Amino Acid	Abbreviation	Code
Alanine	Ala	a
Arginine	Arg	r
Asparagine	Asn	n
Aspartic acid	Asp	d
Cysteine	Cys	c
Glutamine	Glu	q
Glutamic acid	Gln	e
Glycine	Gly	g
Histidine	His	h
Isoleucine	Ile	i
Leucine	Leu	l
Lysine	Lys	k
Methionine	Met	m
Phenylalanine	Phe	f
Proline	Pro	p
Serine	Ser	s
Threonine	Thr	t
Tryptophan	Trp	w
Tyrosine	Tyr	y
Valine	Val	v

FIG. 1: Amino Acid Encoding.

```
1: AAF04328. serine protease D...[gi:6137097]  BLink, Links

LOCUS       AAF04328                 422 aa            linear   PRI 20-APR-
2004
DEFINITION  serine protease DESC1 [Homo sapiens].
ACCESSION   AAF04328
VERSION     AAF04328.1  GI:6137097
DBSOURCE    accession AF064819.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 422)
  AUTHORS   Lang,J.C. and Schuller,D.E.
  TITLE     Differential expression of a novel serine protease homologue in
            squamous cell carcinoma of the head and neck
  JOURNAL   Br. J. Cancer 84 (2), 237-243 (2001)
  MEDLINE   21094880
   PUBMED   11161383
REFERENCE   2  (residues 1 to 422)
  AUTHORS   Lang,J.C. and Schuller,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-1998) Otolaryngology, Ohio State University,
1248
            James Cancer Hospital, 300 West 10th Avenue, Columbus, OH 43210,
            USA
COMMENT     Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4q12-q13; between D4S1619 and WI-7844; 20.21 cR from
                     WI-5548"
     Protein         1..422
                     /product="serine protease DESC1"
                     /name="expressed in normal oral epithelium, but not in
                     squamous cell tongue carcinoma or metastatic neck nodal
                     tissue"
     Site            167
                     /site_type="cleavage"
                     /note="between arginine and isoleucine"
     Site            208
                     /site_type="active"
                     /note="histidine; part of the catalytic triad"
     Site            253
                     /site_type="active"
                     /note="aspartic acid; part of the catalytic triad"
     Site            349
                     /site_type="active"
                     /note="serine; part of the catalytic triad"
     CDS             1..422
                     /gene="DESC1"
                     /coded_by="AF064819.1:56..1324"
ORIGIN
        1 myrpdvvrar krvcwepwvi glvifisliv lavcigltvh yvryngkkty nyystlsftt
       61 dklyaefgre asnnftemsq rlesmvknaf yksplreefv ksqvikfsqq khgvlahmll
      121 icrfhstedp etvdkivqlv lheklqdavg ppkvdphsvk ikkinktetd sylnhccgtr
      181 rsktlgqslr ivggteveeg ewpwqaslqw dgshrcgatl inatwlvsaa hcfttyknpa
      241 rwtasfgvti kpskmkrglr riivhekykh pshdydisla elsspvpytn avhrvclpda
      301 syefqpgdvm fvtgfgalkn dgysqnhlrq aqvtlidatt cnepqaynda itprmlcags
      361 legktdacqg dsggplvssd ardiwylagi vswgdecakp nkpgvytrvt alrdwitskt
      421 gi
//
```

FIG. 2: Cancer Protein Description.

