United States Patent **6,611,039**

Anthony **August 26, 2003**

Vertically oriented ***nano***-fuse and ***nano***-resistor circuit elements

Abstract

Vertically oriented nano-circuits including fuses and resistors allow for significant densities to be achieved. The vertically oriented nano-circuits can be fabricated using standard known processes such as Damascene, wet etching, reactive etching, etc. Thus little additional capital expenditure is required other than to acquire present state-of-the-art equipment. Devices using these vertically oriented nano-circuits are also inexpensive to manufacture.

Inventors: **Anthony; Thomas C.** (Sunnyvale, CA)

Assignee: **Hewlett-Packard Development Company, L.P.** (Houston, TX)

Appl. No.: **964770**

Filed: **September 28, 2001**

Current U.S. Class: **257/529**; 257/734; 257/758; 257/759; 257/774; 257/776; 438/622; 438/625

Intern'l Class: H01L 029/00; H01L 023/52; H01L 029/40; H01L 021/476.3; H01L 023/48

Field of Search: 257/758,759,776,734,774,529 438/622,625

References Cited [Referenced By]

U.S. Patent Documents

6031287	Feb., 2000	Harshfield	257/734.
6344371	Feb., 2002	Fischer et al.	438/106.

Other References

IBM Technical Disclosure Bulletin; vol. 9, Feb. 1992.

Primary Examiner: Flynn; Nathan J.
Assistant Examiner: Mandala, Jr.; Victor A.

Parent Case Text

RELATED APPLICATIONS

The following applications of the common assignee may contain some common disclosure and may relate to the present invention:

U.S. patent application Ser. No. 09/964,768, entitled "ONE TIME PROGRAMMABLE FUSE/ANTI-FUSE COMBINATION BASED MEMORY CELL";

U.S. patent application Ser. No. 09/924,500, filed Aug. 9, 2001, entitled "ONE-TIME PROGRAMMABLE UNIT MEMORY CELL BASED ON VERTICALLY ORIENTED FUSE AND DIODE AND ONE-TIME PROGRAMMABLE MEMORY USING THE SAME"; and

U.S. patent application Ser. No. 09/924,577, filed Aug. 9, 2001, entitled "ONE-TIME PROGRAMMABLE MEMORY USING FUSE/ANTI-FUSE AND VERTICALLY ORIENTED FUSE UNIT MEMORY CELLS".

Claims

What is claimed is:

1. A vertically oriented nano-circuit, comprising:

a top conductor extending in a first direction;

a bottom conductor extending in a second direction so as to define an overlap between said top and bottom conductors, wherein said first and second directions are not parallel, said bottom conductor having electrical connectivity with said top conductor;

a vertically oriented conductive spacer formed in said overlap having electrical connectivity with sad top and bottom conductors; and

an insulating plug substantially occupying a center of said overlap such that said insulating plug directly contacts said top and bottom conductors.

2. The vertically oriented nano-circuit of claim 1, wherein:

said vertically oriented conductive spacer is a vertically oriented nano-fuse.

3. The vertically oriented nano-circuit of claim 1, wherein said vertically oriented conductive spacer may be at least one of a semiconductor, conductor, low melting temperature material, refractory metal, transition metal, silicide, high resistivity material, and carbon.

4. The vertically oriented nano-circuit of claim 3, wherein:

said semiconductor includes at least one of Si and Ge;

said conductor includes at least one of Al, Cu, Ag, Au, Pt, and alloys thereof;

said low melting temperature material includes at least one of In, Zn, Sn, Pb, and alloys thereof;

said refractory metal includes at least one of Ta, W, and alloys thereof;

said transition metal includes at least one of Ni, Cr, and alloys thereof;

said silicide includes at least one of PtSi, WSi, and TaSi; and

said high resistivity material includes at least one of TaN, TaSiN, WN, WsiN.

5. The vertically oriented nano-circuit of claim 1, wherein:

said vertically oriented conductive spacer is shaped such that a void exists in said conductive spacer substantially about a center of said conductive spacer.

