Journal of Clinical Microbiology

Molecular Mimicry Revisited: Gut Bacteria and Multiple Sclerosis

Fred C. Westall

J. Clin. Microbiol. **2006, 44(6):2099. DOI: 10.1128/JCM.02532-05.**

JOURNAL OF CLINICAL MICROBIOLOGY, June 2006, p. 2099–2104 Vol. 44, No. 6
0095-1137/06/$08.00+0 doi:10.1128/JCM.02532-05

Molecular Mimicry Revisited: Gut Bacteria and Multiple Sclerosis

Fred C. Westall*

Institute for Disease Research, P.O. Box 890193, Temecula, California 92589

Received 5 December 2005/Accepted 5 April 2006

Molecular mimicry is a possible explanation for autoimmune side effects of microorganism infections. Protein sequences from a particular microorganism are compared to known autoimmune immunogens. For diseases such as multiple sclerosis (MS), where the infectious agent is unknown, guesses to its identity are made. Mimics are assumed to be rare. This study takes a radically different approach. Reported sequences from all known human bacterial and viral agents were searched for autoimmune immunogen mimics. Three encephalitogenic peptides, whose autoimmune requirements have been studied extensively, were selected for comparison. Mimics were seen in a wide variety of organisms. For each immunogen, the mimics were found predominantly in nonpathogenic gut bacteria. Since the three immunogens used in this study are related to MS, it is suggested that a microorganism responsible for autoimmune activity in MS could be a normally occurring gut bacterium. This would explain many of the peculiar MS epidemiological data and why no infective agent has been identified for MS and supports recently found MS gut metabolism abnormalities.

During the past 20 years molecular mimicry has been proposed as an explanation for autoimmune side effects of microorganism infections (3–5, 13, 15, 19, 41, 53). The process has involved comparing sequences within the proteins of a microorganism, usually a virus, with known human autoimmune immunogens. The probability of finding a particular sequence is 1 in 20^n, where 20 refers to the 20 amino acids and "*n*" is the number of residues within the sequence under consideration. Finding a duplicate sequence in a microorganism which is identical to a human sequence would appear to be impossible. However, it is not that difficult. First, some amino acids, such as leucine and glycine, are used far more than others, such as tryptophan and histidine. Thus, some sequences are more likely to appear than others. Second, amino acid usage within sequences varies with the species. Third, one does not need to find an exact match to a human immunogen. Each peptide immunogen is composed of amino acids that each contribute differently to the overall immunogenicity of the peptide. Therefore, substitutions can be made in the original immunogen without destroying its activity. For example, it can be easily estimated that 10,000 peptides would satisfy the encephalitogenic requirements for the tryptophan peptide, fswgaegqr, of myelin basic protein (43, 44, 47).

Quite often if a sequence, under consideration, contains some amino acids that match the immunogen in type and location, the sequence is considered a possible mimic. Very seldom is there a concern for the location of the microorganism's sequence within the cell or within its protein structure itself, i.e., whether the microorganism's sequence has access to the human immune system. More importantly, the individual contributions to autoimmunity of the individual amino acids of the human immunogen are not considered. Without considering the individual amino acid contributions of the sequence fswgaegqr responsible for autoimmune encephalomyelitis (EAE), for example, which one of the following sequences would be considered a likely encephalitogen (characters in boldface represent changes in the sequence from the original fswgae gqr): fswga**i**gqr, fsw**a**aegqr, or fswgae**g**er? The correct answer is the first sequence (47). Most would pick the second or the third. The glycine allows an essential bend at the glycine-tryptophan bond (17, 25). The glutamate/glutamine substitution results in negating the positive arginine (50). Therefore, to adequately analyze potential microorganism mimics, one needs to have a totally defined human immunogen. This presents a problem.

In 1970 I synthesized and chemically defined the requirements for EAE induction from the first myelin basic protein encephalitogenic sequence (8, 43, 47). The contribution of each of the nine amino acids to immunity was ascertained. Since then, numerous (Table 1) other short encephalitogenic regions have been discovered within the myelin basic protein and the other myelin proteins. With a few exceptions, the other regions have only been isolated and in some cases synthetically

TABLE 1. Major defined encephalitogenic sequences[a]

Encephalitogen	Active EAE species	Reference(s)
Myelin basic protein		
fswgaegqr-tryptophan pep.	Guinea pig, rabbit, monkey	7, 9, 47
tthygslpqk-mid-peptide	Rabbit, DR rat	31, 33
pqksqrtqdenpv-hyperacute	Lewis rat, monkey	45, 52
fklggrdsr	Rabbit, monkey	16, 51
Lipoprotein		
hslgkwlghpdkf	Mouse	39
ntwttcqsiafpsk	Mouse	14
yktticgkglsatv	Mouse	38
dyeylinvihafqyv	Mouse	1
Oligodendrocytic glycoprotein		
mevgwyrppfsrvvhlyrngk (also 1–22, 43–57, and 92–106)	Mouse	2, 24

[a] The decision of what should be included as a "major" encephalitogenic sequence is quite arbitrary. The result depends upon the laboratories testing the sequence and the animal being tested. The purpose of this table is primarily to show that there are several proteins that are encephalitogenic, and within these proteins there are numerous encephalitogenic regions.

* Mailing address: Institute for Disease Research, P.O. Box 890193, Temecula, CA 92589. Phone: (951) 699-5177. Fax: (951) 699-5177. E-mail: fcwestallidr@adelphia.net.

