Virus Research 238 (2017) 124–132

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

A comparative analysis on the physicochemical properties of tick-borne encephalitis virus envelope protein residues that affect its antigenic properties

Yu. S. Bukin[a,b,*], Yu. P. Dzhioev[c,d], S.E. Tkachev[e], I.V. Kozlova[d], A.I. Paramonov[d], D. Ruzek[f,g], Z. Qu[h,i], V.I. Zlobin[c]

[a] Limnological Institute of the Siberian Branch of the Russian Academy of Sciences, Ulan- Batorskaya Str., 3, 664033 Irkutsk, Russia
[b] Irkutsk National Research Technical University, Lermontov Str., 83, 664074 Irkutsk, Russia
[c] Irkutsk State Medical University, Institute of Biomedical Technology, Irkutsk, Russia
[d] Science Center of Family Health Problems and Human Reproduction, Irkutsk, Russia
[e] Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
[f] Faculty of Science, University of South Bohemia and Biology Centre, Institute of Parasitology, Czech Academy of Sciences, Branisovska 31, České Budějovice (Budweis) 37005, Czech Republic
[g] Veterinary Research Institute, Hudcova 70, Brno 62100, Czech Republic
[h] Department of Microbiology, School of Public Health, Harbin Medical University, China
[i] Department of Natural Focus Disease Control, Institute of Environment-Associated Disease, Sino-Russia Joint Medical Research Center, Harbin Medical University, China

ARTICLE INFO

Keywords:
Tick-borne encephalitis virus
E protein
Physicochemical properties amino acid residue
Antigen
Antibody

ABSTRACT

This work is dedicated to the study of the variability of the main antigenic envelope protein E among different strains of tick-borne encephalitis virus at the level of physical and chemical properties of the amino acid residues. E protein variants were extracted from then NCBI database. Four amino acid residues properties in the polypeptide sequences were investigated: the average volume of the amino acid residue in the protein tertiary structure, the number of amino acid residue hydrogen bond donors, the charge of amino acid residue lateral radical and the dipole moment of the amino acid residue. These physico-chemical properties are involved in antigen-antibody interactions. As a result, 103 different variants of the antigenic determinants of the tick-borne encephalitis virus E protein were found, significantly different by physical and chemical properties of the amino acid residues in their structure. This means that some strains among the natural variants of tick-borne encephalitis virus can potentially escape the immune response induced by the standard vaccine.

1. Introduction

One of the most dangerous for humans natural foci neuroinfections in the territory of Northern Eurasia is tick-borne encephalitis (TBE) (Zilber, 1945; Panov, 1956; Dumpis et al., 1999; Charrel et al., 2004; Goodman et al., 2005; Bogovic and Strle, 2015). The pathogen that causes TBE, tick-borne encephalitis virus (TBEV), is divided into three major subtypes, namely Far Eastern, Siberian and European (Zlobin et al., 1996; Ecker et al., 1999a,b; Zlobin et al., 2001a,b). Phylogenetic analysis on the full genome and genome fragments sequences of TBEV show a clear clusterization of TBEV strains to one of three subtypes (Zlobin et al., 2001a,b; Demina et al., 2010). Two genetic variants of the virus (prototype strains 886-84 and 178-79) were also described as possible new TBEV subtypes (Demina et al., 2012; Kozlova et al., 2013).

The standard method of TBE prevention for humans is vaccination. The TBEV envelope protein (E protein) is the main antigen that activates the immune response during both infection and vaccination (Mandl et al., 1989; Rey et al., 1995). The vaccine is produced by inactivated TBEV particles grown in chick embryo cells cultures (Bock et al., 1990; Klockmann et al., 1991; Vorobyova and Rasschepkina, 1992; Leonova et al., 2007; Orlinger et al., 2011; Morozova et al., 2012). Currently, four commercially available vaccines for TBEV prevention are produced, namely FSME Immun Inject (“Baxter”, Austria) based on Neudoerfl strain corresponding to TBEV European subtype; Encepur (“Novartis Vaccines and Diagnostics” Germany) based on K23 strain of European subtype; Entsevir (“Microgen”, Russia); and, the vaccine by M.P. Chumakov (Moscow, Russia) produced by the Institute of Poliomyelitis based on 205 and Viral Encephalitis based on Sofjin strains, corresponding to the Far-Eastern subtype (Wagner et al., 2004a; Zlobin et al., 2009; Zavadska et al., 2013; Morozova et al., 2014).

* Corresponding author at: Limnological Institute of the Siberian Branch of the Russian Academy of Sciences, Ulan- Batorskaya Str., 3, 664033 Irkutsk, Russia.
E-mail address: bukinyura@mail.ru (Y.S. Bukin).

http://dx.doi.org/10.1016/j.virusres.2017.06.006
Received 30 January 2017; Received in revised form 22 May 2017; Accepted 2 June 2017
Available online 15 June 2017
0168-1702/

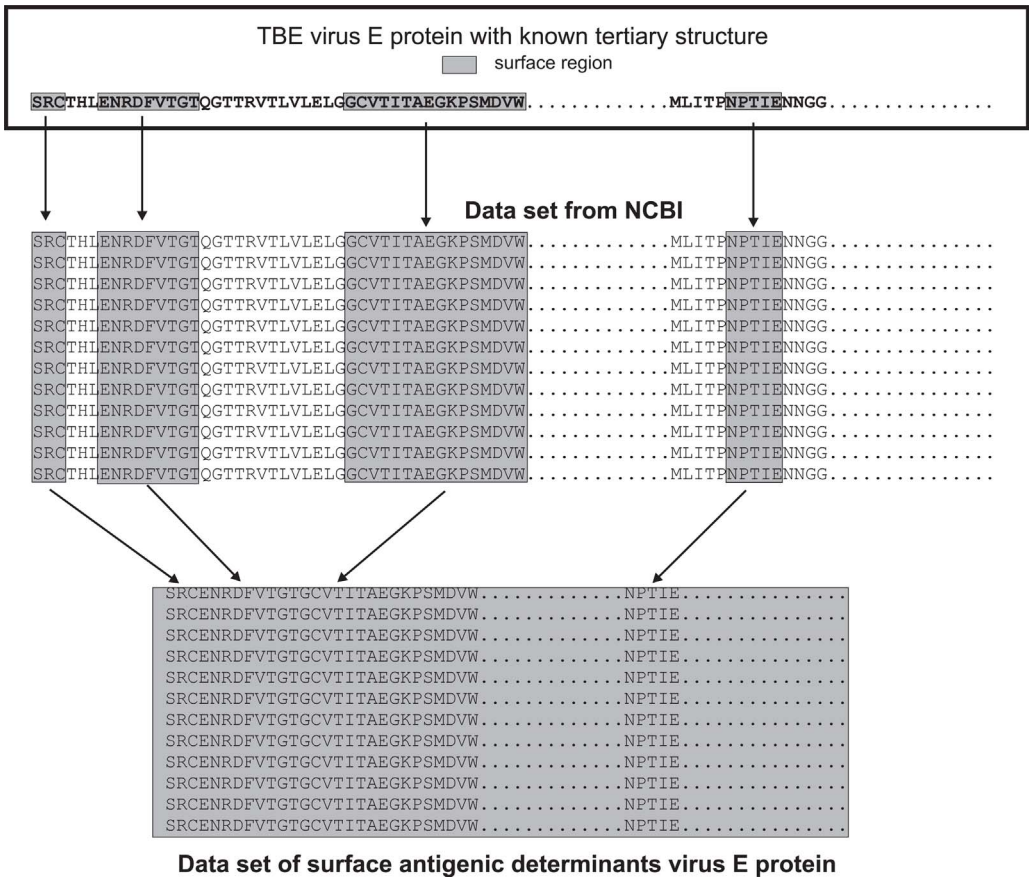

Fig. 1. The scheme for isolating the surface antigenic determinants.