6. The vertically oriented nano-circuit of claim 1, wherein said conductive spacer substantially occupies a closed region near said region of overlap such that an inner wall of said conductive spacer is bounded by said insulating plug, said vertically oriented nano-circuit Further comprising:

an insulator formed around a perimeter of said closed region such that an outer wall of said conductive spacer is bounded by said insulator.

7. The vertically oriented nano-circuit of claim 6, wherein:

each of said top and bottom conductors is composed of at least one of polysilicon, aluminum, copper, gold, tungsten, and any alloys made therefrom; and

each of said insulator and said insulating plug is composed of at least one of silicon oxides, silicon nitrides, aluminum oxides, aluminum nitrides, silicon oxynitrides, and tantalum oxides.

8. The vertically oriented nano-circuit of claim 1, wherein:

a vertical height to lateral thickness ratio of said vertically oriented conductive spacer is unity or greater.

9. The vertically oriented nano-circuit of claim 1, wherein said vertically oriented conductive spacer extends in one of said first and second directions.

10. The vertically oriented nano-circuit of claim 9, further comprising:

an insulator formed at an exterior region of said vertically oriented conductive spacer; and

an insulating plug formed at an interior region of said vertically oriented conductive spacer.

11. The vertically oriented nano-circuit of claim 1, wherein:

tops of said vertically oriented conductive spacer and said insulating plug are coplanar.

12. A method to form a vertically oriented nano-circuit, comprising:

forming a top conductor extending in a first direction;

forming a bottom conductor extending in a second direction so as to define an overlap between said top and bottom conductors, wherein said first an second directions are not parallel, said bottom conductor having electrical connectivity with said top conductor;

forming a vertically oriented conductive spacer in said overlap having electrical connectivity with said top and bottom conductors; and

forming an insulating plug substantially occupying a center of said overlap such that said insulating plug directly contacts said top and bottom conductors.

13. The method of claim 12, wherein said vertically oriented conductive spacer is a vertically oriented nano-fuse.

14. The method of claim 12, wherein forming a void in an interior of said vertically oriented conductive spacer.

15. The method of claim 12, further comprising:

forming said insulating plug substantially in a center of a closed region near said overlap such that an inner wall of said vertically oriented conductive spacer is bounded by said insulating plug; and

forming an insulator around a perimeter of said closed region such that an outer wall of said conductive spacer is bounded by said insulator.

16. The method of claim 12, wherein a vertical height of said vertically oriented conductive spacer is equal to or greater than a width of a closed region near said overlap.

17. The method of claim 12, wherein a vertical height to lateral thickness ratio of said vertically oriented conductive spacer is unity or greater.

18. The method of claim 12, wherein said vertically oriented conductive spacer extends in one of said first and second directions.

19. The method of claim 18, further comprising:

forming an insulator on an exterior of said vertically oriented conductive spacer; and

forming an insulating plug in an interior of said vertically oriented conductive spacer.

20. The method of claim 18, wherein:

forming a void in an interior of vertically oriented conductive spacer.

Description

FIELD OF THE INVENTION

This invention relates generally to nano-circuits. More particularly, the invention relates to vertically oriented nano-fuses and nano-resistors in manufacturing semiconductor devices.

BACKGROUND OF THE INVENTION

The demand for semiconductor devices has increased dramatically in recent years. One can readily observe the pervasiveness of consumer electronic devices in the modern world. Most or all of the consumer electronic devices are made possible because of developments in the semiconductor devices. As the electronic devices become smaller, more sophisticated, and less expensive, increasingly higher

United States Patent **6,625,585**
MacCuish , et al. **September 23, 2003**

Method and system for *artificial intelligence* directed lead discovery though multi-domain agglomerative clustering

Abstract

A system for helping a chemist to identify pharmacophoric mechanisms, based on a set of input data representing many chemical compounds. Given an input data set defining for each compound a feature characteristic and an activity characteristic, a computer agglomeratively clusters representations of the molecules based on their feature characteristics. The result of this process is a multi-domain pyramid structure, made up of a number of nodes each representing one or more molecules. For each node, the computer identifies a representative feature set (such as a largest substructure common among the molecules in the node) and a representative activity level (such as an average of the activity levels of the molecules in the node). The computer then provides as output to a chemist a description of all or part of the pyramid. This process thus converts a large set of raw data into an understandable and commercially useful form, which can assist the chemist in developing beneficial new pharmaceuticals.