Nucleic Acid Base Triplets	Corresponding Amino Acid	Code
GCT GCC GCA GCG	Alanine	a
CGT CGC CGA CGG AGA AGG	Arginine	r
AAT AAC	Asparagine	n
GAT GAC	Aspartic acid	d
TGT TGC	Cysteine	c
CAA CAG	Glutamine	q
GAA GAG	Glutamic acid	e
GGT GGC GGA GGG	Glycine	g
CAT CAC	Histidine	h
ATT ATC ATA	Isoleucine	i
TTG TTA CTT CTC CTA CTG	Leucine	l
AAA AAG	Lysine	k
ATG	Methionine	m
TTT TTC	Phenylalanine	f
CCT CCC CCA CCG	Proline	p
TCT TCC TCA TCG AGT AGC	Serine	s
ACT ACC ACA ACG	Threonine	t
TGG	Tryptophan	w
TAT TAC	Tyrosine	y
GTT GTC GTA GTG	Valine	v
TAA TAG TGA	< stop code >	

FIG. 3: Nucleic Acid Base – Amino Acid Correspondence.

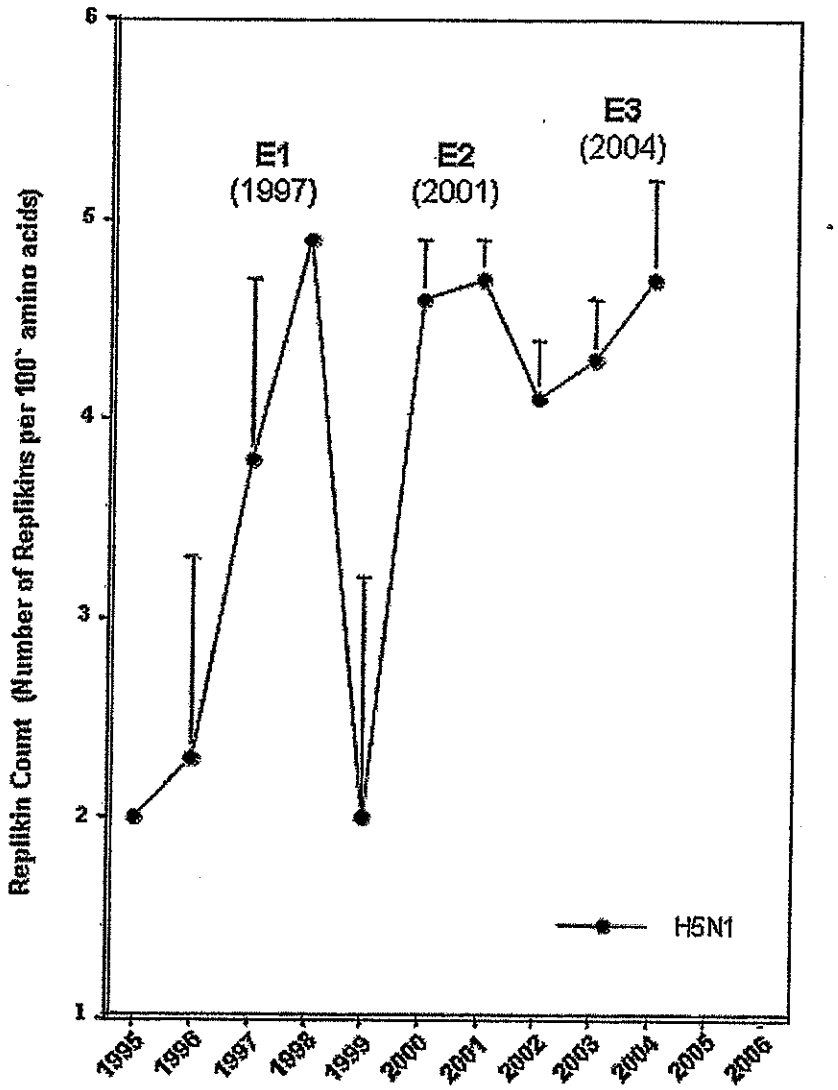


FIG. 4: Increasing Replikin Concentration ("Replikin Count") of Hemagglutinin Protein of H5N1 Prior to Three 'Bird Flu' Epidemics.

