

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
12. April 2012 (12.04.2012)

(10) Internationale Veröffentlichungsnummer
WO 2012/045396 A2

- (51) Internationale Patentklassifikation: Nicht klassifiziert
- (21) Internationales Aktenzeichen: PCT/EP2011/004687
- (22) Internationales Anmeldedatum:
20. September 2011 (20.09.2011)
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität:
10 2010 047 427.4
4. Oktober 2010 (04.10.2010) DE
- (71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): **KARLSRUHER INSTITUT FÜR TECHNOLOGIE** [DE/DE]; Kaiserstrasse 12, 76131 Karlsruhe (DE).
- (72) Erfinder; und
- (75) Erfinder/Anmelder (nur für US): **HUBBUCH, Jürgen** [DE/DE]; Weltzienstr. 37, 76135 Karlsruhe (DE). **HANSEN, Sigrid** [DK/DE]; Weltzienstr. 3, 76135 Karlsruhe

(DE). **SKIBSTED, Erik** [DK/DK]; Diget 12, DK-4300 Holbæk (DK). **KRARUP, Janus C.** [DK/DK]; Lorcksvej 5, DK-2820 Gentofte (DK). **SØRENSEN, Hans Holme-gaard** [DK/DK]; Joachim Rønnows Vej 21, DK-2830 Vi-rum (DK). **STABY, Arne** [DK/DK]; Gunnarsvej 17, DK-2880 Bagsværd (DK).

(74) **Anwalt: HAYDN, Christian**; Müller-Boré & Partner, Grafinger Str. 2, 81671 München (DE).

(81) **Bestimmungsstaaten** (soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[Fortsetzung auf der nächsten Seite]

(54) **Title:** SELECTIVE PROTEIN QUANTIFICATION BY MEANS OF MULTIVARIATE EVALUATION OF UV ABSORPTION SPECTRA

(54) **Bezeichnung :** SELEKTIVE PROTEINQUANTIFIZIERUNG MITTELS MULTIVARIATER AUSWERTUNG VON UV-ABSORPTIONSSPEKTREN

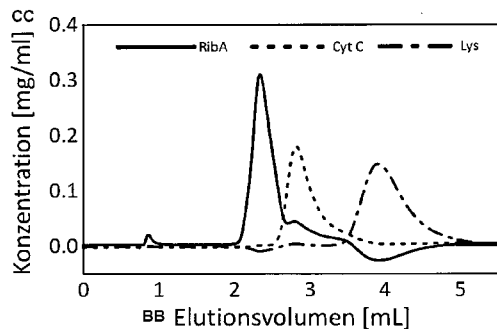
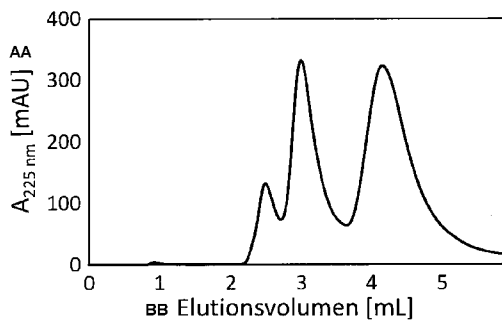


Figure 6

AA $A_{225 \text{ nm}}$ [mAU]
BB Elution volume [ml]
CC Concentration [mg/ml]

(57) **Abstract:** The present invention relates to a method for the selective quantification of individual proteins in a protein mixture. This method is based on the determination of the individual concentrations of the proteins in the protein mixture from a UV absorption spectrum of the protein mixture. Furthermore, the present invention relates to a system, based on this method, for the selective quantification of individual proteins in a protein mixture.

(57) **Zusammenfassung:** Die vorliegende Erfindung betrifft ein Verfahren zur selektiven Quantifizierung einzelner Proteine in einem Proteingemisch. Dieses Verfahren basiert auf der Bestimmung der Einzelkonzentrationen der Proteine in dem Proteingemisch aus einem UV-Absorptionsspektrum des Proteingemisches. Weiter betrifft die vorliegende Erfindung ein auf diesem Verfahren basierendes System zur selektiven Quantifizierung einzelner Proteine in einem Proteingemisch.

WO 2012/045396 A2



(84) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,

RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts (Regel 48 Absatz 2 Buchstabe g)

5 "Selektive Proteinquantifizierung mittels multivariater Auswertung
von UV-Absorptionsspektren"

Beschreibung

10 Die vorliegende Erfindung betrifft ein Verfahren zur selektiven Quantifizierung
einzelner Proteine in einem Proteingemisch. Dieses Verfahren basiert auf der
Bestimmung der Einzelkonzentrationen der Proteine in dem Proteingemisch aus
einem UV-Absorptionsspektrum des Proteingemisches. Weiter betrifft die
vorliegende Erfindung ein auf diesem Verfahren basierendes System zur selektiven
Quantifizierung einzelner Proteine in einem Proteingemisch.

15

Die meisten Proteine absorbieren aufgrund der aromatischen Strukturen der darin
enthaltenen Aminosäuren Tyrosin, Tryptophan und Phenylalanin elektromagnetische
Strahlung im mittleren ultravioletten Bereich (etwa 210 bis 300 nm). Die
Absorptionsspektren dieser Aminosäuren weisen zwei Absorptionsmaxima auf, ein
20 intensives Maximum im Bereich von 220 bis 230 nm und ein weniger intensives
Maximum im Bereich von 250 bis 290 nm (vgl. Fig. 1). Das Absorptionsverhalten von
Proteinen, die diese aromatischen Aminosäuren enthalten, weist die gleichen
charakteristischen Absorptionsmaxima auf. Diese Eigenschaft, UV-Strahlung zu
absorbieren, wird oft zur Quantifizierung von gelösten Proteinen herangezogen.
25 Häufig wird dabei die Absorption bei einem der beiden Absorptionsmaxima, z.B. bei
220 nm oder 280 nm, gemessen. Diese Herangehensweise erlaubt jedoch lediglich
die Messung der Gesamtkonzentration aller Proteine in der Lösung, ohne eine
selektive Aussage, d.h. eine Aussage über die Einzelkonzentrationen bzw. die
Mengen der jeweils einzelnen in der Lösung enthaltenen Proteine, machen zu
30 können.

Für die präparative Aufreinigung von Proteinen, z.B. in der pharmazeutischen
Industrie, wird sehr häufig Säulenchromatographie mit einer flüssigen mobilen Phase
eingesetzt. Hierbei wird standardmäßig die erzielte chromatographische Auftrennung
35 der Proteine mit Hilfe von UV-Absorptionsmessungen bei einer einzelnen
Wellenlänge im mittleren UV-Bereich am Säulenausgang visualisiert bzw. analysiert.

So erstellte Chromatogramme zeigen typischerweise die über das Elutionsvolumen aufgetragene UV-Absorption des Eluats bei einer bestimmten Wellenlänge, wobei diese UV-Absorption mit der Gesamtkonzentration aller Proteine im Eluat korreliert. Weitere prozessanalytische Parameter, die bei einem solchen Verfahren erfaßt werden können, sind beispielsweise pH-Wert, Leitfähigkeit und Temperatur des Eluats am Säulenausgang.

In der Entwicklung von Prozessen zur Proteinaufreinigung oder von Screeningverfahren werden in jüngerer Zeit vermehrt miniaturisierte Experimente automatisiert in Hochdurchsatzverfahren durchgeführt. Für die Auswertung solcher Experimente ist ebenfalls eine selektive Proteinquantifizierung notwendig. Geeignete analytische Verfahren hierfür müßten einen vergleichbaren Durchsatz haben und einfach zu automatisieren sein. Entsprechende analytische Meßverfahren, die auf der Basis photospektrometrischer Daten Proteingemische qualitativ und/oder quantitativ analysieren könnten, sind dabei im Stand der Technik nicht bekannt.

Der Begriff „selektive Proteinanalytik“ deckt Verfahren ab, die in der Lage sind, die verschiedenen Proteine, die in einer Probe vorhanden sind, gleichzeitig nachzuweisen und gegebenenfalls zu quantifizieren. Für die selektive Analytik komplexer Mehrkomponentengemische, die keine Proteine enthalten, sind im Stand der Technik Verfahren bekannt, die multivariate Kalibrierungsmethoden verwenden, um anhand von Spektraldaten die Zusammensetzung dieser komplexen Gemische zu bestimmen. Solche Verfahren wurden allerdings bisher nicht für Proteingemische verwendet, da die Unterschiede zwischen den UV-Absorptionsspektren verschiedener Proteine als zu gering erachtet wurden, um multivariate Kalibrierungsmethoden erfolgreich auf komplexe Proteingemische anwenden zu können.

Der vorliegenden Erfindung liegt daher die Aufgabe zu Grunde, ein einfaches und schnelles Verfahren zur selektiven Quantifizierung einzelner Proteine in einem Proteingemisch bereitzustellen. Dieses Verfahren soll die Nachteile bestehender Verfahren überwinden können. Insbesondere soll das Verfahren ohne aufwendige Probenvorbereitung auskommen, leicht zu automatisieren sein und eine schnelle Analytik von Proteingemischen ermöglichen.

Diese Aufgabe wird durch die in den Ansprüchen gekennzeichneten Ausführungsformen der vorliegenden Erfindung gelöst.

5 Dementsprechend betrifft ein Gegenstand der vorliegenden Erfindung ein Verfahren zur selektiven Quantifizierung einzelner Proteine in einem Proteingemisch, umfassend die Schritte:

- (a) Erstellen eines UV-Absorptionsspektrums des Proteingemisches, und
- (b) Bestimmen der Einzelkonzentration zumindest eines Proteins in dem Proteingemisch aus dem in Schritt (a) erstellten UV-Absorptionsspektrum.

10

Gemäß der vorliegenden Erfindung wurden Nachteile von bestehenden Verfahren zur selektiven Proteinanalytik, wie Massenspektrometrie, Gelelektrophorese und Kapillarelektrophorese überwunden. Beispielsweise ist zur Durchführung der genannten herkömmlichen Verfahren meist eine mehr oder weniger aufwendige
15 Probenvorbereitung von Nöten. Die dadurch bedingte Dauer der Analysen und die mangelhafte Automatisierung macht jedoch eine Anwendung auf dem Gebiet der Hochdurchsatz-basierten Prozessentwicklung bzw. -analytik schwierig, wenn nicht unmöglich. Diese Schwierigkeit wird durch das erfindungsgemäße Verfahren überwunden.

20

Ein weiteres herkömmliches Verfahren zur selektiven Proteinanalytik, die analytische Chromatographie, läßt sich zwar in vielen Fällen ohne Probenvorbereitung durchführen, ist jedoch gerätebedingt ein sequentiell durchgeführtes Verfahren, wodurch sich sehr lange Analysenzeiten ergeben können, welche den zeitlichen
25 Gewinn bei Hochdurchsatz-Experimenten wieder zunichte machen.

Schließlich existieren Verfahren auf dem Gebiet der selektiven Proteinanalytik, deren Anwendung auf Proteine beschränkt ist, die eine spezifische Absorption bei bestimmten Wellenlängen aufweisen (z.B. Cytochrom C oder *Green Fluorescent Protein* (GFP)). Dabei wird herkömmlicherweise die Absorption bei diesen
30 bestimmten Wellenlängen für die selektive Quantifizierung verwendet. Eine generelle Anwendbarkeit, wie dies durch die vorliegende Erfindung ermöglicht wird, ist bei solchen herkömmlichen Verfahren allerdings nicht gegeben.

Das erfindungsgemäße Verfahren weist diese Nachteile nicht mehr auf. Insbesondere bedarf das erfindungsgemäße Verfahren keiner Probenvorbereitung, was eine schnelle und leicht zu automatisierende selektive Proteinanalytik ermöglicht.

5

Der Begriff „selektive Quantifizierung“ bezeichnet die Bestimmung der Einzelkonzentrationen eines, mehrerer oder aller Proteine in dem Proteingemisch nach ihrer jeweiligen Art. In einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens werden in Schritt (b) die Einzelkonzentrationen
10 eines, mehrerer oder aller Proteine in dem Proteingemisch bestimmt. Das erfindungsgemäße Verfahren ist damit in der Lage, nicht lediglich die Gesamtkonzentration aller Proteine in der Gesamtheit, sondern die Einzelkonzentrationen der jeweiligen Proteine zu bestimmen.

15 Verfahren zum Erstellen von UV-Absorptionsspektren sind dem Fachmann bekannt und umfassen die Verwendung eines UV-Detektors, beispielsweise eines Diodenbasierten UV-Detektors.

In einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens werden
20 kontinuierliche UV-Absorptionsspektren erstellt. Der Begriff „kontinuierliches UV-Spektrum“ bezeichnet UV-Spektren, welche die UV-Absorption über einen bestimmten Wellenlängenbereich in angemessener Auflösung, beispielsweise in 10 nm-, bevorzugt in 5 nm-, mehr bevorzugt in 1 nm-Schritten, erfassen, wobei sich die Auflösung über den gemessenen Wellenlängenbereich ändern kann. Eine weitere
25 bevorzugte Ausführungsform besteht in der Nutzung von diskontinuierlichen Absorptionsspektren, wobei die Auflösung nicht über den gesamten Wellenlängenbereich vorliegen muss. Beispielsweise können nur zehn UV-Spektren, bevorzugt nur fünf UV-Spektren, mehr bevorzugt nur drei UV-Spektren gemessen werden.

30

Erfindungsgemäß wurde erkannt, daß sich die UV-Absorptionsspektren verschiedener Proteine unterscheiden, insbesondere wurde erkannt, daß sich die Formen der Absorptionsspektren der individuellen Proteine unterscheiden (vgl. Fig. 2a). Diese Unterschiede sind größtenteils auf die unterschiedliche Aminosäure-

Zusammensetzung der beteiligten Proteine zurückzuführen und beinhalten daher Informationen, die mit der Identität des jeweiligen Proteins verknüpft sind. Ein Proteingemisch aus mehreren Proteinen weist so ein für dieses Proteingemisch charakteristisches UV-Absorptionsspektrum auf. Die Korrelation dieser Informationen mit den Einzelkonzentrationen der in dem Proteingemisch enthaltenen Proteine erfolgt in einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens rechnerisch. Dementsprechend erfolgt das Bestimmen der Einzelkonzentrationen in Schritt (b) des erfindungsgemäßen Verfahren bevorzugt rechnerisch. Dazu kann in einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens ein sogenanntes multivariates Kalibrierungsmodell verwendet werden.

In anderen Worten wurde erfindungsgemäß erkannt, die Unterschiede in den Absorptionsspektren verschiedener Proteingemische ausreichen, daß die Einzelkonzentration zumindest eines Proteins des Proteingemisches, insbesondere aller Proteine des Proteingemisches bestimmt, insbesondere berechnet werden können.

Dementsprechend wird in einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens Schritt (b), also das Bestimmen der Einzelkonzentration der Proteine in dem Proteingemisch aus dem in Schritt (a) erstellten UV-Absorptionsspektrum, mit Hilfe eines multivariaten Kalibrierungsmodells durchgeführt.

Bevorzugt wird das multivariate Kalibrierungsmodell gemäß der vorliegenden Erfindung auf Grundlage von UV-Absorptionsspektren von Proben erstellt, die bekannte Proteine in bekannten Konzentration enthalten. In einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens enthalten die Kalibrierungsproben jeweils nur ein Protein in bekannter Konzentration. In einer weiteren bevorzugten Ausführungsform werden aus den UV-Absorptionsspektren der jeweils nur ein Protein in bekannter Konzentration enthaltenden Kalibrierungsproben durch Linearkombinationen Mischspektren erzeugt, die für die Erstellung des multivariaten Kalibrierungsmodells verwendet werden. Verfahren für das Erzeugen von Mischspektren aus den UV-Absorptionsspektren von jeweils nur ein Protein in bekannter Konzentration enthaltenden Kalibrierungsproben durch

Linearkombinationen sind dem Fachmann bekannt. Vorteile einer solchen Vorgehensweise sind eine verbesserte Genauigkeit des Verfahrens, da Ungenauigkeiten beim Generieren der Proteinproben, z.B. durch fehlerhaftes Pipettieren, minimiert werden können. Außerdem wird nur eine kleine Menge der oft
5 seht teuren und schwer zugänglichen Proteine benötigt.