Inventors: **MacCuish; John D.** (Santa Fe, NM); **Nicolaou; Christodoulos A.** (Limassaol, CY)

Assignee: **Bioreason, Inc.** (Santa Fe, NM)
Appl. No.: **549746**
Filed: **April 14, 2000**

Current U.S. Class:	**706/10**; 706/11
Intern'l Class:	G06E 001/00; G06F 017/00
Field of Search:	706/10,11

References Cited [Referenced By]

U.S. Patent Documents

5025388	Jun., 1991	Cramer, III et al.	
5263120	Nov., 1993	Bickel.	
5307287	Apr., 1994	Cremer, III et al.	
5590218	Dec., 1996	Ornstein	382/157.
5684711	Nov., 1997	Agrafiotis et al.	
5751605	May., 1998	Hurst et al.	
5825909	Oct., 1998	Jang	382/132.

Foreign Patent Documents

WO 98/47087	Oct., 1998	WO.

Other References

Downs, G.M. and Willett, P., Similarity Searching and Clustering of Chemical-Structure Databases Using Molecular Property Data, J. Chem. Inf. Comput. Sci. 34:1094-1102 (1994).
Kearsley, S.K. et al., Chemical Similarity Using Physiochemical Property Descriptors, J. Chem. Inf. Comput. Sci, 36:118-127 (1996).
Brown, R.D., et al., Matching Two-Dimensional Chemical Graphs Using Genetic Algorithms, J. Chem. Inf. Comput. Sci. 34:63-70 (1994).
Brown, R.D. and Martin, Y.C., Use of Structure-Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection, J. Chem. Inf. Comput. Sci. 36:572-584 (1996).
Discriminant Analysis and Clustering--Panel on Discriminant Analysis,

Classification and Clustering. Statistical Science 4: 34-69 (1989).
Barnard, J.M. and Downs, G.M., Chemical Fragment Generation and Clustering Software, Product Descriptions, Jun. 27, 1996.
Regalado, A., Preclinical Strategies--Drug Development's Preclinical Bottleneck. Start-Up, pp26-37 (Dec. 1997).
Longman, R., Marketplace Strategies--Screening the Screeners. Start-Up, pp. 14-22 (Sep. 1997).
Thayer, A.M., Combinatorial Chemistry becoming core technology at drug discovery companies. C&EN, pp. 57-67 (Feb. 1996).
Combinatorial Chemistry--Combinatorial chemists focus on small Molecules, molecular recognition, and automation, C & EN, pp. 28-54 (Feb. 1996).
Kohonen, T., Self-Organizing Maps, Springer, pp. 85-144.
Goodacre, R. et al., Quantitative Alanysis of Multivariate Data using Artificial Neural Networks: A Tutorial Review and Applications to the Deconvolution of Pyrolysis Mass Spectra, Tutorial form <> Zentralblatr fur Bakteriologie (1999).
Chen, X. et al., Recursive Partitioning Analysis of a Large Structure-Activity Data Set Using Three-Dimensional Descriptors. J. Chem Inf. Comput. Sci. (1998).
James, C.A. et al., Daylight Theory Manual Daylight 4.61, Daylight Chemical Information Systems, Inc., Version 11 (Feb. 1997).
Labute, P., Binary QSAR: A new Technology for HTS and UHTS Data Analysis, Chemical Computing Group, Inc. In Journal of the Chemical Computing Group (1998).
From the World Wide Web: www.netsci.org, Network Science--Welcome to NetSci's Lists of Computational Chemistry Software (1999), printed Feb. 8, 1999.
From the World Wide Web: www.netsci.org, Network Science--Welcome to NetSci's Combinatorial Chemistry and Mass Screening YellowPages (1999), printed Feb. 8, 1999.
Weininger, D., SMILES, a Chemical Language and Information System, 1. Introduction to Methodology and Encoding Rules. J. Chem. Inf. Comput. Sci., 28: 31-36 (1988).
Cook, D.J. et al., Knowledge Discovery from Structural Data. Journal of Intelligence and Information Sciences, vol. 5, No., 3, pp. 229-245 (1995).
Djoko, S. et al., An Empirical Study of Domain Knowledge and its Benefits to Substructure Discovery. In IEEE Transactions on Knowledge and Data Engineering, vol. 9, No. 4, pp. 1-13 (1997).