Category	SEQ. ID.	Name	Replikin Pattern
Bacteria:	10	Mycoplasma pulmonic, chromosome replication	kkektthnk
	43	Macrophage infectivity potentiator, L. legionella	kvhffqlkk
Tumor Viruses:	48	Rous sarcoma virus tyrosine-protein kinase	kklrhek
	49	v-yes, avian sarcoma	kklrhdk
	50	c-yes, colon cancer, malignant melanoma	kklrhdk
	51	v-srcC, avian sarcoma	kklrhek
	52	c-src, colon, mammary, panrcreatic cancer	kklrhek
	53	Neuroblastoma RAS viral (v-ras) oncogene	kqahelak
	54	VPI (major capsid protein) [Polyamavirus sp.]	kthrfskh
	55	Sindbis	knlhekik
	56	El [Human papilloamavirus type 71]	khrpllqlk
	57	v-erbB from AEV and c-erb	kspnhvk
	58	v-fms (feline sarcoma)	knihlekk
	59	c-fms (acute and chronic myelomonocytic tumors)	knihlekk
	60	large t-antigen I [Polyomavirus sp.l	kphlaqslek
	61	middle t-antigen [Polyomavirus sp,l-	kqhrelkdk
	62	small t-antigen [Polyomavirus spJ,	kqhrelkdk
	63	v-abl, murine acute leukemia	kvpvlisptlkh
	64	Human T-cell lymphotropic virus typo 2	kslllevdkdish
	65	c-kit, GI tumors, small cell lung carcinoma	kagitimvkreyh
	18	Hepatitis C	hyppkpgcivpak
Trans-forming Proteins:	66	Transforming protein myb	ksgkhlgk
	67	Transforming protein myc, Burkitt lymphoma	krreqlkhk
	68	Ras-related GTP-binding protein	ksfevikvih
	69	Transforming protein ras (teratocarcinoma)	kkkhtvkk
	70	TRAF-associated NF●kB activator TANK	kaqkdhlsk
	71	RFP transforming protein	hlkrvkdlkk
	72	Transforming protein D (S.C.)	kygspkhrlik
	73	Papilloma virus type 11, transforming protein	klkhilgkarfik
	74	Protein tryosine kinasc (EC 2.7.1.ll2slk	kgdhvkhykirk
	75	Transforming protein (axl(-))	keklrdvmvdrhk
	76	Transforming protein (N-myc)	kdqarqqqllkkieh
	77	Fibroblast growth factor 4 (Kaposi sarcoma)	kkgnrvsptmkvth
Cancer Cell Proteins:	78	Matrix metaloproteinase 7 (uterine)	keiplhfrk
	79	Transcription factor 7-like	kkkphikk
	80	Breast cancer antigen NY-BR-87	ktrhdplak
	81	BRCA-1-Associated Ring Domain Protein (breast)	khhpkdnlik
	82	Autoantigen from a breast tumor'	khkrkkfrqk
	83	Glioma Replikin (this study)	kagvaflhkk
	84	Ovarian cancer antigen	khkrkkfrqk
	85	EE L leukemia	kkkskkhkdk
	86	Proto-oncogene tyrosine-protein kinase C-ABLE	hksekpalprk
	87	Adenomatosis polyposis coli	kkkkpsrlkgdnek
	88	Gastric cancer transforming protein	ktkkgnrvsptmkvth
	89	Transforming protein (K-RAS 2B),lung	khkekmskdgkkkkkksk

FIG. 5 Selected examples of Replikins in various organisms.