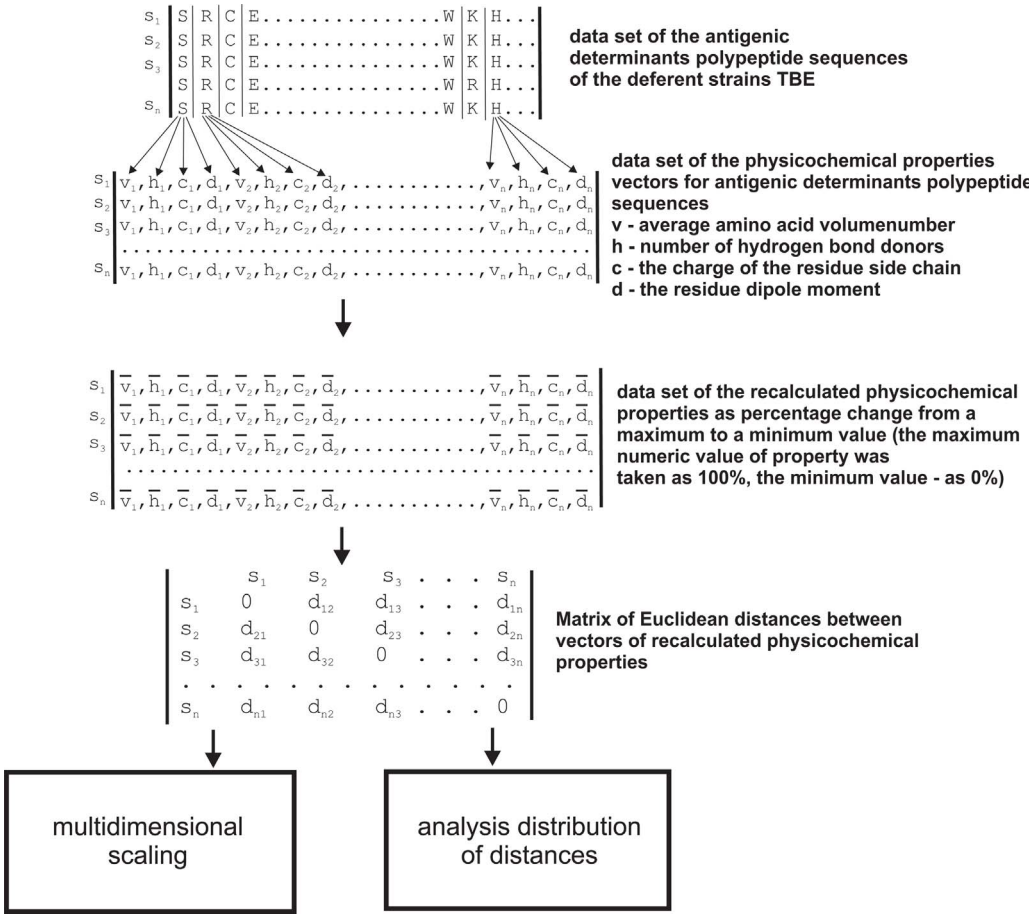

Fig. 2. The scheme of statistical analysis pipeline. s_1, s_2, s_3, ……, s_n – the indices of the amino acid sequence of the studied strains of the virus.

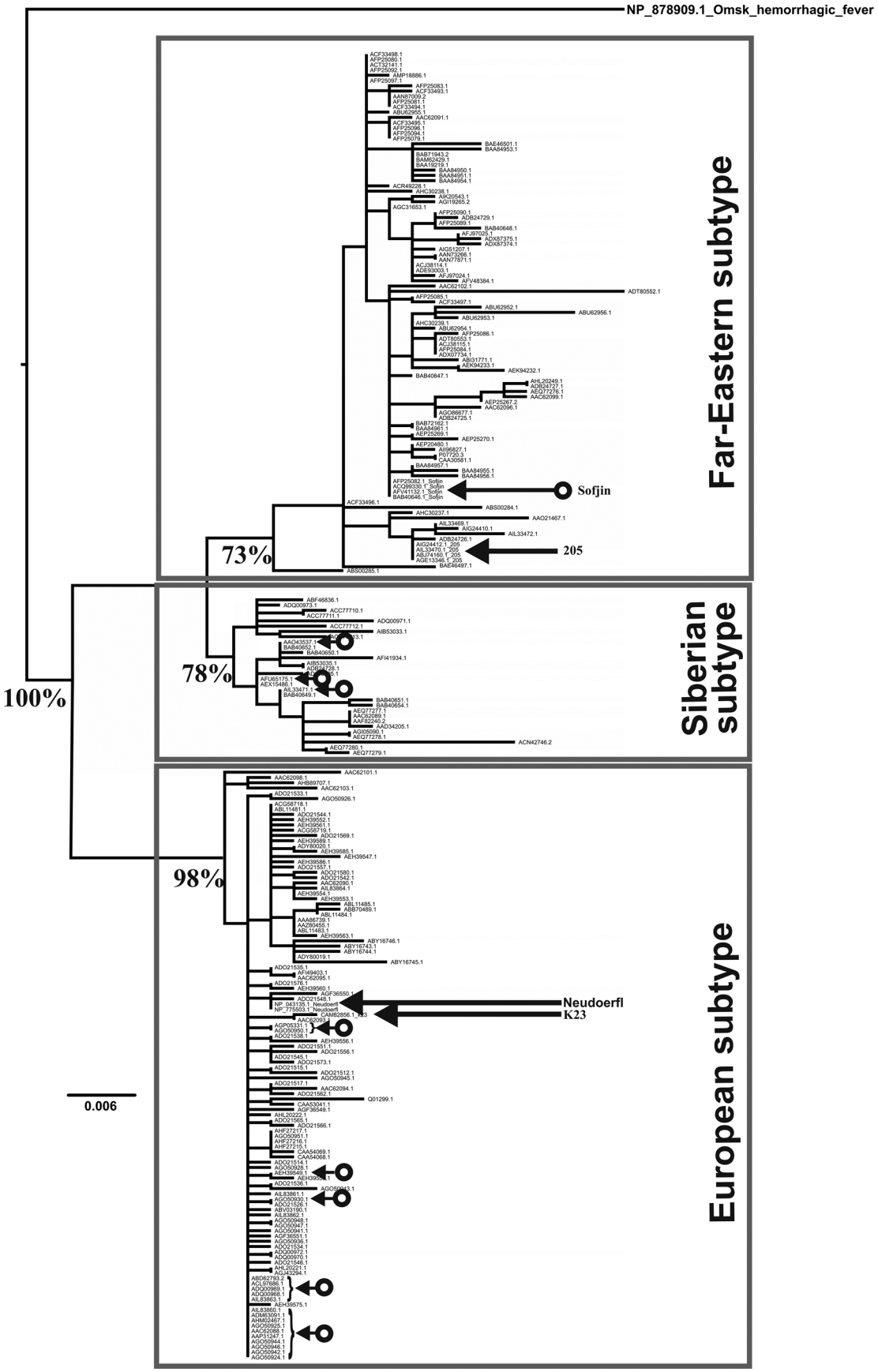

Fig. 3. The phylogenetic tree reconstructed by the maximum likelihood method using the JTT + G model. In the nodes separating the subtypes are the values of bootstrap support. Simple arrows indicate the position of vaccine strains. Arrows with circles indicate the position of the strains most suitable for the creation of polyvalent vaccines, each subtype has its own strains.