Galal, G. et al., Improving Scalability in a Knowledge Discovery System by Exploiting Parallelism. In the Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, pp. 171-174 (1997).
Holder, L. B. and D. J. Cook. Discovery of Inexact Concepts from Structural Data. In IEEE Transactions on Knowledge and Data Engineering, vol. 5,No. 6, pp. 992-994 (1993).
Holder , L. B. et al., Fuzzy Substructure Discovery. In Proceedings of the Ninth International Conference on Machine Learning, pp. 218-223 (1992).

Djoko, S. et al., Analyzing the Benefits of Domain Knowledge in Substructure Discovery. In Proceedings of the First International Conference on Knowledge Discovery and Data Mining, pp. 75-80 (1995).
Cook, D. J. et al., Scalable Discovery of Informative Structural Concepts Using Domain Knowledge. In IEEE Expert, vol. 11, No. 5, pp. 59-68 (1996).
Aude, J.C. et al., Applications of the pyramidal clustering method to biological objects. Computers & Chemistry 23:301-315 (1999).
Bertrand, P., Structural Properties of Pyramidal Clustering. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 19: 35-53 (1995).
Fausett, L., Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. pp. 169-188 (1994).
Godden, Jeffrey W., et al. Combinatorial Preferences Affect Molecular Similarity/Diversity Calculations Using Binary Fingerprints and Tanimoto Coefficients. J. Chem. Inf. Comput. Sci. 2000, vol. 40, pp. 163-166.
"Pyramids" Homepage: Including http://bioweb.pasteur.fr/docs-gensoft/pyramids/QuickStart.
From the World Wide Web: http://www.tripos.com/software/charisma.html, printed Feb. 15, 2000.
From the World Wide Wide: http://www.tripos.com/about/press/1999/1999.03.23.html, printed Feb. 15, 2000.
From the World Wide Web: http://www.daylight.com/meetings/mug2000/JmacCuish/index.html, printed Apr. 5, 2000.
From the World Wide Web: http://www.daylight.com/meetings/mug2000/JmacCuish/MugTies.html, printed Mar. 2, 2000.

van Osdol, W. W. et al., "Use of the Kohonen Self-organizing Map to Study the Mechanisms of Action of Chemotherapeutic Agents", Journal of the National Cancer Institute, 86:1853-1859 (1994).
Ornstein, L., "Computer Learning and the Scientific Method: A Proposed Solution to the Information Theoretical Problem of Meaning", Journal of the Mount Sinai Hospital, XXXII:437-494 (1965).
Barnard, J. M. and Downs, G. M., "Clustering of Chemical Structures on the Basis of Two-Dimensional Similarity Measures", Journal of Chemical Information and Computer Sciences, 32:644-649 (1992).
Grethe, G. and Hounshell, W. D., "Similarity Searching in the Development of New Bioactive Compounts. An Application.", Chemical Structure Proceedings International Conference, pp. 399-407 (1993).
King, R. D. et al., "Comparison of Artificial Intelligence Methods for Modeling Pharmaceutical Qsars", Applied Artificial Intelligence, 9:213-233 (1995).
Jain, K. J. et al., "Algorithms for Clustering Data", Algorithms for Clustering Data, pp. 96-101 (1988).

Primary Examiner: Davis; George B.
Attorney, Agent or Firm: McDonnell Boehnen Hulbert & Berghoff

---

Parent Case Text

RELATED APPLICATIONS

The application claims priority to U.S. Provisional Patent Application No. 60/183,383, entitled "Method and System for Artificial Intelligence Directed Lead Discovery Through Multi-Domain Agglomerative Clustering," filed by John D. MacCuish and Christodoulos A. Nicolaou on Feb. 18, 2000, which is assigned to the owner of the present invention, and the entirety of which is hereby incorporated by reference.