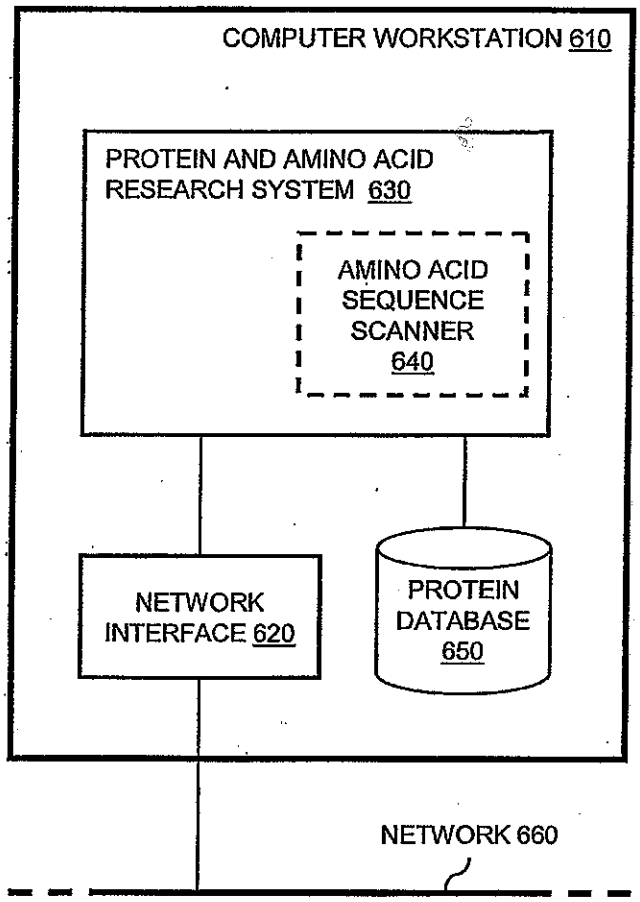

COMPUTER WORKSTATION 610
PROTEIN AND AMINO ACID RESEARCH SYSTEM 630
AMINO ACID SEQUENCE SCANNER 640
NETWORK INTERFACE 620
PROTEIN DATABASE 650
NETWORK 660