Der Begriff "multivariates Kalibrierungsmodell" bezeichnet im Allgemeinen Modelle mit deren Hilfe mehrere miteinander in Zusammenhang stehende statistische Variablen gleichzeitig untersucht werden können. Dabei können Zusammenhänge
10 und Abhängigkeitsstrukturen zwischen den Variablen erkannt werden. Ein multivariates Kalibrierungsmodell verknüpft dabei die Beziehung zwischen mehreren Eingangsvariablen mit einer oder mehreren Zieleigenschaft(en). In manchen Fällen ist der Vorteil einer multivariaten Kalibrierung ein robusteres Modell. In anderen Fällen kann die Zieleigenschaft überhaupt nur untersucht werden, wenn mehrere
15 Variablen gemessen werden. Im Fall der vorliegenden Erfindung sind die Eingangsvariablen die spektralen Daten der aufgenommenen UV-Absorptionsspektren, und die Zieleigenschaften sind die Konzentrationen der in dem Proteingemisch enthaltenen jeweiligen Proteine. Im Weiteren werden die Eingangsdaten als X-Daten und die Zieleigenschaften als Y-Daten bezeichnet.

20

Multilineare Regression (MLR) ist ein klassisches statistisches Verfahren für die multivariate Kalibrierung. Ein auf MLR basierendes Kalibrierungsmodell besteht aus einer linearen Gleichung, in der jede Variable der X-Daten mit einer Gewichtung multipliziert wird. Allerdings ist MLR bekannt dafür, empfindlich gegenüber
25 Kollinearität zu sein. Kollinearität zwischen den gemessenen Variablen kann immer auftreten, wenn die Anzahl der gemessenen Variablen die Anzahl der in der Kalibrierung enthaltenen Proben übersteigt. Weiter ist Kollinearität inhärent in spektralen Daten enthalten, da diese nicht aus unabhängigen Variablen bestehen. Die Hauptkomponentenanalyse (*principal component analysis*, PCA) ist ein
30 Verfahren, das verwendet wird, um die signifikante Varianz innerhalb von X-Daten zu finden, die aus nicht-unabhängigen Variablen bestehen. Durch sie werden die X-Daten in ein neues (Haupt-)Achsensystem transformiert. Dies geschieht auf solche Weise, daß die Transformation auf diese Achsen die größtmögliche Varianz in dem Datensatz widerspiegelt. Diese neuen Variablen werden als latente Variablen

bezeichnet. Auf diese Weise können unabhängige latente Variablen aus den ursprünglichen X-Daten (also den Absorptionsspektren) erzeugt werden, und verwendet werden, um ein Kalibrierungsmodell mittels MLR zu erzeugen, mit dessen Hilfe die Y-Daten (also die spezifischen Proteinkonzentrationen) vorhergesagt werden können. Allerdings ist, wenn das PCA-Verfahren verwendet wird, um latente Variablen zu finden, das einzige Kriterium die Menge der beschreibbaren Varianz. Da es keine Garantie dafür geben kann, daß diese Variablen mit den Zieleigenschaften korrelieren, kann das Verwenden von PCA für die multivariate Kalibrierung zu suboptimalen Kalibrierungen führen. Dieses Problem wird durch Verwenden der PLS-Regression gelöst, da diese die latenten Variablen der X-Daten basierend auf zwei Kriterien extrahiert, nämlich beschriebene Varianz und Korrelation mit den Y-Daten. Im Zusammenhang mit PLS werden die latenten Variablen als Komponenten bezeichnet.

Der Begriff „PLS-Regression“ bezeichnet das beschriebene statistische Regressionsverfahren, das oftmals als *partial least square* Regression, richtiger aber als *projection to latent structure* Regression bezeichnet wird. Das Akronym „PLS“ kann somit beide Bedeutungen haben, wobei das bezeichnete statistische Regressionsverfahren in jedem Fall dasselbe ist. Die PLS-Regression bietet im Vergleich zu anderen multivariaten Kalibrierungsansätzen, wie z.B. der multilinearen Regression, eine höhere Robustheit gegenüber Meßfehlern.

Multivariate Kalibrierungsmodelle, die in Verbindung mit dem erfindungsgemäßen Verfahren eingesetzt werden können, sind dem Fachmann bekannt. Erfindungsgemäß wurde erkannt, daß zur Bestimmung der Einzelkonzentration zumindest eines Proteins in einem Proteingemisch beispielsweise Kalibrierungsmodelle eingesetzt werden können, die mit Hilfe eines MCR (Multivariate Curve Resolution) genannten Verfahrens, multilinearer Regression oder PLS-Regression erstellt werden.

Dementsprechend wird in einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens das multivariate Kalibrierungsmodell mit Hilfe eines statistischen Verfahrens, ausgewählt aus der Gruppe der Methoden zur multivariaten Kalibrierung, insbesondere bestehend aus *Multivariate Curve Resolution* (MCR),

partial least square (PLS)-Regression und Multilinearer Regression (MLR), erstellt. In einer besonders bevorzugten Ausführungsform ist das statistische Verfahren die PLS-Regression.

- 5 Die Berechnung multivariater Kalibrierungsmodelle und ihre Anwendung sind oft sehr aufwendig und werden daher meist computergestützt durchgeführt. Geeignete Software-Pakete, die auch in Verbindung mit dem erfindungsgemäßen Verfahren eingesetzt werden können, sind dem Fachmann bekannt und umfassen beispielsweise das auf MATLAB basierende Paket „PLS Toolbox“ (EigenVector,
10 USA).

Gemäß einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens wird, basierend auf Spektren von Proben bekannter Zusammensetzung bezüglich Protein und Konzentration, ein Kalibrierungsmodell mit Hilfe der PLS-Regression
15 erstellt. Mit diesem Modell können in Proteingemischen unbekannter Zusammensetzung die Einzelkonzentrationen der vorhandenen Proteine anhand des UV-Absorptionsspektrums des jeweiligen Proteingemisches bestimmt werden.

Dabei stammt in einer bevorzugten Ausführungsform des erfindungsgemäßen
20 Verfahrens das Proteingemisch aus einer Säulenchromatographie. Das Erstellen des UV-Spektrums erfolgt dabei bevorzugt am Säulenausgang, wobei bevorzugt mindestens ein UV-Spektrum alle zehn Sekunden, mehr bevorzugt mindestens ein UV-Spektrum pro Sekunde, mehr bevorzugt mindestens vier UV-Spektren pro Sekunde erstellt werden. Es ist auch möglich, daß zwischen 1 und 20 UV-Spektren
25 pro Sekunde, beispielsweise zwischen 2 und 10 UV-Spektren pro Sekunde, insbesondere zwischen 4 und 7 UV-Spektren pro Sekunde am Säulenausgang erstellt werden.

In einer bevorzugten Ausführungsform ist die Säulenchromatographie eine
30 präparative Säulenchromatographie. Säulenchromatographische Verfahren, die in Verbindung mit dem erfindungsgemäßen Verfahren verwendet werden können, sind dem Fachmann bekannt.

In einer weiteren bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens stammt das Proteingemisch aus einem Hochdurchsatzverfahren, wobei bevorzugt mindestens ein UV-Spektrum alle zehn Sekunden, mehr bevorzugt mindestens ein UV-Spektrum pro Sekunde, mehr bevorzugt mindestens vier UV-Spektren pro Sekunde erstellt werden. Es ist auch möglich, daß zwischen 1 und 20 UV-Spektren pro Sekunde, beispielsweise zwischen 2 und 10 UV-Spektren pro Sekunde, insbesondere zwischen 4 und 7 UV-Spektren pro Sekunde erstellt werden. Hochdurchsatzverfahren, die in Verbindung mit dem erfindungsgemäßen Verfahren verwendet werden können, sind dem Fachmann bekannt.

10

Ein Nachteil herkömmlicher Verfahren zur Analyse chromatographischer Auftrennungen von Proteinen ist die fehlende Möglichkeit, die Proteine im Eluat selektiv, d.h. ihrer jeweiligen Art nach, zu bestimmen. Daher kann die genaue Zusammensetzung des Eluats, sowie die Einzelkonzentrationen bzw. die Mengen der einzelnen darin enthaltenen Proteine nicht zeitnah bestimmt werden. Vielmehr müssen erst die gesammelten Fraktionen des Eluats analysiert werden, wodurch nachstehende Prozessschritte verzögert werden. Ebenso ist eine optimale Fraktionierung des Zielproteins, unabhängig davon ob Reinheit oder Ausbeute des Proteins die relevante Zielgröße ist, nicht möglich.

20

Dagegen ist es mit dem erfindungsgemäßen Verfahren zum selektiven Nachweis von Proteinen in einem Proteingemisch möglich, zeitgleich zur Elution die Reinheit des Zielproteins im Eluat zu bestimmen. Somit ist eine deutlich verbesserte Grundlage für die Fraktionierung des Eluats im Hinblick auf eine kontrollierbare Reinheit oder Ausbeute geschaffen.

25

In einer weiteren bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens umfaßt das zu untersuchende Proteingemisch N verschiedene Proteine, wobei N eine natürliche Zahl und bevorzugt $N \geq 2$ oder $N \geq 3$ oder $N \geq 5$ oder $N \geq 7$ oder $N \geq 10$ und/oder $N \leq 100$ und/oder $N \leq 50$ und/oder $N \leq 30$ und/oder $N \leq 20$ ist. Beispielsweise kann für N gelten: $N \in [2, 10]$ oder $N \in [3, 10]$ oder $N \in [4, 7]$. Insbesondere kann für N gelten: $N = 2$ oder $N = 3$ oder $N = 4$ oder $N = 5$ oder $N = 6$ oder $N = 7$ oder $N = 8$ oder $N = 9$ oder $N = 10$, etc.

30

Dabei enthält in einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens das zu untersuchende Proteingemisch kein dem multivariaten Kalibrierungsmodell unbekanntes Protein, bzw. kein Protein, das bei der Erstellung des multivariaten Kalibrierungsmodells nicht berücksichtigt wurde. In einer weiteren
5 bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens enthält das zu untersuchende Proteingemisch dem multivariaten Kalibrierungsmodell unbekannte Proteine zu einem Anteil von höchstens 30% (w/v), mehr bevorzugt höchstens 10% (w/v), und am meisten bevorzugt höchstens 1% (w/v). In einer weiteren bevorzugten Ausführungsform bedarf das erfindungsgemäße Verfahren keinerlei
10 Probenvorbereitung.

In einer weiteren bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens wird das UV-Spektrum in einem mittleren ultravioletten Bereich, mehr bevorzugt in einem Bereich von etwa 210 nm bis etwa 300 nm Wellenlänge, am meisten
15 bevorzugt in einem Bereich von etwa 240 nm bis etwa 300 nm Wellenlänge erstellt. In einer weiteren bevorzugten Ausführungsform wird das UV-Spektrum in einem Wellenlängenbereich erstellt, welcher den Bereich zwischen 240 nm und 300 nm beinhaltet oder Teile davon umfasst, beispielsweise im Bereich zwischen 220 nm und 300 nm, 230 nm und 300 nm, 250 nm und 300 nm, 260 nm und 300 nm, 220 nm
20 und 280 nm, 240 nm und 280 nm oder 250 nm und 290 nm. Dabei ist das Spektrum bevorzugt ein kontinuierliches oder diskontinuierliches UV-Absorptionsspektrum. Das UV-Absorptionsspektrum wird in einer bevorzugten Ausführungsform mit Hilfe eines Dioden-basierten Detektors erstellt. Solche Geräte sind in der Lage, mehrere Spektren pro Sekunde aufzunehmen. Geeignete Detektoren sind dem Fachmann
25 bekannt.

In einer bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens wird das Verfahren automatisiert durchgeführt.

30 Ein weiterer Gegenstand der vorliegenden Erfindung betrifft ein System zur selektiven Quantifizierung einzelner Proteine in einem Proteingemisch, umfassend:
(a) Mittel für eine säulenchromatographische Trennung eines Proteingemisches,
(b) einen Detektor, insbesondere einen Dioden-basierten UV-Detektor, der am Säulenausgang angeordnet ist, und

- (c) Mittel für die automatische Verarbeitung der von dem UV-Detektor erstellten UV-Absorptionsspektren und die Bestimmung der Konzentration der einzelnen Proteine in dem Proteingemisch aus diesen UV-Absorptionsspektren mit Hilfe eines multivariaten Kalibrierungsmodells und eines geeigneten statistischen Verfahrens.

Ein weiterer Gegenstand der vorliegenden Erfindung betrifft ein System zur selektiven Quantifizierung einzelner Proteine in Proteingemischen, umfassend:

- (a) Mittel für die automatisierte Handhabung von Multi-Well-Platten, welche die Proteingemische enthalten,
- (b) einen Detektor, insbesondere einen Dioden-basierten UV-Detektor, der oberhalb oder unterhalb der Multi-Well-Platten angeordnet ist, und
- (c) Mittel für die automatische Verarbeitung der von dem Detektor erstellten UV-Absorptionsspektren und die Bestimmung der Konzentration der einzelnen Proteine in den Proteingemischen aus diesen UV-Absorptionsspektren mit Hilfe eines multivariaten Kalibrierungsmodells und eines geeigneten statistischen Verfahrens.

Mittel für eine säulenchromatographische Trennung eines Proteingemisches sind dem Fachmann bekannt und umfassen beispielsweise geeignete Säulen, Säulenmaterialien, Pumpen, Sammler, mobile Phasen, Fluidik, Steuer- und Kontrolleinrichtungen. Mittel für die automatisierte Handhabung von Multi-Well-Platten sind dem Fachmann ebenfalls bekannt und umfassen beispielsweise übliche Plattenlesegeräte. Die Systeme können insbesondere auch ein oder mehrere Quellen elektromagnetischer Strahlung, insbesondere eine UV-Quelle beinhalten oder damit funktional verbunden sein. Geeignete Dioden-basierte UV-Detektoren sind dem Fachmann ebenfalls bekannt. Zusätzlich oder alternativ ist es auch möglich, das erfindungsgemäße Verfahren bei anderen Wellenlängen als der Wellenlänge des UV zu verwenden. Entsprechend kann das System andere bzw. zusätzliche Quellen elektromagnetischer Strahlung und entsprechende Detektor(en) aufweisen.

Mittel für die automatische Verarbeitung der von dem UV-Detektor erstellten UV-Absorptionsspektren und die Bestimmung der Konzentration der einzelnen Proteine

in dem Proteingemisch aus diesen UV-Absorptionsspektren mit Hilfe eines multivariaten Kalibrierungsmodells sind dem Fachmann ebenfalls bekannt, und umfassen beispielsweise geeignete Computer in Verbindung mit geeigneter Software. Geeignete Software-Pakete, die in diesem Zusammenhang eingesetzt werden können, sind dem Fachmann bekannt und umfassen beispielsweise das auf
5 MATLAB basierende Paket „PLS Toolbox“ (EigenVector, USA).

Anhand der vorliegenden Erfindung konnte überraschend gezeigt werden, daß die Unterschiede zwischen den UV-Absorptionsspektren verschiedener Proteine groß
10 genug sind, um mit Hilfe multivariater Kalibrierungsmethoden aus UV-Absorptionsspektren komplexer Proteingemische die Einzelkonzentrationen der enthaltenen Proteine zu bestimmen. Bislang wurde in der Fachwelt angenommen, daß dies nicht möglich ist.