This application also claims priority to U.S. patent application Ser. No. 09/506,975, entitled "Method and System for Artificial Intelligence Lead Discovery Through Multi-Domain Clustering," filed by Christodoulos A. Nicolaou, Brian P. Kelley, Ruth F. Nutt, and Susan I. Bassett on Feb. 18, 2000, which is also assigned to the owner of the present invention, and the entirety of which is also hereby incorporated

by reference.

In addition, this application relates to the subject matter of U.S. Provisional Patent Application No. 60/120,701, entitled "Artificial Intelligence Directed Lead Discovery," filed by Susan I. Bassett, Andrew P. Dalke, John W. Elling, Brian P. Kelley, Christodoulos A. Nicolaou, and Ruth F. Nutt on Feb. 19, 1999, the entirety of which is also hereby incorporated by reference.

Claims

We claim:

1. A computer-operable method for identifying chemical substructures by analysis of a data set representing a plurality of chemical structures, the method comprising, in combination:

executing a computer program to pyramidally cluster representations of the chemical structures, so as to produce in a data storage medium a hierarchy of clusters each representing one or more of the chemical structures;

with respect to each of at least a plurality of the clusters of the hierarchy, analyzing the one or more chemical structures in the cluster and determining a chemical substructure representative of the one or more chemical structures in the cluster; and

outputting for consideration by a person a description of at least a portion of the hierarchy and an indication of at least one of the representative chemical substructures determined in the preceding step.

2. A method as claimed in claim 1, wherein executing a computer program to pyramidally cluster representations of the chemical structures comprises:

comparing clusters, and merging together pairs of clusters having a greatest similarity; and

determining, at a given level of the hierarchy, that at least two pairs of clusters have substantially the same similarity, and responsively merging each pair respectively, to thereby form at least two new clusters at a next level of the hierarchy.

3. A method as claimed in claim 1, wherein executing a computer program to pyramidally cluster the representations comprises applying a clustering algorithm, the method further comprising, before applying the clustering algorithm, receiving user input defining at least one aspect of the clustering algorithm.

4. A method as claimed in claim 3, wherein the at least one aspect of the clustering algorithm comprises an identification of the clustering algorithm.

5. A method as claimed in claim 3, wherein the at least one aspect of the clustering algorithm comprises a fuzziness parameter.

6. A computer-operable method for identifying pharmacophoric mechanisms through analysis of a plurality of molecules, each molecule having a respective feature characteristic and a respective activity characteristic, the method comprising, in combination:

(a) establishing in a computer memory a plurality of cluster objects, each cluster object representing one of the molecules;

(b) agglomeratively clustering the cluster objects based on comparisons of the feature characteristics of the molecules that the cluster objects represent, to build in the computer memory a hierarchical pyramid comprising a plurality of cluster objects, each cluster object of the pyramid representing a number of the molecules;

(c) with respect to each cluster object of a plurality of the cluster objects of the pyramid, identifying a substructure common to molecules represented the cluster object, each substructure defining a respective pharmacophoric mechanism; and

(d) outputting for viewing by a person a description of at least a portion of the hierarchical pyramid, including at least one substructure identified in step (c).

7. A method as claimed in claim 6, wherein, the step of agglomeratively clustering the cluster objects comprises:

to the extent any given cluster object is determined to be equidistant to a plurality of other cluster objects, merging the given cluster object with each cluster object of the plurality of other cluster objects.

8. A computer-operable method for identifying pharmacophoric mechanisms

through analysis of a plurality of molecules, each molecule defining a feature characteristic and an activity characteristic, the method comprising the following steps:

storing in a computer memory a plurality of data objects, each data object representing one of the molecules and having associated therewith a feature vector representing the feature characteristic of the molecule;

pyramidally clustering the data objects based on their associated feature vectors, to form in the computer memory a pyramidal data structure comprising a number of nodes, each node representing one or more of the molecules;

when pyramidally clustering the data objects, encountering a tie in proximity between a given node and at least two other nodes and responsively merging the given node separately with each of the at least two other nodes;

with respect to each node of the pyramidal data structure, identifying a chemical feature set common to the one or more molecules represented by the node, the chemical feature set defining a pharmacophore; and

providing an output describing at least a portion of the pyramidal data structure and including a description of the chemical feature set identified with respect to at least one node of the pyramidal data structure.

