Romanian Biotechnological Letters *Vol. 14, No.4, 2009, pp. 4589-4596*
Copyright © 2008 Bucharest University Printed in Romania. All rights reserved
Romanian Society of Biological Sciences

ORIGINAL PAPER

Predicting antigenic sites of hemagglutinin-neuraminidase glycoprotein – Newcastle disease virus

Received for publication, January 13, 2008
Accepted, July 5, 2008

DANIELA BOTUŞ
National Society "Pasteur Institute", Calea Giulesti Str., 333, Bucharest, 060269
e-mail: dana_botus@yahoo.com

Abstract

There was carried out a virtual antigenic study of hemagglutinin-neuraminidase (HN) glycoprotein belonging to Newcastle disease virus, using three prediction programs based on algorithms which combine physical-chemical properties as: hydropathy (hydrophobicity/hydrophily), mobility/flexibility, proteic surfaces accessibility, secondary structure (α or β type).

The sequential prediction by ProtScale program suggested the presence of 8 hydrophobic antigenic sites, and the hydrophylic profile indicated 9 antigenic sites. There were identified 12 antigenic sites related to β-sheet structure frequency and 14 sites for β-turn structures. Finally, the comparative analysis and overlapping these sequences suggested the presence of 22 antigenic determinants, which contain both hydrophobic and hydrophylic aminoacids; the predominant structures are of β type. Even if they have antigenic features, some of these sequences are not accessible, suggesting that, because of the folding process of polypeptidic chain, they are located in hidden areas inside the molecule, forming the proteic core.

There was realized an antigenic virtual profile of HN glycoprotein by means of two other programs: Antigenic Prediction – based on hydrophobic properties of proteins, and CTLPred, based on Artificial Neuronal Network – ANN. There are satisfactory correlations between the results obtained by the three prediction methods, with a difference of ± 5-8 aminoacids.

Keywords: antigenic sites, hemagglutinin-neuraminidase, Newcastle disease virus

Introduction

Identification of protein antigenic sites is of vital importance for the development of synthetic vaccines, immunodiagnostic tests and antibody production. For precise identification of antigenic determinants it is necessary to obtain derived proteins and peptidic fragments well characterized, starting from the original protein followed by their testing in order to identify the immunological activity. But such studies are extremely laborious, so that the efforts of researchers have turned to develop more rapid prediction methods, based only on the amino acids sequence of the protein in question.

There are several methods that predict the epitopes position, based on some physico-chemical properties of proteins such as hydrophilicity, mobility / flexibility, surfaces accessibility, helicity, structure etc. There were reported various algorithms which use combinations of these properties, but with an accuracy of 40 - 60%. Kolaskar and Tangaonkar developed a combination scale using hydrophilicity / hydrophobicity, flexibility and surfaces accessibility of protein, known as the *Antigenic Propensity* scale which is considered a gold standard in epitope prediction, with an accuracy of 75% [1].

The hydrophobicity is one of the main forces contributing to antigen – antibody interactions. The hydrophobic interactions generally increase the enthalpy, by removing the water molecules surrounding involved molecules, thus leading to a higher stability of the formed complexes. It also appears a higher energy, as unfavourable interactions like polar groups - nonpolar groups are replaced by others more favourable. Thus, the recognition sites of proteins and ligands are often more hydrophobic than the rest of protein molecule [2].

It is known that the secondary structure of proteins may be of two types: *α-helix* and *β-sheet*. In addition, there are intermediate structures derived from the two ones, such as *β-turn* or *coil*. Regarding the tertiary structure of proteins, it appears as a result of the folding of polypeptidic chains, when hydrophobic areas tend to inner side of the molecule forming the so-called "core", especially for globular proteins. The β-sheet and β-turn structures have an important role for this phenomenon. The β-type structures typically appear on the surface molecule, being involved in the intermolecular interactions. In this way, an antigen - antibody interaction can take place with greater probability at β-turn level than in the compact, rigid structure (α-helix) [3,4]. Moreover, it has been shown that, for the proteic antigens, the antibody binding site is an extended flat area that matches to β-type structures, unlike DNA antigens for which antibodies binding site is a cavity-like [5].

Most viruses, including the Newcastle disease virus (NDV), contain on their surface antigenic glycoproteins, but not the entire proteic structure participates to antigen – antibody interactions [6]. Moreover, some glycoproteins do not take part to these interactions, even if they are able to induce specific antibody synthesis, because of their structure which may be hidden by the presence of other proteins. Each glycoprotein can present several areas that the antibodies may interact so that the overall structure of viruses has antigenic polydeterminants [7]. Not all these antigenic determinants bind the specific antibodies with the same strength, therefore it became imperative to identify the main epitopes of these proteins. For this aim, we proposed a virtual antigenic study, allowing identifying the antigenic determinants of a major glycoprotein of Newcastle disease virus.