TABLE 2. Potential encephalitogenic regions within proteins from known human bacteria: mid-region encephalitogen from myelin basic protein hygslp

Organism[a]	Description[b]	Sequence[c]	Species
Burkholderia cepacia (−)	Selenophosphate synthase 46317954	v**hygslp**wla	
Enterococcus faecium (−)	Peptidase M20/M25/M40 family 29374880	y**qygtlp**vin	
Bacteroides spp. (−)	Hypothetical BT 1743 29347153	v**hfgslp**tye	*B. thetaiotaomicron*
	Hypothetical 53715314	f**hfgtlpvqe**	*B. fragilis*
	Hypothetical 53714072	s**hfgalp**qsi	*B. fragilis*
	DNA primase 60681803	g**lygalp**edl	*B. fragilis*
	Outer membrane-nutrient binding 29341170	d**rygslp**ntg	*B. thetaiotaomicron*
	Outer membrane-nutrient binding 29341066	s**kygnlp**nsl	*B. thetaiotaomicron*
Escherichia coli (−)	Hypothetical 75259460	i**lygslp**vef	
	Hypothetical 26250174	c**hygslt**pvw	
	Intimin 18202007	v**eygalp**vlg	
	ATP-binding D-ribose high-affinity transport rbs A 1790190	v**lygalp**rts	
	Membrane-permease yicL 26250402	a**rygtlp**vvw	
	Membrane uidC 34396001	y**sygslp**yrr	
	K88 fimibrial AC 120423	i**fygglp**rgs	
Streptococcus spp. (+)	Glycogen synthase 41017195	m**rygslp**lvh	*S. agalactiae*
	SepSi6A 58197483	k**hfgtlp**kvs	*S. suis*
	LepA 28895013	r**yygalp**ing	*S. pyogenes*
	Permease 24379316	k**tygtlpsqd**	*S. mutans*
Lactobacillus spp. (+)	Starch synthase 28376995	m**hygtlp**ivh	*L. johnsonii*
	Hypothetical 42518201	g**fygslp**tda	*L. johnsonii*
	Hypothetical 68160961	v**iygslpvwqe**	*L. reuteri*
	Aspartokinase 62515517	g**fygslp**ngv	*L. delbrueckii*
	α-Glucosidase 23003366	g**fygalp**ptn	*L. gasseri*
	DNA helicase 58337296	k**hfgsl**kln**qk**	*L. acidophilus*
Clostridium spp. (+)	Aspartokinase 28211949	g**fygslp**ngd	*C. tetani*
	Glycogen synthase 15025239	l**rygslp**ivr	*C. acetobutylicum*
	Acetoacetate decarboxylase 49036684	l**kygalp**vvt	*C. beijerinckii*
Bifidobacterium spp. (+)	Hypothetical 23466177	t**nygalp**gsi	*B. longum*
Klebsiella spp. (−)	*N*-Carbamyl-L-amino acid amidohydrolase 38016761	s**yygtlp**avd	*K. pneumoniae*
Fusobacterium nucleatum (+)	Hypothetical 34762382	k**hygsip**E**ks**	
Mycobacterium spp. (+)	ABC transport 840831	a**yygalp**liv	*M. tuberculosis*
	SecD 41407141	l**kygslp**lsf	*M. avis*
Salmonella enterica serovar Typhimurium (−)	Inner membrane transport yicl 20141848	a**qygtlp**vvg	
Chlamydophila pneumoniae (−)	Methionine amino peptidase 6647435	f**hygspp**fpk	
Haemophilus influenzae (−)	Glycogen synthase 1169909	l**qygtlp**lvr	
Bacillus cereus (+)	DEAD/DEAH box helicase 42779931	f**hygnlp**lii	
Pseudomonas fluor (−)	Lipopolysaccharide biosynthesis 70732417	q**qygslp**qgy	
Actinobacillus actinomycetemcomitans (+)	Unknown 10880891	t**hygtlp**qdl	

[a] Symbols in parentheses: +, gram positive; --, gram negative.
[b] Numbers represent access identification numbers on the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/).
[c] Boldface characters in the sequences indicate amino acids making major contributions to encephalitogenicity.

produced. Even though EAE is by far the most studied autoimmune disease and is considered to be the autoimmune model for multiple sclerosis (MS) and viral neuropathies, few of its immunogens have been adequately defined, i.e., the contribution of each of the amino acids to immunity has not been examined.

The molecular mimicry process has centered on examining an individual microorganism for a sequence which fulfills the requirements for a given human immunogen. I report here the results of the opposite approach. The proteins of a large number of human bacterial and viral nonpathogenic and pathogenic organisms are compared to the three most studied encephalitogenic sequences. The following questions are asked: (i) How plentiful are the matches? (ii) Are these sequences found more often in bacterial or viral organisms? (iii) Is there a particular species that contains these mimics? (iv) Do matches favor a particular encephalitogen? (v) How are these results relevant to multiple sclerosis? (vi) Are these results applicable to searches for other human immunogen mimics?

MATERIALS AND METHODS

Selection of encephalitogenic sites. EAE has for years been used as the autoimmune model for multiple sclerosis (18). This disease is induced when whole brain, myelin, one of the myelin proteins, or encephalitogenic peptides in an appropriate adjuvant are injected into test animals. The fact that one autoimmune disease, EAE, is initiated even when whole brain is used as the encephalitogen points to the special immunological nature of the myelin proteins. Table 1 lists the major encephalitogenic regions within each of the encephalitogenic myelin proteins. Most of the encephalitogenic regions found in the myelin proteins only have been sequenced. However, the major three encephalitogenic regions of the myelin basic protein have been extensively studied. The encephalitogenic contribution of each of the amino acids from the tryptophan region, fswgaegqr, has been examined thoroughly in the guinea pig (8, 29, 37, 43, 44, 47, 50). The midpeptide, tthygslpqk, has been studied in the DR rat and rabbit (31, 33). The hyperacute EAE site, pqksqrtqdenpv, has been analyzed in the Lewis rat (20–22, 45, 52). Furthermore, the tryptophan peptide is encephalitogenic in rabbits (9), guinea pigs, and monkeys (7) but not in Lewis rats. The midpeptide has been found active in rabbits and DR rats but not in guinea pigs. Finally, the hyperacute site is active in Lewis rats and monkeys but not in guinea pigs.