9. A computer-readable medium embodying a set of machine language instructions executable by a computer to identify pharmacophoric mechanisms through analysis of a plurality of molecules, each molecule defining a feature characteristic and an activity characteristic, wherein the machine language instructions are executable by the computer to perform functions comprising:

storing in a computer memory a plurality of data objects, each data object representing one of the molecules and having associated therewith a feature vector representing the feature characteristic of the molecule;

pyramidally clustering the data objects based on their associated feature vectors, to form in the computer memory a pyramidal data structure comprising a number of nodes, each node representing one or more of the molecules,

when pyramidally clustering the data objects, encountering a tie in proximity between a given node and at least two other nodes and responsively merging the

given node separately with each of the at least two other nodes;

with respect to each node of the pyramidal data structure, identifying a chemical feature set common to the one or more molecules represented by the node, the chemical feature set defining a pharmacophore; and

providing an output describing at least a portion of the pyramidal data structure and including a description of the chemical feature set identified with respect to at least one node of the pyramidal data structure.

10. A computer-operable method for learning pharmacophoric mechanisms through analysis of a plurality of molecules, each molecule having a respective feature characteristic and a respective activity characteristic, the method comprising, in combination:

(a) selecting from the plurality of molecules a group of molecules having at least a threshold activity characteristic;

(b) storing in a data storage medium a plurality of data objects each representing at least one of the molecules of the group, at least a first data object representing a plurality of molecules;

(c) measuring distances between the data objects based on the feature characteristics of the molecules represented by the data objects, and making a determination that the distance between the first data object and a second data object is substantially equal to the distance between the first data object and a third data object;

(d) in response to the determination, (i) storing in the data storage medium a fourth data object representing the molecules cooperatively represented by the first data object and the second data object and (ii) storing in the data storage medium a fifth data object representing the molecules cooperatively represented by the first data object and the third data object;

(e) identifying at least (i) a common feature set among the feature characteristics of the molecules represented by the first data object and (ii) a common feature set among the feature characteristics of the molecules represented by the fourth data object, whereby each common feature set defines a respective pharmacophoric mechanism; and

(f) providing to a person an indication of at least the common feature sets identified

with respect to the molecules of the first and fourth data objects.

11. A method as claimed in claim 10, further comprising representing each feature characteristic as a binary vector having members indicating the presence or absence of respective molecular features.

12. A method as claimed in claim 11, wherein measuring distances between the data objects comprises computing a distance between a pair the data objects based on the binary vectors of the molecules represented by the data objects of the pair.

13. A method as claimed in claim 11, wherein measuring distances between data objects comprises computing a Tanimoto distance between a pair of the data objects.

14. A method as claimed in claim 11, wherein measuring distances between data objects comprises computing a Euclidean distance between a pair of the data objects.

15. A method as claimed in claim 11, wherein representing each feature characteristic as a binary vector comprises generating and storing the binary vector.

16. A method as claimed in claim 11, further comprising:

determining an object activity characteristic representative of the activity characteristics of the molecules represented by the first data object; and

determining an object activity characteristic representative of the activity characteristics of the molecules represented by the fourth data object.

17. A method as claimed in claim 16, further comprising determining a differential between the object activity characteristics determined with respect to the first and fourth data objects.

18. A method as claimed in claim 17, further comprising providing to the person an indication of the differential, whereby the person may correlate the differential with the common feature set identified with respect to the first data object.

19. A method as claimed in claim 10, wherein measuring distances between the data objects comprises measuring a distance between the first data object and the second data object, and, wherein, measuring a distance between the first data object and the

second data object comprises applying a process selected from the group consisting of (i) Wards, (ii) complete-link, (iii) group average link, (iv) single link, and (v) centroid.

20. A method as claimed in claim 10, wherein storing in the data storage medium a plurality of data objects each representing at least one of the molecules of the group comprises developing a representation of each molecule and agglomeratively clustering the representations into the plurality of data objects.