Materials and Methods

For antigenic sites prediction, we started from the amino acid sequence of glycoprotein hemagglutinin-neuraminidase (NDV-HN) of Newcastle disease virus, La Sota strain, taken up from the database Protein Data Bank, with the access number P32884 [8].

Two of the features widely used for antigenic determinants prediction are hydrophiliciy respectively hydrophobicity of proteins.

Hydrophobic profile. First, we used an algorithm based on amino acids hydropathy. A hydropathic scale combines the hydrophilic and hydrophobic properties of the 20 essential amino acids. For each aminoacid each was assigned a hydrophilic or hydrophobic index, and depending on the authors, there are more ways to build those scales [9,10]. The hydrophobic index is a measure of the free energy for an aminoacid while transferring it from aqueous to organic solvents. The most used method, with the highest accuracy is that of Kyte & Doolittle which we applied as well in this study [10].

The hydropathic profile of a protein is graphically expressed in a coordinate system in which the averaged hydrophobic indices of amino acids are plotted versus each amino acid position in the protein sequence. The entire sequence is scanned by a moving "window" spanning 6-21 amino acids, and the average indices for each encountered amino acid are computed. The average hydrophobicity was calculated by the following equation [10,11]:

$$\overline{H} = \frac{\left[\sum_{i=segment} H(i)\right]}{n}$$

where H(i) is the hydropathic index of residue i, *segment* indicates the scanning window with a certain number of amino acids, and n is the number of amino acids throughout the entire structure. The peptidic areas with values of H> 0 are considered hydrophobic.

Hydrophilic profile. In addition to the hydrophobic profile of HN-NDV, we also realized a hydrophilic profile based on the algorithm described by Hopp & Woods [9]. This algorithm states that the highest probability of antigenic determinants occurrence is found at the level of highly hydrophilic areas, as they are located on the surface of protein molecules, therefore being mostly exposed to antibody binding sites.

Such calculations are extremely laborious, so there were developed various softwares that allow the construction of proteins hydropathic profile, simply by introducing those amino acids sequences.

The current study was conducted on such logical program that is available online, namely ExPAsy - ProtScale [12].

The frequency of protein secondary structures has been structures. The profile of α-helix and β-sheet frequencies was achieved using the Chou-Fasman algorithm [13,14,15] by means of ProtScale software. This algorithm takes into account the probability that each of the 20 amino acids may be involved into an α or β-type structure, and computes the likelihood of amino acids to appear in a particular structure (the propensity for a secondary strucure); also it computes the conformational parameters based on the frequency of their occurrence and on the amino acids fraction in the given structure.

In addition, we carried out a virtual profile of antigenicity for HN-NDV glycoprotein through other two programs. The first of them - *Antigenic Prediction* [2] - relies more on hydrophobic features of proteins and the second, *CTLPred* [3], has a complex algorithm, based on the sequential transmission of nerve impulses (Artificial Neuronal Network - ANN).

Results and Discussion

The most important glycoprotein of Newcastle disease virus involved in antigen-antibody reactions is hemagglutinin- neuraminidase. Starting from its amino acid sequence (577 amino acids), we achieved the hydrophobic and hydrophilic profiles through ProtScale software (Figure 1).

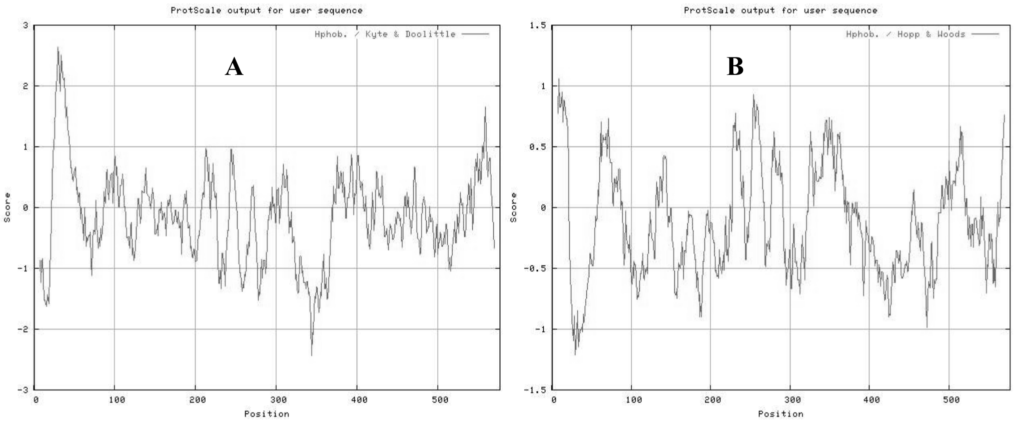

Fig. 1. Hydrophobic (A) and hydrophilic (B) profiles for hemagglutinin-neuraminidase glycoprotein, Newcastle disease virus, La Sota strain (ExPAsy – ProtScale)

Because in general antigenic determinants contain 5-12 amino acids, we used an average moving "window" of 15 amino acids which scanned the entire HN protein.