The relevance of any of these sequences to human disease, and particularly to MS, is questionable. The only way to determine whether any of these sequences are encephalitogenic in humans is to inject humans. This, of course, is not feasible. Many of the encephalitogenic regions are capable of reacting with activated lymphocytes from MS patients. However, this is no proof that they can induce disease in humans. Without reviewing this topic, five pertinent papers can be cited. First, Field and Caspary (10) showed that the tryptophan peptide could react with activated lymphocytes from a variety of cancer patients. These patients had neither EAE nor MS. The human cancer cells possess a surface protein that sequentially is similar to the tryptophan peptide. Second, Weizman et al. (42) reported a cell-mediated autoimmune response to human myelin basic protein in

TABLE 3. Potential encephalitogenic regions within proteins from known human bacteria: tryptophan encephalitogen from myelin basic protein fswgaegqr

Organism[a]	Description[b]	Sequence[c]	Species
Bacteroides spp. (−)	Hypothetical 29348610	li**wg**aeg**qr**l	*B. thetaiotaomicron*
	Outer membrane-nutrient binding 29341398	k**swg**adg**n**me	*B. thetaiotaomicron*
	TonB-dependent outer membrane receptor 60492730	**y**f**wg**ang**q**gn	*B. fragilis*
	Outer membrane-nutrient binding 29341954	f**swg**vdq**s**ny	*B. thetaiotaomicron*
	Outer membrane-nutrient binding 29337665	f**twg**age**n**lp	*B. thetaiotaomicron*
	Outer membrane-nutrient binding 29347092	f**swg**scslde	*B. thetaiotaomicron*
Enterococcus faecalis (−)	Hypothetical 29377638	**q**g**wg**aan**qr**g	
	Methionyl-tRNA synthetase 29375514	f**swg**ipl**kn**d	
	β-Galactosidase 68194203	**n**s**wg**adv**esp**	
Lactobacillus johnsonii (+)	Amino acid transporter 42519891	f**nwg**afkp**s**f	
	β-Galactosidase 42518789	**d**a**wg**aeg**te**t	
Bifidobacterium longum (+)	DNA polymerase III 23464753	**vhwg**tea**qr**r	
	Helicase 23466281	**y**n**wg**aef**tk**f	
Escherichia coli (−)	Sugar phosphate isomerase 26246223	f**gwg**ael**nmr**	
	Fumarase 75257030	**k**l**wg**aqt**qr**s	
Clostridium spp. (+)	Hypothetical 18145026	**dkwg**aeli**td**	*C. perfringens*
	Sugar phosphate isomerase 15895864	f**gwg**ael**n**le	*C. acetobutylicum*
Streptococcus spp. (+)	Ribose 5-phosphate isomerase 50589977	f**gwg**ael**nl**k	*S. suis*
	NADH flavin oxidoreductase 28896405	f**awg**aqy**q**l**e**	*S. pyogenes*
	Methionyl-tRNA synthase 55738410	f**swg**vkv**psd**	*S. thermophilus*
Haemophilus influenzae (−)	Ribose transport ATP binding 1172865	**tswg**ainw**gk**	
Helicobacter pylori (−)	DNA directed RNA polymerase 41017590	**dswg**aika**nr**	
Burkholderia pseudodomallei (−)	TPR repeat 67762886	**wrwg**qeg**qr**c	

[a] Symbols in parentheses: +, gram positive; --, gram negative.
[b] Numbers represent access identification numbers on the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/).
[c] Boldface characters in the sequences indicate amino acids making major contributions to encephalitogenicity.

76% of the autistic children studied. Here again, these children were not suffering from EAE, MS, or cancer. Finally, Spitler et al. (34, 35) showed that guinea pigs sensitized with encephalitogenic peptides having amino acid sequences different from that in the test protein did not show cellular immunity in vivo or in vitro to myelin basic protein, although the animals developed EAE. Einstein (6a) showed that antigen (MBP) suppression of EAE in guinea pigs did not require an intact tryptophan, and yet the tryptophan is an absolute requirement for disease induction. The specificity for disease induction and activated immune recognition are different.

Three sequences were compared in order that any conclusions presented would reflect all of the sites, not just the characteristic of one region. The tryptophan peptide has been studied by far the most. As part of my Ph.D. thesis (43), I systematically examined in guinea pigs the contribution to encephalitogenicity of each amino acid in the sequence. Further studies in later years also added to this initial work (8, 44, 47, 50).

Of the nine residues of the tryptophan peptide, fswgaegqr, five appear considerably more important. The five are the glutamine, a positively charged amino acid adjacent to the glutamine, the tryptophan located six residues from the glutamine toward the amino-terminal end with the glycine C terminally adjacent to it, and a hydrophobic region (preferably a ring structure) eight residues from the glutamine. The glutamine itself can be replaced by serine or asparagine, but not by glutamic acid. The negative charge of the glutamic acid will negate the positive of the adjacent amino acid. The glutamine appears to be contributing a hydrogen-bonding side chain. The phenylalanine and tryptophan form a hydrophobic pocket, called a molecular sandwich (26, 28, 43). The glycine permits a bend in the molecule at the tryptophan-glycine bond (17, 25).