15 Das erfindungsgemäße Verfahren ermöglicht insbesondere eine schnelle Analytik von Proteingemischen, z.B. des Eluats einer säulenchromatographischen Proteinaufreinigung, wodurch eine oft zeit- und kostenaufwendige nachgeschaltete Proteinanalytik überflüssig wird. Insbesondere kann hierbei durch die mit der Elution gleichzeitig stattfindende Bestimmung der Einzelkonzentrationen der im Eluat
20 enthaltenen Proteine eine optimale Fraktionierung des Eluats unmittelbar erfolgen, da eine nachgeschaltete Proteinanalytik überflüssig wird. Das erfindungsgemäße Verfahren bietet weiter die Möglichkeit, Proteine in wäßriger Lösung schnell und automatisiert durch das Erstellen eines UV-Absorptionsspektrums und die Verknüpfung der so gewonnenen spektralen Informationen mit selektiven
25 Proteinkonzentrationen über eine multivariate Kalibrierung zu quantifizieren, ohne daß eine Vorbereitung der Proben nötig wäre.

Das erfindungsgemäße Verfahren ist in vielfältiger Weise auf den Gebieten der Protein- und Prozessanalytik einsetzbar. Insbesondere erlaubt es eine selektive
30 Proteinanalytik bei allen Fragestellungen der Proteinquantifizierung in einem Proteingemisch und/oder der relativen oder absoluten Änderung der Proteinzusammensetzung eines Proteingemisches. Mögliche Einsatzgebiete sind das Hochdurchsatz-Screening in der Biologie, Biochemie, Biotechnologie, Pharmazie oder Medizin, die Mikrofluidik, die Sensorik, die Biochromatographie, die

Bioprozessentwicklung oder die Bioprozessanalytik, wobei das erfindungsgemäße Verfahren beispielsweise sowohl im Batchbetrieb in Küvetten oder Multi-Well-Platten, als auch im Durchflussbetrieb in mikrofluidischen Applikationen, Leitungen oder Meßzellen durchgeführt werden kann. Dies ist bei den im Stand der Technik
5 bekannten Verfahren zur selektiven Proteinanalytik nicht möglich.

Weiter ist die Einfachheit und Schnelligkeit des erfindungsgemäßen Verfahrens auch von großem Nutzen in der Hochdurchsatz-Analytik. Insbesondere ist, entgegen den im Stand der Technik bekannten Verfahren zur selektiven Proteinanalytik, eine
10 Automatisierung ohne großen Aufwand möglich und die Analysezeit und/oder die Vorbereitungszeit wesentlich geringer. Die benötigte Ausstattung ist darüber hinaus preiswert, wenig störanfällig und leicht zu bedienen.

Ein weiterer Gegenstand der vorliegenden Erfindung betrifft die Verwendung von
15 UV-Absorptionsspektren zur Bestimmung der Konzentration einzelner Proteine in dem Proteingemisch mit Hilfe eines multivariaten Kalibrierungsmodells und eines geeigneten statistischen Verfahrens.

Die Figuren zeigen:

20

Fig. 1: UV-Absorptionsspektren von drei proteinogenen Aminosäuren mit aromatischer Struktur, nämlich Tyrosin, Tryptophan und Phenylalanin im mittleren UV-Bereich. Die Konzentration der Aminosäuren war dabei 0,1 mg/ml. Für Phenylalanin ist aufgrund der relativ geringen UV-Absorption
25 auch eine Konzentration von 1,0 mg/ml gezeigt.

Fig. 2: UV-Absorptionsspektren von drei Modellproteinen, nämlich Lysozym (lys), Cytochrom C (cytC) und Ribonuclease A (ribA).

- 30 (a) Rohspektren der reinen Modellproteine bei 0,5 mg/ml.
(b) Rohspektren von fünf Proben, die reines Protein und drei verschiedene Mischungsverhältnisse von cytC und ribA umfassen.
(c) Die fünf Spektren, die basierend auf der Absorption der reinen cytC-Probe bei 280 nm in ihren Offsets verschoben wurden.
(d) Differenzspektren, die aus den fünf Offset-verschobenen Spektren erstellt

wurden.

Fig. 3: (a) 19 Mischungsverhältnisse der drei Modellproteine, die sich aus dem vierschichtigen Design, auf dem das Kalibrierungsset basierte, ergaben.

5 (b) Die Konzentrationen der Proteine reichten von 0 bis 0,7 mg/ml. Die Konzentrationen von ribA und lys sind gezeigt. Weitere drei Proben mit einer Konzentration von 0,05 mg/ml nur jeweils eines Proteins wurden hinzugefügt (Proben 28 bis 30).

10 Fig. 4: Von dem PLS-Kalibrierungsmodell vorhergesagte Proteinkonzentrationen, die als Funktion der nominellen Konzentration für alle drei Protein aufgetragen wurden (obere Reihe).

(a) Ribonuclease A,

(b) Cytochrom C,

15 (c) Lysozym. Eine die ideale Korrelation symbolisierende Gerade ist gezeigt. Ebenso sind die entsprechenden relativen und absoluten Fehler gezeigt (untere Reihe).

Fig. 5: Chromatogramme der Trennungen des ternären Proteingemisches über vier
20 verschiedene Kationentauscherharze. Für jedes Harz ist ein Histogramm dargestellt (oben), das die gesamte Absorption jeder der 24 Elutionsfraktionen bei 280 und 526 nm zeigt, und ein Scatter-Plot dargestellt (unten), der die mit Hilfe des erfindungsgemäßen Verfahrens berechneten Konzentrationen der drei Proteine als Funktion des Elutionsvolumens zeigt.

25

Fig. 6: Chromatogramm einer säulenchromatographischen Auftrennung der drei Modellproteine mittels herkömmlicher Ein-Kanal UV-Absorptionsmessung bei 280 nm am Säulenausgang (oben) und mittels Berechnung gemäß dem erfindungsgemäßen Verfahren (unten). Spektren im Bereich von 240 bis 300
30 nm wurden am Säulenausgang vom Eluat aufgenommen und anhand dessen die Konzentration der einzelnen im Eluat vorhandenen Proteine berechnet.

Fig. 7: Proteinzusammensetzung der Lysozym/BSA und BSA/HSA Kalibrierungssets.

5 Fig. 8: Vorhersagefähigkeit des auf PLS-Regression basierenden Modells für das Lysozym/BSA-System als Funktion der Anzahl der in dem Modell enthaltenen latenten Variablen.

10 Fig. 9 Vorhergesagte Proteinkonzentrationen als Funktion der nominellen Proteinkonzentrationen für Lysozym und BSA. Die Daten wurden vor der PLS-Regression nicht aufbereitet. Das Modell basierte auf drei latenten Variablen. Eine die ideale Korrelation symbolisierende Gerade ist gezeigt.

15 Fig. 10 Vorhersagefähigkeit des auf PLS-Regression basierenden Modells für das Lysozym/BSA-System als Funktion der Anzahl der in dem Modell enthaltenen latenten Variablen. Die Daten wurden vor Erstellen des Modells durch Zentrierung auf den Mittelwert aufbereitet.

20 Fig. 11 Vorhergesagte Proteinkonzentrationen als Funktion der nominellen Proteinkonzentrationen für Lysozym und BSA. Die Daten wurden vor der PLS-Regression durch Zentrierung auf den Mittelwert aufbereitet. Das Modell basierte auf zwei latenten Variablen. Eine die ideale Korrelation symbolisierende Gerade ist gezeigt.

25 Fig. 12 Vorhersagefähigkeit des auf PLS-Regression basierenden Modells für das HSA/BSA-System als Funktion der Anzahl der in dem Modell enthaltenen latenten Variablen. Die Daten wurden vor Erstellen des Modells durch Zentrierung auf den Mittelwert aufbereitet.

30 Fig. 13 Vorhergesagte Proteinkonzentrationen als Funktion der nominellen Proteinkonzentrationen für HSA und BSA. Die Daten wurden vor der PLS-Regression durch Zentrierung auf den Mittelwert aufbereitet. Das Modell basierte auf zwei latenten Variablen. Eine die ideale Korrelation symbolisierende Gerade ist gezeigt.

Fig. 14 UV-Absorptionsspektren im mittleren UV-Bereich von deamidiertem und nicht-deamidiertem Aspart. Die Spektren wurden mit einer Auflösung von 1 nm gemessen.

- 5 Fig. 15 Ein zwei-faktorielles Doehlert-Design wurde verwendet, um die Zusammensetzung der Kalibrierungsproben zu erstellen. Sieben daraus resultierende verschiedene Zusammensetzungen sind gezeigt. Der Mittelwert wurde dreifach erstellt.
- 10 Fig. 16 Vorhergesagte Proteinkonzentration als Funktion der nominellen Proteinkonzentration für deamidiertes und nicht-deamidiertes Aspart. Die hellgrauen größeren Punkte repräsentieren die Vorhersagen aus den Kreuzvalidierungen, die kleineren schwarzen Punkte die aus den Testproben. Eine die ideale Korrelation symbolisierende Gerade ist gezeigt.

15 Die vorliegende Erfindung wird anhand der folgenden nicht-einschränkenden Beispiele näher erläutert.

Material und Methoden

20 *Verbrauchsmaterialien*

Probenvorbereitung: 96-well Platten mit flachem Boden, volumetrische Kapazität 2 ml, Polypropylen (VWR, Deutschland). Messung von UV-Spektren im mittleren Bereich: 96-well UV-Mikrotiterplatten mit flachem Boden, volumetrische Kapazität 360 µl, Polystyrol (Greiner, Deutschland). Chromatographie: Mit 200 µl Harz für die Kationenaustauschchromatographie befüllte RoboColumns[®]. Die folgenden Harze wurden verwendet: CM Ceramic HyperD[®] F, Toyopearl[®] SP 650 M, SP Sepharose[®] FF und Fractogel[®] EMDSO₃ (Atoll, Deutschland).

30 *Modellproteine*

Modellproteine waren Lysozym (lys) aus Hühnereiweiß, Cytochrom C (cytC) aus Pferdeherzen, Ribonuclease A (ribA) aus bovinem Pankreas, bovines Serumalbumin (BSA) und humanes Serumalbumin (HSA) (alle Sigma-Aldrich, Deutschland).

Pipettierstation

Die Herstellung der Kalibrierungs- und Testproben für das Kalibrierungsmodell, sowie die chromatographischen Experimente wurden automatisiert auf einer Tecan Freedom Evo[®] 200 Pipettierstation (Tecan, Crailsheim, Deutschland) durchgeführt.

- 5 Um das Durchführen von Säulenchromatographie auf der Pipettierstation zu ermöglichen, wurde ein Te-Chrom[®] Säulenträger (Tecan, Crailsheim, Deutschland) installiert. Weiter wurde die Pipettierstation mit einem Infinite 200 Plattenleser (Tecan, Crailsheim, Deutschland) ausgerüstet, um UV-Spektren im mittleren UV-Bereich der hergestellten Kalibrierungs- und Testproben, sowie der
10 chromatographischen Fraktionen zu messen.

Software

Für das Design der Probenzusammensetzung für das Kalibrierungsmodell und für die Testproben, die zur Evaluierung des Kalibrierungsmodells verwendet wurden,
15 wurde die Software MODDE[®] (Umetrics, Umea, Schweden) verwendet. Die Pipettierstation wurde mit EVOware 2.0 (Tecan, Crailsheim, Deutschland) betrieben. Die zu importierenden Werte für die zu pipettierenden Volumina und die zu exportierenden Spektraldaten wurden mit Microsoft Excel verwaltet (Microsoft, Redmond, USA). Um die Kalibrierungsmodelle der verschiedenen Proteingemische
20 zu erstellen, und um die unbekannt Proteinkonzentrationen in den Chromatographieproben vorherzusagen, wurde das auf MATLAB basierende Paket PLS Toolbox (EigenVector, USA) verwendet.

Probenvorbereitung und Aufnahmen der Spektren

- 25 Die Kalibrierungs- und Testproben wurden aus Stammlösungen, die 2,1 mg/ml Protein in einem Natriumphosphatpuffer bei pH 7 enthielten, hergestellt. Die Proben wurden in 96-well Platten in einem Volumen von 2 ml hergestellt. Für die Messung der UV-Absorptionsspektren wurden 150 µl der Proben in UV-transparente 96-well UV-Mikrotiterplatten mit flachem Boden überführt. Die Absorptionsspektren wurden
30 mit einer Auflösung von 1 nm im Bereich von 240 nm bis 295 nm mit Hilfe eines Lambda 35 UV-Detektors (Perkin Elmer, Deutschland) oder dem Infinite 200 Plattenleser (Tecan, Crailsheim, Deutschland) gemessen.

Erstellen des multivariaten Kalibrierungsmodells

Für das ternäre Proteingemisch lys/cytC/ribA basierten die Kalibrierungsproben auf einem D-optimalen Zwiebel-Design mit vier Ebenen. Dies führte zu 19 Mischungsverhältnissen (vgl. Fig. 3a) der drei Komponenten und insgesamt 28 Proben einschließlich dreier Mittelpunkte (alle drei Komponenten bei halbmaximaler Konzentration in einer Probe). Weiter wurden drei Proben zu dem Kalibrierungsset hinzugefügt, die jeweils 0,05 mg/ml eines der drei Proteine enthielten. Insgesamt basierte das Kalibrierungsmodell somit auf 31 Proben. Fig. 3b zeigt die Zusammensetzung der Kalibrierungsproben für lys und ribA. Die Testproben basierten auf einem D-optimalen Zwiebel-Design mit drei Ebenen. Sowohl die Kalibrierungsproben, als auch die Testproben wurden aus den selben Protein-Stammlösungen hergestellt. In der Software PLS Toolbox wird per Voreinstellung eine automatische Skalierung für das Aufbereiten der Daten verwendet. Eine solche Skalierung ist nur dann sinnvoll, wenn Datensätze aus Variablen verschiedener Skalierung und Verteilung bestehen, die notwendigerweise transformiert werden müssen. Wird die automatische Skalierung auf spektrale Daten angewendet, können nützliche Informationen verloren gehen. Daher sollte die automatische Skalierung nicht verwendet werden. Statt dessen wurden die Daten durch Zentrierung am Mittelwert aufbereitet. Nach Aufbereitung der Daten wurde ein Kalibrierungsmodell für alle drei Proteine erstellt. Das Modell wurde mit drei latenten Variablen erstellt und mit fünf zufälligen Subsets kreuzvalidiert. Um die Vorhersagefähigkeit des Modells zu gewährleisten, wurde es an einem unabhängigen Probensatz getestet. Die Kalibrierungsmodelle für die binären Proteinsysteme wurden in analoger Weise wie in den entsprechenden Beispielen beschrieben erstellt.

25

Chromatographie

Ein 5 mg/ml jedes Modellproteins enthaltendes Proteingemisch wurde in einem 20 mM Natriumphosphat-Puffer (pH 7) hergestellt. Alle Säulen wurden mit 5 Säulenvolumina (*column volume*, CV) (1 CV = 200 µl) des Ladepuffers äquilibriert, bevor sie mit 50 µl der Proteinlösung, entsprechend 0,25 mg jedes Proteins, beladen wurden. Die Proteine wurden von der Säule mit einem Natriumchlorid-Gradienten von 0 bis 500 mM über 24 CV eluiert. Gerätebedingt wurde dabei ein schrittweiser Gradient mit einem Anstieg der Salzkonzentration von 20,83 mM pro CV verwendet. Jeder Schritt des Gradienten hatte ein Volumen von 200 µl und wurde als eine

30

Fraktion gesammelt und analysiert. Nach der vollständigen Elution wurden 150 µl jeder eluierten Fraktion in frische 96-well Platten überführt und die UV-Absorptionsspektren gemessen.

5 *Deamidierung von Aspart*

Das humane Insulin-Analogon Aspart (Novo Nordisk, Dänemark) wurde in Form von Aspart API bereitgestellt. Für die Deamidierung wurde Aspart für 16 Tage bei 45°C in einem Natriumphosphatpuffer (pH 7,4) inkubiert. Die Lösung wurde alle 24 Stunden durch einen Spritzenfilter (0,2 µm) filtriert, um bakterieller Kontamination
10 vorzubeugen.

Beispiel 1:

UV-Absorptionsspektren der drei Modellproteine lys, cytC und ribA

15 Aufgrund der verschiedenen Verhältnisse der aromatischen Aminosäurereste, die in einem bestimmten Protein vorhanden sind, unterscheiden sich die Intensität und die Form der entsprechenden UV-Absorptionsspektren verschiedener Proteine leicht voneinander. Fig. 2a zeigt die UV-Absorptionsspektren der drei Modellproteine lys,
20 cytC und ribA. Diese Proteine stellen ein günstiges Modellsystem dar, da sie aufgrund ihrer unterschiedlichen isoelektrischen Punkte leicht chromatographisch voneinander getrennt werden können.