Table 1
The parts of E protein amino acid sequence facing to the surface of the viral particle (marked with gray), TBEV Sofjin strain.

SRCTHLENRDFVTGTQGTTRVTLVLELGGCVTITAEGKPSMDVWLDSIYQENPAKTREYCLHAKLSDTKVAARCPT
MGPATLAEEHQSGTVCKRDQSDRGWGNHCGLFGKGSIVTCVKASCEAKKKATGHVYDANKIVYTVKVEPHTGDYVA
ANETHSGRKTASFTVSSEKTILTMGDYGDVSLLCRVASGVDLAQTVILELDKTSEHLPTAWQVHRDWFNDLALPWK
HEGAQNWNNAERLVEFGAPHAVKMDVYNLGDQTGVLLKSLAGVPVAHIDGTKYHLKSGHVTCEVGLEKLKMKGLTY
TMCDKTKFTWKRIPTDSGHDTVVMEVAFSGTKPCRIPVRAVAHGSPDVNVAMLITPNPTIENNGGGFIEMQLPPGD
NIIYVGELSHQWFQKGSSIGRVFQKTRKGIERLTVIGEHAWDFGSTGGFLTSVGKALHTVLGGAFNSLFGGVGFLP
KILVGVVLAWLGLNMRNPTMSMSFLLAGGLVLAMTLGVGA

Table 2
The values of the diversity of TBEV E protein antigenic variants at the level of virus subtypes and population.

Virus subtype	The number of sequences in GenBank database	The number of antigenic variants	The ratio n/N − the variety of antigenic variants	The maximum distance of physical and chemical characteristics	The average distance of physical and chemical characteristics	The coefficient of variation for physical and chemical distances
Far-Eastern	98	42	0.429	2.751	0.657	73.398
Siberian	31	15	0.484	1.611	0.780	48.830
European	112	46	0.411	2.286	0.653	69.094
In total	241	103	0.427	2.922	0.927	46.366

Table 3
Results of analysis significant differences variability of antigenic determinants of protein variants E TBEV at the level subtypes and virus population.

Virus subtype	Far-Eastern	Siberian	European	The total for TBEV
Far-Eastern				
Siberian	P_v = 0.78 (X^2 = 0.07) +			
European	P_v = 0.68 (X^2 = 0.16) +	P_v = 0.66 (X^2 = 0.18) +		
The total for TBEV	P_v = 1.00 (X^2 = 0.00) +	P_v = 0.80 (X^2 = 0.06) +	P_v = 0.84 (X^2 = 0.03) +	

Note: The table is a matrix with the results of the pairwise testing variability differences between selected groups of strains virus. In the cells matrix shows the value probability P_v, acceptance hypothesis H0 about absence reliable values between indexes diversity and criterion values X2, on based which is calculated P_v. Sign + at cell points out that the null hypothesis (absence of reliable differences) It was accepted.

Vaccines based on the TBEV Sofjin strain also prevents against Japanese encephalitis virus (Lai and Monath, 2003). The aforementioned vaccines were shown to be highly immunogenic (Wagner et al., 2004b; Fritz et al., 2012; Amicizia et al., 2013; Domnich et al., 2014). In Austria, after the introduction of a mass vaccination against TBEV, the incidence of TBE decreased by 95% (Romanenko et al., 2007; Heinz et al., 2008; Kunze and Böhm, 2015). In other TBEV endemic countries the vaccination is mainly used to protect specific groups, in case of an occupational risk, for school children and others (Zavadska et al., 2013).

Despite the success of TBEV vaccination, there are still some reports of TBEV infection cases among vaccinated individuals (Romanenko et al., 2007; Pogodina et al., 2009; Konkova-Rejdman and Zlobin, 2012; Subbotina et al., 2014; Pogodina et al., 2015). TBEV immunized individuals may accumulate amino acid substitutions by some TBEV strains in the course of microevolution in natural foci that prevent binding of the main TBEV antigen (the E protein) with the antibodies produced in response to the vaccination. Amino acid substitutions in the protein envelope for some hepatitis B viral strains have also rendered the vaccine ineffective for immunized individuals (Heijtink et al., 2002). A bioinformatics study on various hepatitis B virus variants showed the presence of large number of amino acids substitutions that cause significant physicochemical changes at specific positions of the antigenic protein (Stolbikov et al., 2014). A series of studies were also devoted to the physicochemical properties of the TBEV envelope E protein (Holzmann et al., 1990; Heinz et al., 1991; Rey et al., 1995).

Experiments showed that some monoclonal antibodies produced against the E protein for a particular TBEV strain have a lower affinity for E proteins from other TBEV strains (Guirakhoo et al., 1989; Rey et al., 1995). Earlier X-ray crystallography methods provided detailed tertiary and quaternary structures of the TBEV envelope E protein (Rey et al., 1995). The tertiary structure of the TBEV E protein provided a spatial location of the amino acid substitutions that affect the affinity to antibodies. These substitutions face the outer surface of the E protein that interacts with the antibodies (Rey et al., 1995). These observations indicate that the vaccine based on a particular TBEV strain may not provide reliable cross-protection against other viral strains.

Currently, international databases are rapidly supplemented with determined sequences of viral genomes and proteins. For example, GenBank database (http://www.ncbi.nlm.nih.gov/genbank/) contains a large number of determined amino acid sequences of TBEV E protein. These sequences can be statistically analyzed for amino acid diversity and variability to identify residue-specific, physicochemical properties at various positions of TBEV E protein. Therefore, the aim of this study was to search for sequence variants of the TBEV E protein in GenBank that are associated with the antigenic properties of the virus. A comparative analysis on the variability of vaccine-based or natural TBEV strains was then determined by statistical methods.

2. Materials and methods

The TBEV Sofjin strain (496 amino acid residues) was used for a protein BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to search for homologous TBEV E protein sequences in GenBank (http://www.ncbi.nlm.nih.gov/genbank/) non-redundant Database. For homologous search, the penalty for gap opening was set to 8 and the penalty for gap continuation was set to 2. The maximum number of queries produced as a result of the search was set to 1000. This combination of parameter values allows to identify sequences with both high and low similarity with the original sequence of the Sofjin strain.

All TBEV E protein sequences were visualized with the SeaView software (Gouy et al., 2010). Subsequent statistical analysis of antigenic determinants and the complete amino acid sequences of TBEV E protein were performed with R programming language designed for statistical data and its supplement package, bios2mds, for the polypeptide sequences processing (Pele et al., 2012).