Using the midpeptide, Smeltz et al. (33) showed by alanine replacement in the DA rat the importance of the "hygslp" sequence. Since only alanine substitutions were used, it is difficult to ascertain the variability acceptable. However, it appears that, unlike the tryptophan-glycine requirement, the tyrosine-glycine shows some flexibility. As stated above, the midpeptide is also active in rabbits. The midpeptide structurally is quite similar to the tryptophan peptide, which is also active in rabbits. Therefore, sequences that satisfy the general "hygslp" requirement and contain a glutamine adjacent to a charged amino acid at the C-terminal end are also noted (46).

The hyperacute sequence, pqksqrtqdenpv, is the least studied of the three encephalitogenic regions. The "tqde" sequence appears to be significant (32). A serine for threonine substitution converts the potency of the molecule from regular EAE to hyperacute EAE (20, 52). The aspartic acid appears to be essential. Although not as specific as the aspartic acid, the serine is important (21, 22). Therefore, small hydrophilic residues were included as acceptable for the replacement of the serine. With respect to the glutamic acid, the "tqde" data indicate some variability. For this position, glutamic acid, aspartic acid, glutamine, and asparagine were included.

Using the information on the requirements for the three encephalitogenic regions, bacterial and viral sequences were selected as potential encephalitogens.

Analytical process. The BLAST (basic local alignment search tool) program of the National Center for Biotechnology Information was used to ascertain potential encephalitogenic mimics (www.ncbi.nlm.nih.gov/BLAST). Four databases—nr, refseq, Swiss-Prot, and month—were used.

The proteins in the following human bacterial groups were examined for potential encephalitogenic mimics: *Klebsiella*, *Morganella*, *Proteus*, *Serratia*, *Enterococcus*, *Micrococcus*, *Streptococcus*, *Bifidobacterium*, *Lactobacillus*, *Prevotella*, *Bacteroides*, *Fusobacterium*, *Eubacterium*, *Burkholderia*, *Mycobacterium*, *Salmonella*, *Chlamydophila*, *Haemophilus*, *Bacillus*, *Pseudomonas*, *Actinobacillus*, *Clostridium*, and *Escherichia*.

The proteins in the following human virus groups were examined for potential encephalitogenic mimics: *Morbillivirus*, *Paramyxovirus*, *Rubulavirus*, *Pneumovirus*, *Filoviridae*, *Influenza virus*, *Arenaviridae*, *Bunyaviridae*, *Rotavirus*, *Coltivirus*, *Orthreovirus*, *Coronavirus*, *Torovirus*, *Flaviviridae*, *Togaviridae*, *Calicivirus*, *Astrovirus*, *Enterovirus*, *Rhinovirus*, *Hepatitis A virus*, *Hepatitis B virus*, *Hepatitis D virus*, *Hepatitis E virus*, *Herpesvirus B*, *Varicella-Zoster virus*, *Herpes simplex virus*, *Herpesvirus*, *Cytomegalovirus*, *Epstein-Barr virus*, *Adenoviridae*, *Parvovirus*, *Polyomavirus*, *Echovirus*, *Bluetongue virus*, and *Papillomavirus*.

Obviously, only the data that have been accumulated to date can be examined. The entire genomes of some microorganisms, including strains, have been ascertained. Others have not. Therefore, as more data are revealed more potential encephalitogenic mimics will be found. Comparative analysis of 16S rRNA sequences amplified from human feces indicated that less than 25% of the molecular species identified corresponded to known organisms (36). Obviously many bacteria in the gut, at least, have yet to be identified.

RESULTS AND DISCUSSION

Tables 2 (midregion), 3 (tryptophan peptide), and 4 (hyperacute site) present the potential encephalitogenic mimics within proteins from known human bacteria. Table 5 gives the same data for the potential encephalitogenic mimics within proteins from known human viruses. No attempt was made in these tables to separate pathogenic from nonpathogenic or-

TABLE 4. Potential encephalitogenic regions within proteins from known human bacteria: HEAE encephalitogen from myelin basic protein pqkshqrtqdenpv