Das UV-Absorptionsspektrum von lys unterscheidet sich klar von denen der anderen
25 beiden Proteine. Dagegen erscheinen die UV-Absorptionsspektren von cytC und ribA sehr ähnlich in ihrer Form und scheinen sich nur in ihrer Intensität zu unterscheiden. Um zu zeigen, daß sich die Spektren der beiden Proteine tatsächlich auch in ihrer Form unterscheiden, wurden fünf Proben mit unterschiedlichen Mischungsverhältnissen der beiden Proteine hergestellt und die entsprechenden UV-
30 Absorptionsspektren verglichen. Fig. 2 zeigt die rohen Spektren (Fig. 2b), Spektren die basierend auf der Absorption bei 280 nm übereinandergelegt wurden (Fig. 2c) und Differenz-Spektren, die auf den übereinandergelegten Spektren basieren (Fig. 2d). Die Differenz-Spektren zeigen klar, daß sich die Spektren von Cytochrom C und Ribonuclease A auch in ihrer Form unterscheiden.

Beispiel 2:Multivariates Kalibrierungsmodell für das lys, cytC, ribA System

Die Unterschiede in den UV-Absorptionsspektren der drei Modellproteine wurden
 5 verwendet, um multivariate Kalibrierungsmodelle mit der Fähigkeit, spezifische
 Proteinkonzentrationen vorherzusagen, zu erstellen. Um PLS-Regression im
 Vergleich zu MLR zu testen, wurden Kalibrierungsmodelle mit beiden Ansätzen
 erstellt. Weiter wurde der Einfluß der Datenaufbereitung, hier Zentrierung auf den
 10 Mittelwert, untersucht. In allen vier Modellen wurden die Konzentrationen der drei
 Proteine ko-kalibriert. Ein unabhängiger Probensatz wurde verwendet, um die
 tatsächliche Vorhersagefähigkeit der Modelle zu testen. Die PLS-Modelle basierten
 auf drei Komponenten. Tabelle 1 zeigt die Bestimmungskoeffizienten (R^2 -Werte) für
 jedes, jeweils mit den vier verschiedenen Ansätzen kalibrierte Protein. Die
 Ergebnisse der Kreuzvalidierung (*cross validation*, *cv*), wie auch die Vorhersagen der
 15 Testsets (*test*) sind gezeigt.

Tabelle 1: Bestimmungskoeffizienten aller Kreuzvalidierungen und der
 vorhergesagten Testdaten

Verfahren	Datenaufbereitung	cv ribA	test ribA	cv cytC	test cytC	cv lys	test lys
PLS	Zentr. Mittelwert	0,9911	0,9957	0,9992	0,9989	0,9992	0,9993
MLR	Zentr. Mittelwert	1,0	0,9706	1,0	0,9958	1,0	0,9916
MLR	keine	1,0	0,9149	1,0	0,7998	1,0	0,9702
PLS	keine	0,8506	0,870	0,552	0,5288	0,8119	0,8064

20 Die besten Ergebnisse wurden mit PLS-Regression, basierend auf
 mittelwertzentrierten Daten, erhalten. Dieses Modell lieferte R^2 -Werte über 0,99 für
 alle Proteine. Eine Zentrierung auf den Mittelwert war dabei entscheidend für das
 PLS-Modell, nachdem die Vorhersagefähigkeit von Kalibrierungsmodellen, die auf
 25 rohen Spektraldaten basierten, wesentlich niedriger war. Charakteristisch für die
 MLR-Modelle sind R^2 -Werte von 1,0 bei der Kreuzvalidierung. Aufgrund der hohen
 Anzahl an Variablen im Vergleich zur PLS-Regression, auf denen diese Modelle
 basieren (MLR: 60 Variablen; PLS: 3 Variablen), können die MLR-Modelle angepaßt
 werden, um die exakten Konzentrationen in den Kalibrierungsdatensätzen
 30 unabhängig von der Datenaufbereitung vorherzusagen. Diese Fähigkeit zur

fehlerfreien Vorhersage trifft allerdings nur auf die Kalibrierungsdatensätze zu.

Für alle Testproben wurden absolute (mg/ml) und relative (%) Abweichungen von den Vorhersagen, die das PLS-Modell basierend auf mittelwertszentrierten Daten lieferte, berechnet. Figur 4 zeigt diese Abweichungen als Funktion der nominellen Konzentrationen für jede Probe und jedes Protein. Über den kalibrierten Bereich von 0,0 bis 0,7 mg/ml war der absolute Vorhersagefehler für lys und cytC in Bereich von $\pm 0,025$ mg/ml und für ribA im Bereich von $\pm 0,05$ mg/ml. Die größeren Abweichungen bei ribA waren dabei vermutlich auf die geringe Intensität der UV-Absorptionsspektren von ribA zurückzuführen.

Beispiel 3:

Anwendung des erfindungsgemäßen Verfahrens bei der Säulenchromatographie

Die Anwendbarkeit des erfindungsgemäßen Verfahrens wurde anhand säulenchromatographischer Trennungen der drei Modellproteine untersucht. Die Trennungen wurden automatisiert auf einer Pipettierstation durchgeführt. Dabei wurden vier verschiedene Kationentauscher-Materialien verwendet, und die Elution mit dem gleichen Salzgradienten für jede Säule durchgeführt. Die resultierenden Chromatogramme sind in Fig. 5 gezeigt, wobei für jedes Harz zwei Plots gezeigt sind, zum einen die gesamte UV-Absorption der gesammelten Fraktionen bei 280 nm und 526 nm gegen das Elutionsvolumen und zum anderen die Konzentrationen der einzelnen Proteine gegen das Elutionsvolumen, wie von dem Kalibrierungsmodell vorhergesagt. Die vier Trennungen repräsentieren vier häufige Szenarien in der Trennung von Proteinen, nämlich Co-Elution aller drei Proteine (Fig. 5a), Trennung aller drei Proteine (Fig. 5b), Co-Elution der ersten beiden Proteine (Fig. 5c) und Co-Elution der letzten beiden Proteine (Fig. 5d). Eine Gegenüberstellung eines Chromatogramms, das durch herkömmliche UV-Absorptionsmessung bei 280 nm erstellt wurde, und eines Chromatogramms, das die mit dem erfindungsgemäßen Verfahren berechneten Konzentrationen der einzelnen Proteine zeigt, ist in Fig. 6 dargestellt.

Beispiel 4:Kalibrierungsmodell für BSA/lys

UV-Absorptionsspektren für das binäre Proteinsystem BSA (bovines
5 Serumalbumin)/lys wurden aufgenommen. Die Proben wurden jeweils entsprechend
den in Fig. 7 dargestellten Proteinkonzentrationen erzeugt. Kalibrierungsmodelle
wurden aus den aufgenommenen Spektren mit der Software PLS Toolbox erstellt.

Zunächst wurden Kalibrierungsmodelle ohne vorherige Datenaufbereitung erstellt.
10 Fig. 8 zeigt den quadratischen Mittelwert des Fehlers der Kalibrierung und der
Kreuzvalidierung (*root mean square error of calibration/cross validation*;
RMSEC/RMSECV) als Funktion der in dem Modell enthaltenen latenten Variablen.
Entsprechend diesen Daten wurde ein auf drei latenten Variablen basierendes
15 Modell gewählt. Fig. 9 zeigt die vorhergesagten Proteinkonzentrationen als Funktion
der nominellen Proteinkonzentrationen jeweils für Lysozym und BSA. Die
Bestimmungskoeffizienten (R^2 -Werte) der Kalibrierungsmodelle waren 0,997 für
Lysozym und 0,990 für BSA.

Weitere Kalibrierungsmodelle wurden mit den selben Daten des BSA/lys-Systems
20 erstellt, allerdings wurden die Daten vorab durch Zentrierung auf den Mittelwert
aufbereitet. Basierend auf den in Fig. 10 gezeigten RMSEC- und RMSECV-Werten
in Abhängigkeit der Zahl der in dem Modell enthaltenen latenten Variablen wurden
Modelle mit zwei latenten Variablen ausgewählt. Fig. 11 zeigt die vorhergesagten
Proteinkonzentrationen als Funktion der nominellen Proteinkonzentrationen jeweils
25 für Lysozym und BSA. Die Bestimmungskoeffizienten (R^2 -Werte) der
Kalibrierungsmodelle waren 0,999 für Lysozym und 0,997 für BSA. Obwohl die
Modelle mit mittelwertszentrierten Daten auf weniger latenten Variablen basierten,
lieferten diese somit eine höhere Korrelation.

30

Beispiel 5:Kalibrierungsmodell für BSA/HSA

UV-Absorptionsspektren für das binäre Proteinsystem BSA/HSA (humanes

Serumalbumin) wurden aufgenommen, um die Kalibrierung von strukturell ähnlicheren Proteinen zu testen. Die Proben wurden jeweils entsprechend den in Fig. 7 dargestellten Proteinkonzentrationen erzeugt. Kalibrierungsmodelle wurden aus den aufgenommenen Spektren mit der Software PLS Toolbox erstellt.

5

Für das Erstellen der Kalibrierungsmodelle wurden die Daten vorab mittelwertszentriert. Basierend auf den in Fig. 12 gezeigten RMSEC- und RMSECV-Werten in Abhängigkeit der Zahl der in dem Modell enthaltenen latenten Variablen wurden Modelle mit drei latenten Variablen ausgewählt. Fig. 13 zeigt die vorhergesagten Proteinkonzentrationen als Funktion der nominellen Proteinkonzentrationen jeweils für BSA und HSA. Trotz der großen Ähnlichkeit der beiden Proteine hatte das Modell eine hohe Vorhersagefähigkeit. Die Bestimmungskoeffizienten (R^2 -Werte) des Kalibrierungsmodells waren 0,988 für HSA und 0,986 für BSA.

15

Beispiel 6:

Kalibrierungsmodell für deamidiertes/nicht-deamidiertes Aspart

Um die Grenzen des Auflösungsvermögens des erfindungsgemäßen Verfahrens auszuloten, wurde die Kalibrierung von zwei extrem ähnlichen Proteinen, nämlich dem humanen Insulin-Analogen Aspart in deamidierter und nicht-deamidierter Form getestet. Mögliche Unterschiede in den UV-Absorptionsspektren dieser Proteine sollten, wenn überhaupt detektierbar, äußerst klein sein.

25

Fig. 14 zeigt die UV-Absorptionsspektren beider Proteine. Wie erwartet sind sich die Spektren in Form und Intensität sehr ähnlich, wobei geringfügige Unterschiede festgestellt werden konnten. Um zu prüfen, ob diese Unterschiede ausreichen, um ein stabiles Kalibrierungsmodell zu erstellen, wurden Proben für eine Kalibrierung zubereitet. Fig. 15 zeigt die Proteinkonzentrationen der neun zubereiteten Proben. Ihre Zusammensetzung basiert auf einem zwei-faktoriellen Doehlert-Design mit einem dreifachen Mittelwert.

30

Das auf PLS basierende Kalibrierungsmodell wurde mit zwei latenten Variablen

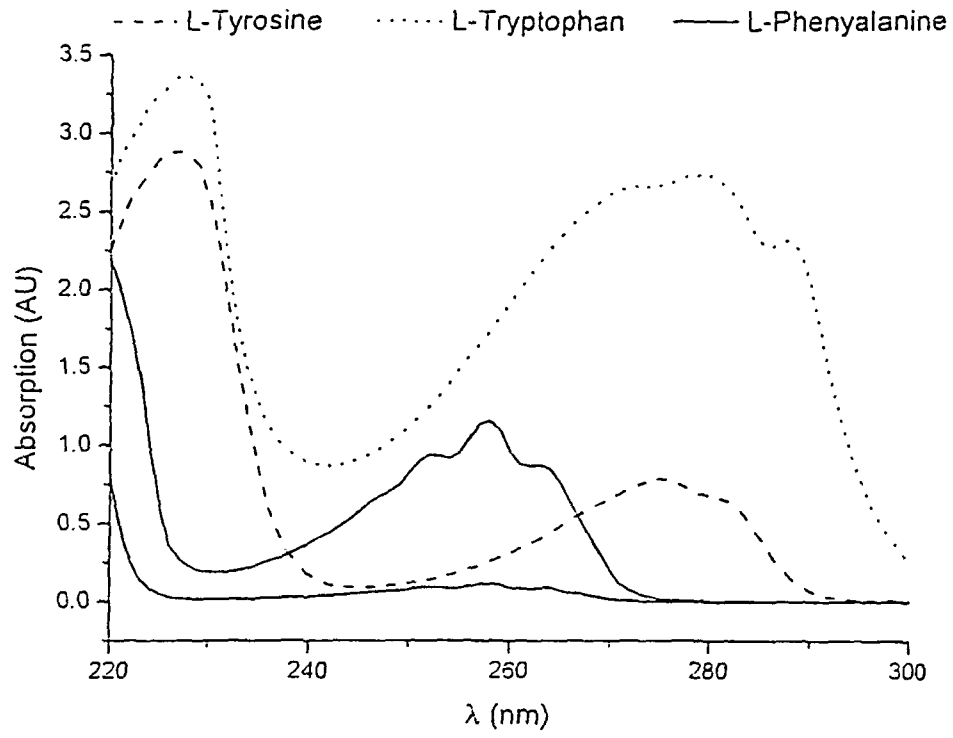
erstellt. Die PLS-Regression wurde mit Kreuzvalidierungen durchgeführt, wobei jede Probe einmal weggelassen und von dem Kalibrierungsmodell, basierend auf den restlichen Proben, vorhergesagt wurde. Fig. 16 zeigt die Vorhersagen der Kreuzvalidierungen, die während des Erstellens des Modells gemacht wurden, sowie die Vorhersagen der Testproben. Die Bestimmungskoeffizienten wurden für die Vorhersagen der Kreuzvalidierungen berechnet, wobei diese auf die Gerade idealer Korrelation bezogen wurden und nicht auf die am besten zu den vorhandenen Datenpunkten passende Gerade. Die entsprechenden Werte waren 0,77 für deamidiertes und 0,88 für nicht-deamidiertes Aspart. Diese Werte waren nicht so hoch wie für die anderen Testsysteme (s.o.). Angesichts der marginalen Unterschiede zwischen beiden Proteinen (vgl. Fig. 14), wurde gemäß der vorliegenden Anmeldung jedoch erkannt, daß eine Kalibrierung überhaupt möglich ist.

