Biologics: Targets & Therapy

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Describing the hexapeptide identity platform between the influenza A H5N1 and *Homo sapiens* proteomes

This article was published in the following Dove Press journal:
Biologics: Targets & Therapy
8 September 2010
Number of times this article has been viewed

Darja Kanduc

Department of Biochemistry and Molecular Biology, University of Bari, Italy

Abstract: We searched the primary sequence of influenza A H5N1 polyprotein for hexamer amino acid sequences shared with human proteins using the Protein International Resource database and the exact peptide matching analysis program. We find that the viral polyprotein shares numerous hexapeptides with the human proteome. The human proteins involved in the viral overlap are represented by antigens associated with basic cell functions such as proliferation, development, and differentiation. Of special importance, many human proteins that share peptide sequences with influenza A polyprotein are antigens such as reelin, neurexin I-α, myosin-IXa, Bardet–Biedl syndrome 10 protein, Williams syndrome transcription factor, disrupted in schizophrenia 1 protein, amyotrophic lateral sclerosis 2 chromosomal region candidate gene 17 protein, fragile X mental retardation 2 protein, and jouberin. That is, the viral-vs-human overlap involves human proteins that, when altered, have been reported to be potentially associated with multiple neurological disorders that can include autism, epilepsy, obesity, dystonia, ataxia–telangiectasia, amyotrophic lateral sclerosis, sensorineural deafness, sudden infant death syndrome, Charcot-Marie-Tooth disease, and myelination. The present data are discussed as a possible molecular basis for understanding influenza A viral escape from immunosurveillance and for defining anti-influenza immune-therapeutic approaches devoid of collateral adverse events.

Keywords: peptide sharing, neurological disorders, host-pathogen relationships, viral escape from immunosurveillance

Introduction

Many fundamental questions remain unanswered in virology. Virus-associated pathologies and phenomena (such as tissue tropism, chronic infection, the viral role in cancer) are yet to be explained. Likewise, we remain ignorant of the basic mechanisms that regulate viral quiescence and activation.[1,2] Moreover, as a fundamental priority step in any possible therapeutic approach, we need to solve the enigma of successful virus escape from immune surveillance.[3,4]

In order to understand the molecular basis of viral immunogenicity and pathogenesis, we analyze, quantitatively and qualitatively, the peptide-sharing platform between viruses and human proteins.[5–7] We use short peptide motifs as scanning units because 5–6 amino acids define functional modules with a central role in cell biology and immunology.[8] Indeed, scientific reports indicate that 5–6 amino acids is a sufficient minimal determinant in antigen–antibody recognition.[8,9] Also, it is known that the optimal T-cell epitope length ranges between 9 and 15 amino acids, with the central 5–6 amino acid residues playing a major role in establishing specific immune interactions.[10] Thus, according to our previous studies too,[11] the structural features of an immunogenic T-cell epitope may

Correspondence: Darja Kanduc
Department of Biochemistry and Molecular Biology, University of Bari, Bari 70126, Italy
Tel +39 080 544 3321
Fax +39 080 544 3321
Email d.kanduc@biologia.uniba.it; dkanduc@gmail.com

be located in a central immune unit formed by 5–6 residues flanked by MHC binding terminal residues.[12–15]

In the present study, we investigate the influenza A H5N1 polyprotein-vs-human proteome overlap at the hexapeptide level. We report that viral matches occur in numerous human proteins associated with fundamental crucial functions. In particular, we describe the viral hexapeptide matches occurring in proteins specifically expressed in the central nervous system (CNS).

Methods

The influenza A virus (A/Goose/Guangdong/1/96/H5N1, polyprotein length: 4,426 aa; 10 proteins) is described at the URL www.expasy.org/viralzone/all_by_species/130.html. Viral proteins, abbreviations, UniProtKB/Swiss-Prot accession numbers, and entry names are as follows: nonstructural protein 1 (NSP1, Q9Q0L6, NS1_I96A0); nuclear export protein (NEP, Q9Q0L7, NEP_I96A0); matrix protein 1 (MP1, Q9Q0L8, M1_I96A0); matrix protein 2 (MP2, Q9Q0L9, M2_I96A0); hemagglutinin (HA, Q9Q0U6, HEMA_I96A0); neuraminidase (NA, Q9Q0U7, NRAM_I96A0); nucleocapsid protein (NC, Q9Q0U8, Q9Q0U8_I96A0); polymerase acidic protein (PA, Q9Q0U9, PA_I96A0); ribonucleic acid (RNA)-directed RNA polymerase catalytic subunit (RDRP, Q9Q0V0, RDRP_I96A0); and polymerase basic subunit 2 (PB2, Q9Q0V1, PB2_I96A0).

The influenza A polyprotein primary sequence was dissected into hexapeptides that were analyzed for exact matching with the human proteome using the Protein Information Resource perfect match program (pir.georgetown.edu/pirwww/search/peptide.shtml).[16] The hexapeptide sequences were offset by one residue, ie, overlapping by 5 residues, MDSNTI, DSNTIT, SNTITS, NTITSF, etc. The human proteome consisted of 36,103 proteins and 8,247,275 unique hexamers at the time of analysis. Human protein sequences corresponding to fragments, duplicates, and obsolete entries were not considered. Functions of human proteins and potential disease associations were derived from the Universal Protein Resource (uniprot.org/uniprot). As a control, the primary sequence of tobacco mosaic virus (TMV) coat protein (UniProtKB/Swiss-Prot P69687, COAT_TMV) was used.

Results

Quantitative analysis of the hexapeptide overlapping between influenza A polyprotein and human proteomes

Figure 1 describes the influenza A overlapping pattern to the human proteome at the hexapeptide level (ie, the number of occurrences in the human proteome for each viral hexapeptide). It can be seen that all the 10 influenza proteins, regardless of their functions or lengths, present perfect hexapeptide matches to human proteins. Moreover, Figure 1 shows that the overlapping pattern has a similar behavior in the 10 influenza A proteins, with viral peptide areas scarcely represented in the human proteins alternating with viral peptide areas highly repeated in human proteins.

Numerically, the viral-vs-human overlap is defined by 1,905 perfect matches (Table 1). Theoretically, the number of times a given viral hexamer might occur at random in the human proteome (calculated on the basis of the unique hexamers in viral and human proteomes) is 569.7. Therefore, the overlapping extent reported in Table 1 is about 3.3-fold higher than the theoretical value.

Qualitative analysis of the hexapeptide overlap of influenza A polyprotein vs human proteins

The number of viral hexapeptide matches found in human proteins is high, and only a short survey is reported in Table 2,

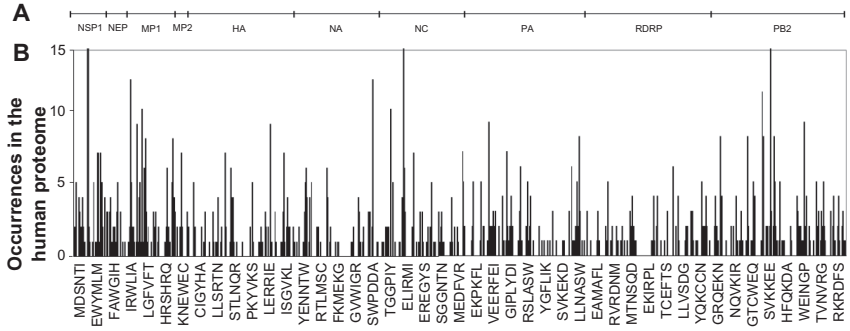

Figure 1 Identity platform shared by influenza A polyprotein and the human proteome at the hexapeptide level. **A**) Location of the 10 viral proteins along the primary sequence of the influenza A polyprotein. Viral protein abbreviations as detailed in Methods section. **B**) Similarity profile of influenza A H5N1 polyprotein vs the human proteome at the hexapeptide level. The columns indicate the number of viral hexapeptide occurrences in the human proteome. Hexapeptide amino acid sequences in one-letter code.

Table 1 Numerical description of the hexapeptide overlap between influenza A polyprotein and *Homo sapiens* proteome

Unique viral hexamers	4,416
Viral hexamers occurring in the human proteins (including multiple occurrences)[a]	1,905
Expected number of viral hexamer occurrences in the human proteome	569

[a]Human protein sequences corresponding to fragments, duplicates, and obsolete entries were filtered out.

describing 74 viral hexapeptides occurring in 82 human proteins. However, Table 2 clearly shows that the human proteins involved in the viral hexapeptide overlap exert key cellular functions, and many of them have a specific localization in the brain. The functional relevance of the human proteins sharing viral hexapeptides is illustrated in the following examples: Alstrom syndrome protein 1 (ALMS1), involved in sensory transduction and expressed in fetal brain; Laforin-interacting protein, associated with progressive myoclonic epilepsy type 2; chloride channel protein 2 (CLCN2), defects in which may be the cause of epilepsy or generalized tonic–clonic seizures in childhood or adolescence; Niemann–Pick C1-like protein 1 (NPCL1), alterations of which may be associated with Niemann–Pick C disease and characterized by cognitive impairment and ataxia; Huntingtin protein and Huntingtin-interacting protein 3; the paroxysmal nonkinesiogenic dyskinesia (PNKD) protein, defects of which are the cause of dystonia type 8; gemin-4, a component of the complex survival of motor neuron (SMN) that is associated with the neurodegenerative disease spinal muscular atrophy; and dehydrogenase enzymes, such as the mitochondrial glutamate dehydrogenase 1 (involved in learning and memory and, when altered, causes hyperinsulinism–hyperammonemia syndrome) and glutamate dehydrogenase 2 (involved in neurotransmission).

Interestingly, viral hexapeptides are found in human proteins that have been specifically associated with sudden death, including sudden infant death syndrome, ie, trifunctional enzyme subunit β (ECHB), ryanodine receptor (RYR) 2 or cardiac muscle-type ryanodine receptor, caveolin-3 (CAV3), and ankyrin-2 (ANK2) or brain ankyrin (Table 2). Of special note, titin, that has been linked to cardiac failure and sudden cardiac death,[17] presents 14 exact matches in common with influenza A H5N1 polyprotein (ie, the matches along the viral polyprotein are, in the order: LKANFS, SDSSDP, SYIVEK, PKSSWS, AELLVL, DPKKTG, PKKTGG, SGAAGA, GAAGAA, SAAFED, LSIEDP, ELQDIE, TNDAGS, KEEEVL).