Organism[a]	Description[b]	Sequence[c]	Species
Chlamydophila pneumoniae (−)	Hypothetical Cpn 0483 15618394	llpr**n**pr**tedq**n-	
Bacteroides spp. (−)	DNA polymerase 60491636	kdlfdef**tqde**ngn	*B. fragilis*
	Exported 60491363	drhpqgq**tedd**rpg	*B. fragilis*
	Hypothetical 60493947	dlekeer**tede**fma	*B. fragilis*
Fusobacterium nucleatum (+)	Calcium-transporting ATPase 19714607	depkdlt**tqde**dsy	
	Methyltransferase 34764012	serl**s**el**tqde**kfl	
	DNA gyrase A 19705415	evtedee**tede**elm	
	Phosphoenolpyruvate protein phosphotransferase 19705098	mekd**s**fp**tede**qfe	
Serratia spp. (−)	Piln 38176558	vysvekr**tqdq**ygi	*S. entomophila*
	Hypothetical 38259461	ryspdfq**tqde**fak	*S. marcescens*
	Afp18 48995204	depndni**tqde**lfr	*S. entomophila*
	Tn7 transposition 38259433	ealpeal**tede**vll	*S. marcescens*
	PTS system 1 21039017	mdrd**s**lp**tede**qfq	*S. marcescens*
Escherichia coli (−)	Adenylate cyclase 581058	iqfftee**tqde**ngf	
	FtsA 21321975	assy**s**vl**tede**rel	
	ATPase 73853188	ndva**s**rr**tede**rrl	
	Hypothetical 26250231	etgr**s**pr**tede**hmi	
	Peptide synthetase 26248278	ctll**n**rm**tede**nsw	
	FotE 29293010	vwlvqtw**tede**nks	
	Outer membrane usher 15800422	shhtede**tede**tfi	
	Ycdy oxidoreductase 15830666	swledq**stede**sea	
Clostridium spp. (+)	Phosphatidylserine synthase 150323686	iarmckr**tede**klf	*C. acetobutylicum*
	Regulatory recX 18145587	dkls**n**id**tede**ndt	*C. perfringens*
	Yqe 15024218	sylpqql**tede**iri	*C. acetobutylicum*
	Hypothetical 28210235	lqrk**s**er**tede**qre	*C. tetani*
	Glucose-inhibited division B 28209871	kinltai**tede**dii	*C. tetani*
	Ribose recycling factor 22001946	kkdnsi**tede**mks	*C. acetobutylicum*
Enterococcus faecalis (−)	ABC transporter-ATP binding 29376690	yeplemv**tqde**kvi	
	Phenomone binding 29376080	gsmd**s**if**tqde**sin	
	Hypothetical 29376056	mvrf**s**lv**tqde**tin	
	Hypothetical 29375363	alhkmfa**tqde**wgn	
	PrgE 59616089	havt**n**fl**tqde**fee	
	Arginine repressor 39931070	imqqeie**tqde**lit	
	Conjugal transfer 29377006	ellierl**tede**lyy	
	V-type sodium ATP synthase F	rkneeev**tede**qhk	
Bifidobacterium longum (+)	Hypothetical 23465454	aqvltem**tqde**snp	
	RpoB 23465772	leketle**tqde**alv	
	α-Galactosidase 23464793	dsygttl**tede**lla	
Streptococcus spp. (+)	Hypothetical 15674763	rvhytft**tedd**npk	*S. pyogenes*
	PTS system II 28896695	kshimtk**tede**akl	*S. pyogenes*
	Cysteine tRNA synthetase 30316153	elga**s**gr**tdee**tar	*S. pneumoniae*
Lactobacillus spp. (+)	Hypothetical 58336799	timncdv**tqde**dgk	*L. acidophilus*
	Hypothetical 42518235	yfyfdpn**tqde**dyq	*L. johnsonii*
	RNase 6 62514503	itke**n**tp**tede**kdn	*L. casei*
	Hypothetical 68160989	iinrqrq**tede**knk	*L. reuteri*
	Holliday junction DNA binding resolvase 62515837	yvae**n**lf**tede**pve	*L. delbrueckii*
	NADH dehydrogenase 28377235	flqk**s**lp**tede**iil	*L. plantarum*
	FtsY 42519391	esaeevt**tede**qer	*L. johnsonii*
	Glutamyl-tRNA synthetase 38258342	kaye**s**ym**tede**lsa	*L. plantarum*

[a] Symbols in parentheses: +, gram positive; --, gram negative.
[b] Numbers represent access identification numbers on the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/).
[c] Boldface characters in the sequences indicate amino acids making major contributions to encephalitogenicity.

ganisms. Moreover, no attempt to separate the organisms by location of potential "infection" was made.

The bacteria present 111 potential encephalitogenic mimics: 39 midregion, 23 tryptophan peptide, and 49 hyperacute site. Thus, there are numerous sequences that are potentially encephalitogenic in bacteria. No one immunogen dominates the group. In fact, the differences probably reflect more on the specific requirements imposed on selection. The tryptophan site, which is by far the most studied (and presents the most restrictions), has the fewest mimics but still presents a large selection. The sites are found in most of the bacteria examined. In fact, the number of sites found in a particular bacterium probably reflects more on the amount of data known about the species than any bacterial restriction. Gram stain characteristics seem to have little significance. Finally, the locations of the sites are in proteins located in all parts of the bacterial cell—inside, outside, etc. The locations of the sites within the proteins are also quite variable.

For molecular mimicry to work the immunogen must be "visible" to the immune system. It has therefore been assumed that it must be located on the surface of a protein that is either excreted by the microorganism or is on its surface. This is not necessarily true. The microorganism itself can initiate an immune response. Cells are hydrolyzed, and their contents are released. A great amount of digestive enzyme activity is present. All this can "expose" potential immunogens (23). It certainly happens within active EAE lesions. These "new" immunogens can be processed by the already-present immune cells.

The one feature of all of the diseases that have been related to molecular mimicry is that the incidence of the primary infection is very high, e.g., measles and influenza. However, the

TABLE 5. Potential encephalitogenic regions within proteins from known human viruses

Section[a]	Virus	Description[b]	Sequence[c]
A	*Hepatitis B virus*	Polymerase 28812222	h**hygtlp**nlh
		Polymerase 33468377	g**cygslp**qdh
		Reverse transcriptase 76253219	g**sygslp**qeh
		Reverse transcriptase 76253625	h**qygtlp**slh
	Influenza C virus	RNA-directed RNA polymerase 133531	k**sygslp**elf
	Papillomavirus	Replication E1 9628545	f**kygtlp**swv
B	*Hepatitis C virus*	Polyprotein 1381032	**ys**w**g**ane**td**
C	*Ornithogalum mosaic virus*	Genome polyprotein 130497	vdpltga**tqde**npl
	Bluetongue virus	Outer capsid 60416215	qrea**s**er**sqde**ikm
	Echovirus type 32	Polyprotein 34485450	kyss**n**at**tqde**qym
	Herpesvirus 1	Capsid 59538	gglv**s**wv**tqde**las

[a] Sections: A, midregion encephalitogen from myelin basic protein hygslp; B, tryptophan encephalitogen from myelin basic protein fswgaegqr; C, HEAE encephalitogen from myelin basic protein pqkshqrtqdenpv.

[b] Numbers represent access identification numbers on the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/).

[c] Boldface characters in the sequences indicate amino acids making major contributions to encephalitogenicity.