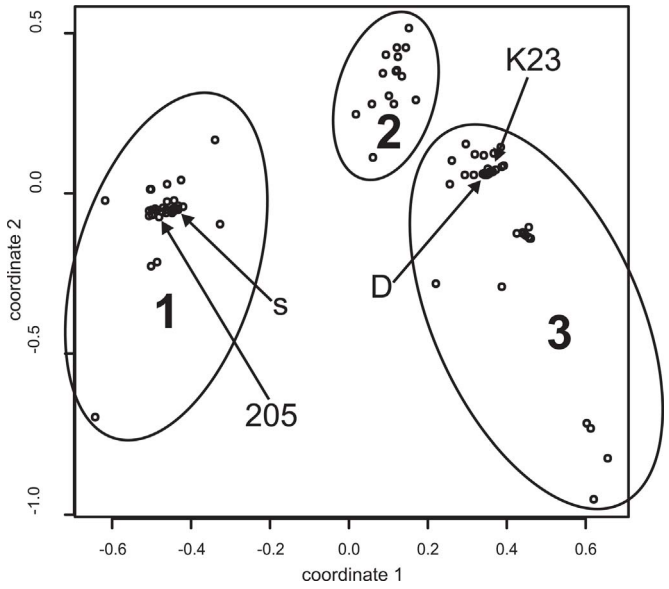

Fig. 4. The distribution of amino acid sequences of different variants of TBEV E protein antigenic determinants in two-dimensional space position constructed on the basis of multidimensional scaling involving physical and chemical characteristics of amino acid residues involved in the interaction of the antigen-antibody complex. The points cloud belonging to the antigenic determinants of Far-Eastern subtype is marked with number 1 and the corresponding ellipse; the points clouds of Siberian and European subtypes are marked with numbers 2 and 3 and corresponding ellipses, respectively. The letter S states the position of antigenic determinant of TBEV Sofjin strain; letter D − the position of epitope of Neudoerfl strain; symbols K23 and 205 define the positions of K23 and 205 strains epitopes, respectively.

Maximum likelihood phylogenetic analysis was performed on the selected amino acid sequences using the SeaView program. The reliability of the phylogenetic tree topology was tested using bootstrap analysis. The 1000 repeats were generated for calculations of bootstrap support. The model for the accumulation of amino acid substitutions for phylogeny was selected using the ProtTest program (Abascal et al., 2005). The protein sequence of the Omsk hemorrhagic fever virus was used as the outgroup as the closest relative of the TBE virus. The most optimal model of amino acid substitutions was chosen on the basis of Bayesian information criterion (BIC). The results of phylogenetic clustering were used to determine the subtype of strains from which the amino acid sequences were isolated.

3 Based on the data on TBEV E protein tertiary structure (Rey et al., 1995) from the data set containing the full-length amino acid sequences, the amount of E protein fragments sequences forming the outer layer of virion envelope was formed with SeaView software. Thus, for each variant of amino acid sequence of E protein the sequences consisted of the set of E protein fragments exposed at the virus surface − the antigenic determinants, − were defined. The scheme for isolating to the antigenic determinants is shown in Fig. 1.

4 A comparison analysis on the diversity of E protein antigenic variants from different TBEV subtypes and virus population was calculated by the ratio of antigenic determinants to the total number of E protein sequences. The coefficient of antigenic variation was determined according to Hudson (Hudson, 2000). Specific virus subtypes with coefficient value approaching 1 suggests an increase in variation in the composition of E protein antigenic variants. Using the statistical criterion based on chi-square distribution, the significance differences were identified between the variability coefficients for the antigenic determinants of virus subtypes and for the entire virus population (Newcombe, 1998).

5 The numerical values of the physicochemical properties of amino acid residues in the polypeptide sequences were obtained from APDbase database (http://www.roskamps.com/bioinfo/APDbase/) (Mathura and Kolippakkam, 2005). To eliminate the effects of unequal contribution of different physicochemical properties in the statistical analysis, the selected numerical physicochemical properties of each feature were recalculated as percentage change from a maximum to a minimum value (the maximum numeric value of feature was taken as 100%, the minimum value − as 0%), i.e., were normalized to 100%.

The amino acid sequence of the antigenic determinant of TBEV E protein can be represented as a vector of dimension N − the length of the sequence. Each coordinate of the vector can take one of 20 literal values (single-letter encoding of amino acid). If n normalized physicochemical properties will match each amino acid, the vector (amino acid sequence of the antigenic determinant of TBEV E protein) will be transformed into a numeric vector of dimension $Np = N \times n$. The numeric vector of dimension Np will characterize the physicochemical properties of the antigenic determinants of the TBEV E protein. By comparing the antigenic determinants of amino acids at particular positions, an array of numerical vectors is obtained that characterize the TBEV E protein antigenic determinants. Statistical analysis pipeline is shown in Fig. 2.

A pairwise distance matrix using the Euclidean distance measure was built based on numerical vectors describing the physicochemical properties of the TBEV E protein antigenic determinants. The resulting pairwise distance matrix was used for multivariate statistics − multidimensional scaling (Gower, 1966). The same distances matrix was used to evaluate the properties of E protein antigenic determinants for TBEV strains currently used for vaccines production, namely Neudoerfl, K23, 205 and Sofjin. The parts of the matrix used contained the distances of antigenic determinant of a definite strain from all other antigenic determinants of the viruses. Based on this information, the average distance of the strain determinant against all other antigenic determinants was calculated and the distribution of distances presented as boxplots. The reliability of the differences between the mean values for the distances spectra of vaccine strains from other strains was determined using a nonparametric version of the single-factor analysis of variance − the Kruskal-Wallis test (Kruskal, Wallis, 1952).

Strains that had the lowest average Euclidean distance from other antigenic determinants calculated on the basis of the physicochemical properties of the amino acid residues were used to isolate the athenic determinants most preferred for the creation of a polyvalent vaccine. It can be assumed that upon introduction into the body viral particles with such an amino acid composition of the surface layer of E protein will stimulate the formation of antibodies that are potentially more likely to bind to other antigenic determinants of different strains of

Table 4
The characteristics of TBEV strains and vaccines based on them.

Vaccine name	The company and the country of origin	Strain	Representation of the antigen in GenBank	The average distance to other antigenic determinants on physical and chemical characteristics
FSME-IMMUN Inject	Baxter, Austria	*Neudoerfl*	2	0.747
FSME-IMMUN Junior	Baxter, Austria	*Neudoerfl*	2	0.747
Encepur-Adults	Novartis, Germany	*K23*	1	0.995
Encepur-Children	Novartis, Germany	*K23*	1	0.995
EnceVir	Microgen, Russia	*205*	9	0.926
TBE-Moscow	Chumakov Institute, Russia	*Sofjin*	37	0.745

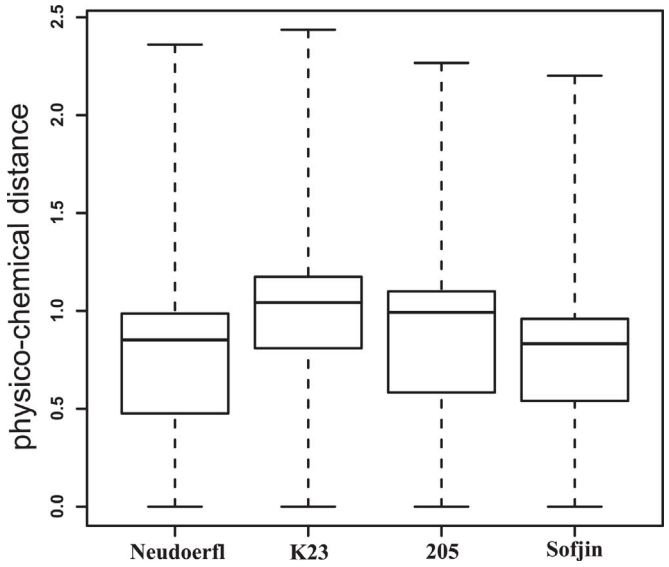

Fig. 5. Boxplots visualization of the distribution of pairwise differences of Neudoerfl, K23, 205 and Sofjin strains antigenic determinants from other E protein antigenic determinants constructed on the basis of differences in physical and chemical characteristics of amino acid sequences involved in the antigen-antibody interaction.

viruses present in the natural population.