21. A computer-operable method for analyzing a plurality of molecules, each molecule having a respective feature characteristic and a respective activity characteristic, wherein the respective activity characteristic of each molecule represents at least a threshold activity level, the method comprising, in combination:

(a) storing in a computer memory a plurality of cluster objects, each cluster object representing at least one of the molecules;

(b) conducting a merging process with respect to the cluster objects, the merging process comprising:

(i) comparing pairs of the cluster objects and, for each pair, measuring a respective dissimilarity between the cluster objects within the pair based on the feature characteristics of the molecules represented by the respective cluster objects;

(ii) of the dissimilarities measured in step (i), identifying a smallest dissimilarity;

(iii) selecting at least one pair of the cluster objects that has the smallest measured dissimilarity; and

(iv) with respect to each of the at least one pair selected in step (iii), merging the cluster objects of the pair to establish a cluster object cooperatively representing the molecules that were represented by the cluster objects of the pair;

(c) if at least two cluster objects have not yet been merged, then repeating step (b) with respect to the cluster objects that have not yet been merged;

(d) with respect to at least each cluster object established in step (b)(iv), identifying a common substructure among the molecules represented by the cluster object; and

(e) outputting a description of at least one cluster object established in step (b)(iv), wherein, the description of each of the at least one cluster object comprises a first portion indicating the common substructure identified in step (d) for the cluster object.

22. A method as claimed in claim 21, further comprising:

establishing for each molecule a feature vector representing the feature characteristic of the molecule,

wherein measuring a respective dissimilarity between the cluster objects within the pair based on the feature characteristics of the molecules represented by the respective cluster objects comprises comparing the feature vectors of molecules represented by the cluster objects of the pair.

23. A method as claimed in claim 22, wherein each feature vector is a bit-string.

24. A method as claimed in claim 22, wherein measuring a respective dissimilarity between the cluster objects within the pair comprises computing a Euclidean distance between the cluster objects within the pair.

25. A method as claimed in claim 22, wherein measuring a respective dissimilarity between the cluster objects within the pair comprises computing a Tanimoto distance between the cluster objects within the pair.

26. A method as claimed in claim 22, wherein measuring a respective dissimilarity between the clusters objects within the pair comprises applying a comparison process selected from the group consisting of (i) Wards, (ii) complete-link, (iii) group average link, (iv) single link, and (v) centroid.

27. A method as claimed in claim 26, wherein identifying a common substructure among the molecules represented by the cluster object comprises identifying a chemical structure present in all of the molecules in the group.

28. A method as claimed in claim 27, wherein the chemical structure comprises an arrangement of atoms and bonds.

29. A method as claimed in claim 28, wherein the arrangement of atoms and bonds is a contiguous arrangement.

30. A method as claimed in claim 21, wherein outputting a description of at least one cluster object established in step (b)(iv) comprises providing output data representing at least all of the cluster objects established in step (b)(iv).

31. A method as claimed in claim 21, wherein outputting a description of at least one cluster object established in step (b)(iv) comprises presenting to a person a graphical depiction of cluster objects, the graphical depiction including for each cluster object an indication of the common substructure established with respect to the cluster object.

32. A method as claimed in claim 21, wherein outputting a description of at least one cluster object established in step (b)(iv) comprises presenting to a person a graphical depiction of cluster objects, the graphical depiction including for each cluster object an indication of a measure of the activity characteristics of the molecules represented by the cluster object.

33. A method as claimed in claim 21, wherein outputting a description of at least one cluster object established in step (b)(iv) comprises displaying a tree structure having nodes reflecting the cluster objects established in step (b)(iv).

34. A method as claimed in claim 21, wherein the description of each of the at least one cluster object further comprises a second portion indicating a measure of the activity characteristics of the molecules represented by the cluster object.

35. A method as claimed in claim 21, further comprising measuring an activity differential between a cluster object established in step (b)(iv) and a cluster object merged into the cluster object established in step (b)(iv).

36. A method as claimed in claim 35, wherein measuring the activity differential comprises comparing a measure of the activity characteristics of the molecules represented by the cluster object established in step (b)(iv) with a measure of the activity characteristics of the molecules represented by the cluster object merged into the cluster object in step (b)(iv).

