

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
3 January 2003 (03.01.2003)

PCT

(10) International Publication Number  
**WO 2003/000849 A3**

- (51) International Patent Classification<sup>7</sup>: **G06N 1/00**
- (21) International Application Number: PCT/US2002/019686
- (22) International Filing Date: 21 June 2002 (21.06.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/299,911 21 June 2001 (21.06.2001) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/299,911 (CIP)  
Filed on 21 June 2001 (21.06.2001)
- (71) Applicant (for all designated States except US): **BIOIN-FORMATICS DNA CODES, LLC** [US/US]; 316 South Halsted Street, Suite 2w, Chicago, IL 60661 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **PANCOSKA, Petr** [US/US]; 901 Hinman Avenue #2C, Evanston, IL 60202 (US). **HOPFINGER, Anton, J.** [US/US]; 1780 Wilson Drive, Lake Forest, IL 60045 (US). **BENIGHT, Albert, S.** [US/US]; 1630 Valley View Drive, Schaumburg, IL 60193 (US). **RICCELLI, Peter, V.** [US/US]; 16830 Richards Drive, Tinley Park, IL 60477 (US).
- (74) Agent: **MYERS, P., Louis**; Fish & Richardson, P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).
- (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— with international search report  
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- (88) Date of publication of the international search report: 25 November 2004
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: METHODS FOR REPRESENTING SEQUENCE-DEPENDENT CONTEXTUAL INFORMATION PRESENT IN POLYMER SEQUENCES AND USES THEREOF

(57) Abstract: The invention includes methods of representing polymer sequences in a way that reveals important position-specific contextual information. The representations can be used to determine a number of properties of polymers, such as protein and nucleic acid sequences, including the identification of secondary domain structures, folding rate constants, and the effects of altering (e.g., mutating) monomers. In addition, the representations can be used to compare polymers and thereby identify important structural and functional characteristics of polymers.



WO 2003/000849 A3

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/19686

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) : G06N 1/00  
 US CL : 703/2

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
 U.S. : 703/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
 EAST

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	US 2001/0034580 A1 (SKOLNICK et al.) 25 October 2001, paragraphs 51, 57, 78-89.	1-6

Further documents are listed in the continuation of Box C.

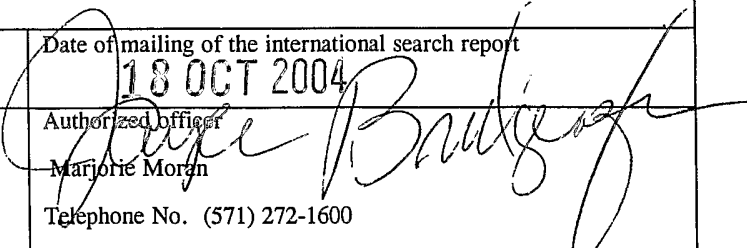
See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search  
 05 October 2004 (05.10.2004)

Date of mailing of the international search report  
**18 OCT 2004**

Name and mailing address of the ISA/US  
 Mail Stop PCT, Attn: ISA/US  
 Commissioner for Patents  
 P.O. Box 1450  
 Alexandria, Virginia 22313-1450  
 Facsimile No. (703) 305-3230

Authorized officer  
  
 Marjorie Moran  
 Telephone No. (571) 272-1600

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/19686

## Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claim Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claim Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3.  Claim Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Continuation Sheet

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.  As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-6

Remark on Protest  The additional search fees were accompanied by the applicant's protest.  
 No protest accompanied the payment of additional search fees.

**BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING**

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-6, drawn to a method of representing a polymer sequence.

Group II, claim(s) 7-8, drawn to a method of predicting the effects of changes in a protein sequence.

Group III, claim(s) 9, drawn to a method of predicting secondary structure boundaries in a protein.

Group IV, claim(s) 11-12, drawn to a method for predicting structural homologs of a protein.

Group V, claim(s) 13, drawn to a method of identifying positions of contextual similarity in a pair of polymers.

Group VI, claim(s) 14-15, drawn to a method of identifying positions of contextual similarity in a polymer.

Group VII, claim(s) 10 and 16-18, drawn to a method of identifying proteins with similar structural folds.

Group VIII, claim(s) 19-21, drawn to a method of identifying positions of contextual similarity in a pair of polymers, comprising method steps different from those of Group V.

Group IX, claim(s) 22-23, drawn to a method of representing a polymer sequence, comprising method steps different from those of Group I.

Group X, claim(s) 24, drawn to a method of characterizing secondary structure segments in a protein.

Group XI, claim(s) 25, drawn to a method of characterizing the contextual similarity of different positions in a polymer.

Group XII, claim(s) 26, drawn to a method of identifying contextually unique positions in a polymer.

Group XIII, claim(s) 27, drawn to a method of predicting the effects of mutations on the structure of a protein.

Group XIV, claim(s) 28, drawn to a method of identifying positions in a nucleic acid sequence.

The inventions listed as Groups I-XIV do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: A method of representing a polymer with position vectors is known in the art (see e.g. COHEN et al., US 5,878,373, columns 14-15), therefore the methods do not recite a special technical feature and are not linked by a single general inventive concept.