3. Results and discussion

The BLAST search identified 350 TBEV E protein sequences. The spectrum of matches found in the database contained sequences with an identity score of 100% to 41%. The least identity with the Sofjin strain for a sequence recorded in the database under TBEV identifier was 95%. This means that the entire variety of TBEV E protein variants contained in the database was presented in this search. The search was conducted according to the state of the GenBank database for October 2016. Sequences less than 496 residues (the length of the TBEV Sofjin strain) or contained unknown residues (X) were deleted from the sequence dataset − 241 sequences remained.

On the basis of testing the data set in the ProtTest program, the JTT model (Jones et al., 1992) with gamma correction (JTT + G) was the most optimal model. The phylogenetic tree reconstructed by the maximum likelihood method using the JTT + G model is shown on Fig. 3. Three clusters with sufficiently high bootstrap supports values (73%–98%%) can be distinguished on the phylogenetic tree. Each of the clusters contained sequences for which only one corresponding subtype in the GenBank database was specified and sequences for which the subtitle was not specified. The first cluster included sequences identified as the Far-Eastern subtype, the second cluster contained the Siberian subtype, and the third formed the European subtype. According to these clusters, the unknown sequences can be appropriately assigned a subtype. The branching topology and subtype distribution of the reconstructed phylogenetic tree corresponded to the tree obtained on the basis of the nucleotide sequences of TBEV E gene in (Ecker et al., 1999a,b). On the phylogenetic tree (Fig. 3), the vaccine strains Neudoerfl and K23 clustered together with other strains of the western genotype, strains 205 and Sofjin clustered together with strains of the Far Eastern TBEV genotype. Of the 241 TBEV E protein sequences, 98 belong to the Far-Eastern subtype, 31 to the Siberian subtype, and 112 to the European subtype. All 241 amino acid sequences of TBEV E protein with GenBank accession number, which were used in the analysis, are presented in supplementary file in fasta format.

Protein fragments that face outwardly from the virion envelope surface were identified (Table 1) based on the tertiary structure of the complete TBEV Sofjin strain E protein (Rey et al., 1995). As a result, an antigenic determinant-based sequence (224 residues) was obtained from the TBEV Sofjin strain E protein. This antigenic determinant was used as the reference sequence for a multiple sequence alignment to detect similar determinants in the TBEV sequence dataset. Further computational analysis showed that the complete set of TBEV E protein sequences contained 103 antigenic determinant variants, differing in at least one amino acid substitution. The Far-Eastern subtype included 42 antigenic determinant variants, the Siberian subtype had 31 variants, and the European subtype had 46 variants (Table 2). The coefficient of antigenic variation was calculated for the total TBEV population (full sequence dataset) and for each virus subtype. The coefficient values are represented in Table 2. The measured values for these coefficients ranged from 0.484 to 0.411. The Far-Eastern, Siberian and European subtypes were not significantly different (Table 3), therefore, all three TBEV E protein subtypes have the same variability of antigenic determinants.

Antibody-antigen binding occurs by space-conformational interaction (Boyden, 1966). Theoretical and experimental studies show that specific physicochemical properties of amino acids of the antigen play key roles in its reaction with an antibody (Absolom and Van Oss, 1985; Mian et al., 1991) These properties include, a) the average amino acid volume of the protein; b) the number of hydrogen bond donors of residues; c) the charge of the residue side chain; and, g) the residue dipole moment. The sequence residues differing in volume on the surface of the protein antigen defines the spatial structure recognized by antibody. The hydrogen bonding, charge and dipole interaction between the antigen and antibody bind them into a single structure. These four physicochemical properties (n = 4) were selected for this study. Each antigenic determinant of TBEV E protein (as described in Materials and Methods section) with the length N = 224 residues was characterized with the numerical vector with the dimension $N_p = 896$ ($N_p = N \times n$). Using these vectors, a pairwise distance matrix was built for the TBEV E protein antigenic determinants. The calculated distances are based on the physicochemical properties of the antigenic determinants. For each TBEV subtype, and TBEV in general, the maximum and average distances on the physicochemical properties and the variation coefficients for the studied distances were calculated (Table 2). The Far-Eastern and European subtypes have similar maximum and average distances values. The Siberian subtype differs since it has a lower maximum distance and a higher average distance values. For TBEV in general, the maximum and the average distances exceeded the corresponding values of each subtype (Table 2).

The estimated values of the coefficients of distance variation based on the numeric values of the physicochemical properties of the TBEV antigenic determinants for Far-Eastern (73.398) and European (69.094) subtypes differed slightly from each other (Table 2). The Siberian subtype had the distance variations coefficient of 48,830. For TBEV in general, the coefficient of distance variation was 46,366. The low value of the distances variation coefficient for Siberian subtype is possibly connected with the small representation of the Siberian subtype in GenBank database − only 31 complete E protein sequences and 15 antigenic determinant sequences (Table 2). An increase in sequences analyzed may equalize the coefficient values of distance variation for different subtypes. However, for the current number of sequences the coefficient of variability of antigenic determinants for all subtypes, and TBEV in general, did not significantly differ. Additionally, amino acid variability depends on positive or purifying selection. In any case, to verify these assumptions it is necessary to increase the number of determined sequences of TBEV E protein of Siberian subtype.

7 Using the distance matrix calculated from the numeric values of the physicochemical properties, an analysis on the TBEV antigenic determinants was carried out by multidimensional scaling. Fig. 4 is two-dimensional scatter − result of the multidimensional scaling from space of physicochemical properties of amino acid residues. The distinguishable three clouds in Fig. 4 are antigenic determinants belonging to the E protein of the three viral subtypes (Far-Eastern, Siberian and European). The characteristic point spread was observed for each

Table 5
Variants of antigenic determinants, the most optimal for the creation of polyvalent vaccines.