On the whole, a result emerging from the analysis given in Table 2 is that the viral-vs-human overlap is physiologically complex with respect to the human proteins involved. As a paradigmatic snapshot, the viral PA protein shares the GVEEGS sequence (aa 627–632) with the human Msx2-interacting protein (hMINT or SMART/histone deacetylase [HDAC1]-associated repressor protein). Human MINT is an essential corepressor protein, which probably regulates different key pathways, such as the Notch pathway. The protein is expressed at high levels in brain, testis, spleen, and thymus and is induced by the 17β-estradiol hormone. The (inter) actions exerted by hMINT are physiologically important. Indeed, Notch dysregulation may underlie some of the distinctive clinical and pathologic features of tuberous sclerosis complex, a multisystem tumor suppressor gene syndrome characterized by neurologic disorders (seizures, mental retardation, autism); pulmonary lymphangioleyomyomatosis; and hamartomatous tumors of the brain, heart, kidney, and skin.[18,19]

Quantitatively, in a few instances, the viral-vs-human overlapping involves multiple occurrences of the same hexapeptide. Table 2 shows that the viral DSNTIT hexapeptide present in the human ALMS1 is present in the influenza A NSP1 and NEP. Likewise, multiple occurrences of a viral hexapeptide may occur in the same human protein. For example, the influenza A $RDRP_{510–515}$PSFGVS hexapeptide occurs 17 times in the human AHNAK2 protein, while the influenza A $PA_{118–123}$IEIGVT peptide occurs 3 times in human neuralized-like protein 4. Also, tubulin and sodium channel protein type 10 have a heptapeptide in common with the influenza A polyprotein, while human ataxia–telangiectasia mutated protein (ATM) has 2 different viral hexapeptides (corresponding to $NC_{375–380}$DSSTLE and $PA_{626–631}$KGVEEG sequences). Similar quantitative and qualitative data were obtained by analyzing the primary sequence of proteins from influenza A H1N1 virus, strain Alabama/02/2009, National Center for Biotechnology Information (NCBI) Taxonomy ID: 645728 (data not shown).

Analysis of the hexapeptide overlap of influenza A vs human proteins associated with behavioral syndromes and crucial function(s) in the brain and/or in the peripheral nervous system

An analysis of Table 2 shows that a number of proteins containing viral hexapeptides are specifically located in selected brain areas, such as the amygdala, substantia nigra, caudate nucleus, thalamus and hypothalamus, hippocampus, and cerebellum (mentioned in boldface in Table 2). Following

Table 2 Examples of human proteins sharing hexapeptide sequence(s) with influenza A H5N1 polyprotein

Influenza A hexapeptide			Human protein description[a]
Protein	**Aa pos**	**Sequence**	
NSP1	2	DSNTIT[b]	**ALMS1:** Alstrom syndrome protein 1. Involved in sensory transduction of extracellular signals, such as light, taste, sound, touch, or smell, into electric signals. Expressed in the pancreas β-cells of the islets in fetal aorta and **the brain.** Defects in ALMS1 are the cause of Alstrom syndrome, characterized by cone-rod retinal dystrophy, neurosensory hearing loss, early childhood obesity, and type 2 diabetes mellitus.
	42	ALKGRG	**EPMIP:** EPM2A interacting protein 1. Laforin-interacting protein. The first recognized Laforin binding partner that may play a critical role in discovering the underlying pathogenesis of progressive myoclonic epilepsy type 2 (Lafora disease), characterized by grand mal seizures and/or myoclonus at about 15 years of age. Mental deterioration follows, often with psychotic features. **Q96A48:** Paroxysmal nonkinesiogenic dyskinesia (PNKD) protein. Myofibrillogenesis regulator 1. PNKD plays a role in the development of cardiac hypertrophy. **Isoform 1 is only expressed in the brain**. Isoform 2 is ubiquitous. Defects in PNKD cause dystonia type 8, a paroxysmal nonkinesigenic dystonia/dyskinesia.
	47	GSTLGL	**GRIP2:** Glutamate receptor-interacting protein 2. Glutamate receptors mediate neurotransmission in the brain and play an important role in neural plasticity, neural development, and neurodegeneration. GRIP1 and GRIP2 are **widely expressed in brain, with the highest levels in the cerebral cortex, hippocampus, and olfactory bulb.** GRIP1 is expressed in early development, specifically postnatal days 8–10, while GRIP2 is expressed during later developmental stages.
	50	LGLDLR	**DHE3:** Glutamate dehydrogenase 1, mitochondrial. Involved in learning and memory. Defects in DHE3 are the cause of hyperinsulinism–hyperammonemia syndrome. **DHE4:** Glutamate dehydrogenase 2, mitochondrial. Important for recycling the chief excitatory neurotransmitter, glutamate, in neurotransmission. Expressed in retina, testis, and brain. **NPCL1:** Niemann–Pick C1-like protein 1. Plays a major role in cholesterol homeostasis. Niemann–Pick C disease is characterized by a disruption of sphingolipid and cholesterol trafficking that produces cognitive impairment, ataxia, and death, often in childhood.
	77	LKIAIA	**P04114:** Apolipoprotein B (APOB)-100. The APOB is a major protein constituent of chylomicrons. Defects in APOB are a cause of familial hypobetalipoproteinemia. Subjects have low plasma low-density lipoprotein (LDL) cholesterol and APOB-100 concentrations. Clinical presentation varies from no symptoms to severe gastrointestinal and neurological dysfunction similar to abetalipoproteinemia.
	82	ASSPAP	**CTRO:** Citron. Rho-interacting, serine/threonine-protein kinase 21. Required for KIF14 localization to the central spindle and midbody. Regulates the development of the CNS. **NCOR1:** Nuclear receptor corepressor 1. Mediates transcription repression of thyroid hormone and retinoic acid receptors. The corresponding gene is located between the CMT and Smith–Magenis syndrome critical regions on chromosome 17. Forms a large corepressor complex that contains SIN3A/B and histone deacetylases HDAC1 and HDAC2.
	83	SSPAPR	**ITPR1:** Inositol 1,4,5-trisphosphate receptor (IP3R) type 1. The IP3R mediates calcium release following stimulation by inositol 1,4,5-trisphosphate. Interacts with RYR1, RYR2, ITPR1, SHANK1, and SHANK3. Defects in ITPR1 are the cause of spinocerebellum ataxia type 15, with incoordination of gait, poor coordination of hands, speech, and eye movements, due to cerebellar degeneration.
	107	PRQKIT	**PMM2:** Phosphomannomutase 2. Involved in the synthesis of GDP-mannose and dolichol-phosphate-mannose. Defects in PMM2 are the cause of Jaeken syndrome, characterized by severe encephalopathy with axial hypotonia, abnormal eye movement, psychomotor retardation, peripheral neuropathy, cerebellar hypoplasia, and retinitis pigmentosa.
	141	LETLVS	**MLL3:** Myeloid/lymphoid or mixed-lineage leukemia protein 3. Histone-lysine N-methyltransferase MLL3. Methylates 'Lys-4' of histone H3. H3 'Lys-4' methylation represents a specific tag for epigenetic transcriptional activation. Highly expressed in testis and ovary, followed by brain and liver. **Within brain, expression is highest in the hippocampus, caudate nucleus, and substantia nigra**.
	178	IGILIG	**IRAK3:** Interleukin-1 receptor-associated kinase 3. Expressed in peripheral blood lymphocytes. Defects in IRAK3 are associated with susceptibility to asthma-related traits type 5, which include clinical symptoms such as coughing, wheezing, dyspnea, serum IgE levels, atopy, and atopic dermatitis.
	205	IHDENG	**Q9HC59:** The α-synuclein-interacting protein (SNCAIP). Synphilin-1c protein. Widely expressed, with **highest levels in the brain,** heart, and placenta. Defects in SNCAIP are a cause of Parkinson disease, characterized by bradykinesia, resting tremor, muscular rigidity, loss of dopaminergic neurons in the substantia nigra, and intraneuronal accumulations of aggregated proteins (ie, Lewy bodies) in surviving neurons.
NEP	18	KMQLES	**GLRB:** Glycine receptor subunit β. A neurotransmitter-gated ion channel. Binding of glycine to its receptor produces hyperpolarization (inhibition of neuronal firing). Defects in GLRB are a cause of startle disease characterized by muscular rigidity of CNS origin, particularly in the neonatal period, and by an exaggerated startle response to unexpected acoustic or tactile stimuli.

(*Continued*)

Table 2 (*Continued*)

Influenza A hexapeptide			Human protein description[a]
Protein	**Aa pos**	**Sequence**	
	22	ESSSVD	**AHI1:** Abelson helper integration site 1 protein homolog. Jouberin. Highly expressed in primitive normal hematopoietic cells. Expressed in the **brain, particularly in neurons that give rise to the crossing axons of the corticospinal tract and superior cerebellum peduncles**. Upregulated in leukemic cells at all stages of differentiation from patients with chronic myeloid leukemia. Defects in AHI1 are the cause of Joubert syndrome type 3. Joubert syndrome is characterized by cerebellum ataxia, oculomotor apraxia, hypotonia, neonatal breathing abnormalities, and psychomotor delay. Neuroradiologically, it is characterized by cerebellum vermian hypoplasia/aplasia, thickened and reoriented superior cerebellum peduncles, and abnormally large interpeduncular fossa (molar tooth sign). Additional variable features include retinal dystrophy and renal disease. Joubert syndrome type 3 shows minimal extra CNS involvement and appears not to be associated with renal dysfunction.
	24	SSVDLN	**VLDLR:** Very low density lipoprotein (VLDL) receptor. Binds VLDL and transports it to cells by endocytosis. Binding to reelin induces tyrosine phosphorylation of Dab1. Deletions involving VLDLR may be the cause of cerebellum hypoplasia, with mental retardation, delayed ambulation, predominantly truncal ataxia, strabismus and pesplanus, seizures (in 40% of patients), and short stature (in 15% of patients).
	36	ERLKIY	**NARG1:** NMDA receptor-regulated protein 1. Important for vascular and neuronal development. Controls retinal neovascularization. Found in **brain (corpus callosum)**.
	101	QALQLL	**SYT9:** Synaptotagmin-9. Integral membrane protein involved in norepinephrine secretion.
	104	QLLLEV	**ABCA1:** Adenosine 5'-triphosphate (ATP)-binding cassette subfamily A member 1. Widely expressed. Defects in ABCA1 cause Tangier disease, a high density lipoprotein deficiency characterized by premature coronary artery disease, hepatosplenomegaly, recurrent peripheral neuropathy, and progressive muscle wasting and weakness. **MESD1:** Mesoderm development candidate 1.
MP1	15	VPSGPL	**SALL1:** Sal-like protein 1. Zinc finger protein SALL1. Involved in organogenesis. Interacts with HDAC1, HDAC2, RBBP4, RBPP7, MTA1, and MTA2. Highest levels expressed in the kidney. **Lower levels in adult brain (enriched in corpus callosum, lower expression in substantia nigra) and liver. In fetal brain, expressed exclusively in neurons of the subependymal region of hypothalamus lateral to the third ventricle**.
	16	PSGPLK	**MYLK:** Myosin light chain kinase (MLCK). Smooth muscle. Telokin. Implicated in smooth muscle contraction. Implicated in the regulation of endothelial as well as vascular permeability. In the nervous system, it has been shown to control the growth initiation of astrocytic processes in culture and to participate in transmitter release at synapses formed between cultured sympathetic ganglion cells.
	20	LKAEIA	**CNGA3:** Cyclic nucleotide-gated cation channel α-3. Essential for the generation of light-evoked electrical responses in the red-, green- and blue-sensitive cones. Prominently expressed in retina. Defects in CNGA3 are the cause of achromatopsia type 2 (or total color blindness or rod monochromacy).
	38	DLEALM	**USH1G:** Usher syndrome type-1G protein. Required for normal hearing. Expressed in vestibule of the inner ear, eye, and small intestine. Defects in USH1G are the cause of Usher syndrome type 1G, characterized by the association of retinitis pigmentosa and sensorineural deafness. Usher syndrome type 1 is characterized by prepubertal onset of progressive retinitis pigmentosa leading to blindness. **Q86U74:** Phosphoglucomutase-1. This enzyme participates in both the breakdown and synthesis of glucose.
	54	PLTKGI	**Q6RJU1:** Rho GTPase-activating protein 20. Expressed predominantly in the **brain**.
	72	RGLQRR	**LAP4:** Protein scribble homolog. Scribble. Involved in different aspects of epithelial and **neuronal morphogenesis**. Localizes to neuronal post- and presynaptic regions.
MP2	59	LKRGPS	**CLCN2:** Chloride channel protein 2. Functions include the regulation of cell volume. Defects in CLCN2 may be the cause of (1) epilepsy with grand mal seizures on awakening; (2) childhood absence epilepsy type 3, with onset at age 6–7 y, frequent absence seizures, and bilateral, synchronous, and symmetric 3-Hz spike waves on electroencephalogram (EEG); and (3) juvenile absence epilepsy characterized by tonic–clonic seizures on awakening.
HA	143	ASSGVS	**DOCK3:** Dedicator of cytokinesis protein 3. Presenilin-binding protein. May be involved in Alzheimer disease. In normal brains, it is localized in the **neuropil,** while in brains affected by Alzheimer disease it is associated with neurofibrillary tangles. A chromosomal aberration involving DOCK3 may be a cause of early onset of behavioral/developmental disorder with features of attention deficit–hyperactivity disorder. **HD:** Huntingtin. Huntington disease protein.
	157	SSFFRN	**TBE[c]:** Tubulin, epsilon 1.
	158	SFFRNV	**TBE[c]:** Tubulin, epsilon 1. See previous .
	330	ATGLRN	**UBP3:** Ubiquitin carboxyl-terminal hydrolase 3. Ubiquitous, with strongest expression in pancreas.
	419	LERRIE	**DESM:** Desmin. Intermediate filament found in muscle cells. Defects in desmin cause (1) desmin-related cardioskeletal myopathy, with skeletal muscle weakness, cardiac conduction blocks, arrhythmias, and restrictive heart failure; (2) cardiomyopathy dilated type 1I, resulting in congestive heart failure and arrhythmia, with risk of premature death; (3) neurogenic scapuloperoneal syndrome Kaeser type.
	534	SIYSTV	**TMEM1:** Trafficking protein particle complex subunit 10. Epilepsy holoprosencephaly candidate 1.