Ansprüche

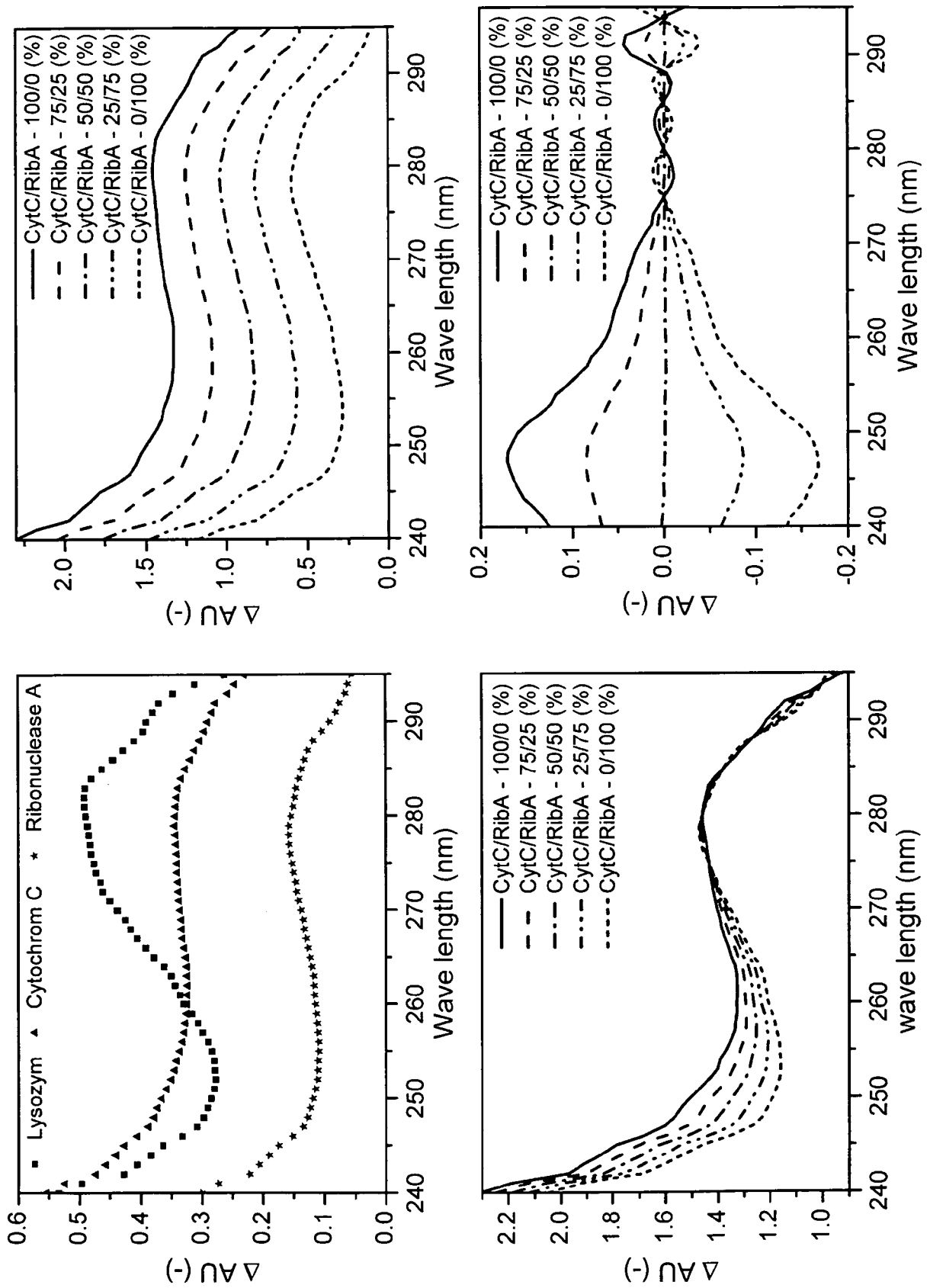
1. Verfahren zur selektiven Quantifizierung einzelner Proteine in einem Proteingemisch, umfassend die Schritte:
 - 5 (a) Erstellen mindestens eines UV-Absorptionsspektrums des Proteingemisches, und
 - (b) Bestimmen der Einzelkonzentration zumindest eines Proteins in dem Proteingemisch aus dem in Schritt (a) erstellten UV-Absorptionsspektrum.
- 10 2. Verfahren nach Anspruch 1, wobei das Bestimmen der Einzelkonzentration(en) in Schritt (b) rechnerisch erfolgt, insbesondere mit Hilfe eines multivariaten Kalibrierungsmodells durchgeführt wird.
3. Verfahren nach Anspruch 2, wobei das multivariate Kalibrierungsmodell auf
15 Grundlage von UV-Absorptionsspektren von Kalibrierungsproben erstellt wird, die bekannte Proteine in bekannten Konzentrationen enthalten.
4. Verfahren nach Anspruch 2 oder 3, wobei aus den UV-Absorptionsspektren der jeweils nur ein Protein in bekannter Konzentration enthaltenden
20 Kalibrierungsproben durch Linearkombinationen Mischspektren erzeugt werden, die für die Erstellung des multivariaten Kalibrierungsmodells verwendet werden.
5. Verfahren nach einem der Ansprüche 2 bis 4, wobei das multivariate
25 Kalibrierungsmodell mit Hilfe eines statistischen Verfahrens zur multivariaten Kalibrierung erstellt wird.
6. Verfahren nach Anspruch 5, wobei das statistische Verfahren zur multivariaten Kalibrierung aus der Gruppe, bestehend aus *Multivariate Curve Resolution*
30 (MCR), *partial least square* (PLS)-Regression und Multilinearer Regression (MLR) ausgewählt ist.
7. Verfahren nach einem der vorangehenden Ansprüche, wobei mindestens vier UV-Absorptionsspektren pro Sekunde erstellt werden.

8. Verfahren nach einem der vorangehenden Ansprüche, wobei das Proteingemisch N verschiedene Proteine umfaßt, wobei N eine natürliche Zahl und bevorzugt $N \geq 2$ ist.
- 5 9. Verfahren nach einem der Ansprüche 2 bis 8, wobei das Proteingemisch kein dem multivariaten Kalibrierungsmodell unbekanntes Protein enthält.
10. Verfahren nach einem der vorangehenden Ansprüche, wobei das UV-Absorptionsspektrum ein kontinuierliches Spektrum ist, vorzugsweise in einem
10 Wellenlängenbereich zwischen 240 und 300 nm.
11. System zur selektiven Quantifizierung einzelner Proteine in einem Proteingemisch, umfassend:
- 15 (a) Mittel für eine säulenchromatographische Trennung eines Proteingemisches,
- (b) einen Detektor, insbesondere einen Dioden-basierten UV-Detektor, der am Säulenausgang angeordnet ist, und
- (c) Mittel für die automatische Verarbeitung der von dem Detektor erstellten UV-Absorptionsspektren und die Bestimmung der Konzentration der
20 einzelnen Proteine in dem Proteingemisch aus diesen UV-Absorptionsspektren mit Hilfe eines multivariaten Kalibrierungsmodells und eines geeigneten statistischen Verfahrens.
12. System zur selektiven Quantifizierung einzelner Proteine in Proteingemischen,
25 umfassend:
- (a) Mittel für die automatisierte Handhabung von Multi-Well-Platten, welche die Proteingemische enthalten,
- (b) einen Detektor, insbesondere einen Dioden-basierten UV-Detektor, der oberhalb oder unterhalb der Multi-Well-Platte angeordnet ist, und
30 (c) Mittel für die automatische Verarbeitung der von dem Detektor erstellten UV-Absorptionsspektren und die Bestimmung der Konzentration der einzelnen Proteine in den Proteingemischen aus diesen UV-Absorptionsspektren mit Hilfe eines multivariaten Kalibrierungsmodells und eines geeigneten statistischen Verfahrens.

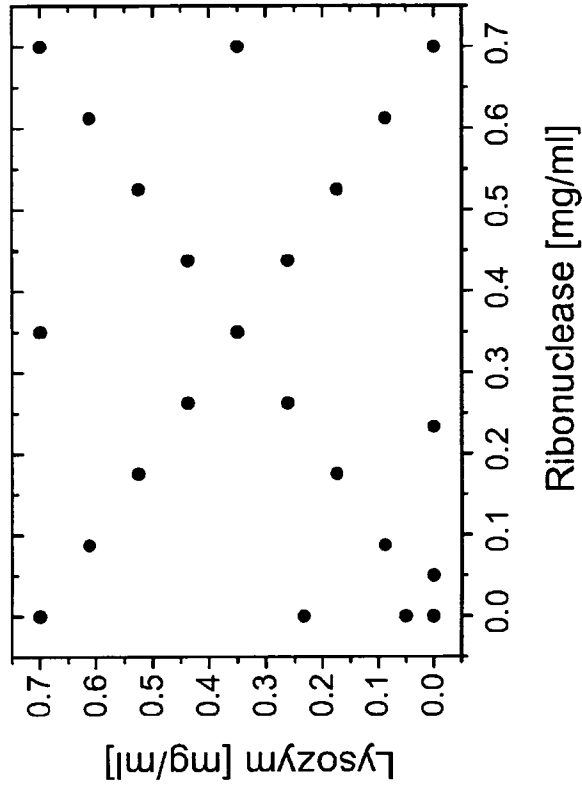
Figuren



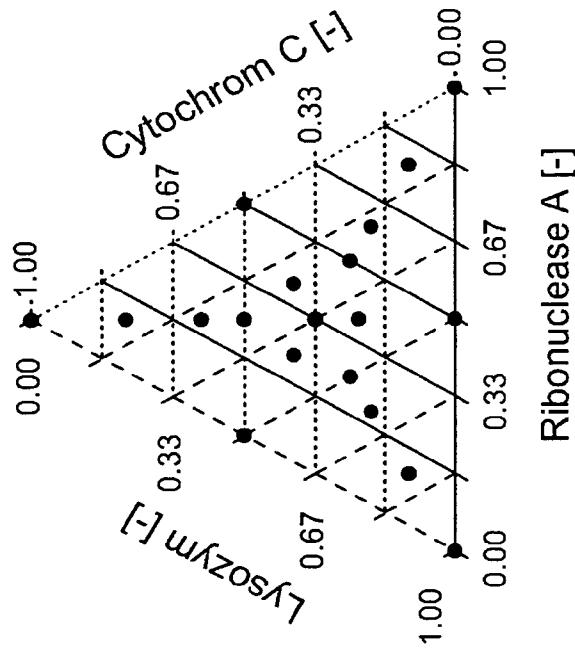
Figur 1



Figur 2



(b) Kalibrierungsdesign - absolute Konzentrationen



(a) Kalibrierungsdesign

Figure 3

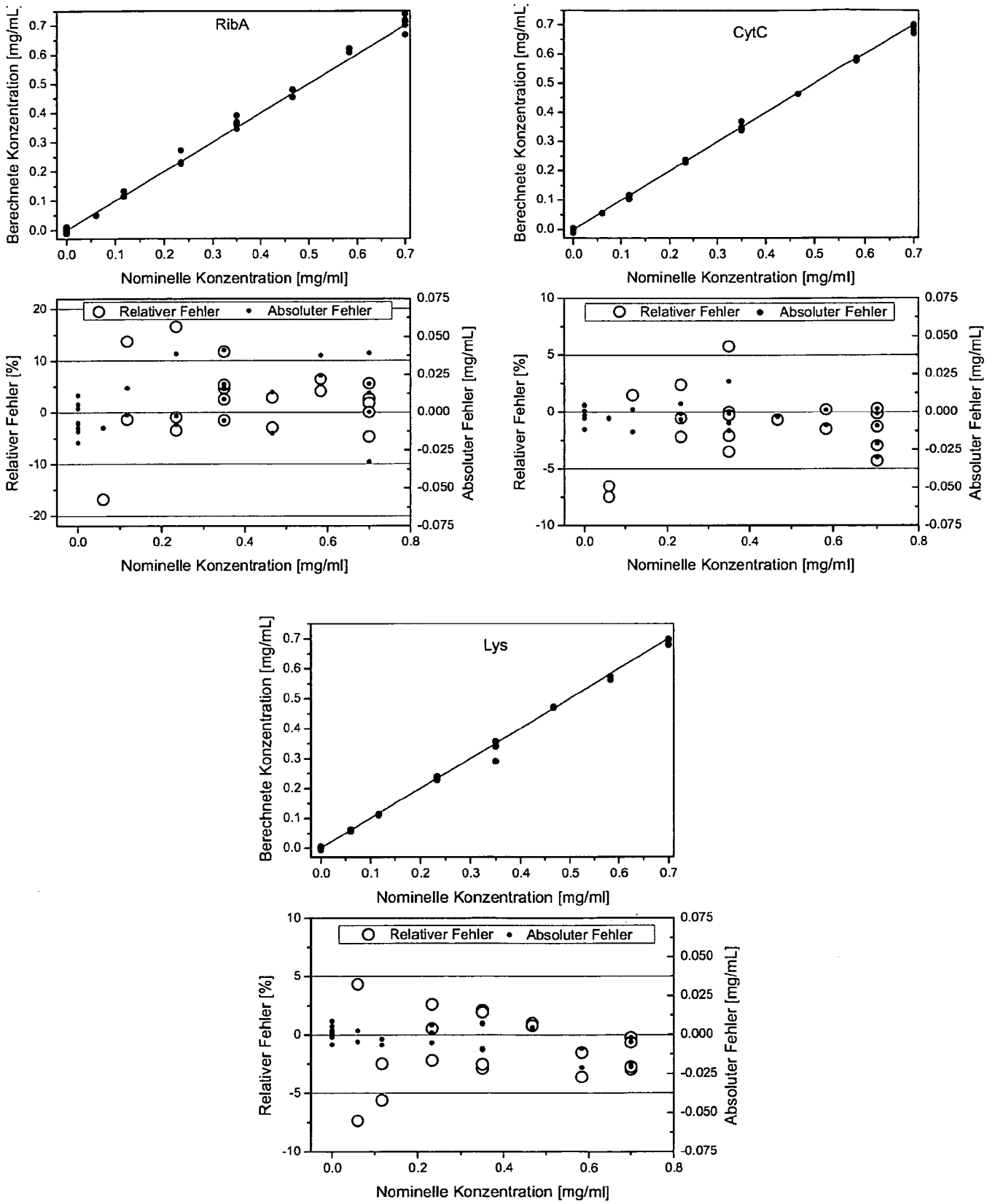
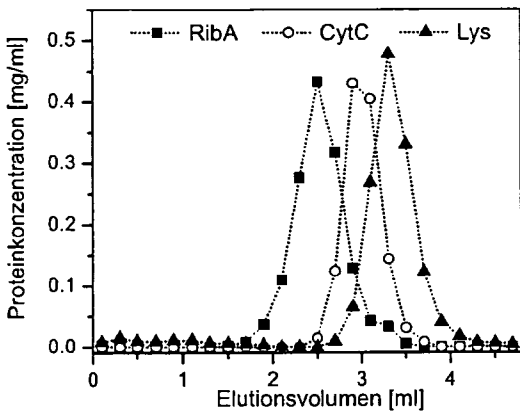
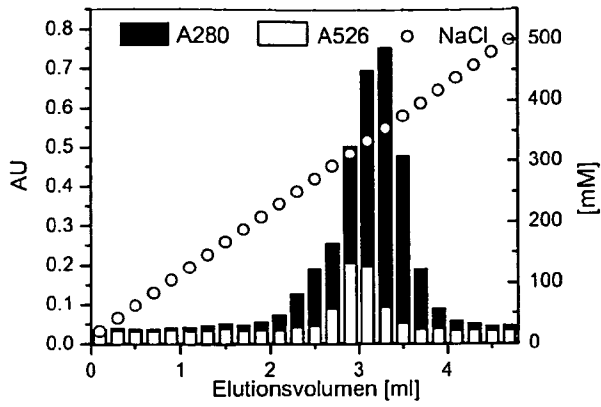
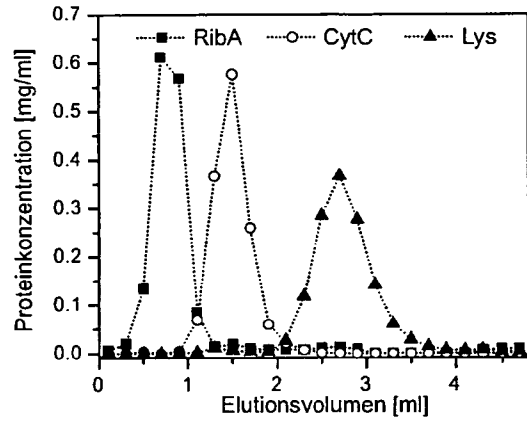
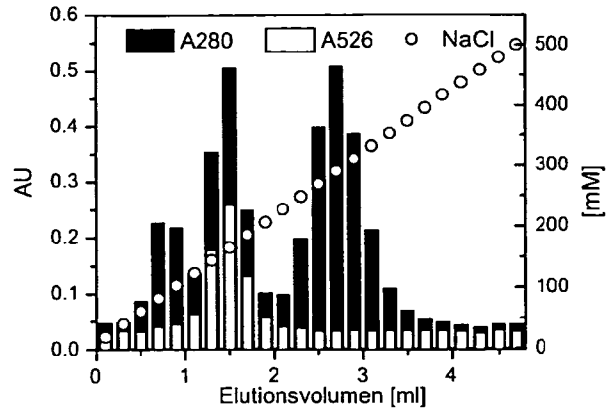