TABLE 6. Comparison of urinary 5-HIAA/tryptophan ratios from MS patients in relapse to those not associated with relapse[a]

Patient	Relapse		Nonrelapse		% Difference
	Sample no.	5-HIAA/Trp ratio	Sample no.	5-HIAA/Trp ratio	
1	3	0.099	4	0.134	−26
2	4	0.102	1	0.249	−59
3	2	0.084	5	0.117	−28
4	2	0.009	4	0.108	−37
5	4	0.167	4	0.186	−10
6	2	0.149	3	0.387	−23
7	6	0.122	2	0.144	−17
8	1	0.245	4	0.307	−20
9	3	0.077	6	0.091	−15
10	4	0.178	7	0.181	−2
11	3	0.000	3	0.071	−∞
12	2	0.123	2	0.090	+27

[a] Relapse average = 0.113; nonrelapse average = 0.172. At each clinic visit, the patients were assigned a "phase" according to the definitions of the Schumacher Committee (30) formulated in 1965 after standard electrophysiological testing. Urine samples were collected periodically for 2 years from the patient's visits. The first voided urine sample on the day of each clinic visit was collected and stored at −20°C in the dark before analysis. Tryptophan (Trp) and serotonin sulfate contents were determined by means of a Beckman automated amino analyzer. The concentrations of 5-HIAA were measured by the procedure of Udenfriend et al. (40).

incidence of the autoimmune "side effect" is extremely low, e.g., the viral neuropathies incidence is 1/10,000. Any hypothesis that attempts to explain this process must take into account the disparity in incidence rates.

One explanation is taken from experimental autoimmune disease research. In order to induce EAE in animals, both the immunogen and the adjuvant must be injected at the same time. In fact, it has been shown that the two form a complex (27, 28, 48). The need for two separate compounds to be present simultaneously is used to explain the low incidence of viral and postvaccinal neuropathies (49). It has been shown that secondary infections are significantly present during the induction of these neuropathies. The cell wall material of the microorganism responsible for the secondary infection is a source of the adjuvant.

How can this be used as an explanation of multiple sclerosis? In this disease it is hard to find one infection (other than the autoimmune reaction at the lesion), much less two. Tables 2 to 5 present more than potential encephalitogenic sites. They show the probability of finding such sites resides in bacteria and not viruses and, in fact, in the bacteria located in the most concentrated bacterial area of the human anatomy—the gut. Most of the gut bacteria are not pathogenic but form a symbiotic relationship with the individual. Therefore, the requirement for the first "infection" is satisfied by the presence of nonpathogenic bacteria.

What about the need for an adjuvant molecule? Adjuvant molecules are portions of the bacterial cell walls. The simplest is *N*-acetylmuramyl dipeptide (MDP) (6). However, MDP does not aid all encephalitogens in producing EAE. Larger portions apparently are needed for some encephalitogens (28). Fox et al. have shown quite convincingly that adjuvant molecules normally are not in human body fluids or tissue, except the gut (11, 12). So, apparently, the two groups of substances, potential immunogens (mimics) and adjuvant molecules, known to be required for an autoimmune response are normally found together in the normal human gut. Although it appears that the adjuvant itself cannot cross the gut, the adjuvant-immunogen complex probably can. In fact the MDP-tryptophan peptide complex forms a spherical ball with one hemisphere hydrophobic and the other hydrophilic. This soap-like structure is ideal for membrane penetration (28, 48).

Why is the incidence of MS so low? Although both the immunogen and adjuvant molecules are present, they must be processed in such fashion that the correct units are made in high enough concentration to form the complex. This will depend entirely upon the nature of the gut and its hydrolytic enzymes. The gut is a dynamically evolving biosystem whose bacterial content is changing and therefore the concentrations and types of hydrolytic enzymes continually vary.

There is other evidence that the gut is involved in MS. Serotonin metabolism is greatly altered during multiple sclerosis relapses. This has been seen by measuring the 5-hydroxy indoleacetic acid (5-HIAA)/tryptophan ratio from urine (Table 6). Ninety percent of the serotonin metabolism is derived from the gut and not the nervous system. Furthermore, one of the adjuvant molecules, MDP, is known to mimic serotonin (28).

Rather than study a single organism, the present study examined a large number of nonpathological and pathological human bacteria and viruses in order to ascertain the prevalence of encephalitogenic mimics. There appear to be many potential encephalitogens within bacteria and viral cells. It also seems that the most probable source of these mimics is the normal gut. This information has been applied to MS. However, the procedure used is just as valid for the location of any potential immunogenic mimic. Therefore, it has a broad application to the field of molecular mimicry.

REFERENCES

1. **Amor, S., D. Baker, N. Groome, and J. L. Turk.** 1993. Identification of a major encephalitogenic epitope of proteolipid protein (residues 50–70) for

the induction of EAE in Biozzi AB/H and nonobese diabetic mice. J. Immunol. **150:**5666–5672.
2. **Amor, S., N. Groome, C. Linington, M. M. Morris, K. Dornmair, M. V. Gardinier, J.-M. Matthieu, and D. Baker.** 1994. Identification of epitopes of myelin oligodendrocyte glycoprotein for the induction of EAE in SJL and Biozzi AB/H mice. J. Immunol. **153:**4339–4356.
3. **Bachmaier, K., N. Neu, L. M. de la Maza, S. Pal, A. Hessel, and J. M. Penninger.** 1999. *Chlamydia* infection and heart disease linked through antigenic mimicry. Science **283:**1335–1343.
4. **Benoist, C., and D. Mathis.** 2001. Autoimmunity provoked by infection: how good is the case for cell mimicry? Nat. Immunol. **2:**797–801.
5. **Carnegie, P. R., and R. Weise.** 1987. Visna and MBP. Nature **329:**294–295.
6. **Chedid, L., F. Audibert, P. Lefrancier, J. Choay, and E. Lelderer.** 1976. Modulation of the immune response by a synthetic adjuvant and analogs. Proc. Natl. Acad. Sci. USA **73:**2472–2475.