FIG. 6

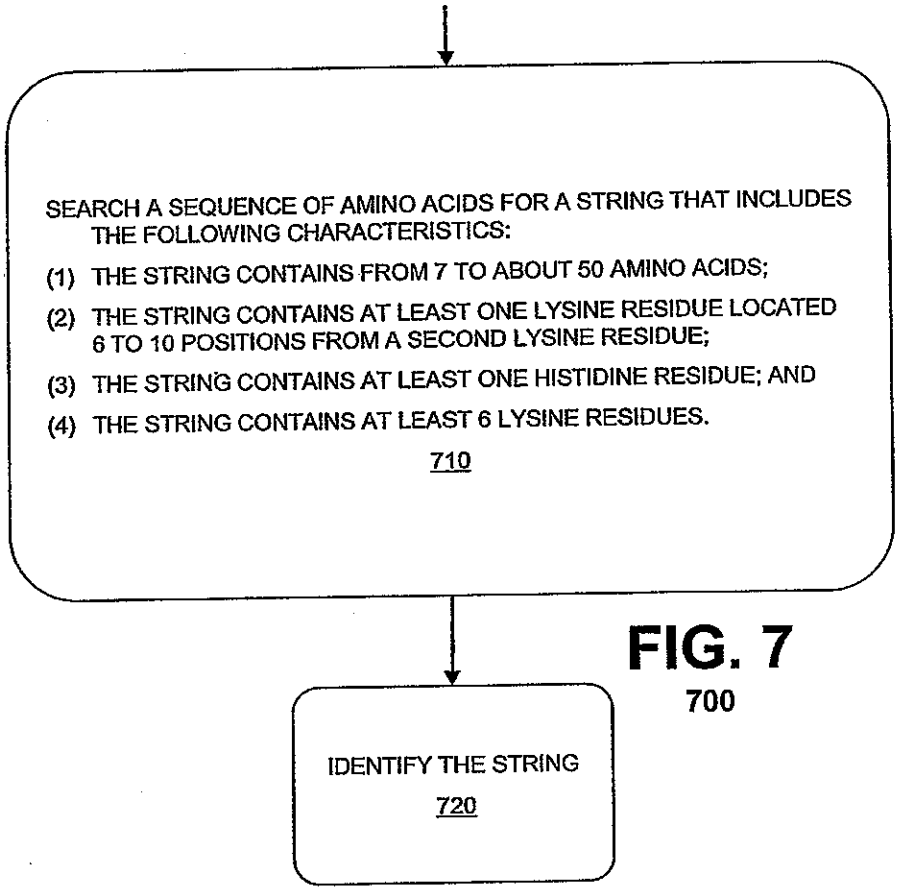

SEARCH A SEQUENCE OF AMINO ACIDS FOR A STRING THAT INCLUDES THE FOLLOWING CHARACTERISTICS:
(1) THE STRING CONTAINS FROM 7 TO ABOUT 50 AMINO ACIDS;
(2) THE STRING CONTAINS AT LEAST ONE LYSINE RESIDUE LOCATED 6 TO 10 POSITIONS FROM A SECOND LYSINE RESIDUE;
(3) THE STRING CONTAINS AT LEAST ONE HISTIDINE RESIDUE; AND
(4) THE STRING CONTAINS AT LEAST 6 LYSINE RESIDUES.
710
IDENTIFY THE STRING
720

FIG. 7
700

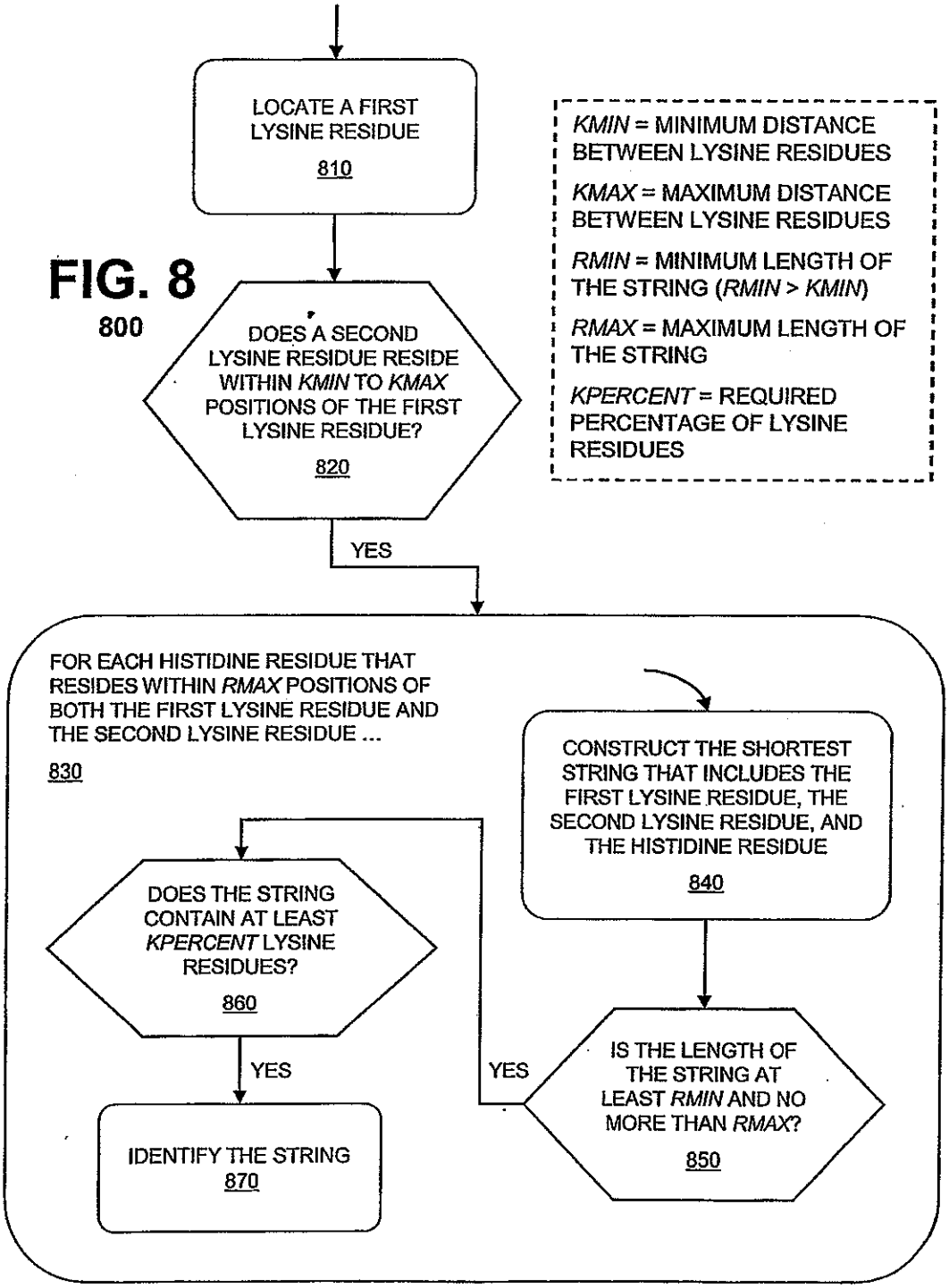

LOCATE A FIRST LYSINE RESIDUE
810
DOES A SECOND LYSINE RESIDUE RESIDE WITHIN KMIN TO KMAX POSITIONS OF THE FIRST LYSINE RESIDUE?
820
KMIN = MINIMUM DISTANCE BETWEEN LYSINE RESIDUES
KMAX = MAXIMUM DISTANCE BETWEEN LYSINE RESIDUES
RMIN = MINIMUM LENGTH OF THE STRING (RMIN > KMIN)
RMAX = MAXIMUM LENGTH OF THE STRING
KPERCENT = REQUIRED PERCENTAGE OF LYSINE RESIDUES
YES
FOR EACH HISTIDINE RESIDUE THAT RESIDES WITHIN RMAX POSITIONS OF BOTH THE FIRST LYSINE RESIDUE AND THE SECOND LYSINE RESIDUE ...
830
CONSTRUCT THE SHORTEST STRING THAT INCLUDES THE FIRST LYSINE RESIDUE, THE SECOND LYSINE RESIDUE, AND THE HISTIDINE RESIDUE
840
IS THE LENGTH OF THE STRING AT LEAST RMIN AND NO MORE THAN RMAX?
850
YES
DOES THE STRING CONTAIN AT LEAST KPERCENT LYSINE RESIDUES?
860
YES
IDENTIFY THE STRING
870

FIG. 8
800

```
#
#   Discover a subsequence h...k...k, k...h...k, or k...k...h such that
#   (1) The distance between ks is in the range kmin..kmax.
#   (2) The distance between an h and the farthest k is in the range kmin+1..hmax.
#   (3) The fraction of k in the subsequence is percent or larger.
#
#   The sequence is searched for all possible subsequences that match,
#   and all these subsequences are returned.

set kmin 6; set kmax 10
set hmax 50
set percent 6

proc match {sequence} {
    global kmin kmax hmax percent

    set pos 0
    set L {}
    array set F {}
    foreach e [regexp -all -indices -inline k $sequence] {
        lappend L [lindex $e 0]
    }

    for {set i 1} {$i<[llength $L]} {incr i} {
        set k0 [lindex $L [expr {$i-1}]]

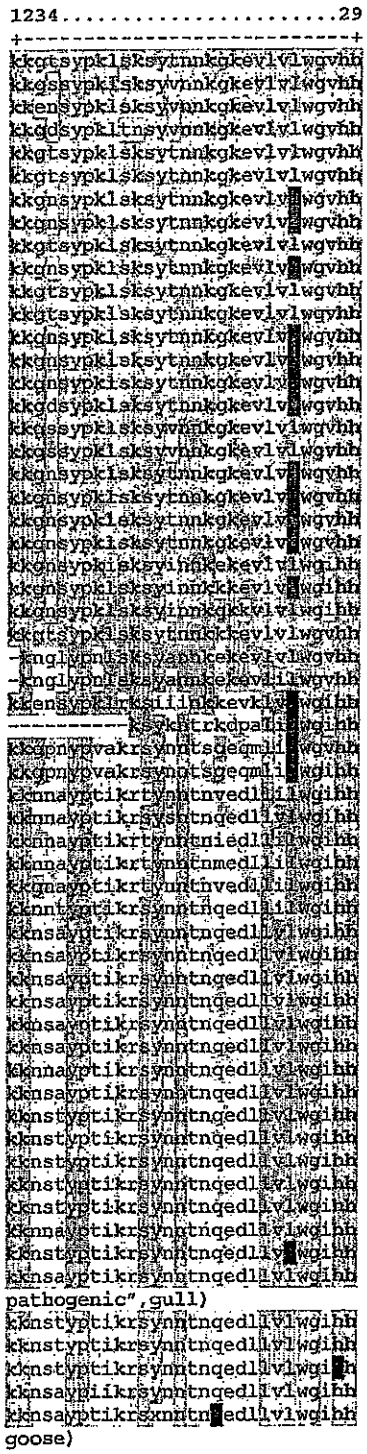
        #  rule 1.
        for {set j $i; set wideenough 0} {!$wideenough && $j<[llength $L]} {incr j} {
            set k1 [lindex $L $j]
            if {$k1-$k0<$kmin} continue
            if {$k1-$k0>$kmax} break

            #  rule 2.
            set offset [expr $k1-$hmax]
            if {$offset<0} {set offset 0}
            while 1 {
                set h [string first h $sequence $offset]
                if {$h<0 || $h>$k0+$hmax} break
                if {$h<$k0} {
                    set b $h
                } else {
                    set b $k0
                }
                if {$h>$k1} {
                    set e $h
                } else {
                    set e $k1
                }

                #  rule 3.
                set subsequence [string range $sequence $b $e]
                set nk [regexp -all k $subsequence]
                if {double($nk)/double([string length $subsequence])*100>=$percent} {
                    set "F($b $e)" 1
                }

                incr offset
            }
        }
    }
    lsort -integer -index 0 [array names F]
}
```

FIG. 9 Replikin Matching Code.



1917	H1N_	Influenza Goose Replikin
1918	H1N1	Human Influenza Pandemic
1930	H1N1	
1933	H0N1	
1976	H1N1	
1977	H1N1	
1979	H1N1	
1980	H1N1	
1980	H1N1	
1981	H1N1	
1981	H1N1	
1985	H1N1	
1991	H1N1	
1992	H1N1	
1996	H1N1	
1996	H1N1	
1997	H1N1	
1998	H1N1	
1999	H1N1	
2000	H1N1	
2001	H1N1	
2002	H1N1	
1999	H1N2	Influenza
2000	H1N2	
2001	H1N2	
2001	H1N2	
2002	H1N2	
2002	H1N2	
1968	H3N2	Human Influenza Pandemic
1979-2003	H7N7	Influenza
1957	H2N2	Human Influenza Pandemic
1957	H2N2	Human Influenza Pandemic
2002	H5N2	Influenza
1959	H5N1	Influenza (Scotland)
1975	H5N1	(Wisconsin)
1981	H5N1	(Minnesota)
1983	H5N1	(Pennsylvania)
1988	H5N1	(Scotland)
1996	H5N1	(China)
1997	H5N1	(China)
1998	H5N1	(China)
1999	H5N1	(China)
2000	H5N1	(China)
2001	H5N1	(China)
2001	H5N1	(China)
2002	H5N1	(China)
2002	H5N1	(Thailand)
2002	H5N1	(Vietnam)
2003	H5N1	(Vietnam)
2003	H5N1	(Thailand)
2003	H5N1	(Sindong, China)
2003	H5N1	(China)
2004	H5N1	(Vietnam, highly pathogenic)
2004	H5N1	(Vietnam,"highly pathogenic",gull)
2004	H5N1	(Vietnam, highly pathogenic)
2004	H5N1	(Thailand, highly pathogenic)
2004	H5N1	(Thailand, highly pathogenic)
2004	H5N1	(China,highlypathogenic)
2004	H5N1	(China,"highly pathogenic", goose)

FIG. 10 Replikin Scaffolding.

Residues identical to original 1917 Goose Replikin residues are shown in medium grey. Amino acid substitutions in light grey and [redacted].

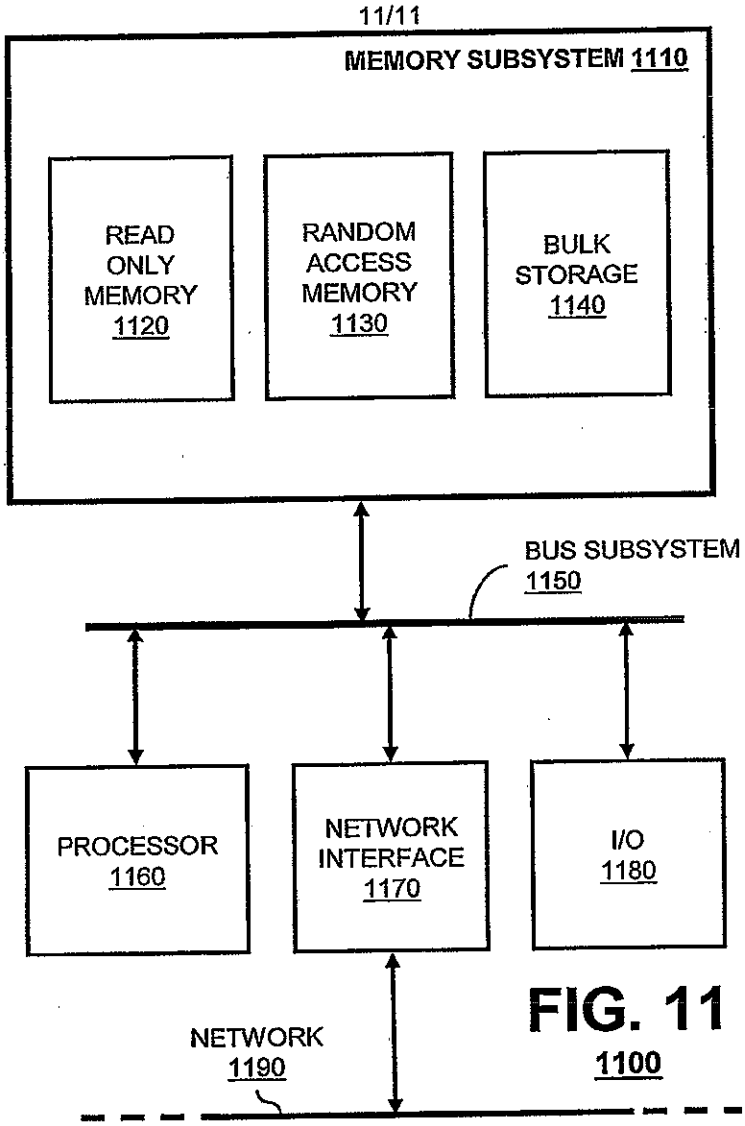

MEMORY SUBSYSTEM 1110
READ ONLY MEMORY 1120
RANDOM ACCESS MEMORY 1130
BULK STORAGE 1140
BUS SUBSYSTEM 1150
PROCESSOR 1160
NETWORK INTERFACE 1170
I/O 1180
NETWORK 1190

FIG. 11
1100