(*Continued*)

Table 2 (*Continued*)

Influenza A hexapeptide			Human protein description[a]
Protein	**Aa pos**	**Sequence**	
NA	17	VGIISL	**RYR1:** Ryanodine receptor 1. Present in skeletal muscle, **cerebellum, and hippocampus.** Defects in RYR1 are the cause of (1) malignant hyperthermia susceptibility type 1, one of the main causes of death due to anesthesia, with accelerated muscle metabolism, contractures, metabolic acidosis, tachycardia, and even death; (2) central core disease, one of the conditions that produces the "floppy" infant, with neonatal hypotonia, delayed motor development, muscle weakness, and amyotrophy.
	23	MLQIGN	**SC10A[c]:** Sodium channel protein type 10, subunit α. **Expressed in the dorsal root ganglia and sciatic nerve.** Plays a role in neuropathic pain mechanisms.
	24	LQIGNI	**SC10A[c]:** Sodium channel protein type 10 subunit α. See previous entry.
	28	NIISIW	**TOP2B:** Deoxyribonucleic acid (DNA) topoisomerase II, β. Reduced activity of this enzyme may also play a role in ataxia–telangiectasia, a neurodegenerative disease that causes poor coordination and dilation of small blood vessels.
	78	KAVASAV	**GEMI4:** Component of gems 4. Gemin-4. P97. Component of the SMN protein complex. SMN is the product of the neurodegenerative disease spinal muscular atrophy gene.
	102	KDNGIR	**ECHB:** Trifunctional enzyme subunit β, mitochondrial. HADHB. Defects in HADHB are a cause of trifunctional protein deficiency. Manifestations include hypoglycemia, cardiomyopathy, and sudden death.
	440	GSSISF	**CSMD2:** CUB and sushi domain-containing protein 2. Expressed at intermediate levels in the **brain, including cerebellum, substantia nigra, hippocampus, and fetal brain**.
	463	LPFTID	**TAAR5:** Trace amine-associated receptor 5. Mediates aspects of movement control. Expressed exclusively in skeletal muscle and selected areas of the **brain (amygdala, hippocampus, caudate nucleus, hypothalamus, and thalamus)**.
NC	63	IERMVL	**HS70L:** Heat shock 70 kDa protein (Hsp70) 1-Hom. Hsp70s stabilize proteins against aggregation.
	127	EDATAG	**DPYS:** Dihydropyrimidinase (DHP) protein. Defects in DPYS are the cause of DHP deficiency. DHP deficiency is characterized by dihydropyrimidinuria and associated with a variable clinical phenotype: epileptic or convulsive attacks, dysmorphic features, developmental delay, and congenital microvillous atrophy.
	375	DSSTLE	**ATM[d]:** Ataxia–telangiectasia mutated protein. Defects in ATM cause ataxia–telangiectasia, characterized by cerebellum ataxia, blood vessel dilation in the conjunctiva, growth retardation, and sexual immaturity.
	453	PEDVSF	**TBCD:** Tubulin-specific chaperone D. Tubulin-folding protein. Involved in the first step of the tubulin-folding pathway. Modulates microtubule dynamics by capturing guanosine triphosphate (GTP)-bound β-tubulin.
PA	13	IVELAE	**RYR2:** Ryanodine receptor 2. RYR-2. Cardiac muscle-type ryanodine receptor. Expressed in heart muscle, **brain (cerebellum and hippocampus),** and placenta. Expressed in myometrium during pregnancy. Defects in RYR2 cause (1) arrhythmogenic right ventricular cardiomyopathy 2 and (2) catecholaminergic polymorphic ventricular tachycardia type that may degenerate into cardiac arrest and cause sudden death. **ZDH17:** Zinc finger palmitoyltransferase specific for a subset of neuronal proteins. Huntingtin-interacting protein 3. **Highest expression in the brain cortex, cerebellum, occipital lobe, and caudate**.
	118	IEIGVT	**Q9H0B0[e]:** Neuralized-like protein 4. Widely expressed at high levels (including brain).
	343	AELQDI	**Q4VY19:** Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, β polypeptide.
	351	EEKIPK	**Q2NKJ7:** Neuronal-specific septin-3. Filament-forming cytoskeletal GTPase. GTPase activity is required for brain-specific filament formation. Up-regulated during neuronal differentiation.
	378	KVDFED	**CABP2:** Caveolin-3. CAV3. Involved in repair mechanism of both skeletal muscle and cardiomyocytes. Defects in CAV3 cause (1) limb-girdle muscular dystrophy type 1C; (2) hyperCKmia (ie, elevated levels of serum creatine kinase); (3) rippling muscle disease; (4) cardiomyopathy familial hypertrophy, with dyspnea, syncope, collapse, palpitations, and chest pain; and (5) long QT syndrome type 9 (long QT syndromes cause syncope and sudden death in response to stress; they can present with a sentinel event of sudden cardiac death in infancy); (6) Sudden infant death syndrome (SIDS), the sudden death of an infant younger than 1yr that remains unexplained after a thorough case investigation, including a complete autopsy, examination of the death scene, and review of clinical history.
	626	KGVEEG	**ATM[d]:** Serine-protein kinase ATM. Ataxia–telangiectasia mutated. See earlier entry.
	627	GVEEGS	**MINT:** Msx2-interacting protein. SMART/HDAC1-associated repressor protein. Regulates different key pathways such as the Notch pathway. **Expressed at high level in brain,** testis, spleen, and thymus.
	671	ALRDNL	**NBEAL:** NBEA L1 protein. Amyotrophic lateral sclerosis 2 chromosomal region candidate gene 17 protein. Neurobeachin-like protein 1. **Highly expressed in brain,** kidney**,** prostate, and testis.
	684	GGLYEA	**DOCK6:** Dedicator of cytokinesis protein 6 guanine nucleotide exchange factor. Widely expressed. Expressed at low level in spleen, **cerebellum, hippocampus, and substantia nigra.**
RDRP	202	QRTIGK	**AFF2: Fragile X mental retardation** 2 protein. Expressed in **hippocampus and amygdala**.

(*Continued*)

Table 2 (*Continued*)

Influenza A hexapeptide			Human protein description[a]
Protein	**Aa pos**	**Sequence**	
	276	NEKKAK	**DCX:** Neuronal migration protein doublecortin. Lissencephalin-X. Involved in cortex lamination during cerebral cortex development. **Highly expressed in neuronal cells of fetal brain (in the majority of cells of the cortical plate, intermediate zone, and ventricular zone).** Not expressed in other fetal tissues. In adults, highly expressed in the **brain frontal lobe.** Defects in DCX cause lissencephaly X-linked type 1, with mental retardation and seizures that are more severe in male patients. Affected boys show a thick cortex with absent or reduced gyri, feeding problems, abnormal muscular tone, seizures, and psychomotor retardation. Females display a less severe phenotype ("doublecortex").
	401	ASLSPG	**CAD22:** Cadherin-22. **Pituitary and brain cadherin.** **SALL3:** Sal-like protein 3. **Expressed in fetal brain of the 24th gestational week (in hippocampus, cortex, mediodorsal and ventrolateral thalamic nuclei, putamen, cerebellum, and brainstem).**
	510	PSFGVS	**Q8IVF2[f]:** Protein AHNAK2. Is a component of the dysferlin protein complex. In dysferlinopathies, reduction or absence of dysferlin correlates with a secondary muscle-specific loss of AHNAK.
	616	MDEDYQ	**VP13A:** Vacuolar protein sorting-associated protein 13A. Chorein. Higher expression is found in **brain**, heart, skeletal muscle, and kidney. Defects in VPS13A are the cause of chorea-acanthocytosis, characterized by hyperkinetic movements, abnormal erythrocyte morphology, psychiatric features, epilepsy, peripheral neuropathy, myopathy, oral self-mutilation, and basal ganglia atrophy in the brain.
	702	SSSYRR	**VIME:** Vimentin. Vimentins are class-III intermediate filaments found especially in mesenchymal cells.
PB2	4	IKELRD	**A0MZ66:** Shootin-1. Localized in **axon**s. Acts upstream of phosphoinositide 3-kinase (PI3K). Involved in the generation of internal asymmetric signals required for neuronal polarization.
	128	GTEGPV	**HEXA:** β-hexosaminidase subunit α. Defects in HEXA are the cause of GM2-gangliosidosis type 1 also known as "Tay–Sachs disease," with accumulation of GM2 gangliosides in the neuronal cells.
	156	AKEAQD	**EZRI:** Ezrin. **Expressed in cerebral cortex, basal ganglia, hippocampus, hypophysis, and optic nerve. Stronger expression in gray matter of frontal lobe compared to white matter. Preferentially expressed in astrocytes of hippocampus, frontal cortex, thalamus, parahippocampal cortex, amygdala, insula, and corpus callosum. Very strong staining of the Purkinje cell layer and in part of the molecular layer of the infant brain compared to adult brain**.
	187	KEKKEE	**RS14:** 40S ribosomal protein S14. **SFR11:** Splicing factor, Arg/Ser-rich 11 that localizes with spliceosome components.
	207	LERELV	**SPTN5:** Spectrin β chain, brain 4. Spectrin, nonerythroid β chain 4. Detected prominently in the outer segments of photoreceptor rods and cones, and in the basolateral membrane and gastric epithelial cells. Expressed in **cerebellum, spinal cord, pituitary gland,** stomach, liver, pancreas, kidney, bladder, and heart.
	227	VYIEVL	**PCDBD:** Protocadherin β-13. Involved in the establishment of neuronal connections in the **brain.** **Q5VY39:** Protocadherin-15. Essential for normal retinal and cochlear function. Found in the inner and outer synaptic layers, and the nerve fiber layer in adult and fetal retinas. Found in the supporting cells, outer sulcus cells, and spiral ganglion of fetal cochlea. Defects in PCDH15 cause (1) Usher syndrome type 1F, characterized by the association of retinitis pigmentosa and sensorineural deafness, (2) Usher syndrome type 1D/F, and (3) nonsyndromic sensorineural deafness autosomal recessive type 23.
	274	ADPLAS	**TECT2:** Tectonic-2. Participates in hedgehog-mediated patterning of the **neural tube**. During neural tube development, mouse tectonic is required for formation of most of the ventral cell types.
	275	DPLASL	**CNTP5:** Contactin-associated protein-like 5. Caspr5. CNTNAP5. May play a role in the correct development and proper functioning of the peripheral and CNS and may be involved in cell adhesion and intercellular communication. Belongs to the neurexin family.
	277	LASLLE	**FA38A:** Protein FAM38A. In normal **brain, expressed exclusively in neurons,** not in astrocytes. In Alzheimer disease, expressed in approximately half of the activated astrocytes located around classical senile plaques.
	316	GLRISS	**MASS1:** G-protein coupled receptor 98. GPR98. Involved in the development of the CNS. Defects in GPR98 may be a cause of (1) Usher syndrome type 2C, characterized by the association of retinitis pigmentosa with sensorineural deafness; and (2) familial febrile convulsions type 4. Febrile convulsions are seizures in childhood without any evidence of intracranial infection or defined pathologic or traumatic cause. Affects 2%–5% of children aged 3 m to 5 yr.
	317	LRISSS	**LEPR:** Leptin receptor. Receptor for obesity factor (leptin). Involved in the regulation of fat metabolism.
	339	KKEEEV	**Q52LR0:** Cingulin-like protein 1. Paracingulin. Localizes to the apical junction complex composed of tight and adherens junctions. Present in smooth muscle, spleen, testis, **fetal brain, amygdala, corpus callosum, cerebellum, thalamus, and the subthalamic nucleus of adult brain.** A chromosomal aberration involving CGNL1 is a cause of aromatase excess syndrome, characterized by estrogen excess.