6a. **Einstein, R. E.** 1972. Suppression of EAF, p. 121–129. *In* E. J. Field, T. M. Bell, and P. R. Carnegie (ed.), Multiple sclerosis: progress in research. North Holland Publishing Co., Amsterdam, The Netherlands.

7. **Eylar, E. H., S. Brostoff, J. Jackson, and H. Carter.** 1972. Allergic encephalomyelitis in monkeys by a peptide from the A1 protein. Proc. Natl. Acad. Sci. USA **69:**617–619.
8. **Eylar, E. H., J. Caccam, J. J. Jackson, F. C. Westall, and A. B. Robinson.** 1970. EAE: synthesis of disease-inducing site of basic protein. Science **168:** 1220–1223.
9. **Eylar, E. H., F. C. Westall, and S. Brostoff.** 1971. Allergic encephalomyelitis. J. Biol. Chem. **246:**3418–3424.
10. **Field, E. J., and E. A. Caspary.** 1971. Lymphocyte sensitization in cancer. Lancet **i:**189–190.
11. **Fox, A., and K. Fox.** 1991. Rapid elimination of a synthetic adjuvant peptide from the circulation after systemic administration and absence of detection of natural muramyl peptides in normal serum at current analytical limits. Infect. Immun. **59:**1202–1205.
12. **Fox, A., K. Fox, B. Christensson, D. Harrelson, and M. Kramer.** 1996. Absolute identification of muramic acid, at trace levels, in human septic synovial fluids in vivo and absence in aseptic fluids. Infect. Immun. **64:**3911–3915.
13. **Fujinami, R. S., and M. B. Oldstone.** 1985. Amino acid homology between the encephalitogenic site of MBP and virus: mechanism for autoimmunity. Science **230:**1043–1045.
14. **Greer, J. M., V. K. Kuchroo, R. A. Sobel, and M. B. Lees.** 1992. Identification and characterization of a second encephalitogenic determinant of myelin proteolipid protein (residues 178–191) for SJL mice. J. Immunol. **149:**783–788.
15. **Jahnke, U., E. D. Fisher, and E. C. Alvord, Jr.** 1985. Sequence homology between certain viral proteins and proteins related to encephalomyelitis and neuritis. Science **229:**282–284.
16. **Karkhanis, Y. D., D. J. Carlo, S. W. Brostoff, and E. H. Eylar.** 1975. Allergic encephalomyelitis. Isolation of an encephalitogenic peptide activity in the monkey. J. Biol. Chem. **250:**1718–1722.
17. **Khanarian, G., S. A. Margeson, W. J. Moore, S. J. Pasaribu, and F. C. Westall.** 1979. Biophys. Biochem. Res. Commun. **87:**236–243
18. **Kies, M. W., and E. C. Alvord, Jr.** 1959. Myelin basic protein, p. 239–299. *In* Allergic encephalomyelitis. Charles C Thomas, Springfield, Ill.
19. **Lang, H. L. E., H. Jacobson, S. Ikemizo, C. Andersson, K. Harla, L. Madsen, P. Hjorth, L. Sondergaard, D. Svejgard, K. Wucherpfennig, D. I. Stuart, J. L. Bell, E. Y. Jones, and L. Fugger.** 2002. A functional and structural basis for the cross-reactivity in multiple sclerosis. Nat. Immunol. **3:**940–943.
20. **Lennon, V. A., F. C. Westall, M. Thompson, and E. Ward.** 1976. Antigen, host, and adjuvant requirements for induction of hyperacute EAE. Eur. J. Immunol. **6:**805–810.
21. **Mannie, M. D., P. Y. Paterson, D. C. U'Prichard, and G. Flouret.** 1989. Encephalitogenic and proliferative responses of Lewis rat lymphocytes distinguished by position 75- and 80-substituted peptides of myelin basic protein. J. Immunol. **142:**2608–2616.
22. **Mannie, M. D., P. Y. Paterson, D. C. U'Pritchard, and G. Flouret.** 1985. Induction of EAE in Lewis rats. Proc. Natl. Acad. Sci. USA **82:**5515–5519.
23. **Matsuo, A., G. C. Lee, K. Terai, K. Takami, W. F. Hickey, E. G. McGeer, et al.** 1997. Unmasking of an unusual MBP epitope during the process of myelin degeneration in humans: a potential mechanism for the generation of autoantigens. Am. J. Pathol. **150:**1253–1266.
24. **Mendel, I., N. Kerlero de Rosbo, and A. Ben-Nun.** 1996. Delineation of the minimal encephalitogenic epitope within the immunodominant region of myelin oligodendrocyte glycoprotein. Eur. J. Immunol. **26:**2470–2479.
25. **Moore, W. J., B. E. Chapman, G. E. James, G. Khanarian, S. A. Margetson, S. J. Pasaribu, and F. C. Westall.** 1981. Conformation of encephalitogenic proteins and peptides. Excerpta Medica **546:**198–208.
26. **Root-Bernstein, R. S., and F. C. Westall.** 1984. Serotonin binding sites. Brain Res. Bull. **12:**425–436.
27. **Root-Bernstein, R. S., and F. C. Westall.** 1986. Complementarity between dual antigens in the induction of EAE. J. Int. Dev. Biol. **2:**1–10.
28. **Root-Bernstein, R. S., and F. C. Westall.** 1990. Serotonin binding sites. II. Muramyl dipeptide binds serotonin binding sites on MBP, LHRH, and MSH-ACTH 4-10. Brain Res. Bull. **25:**827–841.