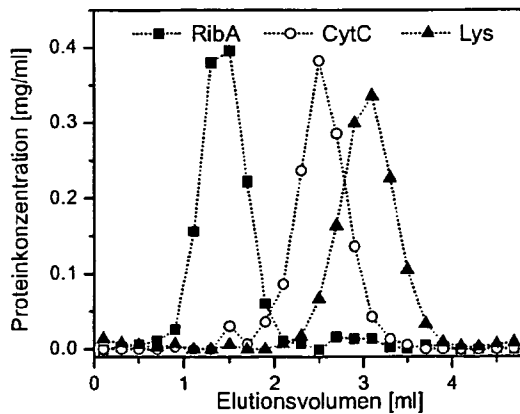
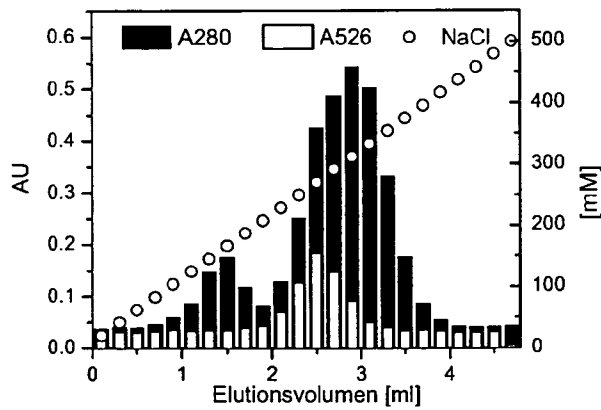
Figure 4



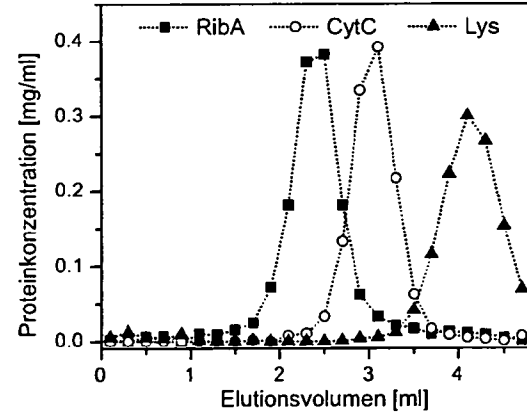
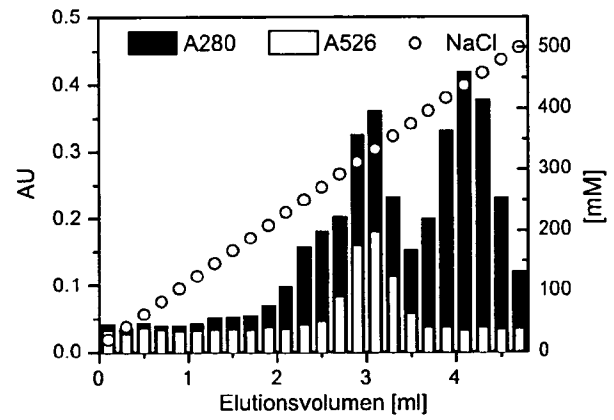
(a) CM Ceramic HyperD F



(b) Toyopearl SP 850 M



(c) SP Sepharose FF



(d) Fractogel EMD₃

Figure 5

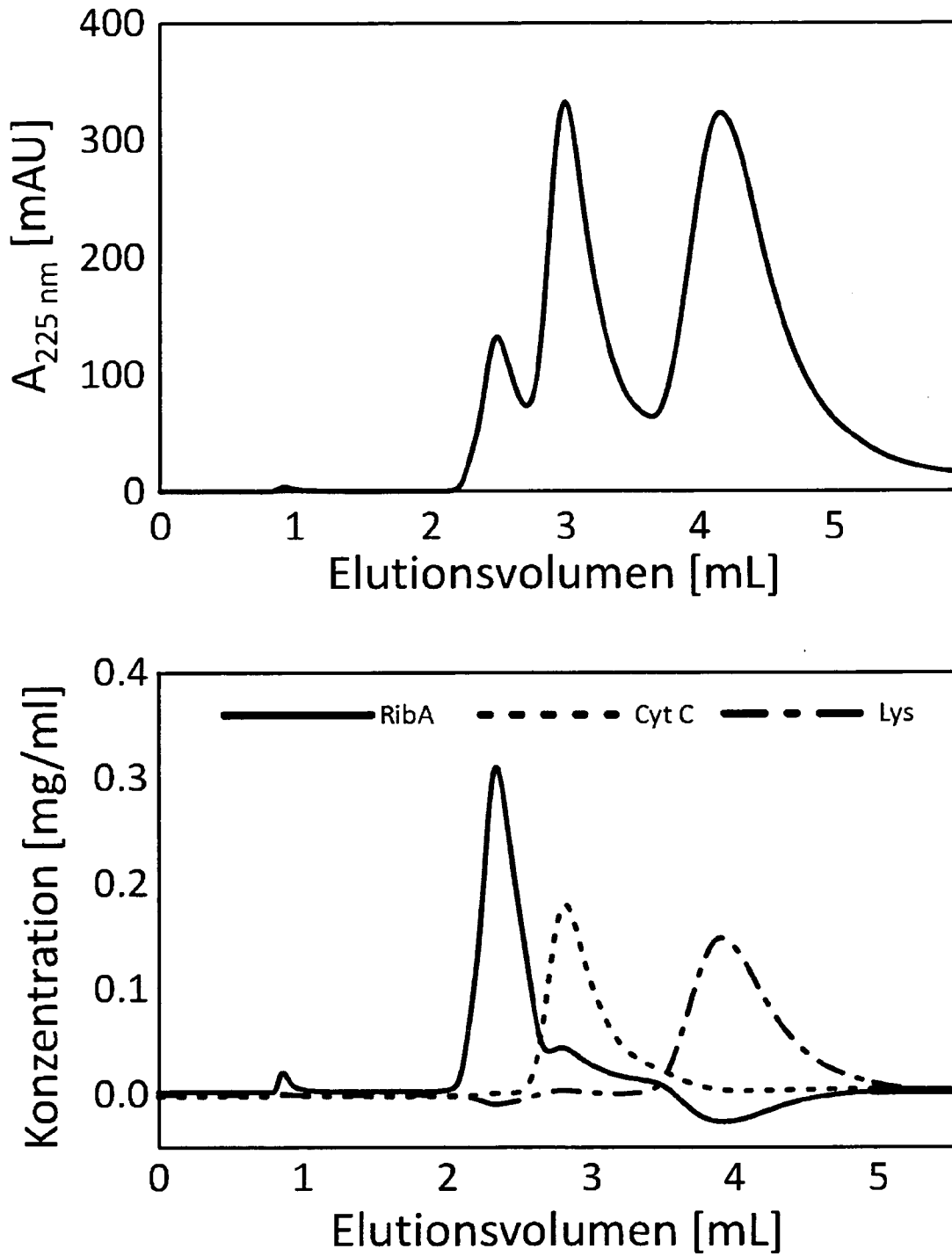
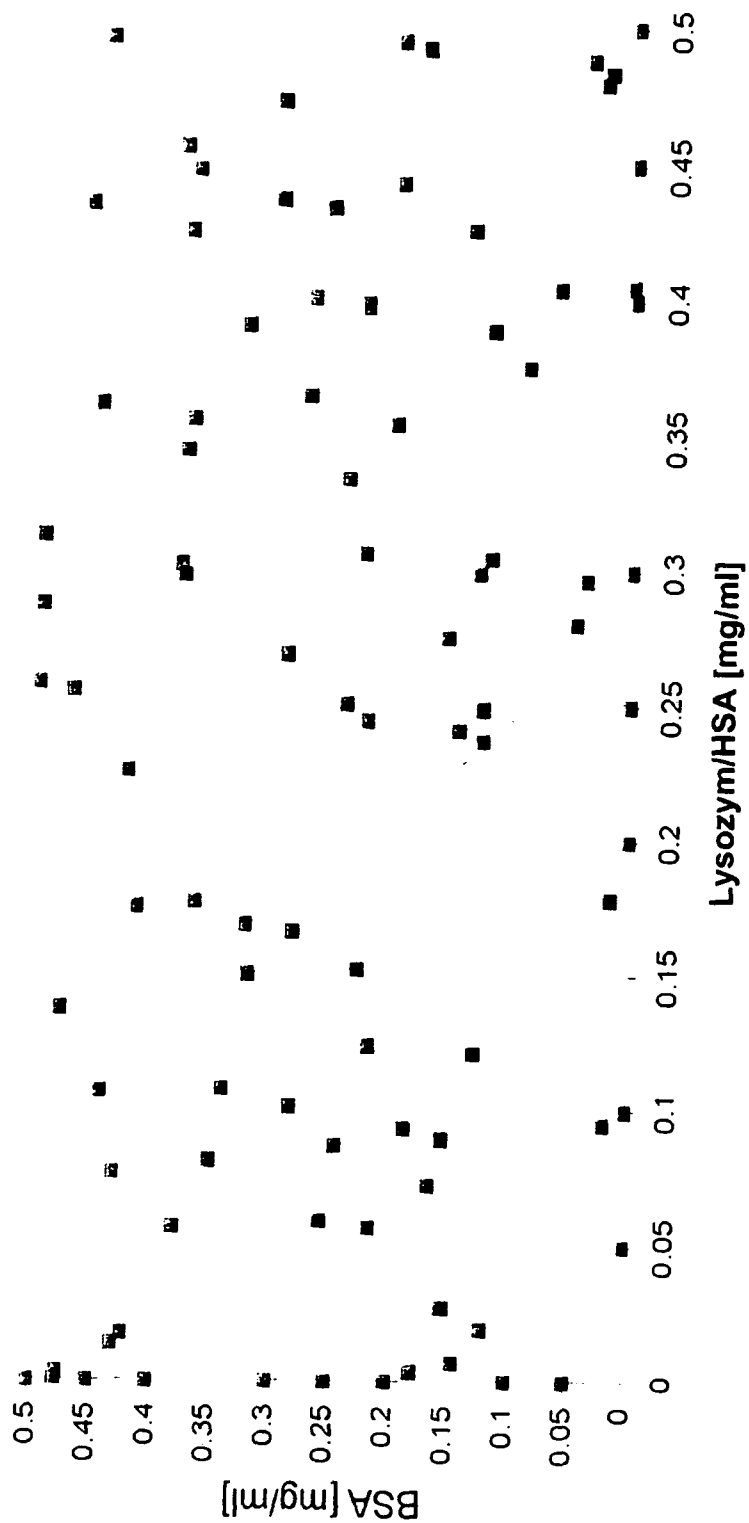
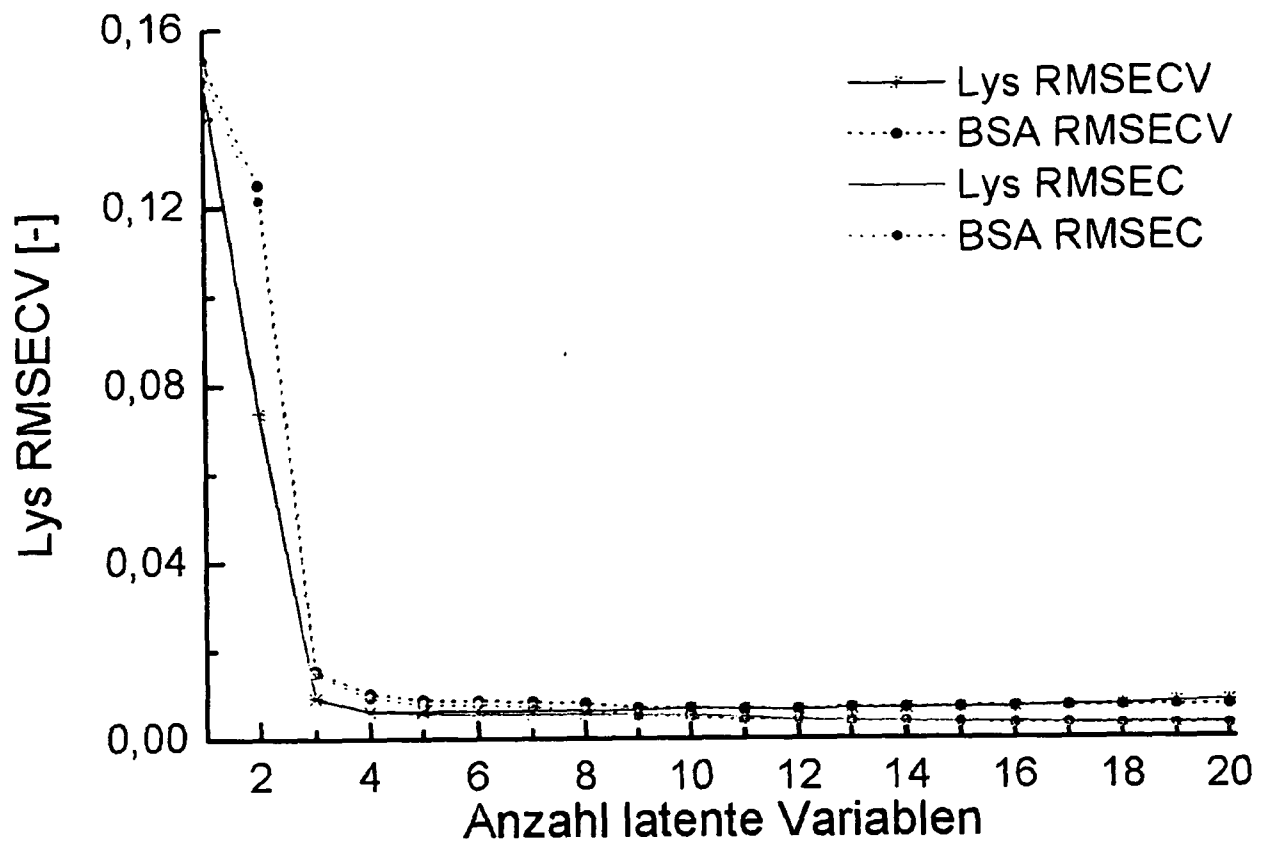


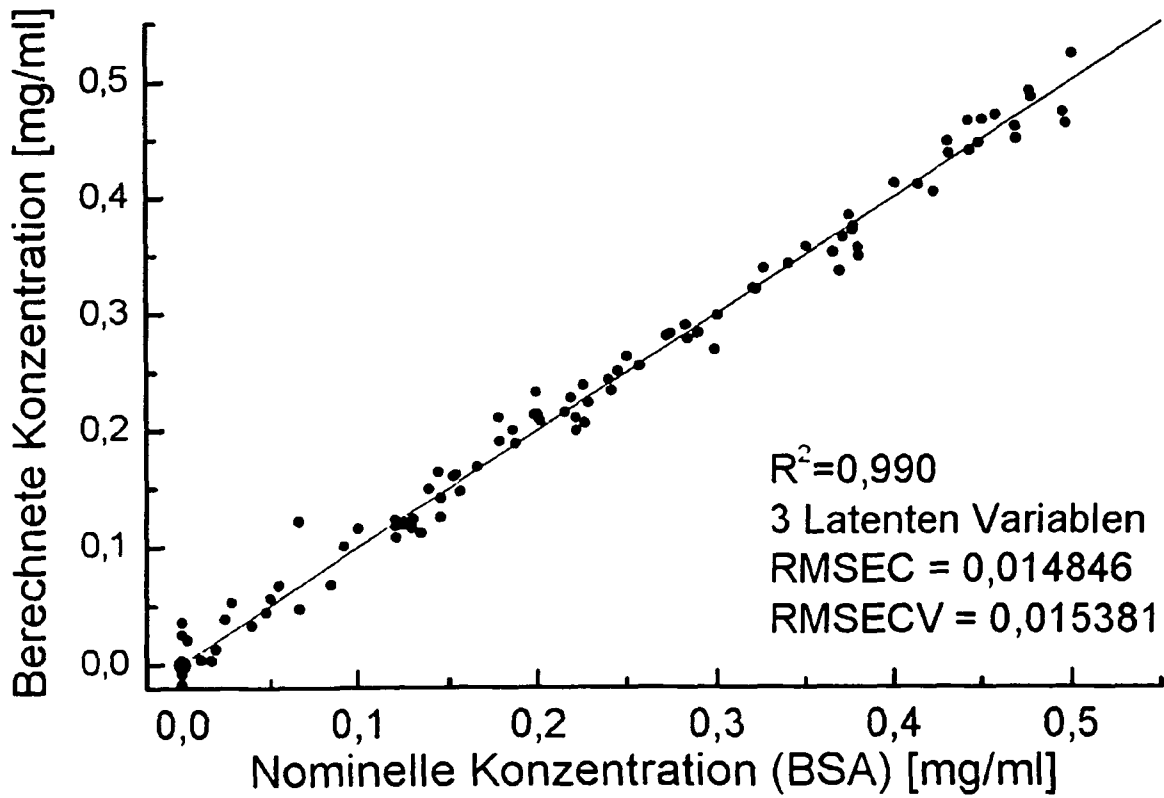
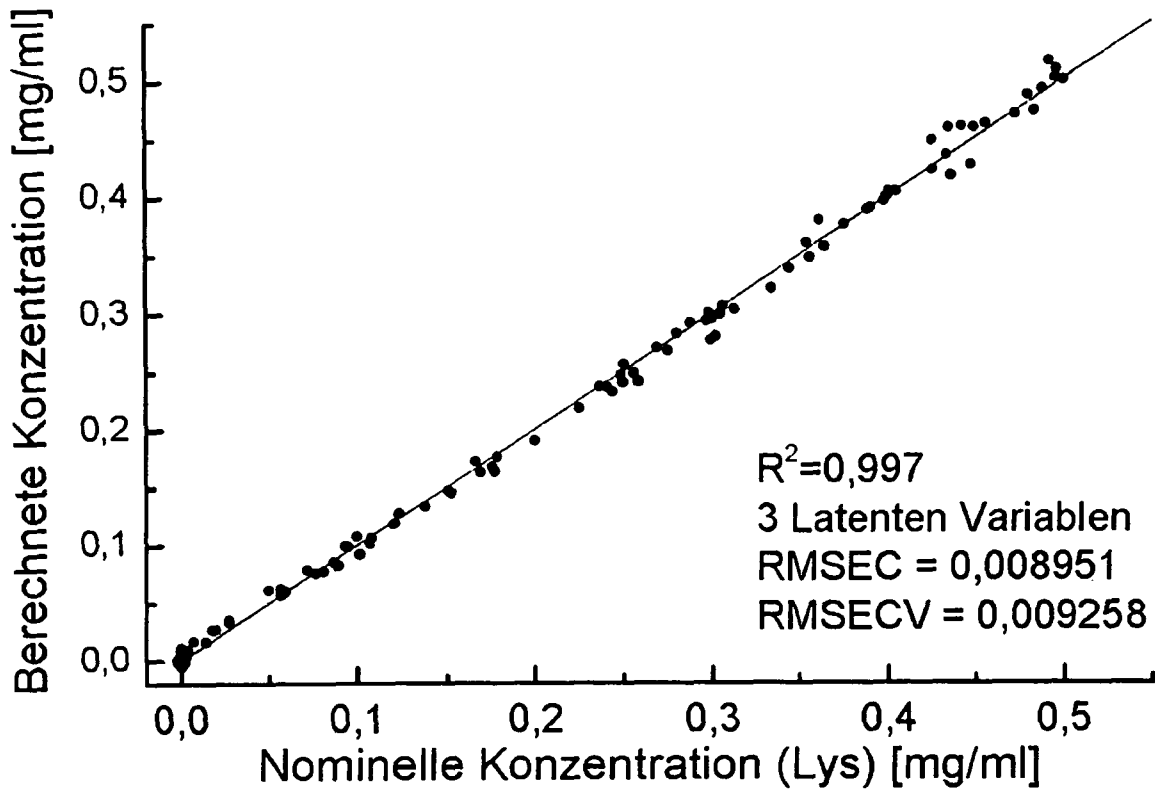
Figure 6



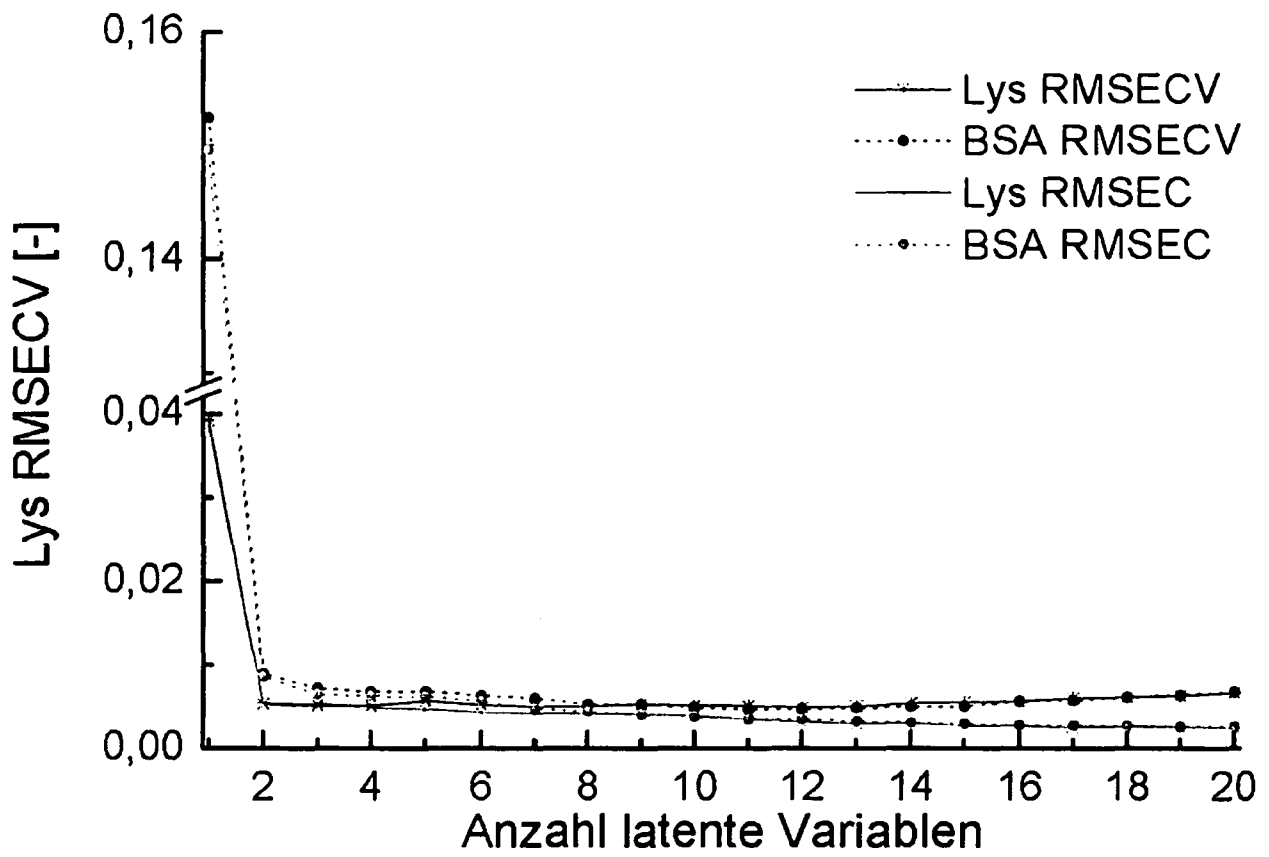
Figur 7



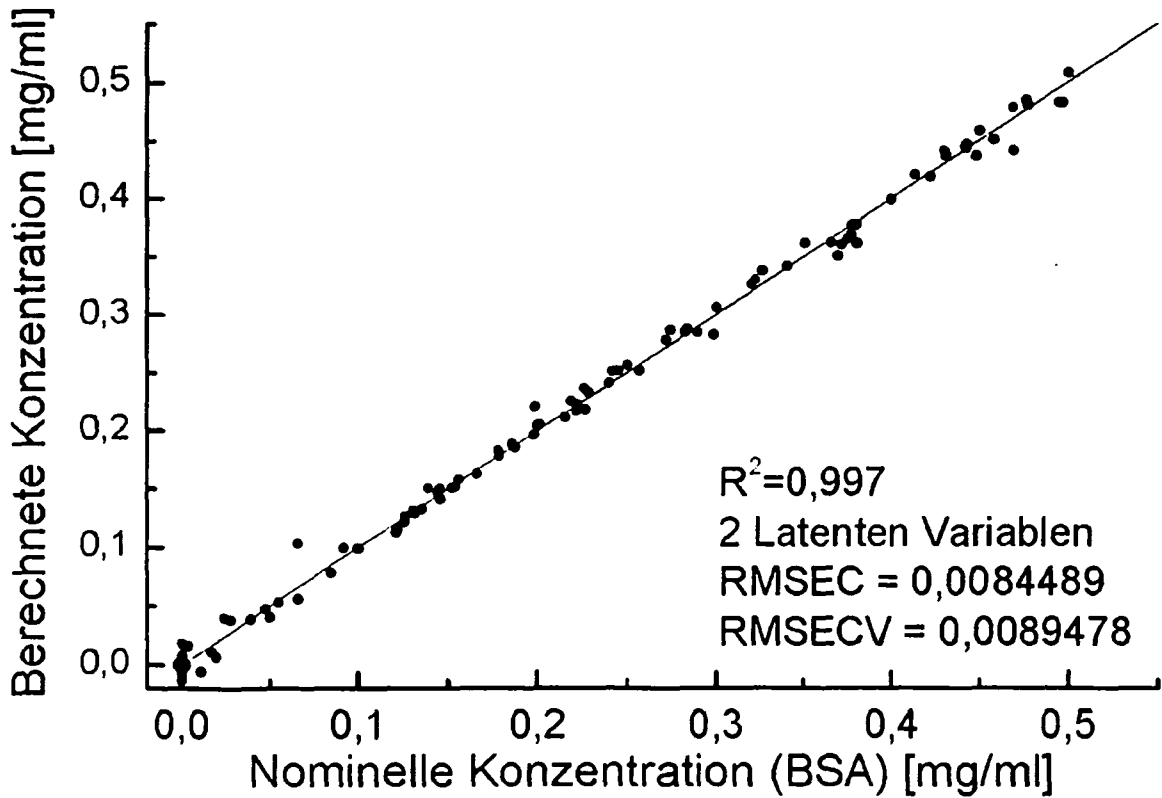
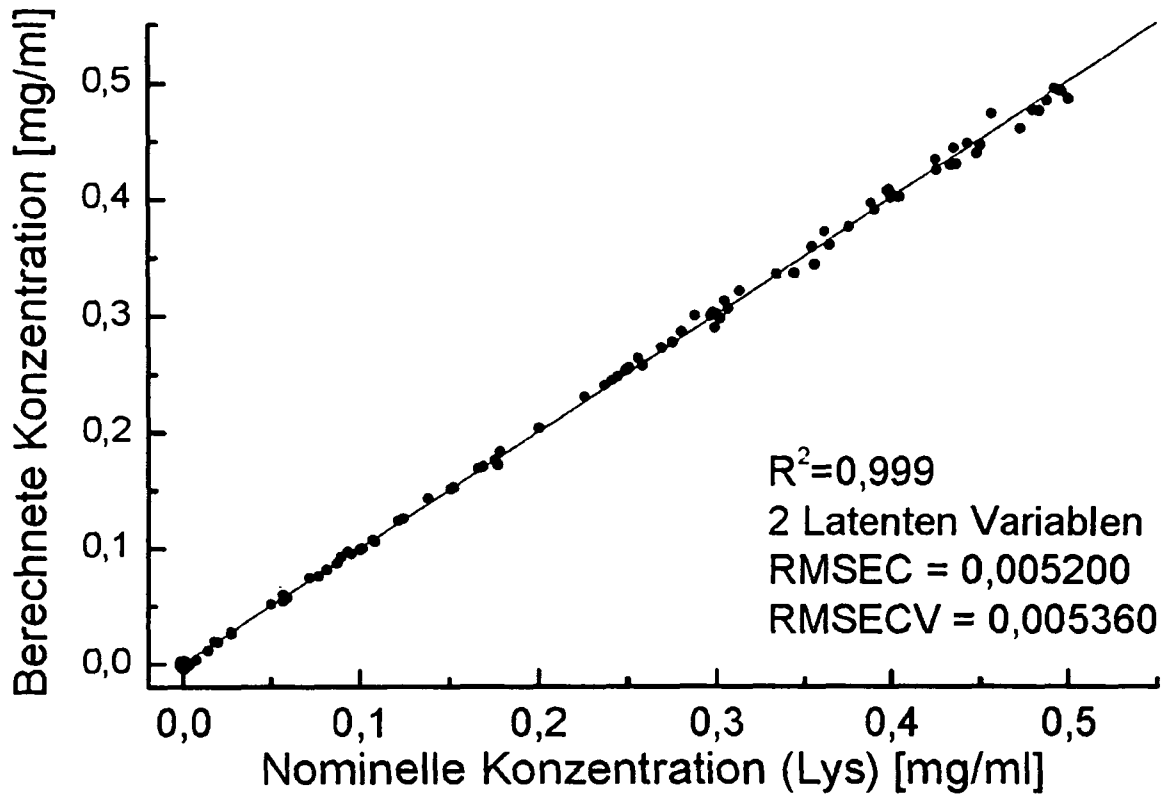
Figur 8



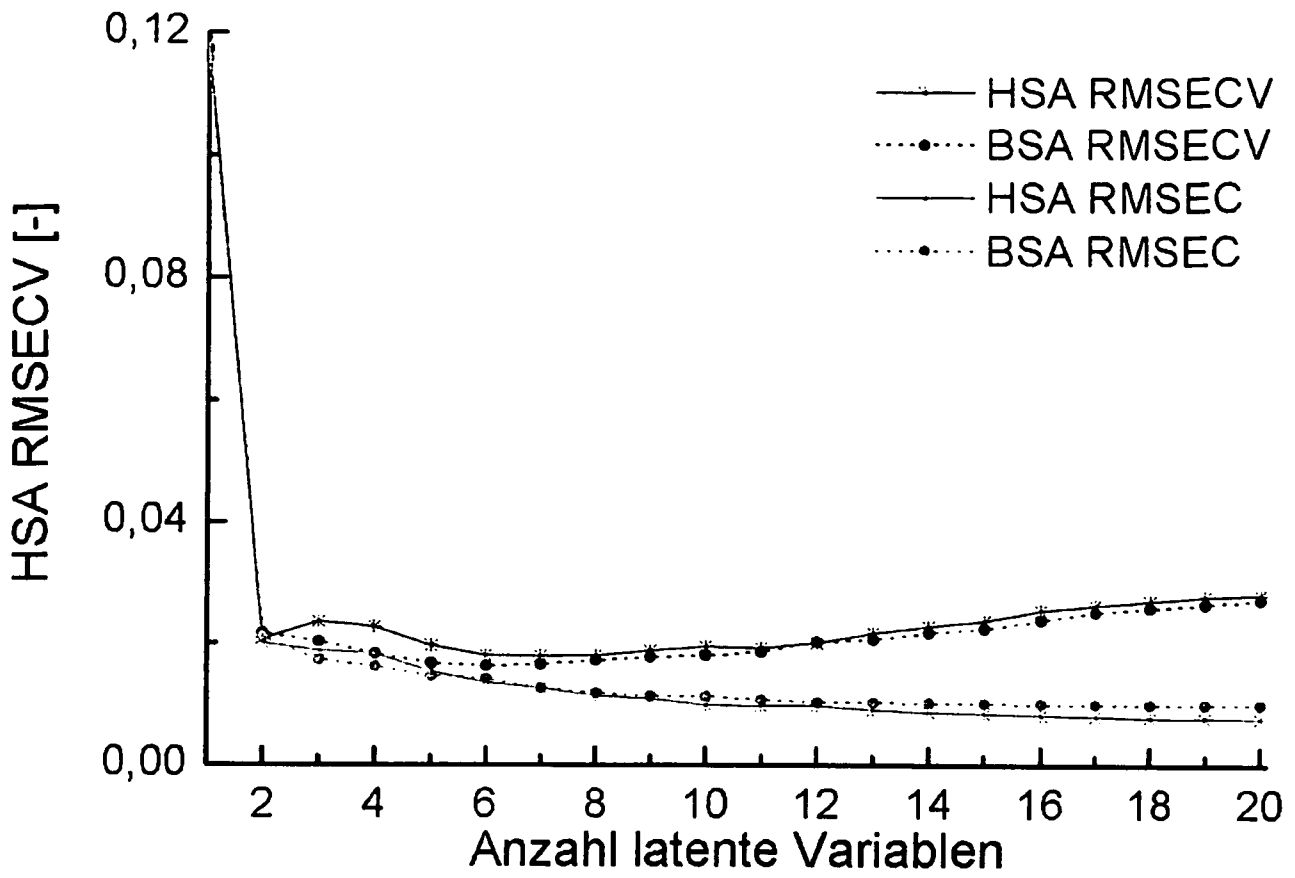
Figur 9



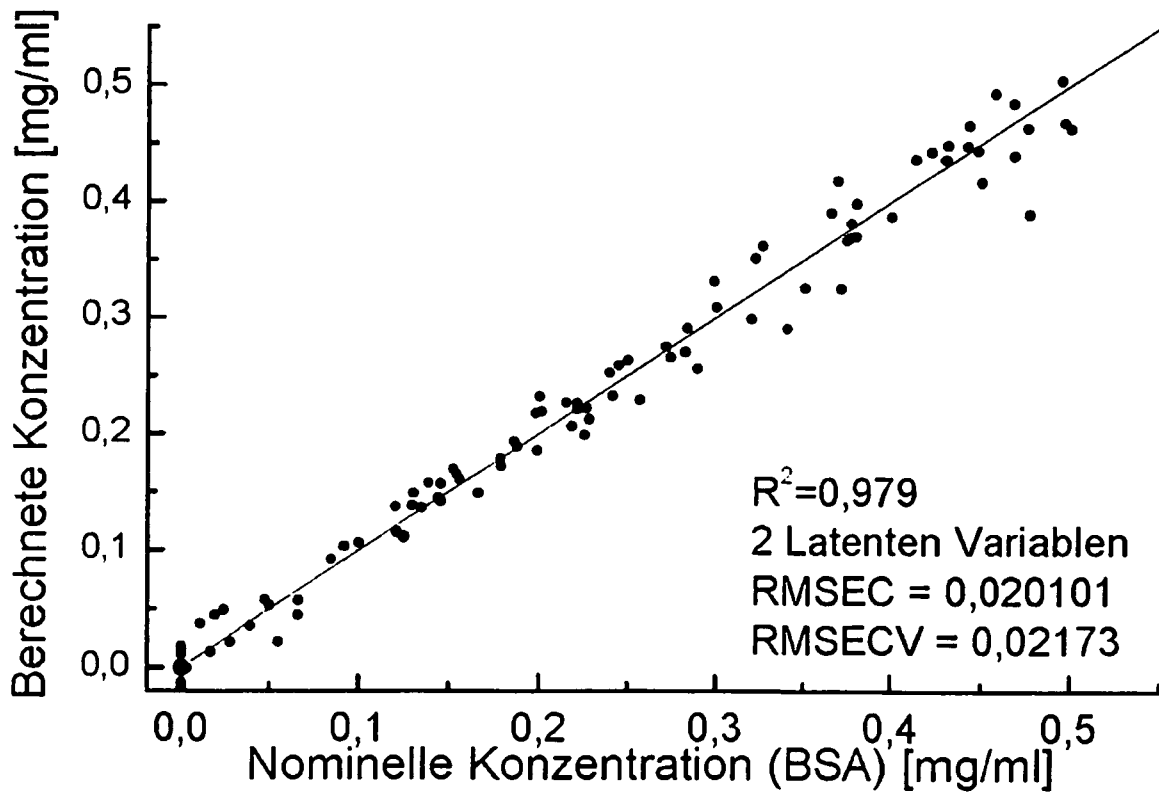
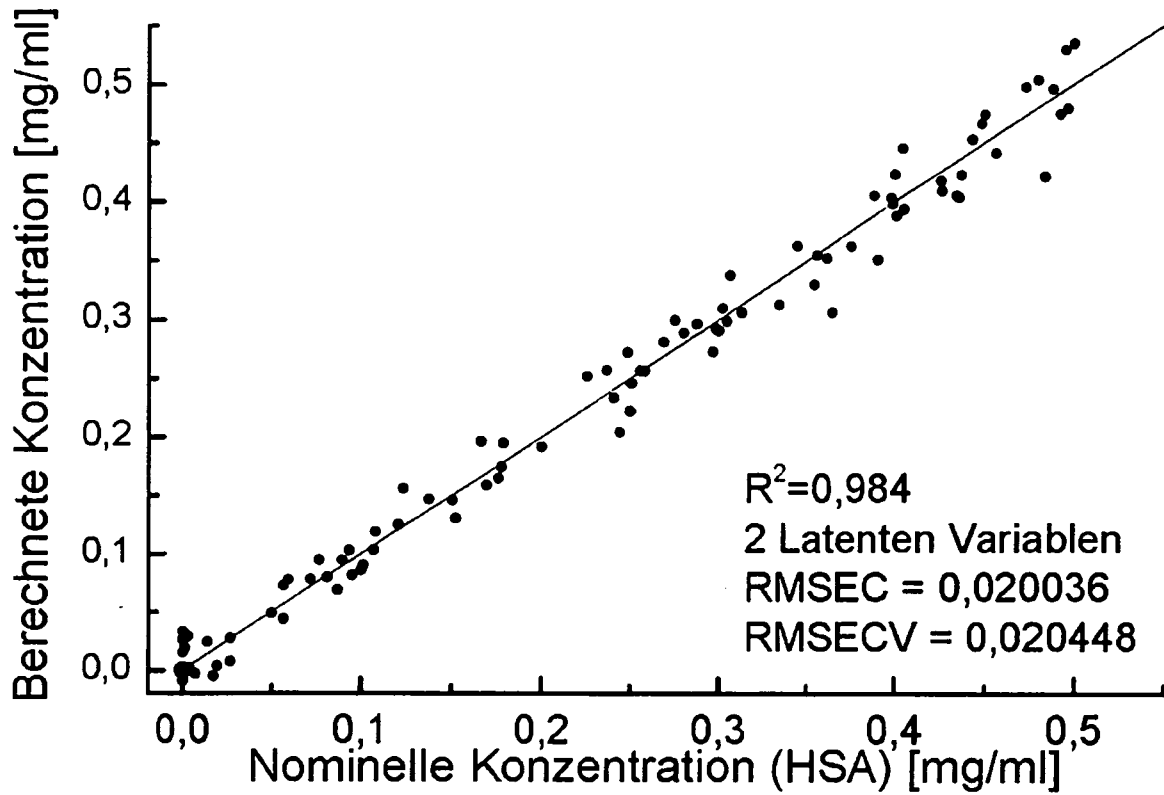
Figur 10



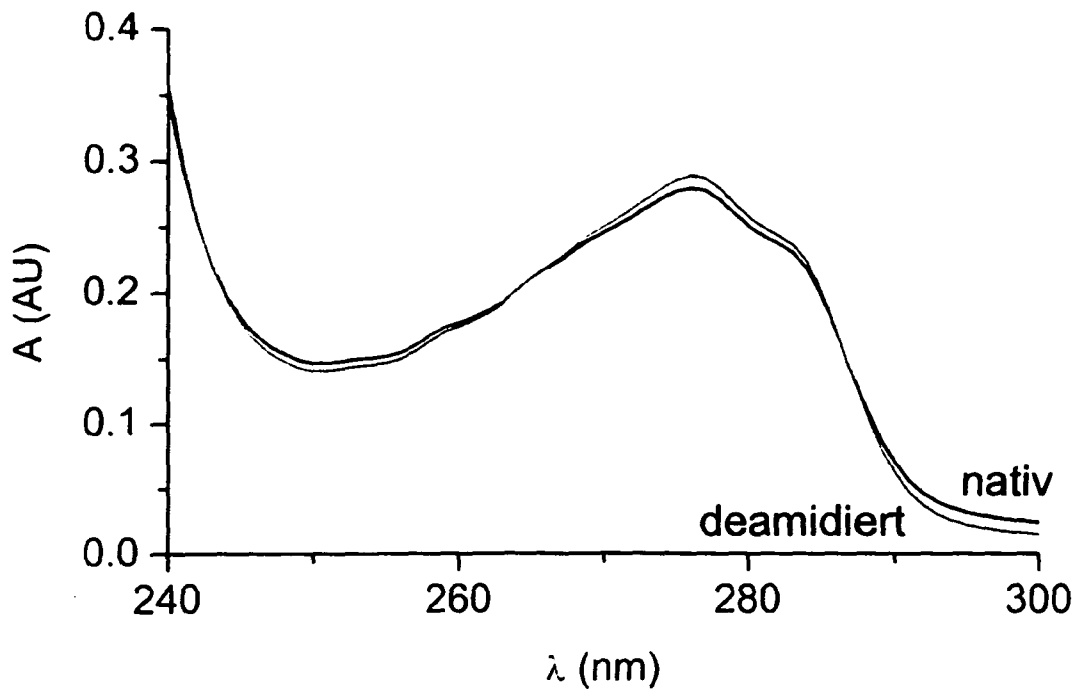
Figur 11



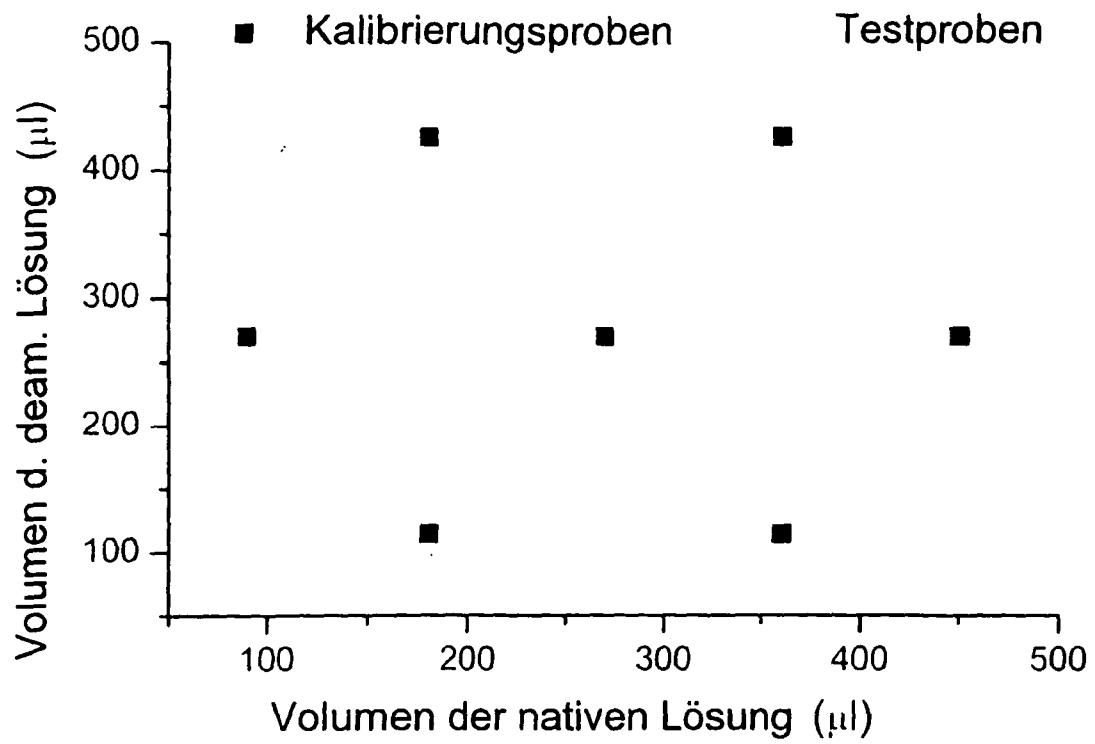
Figur 12



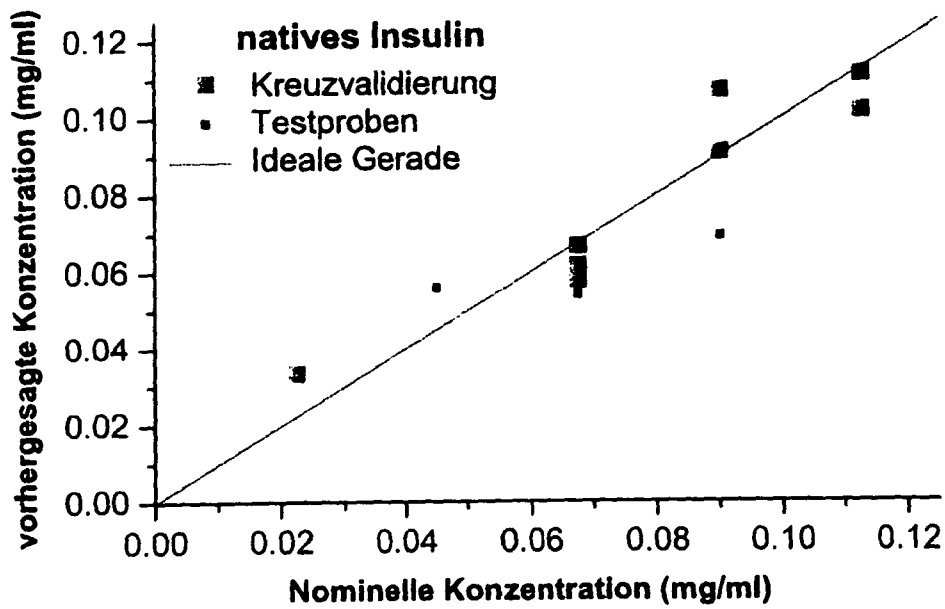
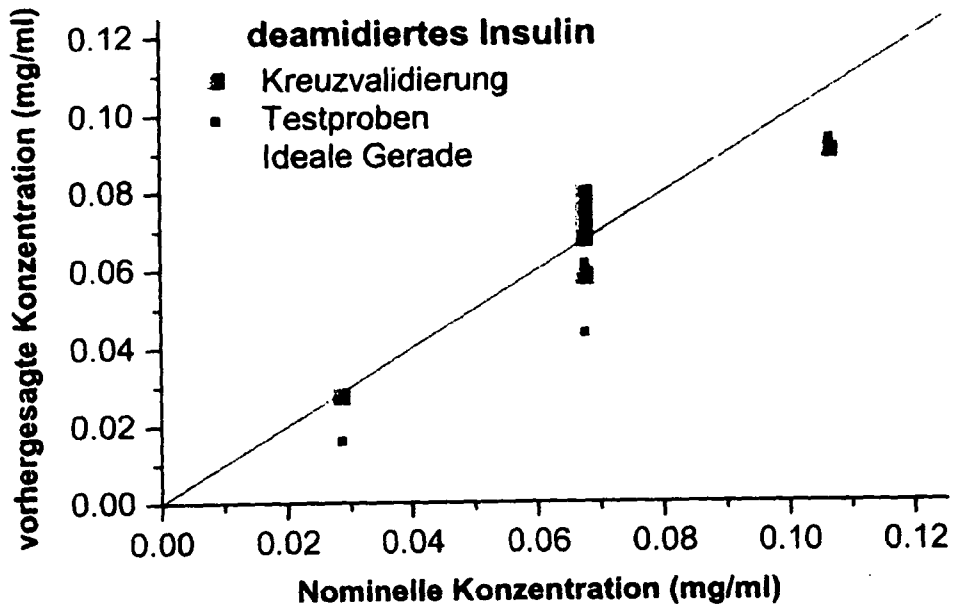
Figur 13



Figur 14



Figur 15



Figur 16