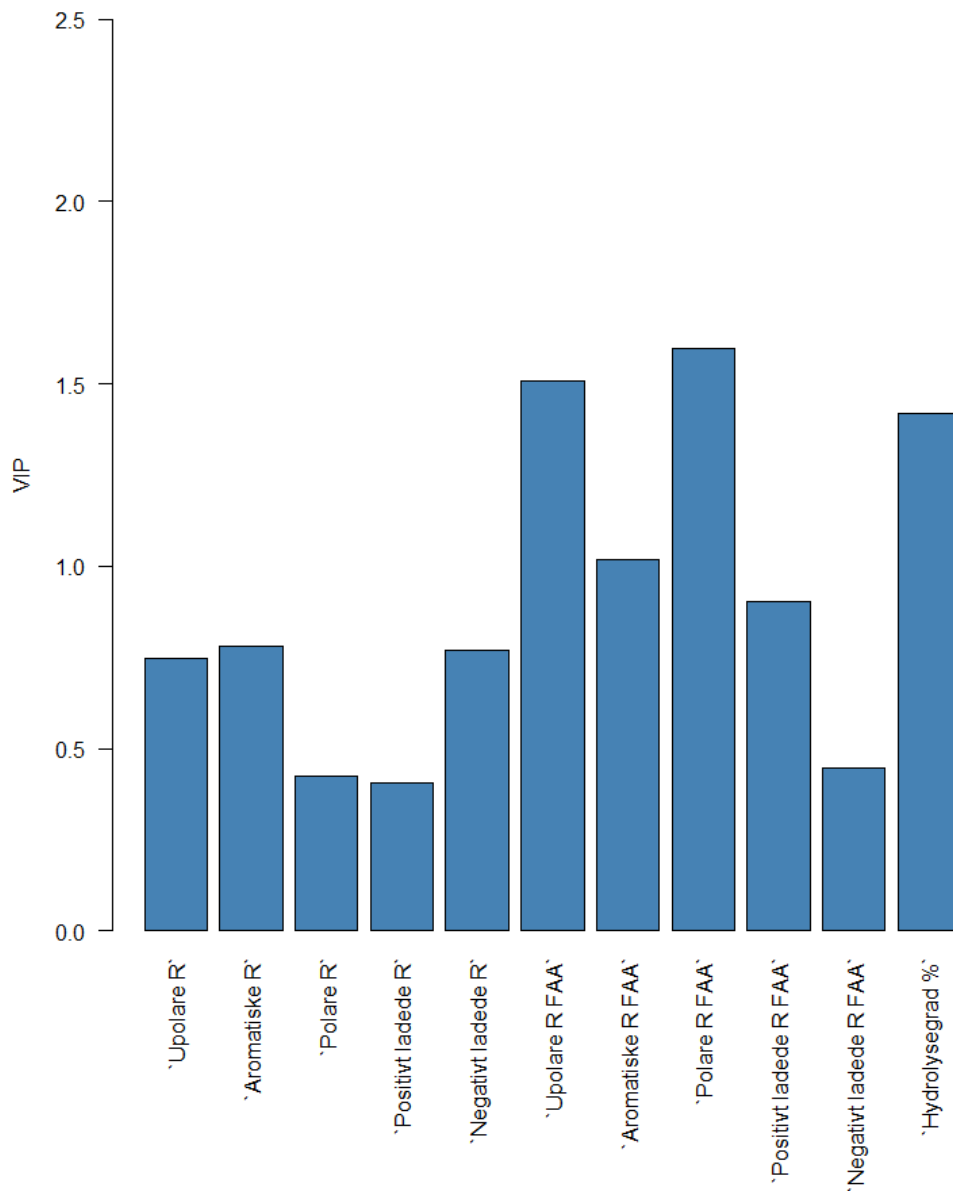


Figur B.66: Korrelasjon-loadings-plott for alle komponentene i PLSR-modellen for matrise 8.





Figur B.67: Regresjonskoeffisientene for variablene i den kryssvaliderte modellen av matrise 8 for tre komponenter.



Figur B.68: VIP-verdiene for variablene i den kryssvaliderte PLSR-modellen av matrise 8 for tre komponenter.

```

> jack.test(plsr.modell,ncomp=3, use.mean=FALSE) #replace n with #comp
Response umolPGperhydrolysat (3 comps):
      Estimate Std. Error Df t value Pr(>|t|)
`Upolare R`      0.106593  0.070871  5  1.5040  0.1929
`Aromatiske R`   0.219941  0.143077  5  1.5372  0.1848
`Polare R`       0.041302  0.065439  5  0.6312  0.5557
`Positivt ladede R` 0.046130  0.166794  5  0.2766  0.7932
`Negativt ladede R` -0.117555  0.098166  5 -1.1975  0.2848
`Upolare R FAA`   0.188226  0.251512  5  0.7484  0.4879
`Aromatiske R FAA` -0.152992  0.166258  5 -0.9202  0.3997
`Polare R FAA`   -0.242476  0.153060  5 -1.5842  0.1740
`Positivt ladede R FAA` -0.148988  0.248947  5 -0.5985  0.5756
`Negativt ladede R FAA` -0.055726  0.154314  5 -0.3611  0.7328
`Hydrolysegrad %` -0.365906  0.257280  5 -1.4222  0.2142
> |

```

Figur B.69: Resultatene av grov signifikanstest (t-test) for regresjonskoeffisientene i matrise 8, gjennom funksjonen jack.test.