(*Continued*)

Table 2 (*Continued*)

Influenza A hexapeptide			Human protein description[a]
Protein	**Aa pos**	**Sequence**	
	341	EEEVLT	**TTBK2:** Tau-tubulin kinase 2. Serine/threonine kinase, which is able to phosphorylate tau on serines. Defects in TTBK2 cause spinocerebellum ataxia type 11. Patients show incoordination of gait and often poor coordination of hands, speech, and eye movements, due to degeneration of the cerebellum with variable involvement of the brainstem and spinal cord.
	347	GNLQTL	**NEB2:** Neurabin-2. Modulates synaptic transmission and dendritic spine morphology. Involved in G-protein coupled receptor signaling, including dopamine D2 receptors and α-adrenergic receptors.
	476	RGVRVS	**ANK2:** Ankyrin-2. **Brain ankyrin**. Required for the regulation of neonatal cardiomyocyte contraction rate. **Expressed in fetal brain and the temporal cortex of adult brain.** Defects in ANK2 cause long QT syndrome type 4. Long QT syndromes cause syncope and sudden death in response to stress.
	565	SQDPTM	**PSD5:** 26S proteasome non-ATPase regulatory subunit 5. 26S proteasome subunit S5B.
	678	DPDEGT	**CELR2:** Cadherin EGF LAG G-type receptor. Flamingo 1. Receptor that may have an important role in cell–cell signaling during nervous system formation. Highest expression in **brain** and testis.
			PCD19: Protocadherin-19. Expressed in **developing cortical plate, amygdala, and subcortical regions and in the ganglionic eminence.** Defects in PCDH19 cause epilepsy (female-restricted), with mental retardation, and convulsive disorder, characterized by seizure onset in infancy and cognitive impairment.
			PCDGM: Protocadherin γ-C5. Involved in the establishment of neuronal connections in the **brain.**

[a]UniProt/Swissprot accession numbers are given in bold. Brain and/or peripheral nervous system localization are given in boldface. For references and potential disease involvement, refer to www.uniprot.org; [b]DSNTIT is present in the viral NEP too; [c]Tubulin and sodium channel protein type 10, subunit α, have 2 consecutive overlapping viral hexapeptides; [d]The ATM has 2 viral hexapeptides: NC375–380DSSTLE and PA626–631KGVEEG; [e]The viral hexapeptide IEIGVT occurs 3 times in the human neuralized-like protein 4; [f]The viral PSFGVS hexapeptide occurs 17 times in the human AHNAK2 protein.

this observation, we restricted our analysis by searching for viral-vs-human overlaps in proteins potentially relevant at the neurological level, focusing in particular on antigens linked to behavior and/or fundamental neurophysiological functions, such as myelin. The data obtained are illustrated in Tables 3 and 4. Specifically, Table 3 reports that influenza A H5N1 polyprotein shares viral hexapeptide sequences with many human proteins that have been associated with schizophrenia, autism, and similar behavioral disorders, such as the Williams–Beuren and Bardet–Biedl syndromes.[20–44] Some among the many human proteins that have influenza A motifs include neurexins, disrupted in schizophrenia 1 protein, and reelin (ie, antigens whose link with autistic syndromes is well documented, both scientifically and clinically).[22,28,35,36] In a few cases, multiple viral hexapeptides occur in the same human protein: eg, human neurexin I-α protein contains 2 viral hexapeptides ($NEP_{25–30}$SVDLNG and $NC_{390–395}$TRSGGN peptides). Likewise, the human Bardet–Biedl syndrome 10 protein contains the viral $HA_{64–69}$KPLILR and $RDRP_{271–276}$LPVGGN peptides (see Table 3). In addition, Table 3 shows that the commonality between the virus polyprotein and a human antigen can also occur at the heptapeptide level (see myosin-IXa and dystrophin).

Table 3 shows that influenza A overlaps occur in a number of human proteins that, when altered, have been reported to be possibly related to autism and behavioral disorders with deficits in social communication and interaction. However, the number of antigens potentially associated with autism and presenting viral motifs may be larger. Indeed, postmortem studies of the brain in autism have shown a broad spectrum of abnormalities in the cerebellum and neocortex, involving limbic regions, such as the anterior cingulate cortex.[45] Following proteomic analysis of the anterior cingulate cortex in major psychiatric disorders and comparing protein spots between control and disease groups, a number of cytoskeletal and mitochondrial proteins appeared dysregulated in psychiatric disorders.[46] Among them, there are glutamate dehydrogenases, described in Table 2.

Finally, Table 4 documents that viral hexapeptide sequences occur in 17 human proteins that have been related to myelin and (de)myelination processes.[47–68] Moreover, many of the proteins described in Table 4 have also been associated with other pathologies, such as bipolar disorder, schizophrenia, congenital cataracts, facial dysmorphism, or neuropathy syndrome. On the whole, again, we observe that the viral-vs-human overlap is physiologically extremely complex regarding the functions and the pathological involvement of the human proteins involved.

Quantitative and qualitative analysis of the hexapeptide overlap of TMV coat protein vs the human proteome

Searching for controls, we also investigated the extent of peptide sharing between human proteins and viruses that do not infect humans, such as TMV and brome mosaic virus (BMV), ie, 2 viruses that specifically infect plants. We found that the TMV and BMV proteomes overlapped the human proteins at a

Table 3 Hexapeptide sharing between influenza A polyprotein and human proteins potentially associated with autism, Williams–Beuren syndrome, Bardet–Biedl syndrome, and schizophrenia

Influenza A hexapeptide			Human protein description[a]
Protein	**Aa pos**	**Sequence**	
NSP1	145	VSLRAF	**Q6ZN28:** Metastasis-associated in colon cancer protein 1 (MACC1). Acts as a transcription activator for MET and as a key regulator of HGF-MET signaling. Alterations in MET and HGF-MET signaling have been associated with autism.[20]
	162	PIPSVP	**VASH:** Vasohibin-1. Angiogenesis inhibitor. Highly expressed in fetal organs. Expressed in brain and placenta. Associated with autism.[21]
NEP	25	SVDLNG	**NRX1A[b]:** Neurexin I-α. Neuronal cell surface protein that may be involved in cell recognition and adhesion. Tissue specificity: heart and brain. Associated with autism.[22]
	42	RDSLGE	**Q9H8T5[c]:** Myosin-IXa. Unconventional myosin-9a. Regulates Rho activity in neurons, and has a role in the regulation of neuronal morphology and function. Candidate gene for the Bardet–Biedl syndrome, which may be characterized by retinal dystrophy, obesity, mental retardation, repetitive behavior, and some autistic-like features.[23]
	43	DSLGES	**Q9H8T5[c]:** Myosin-IXa see previous entry.[23]
	247	SNGQAS	**CSMD1:** CUB and sushi domain-containing protein 1. Expressed at intermediate level in brain, including cerebellum, substantia nigra, hippocampus, and fetal brain. Disruption of the CSMD1 gene is associated with speech delay, autism, and learning difficulties.[24]
HA	64	KPLILR	**BBS10[b]:** Bardet–Biedl syndrome 10 protein. Probable molecular chaperone. Defects in BBS10 are the cause of Bardet–Biedl syndrome type 10, characterized by severe retinopathy, early onset of obesity, polydactyly, hypogenitalism, renal malformation, mental retardation, repetitive behavior, and some autistic-like features.
NC	25	IRASVG	**C1TC:** C-1-tetrahydrofolatesynthase. Includes methylenetetrahydrofolate dehydrogenase, methenyltetrahydrofolate cyclohydrolase, and formyltetrahydrofolate synthetase. Defects in this trifunctional enzyme may cause susceptibility to folate-sensitive neural tube defects, such as open spina bifida and anencephaly. Aberrations in folate metabolic pathway might have a role in the altered susceptibility to schizophrenia, autism, and depression.[25,26]
	335	SAAFED	**BAZ1B:** Williams syndrome transcription factor. Tyrosine-protein kinase BAZ1B. Involved in chromatin remodeling and acts as a transcription regulator. Ubiquitous. Expressed at equal levels in 19–23 week-old fetal tissues. Chromosomal aberrations involving BAZ1B may be the cause of certain cardiovascular and musculoskeletal abnormalities observed in Williams–Beuren syndrome.
	390	TRSGGN	**NRX1A[b]:** Neurexin I-α. Associated with autism. See earlier entry.[22] **NRX1B:** Neurexin-1-β. Involved in cell adhesion by forming intracellular junctions through binding to neuroligins. May play a role in formation or maintenance of synaptic junctions. Disruption of NRX1B has been associated with autism spectrum disorder.[27]
	401	ASAGQI	**DISC1:** Disrupted in schizophrenia 1 protein. Expressed in the dentate gyrus of the hippocampus, and temporal and parahippocampal cortices and cells of the white matter. Expression rises within the dentate gyrus and temporal cortex from the neonatal period to infancy, declines markedly in adolescence, and declines further with aging. Defects in DISC1 may be associated with (1) susceptibility to schizophrenia, and psychosis; (2) susceptibility to schizoaffective disorder characterized by the co-occurrence of symptoms of both mood disorder and psychosis; (3) autism and Asperger syndrome.[28]
PA	2	EDFVRQ	**MAGG1:** Melanoma-associated antigen G1. Ubiquitous. Associated with autism.[29]
	131	YLEKAN	**DMD[c]:** Dystrophin. May play a role in anchoring the cytoskeleton to the plasma membrane. Defects in dystrophin are the cause of the following: (1) Duchenne muscular dystrophy, with proximal muscle weakness causing waddling gait, toe-walking, lordosis, frequent falls, and difficulty in standing up and climbing up stairs (most patients are confined to a wheelchair by the age of 10 or 12 y; contractures and scoliosis ultimately occur); (2) Becker muscular dystrophy, later in onset and more benign; and (3) cardiomyopathy dilated X-linked type 3B. Associated with autism spectrum disorder.[30,31]
	132	LEKANK	**DMD[c]:** Dystrophin. Associated with autism spectrum disorder. See previous entry.[30,31]
	377	EKVDFE	**O75129**: Astrotactin-2. A neuronal cell surface protein expressed on postmitotic neuronal precursors in the cerebellum, hippocampus, cerebrum, and olfactory bulb. Astrotactin mediates neuronastroglial interactions and is also implicated in synaptic development as well as many other neuronal activities. Mutations in the astrotactin gene are linked to neuronal migration. Involved in autism and schizophrenia.[32,33]
	389	DLRQYD	**SBP1:** Selenium-binding protein 1. Highly expressed in liver, lung, colon, prostate, kidney, and pancreas. Present in neurons and glia. Upregulated in brain and blood from schizophrenia patients.[34]

(*Continued*)

Table 3 (*Continued*)

Influenza A hexapeptide			Human protein description[a]
Protein	**Aa pos**	**Sequence**	
	400	PRSLAS	**RELN:** Reelin. Plays a role in layering of neurons in the cerebral cortex and cerebellum. Abundantly produced during brain ontogenesis by the Cajal–Retzius cells and other pioneer neurons located in the telencephalic marginal zone and by granule cells of the granular layer of the cerebellum. In adult brain, preferentially expressed in GABAergic interneurons of prefrontal cortices, temporal cortex, hippocampus, and glutamatergic granule cells of cerebellum. Expression in postnatal human brain is high in the cerebellum. Defects in RELN are the cause of lissencephaly type 2 (Norman–Roberts syndrome). Individuals are ataxic and mentally retarded, and suffer from epilepsy. Defects in RELN may contribute to susceptibility to schizophrenia. Expression of the protein is reduced to about 50% in patients with schizophrenia. Defects in RELN may predispose to autistic disorder.[35,36]
	513	NDTDVV	**DPP10:** Inactive dipeptidyl peptidase 10. May modulate activity of the K+ channels KCND1 and KCND2. Defects in DPP10 may be a cause of susceptibility to asthma. Associated with autism.[37]
RDRP	271	LPVGGN	**BBS10[b]:** Bardet–Biedl syndrome 10 protein. See earlier entry.
	299	SFTITG	**L1CAM:** Neural cell adhesion molecule L1. Involved in the development of the nervous system. Involved in neuron–neuron adhesion, neurite fasciculation, outgrowth of neurites, etc. Defects in L1CAM are the cause of (1) hydrocephalus due to stenosis of the aqueduct of Sylvius, with abnormal accumulation of cerebrospinal fluid in the brain; (2) CRASH syndrome (Corpus callosum hypoplasia, psychomotor Retardation, Adducted thumbs, Spastic paraparesis, and Hydrocephalus), with spasticity and hyperreflexia of lower limbs, shuffling gait, mental retardation, aphasia, and adducted thumbs; and (3) spastic paraplegia X-linked type 1, characterized by progressive weakness and spasticity of the lower limbs. Associated with neurological disorders such as autism, 3p syndrome, and schizophrenia.[38]
PB2	295	VDILRQ	**BAZ1A:** Bromodomain adjacent to zinc finger domain protein 1A. Williams syndrome transcription factor-related chromatin-remodeling factor 180. Involved in the formation or maintenance of heterochromatin playing a critical role in developmental control. Expressed in the testis. Williams syndrome patients are diagnosed by their unique profile of physical and mental abnormalities. Some degree of mental retardation, visual–spatial processing defects, as well as characteristic "elfin" features are obvious traits of most patients.[39]
	314	AMGLRI	**CECR1:** Adenosine deaminase CECR1. Cat eye syndrome critical region protein 1. In embryo, expressed in the outflow tract and atrium of the developing heart, the VII/VIII cranial nerve ganglion, and the notochord. Candidate gene for the cat eye syndrome, a developmental disorder characterized by coloboma of the iris, anal atresia with fistula, downslanting palpebral fissures, heart and renal malformations, and normal or near-normal mental development. Has been associated with autism.[40]
	319	ISSSFS	**WBS23:** Williams–Beuren syndrome chromosomal region 23 protein. **PAX6:** Paired box protein Pax-6. Oculorhombin. Aniridia type II protein. Required for the differentiation of pancreatic islet α cells. Regulates specification of the ventral neuron subtypes. Present in fetal eye, brain, spinal cord, and olfactory epithelium. Defects in PAX6 are the cause of (1) aniridia type II, characterized by absence of the iris, absence of the fovea, malformations of the lens and anterior chamber; (2) Peters anomaly, consisting of a central corneal leukoma; (3) ectopia pupilla; (4) autosomal dominant keratitis, with corneal opacification and vascularization, and by foveal hypoplasia; (5) ocular coloboma (may cause as much as 10% of the childhood blindness); and (6) Gillespie syndrome, with aniridia cerebellum ataxia and mental deficiency. Has been associated with autism.[41]
	346	TGNLQT	**Q8N2F4:** Cell adhesion molecule 1. CADM1. May act in synapse assembly. May be involved in neuronal migration, axon growth, path finding, and fasciculation on the axons of differentiating neurons. Mutations in the gene encoding CADM1 are associated with autism spectrum disorder.[42]
	371	TAILRK	**NBEA:** Neurobeachin. Lysosomal-trafficking regulator 2. Protein BCL8B. LYST2. Expressed predominantly in the brain. Neurobeachin has been associated with idiopathic autism and is disrupted in patients with multiple myeloma.[43]
	388	GRDEQS	**MYH9:** Myosin-9. Nonmuscle myosin heavy chain IIa. Defects in MYH9 are a cause of Hegglin anomaly, Sebastian syndrome, Fechtner syndrome, Alport syndrome, nonsyndromic sensorineural deafness, and macrothrombocytopenia with progressive sensorineural deafness. May be related to many neurological disorders such as depression and autism with only a subclinical platelet defect.[44]

[a]UniProt/Swissprot accession numbers are given in bold. For additional references and potential disease involvement, refer to www.uniprot.org; [b]Neurexin I-α and Bardet–Biedl syndrome-10 protein have 2 viral hexamers; [c]Myosin-IXa and dystrophin have a viral heptamer.

hexapeptide extent comparable to that of the influenza A virus (Kanduc, unpublished data). As an illustrative example, Tables 5 and 6 detail the quantitative and qualitative hexapeptide overlap of TMV coat protein vs the human proteome.

Table 5 shows that the TMV coat protein has a hexapeptide overlap with human proteins about 3.8-fold higher than the theoretical value. That is, it is even higher than the overlap value found with influenza A virus (Table 1). Moreover, the TMV coat protein-vs-human proteome overlap involves numerous human proteins specifically expressed or highly abundant in the human brain (Table 6).

Table 4 Influenza A hexapeptide occurrences in human proteins related to myelination

Influenza A hexapeptide			Human protein description[a]
Protein	Aa pos	Sequence	
HA	116	ELKHLL	**Q5VUM2:** Laminin subunit α-2. Merosin heavy chain. LAMA2. Involved in the attachment, migration, and organization of cells into tissues during embryonic development. Defects in LAMA2 are the cause of merosin-deficient congenital muscular dystrophy type 1A, characterized by difficulty in walking, hypotonia, hyporeflexia, and white matter hypodensity on MRI. A single point mutation in the LN domain of LAMA2 may cause muscular dystrophy and peripheral amyelination.[47]
	442	AELLVL	**NFL:** Neurofilament light polypeptide. NEFL. Involved in the maintenance of neuronal caliber. Defects in NEFL are the cause of CMT disease type 1F (CMT1F), characterized by severely reduced nerve conduction velocities (less than 38 m/s), segmental demyelination and remyelination with onion bulb formations on nerve biopsy, slowly progressive distal muscle atrophy and weakness, absence of deep tendon reflexes, and hollow feet. The CMT1F condition is characterized by onset in infancy or childhood (range 1–13 yr). Defects in NEFL are the cause of CMT disease type 2E, characterized by signs of axonal regeneration in the absence of obvious myelin alterations and progressive distal muscle weakness and atrophy.[48]
NA	26	IGNIIS	**CH60:** 60 kDa heat shock protein, mitochondrial. HSP-60. CPN60. Mitochondrial matrix protein P1. HSPD1. Defects in HSPD1 are a cause of (1) spastic paraplegia, a degenerative spinal cord disorder characterized by a progressive weakness and spasticity of the lower limbs and (2) leukodystrophy hypomyelinating type 4, a severe hypomyelinating leukodystrophy, characterized by infantile-onset rotary nystagmus, progressive spastic paraplegia, neurologic regression, and mental retardation.[49,50]
	107	RIGSKG	**MYEF2:** Myelin expression factor 2. Transcriptional repressor of the myelin basic protein gene.
	201	GAVAVL	**Q4KMR1:** Synaptojanin-1. Synaptic inositol-1,4,5-trisphosphate 5-phosphatase 1. Isoform 1 is more enriched than isoform 2 in developing brain. Isoform 2 is abundant in nerve terminals. Possibly involved in the development of chromosome 21q22-linked bipolar disorders and schizophrenia. May contribute to brain dysfunction and cognitive disabilities in Down's syndrome. Involved in the interaction of myelin basic protein with the plasma membrane in oligodendroglial cells.[51–53]
	202	AVAVLK	**ARHGA:** Rho guanine nucleotide exchange factor 10. ARHGEF10. Defects in ARHGEF10 cause slow nerve conduction, without any clinical signs of peripheral or CNS dysfunction. Involved in developmental myelination of peripheral nerves.[54]
NP	178	AAGAAV	**CTDP1:** RNA polymerase II subunit A C-terminal domain phosphatase. TFIIF-associating CTD phosphatase. Processively dephosphorylates "Ser-2" and "Ser-5" of the heptad repeats YSPTSPS in the COOH domain of the largest RNA polymerase II subunit. This promotes the activity of RNA polymerase II. The phosphorylation is required for the physical interaction with GTF2F1. Defects in CTDP1 may cause congenital cataracts facial dysmorphism neuropathy syndrome, with hypogonadism, hypomyelination of the peripheral nervous system, and delayed motor and intellectual development. Central nervous system involvement, with cerebral and spinal cord atrophy, may be the result of disrupted development. Affected individuals are prone to severe rhabdomyolysis after viral infections and to serious complications related to general anesthesia as pulmonary edema and epileptic seizures.[55]
	282	GLAVAS	**ATRN:** Attractin. Mahogany homolog. Involved in the initial immune cell clustering during inflammatory response. Involved in melanocortin signaling pathways that regulate energy homeostasis and hair color. Has a critical role in normal myelination in the CNS.[56,57]
	283	LAVASG	**SOX10:** Transcription factor SOX-10. Could confer cell specificity to the function of other transcription factors in developing and mature glia. Expressed in fetal brain and in adult brain, heart, small intestine, and colon. Defects in SOX10 are the cause of (1) Waardenburg syndrome type 2E, with sensorineural deafness, pigmentary disturbances, and absence of dystopia canthorum; (2) Waardenburg syndrome type 4, with association of Waardenburg features (depigmentation and deafness) and the absence of enteric ganglia in the distal part of the intestine (Hirschsprung disease); (3) Yemenite deaf–blind hypopigmentation syndrome, with cutaneous hypopigmented or hyperpigmented spots and patches, microcornea, coloboma, and severe hearing loss; and (4) PCWH, with features of Peripheral demyelinating neuropathy, Central dysmyelinating leukodystrophy, Waardenburg syndrome, and Hirschsprung disease.[58,59]
	372	ETMDSS	**Q16301:** Replication initiation-like protein. Myelin transcription factor 2. MyT2. Cerebrin-50. Binds to the promoter of the myelin proteolipid protein. Strongly detected in cerebrospinal fluid. Expressed in oligodendrocyte progenitor cells, pyramidal and Purkinje neurons, neurons of the hypothalamus, hippocampus, ventromedial thalamus, cells of the choroid plexus, and ciliated ependymal cells.[60]
	389	RTRSGG	**Q9P0W5:** Schwannomin-interacting protein 1. SCHIP-1. Component of both axon initial segments and nodes of Ranvier, essential regions for saltatory conduction of the action potential along the axon, in the central and peripheral nervous systems. Isoforms 5 and 6 are highly expressed in the brain.[61]
RDRP	363	KLRTQI	**MYT1L:** Myelin transcription factor 1-like protein. MyT1L protein. May function as a panneural transcription factor associated with neuronal differentiation with a role in the development of neurons and oligodendrogalia in the CNS. Involved in schizophrenia.[33]

(*Continued*)

Table 4 (*Continued*)

Influenza A hexapeptide			Human protein description[a]
Protein	**Aa pos**	**Sequence**	
	513	GVSGIN	**PRAX:** Periaxin. PRX. Involved in the early phases of myelin deposition, in the maintenance of peripheral nerve myelin sheath, and in axon–glial interactions, by interacting with the cytoplasmic domains of integral membrane proteins such as myelin-associated glycoprotein in the periaxonal regions of the Schwann cell plasma membrane. Isoforms 1 and 2 are found in sciatic nerve and Schwann cells. Defects in PRX are a cause of Dejerine–Sottas syndrome, a neuropathy of the demyelinating CMT disease category, with onset by age 2 yr, characterized by motor and sensory neuropathy with slow nerve conduction velocities, increased cerebrospinal fluid protein concentrations, hypertrophic nerve changes, and delayed age of walking as well as areflexia.[62,63]
PB2	113	KVYKTY	**TSN3:** Tetraspanin-3. Tspan-3. Regulates the proliferation and migration of oligodendrocytes, a process essential for normal myelination and repair.[64]
	185	ITKEKK	**MTMRD:** Myotubularin-related protein 13. SBF2. Expressed in spinal cord. Defects in SBF2 are the cause of CMT4B2, a recessive form of CMT disease, the most common inherited disorder of the peripheral nervous system, characterized by abnormal folding of myelin sheaths with pathological hallmarks on nerve biopsies including concentric sheaths (onion bulbs).[65,66]
	190	KEELQD	**SYYC:** Tyrosyl-tRNA synthetase, cytoplasmic. TyrRS. YARS. Defects in YARS are the cause of CMT disease dominant intermediate type C, a form of CMT disease characterized by clinical and pathologic features intermediate between demyelinating and axonal peripheral neuropathies, and motor median nerve conduction velocities ranging from 25 to 45 m/s.[67]
	746	DSQTAT	**IKAP:** IκB kinase complex-associated protein. IKBKAP. May play a role in chromatin remodeling and is involved in acetylation of histones H3 and probably H4. Defects in IKBKAP are the cause of familial dysautonomia or Riley–Day syndrome or hereditary sensory and autonomic neuropathy type III due to progressive degeneration of the sensory, sympathetic and parasympathetic neurons. FD individuals are affected with a variety of symptoms such as decreased sensitivity to pain and temperature, cardiovascular instability, recurrent pneumonias, vomiting crises, and gastrointestinal dysfunction. IKAP/hELP1 deficiency in the cerebrum of familial dysautonomia patients results in down regulation of genes involved in oligodendrocyte differentiation and in myelination.[68]

[a]UniProt/Swissprot accession numbers are given in bold. For additional references and potential disease involvement, refer to www.uniprot.org

Discussion

The present study shows that a high number of influenza A H5N1 hexapeptides are seen in numerous human proteins that play a crucial role in neural networks and functions. This finding might help one to understand the neuropathogenesis of influenza-associated encephalopathy that still remains unclear.[69,70] In this regard, a recent report has described significant expression changes in cytoskeleton proteins, proteins associated with the ubiquitin–proteasome pathway, and neural signal transduction proteins in chicken brain following influenza H5N1 infection.[71] Interestingly, many proteins, the expression of which changed in the H5N1-infected brain (ie, tubulin, phosphoglucomutase-1, ubiquitin carboxyl-terminal hydrolase 3, vimentin, 26S proteasome subunit S5B, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, β polypeptide, septin, and tubulin-specific chaperone D), are listed in Table 2 as proteins that share hexapeptide motifs with the viral H5N1 polyprotein. From this point of view, the present report offers a molecular basis to investigate the possible link(s) between influenza viruses and neuropathogenesis, especially considering that short peptides are functional modules playing a central role in enzyme activity, cell–cell adhesion, hormone activity, and regulation of oncoprotein expression.[8,72,73]

Table 5 Description in numbers of the hexapeptide overlap between TMV coat protein and the human proteome

Unique viral hexamers	154
Viral hexamers occurring in the human proteins	75
Expected number of viral hexamer occurrences in the human proteome	19.84

The remarkable viral-vs-human peptide commonality is not exclusive of the influenza A H5N1 virus, but this phenomenon has been found in a number of different viruses, with different tissue tropism and pathogenicity.[5] In addition, as a datum of relevant interest, for the first time this study demonstrates that there is a high level of peptide sharing between the human proteome and the coat protein from TMV, a virus that does not infect humans. Even more surprisingly, many of the TMV coat protein hexapeptides occur in human-brain-specific proteins. From an evolutionary point of view, these findings raise the fundamental question of what are the phenetic links between viruses and *Homo sapiens*. Indeed, Tables 5 and 6 seem to indicate that the phenetic similarities at the protein level between humans and viruses do not derive from a convergent evolution process catalyzed by a selective pressure. Or, in other words, the phenetic amino acid sequence identity (ie, the peptide

Table 6 Human proteins sharing hexapeptide sequence(s) with TMV coat protein

Sequence	Matches	Human protein(s) description[a]
TTPSQF	2	**BIN1:** Myc box-dependent-interacting protein 1. **NIPBL:** Nipped-B-like protein.
VFLSSA	2	**FRAS1:** Extracellular matrix protein FRAS1. Highest levels found in kidneys, pancreas, and **thalamus**. **R3HC1:** R3H and coiled-coil domain-containing protein 1.
LSSAWA	1	**SC6A3:** Sodium-dependent dopamine transporter. Terminates the action of dopamine by its high affinity sodium-dependent reuptake into **presynaptic terminals**.
SAWADP	1	**NACAD:** NAC-alpha domain-containing protein 1.
WADPIE	1	**HNRPQ:** Heterogeneous nuclear ribonucleoprotein Q. Detected in the heart, brain, pancreas, and placenta.
LINLCT	1	**GP112:** Probable G-protein coupled receptor 112. **Detected in fetal retina.** Highly expressed in normal enterochromaffin cells and in **neuroendocrine carcinoma**.
CTNALG	1	**EMR1[b]:** EGF-like module-containing mucin-like hormone receptor-like 1.
NALGNQ	1	**TACC2:** Transforming acidic coiled-coil-containing protein 2. Acts in the **nuclear** migration of neural progenitors.
FQTQQA	2	**TM40L:** Mitochondrial import receptor subunit TOM40B. **ZN687:** Zinc finger protein 687.
QARTVV	1	**FND3A:** Fibronectin type-III domain-containing protein 3A expressed in odontoblasts.
TVVQRQ	1	**RBM22:** Pre-mRNA-splicing factor RBM22.
RQFSEV	1	**Q96CH6:** MGC20647 protein.
PSPQVT	1	**NBAS: Neuroblastoma**-amplified gene protein.
SPQVTR	2	**NPHP1:** Juvenile nephronophthisis 1 protein. **VRK3:** Serine/threonine-protein kinase VRK3.
PQVTRF	1	**Q6ZR66:** The cDNA FLJ46600 fis, clone THYMU3047144.
TVRFPD	1	**ACHA7:** Neuronal acetylcholine receptor subunit α-7.
DSDFKV	1	**PHF14:** Putative uncharacterized protein PHF14.
FKVYRY	1	**TRIM2:** Tripartite motif-containing protein 2.
YNAVLD	1	**QSOX2:** Neuroblastoma-derived sulfhydryl oxidase 2. Expressed in pancreas, **brain,** placenta, and fetal tissues.
VLDPLV	1	**LPAR5:** Lysophosphatidic acid receptor 5.
LDPLVT	2	**LZTS1:** Leucine zipper putative tumor suppressor 1. Involved in the regulation of cell growth. Highly expressed in testis, prostate, spleen, thymus, ovary, and **brain.** **Q08E86:** KIAA0100 protein.
LVTALL	2	**KLRAQ:** KLRAQ motif-containing protein 1. **TENS4:** C-terminal tensin-like protein.
VTALLG	3	**KDM5C:** Lysine-specific demethylase 5C. Defects in KDM5C are the cause of **mental retardation.** Syndromic X-linked JARID1C-related. **O52B4:** Olfactory receptor, family 52, subfamily B, member 4. **MSLNL:** Mesothelin-like protein.
TALLGA	1	**STRA6:** Stimulated by retinoic acid gene 6 protein homolog.
ALLGAF	1	**TRUA:** The tRNA pseudouridine synthase A.
TRNRII	1	**A6NIH1:** Putative uncharacterized protein ENSP00000350479.
NRIIEV	1	**TIF1A:** Transcription intermediary factor 1-alpha.
RIIEVE	1	**B7Z872:** cDNA FLJ56895, highly similar to Echinoderm microtubule-associated protein-like.
IIEVEN	3	**SEP12:** Septin-12. Widely expressed. **CC110:** Coiled-coil domain-containing protein 110. **SEPT3: Neuronal-specific** septin-3. **Upregulated during neuronal differentiation**.
EVENQA	2	**COG5:** Conserved oligomeric golgi complex subunit 5.
PTTAET	1	**SFRS8:** SFRS8 protein.
TTAETL	2	**REEP2:** Receptor expression-enhancing protein 2. **Detected in the brain,** heart, and skeletal muscle. **BPI[c]:** Bactericidal permeability-increasing protein.
TAETLD	2	**KLD8B:** Kelch domain-containing protein 8B. **BPI[c]:** Bactericidal permeability-increasing protein.
ETLDAT	1	**SYFA:** Phenylalanyl-tRNA synthetase α chain.
DDATVA	1	**FBXW8:** F-box/WD repeat-containing protein 8.
DATVAI	1	**S13A5:** Solute carrier family 13 member 5.
AIRSAI	1	**B4DVQ1:** The cDNA FLJ59288. Highly similar to roundabout homolog 2.
IRSAIN	1	**RAD17:** Cell cycle checkpoint protein RAD17.
NNLIVE	1	**D3DXA2: Brain expressed** X-linked 2 protein, isoform CRA_a.
NLIVEL	1	**CS051:** UPF0470 protein C19orf51.

(*Continued*)

Table 6 (*Continued*)

Sequence	Matches	Human protein(s) description[a]
LIRGTG	1	**SRPR[c]:** Signal recognition particle receptor subunit α.
IRGTGS	2	**SRPR[c]:** Signal recognition particle receptor subunit α.
		PSTK: Putative uncharacterized protein.
RSSFES	2	**PRUN2:** BNIP2 motif-containing molecule at the C-terminal region 1.
		Q5JUB6: Prune homolog 2.
SSFESS	4	**ARMC2:** Armadillo repeat-containing protein 2.
		ANKS6: Ankyrin repeat and SAM domain-containing protein 6. Required for renal function.
		ZEP3: Transcription factor HIVEP3.
		CA031: Uncharacterized protein C1orf31.
FESSSG	2	**PTN13[c]:** Protein–tyrosine phosphatase 1E. Most abundant in the lung, kidney, and **fetal brain**.
		EMR1[b]: EGF-like module-containing mucin-like hormone receptor-like 1.
ESSSGL	4	**PTN13[c]:** Protein–tyrosine phosphatase 1E. Most abundant in the lung, kidney and **fetal brain**.
		NF2L1: Nuclear factor erythroid 2-related factor 1.
		RGS9: Regulator of G-protein signaling 9. Highly expressed in the **caudate and putamen**.
		Q9Y628: Putative uncharacterized protein.
SGLVWT	1	**AL8A1:** Aldehyde dehydrogenase family 8 member A1.
LVWTSG	1	**EIF3L:** Eukaryotic translation initiation factor 3 subunit L.
TSGPAT	5	**MUC16:** Ovarian carcinoma antigen CA125.
		K1755: Uncharacterized protein KIAA1755.
		SF01: Splicing factor 1.
		MAP2: Microtubule-associated protein 2.
		A2RRB4: Regulation of transcription, DNA dependent.

[a]UniProt/Swissprot accession numbers are given in bold. Brain and/or peripheral nervous system localization given in boldface. For references and potential disease involvement, visit www.uniprot.org; [b]The EMR1, EGF-like module-containing mucin-like hormone receptor-like 1 contains 2 viral hexapeptides; [c]The bactericidal permeability-increasing protein (BPI); signal recognition particle receptor (SRPR) subunit α; and protein-tyrosine phosphatase 1E (PTN13) have 2 consecutive overlapping viral hexamers.

sharing between viruses and humans) appears to occur independently of the host's susceptibility to a viral infection. In this frame, the remarkable viral-vs-human peptide commonality suggests that one should rethink and redefine the evolutionary links between entities that are phylogenetically different such as viruses and *Homo sapiens*.

In addition, the viral-vs-human peptide sharing may also offer an interpretation key for the enigmatic ability of influenza viruses to escape from humoral and cellular immune recognition.[74,75] Antigenic drift (ie, accumulation of mutations in influenza surface glycoproteins, HA, and NA, and/or in cytotoxic T-lymphocyte epitopes) and antigenic shift (ie, introduction of new viral subtypes) have been proposed as the main mechanisms allowing escape from host immune surveillance.[76] In addition, the peptide identity platform existing between the influenza A polyprotein and human proteome may be a concomitant factor in successful influenza pandemics. As a matter of fact, a human anti-influenza immune reaction might be prevented a priori by powerful self-tolerance mechanisms fully operating in our body. According to Ehrlich, the immune system is prevented from mounting an attack against self-antigens, thereby avoiding the harmful autodestruction of the tissues it was designed to protect: "It would be exceedingly dysteleologic if in this situation autotoxins were formed". Ehrlich named this concept "horror autotoxicus".[77]

Hence, as already suggested,[78,79] it is possible that the massive peptide overlap existing between microbes and the human host is the main obstacle to an immune reaction against infectious agents, thus explaining the scarce immunogenicity of influenza viruses in humans.[74,75] Accordingly, in vaccination procedures and protocols, the simple administration of an influenza virus generally does not elicit a powerful immune response. To increase immunogenicity, often anti-influenza vaccines contain adjuvants,[80] ie, compounds that appear to act by increasing antigen availability and uptake by immune cells, activating innate immunity pathways *in vivo*, and generating an immunocompetent environment at the injection site.[81] Through these multiple and not yet fully understood mechanisms, adjuvants bypass the host immunotolerance mechanisms and determine powerful immune responses. This means as a logical consequence of the peptide commonality between the virus and the human host, the adjuvant-potentiated anti-influenza reaction

may enhance cross-reactivity with sequences shared with human proteins (Tables 2–4). In this context, it is worth recalling that 5–6 amino acids can form a sufficient minimal determinant for an epitope–paratope interaction, so that the hexapeptides described in Tables 2–4 may act as immune units and play a role in cell immunoreactivity and antigen–antibody recognition.[8,9]

In conclusion, the findings presented here may be of help not only for a more thorough analysis of the relationship(s) between influenza viruses and humans but also for designing safe peptide-based anti-influenza vaccines.[82] Indeed, selecting unique viral sequences and, conversely, discarding viral sequences shared with human proteins may eliminate side effects due to cross-reactivity.[15,83–85]

Disclosure

The author reports no conflict of interest in this work.

References

1. Oldstone MB, Levine AJ. Virology in the next millennium. *Cell*. 2000; 100:139–142.
2. Oldstone MB. Anatomy of viral persistence. *PLoS Pathog*. 2009;5: e1000523.
3. Oldstone MB. A suspenseful game of 'hide and seek' between virus and host. *Nat Immunol*. 2007;8:325–327.
4. Kim WM, Sigalov AB. Viral pathogenesis, modulation of immune receptor signaling and treatment. *Adv Exp Med Biol*. 2008;640: 325–349.
5. Kanduc D, Stufano A, Lucchese G, Kusalik A. Massive peptide sharing between viral and human proteomes. *Peptides*. 2008;29:1755–1766.
6. Kanduc D. Quantifying the possible cross-reactivity risk of an HPV16 vaccine. *J Exp Ther Oncol*. 2009;8:65–76.
7. Ricco R, Kanduc D. Hepatitis B virus and *Homo sapiens* proteome-wide analysis: a profusion of viral peptide overlaps in neuron-specific human proteins. *Biologics*. 2010;4:75–81.
8. Lucchese G, Stufano A, Trost B, Kusalik A, Kanduc D. Peptidology: short amino acid modules in cell biology and immunology. *Amino Acids*. 2007;33:703–707.
9. Oldstone MB. Molecular mimicry and immune-mediated diseases. *FASEB J.* 1998;12:1255–1265.
10. Rothbard JB, Gefter ML. Interactions between immunogenic peptides and MHC proteins. *Ann Rev Immunol.* 1991;9:527–565.
11. Tiwari R, Geliebter J, Lucchese A, Mittelman A, Kanduc D. Computational peptide dissection of Melan-a/MART-1 oncoprotein antigenicity. *Peptides.* 2004;25:1865–1871.
12. Reddehase MJ, Rothbard JB, Koszinowski UH. A pentapeptide as minimal antigenic determinant for MHC class I-restricted T lymphocytes. *Nature.* 1989;337:651–653.
13. Hemmer B, Kondo T, Gran B, et al. Minimal peptide length requirements for CD4(+) T-cell clones – implications for molecular mimicry and T-cell survival. *Int Immunol.* 2000;12:375–383.
14. Niman HL, Houghten RA, Walker LE, et al. Generation of protein-reactive antibodies by short peptides is an event of high frequency: implications for the structural basis of immune recognition. *Proc Natl Acad Sci U S A*. 1983;80:4949–4953.
15. Lucchese G, Stufano A, Kanduc D. Proposing low-similarity peptide vaccines against *Mycobacterium tuberculosis*. *J Biomed Biotechnol.* doi:10.1155/2010/832341.
16. Wu CH, Yeh LS, Huang H, et al. The protein information resource. *Nucleic Acids Res.* 2003;31:345–347.
17. Satoh M, Takahashi M, Sakamoto T, Hiroe M, Marumo F, Kimura A. Structural analysis of the titin gene in hypertrophic cardiomyopathy: identification of a novel disease gene. *Biochem Biophys Res Commun*. 1999;262:411–417.
18. Karbowniczek M, Zitserman D, Khabibullin D, et al. The evolutionarily conserved TSC/Rheb pathway activates Notch in tuberous sclerosis complex and Drosophila external sensory organ development. *J Clin Invest*. 2010;120:93–102.
19. Crino PB, Henske EP. New developments in the neurobiology of the tuberous sclerosis complex. *Neurology*. 1999;53:1384–1390.
20. Campbell DB, D'Oronzio R, Garbett K, et al. Disruption of cerebral cortex MET signaling in autism spectrum disorder. *Ann Neurol*. 2007;62:243–250.
21. Wang K, Zhang H, Ma D, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*. 2009;459: 528–533.
22. Yan J, Noltner K, Feng J, et al. Neurexin 1alpha structural variants associated with autism. *Neurosci Lett*. 2008;438:368–370.
23. Gorman SW, Haider NB, Grieshammer U, et al. The cloning and developmental expression of unconventional myosin IXA (MYO9A) a gene in the Bardet-Biedl syndrome (BBS4) region at chromosome 15q22-q23. *Genomics*. 1999;59:150–160.
24. Glancy M, Barnicoat A, Vijeratnam R, et al. Transmitted duplication of 8p23.1-8p23.2 associated with speech delay, autism, and learning difficulties. *Eur J Hum Genet*. 2009;17:37–43.
25. Mohammad NS, Jain JM, Chintakindi KP, Singh RP, Naik U, Akella RR. Aberrations in folate metabolic pathway and altered susceptibility to autism. *Psychiatr Genet*. 2009;19:171–176.
26. Adams M, Lucock M, Stuart J, Fardell S, Baker K, Ng X. Preliminary evidence for involvement of the folate gene polymorphism 19bp deletion-DHFR in occurrence of autism. *Neurosci Lett*. 2007;422: 24–29.
27. Kim HG, Kishikawa S, Higgins AW, et al. Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet*. 2008;82:199–207.
28. Kilpinen H, Ylisaukko-Oja T, Hennah W, et al. Association of DISC1 with autism and Asperger syndrome. *Mol Psychiatry*. 2008;13: 187–196.
29. Chibuk TK, Bischof JM, Wevrick R. A necdin/MAGE-like gene in the chromosome 15 autism susceptibility region: expression, imprinting, and mapping of the human and mouse orthologues. *BMC Genet*. 2001;2:22.
30. Wu JY, Kuban KC, Allred E, Shapiro F, Darras BT. Association of Duchenne muscular dystrophy with autism spectrum disorder. *J Child Neurol*. 2005;20:790–795.
31. Dosé AC, Burnside B. A class III myosin expressed in the retina is a potential candidate for Bardet-Biedl syndrome. *Genomics*. 2002;79:621–624.
32. Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009;459:569–573.
33. Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, et al. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am J Hum Genet*. 2008;83:504–510.
34. Glatt SJ, Everall IP, Kremen WS, et al. Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. *Proc Natl Acad Sci U S A*. 2005;102:15533–15538.
35. Persico AM, D'Agruma L, Maiorano N, et al. Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder. *Mol Psychiatry*. 2001;6:150–159.
36. Fatemi SH. The role of reelin in pathology of autism. *Mol Psychiatry*. 2002;7:919–920.

37. Marshall CR, Noor A, Vincent JB, et al. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*. 2008;82:477–488.
38. Chen L, Zhou S. "CRASH"ing with the worm: insights into L1CAM functions and mechanisms. *Dev Dyn*. 2010;239:1490–1501.
39. Castermans D, Wilquet V, Parthoens E, et al. The neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism. *J Med Genet*. 2003;40:352–356.
40. Stubbs G, Litt M, Lis E, et al. Adenosine deaminase activity decreased in autism. *J Am Acad Child Psychiatry*. 1982;21:71–74.
41. Maekawa M, Iwayama Y, Nakamura K, et al. A novel missense mutation (Leu46Val) of PAX6 found in an autistic patient. *Neurosci Lett.* 2009;462:267–271.
42. Bochar DA, Savard J, Wang W, et al. A family of chromatin remodeling factors related to Williams syndrome transcription factor. *Proc Natl Acad Sci U S A*. 2000;97:1038–1043.
43. Zhiling Y, Fujita E, Tanabe Y, Yamagata T, Momoi T, Momoi MY. Mutations in the gene encoding CADM1 are associated with autism spectrum disorder. *Biochem Biophys Res Commun*. 2008;377:926–929.
44. Freson K, Labarque V, Thys C, Wittevrongel C, Geet CV. What's new in using platelet research? To unravel thrombopathies and other human disorders. *Eur J Pediatr*. 2007;166:1203–1210.
45. Simms ML, Kemper TL, Timbie CM, Bauman ML, Blatt GJ. The anterior cingulate cortex in autism: heterogeneity of qualitative and quantitative cytoarchitectonic features suggests possible subgroups. *Acta Neuropathol*. 2009;118:673–684.
46. Beasley CL, Pennington K, Behan A, Wait R, Dunn MJ, Cotter D. Proteomic analysis of the anterior cingulate cortex in the major psychiatric disorders: evidence for disease-associated changes. *Proteomics*. 2006;6:3414–3425.
47. Patton BL, Wang B, Tarumi YS, Seburn KL, Burgess RW. A single point mutation in the LN domain of LAMA2 causes muscular dystrophy and peripheral amyelination. *J Cell Sci*. 2008;121:1593–1604.
48. Jordanova A, de Jonghe P, Boerkoel CF, et al. Mutations in the neurofilament light chain gene (NEFL) cause early onset severe Charcot-Marie-Tooth disease. *Brain*. 2003;126:590–597.
49. Hansen JJ, Durr A, Cournu-Rebeix I, et al. Hereditary spastic paraplegia SPG13 is associated with a mutation in the gene encoding the mitochondrial chaperonin Hsp60. *Am J Hum Genet*. 2002;70:1328–1332.
50. Magen D, Georgopoulos C, Bross P, et al. Mitochondrial Hsp60 chaperonopathy causes an autosomal-recessive neurodegenerative disorder linked to brain hypomyelination and leukodystrophy. *Am J Hum Genet*. 2008;83:30–42.
51. Saito T, Guan F, Papolos DF, et al. Mutation analysis of SYNJ1: a possible candidate gene for chromosome 21q22-linked bipolar disorder. *Mol Psychiatry*. 2001;6:387–395.
52. Nawaz S, Kippert A, Saab AS, et al. Phosphatidylinositol 4,5-bisphosphate-dependent interaction of myelin basic protein with the plasma membrane in oligodendroglial cells and its rapid perturbation by elevated calcium. *J Neurosci*. 2009;29:4794–4807.
53. Voronov SV, Frere SG, Giovedi S, et al. Synaptojanin 1-linked phosphoinositide dyshomeostasis and cognitive deficits in mouse models of Down's syndrome. *Proc Natl Acad Sci U S A*. 2008;105:9415–9420.
54. Verhoeven K, de Jonghe P, van de Putte T, et al. Slowed conduction and thin myelination of peripheral nerves associated with mutant rho Guanine-nucleotide exchange factor 10. *Am J Hum Genet*. 2003;73:926–932.
55. Varon R, Gooding R, Steglich C, et al. Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome. *Nat Genet*. 2003;35:185–189.
56. Kuramoto T, Kitada K, Inui T, et al. Attractin/mahogany/zitter plays a critical role in myelination of the central nervous system. *Proc Natl Acad Sci U S A*. 2001;98:559–564.
57. Di Giovanni S, Faden AI, Yakovlev A, et al. Neuronal plasticity after spinal cord injury: identification of a gene cluster driving neurite outgrowth. *FASEB J*. 2005;19:153–154.
58. Li H, Lu Y, Smith HK, Richardson WD. Olig1 and Sox10 interact synergistically to drive myelin basic protein transcription in oligodendrocytes. *J Neurosci*. 2007;27:14375–14382.
59. Takada N, Kucenas S, Appel B. Sox10 is necessary for oligodendrocyte survival following axon wrapping. *Glia*. 2010;58:996–1006.
60. Li AH, Silvestrini B, Leone MG, Giacomelli S, Cheng CY. Cerebrin-50, a human cerebrospinal fluid protein whose mRNA is present in multiple tissues but predominantly expressed in the lymphoblastoid cells and the brain. *Biochem Mol Biol Int*. 1995;35:135–144.
61. Martin PM, Carnaud M, Garcia del Caño G, et al. Schwannomin-interacting protein-1 isoform IQCJ-SCHIP-1 is a late component of nodes of Ranvier and axon initial segments. *J Neurosci*. 2008;28:6111–6117.
62. Boerkoel CF, Takashima H, Stankiewicz P, et al. Periaxin mutations cause recessive Dejerine-Sottas neuropathy. *Am J Hum Genet*. 2001;68:325–333.
63. Guilbot A, Williams A, Ravise N, et al. A mutation in periaxin is responsible for CMT4F, an autosomal recessive form of Charcot-Marie-Tooth disease. *Hum Mol Genet*. 2001;10:415–421.
64. Tiwari-Woodruff SK, Kaplan R, Kornbl um HI, Bronstein JM. Developmental expression of OAP-1/Tspan-3, a member of the tetraspanin superfamily. *J Neurosci Res*. 2004;77:166–173.
65. Goryunov D, Nightingale A, Bornfleth L, Leung C, Liem RK. Multiple disease-linked myotubularin mutations cause NFL assembly defects in cultured cells and disrupt myotubularin dimerization. *J Neurochem*. 2008;104:1536–1552.
66. Berger P, Berger I, Schaffitzel C, Tersar K, Volkmer B, Suter U. Multi-level regulation of myotubularin-related protein-2 phosphatase activity by myotubularin-related protein-13/set-binding factor-2. *Hum Mol Genet*. 2006;15:569–579.
67. Berger P, Niemann A, Suter U. Schwann cells and the pathogenesis of inherited motor and sensory neuropathies (Charcot-Marie-Tooth disease). *Glia*. 2006;54:243–257.
68. Cheishvili D, Maayan C, Smith Y, Ast G, Razin A. IKAP/hELP1 deficiency in the cerebrum of familial dysautonomia patients results in down regulation of genes involved in oligodendrocyte differentiation and in myelination. *Hum Mol Genet*. 2007;16:2097–2104.
69. Yao D, Kuwajima M, Chen Y, et al. Impaired long-chain fatty acid metabolism in mitochondria causes brain vascular invasion by a non-neurotropic epidemic influenza A virus in the newborn/suckling period: implications for influenza-associated encephalopathy. *Mol Cell Biochem.* 2007;299:85–92.
70. Rubin S, Liu D, Pletnikov M, et al. Wild-type and attenuated influenza virus infection of the neonatal rat brain. *J Neurovirol.* 2004;10:305–314.
71. Zou W, Ke J, Zhang A, et al. Proteomics analysis of differential expression of chicken brain tissue proteins in response to neurovirulent H5N1 avian influenza virus infection. *J Proteome Res*. doi:10.1021/pr100080x.
72. Kanduc D. Translational regulation of human papillomavirus type 16 E7 mRNA by the peptide SEQIKA, shared by rabbit alpha(1)-globin and human cytokeratin 7. *J Virol*. 2002;76:7040–7048.
73. Lucchese A, Guida A, Serpico R, Crincoli V, Scully C, Kanduc D. Interkeratin peptide-protein interactions that promote HPV16 E7 gene expression. *Int J Immunopathol Pharmacol*. In press 2010.
74. van Nimwegen E. Epidemiology: influenza escapes immunity along neutral networks. *Science*. 2006;314:1884–1886.
75. Park AW, Daly JM, Lewis NS, Smith DJ, Wood JL, Grenfell BT. Quantifying the impact of immune escape on transmission dynamics of influenza. *Science*. 2009;326:726–728.

76. Voeten JT, Bestebroer TM, Nieuwkoop NJ, Fouchier RA, Osterhaus AD, Rimmelzwaan GF. Antigenic drift in the influenza A virus (H3N2) nucleoprotein and escape from recognition by cytotoxic T lymphocytes. *J Virol*. 2000;74:6800–6807.
77. Silverstein AM. *A history of immunology*. 2nd ed. London: Elsevier; 2009:153–176.
78. Natale C, Giannini T, Lucchese A, Kanduc D. Computer-assisted analysis of molecular mimicry between human papillomavirus 16 E7 oncoprotein and human protein sequences. *Immunol Cell Biol.* 2000;78:580–585.
79. Kanduc D. Immunogenicity in peptide-immunotherapy: from self/nonself to similar/dissimilar sequences. *Adv Exp Med Biol.* 2008;640:198–207.
80. Wack A, Baudner BC, Hilbert AK, et al. Combination adjuvants for the induction of potent, long-lasting antibody and T-cell responses to influenza vaccine in mice. *Vaccine*. 2008;26:552–561.
81. Tritto E, Mosca F, de Gregorio E. Mechanism of action of licensed vaccine adjuvants. *Vaccine*. 2009;27:3331–3334.
82. Lucchese G, Stufano A, Kanduc D. Proteome-guided search for influenza A B-cell epitopes. *FEMS Immunol Med Microbiol*. 2009;57:88–92.
83. Kanduc D. "Self-nonself" peptides in the design of vaccines. *Curr Pharm Des*. 2009;15:3283–3289.
84. Kanduc D. Epitopic peptides with low similarity to the host proteome: towards biological therapies without side effects. *Expert Opin Biol Ther*. 2009;9:45–53.
85. Stufano A, Capone G, Pesetti B, Polimeno L, Kanduc D. Clustering of rare peptide segments in the HCV immunome. *Self/Nonself.* 2010;1:154–162.