29. **Sasaki, Y.** 1974. Studies on encephalitogenic fragments of myelin basic protein. III. Synthesis of H-Arg-Phe-Trp-Gly-Ala-glu-Gly-Asn-Arg-OH as an analog of encephalitogenic decapeptide. Chem. Pharm. Bull. **22:**2188–2191.
30. **Schumacher, G. A., G. Beebe, R. F. Kibler, L. T. Kurland, J. F. Kurtzke, F. McDowell, B. Nagler, W. A. Sibley, W. W. Tourtellotte, and T. L. Wilmon.** 1965. Problems of experimental trials in multiple sclerosis. Ann. N. Y. Acad. Sci. **122:**552–568.
31. **Shapira, R., F. C. H. Chou, S. McKneally, E. Urban, and R. F. Kibler.** 1971. Biological activity and synthesis of an encephalitogenic determinant. Science **172:**736–738.
32. **Smeltz, R. B., M. H. M. Wauben, N. A. Wolf, and R. H. Swanborg.** 1999. Critical requirement for aspartic acid at position 82 of myelin basic protein 73–86. J. Immunol. **162:**829-836.
33. **Smeltz, R. B., N. A. Wolf, and R. H. Swanborg.** 1998. Delineation of two encephalitogenic myelin basic protein epitopes for DA rats. J. Neuroimmunol. **87:**43–48.
34. **Spitler, L. E., C. M. Von Muller, H. H. Fudenberg, and E. H. Eylar.** 1972. EAE: dissociation of cellular immunity to brain protein and disease. J. Exp. Med. **136:**156–174.
35. **Spitler, L. E., C. M. Von Muller, and J. Young.** 1975. Study of cellular immunity to the encephalitogenic determinant. Cell. Immun. **15:**143–151.
36. **Suau, A., R. Bonnet, M. Sutren, J. J. Godon, G. R. Gibson, M. D. Collins, and J. Dore.** 1999. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Appl. Environ. Microbiol. **65:**4799–4807.
37. **Suzuki, K., and Y. Sasaki.** 1974. Studies on encephalitogen fragments of myelin basic protein. IV. Synthesis of glycine analogs of tryptophan containing fragment. Chem. Pharm. Bull. **22:**2181–2187.
38. **Tuohy, V. K., Z. Lu, R. A. Sobel, R. A. Laursen, and M. B. Lees.** 1988. A synthetic peptide from myelin proteolipid induces EAE. J. Immunol. **141:** 1126–1130.
39. **Tuohy, V. K., Z. Lu, R. A. Sobel, R. A. Laursen, and M. B. Lees.** 1989. Identification of an encephalitogenic determinant for myelin proteolipid protein for SJL mice. J. Immunol. **142:**1523–1527.
40. **Udenfriend, S., E. Titus, and H. Weissbach.** 1955. The identification of 5-hydroxy-3-indoleacetic acid in normal urine and a method for its assay. J. Biol. Chem. **261:**499–505.
41. **Weise, M. J., and P. R. Carnegie.** 1988. An approach to searching protein sequences for superfamily relationships or chronic similarities relevant to the molecular mimicry hypothesis: applications to the basic protein of myelin. J. Neurochem. **51:**1267–1272.
42. **Weizman, A., R. Weizman, G. A. Szekely, H. Wijsenbeak, and E. Livni.** 1982. Abnormal immune response to brain tissue antigen in the syndrome of autism. Am. J. Psychiatr. **139:**1462–1465.
43. **Westall, F. C.** 1970. Applications of solid phase peptide synthesis. Ph.D. thesis. University of California, San Diego.
44. **Westall, F. C.** 1972. Solid phase peptide synthesis as applied to EAE, p. 72–77. *In* E. J. Field, T. M. Bell, and P. R. Carnegie (ed.), Multiple sclerosis: progress in research. North Holland, Amsterdam, The Netherlands.
45. **Westall, F. C.** 1977. Hyperacute allergic encephalomyelitis: a single determinant. Immunol. Commun. **6:**227–237.
46. **Westall, F. C.** 1978. Possible change of a recognition mechanism for determining encephalotogenicity. J. Theor. Biol. **71:**465–468.
47. **Westall, F. C., A. B. Robinson, J. Caccam, J. J. Jackson, and E. H. Eylar.** 1971. Essential chemical requirement for induction of allergic encephalomyelitis. Nature **229:**22–24.
48. **Westall, F. C., and R. S. Root-Bernstein.** 1983. An explanation of prevention and suppression of EAE. Mol. Immunol. **20:**169–177.
49. **Westall, F. C., and R. S. Root-Bernstein.** 1986. Cause and prevention of postinfectious and post-vaccinal neuropathies in light of a new theory of autoimmunity. Lancet **ii:**251–252.
50. **Westall, F. C., and M. Thompson.** 1977. Further definition of the encephalitogenic region in guinea pigs. Immunol. Commun. **6:**23–31.
51. **Westall, F. C., and M. Thompson.** 1978. An encephalitogenic region for rabbits. Immunochemistry **15:**189–191.
52. **Westall, F. C., M. Thompson, and V. A. Lennon.** 1977. Hyperacute autoimmune encephalomyelitis-unique determinant conferred by serine in a synthetic autoantigen. Nature **269:**425–427.
53. **Zhao, Z.-S., F. Granucci, L. Yeh, P. Schaffer, and H. Cantor.** 1998. Molecular mimicry by herpes simplex virus-type I: autoimmune disease after viral infection. Science **230:**1043–1051.