Variant of the determinant	Amino acid sequence	Accession numbers from the GenBank database
a1_Sofjin Far-Eastern subtyp	SRCENRDFVTGTGCVTITAEGKPSMDVWLDSIYQENPAKTREYCLAKLSDTKVAARCPTMGPATLAEEHQSGTVCKRDQSDRGWGNHCGLFGKGSIV-TCVKASCEAKKKATGHVYIVYTVKVEPHTGDYVAANETHSGRKTASFTVAQTVILELDKTSEHLPTAWQVHRDPWKHEHIDGTKYHLKYTMCDPCRIPVRAVAHGNPTIEPGDNIIYVGELSHQWFQ	ACF33496.1, AFP25092.1, ACT32141.1, ACF33498.1, AFP25080.1, AFP25097.1, AFP25082.1, ACQ99330.1, AAN87009.2, ACF33495.1, AHC30237.1, AFP25096.1, AFP25094.1, AFP25081.1, AFP25079.1, ACF33494.1, AFV41132.1, AEP20480.1, AFP25085.1, AGC31653.1, AHC30239.1, ACJ38114.1, AFP25083.1, AII96827.1, P07720.3, ADT80553.1, ADE93003.1, ACJ38115.1, CAA30581.1, BAB40646.1, ABU62955.1, AFP25084.1, ADX07734.1, ABI31771.1, AIK20543.1, BAA84957.1, AFP25086.1
a43 European subtype	SRCENRDFVTGTGCVTITAEGKPSMDVWLDAIYQENPAKTREYCLAKLSDTKVAARCPTMGPATLAEEHQGGTVCKRDQSDRGWGNHCGLFGKGSIV-ACVKAACEAKKKATGHVYIVYTVKVEPHTGDYVAANETHSGRKTASFTVAQTVILELDKTVEHLPTAWQVHRDPWKHEHIEGTKYHLKYTMCDPCRIPVRAVAHGNPTIEPGDNIIYVGELSHQWFQ	AGP05331.1, AGO50950.1, ABD62793.2, ACL97686.1, ADQ00969.1, ADQ00968.1, AIL83863.1, AIL83860.1, ADM63091.1, AHM02467.1, AIL83862.1, ADQ00972.1, AIL83861.1, AHF27215.1, ADQ00970.1, AHF27216.1, AHF27217.1, AGO50925.1, AAC62088.1, ADO21535.1, AAP31247.1, AGO50944.1, AGO50946.1, AGO50942.1, AGO50924.1, AEH39549.1, ADO21538.1, AGO50936.1, AGF36551.1, ADO21576.1, ADO21515.1, AGO50941.1, AEH39550.1, ADO21565.1, ADO21514.1, AGO50951.1, AGF36549.1, ADO21566.1, AGO50928.1, CAA54068.1, CAA53041.1
(optimal for the whole variability of the virus subtyp) a92 Siberian subtype	SRCENRDFVTGTGCVTITAEGKPSMDVWLDSIYQENPAKTREYCLAKLSDTKVAARCPTMGPATLAEEHQSGTVCKRDQSDRGWGNHCGLFGKGSIV-TCVKAACEAKKKATGHVYIVYTVKVEPHTGDYVAANETHSGRKTASFTVAQTVILELDKTLEHLPTAWQVHRDPWKHEHIDGTKYHLKYTMCDPCRIPVRAVAHGNPTIEPGDNIIYVGELSHQWFQ	AIL33471.1, AFU65175.1, AAO43537.1, BAB40649.1, AEX15486.1, ADQ00973.1, ADJ96605.1, BAB40652.1, BAB40650.1

subtype indicating that the lowest variation was found in the Siberian subtype. This low variation is consistent with the low coefficient in the distance variation calculated from the numeric values of the physicochemical properties. It should be noted that the extreme distant points in each cloud are comparable with the distances between the closest points of different virus subtypes. This means either the E protein antigenic determinants of different TBEV subtypes have similar antigenic properties or the antigenic determinants of the same subtype significantly differ in antibody affinity. This fact points to the wide variability of the antigenic properties of TBEV E protein in different strains.

The TBEV E protein from vaccination commercial strains Neudoerfl, K23, 205 and Sofjin were among the amino acid variants found in the sequence dataset. The full characterization of these strains is given in Table 4. From the sequence dataset, two antigenic determinants were identical to those in the Neudoerfl strain, one was identical to the K23 strain, nine were similar to the 205 strain, and thirty-seven were identical to the Sofjin strain. As severe causing agents of TBE in the Far-East (Holzmann et al., 1992), both 205 and Sofjin strains are of great research interest. Thus, there is a broad representation of 205 and Sofjin antigenic determinants in the sequence dataset due to the large number of TBEV Far-Eastern subtype E sequences isolated from patients, as carried out by the Russian Federation.

Analysis on the physicochemical properties for Neudoerfl, K23, 205 and Sofjin strains demonstrated that the average distances from the antigenic determinants of these strains to other strains significantly differ from each other. According to the results of the Kruskal-Wallis test the observed differences between the mean distances in the samples are reliable (the value of the test statistic is $W = 81.637$, P_value = 2.2e–16 < 0.05). The TBEV Sofjin and Neudoerfl strain had the lowest average distance (~ 0.75), and the TBEV 205 (0.926) and K23 (0.995) strains had the highest average distances. The average distribution distances between antigenic determinants of Neudoerfl, K23, 205 and Sofjin strains to other strains is shown in Fig. 5. The smaller range of antigenic determinant distances in the Sofjin and Neudoerfl strains compared with 205 and K23 strains indicates that antibodies developed during vaccination will bind with antigens from other virus variants differing in the substitutions of E protein antigenic determinants. For each Neudoerfl, K23, 205 and Sofjin TBEV strain, approximately 20% of the sequence dataset had few substitutions (from 2 to 3) that results in a change of charge, polar characteristics, the ability to form the hydrogen bonds, or a significant change of amino acid volume. Such substitutions significantly decrease the affinity between antigens and antibodies, thereby reducing the effectiveness of a vaccine (Absolom and Van Oss, 1985; Mian et al., 1991).

It should be noted that the vaccine production against TBEV are designed within a finite regional connection for that particular prevalent strain (Klockmann et al., 1991; Morozova et al., 2012). Our results demonstrate that the industrial TBEV strains, comprising into three commercial vaccines, do not fully cover the entire spectrum of antigenic determinants of TBEV E protein within its regional connection. We not only examine antigenic epitopes of linear structures at the virion surface, but to the epitopes of arbitrary cross configuration. Since the investigated antigen is the unified polymer structure of the protein (E protein of TBEV), all differences between the monomers (amino acids in the linear structure between the antigenic determinants of TBEV E protein) will have approximately the same position in space on the virus surface.

Optimal variants of antigenic determinants for the creation of a polyvalent vaccine were determined using the pairwise distances matrix calculated from the Euclidean metric on the basis of the properties of amino acid residues. Such antigenic determinants were determined for each genotype of the virus. Variants of the amino acid sequence for the isolated determinants are presented in Table 5. Each antigenic determinant is represented by several sequences from the GenBank database. Sequences with identical antigenic determinants did not differ in the amino acid composition of the surface layer of E protein, but had amino acid substitutions in other parts of the protein. For the Far Eastern genotype, the antigenic determinant of the Sofjin strain was optimal. For all sets of virus sequences of all genotypes, the smallest average distance from all other determinants had sequence that, by the same characteristics, is optimal for the western genotype of the virus. Strains of the virus with this variant of the antigenic determinant may be optimal for the creation of a universal vaccine.

Currently as a primary target for the vaccination, the TBEV E protein has wide variability in physicochemical properties responsible for antigen-antibody binding. This hypothesis has been supported by experimental studies to find epitopes of antigenic sites in the amino acid sequences of TBEV E protein presented in (Kuivanen et al., 2014). We report that they have first defined the linear epitopes in TBEV E protein and in non-structural NS5 protein; only two (in E protein) of 11 identified epitopes were shown to be potential antigenic determinants for TBEV diagnostics and can serologically differentiate the *Flavivirus* infections apart. In (Kiermayr et al., 2009) binding of various variants of monoclonal antibodies to mutant strains of TBEV E protein was studied. As a result, it was shown that the emerging mutations can neutralize the process of binding of antibodies to the corresponding epitopes on the surface of the virus. Epitopes for binding of antibodies in this case can be in different places of the surface part of the TBEV E protein. Therefore, our theoretical calculations and experimental results reported in (Kiermayr et al., 2009; Kuivanen S. et.al, 2014) show the need for more careful approach to the assessment of the effectiveness of both existing and newly developed vaccines against not only for TBEV, but also against other *Flavivirus* pathogens.

The theoretical study carried out is a preliminary attempt to propose a calculation method for the study and prediction of various variants of antigens for the creation of universal vaccines. It is necessary to conduct an experimental study to confirm the findings of the work. Such a study may consist in analyzing the correlation relationship between distances calculated by the physicochemical properties of amino acid residues and the intensity of the antigen-antibody interaction estimated under experimental conditions.

The authors express their gratitude to James Valdés from Biology Center, ASCR, v.v.i. Institute of Parasitology, Czech Academy of Sciences for advice and useful comments in the course of the work.

Fundings

The work was supported by the Program of Fundamental Research of the Presidium of Russian Academy of Sciences No. 1 on the strategic directions of science development in 2014. “Fundamental problems of mathematical modeling” and Russian Scientific Foundation No. 14-15-00615 (analysis of TBEV strains of European subtype).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virusres.2017.06.006.

References

Absolom, D.R., Van Oss, C.J., 1985. The nature of the antigen-antibody bond and the factors affecting its association and dissociation. CRC Crit. Rev. Immunol. 6 (1), 1–46.
Amicizia, D., Domnich, A., Panatto, D., Lai, P.L., Cristina, M.L., Avio, U., Gasparini, R., 2013. Epidemiology of tick-borne encephalitis (TBE) in Europe and its prevention by available vaccines. Hum. Vaccin. Immunother. 9 (5), 1163–1171.
Bock, H.L., Klockmann, U., Jüngst, C., Schindel-Künzel, F., Theobald, K., Zerban, R., 1990. A new vaccine against tick-borne encephalitis: initial trial in man including a dose-response study. Vaccine 8 (1), 22–24.
Bogovic, P., Strle, F., 2015. Tick-borne encephalitis: a review of epidemiology, clinical characteristics, and management. World J. Clin. Cases 16 (5), 430–441 (3).
Boyden, S.V., 1966. Natural antibodies and the immune response. Adv. Immunol. 5, 1–28.
Charrel, R.N., Attoui, H., Butenko, A.M., Clegg, J.C., Deubel, V., Frolova, T.V., et al., 2004. Tick-borne virus diseases of human interest in Europe. Clin. Microbiol. Infect. 10 (12), 1040–1055.

Demina, T.V., Dzhioev, Y.P., Verkhozina, M.M., Kozlova, I.V., Tkachev, S.E., Plyusnin, A., et al., 2010. Genotyping and characterization of the geographical distribution of tick-borne encephalitis virus variants with a set of molecular probes. J. Med. Virol. 82 (6), 965–976.
Demina, T.V., Dzhioev, Y.P., Kozlova, I.V., Verkhozina, M.M., Tkachev, S.E., Doroshchenko, E.K., al, et., 2012. Genotypes 4 and 5 are borne encephalitis virus: the structural features of the genomes of the possible scenario of their formation. Voprosy Virusol. 4, 13–19 (In Russian).
Domnich, A., Panatto, D., Arbuzova, E.K., Signori, A., Avio, U., Gasparini, R., Amicizia, D., 2014. Immunogenicity against Far Eastern and Siberian subtypes of tick-borne encephalitis (TBE) virus elicited by the currently available vaccines based on the European subtype: systematic review and meta-analysis. Hum. Vaccin. Immunother. 10 (10), 2819–2833.
Dumpis, U., Crook, D., Oksi, J., 1999. Tick-borne encephalitis. Clin. Infect. Dis. 28 (4), 882–890.
Ecker, M., Allison, S.L., Meixner, T., Heinz, F.X., 1999a. Sequence analysis and genetic classification of tick-borne encephalitis viruses from Europe and Asia. J. Gen. Virol. 80 (1), 179–185.
Ecker, M., Allison, S.L., Meixner, T., Heinz, F.X., 1999b. Sequence analysis and genetic classification of tick-borne encephalitis viruses from Europe and Asia. J. Gen. Virol. 80 (1), 179–185.
Fritz, R., Orlinger, K.K., Hofmeister, Y., Janecki, K., Traweger, A., Perez-Burgos, L., al, et., 2012. Quantitative comparison of the cross-protection induced by tick-borne encephalitis virus vaccines based on European and Far Eastern virus subtypes. Vaccine 30 (6), 1165–1169.
Goodman, J.L., Dennis, D.T., Sonenshine, D.E., 2005. Tick-Borne Encephalitis. Tick-Borne Diseases of Humans. ASM Press, Washington DC.
Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27 (2), 221–224.
Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325–328.
Guirakhoo, F., Heinz, F.X., Kunz, C., 1989. Epitope model of tick-borne encephalitis virus envelope glycoprotein E: analysis of structural properties, role of carbohydrate side chain, and conformational changes occurring at acidic pH. Virology 169 (1), 90–99.
Heijtink, R.A., Bergen, P., Melber, K., 2002. Hepatitis B surface antigen (HBsAg) derived from yeast cells (Hansenula polymorpha) used to establish an influence of antigenic subtype (adw2, adr, ayw3) in measuring the immune response after vaccination. Vaccine 20, 2191–2196.
Heinz, F.X., Mandl, C.W., Holzmann, H., Kunz, C.H., Harris, B.A., Rey, F.E., Harrison, S.C., 1991. The *Flavivirus* envelope protein E: isolation of a soluble form from tick-borne encephalitis virus and its crystallization. J. Virol. 65 (10), 5579–5583.
Heinz, F., Holtzmann, H., Essl, A., Kundt, M., 2008. Analysis of the efficiency of tick-borne encephalitis vaccination in the population in the natural foci of Austria. Voprosy Virusol. 53 (2), 19–27.
Holzmann, H., Heinz, F.X., Mandl, C.W., Guirakhoo, F., Kunz, C., 1990. A single amino acid substitution in envelope protein E of tick-borne encephalitis virus leads to attenuation in the mouse model. J. Virol. 64 (10), 5156–5159.
Holzmann, H., Vorobyova, M.S., Ladyzhenskaya, I.P., Ferenczi, E., Kundi, M., Kunz, C.H., Heinz, F.X., 1992. Molecular epidemiology of tick-borne encephalitis virus: cross-protection between European and Far Eastern subtypes. Vaccine 10 (5), 345–349.
Hudson, R.R., 2000. A new statistic for detecting genetic differentiation. Genetics 155 (4), 2011–2014.
Klockmann, U., Křivanec, K., Stephenson, J.R., Hilfenhaus, J., 1991. Protection against European isolates of tick-borne encephalitis virus after vaccination with a new tick-borne encephalitis vaccine. Vaccine 9 (3), 210–212.
Konkova-Rejdman, A.B., Zlobin, V.I., 2012. Specific and nonspecific prevention of tick-borne encephalitis and Ixodes tick borreliosis in the southern Urals. Sibirskij Medicinskij Zurnal (Irkutsk) 4, 71–74 (In Russian).
Kozlova, I.V., Verkhozina, M.M., Demina, T.V., Dzhioev, Y.P., Tkachev, S.E., Karan, L.S., et al., 2013. Genetic and Biological Properties of Original TBEV Strains Group Circulating in Eastern Siberia. «Encephalitis» Publisher: InTech, Published: January 09. pp. 95–112.
Kuivanen, S., Hepojoki, J., Vene, S., Vaheri, A., Vapalahti, O., 2014. Identification of linear human B − cell epitopes of tick-borne encephalitis virus. Virol. J. 11 (1), 115.
Kunze, U., Böhm, G., 2015. Tick-borne encephalitis (TBE) and TBE-vaccination in Austria. Wien Med Wochenschr. Epub ahead of print.
Lai, C.J., Monath, T.P., 2003. Chimeric flaviviruses: novel vaccines against dengue fever, tick-borne encephalitis, and Japanese encephalitis. Adv. Virus Res. 61, 469–509.
Leonova, G.N., Ternovoi, V.A., Pavlenko, E.V., Maistrovskaya, O.S., Protopopova, E.V., Loktev, V.B., 2007. Evaluation of vaccine Encepur Adult for induction of human neutralizing antibodies against recent Far Eastern subtype strains of tick-borne encephalitis virus. Vaccine 25 (5), 895–901.
Mandl, C.W., Guirakhoo, F., Holzmann, H., Heinz, F.X., Kunz, C., 1989. Antigenic structure of the *Flavivirus* envelope protein E at the molecular level, using tick-borne encephalitis virus as a model. J. Virol. 63 (2), 564–571.
Mathura, V.S., Kolippakkam, D., 2005. APDbase: amino acid physicochemical properties database. Bioinformation 1 (1), 2–4.
Mian, I.S., Bradwell, A.R., Olson, A.J., 1991. Structure, function and properties of antibody binding sites. J. Mol. Biol. 217 (1), 133–151.
Morozova, O.V., Grishechkin, A.E., Bakhvalova, V.N., Isaeva, E.I., Podcherniaeva, RIa., 2012. Changes in the reproduction of tick-borne encephalitis virus in cell cultures. Voprosy Virusol. 57 (2), 40–43 (In Russian).
Morozova, O.V., Bakhvalova, V.N., Potapova, O.F., Grishechkin, A.E., Isaeva, E.I., Aldarov, K.V., al, et., 2014. Evaluation of immune response and protective effect of four vaccines against the tick-borne encephalitis virus. Vaccine 32 (25), 3101–3106.
Newcombe, R.G., 1998. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat. Med. 17, 873–890.
Orlinger, K.K., Hofmeister, Y., Fritz, R., Holzer, G.W., Falkner, F.G., Unger, B., al, et., 2011. A tick-borne encephalitis virus vaccine based on the European prototype strain induces broadly reactive cross-neutralizing antibodies in humans. J. Infect. Dis. 203 (11), 1556–1564.
Panov, A.G., 1956. Tick-borne encephalitis. Medgiz, Leningrad In Russian.
Pele, J., Jean-Michel Bécu, J., Abdi, H., Chabbert, M., 2012. Bios2mds: an R package for comparing orthologous protein families by metric multidimensional scaling. BMC Bionf. 13 (1), 33.
Pogodina, V.V., Levin, L.S., Karan, L.S., Kolyasnikova, N.M., Bezrukov, E.G., Druzhinin, T.A., al, et., 2009. Encephalitis deaths caused by Siberian subtype of the pathogen in European Russia and the Urals. Medicinskaya Virusol. 14, 121–122. In Russian.
Pogodina, V.V., Luchinina, S.V., Stepanova, O.N., Stenko, E.A., Gorfinkel, A.N., Karmysheva, V.J., et al., 2015. Unusual case of fatal outcomes tick-borne encephalitis in patients vaccinated with different genotypes (Chelyabinsk Region). Epidemiologya I Infekcionia Zabolevania 1, 56–64 (In Russian).
Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Aresolution. Nature 375, 291–298.
Romanenko, V.V., Kovtun, O.P., Shelkova, E.S., 2007. Clinical and epidemiological features of tick-borne encephalitis in the Sverdlovsk region in the period of mass immunization. Neurological Vestnik 1, 75–79 (In Russian).
Stolbikov, A.S., Bukin, Y.S., Dzhioev, Y.P., Zlobin, V.I., 2014. Bioinformatic search options α-S protein epitope of hepatitis B virus to create a multivalent vaccine. Epidemiol. Vacc. Prevent. 3, 94–104 (In Russian).
Subbotina, A.V., Semenov, V.A., Smirnov, V.D., Shcherbinina, M.S., Pogodina, V.V., 2014. A case of chronic tick-borne encephalitis in vaccinated person. Epidemiol. Vacc. Prevent. 33, 113–117 (In Russian).
Vorobyova, M.S., Rasschepkina, M.N., 1992. The study of the relationship antigenic characteristics of the virus encephalitis level protective activity of inactivated vaccines culture. Zhurnal Mikrobiol. 11–12, 35–37 (In Russian).
Wagner, R.P., Kundi, M., Zent, O., Banzhoff, A., Jaehnig, P., Stemberger, R., al, et., 2004a. Immunogenicity and safety of a booster vaccination against tick-borne encephalitis more than 3 years following the last immunisation. Vaccine 23 (4), 427–434.
Wagner, R.P., Kundi, M., Zent, O., Dvorak, G., Jaehnig, P., Holzmann, H., et al., 2004b. Persistence of protective immunity following vaccination against tick-borne encephalitis-longer than expected? Vaccine 22, 2743–2749.
Zavadska, D., Anca, I., André, F., Bakir, M., Chlibek, R., Cižman, M., al, et., 2013. Recommendations for tick-borne encephalitis vaccination from the central european vaccination awareness group (CEVAG). Hum. Vaccin. Immunother. 9 (2), 362–374.
Zilber, L.A., 1945. Epidemic Encephalitis. Medgiz, Moscow (In Russian).
Zlobin, V.I., Mamaev, L.V., Dzhioev, Y.P., Kozlova, I.V., 1996. Genetic types of viral encephalitis. Z. Infect. Pathol. 4, 13–17 (In Russian) ces involved in the antigen-antibody interaction.
Zlobin, V.I., Belikov, S.I., Dzhioev, Y.P., Demina, T.V., Kozlova, I.V., Verkhozina, M.M., et al., 2001a. The new concept of the natural genetic variability of the tick-borne encephalitis virus. Tihookeansky Medicinsky Z. Vladivostok 2 (7), 75–78 (In Russian).
Zlobin, V.I., Gaso, M.H., Belikov, S.I., Dzhioev, Y.P., Demina, T.V., Verkhozina, M.M., al, et., 2001b. Molecular probes for genetic typing of tick-borne encephalitis virus. Voprosy Virusol. 4, 43–47. In Russian.
Zlobin, V.I., Lvov, D.K., Vorobyov, M.S., Shashin, N.I., 2009. Epidemiology and Prevention of Tick-Borne Encephalitis and Other Tick-Borne Infections. Medgiz, Moscow (In Russian).