The software recorded a hydrophobicity maximum of 2.640 with a minimum value of -2.440, whereas the hydrophilic maximum was of 1.060 and the minimum of -1.213. The lowest hydrophobicity values correspond to α-helix protein structures, while the maximum values indicate a high likelihood of hydrophobic interactions leading to "core" protein formation. Areas with intermediate values are accessible to the solvents from the outer molecule surface. Considering the data from literature, it appears that the highest values are found near or within an antigenic determinant.

The N and C-terminal ends of hemagglutinin-neuraminidase were characterized by the presence of hydrophilic amino acids, while the middle area presented an alternate hydrophobic-hydrophilic surface. A rich hydrophobic area was approximately located at level of residues 25 - 50. In exchange, the poorest hydrophobic area was located to residues 325 - 375, where, in fact, there was recorded an area of higher hydrophilicity. In terms of hydrophobicity there were observed 8 potential antigenic sites at amino acids residues: 25-50, 80-115, 125-150, 210-225, 240-255, 290-315, 375-415 (3 picks), 535-560. The hydrophilic profile showed 9 antigenic sites at residues: 1-25, 50-90, 125-150, 225-240, 250-270, 280-290, 325-375, 500-510 and 560-577. Generally, the hydrophobic sites are adjacent to the hydrophilic ones.

However, not all the antigenic determinants presented high values of hydropathy, there as the data should be correlated as well with other physical-chemical and structural properties of studied protein. Thus, we analyzed the frequency of α and β secondary structures by means of ProtScale software (Figure 2).

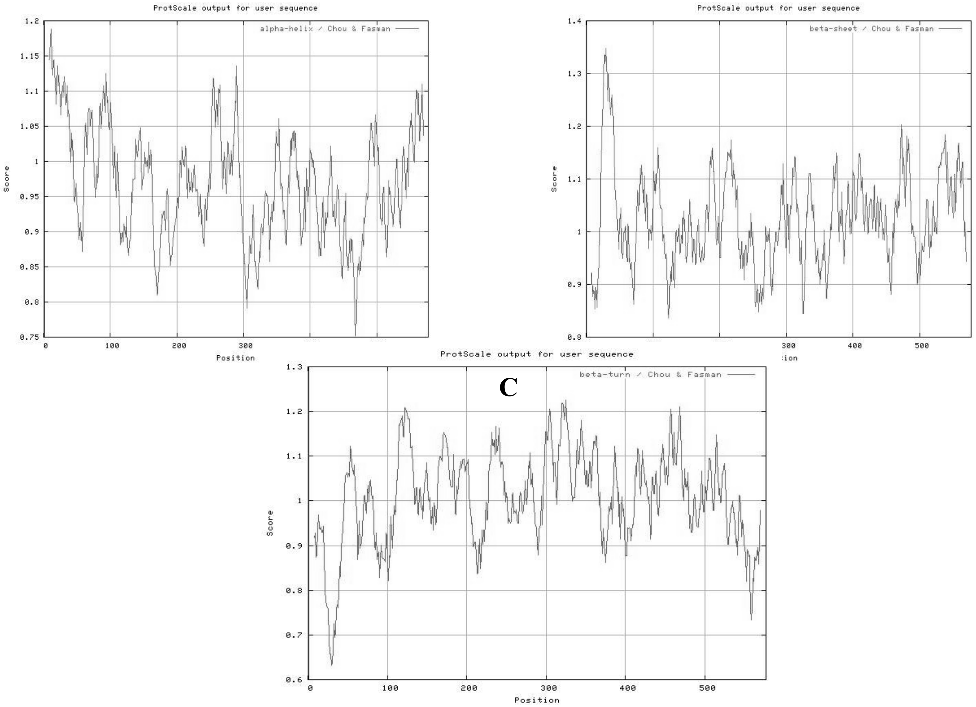

Fig. 2. The frequency of secondary structures for hemagglutinin-neuraminidase, Newcastle disease virus, LaSota strain (ExPAsy – ProtScale); (A – α helix; B – β sheet; C – β turn)

Due to the large length of the protein chain, we noticed a big number of picks for secondary structures. Each α-helix ascendant area corresponds to a descendent area of β-sheet. All secondary structures are possible, because of their frequency index which has an average of about 1. