



(19) Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) DE 11 2005 002 331 T5 2007.11.08

(12)

Veröffentlichung

der internationalen Anmeldung mit der
(87) Veröffentlichungs-Nr.: **WO 2005/106764**
in deutscher Übersetzung (Art. III § 8 Abs. 2 IntPatÜG)
(21) Deutsches Aktenzeichen: **11 2005 002 331.1**
(86) PCT-Aktenzeichen: **PCT/US2005/000202**
(86) PCT-Anmeldetag: **06.01.2005**
(87) PCT-Veröffentlichungstag: **10.11.2005**
(43) Veröffentlichungstag der PCT Anmeldung
in deutscher Übersetzung: **08.11.2007**

(51) Int Cl.⁸: **G06F 19/00 (2006.01)**

(30) Unionspriorität:
60/535,352 **09.01.2004** **US**
10/794,407 **05.03.2004** **US**

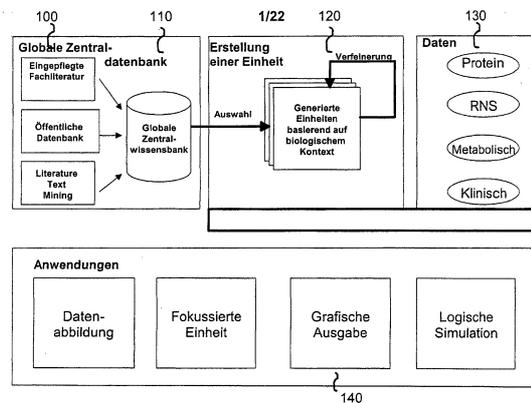
(71) Anmelder:
Genstruct, Inc., Cambridge, Mass., US

(74) Vertreter:
COHAUSZ & FLORACK, 40211 Düsseldorf

(72) Erfinder:
Sun, Justin, Norwood, Mass., US; Pratt, Dexter R., Reading, Mass., US; Levy, Joshua, Chapel Hill, N.C., US; Kightley, David A., York, Me., US; Chandra, Navin D., , ZZ

(54) Bezeichnung: **Verfahren, System und Vorrichtung zur Zusammenstellung und Nutzung von biologischem Wissen**

(57) Hauptanspruch: Verfahren, um neues biologisches Wissen zu generieren, das die folgenden Schritte umfasst:
(a) die Bereitstellung einer Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, die biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen den Knoten charakterisieren;
(b) das Gewinnen einer Teilmenge von Assertionen aus der Datenbank, die eine Reihe von biologischen Kriterien erfüllen, die ein Nutzer so spezifiziert hat, dass sie ein ausgewähltes biologisches System definieren;
(c) die Kompilierung der erfassten Daten, um eine Einheit zu erzeugen, die eine biologische Wissensbank von Assertionen bildet, die potenziell relevant für das genannte ausgesuchte biologische System ist; und
(d) die Transformierung der genannten Einheit, um neues biologisches Wissen über das genannte biologische System zu generieren.



Beschreibung

Verwandte Anmeldungen

[0001] Diese Anmeldung erhebt Anspruch auf die Nutzung der provisorischen Anmeldung Nr. 60/535,352 mit dem Titel „Verfahren, System und Vorrichtung zur Zusammenstellung und Nutzung von biologischem Wissen“, beantragt am 9. Januar 2004, deren Offenlegung hierin durch Bezugnahme eingefügt ist.

Technisches Gebiet

[0002] Die Erfindung betrifft Verfahren, Systeme und Vorrichtungen, um neues biologisches Wissen zu ermitteln, insbesondere Verfahren, Systeme und Vorrichtungen, um eine biologischen Wissensbank zusammenzustellen, Verfahren, Systeme und Vorrichtungen, um Teilmengen aus biowissenschaftlichen Daten und Informationen zu bilden und in biologische Modelle zu transformieren sowie Verfahren, Systeme und Vorrichtungen zur Vereinfachung der elektronischen Verarbeitung von und des Schlussfolgerns aus biologischen Informationen.

Hintergrund

[0003] Die in der heutigen Welt erzeugte Menge biologischer Informationen nimmt dramatisch zu. Man schätzt, dass die Menge der Information sich derzeit alle vier bis fünf Jahre verdoppelt. Mit der riesigen Informationsmenge, die verarbeitet und analysiert werden muss, sind traditionelle Verfahren (bzw. Methoden) des Erkennens und Verstehens der Bedeutung dieser Informationen, vor allem im Bereich der Biowissenschaften, überfordert.

[0004] Um zu einem tatsächlichen Verständnis eines biologischen Systems zu gelangen, muss ein Forscher in den Biowissenschaften Informationen aus vielen Quellen zusammenfügen. Das Begreifen biologischer Systeme wird durch die interdisziplinäre Natur der Biowissenschaften erschwert. Um biologische Systeme verstehen zu können, kann umfassendes Wissen in Genetik, Zellbiologie, Biochemie, Medizin und vielen anderen Fachgebieten erforderlich sein. Um biologische Systeme zu verstehen, kann es nötig sein, viele verschiedene Informationen zu verknüpfen. Biowissenschaftliche Informationen können Material über grundlegende Chemie, Proteine, Zellen, Gewebe und Auswirkungen auf Organismen und Populationen umfassen – die alle miteinander in Wechselbeziehung stehen könnten. Diese Wechselbeziehungen können komplex, kaum verstanden oder verborgen sein.

[0005] Es wird bereits versucht, elektronische Modelle biologischer Systeme zu erschaffen. Dazu gehören die Zusammenstellung und Organisation enor-

mer Datenmengen und der Aufbau eines Systems, das die Daten dazu verwenden kann, das Verhalten eines biologischen Systems zu simulieren. Wegen der Komplexität der Biologie und der schier unendlichen Datenmenge kann der Aufbau eines solchen Systems Hunderte von Jahren dauern und mehrere zehn Millionen Dollar kosten. Dazu stehen diejenigen, die in den Biowissenschaften neue Einsichten und neue Erkenntnisse suchen, vor der immer schwierigeren Aufgabe, die richtigen Daten aus den Bergen von Informationen, zusammengetragen aus vollkommen unterschiedlichen Quellen, zu verknüpfen. Unternehmen, die bereit waren, dafür Mittel zu investieren, hatten bisher keinen Erfolg beim Erarbeiten von tatsächlich brauchbaren Modellen, die Forscher wesentlich dabei unterstützen, biowissenschaftliches Wissen zu vergrößern. Deshalb sind die derzeitigen zur Verarbeitung und Analyse von Wissen entwickelten Systeme zur Verarbeitung und Aufzeichnung biowissenschaftliche Daten bei weitem nicht optimal und maßgebliche neue Leistungen werden benötigt.

[0006] Genauer ausgedrückt, braucht die Branche eine Methode, um die riesigen Mengen von Wissen aus den verschiedenen Bereichen der Biowissenschaften zusammenzustellen, und mit deren Hilfe kenntnisreiche und sinnvolle Modelle zu schaffen, die geprüft und abgefragt werden können, um neue biologische Beziehungen, Pfade, Ursachen und Wirkungen und andere Einsichten effizient und einfach zu erkennen.

Zusammenfassung der Erfindung

[0007] In Übereinstimmung mit der Erfindung wird erkannt, dass die Bereitstellung von Mitteln für den raschen und effektiven Aufbau von Teilwissensbanken und abgeleiteten Wissensbanken ein Schlüssel ist zur Bereitstellung nützlicher und leichter zu bedienenden biologischer Wissensbanken, die biologische Systeme effektiv nachstellen können. Diese spezialisierten Wissensbanken können von einer globalen Wissensbank ausgehend aufgebaut werden, indem zu Beginn eine potenziell relevante Teilmenge von biowissenschaftlich relevanten Daten, die von einem Benutzer spezifizierte Kriterien erfüllen, erfasst wird eine speziell ausgerichtete Wissensbank mit der hier dargelegten Struktur zusammengestellt wird.

[0008] Diese können in verschiedenen Formaten verfeinert, augmentiert, untersucht und in verschiedenen Formaten dargestellt, durch menschliche Beobachtung und Analyse genutzt werden; dazu können mithilfe einer Vielzahl von Instrumenten Verständnis und Aufdeckung versteckter Interaktionen und Beziehungen in biologischen Systemen erleichtert werden, d. h. neues biologisches Wissen wird geschaffen. Das wiederum erlaubt das Aufstellen neuer Hypothesen über biologische Pfadverläufe, die auf dem neuen biologischen Wissen basieren, und erlaubt dem

Nutzer, eine Hypothese zu bestätigen oder zu widerlegen, indem er biologische Experimente mit Biomolekülen, Zellen und Tiermodellen oder einen klinischen Versuch entwirft und durchführt.

[0009] Die Erfindung bietet also ein neues Paradigma, neue Verfahren, Vorrichtungen und Instrumente, die auf eine globale Wissensbank angewendet werden können. Die Instrumente und Verfahren ermöglichen eine effiziente Durchführung von Erkundungsprojekten in den für die Biowissenschaften relevanten Bereichen. Die Erfindung stellt neue Verfahren und Instrumente bereit, die es erlauben, eine Wissensbank zu konditionieren, um bei einem Projekt oder einer Aufgabe sowohl Fokussierung als auch Flexibilität zu ermöglichen. Die Erfindung erlaubt es auch, jedes biologische Thema zu behandeln, egal wie obskur oder esoterisch es sein mag, vorausgesetzt, es gibt zumindest einige Assertionen in einer globalen Wissensbank, die für dieses Thema relevant sind. Assertionen sind wahre Aussagen, die sich auf ein existierendes Objekt in einem System oder eine wahre Aussage über ein Objekt in diesem System und einen Literalwert oder eine beliebige Kombination daraus beziehen. Jede wahre Aussage in einer Wissensbank oder einer Einheit wird hier als eine Assertion bezeichnet.

[0010] Ein Aspekt der vorliegenden Erfindung ist die Erfassung einer Teilmenge von Daten, die notwendig oder hilfreich ist, aus einer globalen Wissensbank oder einer zentralen Datenbank und der Neuaufbau einer spezialisierteren Teilwissensbank, die speziell für den vorliegenden Zweck konzipiert wird. Dazu ist es wichtig, dass die Struktur der globalen Wissensbank so konzipiert wird, dass man eine Teilwissensbank erfassen kann, die die relevanten Beziehungen zwischen den Informationen der Teilwissensbank bewahrt. Die Teilwissensbank, oder was hier der Einfachheit halber als eine Einheit bezeichnet wird, erlaubt eine Selektion und die logische Organisation von scheinbar verschiedenen Daten in ein kohärentes Modell eines beliebigen biologischen Systems, definiert durch irgendeine gewünschte Kombination von Kriterien. Diese Einheiten sind Mikrokosmen der globalen Wissensbank, sie können in dem Gebiet, das sie abdecken, detaillierter und umfassender als die globale Wissensbank sein und einfacher, produktiver und effizienter genutzt werden. Einheiten können zusammengeführt werden, sie können sich gegenseitig augmentieren oder wieder zur globalen Wissensbank hinzugefügt werden. Wie bereits dargestellt, sind die Begriffe Einheit und Wissensbank austauschbar.

[0011] Ein wichtiger Aspekt der Erfindung ist, dass sie die Generierung von abgeleiteten Einheiten ermöglicht. Abgeleitete Einheiten sind solche, in denen auf der Basis von logischen Inferenzen aus Assertionen neue Assertionen erzeugt werden. Abgeleitete

Einheiten können durch das Ziehen von Schlüssen und andere Algorithmen augmentiert werden. Die Augmentierung wird durch das Hinzufügen neuen Wissens erreicht, das möglicherweise, aber nicht notwendigerweise Teil der ursprünglichen Einheit oder der globalen Wissensbank ist. Die Augmentierung beinhaltet das Ziehen von Schlüssen aus der Einheit und den Abgleich der Einheit mit externen Daten (z. B. Labordaten, klinischen Daten, Literaturdaten usw.), ist aber nicht darauf beschränkt.

[0012] Die Erfindung stellt Verfahren zum Aufbau einer Wissensbank, Mittel zur deren Erstellung und Instrumente zu deren Verfeinerung bereit. Im Einzelnen bietet die Erfindung Verfahren an, um eine biologische Wissensbank aufzubauen, indem zuerst eine Datenbank biologischer Assertionen oder Mittel wie eine Benutzeroberfläche zum Zugriff auf eine solche Wissensbank bereitgestellt werden, die eine Vielzahl von Knoten umfasst, welche biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen diesen charakterisieren. Eine bevorzugte Wissensbank ist in der parallel anhängigen US-Patentanmeldung mit der Seriennummer 10/644,582 des selben Anmelders dargelegt, deren Offenlegung hierin durch Bezugnahme eingefügt ist. Als nächstes erfasst die Methode eine Teilmenge von Assertionen aus der Wissensbank, die eine Reihe biologischer Kriterien erfüllt, die ein Benutzer spezifiziert hat, um ein ausgewähltes biologisches System zu definieren. Dann werden die erfassten Daten kompiliert, um eine Einheit, d. h. eine biologische Wissensbank von Assertionen, die für das ausgewählte biologische System relevant sein könnten, zu erstellen.

[0013] Die Erfindung stellt Verfahren zur Ermittlung von neuem biologischen Wissen bereit. Die Verfahren beinhalten die Bereitstellung einer Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, die biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen diesen charakterisieren. Die Verfahren beinhalten auch die Erfassung einer Teilmenge von Assertionen, die eine Reihe von biologischen Kriterien erfüllen, die ein Benutzer spezifiziert hat, um ein ausgewähltes biologisches System zu definieren. Die Verfahren beinhalten darüber hinaus die Kompilierung der erfassten Assertionen, um eine biologische Wissensbank von Assertionen, die für das ausgewählte biologische System relevant sein könnten, zu erstellen, und die biologische Wissensbank danach zu analysieren, um neues biologisches Wissen zu ermitteln. Die Erfindung stellt auch Verfahren zur Generierung von neuem biologischen Wissen bereit, indem eine Datenbank biologischer Assertionen bereitgestellt wird, die eine Vielzahl von Knoten umfasst, welche biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen diesen charakterisieren, und dann eine

Vielzahl der biologischen Assertionen zu transformieren, um ein abgeleitetes Wissensnetzwerk zu erzeugen.

[0014] Die Erfindung bietet Verfahren, um eine biologische Wissensbank zu nutzen, darunter auch die Bereitstellung einer Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, welche biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen diesen charakterisieren, um eine Vielzahl der biologischen Assertionen zu transformieren und so ein abgeleitetes Wissensnetzwerk zu erzeugen und neues biologisches Wissen aus der Einheit zu schöpfen.

[0015] Die Erfindung stellt Systeme bereit, um eine biologische Wissensbank zusammenzustellen. Die Systeme beinhalten eine Datenbank biologischer Assertionen in elektronischer Form, die eine Vielzahl von Knoten umfasst, welche biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen diesen charakterisieren. Die Systeme beinhalten auch eine Anwendung zur Erfassung einer Teilmenge von Assertionen, die eine Reihe biologischer Kriterien erfüllt, die ein Benutzer spezifiziert hat, um ein ausgewähltes biologisches System zu definieren, aus der Datenbank. Die Systeme beinhalten außerdem noch einen Wissens-Assembler, der für die Kompilierung der erfassten Assertionen konfiguriert ist, um so eine biologische Wissensbank von Assertionen zu erzeugen, die für das ausgewählte biologische System potenziell relevant sind. Die Erfindung stellt auch Systeme bereit, um eine biologische Wissensbank einschließlich einer Datenbank biologischer Assertionen zusammenzustellen, die eine Vielzahl von Knoten haben, welche biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen diesen charakterisieren, sowie eine Anwendung zur Transformation einer Vielzahl von biologischen Assertionen, um ein abgeleitetes Wissensnetz zu erzeugen.

[0016] Die Erfindung stellt Rechner zur Zusammenstellung einer biologischen Wissensbank und zur Ermittlung von neuem biologischen Wissen bereit. Die Rechner umfassen Mittel zum Zugriff auf eine elektronische Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, welche biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen diesen charakterisieren, sowie eine Benutzeroberfläche für spezifische biologische Kriterien, die von diesem Rechner zum Aufbau einer Einheit verwendet werden, die ein ausgewähltes biologisches System darstellt. Die Rechner umfassen außerdem eine Computeranwendung zur Erfassung einer Teilmenge von Assertionen, die vom Benutzer spezifizierten biologischen Kriterien erfüllen, aus der Datenbank, und einen Wissens-Assembler, für die Kompilierung der erfass-

ten Assertionen konfiguriert ist, um so eine biologische Wissensbank von Assertionen zu erzeugen, die für das ausgewählte biologische System potenziell relevant sind. Die Erfindung stellt auch einen Herstellungsartikel mit einem computerlesbaren Programmträger bereit, auf dem computerlesbare Instruktionen enthalten sind, um die oben genannten Verfahren und Systeme anzuwenden.

[0017] In verschiedenen Ausgestaltungen beinhaltet die Erfindung methodische Schritte, Anwendungen und Rechner, um aus den erfassten Assertionen Schlüsse zu ziehen, damit logische Inkonsistenzen in der Wissensbank beseitigt werden können; um aus den erfassten Assertionen Schlüsse zu ziehen, damit dadurch neues biologisches Wissen generiert wird; um aus den erfassten Assertionen Schlüsse zu ziehen, um so die Assertionen darin zu augmentieren, indem man der Wissensbank zusätzliche Assertionen aus der Datenbank, die den Auswahlkriterien entsprechen, hinzufügt; oder um die Assertionen darin zu augmentieren, indem man zusätzliche Assertionen aus Datenquellen außerhalb der Datenbank zur Wissensbank hinzufügt.

[0018] In verschiedenen Ausgestaltungen beinhaltet die Erfindung methodische Schritte, Anwendungen und Rechner, um aus den erfassten Assertionen Schlüsse zu ziehen, um die Assertionen darin zu augmentieren, indem man zusätzliche Assertionen, die in der Einheit neu sind, zur Wissensbank hinzufügt; indem man eine Pfadanalyse für die Wissens-einheit durchführt, um einen oder mehrere Pfade zu erfassen, der sich auf experimentelle oder klinische Daten bezieht; indem man auf die erfassten Assertionen eine homologe Transformation anwendet; indem man mit den erfassten Assertionen eine logische Simulation durchführt oder indem man der Einheit zusätzliche Assertionen aus Datenquellen außerhalb der Datenbank hinzufügt.

[0019] In verschiedenen Ausgestaltungen beinhaltet die Erfindung methodische Schritte, Anwendungen und Rechner, um aus den biologischen Assertionen auf neue Assertionen zu schließen; um eine Teilmenge von Assertionen aus der Datenbank zu erfassen, die eine vom Benutzer spezifizierte Reihe von biologischen Kriterien zur Festlegung eines ausgewählten biologischen Systems erfüllt; um mathematische Operationen auf Reihen biologischer Assertionen anzuwenden, um dadurch neue Assertionen zu erzeugen; und um biologische Assertionen zusammenzufassen, um neue Assertionen zu erzeugen.

[0020] In verschiedenen Ausgestaltungen sind Knoten Enzyme, Kofaktoren, Enzymsubstrate, Enzyminhibitoren, DNS, RNS, Transkriptionsregulatoren, DNS-Aktivatoren, DNS-Repressoren, Signalmoleküle, Transmembran-Moleküle, Transportmoleküle, Fressmoleküle, Regulationsmoleküle, Hormone, Zy-

tokine, Chemokine, Antikörper, Strukturmoleküle, Metaboliten, Vitamine, Toxine, Nährstoffe, Mineralien, Agonisten, Antagonisten, Liganden, Rezeptoren oder Kombinationen daraus. In anderen Ausgestaltungen sind Knoten Protonen, Gasmoleküle, organische Moleküle, Aminosäuren, Peptide, Proteinbereiche, Proteine, Glykoproteine, Nukleotide, Oligonukleotide, Polysaccharide, Lipide, Glykolipide oder Kombinationen daraus. In weiteren Ausgestaltungen umfassen Knoten Zellen, Gewebe oder Organe oder Moleküle, die als Medikamente infrage kommen.

[0021] In verschiedenen Ausgestaltungen können die durch die Knoten und Assertionen repräsentierten biologischen Informationen experimentelle Daten, Wissen aus der Fachliteratur, Patientendaten, Daten aus klinischen Versuchen, Compliance-Daten, chemische Daten, medizinische Daten oder hypothetische Daten umfassen. In anderen Ausgestaltungen können die biologischen Informationen wahre Aussagen über ein Molekül, eine biologische Struktur, einen physiologischen Zustand, eine Eigenschaft, einen Phänotyp oder einen biologischen Prozess repräsentieren.

[0022] In verschiedenen Ausgestaltungen repräsentieren die biologischen Informationen ein Molekül, eine biologische Struktur, einen physiologischen Zustand, eine Eigenschaft, einen Phänotyp, einen biologischen Prozess, klinische Daten, medizinische Daten, Daten über Krankheiten oder Unverträglichkeiten. In verschiedenen Ausgestaltungen beinhaltet die biologische Information einen Deskriptor für Zustand, Ort, Menge oder Unterstruktur eines Moleküls, biologische Struktur, physiologischen Zustand, Eigenschaft, Phänotyp, biologischen Prozess, klinische Daten, medizinische Daten, Daten über Krankheiten oder Unverträglichkeiten.

[0023] In verschiedenen Ausgestaltungen beinhaltet das durch die Methode erzeugte neue biologische Wissen Vorhersagen über physiologisches Verhalten bei Menschen, z. B. aus der Analyse von an Tieren durchgeführten Experimenten wie Studien über Medikamentenwirksamkeit und/oder Medikamententoxizität oder über die Ermittlung von Biomarkern, die auf die Prognose, Diagnose, Empfänglichkeit für Medikamente, Medikamententoxizität sowie Schwere oder Stadium der Krankheit schließen lassen. In einigen Ausgestaltungen schließt die Methode das Vergleichen verschiedener Einheiten ein, in anderen Datenabbildung und in wiederum anderen die grafische Darstellung aller oder mehrerer Teile der Einheit, um so das Verständnis, die Extrapolation, die Interpolation und das Ziehen von Schlüssen zu erleichtern.

[0024] Die vorhergehenden und andere Funktionen und Vorteile der vorliegenden Erfindung sowie die Erfindung an sich werden aus der Beschreibung, den Zeichnungen und Ansprüchen, die noch folgen, bes-

ser verständlich.

Kurze Beschreibung der Zeichnungen

[0025] In allen Ansichten der Zeichnungen beziehen sich gleiche Bezugszeichen durchgehend auf dieselben Teile. Die Zeichnungen sind nicht unbedingt maßstabsgetreu, das Gewicht wurde stattdessen ganz allgemein auf die Darstellung der Grundlagen der Erfindung gelegt. In der folgenden Beschreibung werden verschiedene Ausgestaltungen mit Verweis auf die folgenden Zeichnungen beschrieben, in denen:

[0026] [Fig. 1](#) ein Übersichtsdiagramm ist, das eine erläuternde Ausgestaltung der Erfindung darstellt.

[0027] [Fig. 2A](#) ein Originalnetzwerk zeigt und [Fig. 2B](#) eine Teilmenge eines Netzwerkes in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0028] [Fig. 3](#) einen Wissenseinheitsgraphen in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0029] [Fig. 4](#) die Vereinigung von zwei Pfaden in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0030] [Fig. 5](#) einen Wissenseinheitsgraphen in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0031] [Fig. 6](#) einen Wissenseinheitsgraphen in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0032] [Fig. 7](#) ein transformiertes Netzwerk in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0033] [Fig. 8](#) die Darstellung einer zusammengefassten Stoffwechselreaktion in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0034] [Fig. 9](#) ein abgeleitetes Netzwerk in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0035] [Fig. 10](#) ein erläuterndes Beispiel von Datenabbildung in Übereinstimmung mit einer Ausgestaltung der Erfindung zeigt.

[0036] [Fig. 11](#) Inferenzpfade für übergeordnete Ursachen zeigt, beginnend mit einer Veränderung der Boten-RNS-Niveaus für ein spezielles Gen in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung.

[0037] [Fig. 12](#) ein Diagramm ist, das in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung die Ausbreitung von vorausgesagten Veränderungen in einer Vorwärtssimulation zeigt, die mit beobachteten Expressionsänderungen verglichen, werden.

[0038] [Fig. 13](#) ein Diagramm ist, das in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung durch eine Rückwärtssimulation aus neun Expressionsdatenpunkten generiert wurde, gefolgt von der Beschneidung des Graphen, um nur die Folgerungsketten zu zeigen, die die Primärhypothesen stützen.

[0039] [Fig. 14](#) in Übereinstimmung mit der vorliegenden Erfindung ein erläuterndes Beispiel einer Visualisierungstechnik zeigt, die auf einer Vorwärtssimulation basiert, die vorausgesagte Ergebnisse mit tatsächlichen Labordaten vergleicht.

[0040] [Fig. 15](#) einen Einheitsübersichtsgraphen in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0041] [Fig. 16](#) ein Graph ist, der in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung Simulationsergebnisse zeigt.

[0042] [Fig. 17](#) eine Darstellung von Zeitserienexpression und einer Abbildung proteometrischer Daten auf ein Segment eines bekannten metabolischen Pfadverlaufs in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung zeigt.

[0043] [Fig. 18](#) ein Diagramm zeigt, das in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung ein Mittel zur Zusammenfassung von Zeit, Dosis oder Daten anderer Datenreihen aus vielen Experimenten rund um ein bestimmtes Gen oder Protein anzeigt.

[0044] [Fig. 19](#) ein Tortendiagramm zeigt, die in Übereinstimmung mit einer erläuternden Ausgestaltung der Erfindung die Übereinstimmung einer Hypothese mit beobachteten Daten zusammenfasst.

[0045] [Fig. 20](#) in Übereinstimmung mit der Erfindung ein Beispiel für einen Algorithmus zur Verwendung bei der Bewertung eines biologischen Modells durch Vergleichen vorausgesagter Ergebnisse mit tatsächlichen Ergebnissen zeigt.

[0046] [Fig. 21](#) in Übereinstimmung mit der Erfindung ein Beispiel für einen Algorithmus zur Identifizierung eines Biomarkers zeigt.

Beschreibung

[0047] Um die gegenwärtige Erfindung umzuset-

zen, wird eine globale Wissensbank oder zentrale Datenbank so strukturiert, dass sie eine Vielzahl von Knoten und Deskriptoren umfasst, und diese Knoten und Deskriptoren kopiert oder transferiert werden können, ohne dass dabei interne Konsistenz oder biologischer Kontext verloren geht. Knoten sind Elemente biologischer Systeme, sowohl physisch als auch funktionell, und schließen zum Beispiel solche Dinge wie bestimmte Organe, Gewebe, Zellen, Organellen, Zellkompartimente, Membrane, Proteine, DNS, RNS, kleine Moleküle, Medikamente und Metaboliten ein. Die Deskriptoren sind Dateneinträge, welche die Knoten funktionell und/oder strukturell miteinander verbinden (z. B. Kasusrahmen, die „Verben“ sind, welche die Wechselbeziehung von Knoten bestimmen) und Dateneinträge, die zusätzliche Informationen mit einem oder beiden Knoten und ihren Wechselbeziehungen in Verbindung bringen (z. B. Aufzeichnen der Spezies oder des Organs, wo das Protein gefunden wird, Bestimmen der Zeitschrift, in der die Daten veröffentlicht wurden, Vermerk von tertiärer struktureller Information über das behandelte Protein, Vermerk, dass das Protein bei Patienten mit Hypertonie erhöht ist, usw.). Die globale Zentralwissensbank kann eine große Menge an Informationen enthalten, und tut das meist auch, die für die vorliegende Aufgabe nicht relevant sind, doch hat sie eine Struktur, die eine Erfassung von potenziell relevanten Assertionen erlaubt, die auf der Anwendung von Benutzer spezifizierten biologischen Kriterien basieren.

[0048] Knoten können, wobei die aufgeführten Beispiele nicht erschöpfend sind, biologische Moleküle sein, darunter Proteine, kleine Moleküle, Ionen, Gene, ESTs, DNS, Transkriptionsfaktoren, Metaboliten, Liganden, Transmembran-Proteine, Transportmoleküle, Fressmoleküle, Regulierungsmoleküle, Hormone, Zytokine, Chemokine, Histone, Antikörper, Strukturmoleküle, Metaboliten, Vitamine, Toxine, Nährstoffe, Mineralien, Agonisten, Antagonisten, Liganden oder Rezeptoren. Die Knoten können Arzneimittelsubstanzen, Verbindung, die als Kandidaten für Medikamente angesehen werden, Antisense-Moleküle, RNS, RNS-Interferenz, Haarnadel-RNS, Doppelstrang-RNS oder chemogenetische oder chemo-proteomische Untersuchungen sein. Chemisch betrachtet können die Knoten Protonen, Gasmoleküle, kleine organische Moleküle, Aminosäuren, Peptide, Proteindomänen, Proteine, Glykoproteine, Nukleotide, Oligonukleotide, Polysaccharide, Lipide oder Glykolipide sein. In Modellen höherer Ordnung können die Knoten Proteinkomplexe, Protein-Nukleotid-Komplexe wie Ribosomen, Zellkompartimente, Organellen oder Membrane sein. Strukturell betrachtet können sie verschiedene Nanostrukturen wie Filamente, interzelluläre Lipid-Doppellagen, Zellmembrane, Lipid Rafts, Zelladhäsionsmoleküle, Gewebeschränken und halbdurchlässige Membrane, Kollagenstrukturen, mineralisierte Strukturen oder Bindegewebe sein. In noch höheren Ordnungen sind die Knoten

Zellen, Gewebe, Organe oder andere anatomische Strukturen. So kann ein Modell des Immunsystems zum Beispiel Immunglobuline, Zytokine, verschiedene Leukozyten, Knochenmark, Thymusdrüse, Lymphknoten und Milz einschließen. Bei der Simulation von klinischen Versuchsreihen können die Knoten zum Beispiel Personen, deren klinische Prognose oder derzeitigen Symptome, Medikamente, Niveaus der Medikamentendosierung und klinische Endpunkte sein. Bei der Simulation von Epidemiologie können die Knoten zum Beispiel Personen, deren Symptome, physiologische oder gesundheitliche Charakteristika, Umwelteinflüsse, denen sie ausgesetzt sind, Substanzen, die sie einnehmen, sowie Krankheitsdiagnosen sein. Knoten können auch Ionen, physiologische Prozesse, Krankheiten, Krankheitsprozesse, Translokationen, Reaktionen, molekulare Komplexe, Zellkomponenten, Zellen, anatomische Teile, Gewebe, Zelllinien und Proteindomänen sein.

[0049] Deskriptoren können biologische Beziehungen zwischen Knoten darstellen und beinhalten nichtkovalente Bindung, Adhärenz, kovalente Modifikation, multimolekulare Interaktionen (Komplexe), Spaltung einer kovalenten Bindung, Konversion, Transport, Zustandsänderung, Katalyse, Aktivierung, Stimulierung, Agonismus, Antagonismus, Aufwärtsregulierung, Repression, Inhibition, Abwärtsregulierung, Expression, posttranskriptionale Modifikation, posttranslationale Modifikation, Internalisierung, Abbau, Steuerung, Regulierung, Chemoattraktion, Phosphorylation, Azetylierung, Dephosphorylation, Deazetylierung, Transport und Transformation, sind aber nicht auf diese beschränkt.

[0050] Eine bevorzugte Form von Deskriptoren zur Verwendung bei der Erfindung sind zum Beispiel Kausurahmen, die aus der Darstellungsstruktur erfasst werden, was eine unmittelbare und allgemeine Anwendung der Modelle auf eine Vielzahl von biowissenschaftlichen oder anderen Systemen erlaubt. Kausurahmen werden im Einzelnen in der anhängigen US-Patentanmeldung mit der Seriennummer 10/644,582 des selben Anmelders beschrieben, deren Offenlegung hierin durch Bezugnahme eingefügt ist. Deskriptoren können quantitative Funktionen wie Differenzialgleichungen umfassen, die mögliche quantitative Beziehungen zwischen Knotenpaaren darstellen, welche genutzt werden können, um das Netzwerkweiter zu verfeinern. Deskriptoren können auch qualitative Eigenschaften umfassen, die entweder nicht gemessen oder analytisch oder quantitativ nur schwer beschrieben werden können, oder aufgrund von ungenügender Kenntnis eines Systems im Allgemeinen oder der Eigenschaft an sich unmöglich beschrieben werden können.

[0051] Der Prozess der Sammlung von Wissen kann auf ungleichartige Systeme angewendet wer-

den, und das Ergebnis kann zu einer Sammeleinheit zusammengefasst werden, die ein Modell begründet. Darüber hinaus kann auf eine Wissensseinheit, die auf ungleichartigen Systemen aufgebaut ist, als geschlossenes Modell zugegriffen werden, indem auf die Fragmente des Modells dezentral zugegriffen wird. Ein Modell stellt eine Hypothese dar, die die Arbeitsweise des Systems erklärt, d. h. basierend auf einer Simulation ist es in der Lage, vorausgesagte Daten zu erzeugen, die zu den tatsächlichen Daten passen, die als Eignungskriterien dienen. Die Hypothese kann mittels weiterer Experimente getestet werden, sie kann mit anderen Modellen oder Netzen kombiniert werden, sie kann verfeinert, verifiziert, reproduziert, modifiziert, perfektioniert, korrigiert oder mit neuen Knoten und neuen Assertionen, die auf manuellen oder computergestützten Analysen neuer Daten beruhen, erweitert und produktiv als biologische Wissensbank genutzt werden Modelle von Teilen eines physiologischen Pfades oder Teilnetzwerks in einem Zellkompartiment, einer Zelle, einem Organismus, einer Population oder Ökologie können zu einem geschlossenen Modell verknüpft werden, indem ein oder mehrere Knoten in einem Modell mit einem oder mehreren Knoten in einem anderen Modell verbunden werden.

[0052] Jede wahre Aussage in einer Wissensbank oder einer Einheit wird hier als eine Assertion bezeichnet. Assertionen sind wahre Aussagen, die sich auf ein existierendes Objekt in einem System beziehen, auf eine wahre Aussage über ein Objekt in diesem System und einen Literalwert oder jede beliebige Kombination daraus. In verschiedenen Ausgestaltungen können Assertionen Wissen wie RNS, proteomisches, metabolisches oder klinisches Wissen aus Quellen wie medizinischen Fachpublikationen, Patientendaten, Daten aus klinischen Versuchen, Compliance-Daten, chemische Daten, medizinische Daten, hypothetische Daten oder Daten aus biologischen Datenbanken sein.

[0053] Der Aufbau einer Einheit beginnt, sobald ein Individuum durch Eingabe über eine Benutzeroberfläche biologische Kriterien spezifiziert, die so gestaltet sind, dass sie aus der zentralen Wissensbank alle Assertionen abrufen, die für das behandelte Thema als potenziell relevant erachtet werden. Die Beispiellklassen von Kriterien, die auf die zentrale Wissensbank angewendet werden, um die unaufbereitete Einheit zu erzeugen, umfassen Zuordnungen, spezielle Netzwerke (z. B. transkriptionelle Steuerung, metabolisch) und biologische Kontexte (z. B. Spezies, Gewebe, Entwicklungsstadium), sind aber nicht auf diese beschränkt. Zusätzliche Beispiellklassen von Kriterien beinhalten Assertionen, die auf Beziehungs-Deskriptoren basieren, Assertionen, die auf Textabgleich regulärer Begriffe basieren, Assertionen, die mithilfe von Algorithmen zur Vorwärtsverketzung berechnet wurden, Assertionen, die auf Homo-

logie-Berechnungen basieren, und alle Kombinationen dieser Kriterien, sind aber nicht auf diese beschränkt. Oft werden Schlüsselbegriffe oder Wortwurzeln verwendet, aber auch andere Kriterien sind wertvoll. Man kann z. B. Assertionen wählen, die auf verschiedenen strukturbezogenen Algorithmen beruhen, z. B. kann man Algorithmen zur Vorwärtsverketzung oder Rückwärtsverketzung anwenden (z. B. alle verknüpften Assertionen drei oder mehreren Schritten unterhalb von allen Serinkinasen in Mastzellen erfassen). Verschiedene logische Operationen wie „oder“, „und“ und „nicht“ können auf alle beliebigen Auswahlkriterien angewendet werden, um komplexe Auswahlen zu spezifizieren. Es ist die Unterschiedlichkeit der Kriterienreihen, die ausgearbeitet werden können, und die Tiefe der Assertionen in der globalen Wissensbank, die die Flexibilität ermöglichen, die der Erfindung eigen ist.

[0054] Assertionen, die in Übereinstimmung mit der Erfindung in der Form der Dateneinträge, die eine Reihe bestimmter Kriterien erfüllen, ausgewählt wurden, werden von der Wissensbank abgerufen und dann in einer Teilwissensbank oder Einheit, die aus einer Teilmenge von Knoten mit Wechselbeziehungen und Deskriptoren, die für das untersuchte System potenziell relevant sind, besteht, wieder zusammengesetzt. Durch diese Teilmengenbildung wird ein neues biologisches Modell erstellt. Dieses Modell umfasst typischerweise weit weniger Assertionen als die globale Wissensbank und dient als Ausgangspunkt auf dem Weg, eine nützlichere Einheit mit kleinerem Fokus zu erstellen. Danach wird es von automatischen Routinen in der Softwareanwendung, die es erstellt hat, und dadurch, dass die Person, die die Anwendung ausführt, Werkzeuge anwendet, transformiert und verfeinert. Es kann durch andere Informationen augmentiert und andere Informationen können darin eingegliedert werden, einschließlich, aber nicht nur Assertionen, die aus der Literatur stammen und von der Person, die die Daten einpflegt, als relevant für das biologische System erachtet werden.

[0055] Einheiten, die durch die vorliegende Erfindung erstellt wurden, sind für gewöhnlich besser als die globale Wissensbank oder die zentrale Datenbank, aus der sie abgeleitet wurden, da sie die tatsächliche Biologie typischerweise besser voraussagen und beschreiben. Diese Errungenschaft der Erfindung beruht auf der Anwendung von Logik während oder nach der Kompilierung der Rohdatenreihe, um die anfangs abgerufenen Daten zu augmentieren und die resultierende Struktur wie hier angemerkt zu verbessern und zweckmäßiger zu gestalten. Das kann automatisch bei der Erstellung der Einheit erfolgen, zum Beispiel durch Programme, die in Computersoftware eingebettet sind, oder durch die Anwendung von Softwareinstrumenten, welche die Person, die die Anwendung durchführt, auswählt und steuert.

[0056] Eine Einheit ist in vielerlei Hinsicht mit einer globalen Zentralwissensbank strukturell identisch, sie ist jedoch kleiner und viel stärker auf das betrachtete Thema oder Problem fokussiert, rechnerisch leichter lenkbar und entweder physisch oder virtuell isoliert, damit sie auf ein bestimmtes Projekt zugeschnitten ist, und erleichtert die Einhaltung von Auflagen bei beschränkter Nutzung oder Offenlegungsvorschriften, die möglicherweise durch eine Datenquelle auferlegt sind. Außerdem wird eine Einheit oft die Kennzeichen einer laufenden Arbeit aufweisen, die im Lauf der Anwendung geändert, verbessert, geprüft und korrigiert wird. Eine Einheit kann jederzeit oder bei jedem Schritt in einem berechenbaren Format gespeichert und der globalen Wissensbank wieder hinzugefügt werden.

[0057] Die Erstellung einer wertvollen Einheit beinhaltet daher einen Prozess der Teilmengenbildung oder Segmentierung, der auf eine globale Zentralwissensbank angewendet wird, gefolgt von Datentransformationen oder -manipulationen, um die erste erstellte Einheit zu verbessern, zu verfeinern und/oder zu augmentieren, um die Einheit zu vervollständigen und die Einheit an die Analyse anzupassen. Das wird durch die Umsetzung eines Prozesses wie der Anwendung von Logik auf die resultierende Datenbank, um sie mit der tatsächlichen Biologie abzustimmen, erreicht. So können die Kriterien nach allen Proteinen fragen, die in menschlichen Muskelzellen vorkommen, und die zentrale Datenbank schließt vielleicht Muskelzellenproteine mit ein, die bei Mäusen vorkommen und von denen einige nicht in menschlichen Muskelzellen vorkommen, dann werden diese Daten aus einer Einheit, die die Physiologie menschlicher Muskelzellen untersucht, entfernt. Eine Einheit kann durch das Einfügen neuer Knoten und Beziehungs-Deskriptoren, die aus der Wissensbank stammen und auf den weiter oben festgelegten Annahmen basieren (und vieler anderer logischer Annahmen, die möglich sind), augmentiert werden. Eine Einheit kann gefiltert werden, indem man Teilmengen von Daten, die auf anderen biologischen Kriterien basieren, ausschließt. Die Körnung des Systems kann so verfeinert oder vergrößert werden, wie es für die anstehende Analyse angemessen ist (was wesentlich ist für die Fähigkeit, gültige Extrapolationen zwischen Arten oder Generalisierungen innerhalb einer Art machen zu können, da Datenreihen in ihrer Körnigkeit variieren). Eine Einheit kann kompakter und relevanter gemacht werden, indem detailliertes Wissen zu stärker folgernden Assertionen zusammengefasst wird, die für eine Untersuchung mittels Datenanalysealgorithmen oder für den Einsatz von generischen Analyseinstrumenten wie Instrumenten zur Analyse von Häufungen besser geeignet sind.

[0058] Eine Einheit kann dem Wissenszuwachs folgend regelmäßig aktualisiert werden und die daraus entstehenden Einheiten können gespeichert werden,

um die Progression des Wissens auf diesem Gebiet zu zeigen. Eine Einheit kann auf verschiedene Arten augmentiert werden, zum Beispiel kann die Person, die die Daten einpflegt, neue Daten aus einer strukturierten oder unstrukturierten Datenbank oder Daten aus der Fachliteratur hinzufügen. Eine Einheit kann auch wieder in eine zentrale Datenbank zurückgeführt werden, damit neue Assertionen als Rohmaterial für die Erstellung einer anderen Einheit verwendet werden können.

[0059] Die zugrunde liegende Darstellung des Wissens einer zentralen Wissensbank ist konzipiert, um Wissen detailliert und ohne Verzerrung in Bezug auf die Nutzung des Wissens zu erfassen. Bei einem Netzwerk von solcher Komplexität kann das Schlussfolgern schwierig sein. Deshalb umfassen die Verfahren und Systeme der Erfindung einen flexiblen Rahmen, um das Wissen stufenweise zu bearbeiten und durch die Anwendung genau definierter Regeln und Prozeduren abgeleitete Einheiten zu erstellen. Diese abgeleiteten Einheiten sind so aufgebaut, dass sie aufeinander folgende Durchgänge des Schlussfolgerns in den Einheiten ermöglichen.

[0060] Einheiten können verwendet werden, um ein beliebiges biologisches System nachzubilden, egal wie es definiert ist und so detailliert wie gewünscht, und werden nur durch den Wissensstand in dem betreffenden Fachgebiet, dem Zugang zu Daten und (bei neuen Daten) der Zeit, die benötigt wird, um sie einzupflegen und zu importieren, begrenzt. In einer Ausgestaltung können Einheiten verwendet werden, um Modelle kontinuierlich oder in Abständen zu aktualisieren, sobald neue Daten zur Erfassung verfügbar sind, und um ein Mittel zum besseren Verständnis der Biologie bereitzustellen. In einer anderen Ausgestaltung können Einheiten zur Darstellung biologischer Systeme im Ganzen oder in Teilen in verschiedenen Formaten verwendet werden, damit Menschen sie betrachten und analysieren können.

[0061] Einheiten können auch verwendet werden, um auf verschiedene Arten Daten über biologische Systeme abzufragen, um aus neuem biologischen Wissen (z. B. Überlagerung verschiedener Einheiten, um Unterschiede zu erkennen) zu schöpfen. In verschiedenen Ausgestaltungen können Einheiten verwendet werden, um: (a) aus der Analyse von Experimenten mit Tieren physiologisches Verhalten bei Menschen vorauszusagen (z. B. Wirksamkeit und Toxizität von Medikamenten); (b) ideale Biomarker zu finden (leicht erkennbare oder leicht quantifizierbare Substanzen in Körperflüssigkeiten, um Voraussagen über das Vorhandensein einer Krankheit, ihre Prognose, ob der Patient auf Medikament X ansprechen wird, Schwere der Krankheit usw. zu machen); oder (c), um zu lernen, wie man Mitglieder einer Population segmentiert, um Ergebnisse zu verbessern und Komplikationen in klinischen Versuchen zu vermei-

den.

[0062] Einheiten können außerdem verwendet werden, um die Biologie zu untersuchen, indem man verschiedene Einheiten vergleicht (z. B. Mensch mit Maus, krankes Gewebe mit gesundem, die adipöse Physiologie unter mehreren unterschiedlichen Ernährungsbedingungen). Einheiten können verwendet werden, um die Biologie von Geweben zu verschiedenen Zeitpunkten während der Entwicklung, des Fortschreitens oder der Heilung einer Krankheit zu vergleichen, oder um die Wirkung verschiedener Störungen wie die Wirkungen von Medikamenten oder die Wirkung eines anderen Umwelteinflusses innerhalb eines beliebigen biologischen Systems zu bestimmen. Einheiten können verwendet werden, um Daten abzubilden (um z. B. durch den Import experimenteller Daten die Auswirkung von Störungen auf ein biologisches System auf eine oder mehrere Komponenten des Systems zu zeigen). In weiteren Ausgestaltungen können Einheiten verwendet werden, um logische Simulationen umzusetzen, um Datenreihen zu evaluieren, die in einer globalen zentralen Datenbank zur Zeit der Erstellung der ursprünglichen Einheit nicht enthalten sind (um z. B. eine Hypothese, die auf neuen experimentellen Daten basiert, nochmals zu testen), um Mutmaßungen über Pfadverläufe anzustellen und komplexe und subtile Kausalbeziehungen innerhalb eines biologischen Systems zu erkennen, und um eine Krankheitsursache zu erkennen, toxische biochemische Mechanismen zu verstehen und toxische Reaktionen vorherzusagen.

[0063] Neues Wissen kann ermittelt werden, indem die Einheiten zum Beispiel mit epistemischen Motoren verwendet werden. Epistemische Motoren werden im Einzelnen in der ebenfalls angemeldeten und gehaltenen US-Patentanmeldung mit der Seriennummer 10/717,224 beschrieben, deren Offenlegung hierin durch Verweis eingefügt ist. Epistemische Motoren sind programmierte Computer, die biologische Daten von tatsächlich oder gedanklich durchgeführten Experimenten zur Untersuchung eines biologischen Systems akzeptieren und sie zur Erzeugung eines Netzwerkmodells von Protein-Interaktionen, Gen-Interaktionen und Gen-Protein-Interaktionen verwenden, das mit den Daten und dem vorherigen Wissen über das System übereinstimmt, und dadurch die biologische Wirklichkeit dekonstruieren und überprüfbar Erklärungen (Modelle) der Arbeitsweise natürlicher Systeme vorschlagen. Die Motoren identifizieren neue Wechselbeziehungen zwischen biologischen Strukturen, zum Beispiel zwischen Biomolekülen, welche die Substanz des Lebens darstellen. Diese neuen Beziehungen erklären für sich allein genommen oder gemeinsam das Verhalten des Systems. Sie können zum Beispiel die beobachtete Wirkung einer Störung im System erklären, Faktoren identifizieren, die die Homöostase aufrechterhalten, Wirkungsweise und Nebenwirkungen von Medika-

menten erklären, epidemiologische und klinische Daten zweckmäßiger gestalten, Gründe für den Erfolg einer Art offenbaren, embryologische Prozesse entschleiern und die Mechanismen von Krankheiten erkennen. Die Programme enthüllen feine Muster in komplexen Datenreihen, die der menschliche Verstand ohne Hilfe nicht aufspüren kann. Der Output des epistemischen Motors ermöglicht es, das untersuchte System besser zu verstehen, Hypothesen aufzustellen, das untersuchte System in andere Systeme zu integrieren, komplexere und verständlichere Modelle zu bilden und neue Experimente vorzuschlagen, um die Gültigkeit von Hypothesen zu testen.

[0064] Die Funktionsweise der hier dargelegten Systeme und Verfahren kann als Software auf einem Computer für allgemeine Zwecke implementiert werden. In manchen Ausgestaltungen kann ein Computerprogramm in einer der höheren Computersprachen wie FORTRAN, PASCAL, C, C++, LISP, JAVA oder BASIC geschrieben werden. Darüber hinaus kann ein Computerprogramm als Script, Makro oder in anderen Funktionsweisen, die in handelsüblicher Software eingebettet sind wie EXCEL oder VISUAL BASIC, geschrieben werden. Zusätzlich könnte Software in einer Assemblersprache, die auf einen Mikroprozessor gerichtet ist, der auf einem Computer speicherresident ist, implementiert werden. Zum Beispiel könnte Software in der Assemblersprache Intel 80x86 implementiert werden, wenn sie so konfiguriert wird, dass sie auf einem IBM-PC oder einem PC-Klon läuft. Software kann in einen Herstellungsartikel eingebettet werden, einschließlich eines Speichermediums oder computerlesbaren Mediums wie einer Diskette, einer Festplatte, einer Bildplatte, eines Tonbands, eines PROM, eines EPROM oder einer CD-ROM, aber nicht nur auf diese beschränkt.

Zusammenziehung von Einheiten

[0065] Die Erfindung erlaubt die Erstellung von Wissensseinheiten mittels Erfassung aus einer globalen Zentraldatenbank und dem nachfolgenden Hinzufügen von neuem Wissen durch Einpflegen von Daten und andere Verfahren. In einem Beispiel wird neues Wissen einer globalen Zentraldatenbank in einem schrittweisen Prozess hinzugefügt, dessen Fokus auf der Anwendung liegt. Zuerst wird der globalen Zentraldatenbank allgemeines Wissen, das in der globalen Zentraldatenbank noch nicht vorhanden ist, hinzugefügt (z. B. zusätzliches Wissen über Krebs). Zweitens wird grundlegendes Wissen im Abfragebereich der beabsichtigten Anwendung (z. B. Prostatakrebs) aus der Fachliteratur gesammelt, dazu gehören unter anderem Lehrbücher, wissenschaftliche Schriften und Rezensionen von Artikeln. Drittens wird der spezielle Fokus des Projekts (z. B. Androgenunabhängigkeit bei Prostatakrebs) genutzt, um noch speziellere Informationsquellen auszuwählen. Dem folgt die Verwendung von experimentellen Daten, um

den nächsten Schritt beim Einpflegen von Daten und Sammeln von Wissen zu lenken. Zum Beispiel können experimentelle Daten zeigen, welche Gene und Proteine im fokussierten Bereich beteiligt sind. Durch das Einpflegen von Daten aus der Fachliteratur, die sich auf Gene und Proteine beziehen, kann eine Untereinheit erstellt werden, die sich auf das spezielle Interessengebiet konzentriert.

[0066] [Fig. 1](#) zeigt in Übereinstimmung mit der Erfindung eine erläuternde Übersicht über ein System. In diesem Diagramm wird das System **100** benutzt, um neues biologisches Wissen zu ermitteln. In Phase **110** wird eine globale Zentralwissensbank erstellt, indem Informationen (z. B. eingepflegte wissenschaftliche Daten aus der Fachliteratur, aus öffentlichen Datenbanken und Informationen aus dem Studium von Publikationen) in eine Computerdatenbank eingegeben werden. In Phase **120** wird eine Teilmenge der Informationen in der globalen Zentralwissensbank erfasst, um Wissensseinheiten zu generieren, die auf biologischen Inhalten basieren. Die Wissensseinheiten sind dann verfeinert. In Phase **130** werden experimentelle Daten (z. B. Daten, die sich auf Proteine, RNS, Stoffwechselaktivität, klinische Informationen usw. beziehen) verwendet, um das Einpflegen und Sammeln von Wissen anzuleiten. In Phase **140** können Wissensseinheiten in verschiedenen Anwendungen verwendet werden, darunter zum Beispiel auch Datenabbildung, fokussierter Zusammenbau durch Anwendung von Pfadsuche, grafische Ausgabe und logische Simulation.

[0067] Um abgeleitete Einheiten zu erstellen, können Algorithmen eingesetzt werden. In einigen Ausgestaltungen können Algorithmen als Computerprogramme zum Ausdruck kommen und dazu verwendet werden, abgeleitete Einheiten als Datenobjekte innerhalb eines Programmierungsrahmens zu erstellen. Ein Beispielalgorithmus führt eine oder mehrere Transformationen an den bestehenden Einheiten durch, um eine neue Einheit zu generieren. Transformationen können zum Beispiel durch irgendeine der folgenden Techniken erreicht werden: (a) indem man Assertionen aus bestehenden Einheiten auswählt und die ausgewählten Assertionen in eine neue, im Aufbau befindliche Einheit einfügt; (b) indem man Knoten und Assertionen aus bestehenden Einheiten zusammenfasst und die zusammengefassten Knoten und Assertionen in eine Einheit einfügt; (c) indem man mathematische Mengenlehreoperationen auf die Knoten und Assertionen aus bestehenden Einheiten anwendet und die Knoten und Assertionen, welche aus diesen Operationen resultieren, in eine Einheit einfügt; (d) indem man Zusammenstellungsoperationen auf bestehende Einheiten anwendet, um eine Einheit zu erstellen, die für weitere Transformationen verwendet wird, oder (e) indem man eine beliebige Kombination der oben genannten Techniken anwendet.

[0068] Die einfachste Form der Transformation einer Einheit ist die Erstellung einer Teilmenge der Einheit. So kann zum Beispiel eine Teilmenge einer Einheit eine Teilmenge der Knoten und Deskriptoren in der ursprünglichen Einheit enthalten. Eine Teilmenge ist im Wesentlichen das Ergebnis einer Anfrage, die Knoten und Assertionen anhand einer Reihe von Kriterien auswählt. Diese Kriterien können verfahrensmäßig definiert werden, z. B. kann die Auswahl das Resultat eines Algorithmus sein, der iterativ oder rekursiv Knoten und Deskriptoren, die die Einheit verkörpern, erkundet. Wie zum Beispiel in [Fig. 2A](#) gezeigt, wurde ein Originalnetzwerk **200** von Knoten **210** und Deskriptoren **220** wie in [Fig. 2B](#) gezeigt transformiert, um ein Teilmengennetzwerk **205** aus Knoten **210** und Deskriptoren **220** nur vom Typ „A BindungInput B“ zu erstellen und deswegen alle anderen auszuschließen. „A BindungInput B“ ist eine Assertion, die eine Klasse A von molekularen Bindungsprozessen mit einer Klasse B von molekularen Einheiten verbindet (z. B. Molekül oder Komplex).

[0069] In manchen Ausgestaltungen kann eine Einheit die Form einer oder mehrerer Datenbanktabellen mit Spalten und Zeilen annehmen. In diesen Ausgestaltungen kann die Transformation oder Teilmengebildung einer globalen Wissensbank zu einer Einheit zum Beispiel durch die Auswahl von Zeilen erreicht werden, die Assertionen aus einer Datenbanktabelle darstellen, die zu den Auswahlkriterien des Benutzers passen. Es sollte sich verstehen, dass eine Wissensbank oder eine Einheit in Form einer Datenbank nur eine Möglichkeit ist, in der Information in einem Computer dargestellt werden kann. Informationen könnten stattdessen auch als Vektor, als mehrdimensionale Anordnung von Daten, als verknüpfte Datenstruktur oder in Form vieler anderer geeigneter Strukturen oder Ausgestaltungen von Daten dargestellt sein.

[0070] Ein Aspekt einer Assertion ist eine Zuordnung. Eine Zuordnung stellt die Quelle der Assertion dar, zum Beispiel einen wissenschaftlichen Artikel, eine Zusammenfassung (z. B. Medline oder PubMed), ein Kapitel eines Buches, Tagungsberichte, eine persönliche Mitteilung oder ein internes Memo. Eine Einheit kann durch die Auswahl von Deskriptoren erstellt werden, deren Zuordnung bestimmte Kriterien erfüllen, z. B. Übereinstimmung nach Art der Zuordnungsquelle, Name der Zuordnungsquelle oder Datum der Zuordnungsquelle. Man könnte zum Beispiel alle Assertionen auswählen, deren Zuordnung ein Knoten ist, der einen im Jahr 2001 oder später veröffentlichten Zeitschriftenartikel repräsentiert.

[0071] Ein weiterer Aspekt einer Assertion ist ihr biologischer Kontext. So können Assertionen ausgewählt werden, die mit einem bestimmten biologischen Kontext verknüpft sind.

[0072] Biologischer Kontext bezieht sich beispielsweise auf Spezies, Gewebe, Körperteile, Zelllinien, Tumor, Krankheit, Probe, Virus, Organismus, Entwicklungsstadium oder jede beliebige Kombination daraus. Ein weiterer Aspekt einer Assertion ist ihre Vertrauenswürdigkeit, ein Maß des Vertrauens darauf, dass die Assertion wahrhaft repräsentative, tatsächliche Biologie widerspiegelt und reproduzierbar ist. Assertionen können auch nach ihrer Vertrauenswürdigkeit ausgewählt werden. Dabei wird ein Mindestschwellenwert gesetzt und alle Assertionen, die diesem Schwellenwert entsprechen oder ihn übersteigen, werden ausgewählt.

[0073] Teilmengen einer Wissensbank können auch mithilfe von Spezifizierungen, die ein komplexes Muster von Assertionen zwischen Knoten definieren, gebildet werden. Alle Reihen von Knoten und Assertionen, die den Kriterien des Musters entsprechen, bilden die Teilmenge. In einer Ausgestaltung kann ein Suchalgorithmus die Datenbank filtern, um eine Liste von biologischen Entitäten, die dem vorher festgelegten Muster entsprechen, zu generieren. Beispielsweise kann eine Struktursuche verwendet werden, um die Teilmenge aller Reaktionen zu generieren, die ein Produkt, das phosphoryliert wird, und einen Molekülkomplex als Katalysator haben. Diese Suche wird alle durch einen Molekülkomplex katalysierten phosphorylierten Reaktionen finden und durch ein einzelnes Protein katalysierte phosphorylierte Reaktionen auslassen.

[0074] In einer weiteren Ausgestaltung können Teilmengen mithilfe von Algorithmen zur Pfadsuche generiert werden, dazu gehören radiale Pfadsuche, Suche nach dem kürzesten Pfad und Suchen aller Pfade. Radiale Pfadsuche ist hilfreich, um herauszufinden, wie eine biologische Entität funktionell oder strukturell mit einer anderen biologischen Entität in Verbindung steht. Enthält zum Beispiel irgendeine gegebene Zelle eine mutierte Form von P53, könnte jemand daran interessiert sein, ihre Wirkung auf über- und untergeordnete Moleküle des mutierten Genprodukts aufzuklären. Ein Algorithmus, um diese Informationen zu finden, kann bei einem bestimmten Knoten beginnen und alle Knoten finden, die mit diesem Knoten durch eine vorher festgelegte Anzahl vom Knoten entfernter Schritte verbunden sind. Falls Richtcharakteristik von Bedeutung ist (wie z. B. bei Reaktionen), kann der Algorithmus so instruiert werden, dass er Verknüpfungen nur in die Richtung folgt, die von den Kriterien für die Pfadsuche vorgegeben wird. Die radiale Pfadsuche kann in mehreren Schritten durchgeführt werden. Eine radiale Pfadsuche in zwei Schritten wird beispielsweise beinhalten, dass von einem Knoten ausgehend begonnen wird, die unmittelbar mit ihm verknüpften Knoten zu suchen und danach die unmittelbar mit diesen Knoten verknüpften Knoten zu suchen. Dieser Prozess kann auf so viele Schritte wie nötig angewendet werden. Diese

Analyse kann verwendet werden, um die erwarteten Veränderungen zu bestimmen und vorauszusagen, wenn ein vorhandener Knoten gestört wird. Diese Analyse kann dem Benutzer gezeigt werden, um zu erläutern, wie eine Veränderung sich in der Wissensbank ausbreiten könnte und dadurch ihre tatsächliche Wirkung auf ein biologisches System festzustellen. [Fig. 3](#) zeigt ein Beispiel für die Progression einer radialen Pfadsuche in zwei Schritten, die von einem spezifizierten Knoten **300** ausgeht. Im ersten Schritt der Suche werden die verbundenen Knoten **310** gefunden. Im zweiten Schritt der Suche werden die verbundenen Knoten **320** gefunden. Das Ergebnis dieser radialen Pfadsuche ist die Kombination aller Knoten und Assertionen wie in [Fig. 3](#) gezeigt. 3. Eine Pfadsuche kann optional auch so konfiguriert werden, dass sie nur bestimmten Deskriptoren folgt, bestimmte Knoten, die vielleicht ubiquitär oder uninformativ sind, ignoriert oder aufhört, neue Knoten zu suchen, sobald sie auf bestimmte Knoten gestoßen ist.

[0075] In großen biologischen Netzen gibt es zwischen zwei Entitäten für gewöhnlich mehrere Pfade. Oftmals ist der kürzeste Pfad der nützlichste für eine Analyse. Ein Algorithmus zur Bestimmung des kürzesten Pfades in einem Netzwerk beginnt damit, eine radiale Breitensuche von jedem der beiden Startknoten durchzuführen. Sobald ein gemeinsamer Knoten gefunden ist, wird der Pfad als kürzester Pfad zwischen den Knoten bekannt gegeben. Um die Pfadverläufe zwischen mehreren Knoten zu bestimmen, kann der oben besprochene Algorithmus zum Auffinden des kürzesten Pfades angewendet werden, bis man alle Pfadverläufe zwischen den Knoten gefunden hat. Bei dieser Technik beginnt man von jedem einzelnen Startknoten aus eine radiale Pfadsuche. Dann werden die Pfade aufgezeichnet, denen man in jeder der radialen Suchen gefolgt ist. Das Ergebnis dieses Algorithmus ist die Vereinigung aller Pfade von den Startknoten zu den Zielknoten. Da dieser Ansatz dazu neigt, in Bezug auf die Anzahl von Pfaden und Knoten exponentiell zu wachsen, kann der Algorithmus beschränkt werden, damit er einer vorher festgelegten Anzahl von Schritten folgt. So wird eine Suche in drei Schritten nur all die Pfade generieren, die zwischen den gegebenen Ursprungsknoten bestehen, indem von jedem Knoten aus eine radiale Suche in drei Schritten durchgeführt wird. Das Ergebnis dieses Pfadverlaufsalgorithmus kann zum Beispiel als sortierte Liste von Pfadverläufen angezeigt werden, beginnend mit dem kürzesten oder längsten, oder als integrierter Graph.

[0076] Ein integrierter Graph wird erzeugt, indem im Falle einer radialen Pfadsuche alle genutzten Pfadverläufe bis zu einer bestimmten Länge zusammengefasst werden oder indem die Reihe von Pfaden zusammengefasst wird, die irgendeinen der Quellknoten mit irgendeinem der Zielknoten verknüpfen. Das erreicht man, indem man zwei Pfadverläufe zugleich

zusammenfasst, bis nur ein einzelner Graph entsteht, der alle Knoten und Assertionen enthält. Ein Beispiel für das Zusammenfassen von zwei Pfadverläufen beinhaltet, alle gemeinsamen Knoten und Assertionen, wie in [Fig. 4](#) gezeigt, in einem kombinierten Pfad zusammenzufassen. Da die Knoten A, B und D in diesem Diagramm sowohl zu Pfadverlauf **410** als auch zu Pfadverlauf **420** gehören, werden diese Knoten nur einmal im kombinierten Pfadverlauf **430** dargestellt. Knoten B taucht in Pfadverlauf **410** auf und Knoten E in Pfadverlauf **420**, und sie werden auch im kombinierten Pfadverlauf **430** dargestellt. [Fig. 5](#) zeigt das Ergebnis der Zusammenführung aller Pfadverläufe in einen einzigen Graphen, basierend auf einer Pfadsuche zwischen dem Startknoten „FXR“ (in der oberen linken Ecke des Diagramms) und einem Zielknoten „LDL“ (in der unteren rechten Ecke des Diagramms). Diese Art der Analyse erlaubt auch die Untersuchung der Implikationen von beobachteten Veränderungen in Genexpressionsstudien oder von Veränderungen bei Konzentrationen von Proteinen und Metaboliten. Anhand der Analyse wird gezeigt, in welchem Zusammenhang die veränderten Entitäten stehen, damit man die abhängigen Veränderungen erkennen und Veränderungen finden kann, die für das durchgeführte Experiment zentral sind.

[0077] Die Matrixmethode ist eine andere Möglichkeit, um die Veränderungen in einem Wissensgraphen zu untersuchen. Anhand einer Liste von Knoten von Interesse (z. B. statistisch signifikante, hoch modulierte RNS in einem Experiment) werden die Knoten in einer Matrix platziert, wobei jeder Knoten mit einem Eintrag in einer Spalte und einer Reihe platziert wird. Dann wird für jedes Knotenpaar der kürzeste Pfad generiert (redundante Paarungen werden ignoriert). Dann werden alle generierten Pfadverläufe wie oben erklärt zusammengeführt. Die Matrixmethode kann auch angewendet werden, indem man nicht nur einen Pfad für jede Zelle in der Matrix sucht, sondern indem man mehrere Pfadverläufe generiert. Das kann auf verschiedene Arten geschehen: (1) man generiert alle Pfade für jedes Paar; (2) man generiert die oberen „n“ Pfadverläufe beginnend mit dem kürzesten oder längsten; (3) man generiert alle oberen „n“ Pfadverläufe, die nicht länger sind als eine vorher festgelegte Anzahl von Schritten. Die Matrixmethode ist auch hilfreich, um zu bestimmen, wie eine Reihe biologischer Entitäten miteinander in Zusammenhang steht. [Fig. 6](#) zeigt das Ergebnis einer Matrixmethodenanalyse zwischen drei Knoten, „Acox1“, „LDL“ und „FXR“, nachdem alle kürzesten Pfade zwischen jedem Knotenpaar zusammengeführt wurden.

[0078] Ein abgeleitetes Netzwerk ist nicht auf Operationen beschränkt, die das Anfangsnetzwerk unterteilen, vereinfachen oder zusammenfassen. Die Ableitung kann eine Theorie über das Wissen verkörpern, die es ermöglicht, einen Schluss aus neuen

Fakten zu ziehen, die auf anderen Fakten beruhen. Ein Primärbeispiel dafür ist die Theorie, dass biologische Mechanismen erhalten werden und dieser Mechanismus von Gen- und Proteinsequenzen abhängt. Falls also ein Mechanismus einer Spezies bekannt ist, kann geschlossen werden, dass dieser Mechanismus in anderen Spezies existiert, wenn alle beteiligten Gene/Proteine in diesem Mechanismus die gleichen – homologe – Gegenstücke in der zweiten Spezies haben. Diese Technik wird verwendet, um Wissenseinheiten, die sich auf einen einzigen Organismustyp konzentrieren, zu augmentieren. So kann eine Einheit, die auf Humanbiologie ausgerichtet ist, durch Fakten aus der Biologie der Maus vergrößert werden, indem man festlegt, welche Fakten über die Maus die Kriterien für eine Homologie mit dem Menschen erfüllen, und dann in der Einheit die homologen menschlichen Fakten erstellt. Der Grad der Homologie wird durch Homologiepunktwerte bestimmt, die berechnet werden, indem man die Sequenzen der Gene oder Proteine vergleicht. Diese Punktwerte erlauben es, für einen bestimmten Zweck Grenzwerte für die Gemeinsamkeiten festzulegen – in manchen Ausgestaltungen können die Homologiekriterien weit gefasst werden, damit man Fakten, die in Zusammenhang mit anderen Organismen stehen, importieren kann. In anderen Ausgestaltungen können die Grenzwerte eng gefasst werden, damit nur Mechanismen erfasst werden, die auf den ähnlichsten Genen und Proteinen beruhen.

[0079] Eine einfache Beispiel für ein abgeleitetes Netzwerk ist ein Netzwerk aus kollabierenden Knoten, die nicht als eigenständige Begriffe unterschieden werden müssen. So unterscheidet die Darstellung den Akt des "Bindens" – ein Vorgang, bei dem Entitäten einen Komplex formen – von einem „Komplex“ – dem Ergebnis dieses Vorgangs des Bindens. Diese Unterscheidung ist in vielen Zusammenhängen irritierend – vor allem, wenn man ein Netzwerk in einem Graphen darstellt oder Proteine nach ihren Bindungsbeziehungen in Gruppen einteilt. [Fig. 7](#) zeigt ein Beispielnetzwerk, das aus kollabierenden Knoten gebildet wird. In diesem Diagramm wird die Bindung von A und B mit dem Knoten, der den Komplex von A und B darstellt, zusammengelegt, und der neue Knoten ersetzt in allen Fällen jeden der Ursprungsknoten.

[0080] Eine Einheit kann durch einen Zusammenfassungsprozess transformiert werden. Eine Zusammenfassung beginnt mit dem Prozess der Teilmengebildung, wobei Reihen von Knoten, die einer Spezifikation entsprechen, ausgewählt werden. Jede dieser Reihen kann durch eine neue Reihe von Knoten und Assertionen ersetzt werden, für gewöhnlich ein einfaches Muster wie eine einzelne Assertion zwischen zwei Knoten. [Fig. 8](#) zeigt ein Beispiel für eine Zusammenfassung von zwei Reaktionen, dargestellt als R1 und R2, die einen gemeinsamen Metaboliten

CoA gemeinsam haben. Die Assertionen in diesem Beispiel sind "R1 Reaktionspartner M" und „R2 Produkt M“. Die zusammengefasste Verbindung zwischen den Reaktionen R1 und R2 wird als die Assertion "R1 neue Beziehung R2" dargestellt. Eine komplexere Ableitung kann verwendet werden, um ein Netzwerk einfacher Verknüpfungen zu schaffen und eine einfache Verknüpfung an die Stelle eines komplexen Musters aus Beziehungen zwischen zwei Verknüpfungen zu setzen. Das kann als Prozess des "Zusammenfassens" betrachtet werden. In diesem Beispiel wird zwischen zwei Genen eine Beziehung geschaffen, wenn sie die folgenden Kriterien erfüllen: (1) jedes Gen hat ein Produkt, das als Enzym in einer Reaktion agiert und (2) bei einer Reaktion, bei der ein Genprodukt als Katalysator agiert, entsteht ein Produkt, das seinerseits als Reaktionspartner in einer anderen Reaktion agiert, bei der das Produkt des anderen Gens als Katalysator agiert. Das resultierende abgeleitete Netzwerk, wie in [Fig. 9](#) gezeigt, verknüpft die Gene G1 und G2, die in einer abgeleiteten Einheit aneinander angrenzen. Für diese abgeleitete Einheit gibt es viele Anwendungen. Enthält sie zum Beispiel Vermerke zu Genexpressionsdaten, kann ein Algorithmus Gruppen von koregulierten Genen finden, die in der abgeleiteten Einheit benachbart sind. Das entspricht dem Finden von Reaktionspfadverläufen, die gemeinsam reguliert werden.

[0081] Transformationen der Einheit können durch mathematische Mengenlehreoperationen durchgeführt werden. Zu diesen Operationen gehört zum Beispiel das Bilden der Schnittmenge, der Differenzmenge und der Vereinigungsmenge. Mithilfe von Mengenlehreoperationen kann man Einheiten vergleichen. Alle Mengenlehreoperationen setzen voraus, dass es zwei bestehende Einheiten gibt. Wendet man die Schnittmengenoperation für jede Assertion in einer ersten Einheit an, wird dieselbe Assertion überprüft, um zu erkennen, ob sie nicht in einer zweiten Einheit vorkommt. Taucht die Assertion in einer zweiten Einheit auf, wird sie zur Schnittmengeneinheit hinzugefügt. Knoten, die in irgendeiner Assertion in der Schnittmengeneinheit erwähnt werden, werden ebenfalls aus der ersten Einheit ausgewählt und zur Schnittmengeneinheit hinzugefügt. Wendet man die Differenzoperation für jede Assertion in einer ersten Einheit an, wird dieselbe Assertion überprüft, um zu erkennen, ob sie nicht in einer zweiten Einheit vorkommt. Taucht die Assertion in der zweiten Einheit nicht auf, wird sie zur Differenzeinheit hinzugefügt. Knoten, die in irgendeiner Assertion in der Differenzeinheit erwähnt werden, werden auch aus den ersten oder zweiten Einheiten ausgewählt und zur Differenzeinheit hinzugefügt. Mit einer Vereinigungsmengenoperation wird eine Vereinigungsmengeneinheit erzeugt. Alle Assertionen in einer ersten Einheit werden zur Vereinigungsmengeneinheit hinzugefügt. Denn jede Assertion in einer zweiten Einheit wird nun der Vereinigungsmengeneinheit hinzugefügt, falls sie

in der Vereinigungsmengeneinheit nicht existiert. Knoten, die in der Vereinigungsmengeneinheit erwähnt werden, werden aus den ersten oder zweiten Einheiten ebenfalls ausgewählt. Die Vereinigungsmengenoperation ist eine weitere Möglichkeit, um darzulegen, dass zwei oder mehr Einheiten zusammengeführt werden können.

[0082] Ein Beispiel für eine Vergleichstechnik in Übereinstimmung mit der Erfindung ist die Messung der Fortentwicklung einer Wissensseinheit über einen bestimmten Zeitraum. Das kann erreicht werden, indem man eine Sequenz von Einheiten nimmt, die im Lauf der Zeit erstellt werden, und den Unterschied zwischen jedem Paar in der Sequenz bestimmt. Außerdem können in Übereinstimmung mit der Erfindung zwei oder mehr Einheiten miteinander verglichen werden. Verwendet man zum Beispiel eine Schnittmenge aus zwei Einheiten, wobei die beiden Einheiten nicht identisch sind, wird die Schnittmenge der Assertionen in den beiden Einheiten bestimmt. Die Schnittmenge enthält die Assertionen, die in beiden Einheiten vorkommen. Verwendet man zum Beispiel die Differenzmenge von zwei Einheiten, wobei die beiden Einheiten nicht identisch sind, wird die Differenzmenge der Assertionen in den beiden Einheiten bestimmt. Die Differenzmenge enthält die Assertionen, die in einer der Einheiten vorkommen, nicht aber in der anderen. Vergleiche zwischen Einheiten können hilfreich sein, um Gemeinsamkeiten und Unterschiede biologischer Systeme zu erklären. So könnte beispielsweise eine Einheit ein normales System repräsentieren und eine andere Einheit ein krankes System. Für einen Wissenschaftler wäre es sehr aufschlussreich, die Gemeinsamkeiten und Unterschiede von zwei Systemen zu bestimmen.

[0083] Instrumente, um aus Einheiten Wissen zu schöpfen Die vorliegende Erfindung kann die Analyse einer Einheit zur Ermittlung von neuem biologischen Wissen beinhalten. Analyse schließt unter anderem eine algorithmische Analyse ein, die von Computern oder Personen durchgeführt werden kann. Algorithmen, die Pfadsuche, homologes Schlussfolgern oder simulationsbasiertes Schlussfolgern beinhalten, können neue Assertionen ableiten, die danach der Einheit zugeführt werden können, um diese zu augmentieren. Einheiten können durch Homologietransformation auch verfeinert und augmentiert werden, unter der Annahme, dass (1) die Physik und die fundamentalen biochemischen Eigenschaften und Wechselbeziehungen von Substanzen unter typischen biologischen Bedingungen konstant bleiben und (2) homologe Strukturen identische oder analoge Funktion haben. Enthält eine globale Wissensbank beispielsweise Daten darüber, dass Komplex C produziert wird, wenn Molekül A mit Molekül B in einer Nervenzelle kollidiert, dann kann angenommen werden, dass $A + B = C$ auch dann gilt, wenn A und B in einer Leberzelle kollidieren. Wenn die Leberzellenas-

sertion der globalen Wissensbank Knoten A und Knoten B enthält, aber nicht den Deskriptor, der besagt, dass sie zusammen den Komplex C bilden, kann diese Information während der Kompilierung einer Lebereinheit in dieses importiert werden. Ganze Kaskaden biologischer Aktivitäten können mithilfe solcher Logik in eine Einheit importiert werden. Ähnlich verhält es sich, wenn eine globale Wissensbank die Information enthält, dass ein Mäuseprotein M sich mit dem Mäuserzeptor R verbindet, um die Reparatur der Nierentubuli zu starten, und humanbiologische Assertionen in der Wissensbank einen Knoten homolog zu Mäuseprotein M und einen weiteren homologen Knoten zu Rezeptor R enthalten, dann können die Wechselwirkung und möglichen nachgeordneten Ereignisse aus der Mauseinheit in eine Einheit importiert werden, die auf ein humanbiologisches System ausgerichtet ist. Darüber hinaus kann eine Einheit mit einer anderen, die mit anderen Kriterien generiert wurde, kombiniert und danach die logischen Inkonsistenzen und Redundanzen entfernt werden, damit ein noch besseres, vollständigeres oder enger fokussiertes biologisches Modell erstellt werden kann.

Techniken zur grafischen Ausgabe

[0084] Eine Wissensseinheit kann visuell als ein Graph aus Knoten dargestellt werden, die durch Verknüpfungen, die biologische Beziehungen zwischen und unter Knoten repräsentieren, verbunden sind. Diese Graphen können von einem Wissenschaftler untersucht werden, damit er das biologische System besser versteht und ihm die Ermittlung von neuem biologischen Wissen über Systeme, die mit den Biowissenschaften in Beziehung stehen, erleichtert wird. Die Nutzung von Einheiten, um biologisch relevante Einsichten darüber zu erhalten, wie ein System sich verhält, kann in der Arzneimittelforschung und -entwicklung sowie bei der Entwicklung einer Vielzahl von Therapien äußerst wertvoll sein. Mit den hier beschriebenen Techniken kann man durch die Nutzung von Einheiten, die durch Verfahren und Systeme dieser Erfindung geschaffen wurden, zu biologisch relevanten Einsichten gelangen. Visualisierungstechniken können auch genutzt werden, um Wissen und verwandte Daten darzustellen, um so das Verständnis des Nutzers und das Erkennen von Beziehungen zwischen Entitäten, die als Muster und Häufungen vorkommen können, zu verbessern.

[0085] Nachdem man mit irgendeiner der oben genannten Techniken Graphen generiert hat, möchte man vielleicht eine bessere Vorstellung vom biologischen Kontext der Pfadverläufe bekommen. Das kann erreicht werden, indem man bei jedem Knoten im Eingabe-Graphen beginnt und von jedem Knoten aus eine Radialsuche mit n-Schritten durchführt. Dieser Schritt „dehnt“ die Knoten und die Größe des Graphen „aus“. Durch farbliche Codierung der Knoten, um eine Modulation anzuzeigen (wie durch experi-

mentelle Daten festgelegt) kann man die relevanten Veränderungen erkennen, die funktionell oder strukturell proximal zum relevanten Graphen stehen, oder anders gesagt, zum biologischen Kontext.

[0086] Experimentelle Daten können in einer Einheit kartografiert werden, indem Messungen aus Experimenten an die Assertionen in der Einheit, die die gemessenen Quantitäten darstellen, angepasst werden. In diesem Kontext bedeutet kartografieren, visuell wiedererkennbare Indizien wie Farbe auf eine Karte von Pfadverläufen zu legen, um so kenntlich zu machen, welche Knoten bei einem Prozess beteiligt sind. Das kann zum Beispiel gemacht werden, indem Knoten, die Genexpressionsprozesse darstellen, an die durch Mikrobereiche oder andere Techniken wie RT-PCR gemessenen Niveaus von Boten-RNA angepasst werden. Knoten, die eine Fülle von Proteinen repräsentieren, können an Daten aus proteometrischen Messungen angepasst werden. Knoten, die eine Fülle von chemischen Substanzen repräsentieren, können an Daten aus metabolomischen Messungen angepasst werden. Einmal kartografiert können die Daten verarbeitet werden, um einfachere Qualitätsmerkmale des Knotens zu erstellen, die den Einsatz von Algorithmen für die Darstellung oder Analyse vereinfachen. Zum Beispiel können Faltungsveränderungsdaten basierend auf Benutzer gesteuerten Grenzwerten zusammengefasst werden und Knoten mit Vermerken zu zusätzlichen Qualitätsmerkmalen wie „nach oben“ oder „nach unten“ versehen werden, damit die Anwendung einfacher Algorithmen zur Darstellung oder Analyse möglich ist. Faltungsveränderungsdaten können auch durch Schattierung gezeigt werden, wie in [Fig. 10](#) dargestellt, wo die Schattierung jedes ausgedrückten Gens im Diagramm (z. B. Mat1a, Mat2b, Pemt, Ahcy11, Bhmt, Bhmt2, Mfmt, Shmt und Mthdf) in einem Experiment seine Faltungsveränderung anzeigt (d. h., je dunkler die Schattierung, desto größer der Faltungsveränderung).

[0087] Logische Simulation kann in Übereinstimmung mit der Erfindung ebenfalls eingesetzt werden. Als logische Simulation bezeichnet man eine Klasse von Operationen, die auf eine Einheit angewendet werden, in der beobachtete oder hypothetische Änderungen auf einen oder mehrere Knoten in der Einheit gültig werden und in der die Implikationen dieser Veränderungen basierend auf den kausalen Beziehungen, die als Assertionen in der Einheit dargestellt werden, im Netzwerk verbreitet werden. Eine logische Simulation kann entweder vorwärts gerichtet sein, dann werden die Auswirkungen der Änderungen von den Ursprungspunkten der Änderung ausgehend in Richtung nach unten verbreitet und wirken dort, oder sie kann rückwärts gerichtet sein, dann werden die Auswirkungen der Änderungen von den Ursprungspunkten der Änderung ausgehend in Richtung nach oben verbreitet und wirken dort. In jedem

Fall ist ein Ergebnis einer logischen Simulation ein neues abgeleitetes Netzwerk, das aus den Konten und Assertionen besteht, die bei der Verbreitung der Ursache oder der Wirkung beteiligt waren. Dieses abgeleitete Netzwerk enthält eine Hypothese über das zu untersuchende System.

[0088] Im Fall einer auf beobachteten Änderungen in RNS-Expressionsniveaus basierenden Rückwärts-simulation zeigt [Fig. 11](#) Inferenzpfade, um übergeordnete Ursachen zu finden, die mit einer beobachteten Änderung in den Boten-RNS-Niveaus bei einem bestimmten Gen beginnen. Eine spezielle Ursachenkette könnte wie folgt aussehen: Eine Phosphorylierung eines Transkriptionsfaktors durch eine Kinase, bei der die Kinase die Aktivität des Transkriptionsfaktors verändert, kann umgekehrt Änderungen in der Expression von Genen induzieren, die durch diesen Transkriptionsfaktor gesteuert werden. Dieses Diagramm gibt eine "Pseudocode"-Beschreibung der Inferenzen, die dann durchgeführt werden, um mögliche Ursachen für jede der beobachteten RNS-Veränderungen zu finden. Die zu untersuchenden Assertionstypen sind nicht auf die in diesem Diagramm dargestellten beschränkt. Jede Assertion in der Einheit, die eine kausale biologische Verknüpfung darstellt, kann in diese Art der Analyse eingeschlossen werden. Umgekehrt kann dann jede der möglichen Ursachen untersucht werden, um deren jeweilige Ursache zu finden. Der Prozess kann für jede gewünschte Anzahl von Schritten wiederholt werden, wobei Knoten in der Einheit mit Vermerken zu ihrer möglichen Rolle in der Kausalität der beobachteten Veränderungen versehen werden.

[0089] Das resultierende abgeleitete Netzwerk enthält eine Hypothese über die möglichen Ursachen der beobachteten Daten. Dazu kann es außerdem, abhängig von den Verfahren zur Verbreitung der Kausalität, als Hypothese über die am stärksten mitverantwortlichen und konsistenten möglichen Ursachen der beobachteten Daten werden, d. h. einer Reihe von möglichen Ursachen, die nach objektiven Kriterien angeordnet sind. Diese Technik ist nicht auf RNS-Expressionsdaten beschränkt, sondern kann auf jede Reihe von Veränderungen angewendet werden, die als Darstellungssystem ausgedrückt werden kann, darunter, aber nicht ausschließlich, proteometrische Daten, metabolomische Daten, posttranslationale Modifikationsdaten oder sogar Reaktionszeitdaten.

[0090] [Fig. 12](#) ist ein manuell zusammengestelltes Diagramm, das die Ausbreitung von vorausgesagten Veränderungen **1210** in einer Vorwärtssimulation zeigt, die mit beobachteten Expressionsänderungen **1220** verglichen werden. Dieses Diagramm zeigt die Ausbreitung von vorausgesagten Proteinveränderungen **1210** basierend auf einer Zunahme der Menge einer Verbindung **1230** über einen bekannten Pfad-

verlauf. In diesem Diagramm stellen Kugeln **1240** Proteine dar. Paare von angrenzenden Kugeln **1250** stellen Proteinkomplexe dar. Dünne Pfeile mit T-förmigen Spitzen **1260** zeigen Hemmungen oder kausale Abnahmen an. Dünne Pfeile mit Spitzen **1270** zeigen eine Aktivierung oder eine kausale Zunahme an. Genexpressionsbeziehungen werden durch die Pfeile **1280** angezeigt. Das Diagramm soll die Art und Weise verdeutlichen, auf die Veränderungen, die durch eine Hypothese vorausgesagt wurden, mit beobachteten Daten verglichen werden können.

[0091] [Fig. 13](#) ist ein Diagramm, das durch Rückwärtssimulation aus neun beobachteten Expressionsdatenpunkten **1320** erstellt wurde, nachfolgend wurde der Graph gekappt, um nur die Verbindungen **1330** zu zeigen, die die Primärhypothesen stützen. Jeder Knoten **1310** in dieser Abbildung repräsentiert entweder ein Gen, ein Protein oder eine Verbindung. Neun dieser Knoten **1320** repräsentieren Veränderungen in der Expression von Genen als Reaktion auf eine Ernährung mit mehrfach ungesättigten Fettsäuren. Der Rest des Diagramms wird durch die Durchsuchung der Einheit nach möglichen Knoten **1310** erzeugt, die würden sie verändert, eine oder mehrere der neun beobachteten Veränderungen **1320** erklären könnten, und durch die Entfernung von Knoten **1310** und Verbindungen **1330**, damit nur die besten Erklärungen gezeigt werden.

[0092] Abgeleitete Netzwerke können als Datenobjekte innerhalb eines Programmierungsrahmens für allgemeine Zwecke wie einer Skriptsprache erstellt werden. Spezielle Operatoren wie die oben beschriebenen Verfahren zur Suche nach Pfadverläufen oder logischen Simulation ermöglichen es, diese Datenobjekte zu speichern, zu laden und auf sie zuzugreifen. Außerdem können die Objekte mithilfe der Standardfunktionen des Programmierungsrahmens bearbeitet werden. Weil sowohl die Eingabe als auch die Ausgabe dieser Operationen die abgeleiteten Netzwerke umfasst, können in größeren Verfahren, in Verfahren, die biologisch signifikante Inferenzen umfassen, und in Verfahren, die Theorien und Techniken für automatische Verarbeitung biologischer Datenreihen und biologischen Wissens umfassen, mehrere Verarbeitungsschritte kombiniert werden. Durch unterschiedliche Kriterien können vielfältige abgeleitete Netzwerke erstellt und danach verglichen, zusammengeführt oder auf andere Weise bearbeitet werden. Eine Vielzahl von Hypothesen, wie sie in diesen Netzwerken enthalten sind, kann evaluiert, verglichen und eingestuft werden.

[0093] Ein Beispiel für eine Methode, die aus hier genannten Techniken besteht, würde wie folgt aussehen: (1) Laden Sie eine Reihe von Expressionsdaten über Faltungsveränderung in die Einheit; (2) führen Sie eine auf den Faltungsveränderungsdaten basierende logische Rückwärtssimulation durch; (3) unter-

suchen Sie das resultierende abgeleitete Netzwerk und wählen Sie die am häufigsten betroffenen Knoten – diejenigen, welche die als am wahrscheinlichsten eingestuft möglichen Ursachen der beobachteten Daten sind; (4) kehren Sie für diese Reihe von Knoten zurück zur Ursprungseinheit und führen Sie einen Algorithmus zur Pfadsuche aus, um das abgeleitete Netzwerk zu finden, das der kleinste Graph ist, der die Knoten verbindet; und (5) geben Sie das resultierende abgeleitete Netzwerk als Graph aus. Verfahren wie dieses Beispiel können als Funktionen im Programmierungsrahmen enthalten sein und benannt und erneut verwendet werden.

[0094] [Fig. 14](#) zeigt eine Visualisierungstechnik, die einen Aspekt der vorliegenden Erfindung umfasst, der auf einer Vorwärtssimulation basiert, die vorausgesagte Ergebnisse mit tatsächlichen Labordaten vergleicht. Dieses Diagramm zeigt die direkten untergeordneten Effekte einer Störung. Die Spalte ganz rechts zeigt das erwartete Ergebnis einer Störung in dem System. Jeder vorausgesagte Wert wird mit den tatsächlichen Werten verglichen, um zu bestimmen, wie genau die Voraussagen die Labordaten erklären. Eine Korrelation zwischen dem vorausgesagten Ergebnis und der tatsächlichen Wirkung jeder Behandlung kann berechnet werden. In [Fig. 14](#) zeigen die mit horizontalen Linien markierten Zellen eine signifikante Zunahme, die mit vertikalen Linien markierten Zellen eine signifikante Abnahme, die dunkel gezeichneten Zellen zeigen keine Veränderung und die nicht dunkel gezeichneten Zellen sind nicht signifikant. Störungen können unter anderem die Zunahme oder Abnahme der Konzentration eines Transkriptionsfaktors, eines kleinen Moleküls oder eines biochemischen Katalysator einschließen.

[0095] [Fig. 15](#) zeigt einen Einheitsübersichtsgraphen, der die Verknüpfbarkeit der zugrunde liegenden Einheit veranschaulicht, aus der er generiert wurde. Er kann einem Biologen einen schnellen visuellen Überblick geben über die Anzahl der Assertionen, die Verteilung verschiedener Arten von Assertionen in der Einheit und die Dichte oder den Grad, mit der die zugrunde liegende Einheit verknüpft ist. Die visuelle Übersicht kann verwendet werden, um zu bestimmen, ob die zugrunde liegende Einheit auf einem bestimmten Gebiet über ausreichendes Wissen, über genug verschiedene Assertionsarten oder über eine ausreichende Dichte von Assertionen verfügt. Zwei Diagramme, die zwei verschiedene Einheiten repräsentieren, können nebeneinander verglichen werden, um festzustellen, ob eine Einheit mehr Wissen enthält als die andere. Eine Vergleichsart wäre es, zwei Diagramme zu vergleichen, die dieselbe Wissensbank zu zwei verschiedenen Zeitpunkten repräsentieren, um das Wachsen des Wissens optisch zu untersuchen. Die Mechanik zum Erstellen des Diagramms in [Fig. 15](#) ist wie folgt: Alle Knoten und Assertionen in der Einheit werden in ein Diagramm um-

gewandelt, indem ein Algorithmus für den Aufbau des Graphen angewendet wird, um ein zweidimensionales Diagramm der Einheit zu erstellen. Das resultierende einfarbige Diagramm zeigt die Skala des Wissens, das in der Einheit enthalten ist, und kann als Ausgangspunkt für andere Visualisierungen verwendet werden. Der Einheitsübersichtsgraph kann verbessert werden, indem man die Assertionen, die einen bestimmten Beziehungs-Deskriptor enthalten, mit einer speziellen Farbe hervorhebt.

[0096] Eine Variation des Einheitsübersichtsgraphen ist es, einen Graphen zu generieren, der Simulationsergebnisse zeigt, wie in [Fig. 16](#) dargestellt. Dieses Diagramm kann erstellt werden, indem man mit einem einfarbigen Einheitsübersichtsgraphen beginnt. Die Ergebnisse einer Simulation werden dann über dieses Diagramm gelegt. Inferenzkausalketten können markiert werden, indem Knoten gemäß ihres Implikationsgrades mit Vermerken versehen werden. So können zum Beispiel alle Knoten, die impliziert sind und von der die Hypothese voraussagt, dass sie abnehmen werden, mit einem Vermerk versehen werden, indem die Knoten rot gefärbt werden oder indem das Symbol des Knotens durch ein anderes Symbol wie einem abwärts zeigenden Pfeil ersetzt wird. Andere Knotenzustände können durch entsprechende Wahl der Farben oder Symbole angezeigt werden. Das Erscheinungsbild der Assertionen zwischen Knoten kann auch verändert werden, um ihre kausale Signifikanz zu markieren. [Fig. 16](#) zeigt die Ergebnisse einer Rückwärtssimulation dunkelgrau markiert, wohingegen der Rest der Einheit hellgrau markiert ist. Die Ausgabe als Graph kann einem Biologen helfen, das Ausmaß der Wirkungen einer gegebenen Störung auf das System zu bestimmen.

[0097] [Fig. 17](#) zeigt eine Visualisierung einer Zeitreihenexpression und proteometrischen Daten, abgebildet auf ein Segment eines bekannten metabolischen Pfadverlaufs. In einigen Ausgestaltungen können Hintergrundfarben Quantität und Richtung von Veränderung relativ zu Kontrollen anzeigen. Jede gefärbte Zelle korrespondiert mit einem bestimmten Protein und zeigt entweder die Veränderungen des Expressionsniveaus des korrespondierenden Gens oder die Veränderungen in seiner beobachteten Proteinhäufigkeit. Jede mit einem Zeitpunkt beschriftete Spalte kann Datenwerte für ein bestimmtes Experiment in der Zeitreihe anzeigen. Diese Darstellungsmethode soll die Veränderungen in der Modulation eines Pfadverlaufs während einer Reihe von Experimenten zeigen, in diesem Fall einen Behandlungszeitverlauf. In [Fig. 17](#) wird Schattierung verwendet, um die Expressionsniveaus im Zeitverlauf zu zeigen (d. h. je dunkler die Schattierung, desto größer die Genexpression).

[0098] [Fig. 18](#) zeigt ein Diagramm, das ein Mittel zum Zusammenfassen von Zeit, Dosis oder anderen

Reihendaten aus vielen Experimenten für ein bestimmtes Gen oder Protein anzeigt. Ein Hauptpunkt ist, dass jeder horizontale Block eine spezielle Art der Messung anzeigt, die dem Gen oder Protein zugeordnet werden kann. In diesem Beispiel ist das Protein Anx7 (Mus musculus) mit fünf Arten der Messung verbunden – zwei sind proteometrische Messungen per 2D-Gel, drei sind Untersuchungsreihendaten im Mikrobereich, die Genexpressionsmessdaten ergeben. In diesem Fall werden die Daten als Faltungsveränderungen versus Steuerungen ausgedrückt, aber in anderen Fällen kann es wünschenswert sein, absolute Werte grafisch darzustellen. Für jeden Messtyp werden acht Faltungsveränderungen als Histogrammbalken dargestellt. Generell kann jede Anzahl von Datenpunkten bis zu einer praktischen Grenze, die von der Auflösung des Darstellungsmediums abhängt, auf diese Art dargestellt werden. Die Balken können mit Farbe gekennzeichnet werden – zum Beispiel mit Rot, um Abwärtsveränderungen zu zeigen, und mit Grün, um Aufwärtsveränderungen zu zeigen – um für den Nutzer, der möglicherweise Hunderte dieser Datenreihen durchsuchen muss, wenn er einen Datensatz überprüft, die allgemeine Tendenz jeder Messreihe besser zu verdeutlichen. Auch die Hintergrundfarben jedes Balken können die Signifikanz der Daten zeigen. So sind die Expressionsdaten in dem Experiment eigentlich der Durchschnitt aus multiplen Replikationen jedes Experiments, daher kann jedem Datenpunkt eine statistische Signifikanzmessung zugeordnet werden. In einer Ausgestaltung kann ein blauer Hintergrund die signifikantesten Daten anzeigen, p -Wert $< 0,01$, wohingegen ein purpurroter Hintergrund einen p -Wert $< 0,05$ anzeigen kann. Ein gelber Hintergrund könnte außerdem irgendeinen höheren p -Wert anzeigen. Diese Technik ermöglicht dem Nutzer, die Details der Daten leicht zu erkennen, Details, die in abstrakteren Darstellungen wie einem Netzwerkgraphen vielleicht unterdrückt worden wären, wo Knoten einfach eingefärbt sind, um „nach oben“ oder „nach unten“ anzuzeigen, diese Bestimmungen aber aus multiplen Datenpunkten stammen.

[0099] [Fig. 19](#) zeigt ein Tortendiagramm, das die Übereinstimmung der Veränderungen zusammenfasst, die durch eine Hypothese über die Veränderungen, die in einem großen Datensatz beobachtet wurden, vorausgesagt wurde. Der Datensatz in diesem Beispiel besteht aus Expressionsänderungen aufgrund einer Behandlung von Leberzellen mit Fenofibrat. Die Hypothese ist, dass die Veränderungen auf eine Zunahme der Aktivität des Transkriptionsfaktors PPARA zurückzuführen sind. Das Tortendiagramm in [Fig. 19](#) zeigt die folgenden fünf Kategorien: (1) korrekte Voraussagen (17%), die durch die Daten bestätigt werden; (2) gegenteilige Voraussagen (1%), die durch die Daten widerlegt werden; (3) Voraussagen (27%), die bei den Daten nicht beobachtet wurden; (4) Datenbeobachtungen (26%), für die keine über-

einstimmenden Voraussagen vorliegen; und (5) widersprechende Voraussagen (3%), denen keine Netto-Änderung in den Daten zugeschrieben werden kann.

BEISPIEL 1

Validationsalgorithmus für biologische Modelle

[0100] Ein Beispiel für einen Algorithmus, der zur Validation eines biologischen Modells durch Vergleich von vorausgesagten mit tatsächlichen Ergebnissen verwendet wird, wird weiter unten und im Pseudocode in [Fig. 20](#) beschrieben. Dieser Algorithmus nimmt an, dass eine Wissensbank existiert, die ein biologisches System mit auf die Wissensbank abgebildeten Daten aus Experimenten zur Genexpression repräsentiert.

[0101] Die vorausgesagten Ergebnisse können in zwei Stufen bestimmt werden. Zuerst wird eine Rückwärtssimulation, wie sie hier beschrieben ist, an einer Wissensbank durchgeführt, um potenzielle Ursachen der Genexpressionsänderungen zu bestimmen. Die Rückwärtssimulation erzeugt eine Liste von Genen und einen Punktwert für jedes Gen. Der Punktwert für jeden Knoten basiert auf den „Voten“, die er während der Rückwärtssimulation erhalten hat. Beim Beginn der Rückwärtssimulation werden Knoten, die die signifikant heraufregulierte Gene repräsentieren, positive Voten zugeordnet, während solche, die signifikant herunterreguliert sind, negative Voten bekommen. Während der Simulation werden Voten gemäß einer Reihe von Regeln, die den in der Wissensbank dargestellten Kausalbeziehungen entsprechen, von Knoten zu Knoten kopiert. Am Ende der Simulation wird der Punktwert für jeden Knoten als ein Satz aus drei Ziffern berechnet: die Summe der positiven Voten, die Summe der negativen Voten und eine Gesamtsumme, die die Summe der positiven und negativen Voten ist. An diesem Punkt kann der Satz von Knoten, die potenzielle Ursachen („die Gründe“) darstellen, für den nächsten Schritt verwendet und basierend auf dem Punktwert eines jeden Knotens ausgewählt werden, oder die Reihe potenzieller Ursachen kann manuell bestimmt werden. Auf der zweiten Stufe werden die Voten für alle Knoten auf Null gesetzt und mit der ausgewählten Reihe von Ursachen wird wie hier beschrieben eine Vorwärtssimulation durchgeführt. Mit den Voten wird auf dieselbe Weise verfahren, außer dass sie von Ursachen auf potenzielle Wirkungen ausgebreitet werden. Am Ende der Vorwärtssimulation werden Knoten überprüft, die die Expression von Genen repräsentieren. Solche mit einem positiven Gesamtpunktwert sind diejenigen, von der die Vorwärtssimulation voraussagt, dass sie nach oben reguliert werden, und solche mit einem negativen Gesamtpunktwert sind diejenigen, die laut Voraussage nach unten reguliert werden. Die Ergebnisse der Vorwärtssimulation stellen die Gesamtheit der

vorausgesagten Ergebnisse dar.

[0102] Die tatsächlichen Resultate werden in zwei Kategorien eingeteilt, die auf den Genexpressionsdaten basieren. Eine Liste enthält nach oben regulierte Gene und die andere Liste enthält nach unten regulierte Gene. Die in diesen Listen enthaltenen Gene können durch verschiedene statistische Verfahren generiert werden, wobei absolute Größenordnung der Veränderung (z. B. Signalniveau), relative Größenordnung der Veränderung (z. B. Faltungswerte), statistische Signifikanz usw. mit berücksichtigt werden. Die Gene können aber auch manuell ausgewählt werden.

[0103] Nachdem die vorausgesagten und tatsächlichen Ergebnisse generiert worden sind, werden die Ergebnisse für jedes Gen in den folgenden drei Fällen tabellarisch angeordnet. Im ersten Fall soll das Gen laut Voraussage nach oben reguliert werden. Wenn das Gen sich in der aktuellen Liste der nach oben regulierter Gene befindet, wird der „Zähler für korrekte Voraussagen“ erhöht. Wenn das Gen sich aber in der aktuellen Liste der nach unten regulierter Gene befindet, wird der „Zähler für gegenteilige Voraussagen“ erhöht. Wenn das Gen sich aber in keiner der Listen für tatsächliche Genexpressionsänderungen befindet, wird der „Zähler für nicht beobachtete Voraussagen“ erhöht. Im zweiten Fall soll das Gen laut Voraussage nach unten reguliert werden. Wenn das Gen sich in der Liste der tatsächlich nach oben regulierter Gene befindet, wird der „Zähler für gegenteilige Voraussagen“ erhöht. Wenn das Gen sich aber in der Liste der tatsächlich nach unten regulierter Gene befindet, wird der „Zähler für korrekte Voraussagen“ erhöht. Wenn das Gen sich aber in keiner der Listen für tatsächliche Genexpressionsänderungen befindet, wird der „Zähler für nicht beobachtete Voraussagen“ erhöht. Im dritten Fall existiert keine Voraussage für das Gen und der „Zähler für keine Netto-Veränderung“ wird erhöht.

[0104] Für jedes Gen, das sich in einer der Listen der Gene befindet, die tatsächlich nach oben oder unten reguliert werden, für das es aber keine Voraussage gibt, wird der „Zähler für nicht vorausgesagte Beobachtungen“ erhöht. Die fünf Zähler werden dann wie folgt ausgegeben: (1) „Zähler für korrekte Voraussagen“, (2) „Zähler für gegenteilige Voraussagen“, (3) „Zähler für nicht beobachtete Voraussagen“, (4) „Zähler für nicht vorausgesagte Beobachtungen“ und (5) „Zähler für keine Netto-Veränderung“. Diese Zähler können zum Beispiel in einem Histogramm oder eines Tortendiagramms, wie in [Fig. 19](#) gezeigt, bildlich dargestellt werden. Solche bildlichen Darstellungen bieten dem Wissenschaftler ein intuitives Mittel, um zu bestimmen, inwieweit die aufgestellte Hypothese mit den beobachteten Daten übereinstimmt.

BEISPIEL 2

Algorithmus zur Identifizierung von Biomarkern

[0105] Ein Beispiel für einen Algorithmus zur Identifizierung von Biomarkern in Übereinstimmung mit der Erfindung wird weiter unten und im Pseudocode in [Fig. 21](#) beschrieben. Generell befasst sich der Algorithmus mit Daten, die ein infrage kommendes Protein charakterisieren, und bewertet es, indem er eine Reihe von Faktoren berücksichtigt, durch die das Protein zu einem geeigneten Biomarker würde. Der Algorithmus bringt Maße aus einer Reihe von Quellen zusammen, ordnet ihnen einen numerischen Wert zu und bündelt sie, um dann einen Gesamtpunktwert auszugeben, der zur Bewertung eines Proteins herangezogen werden kann. Insbesondere die Proteine, die den höchsten absoluten Punktwert haben, weisen die höchste Anzahl von Ähnlichkeiten mit einem idealen Biomarker auf. Die in diesem Beispiel verwendeten Faktoren sind Genexpressionsänderungen durch ein Medikament, vorhandenes Wissen über die Natur des Genprodukts und Nähe zu einem bekannten Biomarker. Der Algorithmus wurde auf Datensätze angewendet, die aus einem Experiment stammen, in dem Genexpressionsänderungen als Reaktion auf ein Medikament über drei Zelllinien von unterschiedlicher Empfänglichkeit für dieses Medikament gemessen wurden.

[0106] Der erste Schritt des Biomarker-Algorithmus ist es, ausgehend von einer Liste bekannter sezernierter Proteine eine Pfadverlaufssuche auszuführen. Bei jedem Suchschritt werden Knoten mit dem Mindestabstand zu einem Quellknoten gekennzeichnet, d. h. der Anzahl von Schritten, die sie von einem sezernierten Protein entfernt sind. Der zweite Schritt ist, an der Liste der Proteine in der Einheit eine Iteration vorzunehmen. Für jedes Protein auf der Liste wird wie folgt eine Liste von Maßen berechnet: Berechnung von Richtungskoeffizient und Faltung, Biomarker- und Sekretionspunktwert, Entfernung von einem sezernierten Protein (wurde im ersten Schritt berechnet). Diese Maße werden in eine Zeile in eine Ausgabedatei geschrieben. Faltungsberechnungen beziehen sich auf Daten, die als Faltungsveränderungen gegenüber Steuerungen ausgedrückt werden, und können auf verschiedene Weise berechnet werden, zum Beispiel (1) Krankheit gegenüber Normalzustand; (2) Behandlung mit Medikamenten gegenüber keine Medikamentengabe; (3) Widerstandsfähigkeit gegenüber Suszeptibilität. Der Richtungskoeffizient ist ein Maß für die Veränderungsrate einer Reihe von Datenpunkten. Eine Datenreihe kann zum Beispiel zu verschiedenen Zeitpunkten oder bei unterschiedlichen Dosierungsniveaus gemessen werden. Ein Verfahren, um den Richtungskoeffizienten einer Reihe zu bestimmen, ist die Anwendung einer linearen Regression, die in einer geraden Linie resultiert, die am besten zu der Datenreihe passt.

[0107] Punktwerte für den Richtungskoeffizienten werden gemessen, indem man die Genexpressionsmessungen für jede Untersuchung, die mit dem Protein korrespondiert, über drei Zelltypen betrachtet. Untersuchungen, die einer Kreuzbindung unterliegen, werden ignoriert. Die restlichen Werte werden mit einem Referenzniveau verglichen, wobei ein Wert von 2 zugewiesen wird, wenn der Richtungskoeffizient dieses überschreitet, ein Wert von 1, wenn er nur die halbe Höhe des Referenzniveaus erreicht, oder 0, falls der Richtungskoeffizient weniger als die halbe Höhe des Referenzniveaus erreicht. Für negative Richtungskoeffizienten werden negative Werte zugewiesen. Über die Zelllinien wird, je nachdem welche benutzt wird, nach drei Mustern gesucht und die Untersuchungspunktwerte werden berechnet. Für ein Dosis abhängiges Muster werden die Werte über die Zelltypen addiert. Für ein Resistenzmuster wird der Wert für die resistente Zelllinie mit 2 multipliziert und davon die

[0108] Summe der Werte der beiden sensiblen Zelllinien subtrahiert. Für ein Wirksamkeitsmuster wird der Wert der sensibelsten Zelllinie verdoppelt, der Wert der partiell sensiblen Zelllinie hinzuaddiert und davon der Wert für die resistente Zelllinie subtrahiert. Die Punktwerte aller Untersuchungen werden verglichen, und falls bei einem Paar ein Vorzeichenkonflikt auftritt, wird ein Gesamtpunktzahl von 0 gemeldet, um einen Konflikt anzuzeigen. In allen anderen Fällen wird der höchste oder negativste Punktwert gemeldet. Berechnungen für die Faltungswerte werden auf dieselbe Weise durchgeführt.

[0109] Für die Bewertung von Biomarkern wird eine Punktzahl von 2 aufgezeichnet, wenn das Protein ein bekannter Biomarker ist, oder eine 1, falls es keiner ist. In gleicher Weise wird für sezernierte Proteine ein Punktwert von 2 aufgezeichnet, wenn es ein (maßlich) sezerniertes Protein ist, ansonsten wird ein Punktwert von 1 aufgezeichnet.

[0110] Die Ausgabedatei wird mithilfe eines Algorithmus sortiert, der aus den Werten der Maße einen Gesamtpunktwert berechnet. Im aktuellen Beispiel wird nur der Faltungspunktwert verwendet. Proteine, die die höchsten absoluten Werte haben (d. h. solche, die ganz oben oder ganz unten auf der sortierten Liste stehen), werden für die weitere Evaluation, ob sie gute Kandidaten für Biomarker wären, ausgewählt.

[0111] Die Hauptkomponenten des Punktwerts des Algorithmus basieren auf Genexpressionsdaten. Für jede Locus-ID gibt es Werte für multiple Untersuchungsreihen, die verarbeitet werden, um Werte über Richtungskoeffizienten und Faltungsveränderung zu liefern. Die Maße für jede Locus-ID werden berechnet, indem die Daten aus den Untersuchungen gebündelt werden, während man nach Zeichen sucht,

die sich widersprechen (sich widersprechende Vorzeichen hätten den Punktwert 0 zur Folge). Der Algorithmus kann Dosisdependenz, Sensibilität, Resistenz und Wirksamkeit des Medikaments überprüfen, und das Punktwertmaß berechnet sich für jede dieser Eigenschaften anders. Sucht man zum Beispiel nach einem Resistenzmuster, würde der Richtungskoeffizient positiv bewertet, falls die beiden resistenten Zelllinien gleich wären und die sensible Zelllinie andersartig wäre, wohingegen man für die Dosis-Wirkungs-Beziehung nach einer parallelen Veränderung über alle Zelllinien suchen würde. Der oben näher erläuterte Algorithmus liefert eine Liste, die dann nach Spalten sortiert wird, und die Gene, die ganz oben (Faltung) stehen, werden als gute Biomarker eingeschätzt.

[0112] Obwohl die Erfindung besonders unter Bezugnahme auf spezielle Ausgestaltungen vorgestellt und beschrieben worden ist, sollte jedoch klar sein, dass von denen, die in dieser Kunst besonders geschickt sind, vielfältige Veränderungen in Form und Detail gemacht werden können, ohne dass dadurch vom Geist und Umfang der Erfindung, wie sie durch die Ansprüche im Anhang dargelegt sind, abgewichen wird. Der Umfang der Erfindung wird deshalb durch die im Anhang dargelegten Ansprüche angezeigt, und es ist beabsichtigt, hier alle Änderungen, die innerhalb von Bedeutung und Umfang der Bedeutungsgleichheit der Ansprüche auftreten, ebenfalls einzuschließen.

ZUSAMMENFASSUNG

[0113] Offenbart sind Verfahren, Systeme und Vorrichtungen zum Aufbauen von Zusammenstellungen biologischen Wissens, die eine biologische Wissensbank bilden, und zum Auswählen und Übertragen von Biowissenschaft betreffenden Daten und Informationen in biologische Modelle, um die elektronische Verarbeitung und elektronisches Schlussfolgern auf biologische Informationen zu vereinfachen. Eine Teilmenge von Daten wird von einer umfassenden Wissensbank oder einem Wissensspeicher extrahiert, um eine spezialisiertere untergeordnete Wissensbank oder -einheit zu erzeugen, die speziell für die vorliegenden Zwecke ausgebildet ist. Mittels der Erfindung generierte Zusammenstellungen erlauben die Selektion und rationale Organisation von scheinbar verschiedenen Daten in einem Modell eines beliebig ausgewählten biologischen Systems, das durch beliebig gewünschte biologische Kriterien definiert ist. Aus diesen Zusammenstellungen kann auf einfache Weise geschöpft und mit großer Leistungsfähigkeit und Effizienz logisch geschlussfolgert werden.

Patentansprüche

1. Verfahren, um neues biologisches Wissen zu

generieren, das die folgenden Schritte umfasst:

(a) die Bereitstellung einer Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, die biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen den Knoten charakterisieren;

(b) das Gewinnen einer Teilmenge von Assertionen aus der Datenbank, die eine Reihe von biologischen Kriterien erfüllen, die ein Nutzer so spezifiziert hat, dass sie ein ausgewähltes biologisches System definieren;

(c) die Kompilierung der erfassten Daten, um eine Einheit zu erzeugen, die eine biologische Wissensbank von Assertionen bildet, die potenziell relevant für das genannte ausgesuchte biologische System ist; und

(d) die Transformierung der genannten Einheit, um neues biologisches Wissen über das genannte biologische System zu generieren.

2. Verfahren nach Anspruch 1, worin Schritt (d) die Anwendung des Schlussfolgerns auf genannte erfasste Assertionen zur Generierung von neuem biologischen Wissen umfasst.

3. Verfahren nach Anspruch 1, das den zusätzlichen Schritt der Anwendung des Schlussfolgerns auf genannte erfasste Assertionen zur Beseitigung logischer Widersprüche in genannter Einheit umfasst.

4. Verfahren nach Anspruch 1, das den zusätzlichen Schritt der Anwendung des Schlussfolgerns auf genannte erfasste Assertionen zur Augmentierung der Assertionen darin umfasst, indem der genannten Einheit zusätzliche Assertionen aus der genannten Datenbank hinzugefügt werden.

5. Verfahren nach Anspruch 1, das den zusätzlichen Schritt der Anwendung des Schlussfolgerns auf genannte erfasste Assertionen zur Augmentierung der Assertionen darin umfasst, indem der genannten Einheit aus der genannten Datenbank zusätzliche Assertionen, die neu sind, hinzugefügt werden.

6. Verfahren nach Anspruch 1, das den zusätzlichen Schritt der Anwendung der Pfadverlaufsanalyse auf genannte Wissensseinheit zur Gewinnung eines Pfadverlaufs oder mehrerer Pfadverläufe umfasst.

7. Verfahren nach Anspruch 1, das den zusätzlichen Schritt der Anwendung der homologen Transformation auf genannte erfasste Assertionen umfasst.

8. Verfahren nach Anspruch 1, das den zusätzlichen Schritt der Anwendung logischer Simulation auf genannte erfasste Assertionen umfasst.

9. Verfahren nach Anspruch 1, das den zusätzlichen Schritt des Hinzufügens zusätzlicher Assertio-

nen aus Datenquellen außerhalb der genannten Datenbank umfasst.

10. Verfahren nach Anspruch 1, worin genannte Knoten Enzyme, Kofaktoren, Enzymsubstrate, Enzyminhibitoren, DNS, RNS, Transkriptionsregulatoren, DNS-Aktivatoren, DNS-Repressoren, Signalmoleküle, Transmembranmoleküle, Transportmoleküle, Fressmoleküle, Regulationsmoleküle, Hormone, Zytokine, Chemokine, Antikörper, Strukturmoleküle, Metaboliten, Vitamine, Toxine, Nährstoffe, Mineralien, Agonisten, Antagonisten, Liganden, Rezeptoren oder Kombinationen daraus einschließen.

11. Verfahren nach Anspruch 1, worin die genannten Knoten Protonen, Gasmoleküle, organische Moleküle, Aminosäuren, Peptide, Proteindomänen, Proteine, Glykoproteine, Nukleotide, Oligonukleotide, Polysaccharide, Lipide, Glykolipide oder Kombinationen daraus einschließen.

12. Verfahren nach Anspruch 1, worin die genannten Knoten Zellen, Gewebe oder Organe einschließen.

13. Verfahren nach Anspruch 1, worin die genannten Knoten Moleküle, die mögliche Anwärter für ein Medikament sind, einschließen.

14. Verfahren nach Anspruch 1, worin die genannten biologischen Assertionen Informationen umfassen, die experimentelle Daten, Wissen aus der Fachliteratur, Patientendaten, Daten aus klinischen Versuchen, Compliance-Daten, chemische Daten, medizinische Daten oder hypothetische Daten darstellen.

15. Verfahren nach Anspruch 1, worin die genannten biologischen Assertionen Informationen umfassen, die ein Molekül, eine biologische Struktur, einen physiologischen Zustand, eine Eigenschaft, einen Phänotyp oder einen biologischen Prozess darstellen.

16. Herstellungsartikel mit einem computerlesbaren Programmträger, auf dem computerlesbare Instruktionen enthalten sind, um das Verfahren nach Anspruch 1 auszuführen.

17. Verfahren, um neues biologisches Wissen zu generieren, das die folgenden Schritte umfasst:

(a) Bereitstellung einer Datenbank, die eine Vielzahl von Knoten umfasst, die biologische Elemente repräsentieren und Kasusrahmen, die Wechselbeziehungen zwischen den Knoten repräsentieren;

(b) Gewinnen einer Teilmenge von Kasusrahmenstrukturen aus der Datenbank, die eine Reihe biologischer Kriterien erfüllen, die ein Nutzer so spezifiziert hat, dass sie ein ausgewähltes biologisches System definieren;

(c) Kompilierung der erfassten Kasusrahmenstrukturen, um eine Einheit zu erzeugen, die eine biologische Wissensbank von Assertionen bildet, die potenziell relevant für das genannte ausgesuchte biologische System sind.

18. System, um eine neue biologische Wissensbank zusammenzustellen, enthalten:

(a) eine Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, die biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen den Knoten charakterisieren;

(b) eine Anwendung, um eine Teilmenge von Assertionen aus der Datenbank zu erfassen, die eine Reihe von biologischen Kriterien erfüllen, die ein Nutzer so spezifiziert hat, dass sie ein ausgewähltes biologisches System definieren; und

(c) einen Wissensassembler, der konfiguriert ist, um die erfassten Daten zu kompilieren, um so eine Einheit zu erzeugen, die eine biologische Wissensbank von Assertionen bildet, die potenziell relevant für das genannte ausgewählte biologische System ist.

19. System nach Anspruch 18, das außerdem eine Anwendung umfasst, die es ermöglicht, aus den genannten erfassten Assertionen Schlüsse zu ziehen, um in genannter Einheit logische Widersprüche zu beseitigen.

20. System nach Anspruch 18, das außerdem eine Anwendung umfasst, die es ermöglicht, aus genannten erfassten Assertionen Schlüsse zu ziehen, um neues biologisches Wissen zu generieren.

21. System nach Anspruch 18, das außerdem eine Anwendung umfasst, die es ermöglicht, aus genannten erfassten Assertionen Schlüsse zu ziehen, um die Assertionen darin zu augmentieren, indem der genannten Einheit zusätzliche Assertionen aus der genannten Datenbank, die genannte biologische Kriterien erfüllen, hinzugefügt werden.

22. System nach Anspruch 18, das außerdem eine Anwendung umfasst, die es ermöglicht, aus genannten erfassten Assertionen Schlüsse zu ziehen, um die Assertionen darin zu augmentieren, indem der genannten Wissensbank zusätzliche Assertionen hinzugefügt werden, die in der genannten Einheit neu sind.

23. System nach Anspruch 18, das außerdem eine Anwendung umfasst, um an der genannten Wissensbank eine Pfadverlaufsanalyse durchführen, um einen Pfadverlauf oder mehrere Pfadverläufe zu erfassen.

24. System nach Anspruch 18, das außerdem eine Anwendung umfasst, um an den genannten erfassten Assertionen eine homologe Transformation

durchzuführen.

25. System nach Anspruch 18, das außerdem eine Anwendung umfasst, um an den genannten erfassten Assertionen eine logische Simulation durchzuführen.

26. System nach Anspruch 18, das außerdem eine Anwendung umfasst, um der genannten Einheit zusätzliche Assertionen aus Datenquellen außerhalb der genannten Datenbank hinzuzufügen.

27. System nach Anspruch 18, worin genannte Knoten Enzyme, Kofaktoren, Enzymsubstrate, Enzyminhibitoren, DNS, RNS, Transkriptionsregulatoren, DNS-Aktivatoren, DNS-Repressoren, Signalmoleküle, Transmembranmoleküle, Transportmoleküle, Fressmoleküle, Regulationsmoleküle, Hormone, Zytokine, Chemokine, Antikörper, Strukturmoleküle, Metaboliten, Vitamine, Toxine, Nährstoffe, Mineralien, Agonisten, Antagonisten, Liganden, Rezeptoren oder Kombinationen daraus einschließen.

28. System nach Anspruch 18, worin die genannten Knoten Protonen, Gasmoleküle, organische Moleküle, Aminosäuren, Peptide, Proteindomänen, Proteine, Glykoproteine, Nukleotide, Oligonukleotide, Polysaccharide, Lipide, Glykolipide oder Kombinationen daraus einschließen.

29. System nach Anspruch 18, worin genannte Knoten Zellen, Gewebe oder Organe einschließen.

30. System nach Anspruch 18, worin die genannten Knoten Moleküle, die mögliche Anwärter für ein Medikament sind, einschließen.

31. System nach Anspruch 18, worin die genannten biologischen Assertionen Informationen umfassen, die experimentelle Daten, Wissen aus der Fachliteratur, Patientendaten, Daten aus klinischen Versuchen, Compliance-Daten, chemische Daten, medizinische Daten oder hypothetische Daten darstellen.

32. System nach Anspruch 18, worin die genannten biologischen Assertionen Informationen umfassen, die ein Molekül, eine biologische Struktur, einen physiologischen Zustand, eine Eigenschaft, einen Phänotyp oder einen biologischen Prozess darstellen.

33. Herstellungsartikel mit einem computerlesbaren Programmträger, auf dem computerlesbare Instruktionen enthalten sind, um das System nach Anspruch 18 anzuwenden.

34. Rechner, um eine neue biologische Wissensbank zusammenzustellen, mit:
(a) einem Mittel zum Zugriff auf eine elektronische Datenbank biologischer Assertionen, die eine Viel-

zahl von Knoten umfasst, die biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen diesen charakterisieren;
(b) einer vom Nutzer adressierbaren Computeranwendung zur Gewinnung einer Teilmenge von Assertionen aus der Datenbank, die eine Reihe von biologischen Kriterien erfüllen, die ein Nutzer so spezifiziert hat, dass sie ein ausgewähltes biologisches System definieren; und
(c) einem Wissensassembler, der konfiguriert ist, um die erfassten Daten zu kompilieren, um so eine Einheit zu erzeugen, die eine biologische Wissensbank von Assertionen bildet, die potenziell relevant für das genannte ausgewählte biologische System sind.

35. Rechner nach Anspruch 30, der außerdem eine Computeranwendung enthält, die es ermöglicht, aus genannten erfassten Assertionen Schlüsse zu ziehen, um in genannter Einheit logische Widersprüche zu beseitigen.

36. Rechner nach Anspruch 30, der außerdem eine Computeranwendung enthält, die es ermöglicht, aus genannten erfassten Assertionen Schlüsse zu ziehen, um neues biologisches Wissen zu generieren.

37. Rechner nach Anspruch 30, der außerdem eine Computeranwendung enthält, die es ermöglicht, aus genannten erfassten Assertionen Schlüsse zu ziehen, um die Assertionen darin zu augmentieren, indem der genannten Einheit zusätzliche Assertionen aus der genannten Datenbank, die genannte biologische Kriterien erfüllen, hinzugefügt werden.

38. Rechner nach Anspruch 30, der außerdem eine Computeranwendung enthält, die es ermöglicht, aus genannten erfassten Assertionen Schlüsse zu ziehen, um die Assertionen darin zu augmentieren, indem der genannten Wissensbank zusätzliche Assertionen hinzugefügt werden, die in der genannten Einheit neu sind.

39. Rechner nach Anspruch 30, der außerdem eine Computeranwendung enthält, um an der genannten Wissensbank eine Pfadverlaufsanalyse durchführen, um einen Pfadverlauf oder mehrere Pfadverläufe zu erfassen.

40. Rechner nach Anspruch 30, der außerdem eine Computeranwendung enthält, um an den genannten erfassten Assertionen eine homologe Transformation durchzuführen.

41. Rechner nach Anspruch 30, der außerdem eine Computeranwendung enthält, um mit den genannten erfassten Assertionen eine logische Simulation durchzuführen.

42. Rechner nach Anspruch 30, der außerdem

eine Computeranwendung enthält, um der genannten Einheit zusätzliche Assertionen aus Datenquellen außerhalb der genannten Datenbank hinzuzufügen.

43. Rechner nach Anspruch 30, worin genannte Knoten Enzyme, Kofaktoren, Enzymsubstrate, Enzyminhibitoren, DNS, RNS, Transkriptionsregulatoren, DNS-Aktivatoren, DNS-Repressoren, Signalmoleküle, Transmembranmoleküle, Transportmoleküle, Fressmoleküle, Regulationsmoleküle, Hormone, Zytokine, Chemokine, Antikörper, Strukturmoleküle, Metaboliten, Vitamine, Toxine, Nährstoffe, Mineralien, Agonisten, Antagonisten, Liganden, Rezeptoren oder Kombinationen daraus umfassen.

44. Rechner nach Anspruch 30, worin die genannten Knoten Protonen, Gasmoleküle, organische Moleküle, Aminosäuren, Peptide, Proteindomänen, Proteine, Glykoproteine, Nukleotide, Oligonukleotide, Polysaccharide, Lipide, Glykolipide oder Kombinationen daraus umfassen.

45. Rechner nach Anspruch 30, worin genannte Knoten Zellen, Gewebe oder Organe umfassen.

46. Rechner nach Anspruch 30, worin die genannten Knoten Moleküle, die mögliche Anwärter für ein Medikament sind, umfassen.

47. Rechner nach Anspruch 30, worin die genannten biologischen Assertionen Informationen umfassen, die experimentelle Daten, Wissen aus der Fachliteratur, Patientendaten, Daten aus klinischen Versuchen, Compliance-Daten, chemische Daten, medizinische Daten oder hypothetische Daten darstellen.

48. Rechner nach Anspruch 30, worin die genannten biologischen Assertionen Informationen umfassen, die ein Molekül, eine biologische Struktur, einen physiologischen Zustand, eine Eigenschaft, einen Phänotyp oder einen biologischen Prozess darstellen.

49. Verfahren, um neues biologisches Wissen zu ermitteln, das folgende Schritte umfasst:

(a) Bereitstellung einer Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, die biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen den Knoten charakterisieren;

(b) Gewinnung einer Teilmenge von Assertionen aus der Datenbank, die eine Reihe von biologischen Kriterien erfüllen, die ein Nutzer so spezifiziert hat, dass sie ein ausgewähltes biologisches System definieren;

(c) Kompilieren der erfassten Daten, um eine Einheit zu erzeugen, die eine biologische Wissensbank von Assertionen bildet, die potenziell relevant für das genannte ausgesuchte biologische System sind; und

(d) Analyse der genannten Einheit, um neues biologisches Wissen zu ermitteln.

50. Verfahren nach Anspruch 49, worin genanntes biologisches Wissen Voraussagen über das physiologische Verhalten beim Menschen aus der Analyse von an Tieren durchgeführten Experimenten umfasst.

51. Verfahren nach Anspruch 50, worin das genannte physiologische Verhalten die Wirksamkeit oder Toxizität von Medikamenten umfasst.

52. Verfahren nach Anspruch 50, worin das genannte neue biologische Wissen die Ermittlung eines Biomarkers umfasst.

53. Verfahren nach Anspruch 49, worin Schritt (d) die Wiederholung der genannten Methode mit einer anderen Reihe von biologischen Kriterien, um eine andere Einheit zu erstellen und verschiedene Einheiten zu vergleichen, umfasst.

54. Verfahren nach Anspruch 49, worin Schritt (d) die Abbildung experimenteller Daten auf eine Einheit zur Erstellung einer grafischen Ausgabe umfasst.

55. Verfahren nach Anspruch 49, das zudem das Hinzufügen mutmaßlicher Assertionen zur Wissensbank umfasst, die durch Zuordnung oder einen niedrigeren Vertrauenswert unterschieden werden, damit in Schritt (d) spekulativere Resultate produziert werden können.

56. Verfahren nach Anspruch 49, worin Schritt (d) die Anwendung der Pfadverlaufanalyse auf die genannte Wissensbank zur weiteren Erfassung von einem Pfadverlauf oder mehreren Pfadverläufen umfasst.

57. Verfahren nach Anspruch 49, worin Schritt (d) die Anwendung von Algorithmen zur Mechanismusbestimmung umfasst.

58. Verfahren nach Anspruch 49, worin Schritt (d) die Anwendung von Techniken zur Darstellung umfasst, um Wissen und verwandte Daten darzustellen, um so das Verständnis des Nutzers sowie das Erkennen von Mustern und Häufungen zu verbessern.

59. Verfahren nach Anspruch 49, das den zusätzlichen Schritt der Anwendung des Schlussfolgerns auf genannte erfasste Assertionen zur Beseitigung logischer Widersprüche in genannter Einheit umfasst.

60. Verfahren nach Anspruch 49, das den zusätzlichen Schritt der Anwendung des Schlussfolgerns auf genannte erfasste Assertionen zur Augmentierung der Assertionen darin, indem der genannten

Einheit zusätzliche Assertionen aus der genannten Datenbank hinzugefügt werden, umfasst.

61. Verfahren nach Anspruch 49, das den zusätzlichen Schritt der Anwendung des Schlussfolgerns auf genannte erfasste Assertionen zur Augmentierung der Assertionen darin, indem der genannten Einheit zusätzliche Assertionen aus der genannten Datenbank, die neu sind, hinzugefügt werden, umfasst.

62. Verfahren nach Anspruch 49, das den zusätzlichen Schritt der Anwendung der Pfadanalyse auf genannte Wissensseinheit zur Gewinnung eines oder mehrere Pfade umfasst.

63. Verfahren nach Anspruch 49, das den zusätzlichen Schritt der Anwendung der homologen Transformation auf genannte erfasste Assertionen umfasst.

64. Verfahren nach Anspruch 49, das den zusätzlichen Schritt der Anwendung der Ergebnisse aus einer logischen Simulation auf genannte erfasste Assertionen umfasst.

65. Verfahren nach Anspruch 49, das den zusätzlichen Schritt des Hinzufügens zusätzlicher Assertionen aus Datenquellen außerhalb der genannten Datenbank umfasst.

66. Verfahren nach Anspruch 49, worin genannte Knoten Enzyme, Kofaktoren, Enzyme, Substrate, Enzyminhibitoren, DNS, RNS, Transkriptionsregulatoren, DNS-Aktivatoren, DNS-Repressoren, Signalmoleküle, Transmembranmoleküle, Transportmoleküle, Fressmoleküle, Regulationsmoleküle, Hormone, Zytokine, Chemokine, Antikörper, Strukturmoleküle, Metaboliten, Vitamine, Toxine, Nährstoffe, Mineralien, Agonisten, Antagonisten, Liganden, Rezeptoren oder Kombinationen daraus einschließen.

67. Verfahren nach Anspruch 49, worin die genannten Knoten Protonen, Gasmoleküle, organische Moleküle, Aminosäuren, Peptide, Proteindomänen, Proteine, Glykoproteine, Nukleotide, Oligonukleotide, Polysaccharide, Lipide, Glykolipide oder Kombinationen daraus einschließen.

68. Verfahren nach Anspruch 49, worin die genannten Knoten Zellen, Gewebe oder Organe einschließen.

69. Verfahren nach Anspruch 49, worin die genannten Knoten Moleküle, die mögliche Anwärtler für ein Medikament sind, umfassen.

70. Verfahren nach Anspruch 49, worin die genannten biologischen Assertionen Informationen umfassen, die experimentelle Daten, Wissen aus der

Fachliteratur, Patientendaten, Daten aus klinischen Versuchen, Compliance-Daten, chemische Daten, medizinische Daten oder hypothetische Daten darstellen.

71. Verfahren nach Anspruch 49, worin die genannten biologischen Assertionen Informationen umfassen, die ein Molekül, eine biologische Struktur, einen physiologischen Zustand, eine Eigenschaft, einen Phänotyp oder einen biologischen Prozess darstellen.

72. Herstellungsartikel mit einem computerlesbaren Programmträger, auf dem computerlesbare Instruktionen enthalten sind, um das Verfahren nach Anspruch 49 auszuführen.

73. Verfahren nach Anspruch 49, das den zusätzlichen Schritt der Anwendung des Aufstellens einer auf dem genannten neuen Wissen basierenden Hypothese eines Pfadverlaufes und der Durchführung eines biologischen Experiments mit Biomolekülen, Zellen und Tiermodellen oder eines klinischen Versuchs, um die genannte Hypothese zu bestätigen oder zu widerlegen, umfasst.

74. Verfahren um neues biologisches Wissen zu generieren, das die folgenden Schritte umfasst:
(a) die Bereitstellung einer Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, die biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen den Knoten charakterisieren;
(b) die Transformation einer Vielzahl der genannten biologischen Assertionen, um eine Einheit zu erstellen, die ein abgeleitetes Netzwerk umfasst.

75. Verfahren nach Anspruch 74, worin der genannte Transformationsschritt das Schließen auf neue Assertionen aus den genannten biologischen Assertionen umfasst.

76. Verfahren nach Anspruch 75, worin genannte neue Assertionen eine Vielzahl neuer Knoten umfassen, die neue biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen den neuen Knoten charakterisieren.

77. Verfahren nach Anspruch 74, worin genannter Transformationsschritt die Erfassung einer Teilmenge von Assertionen, die eine Reihe von biologischen Kriterien erfüllen, die ein Benutzer spezifiziert hat, um ein ausgewähltes biologisches System zu definieren, aus der Datenbank umfasst.

78. Verfahren nach Anspruch 74, worin der genannte Transformationsschritt das Anwenden mathematischer Mengenlehreoperationen auf genannte biologische Assertionen umfasst, um neue Assertionen zu erstellen.

79. Verfahren nach Anspruch 78, worin genannte neue Assertionen eine Vielzahl neuer Knoten umfassen, die neue biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen neuen Knoten charakterisieren.

80. Verfahren nach Anspruch 74, worin der genannte Transformationsschritt das Zusammenfassen der genannten biologischen Assertionen zur Erstellung neuer Assertionen umfasst.

81. Verfahren nach Anspruch 80, worin genannte neue Assertionen eine Vielzahl neuer Knoten umfassen, die neue biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen neuen Knoten charakterisieren.

82. System, um eine neue biologische Wissensbank zusammenzustellen, mit:

(a) einer Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, die biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen den Knoten charakterisieren; und

(b) eine Anwendung zur Transformation einer Vielzahl der genannten biologischen Assertionen, um eine Einheit zu erstellen, die ein abgeleitetes Netzwerk umfasst.

83. System nach Anspruch 82, worin die genannte Anwendung aus den genannten biologischen Assertionen auf neue Assertionen schließt.

84. System nach Anspruch 83, worin genannte neue Assertionen eine Vielzahl neuer Knoten umfassen, die neue biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen neuen Knoten charakterisieren.

85. System nach Anspruch 82, worin die genannte Anwendung aus der Datenbank eine Teilmenge von Assertionen erfasst, die eine Reihe biologischer Kriterien erfüllen, die ein Benutzer spezifiziert hat, um ein ausgewähltes biologisches System zu definieren.

86. System nach Anspruch 85, worin die genannten erfassten Assertionen eine Vielzahl neuer Knoten umfassen, die neue biologische Elemente repräsentieren, und Deskriptoren, welche die Elemente oder Beziehungen zwischen neuen Knoten charakterisieren.

87. System nach Anspruch 82, worin der genannte Anwendung mathematische Mengenlehreoperationen auf genannte biologische Assertionen anwendet, um neue Assertionen zu erstellen.

88. System nach Anspruch 87, worin genannte neue Assertionen eine Vielzahl neuer Knoten umfassen, die neue biologische Elemente repräsentieren,

und Deskriptoren, die die Elemente oder Beziehungen zwischen neuen Knoten charakterisieren.

89. System nach Anspruch 82, worin die genannte Anwendung genannte biologische Assertionen zusammenfasst, um neue Assertionen zu erstellen.

90. System nach Anspruch 89, worin genannte neue Assertionen eine Vielzahl neuer Knoten umfassen, die neue biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen Knoten charakterisieren.

91. Verfahren zum Entnehmen aus einer biologischen Wissensdatenbank, das Folgendes umfasst:

(a) die Bereitstellung einer Datenbank biologischer Assertionen, die eine Vielzahl von Knoten umfasst, die biologische Elemente repräsentieren, und Deskriptoren, die die Elemente oder Beziehungen zwischen Knoten charakterisieren;

(b) die Transformation einer Vielzahl der genannten biologischen Assertionen, um eine Einheit zu erstellen, die ein abgeleitetes Netzwerk umfasst; und

(c) das Entnehmen aus der genannten Einheit, um neues biologisches Wissen zu ermitteln.

92. Verfahren nach Anspruch 92, worin Schritt (c) die Abbildung experimenteller Daten auf genannte Einheit zur Erstellung einer grafischen Ausgabe umfasst.

93. Verfahren nach Anspruch 91, das zudem das Hinzufügen mutmaßlicher Assertionen zur Wissensbank umfasst, die durch Zuordnung oder einen niedrigeren Vertrauenswert unterschieden werden, damit in Schritt (c) spekulativere Resultate produziert werden können.

94. Verfahren nach Anspruch 91, worin Schritt (c) die Anwendung der Pfadverlaufsanalyse auf die genannte Wissensseinheit zur weiteren Erstellung von einem Pfadverlauf oder mehreren Pfadverläufen, die sich auf experimentelle oder klinische Daten beziehen, umfasst.

95. Verfahren nach Anspruch 91, worin Schritt (c) die Anwendung von Algorithmen zur Mechanismusbestimmung umfasst.

96. Verfahren nach Anspruch 91, worin Schritt (c) die Anwendung von Techniken zur Darstellung umfasst, um wissen und verwandte Daten darzustellen, um so das Verständnis des Nutzers sowie das Erkennen von Mustern und Häufungen zu verbessern.

97. Verfahren nach Anspruch 91, das den zusätzlichen Schritt der Anwendung des Schlussfolgerns auf genannte Assertionen zur Augmentierung der Assertionen darin umfasst, indem der genannten Einheit aus der genannten Datenbank zusätzliche As-

sertionen, die neu sind, hinzugefügt werden.

98. Verfahren nach Anspruch 91, das den zusätzlichen Schritt der Anwendung der Pfadverlaufsanalyse auf die genannte Einheit zur Erstellung von einem Pfadverlauf oder mehreren Pfadverläufen, die sich auf experimentelle oder klinische Daten beziehen, umfasst.

99. Verfahren nach Anspruch 91, das den zusätzlichen Schritt der Anwendung von homologer Transformation auf genannte Assertionen umfasst.

100. Verfahren nach Anspruch 91, das den zusätzlichen Schritt der Anwendung der Ergebnisse aus einer logischen Simulation auf genannte Assertionen umfasst.

101. Verfahren nach Anspruch 91, das den zusätzlichen Schritt des Hinzufügens zusätzlicher Assertionen aus Datenquellen außerhalb der genannten Datenbank umfasst.

Es folgen 22 Blatt Zeichnungen

Anhängende Zeichnungen

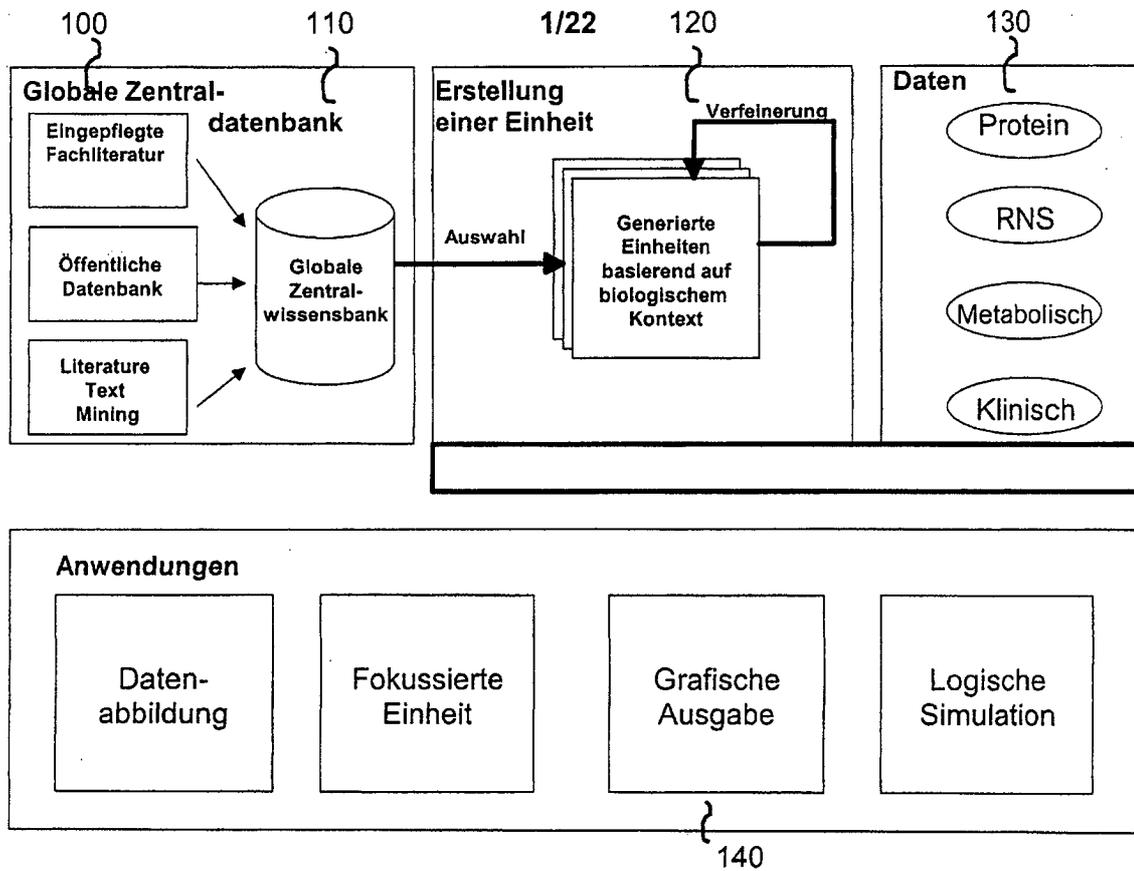


Fig. 1

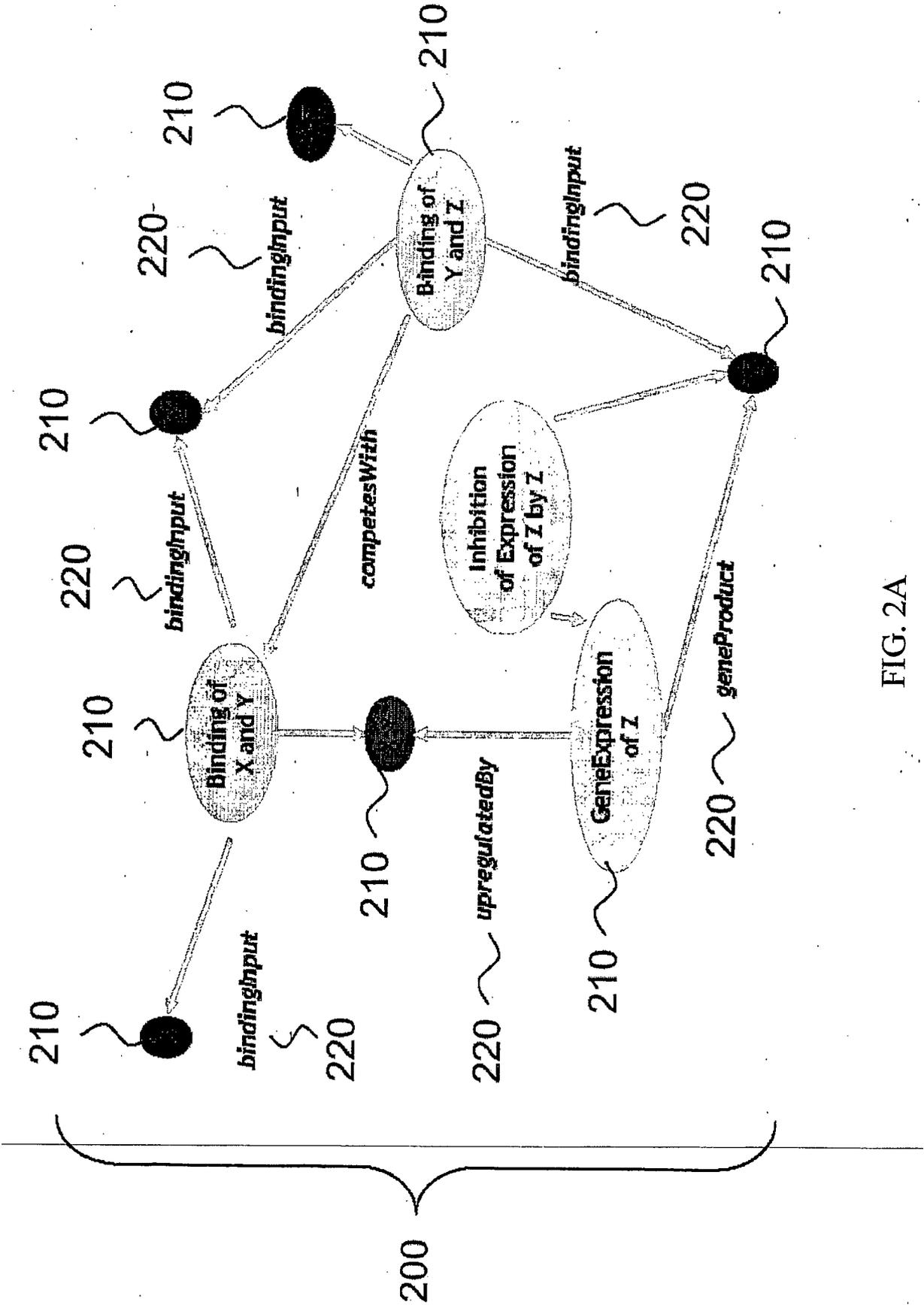


FIG. 2A

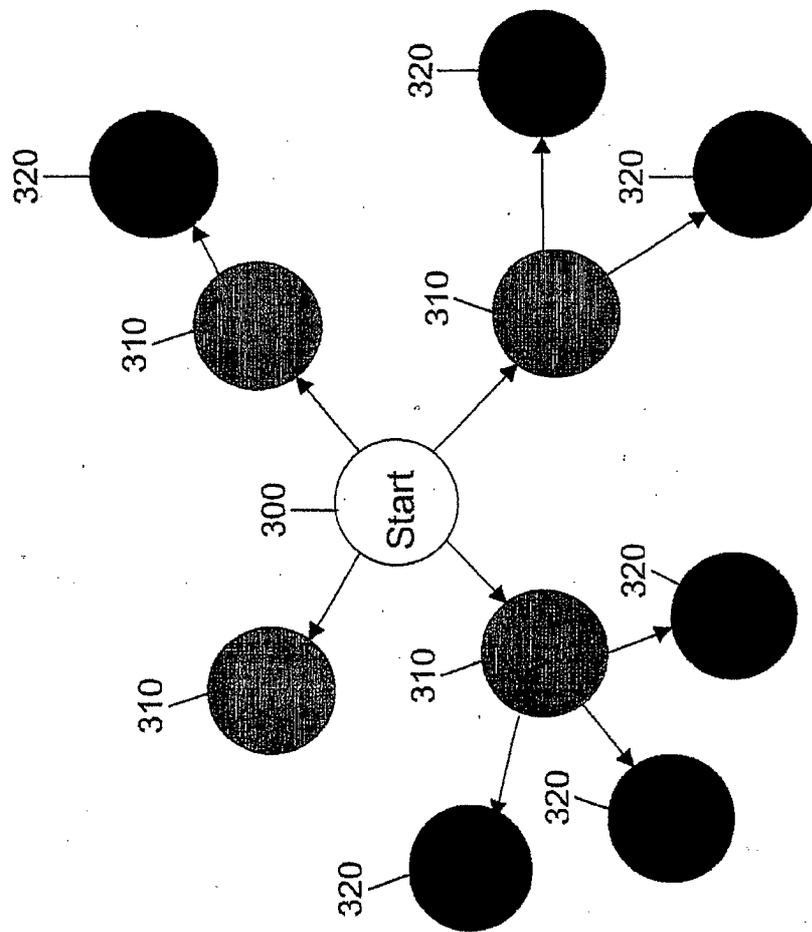


Fig. 3

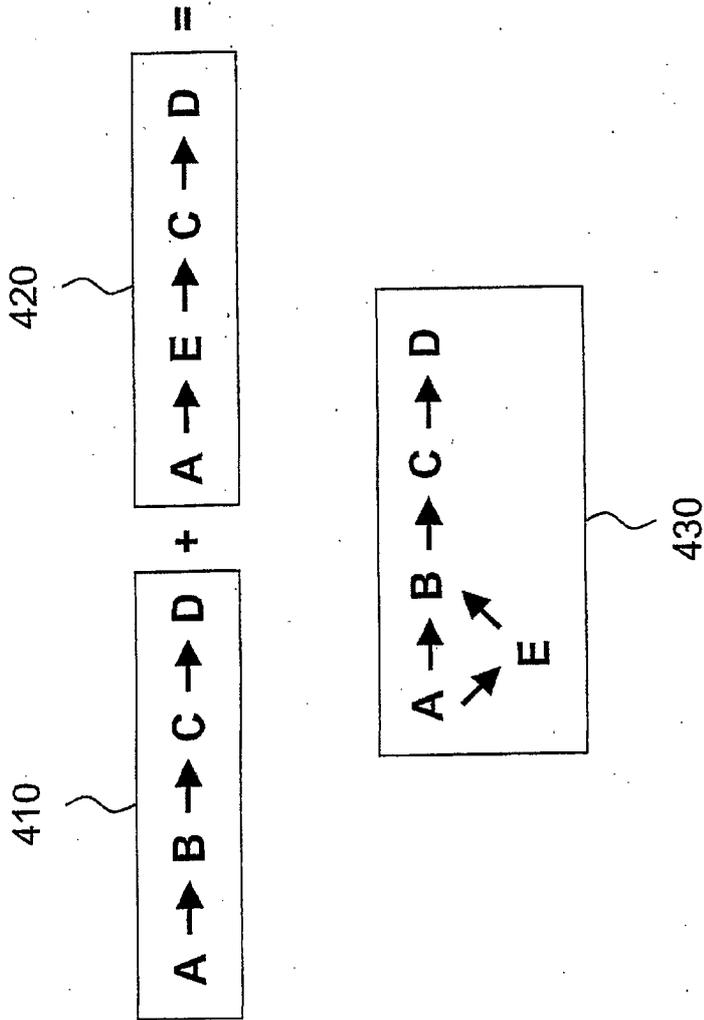


Fig. 4

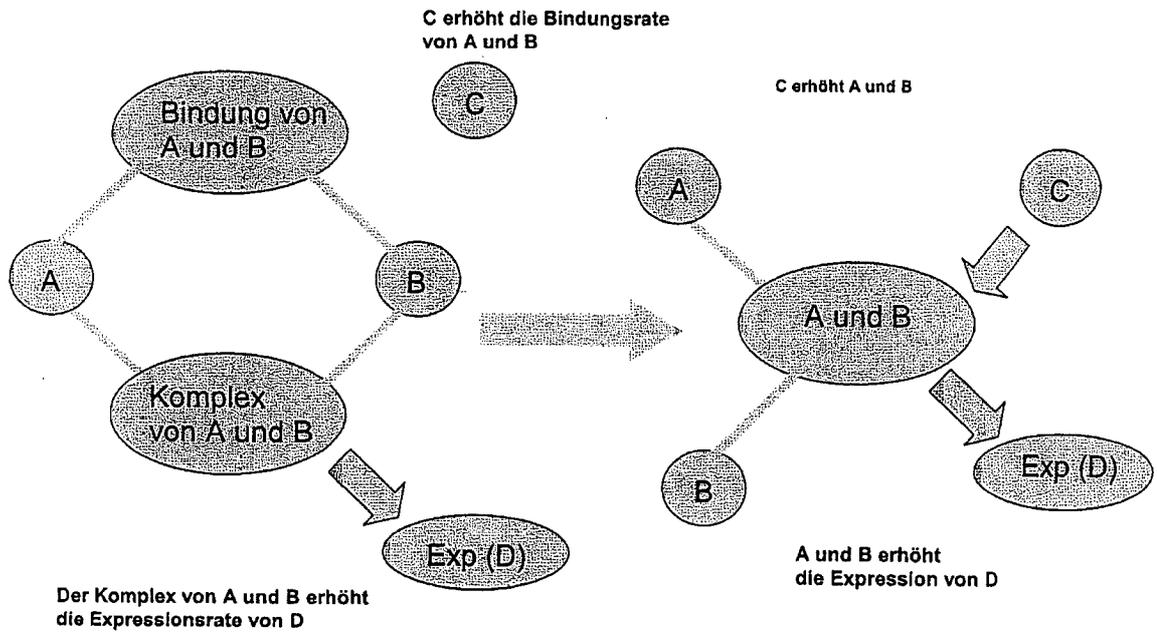


Fig. 7

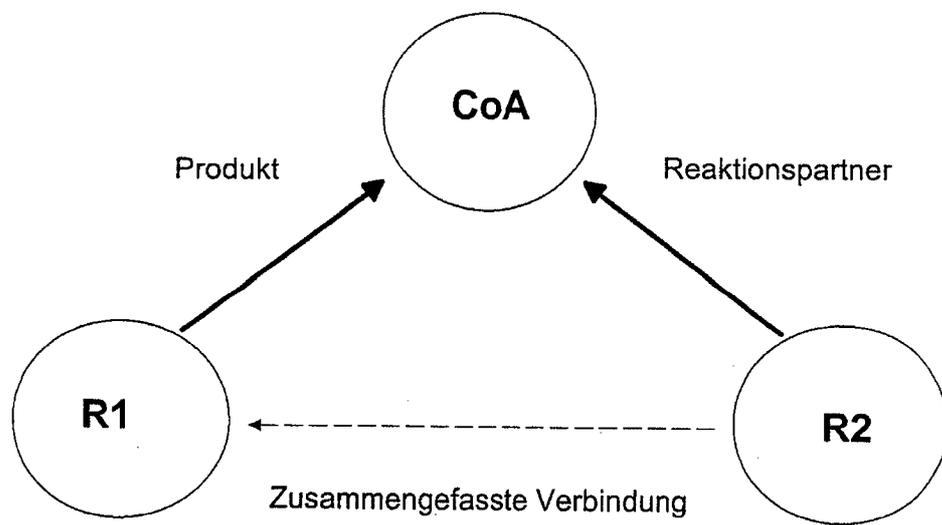


Fig. 8

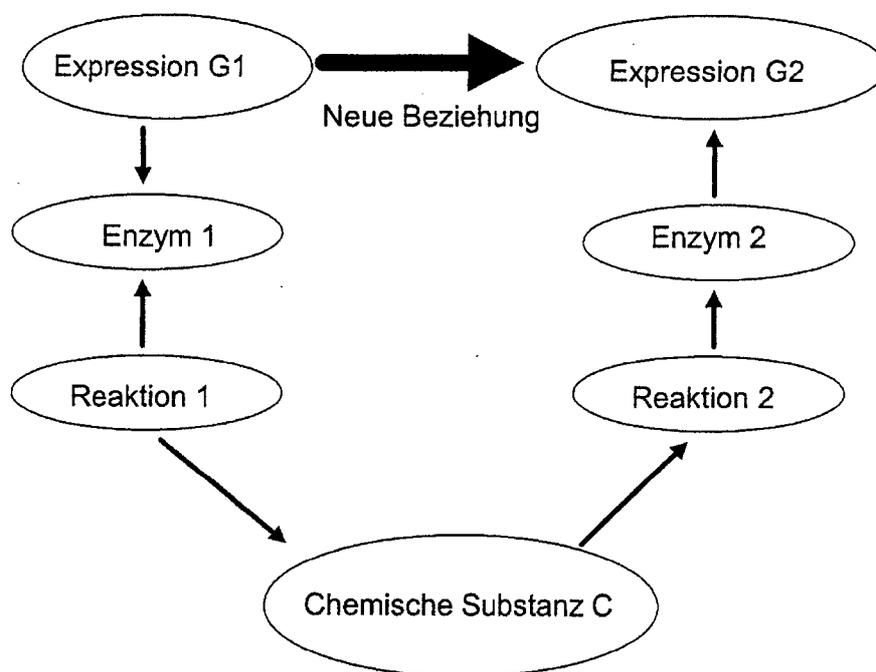


Fig. 9

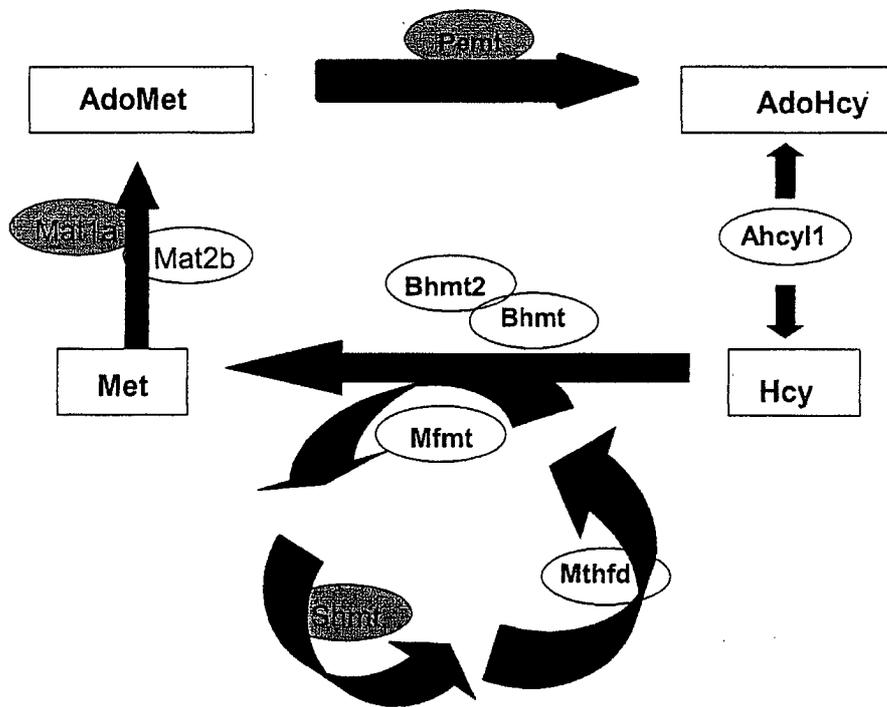


Fig. 10

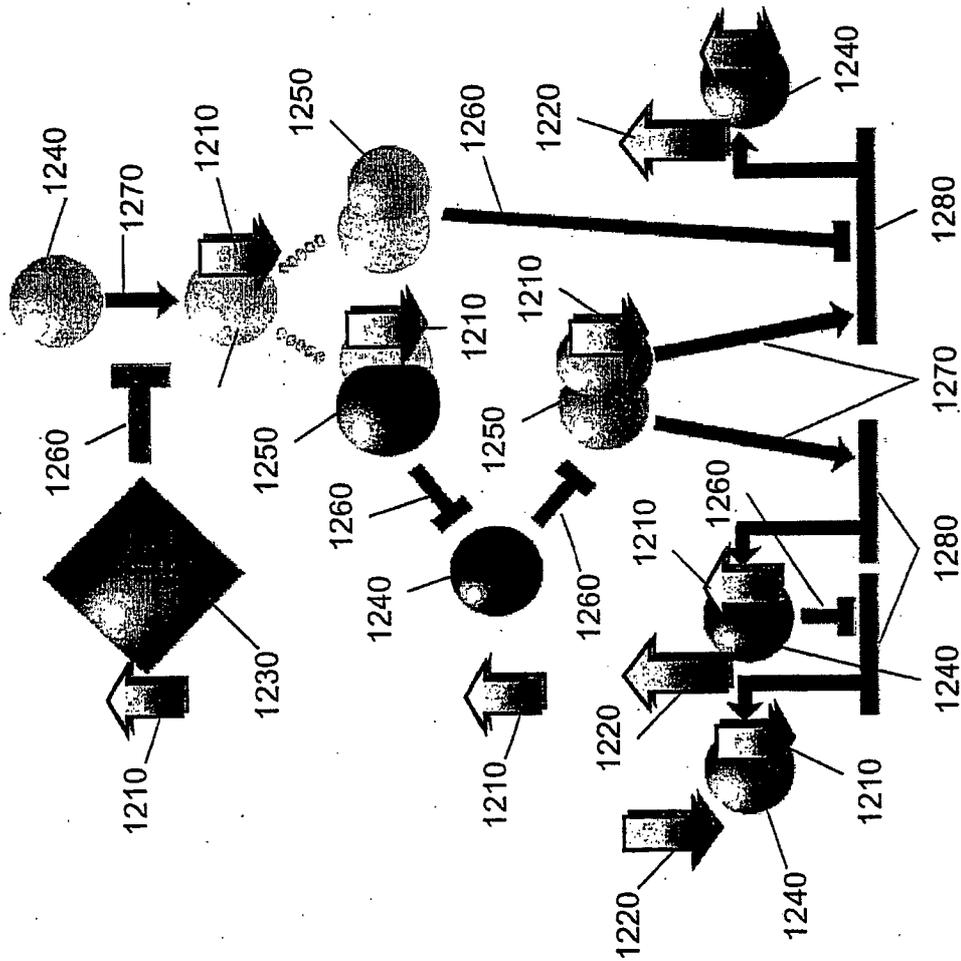


FIG. 12

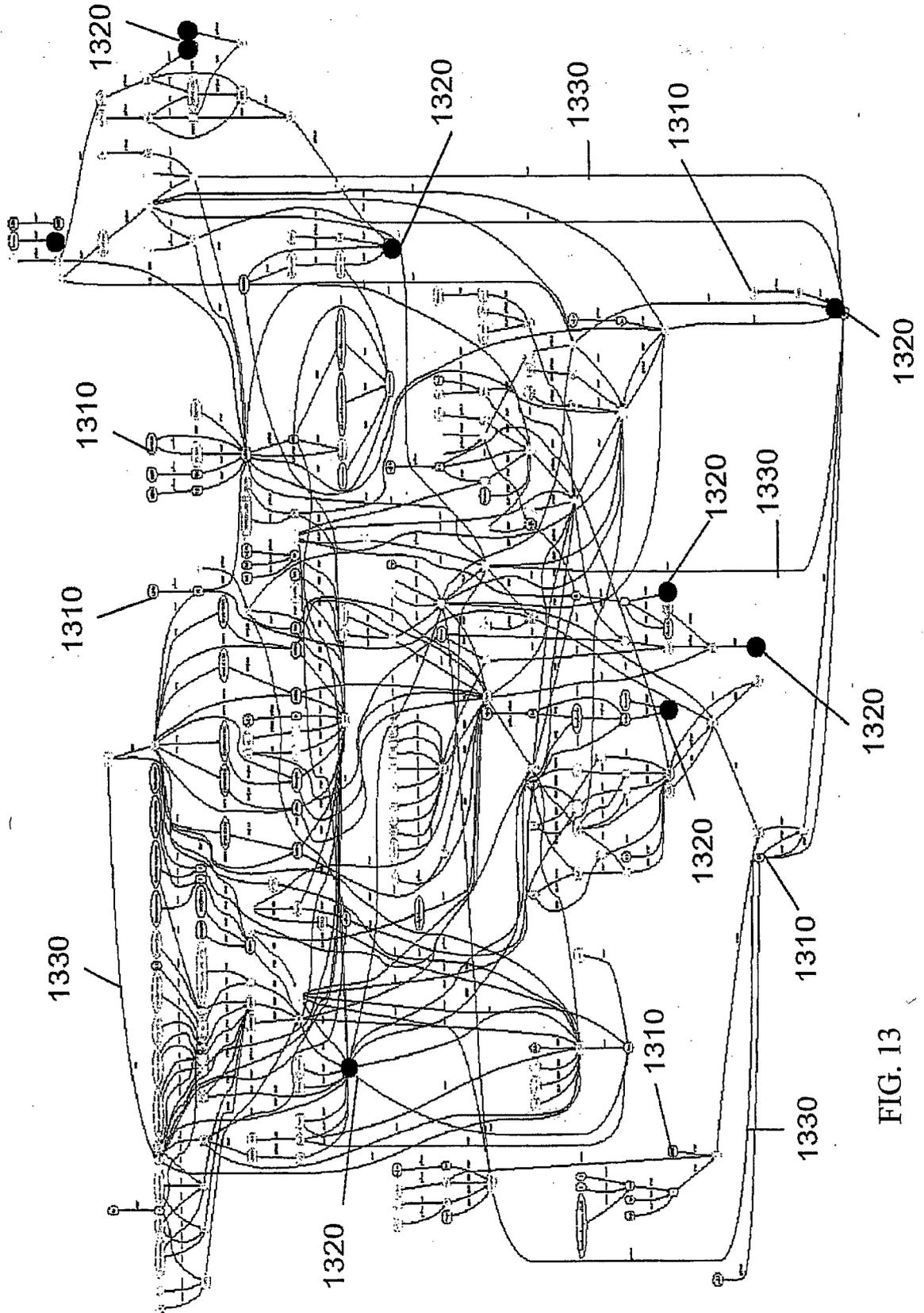


FIG. 13

Behandlung 1	Behandlung 2	Behandlung 3	Behandlung 4
Gen1	Gen1	Gen1	Gen1
Gen2	Gen2	Gen2	Gen2
Gen3	Gen3	Gen3	Gen3
Gen4	Gen4	Gen4	Gen4
Gen5	Gen5	Gen5	Gen5
Gen6	Gen6	Gen6	Gen6
Gen7	Gen7	Gen7	Gen7
Gen8	Gen8	Gen8	Gen8
Gen9	Gen9	Gen9	Gen9

Fig. 14

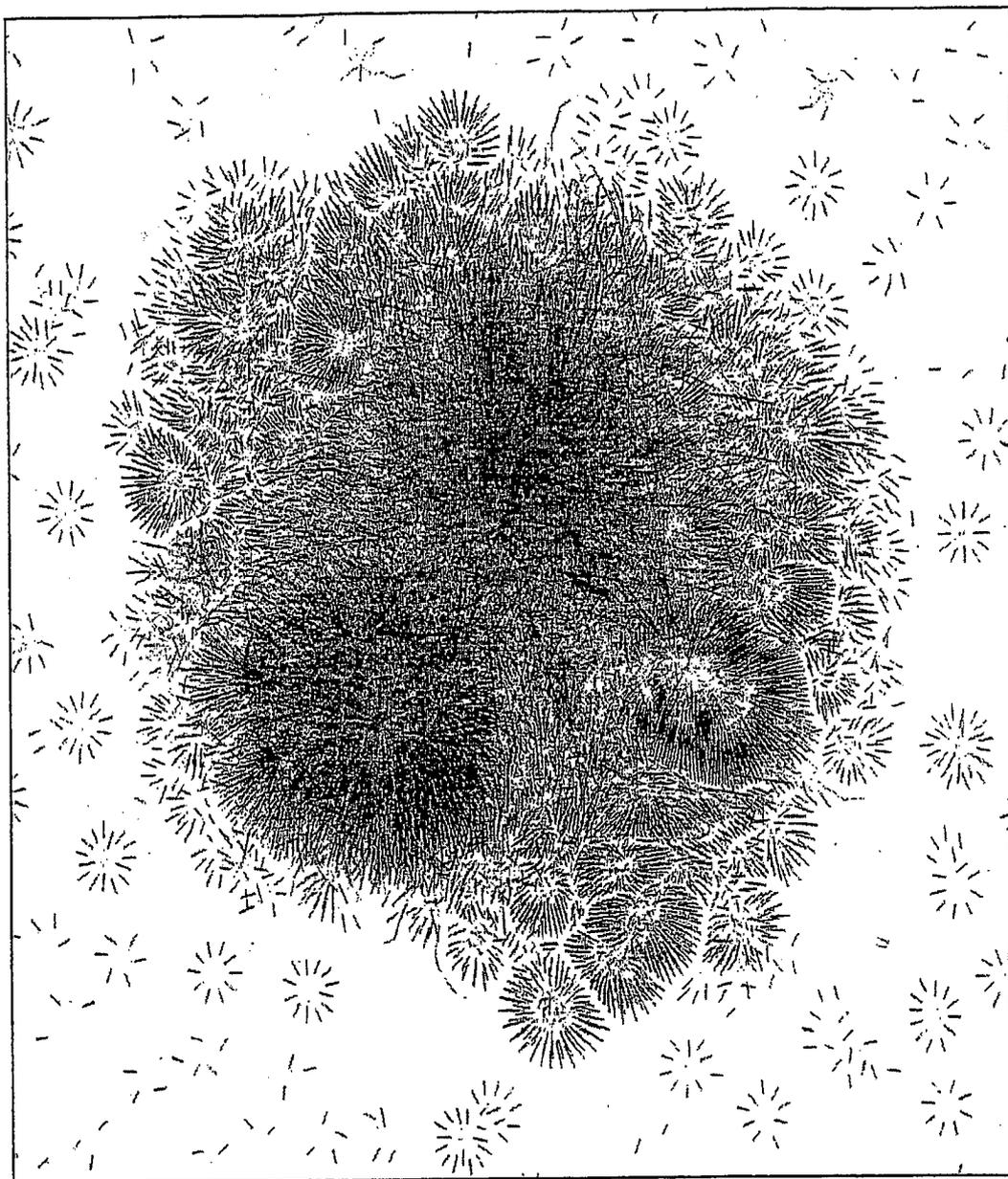


FIG. 15

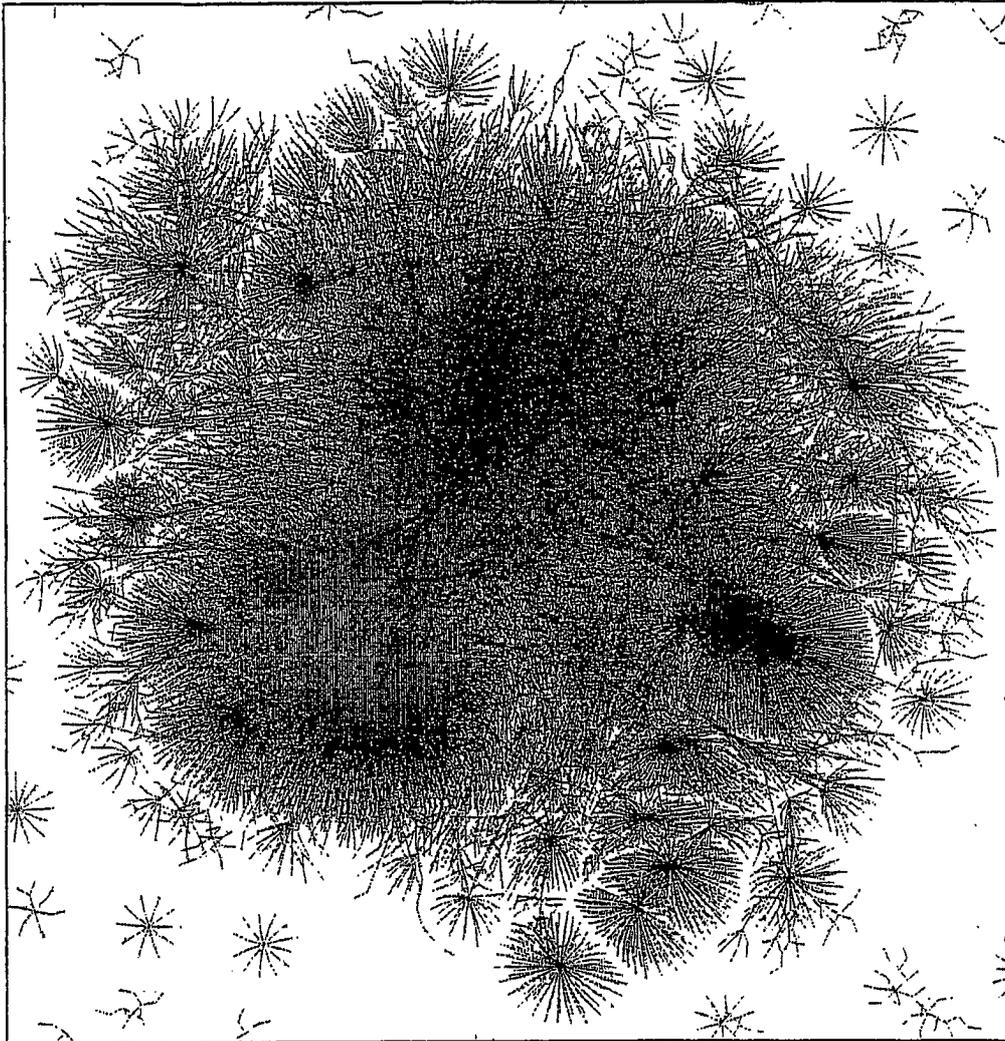


FIG. 16

Beta-Oxidation

	Methylverzweigte carboxylierte Fettsäure		
Reaktionspartner	Methylverzweigtes carboxyliertes Fettsäureacyl-CoA		
1 - Oxidation	Nach 8 Stunden	Nach 24 Stunden	Nach 8 Tagen
Insgesamt	Acox3	Acox3	Acox3
Produkt/Reaktionspartner	Methylverzweigtes carboxyliertes 2-trans-Enoyl-CoA		
2 - Hydratation	Nach 8 Stunden	Nach 24 Stunden	Nach 8 Tagen
	Hsd17b4	Hsd17b4	Hsd17b4
Produkt/Reaktionspartner	Methylverzweigtes carboxyliertes D-3-Hydroxyacyl-CoA		
3 - Oxidation	Nach 8 Stunden	Nach 24 Stunden	Nach 8 Tagen
	Hsd17b4	Hsd17b4	Hsd17b4
Produkt/Reaktionspartner	Methylverzweigtes carboxyliertes 3-Ketoacyl-CoA		
4 - Thiololyse	Nach 8 Stunden	Nach 24 Stunden	Nach 8 Tagen
	Scp2	Scp2	Scp2
Produkt	Methylverzweigtes carboxyliertes Fettsäure-CoA (-C2) + Acetyl-CoA	Scp2 Protein	Scp2 Protein
			Widersprüchlicher Untersuchungs-aufbau

Fig. 17

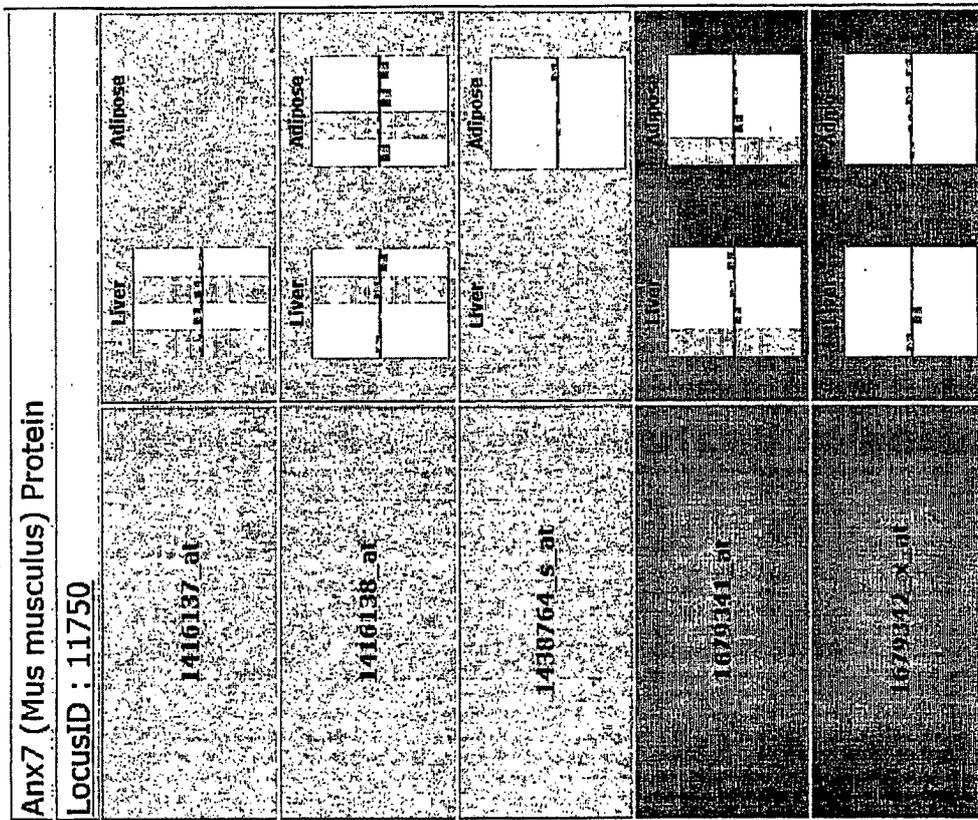
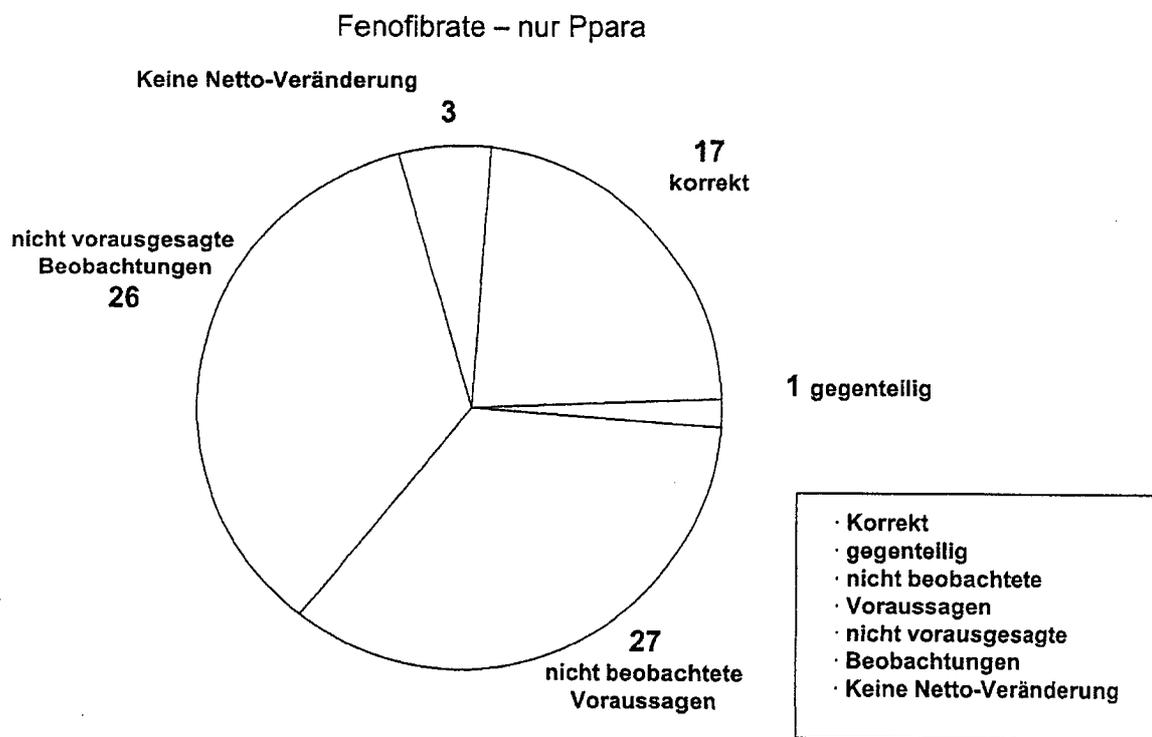


FIG. 18



Ausführen einer Rückwärtssimulation aus
Genexpressionsresultaten, um
Gene zu bekommen, deren Aktivitäten Resultate erklären
können
Filtern nach Votum und Wissen des Systems
Speichern der Liste von Genen als
Rückwärtssimulationsresultate
Abrufen von experimentellen Genexpressionsresultaten
Speichern von nach oben regulierten Genen in upTargets-
Liste
Speichern von nach unten regulierten Genen in
downTargets-Liste
Ausführen einer Vorwärtssimulation unter Verwendung von
Rückwärtssimulationsresultaten, um vorausgesagte
Veränderungen im System zu erhalten
Filtern der vorausgesagten, um nur die
Genexpressionsknoten zu haben
Sortieren der Gene nach positiven und negativen
Gesamtvotes (TPV und TNV) nach folgenden Kategorien:
 TPV-TNV > 0: wird laut Voraussage nach oben
reguliert
 Wenn in upTargets: korrekte Voraussage
 Wenn in downTargets: gegenteilige Voraussage
 Wenn in keiner der beiden: Voraussage wurde nicht
beobachtet
 TPV-TNV < 0: wird laut Voraussage nach oben
reguliert
 Wenn in downTargets: korrekte Voraussage
 Wenn in upTargets: gegenteilige Voraussage
 Wenn in keiner der beiden: Voraussage wurde nicht
beobachtet
Sonst keine Netto-Veränderung
Jetzt gehe durch upTargets and downTargets und ordne die
Gene, die nicht in den Voraussageergebnissen sind, in die
Kategorie Nicht Vorausgesagte Beobachtung
Gib die folgenden Kategorien aus, stelle sie als
Tortendiagramme oder Histogramme dar
 Anzahl korrekter Voraussagen
 Anzahl der gegenteiligen Voraussagen
 Anzahl der nicht beobachteten Voraussagen
 Anzahl der nicht vorausgesagten Beobachtungen
 Zähler für keine Netto-Veränderung

Fig. 20

```

HAUPTFUNKTION
Lade Liste der sezernierten Proteine
Führe eine Pfadsuche mit sezernierten Proteinen als Quellknoten aus (Suche in
drei Schritten)
Kennzeichne die Knoten mit dem Mindestabstand zu einem Quellknoten, (d. h.
einem Mindestabstand zu einem sezernierten Protein)

Öffne eine Ausgabedatei
Hole eine Liste der Knoten vom Graphen
  FOR jeden Knoten
  IF er keine Locus-ID hat, überspringen
  IF er > 3 Schritte von einem sezernierten Protein entfernt ist, überspringen
Schreibe Locus-ID, Name und Messdaten für die Punktwerte von
Richtungskoeffizient, Faltung, Biomarker, Sezernierung, Tabellenauflistung
des Mindestabstands zum Sezernierten Protein

Lade eine Ausgabedatei
Fasse die Werte der Maße zusammen
Sortiere die Spalten nach Gesamtpunktwert
Wähle die Proteine am Anfang und am Ende der Liste und überprüfe sie darauf,
ob sie gute Kandidaten für Biomarker abgäben
Positive Punktwerte werden bevorzugt (d. h. das gemessene Niveau wäre
proportional zum Fortschreiten einer Krankheit)

BERECHNUNGEN VON RICHTUNGSKOEFFIZIENTEN UND FALTUNG
FOR jeden Knoten
Hole eine Liste von Untersuchungen und die entsprechenden Daten
IF Untersuchung ist ein 'x_at' (d. h. Neigung zu Kreuzbindung), dann
ignorieren
Berechne Punkt看wert für Untersuchung
FOR jeden Zelltyp
IF der Wert über dem Grenzwert ist, Punkt看wert 2
IF der Wert unter dem Grenzwert ist, Punkt看wert 1
ELSE Punkt看wert 0
IF es einen Vorzeichenkonflikt bei den Werten über Zelltypen gibt
(abhängig von der gestellten Frage), setze Markierung für einen Konflikt
IF Muster DOSIS benutzt wird, fasse die Punktwerte über Zelltypen zusammen
IF Muster RESISTENZ verwendet wird, multipliziere den Punkt看wert für
resistente Zelllinien mit 2 und subtrahiere die Werte für die sensiblen
Zelllinien
IF Muster WIRKSAMKEIT verwendet wird, multipliziere den Punkt看wert für die
sensibelste Zelllinie mit 2, addiere den Punkt看wert für die partiell sensible
Zelllinie und subtrahiere die Werte für die resistente Zelllinie
Überprüfe den Punkt看wert mit dem fortlaufenden Punkt看wert über alle
Untersuchungen
IF der Punkt看wert im Widerspruch (falsches Vorzeichen) zum fortlaufenden
Punkt看wert steht, weise eine 0 für den Punkt看wert des Knotens zu
Gib den höchsten (absoluten) Wert über die Knoten aus

ZUWEISUNG DER PUNKTWERTE FÜR BIOMARKER UND SEZERNIERTE PROTEINE
Wenn das Protein als Biomarker bekannt ist oder (mutmaßlich) ein sezerniertes
Protein ist, gib 2 aus
Sonst gib 1 aus

TABELLENLISTE
Gib eine Liste von Tabellen aus, auf der die Gene stehen

MINDESTABSTAND
Gib den Mindestabstand zum sezernierten Protein aus (gib eine 3 aus, wenn
seine Entfernung von einem SP im Bereich der Werte 0 bis 3 liegt)

```

Fig. 21