37. A method as claimed in claim 35, wherein the description further comprises a second portion indicating the measure of activity differential between a cluster object established in step (b)(iv) and a cluster object merged into the cluster object established in step (b)(iv).

38. A computer-readable medium embodying a set of machine language instructions

executable by a computer to analyze a plurality of molecules, each molecule having a respective feature characteristic and a respective activity characteristic, wherein the respective activity characteristic of each molecule represents at least a threshold activity level, wherein the machine language instructions are executable by the computer to perform functions comprising:

(a) storing in a computer memory a plurality of cluster objects, each cluster object representing at least one of the molecules,

(b) conducting a merging process with respect to the cluster objects, the merging process comprising:

(i) comparing pairs of the cluster objects and for each pair, measuring a respective dissimilarity between the cluster objects within the pair based on the feature characteristics of the molecules represented by the respective cluster objects;

(ii) of the dissimilarities measured in step (i), identifying a smallest dissimilarity,

(iii) selecting at least one pair of the cluster objects that has the smallest measured dissimilarity; and

(iv) with respect to each of the at least one pair selected in step (iii), merging the cluster objects of the pair to establish a cluster object cooperatively representing the molecules that were represented by the cluster objects of the pair;

(c) if at least two cluster objects have not yet been merged, then repeating step (b) with respect to the cluster objects that have not yet been merged;

(d) with respect to at least each cluster object established in step (b)(iv), identifying a common substructure among the molecules represented by the cluster object; and

(e) outputting a description of at least one cluster object established in step (b)(iv), wherein, the description of each of the at least one cluster object comprises a first portion indicating the common substructure identified in step (d) for the cluster object.

39. A processing system for screening a data set representing a plurality of molecules, in order to assist in identifying sets of molecular features that are likely to correlate with specified activity, the data set defining, for each represented molecule, a feature characteristic and an activity characteristic, the processing

system comprising, in combination:

at least one processor;

at least one data storage medium;

machine-language instructions stored in the at least one data storage medium and executable by the at least one processor to perform the following functions:

(a) storing in a computer memory a plurality of cluster objects, each cluster object representing one of the molecules;

(b) with respect to the cluster objects, conducting a merging process comprising:

(i) comparing pairs of the cluster objects and, for each pair, measuring a respective dissimilarity between the cluster objects within the pair, based on the feature characteristics of the molecules represented by the cluster objects of the pair;

(ii) of the dissimilarities measured in step (i), identifying a smallest dissimilarity;

(iii) selecting at least one pair of the cluster objects that has the smallest respective measured dissimilarity;

(iv) with respect to each pair of the at least one pair selected in step (iii), merging the cluster objects of the pair to establish a cluster object cooperatively representing the molecules that were represented by the cluster objects of the pair;

(c) if at least two cluster objects have not yet been merged, then repeating step (b) with respect to the cluster objects that have not yet been merged;

(d) with respect to at least each cluster object established in step (b)(iv) identifying a common substructure among the molecules represented by the cluster object; and

(e) outputting a description of at least one cluster object established in step (b)(iv), wherein, the description of each of the at least one cluster object comprises a first portion indicating the common substructure identified in step (d) for the cluster object.

40. A processing system as claimed in claim 39, wherein the computer memory is the data storage medium.

41. A processing system as claimed in claim 39, wherein the description of each of the at least one cluster object further comprises a second portion indicating a measure of the activity characteristics of the molecules represented by the cluster object.

Description

COPYRIGHT

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to computer-based analysis of data and to computer-based correlation of data features with data responses, in order to determine or predict which features correlate with or are likely to result in one or more responses. The invention is particularly suitable for use in the fields of chemistry, biology and genetics, such as to facilitate computer-based correlation of chemical structures with observed or predicted pharmacophoric activity. More particularly, the invention is useful in facilitating identification and development of potentially beneficial new drugs.

2. Description of Related Art

The global biotech and pharmaceutical industry is a $200 billion/year business. Most of the estimated $13 billion R&D spending in this industry is focused on discovering and developing prescription drugs. Current R&D effort is characterized by low drug discovery rates and long time-to-market.

In an effort to accelerate drug discovery, biotech and pharmaceutical firms are turning to robotics and automation. The old methods of rationally designing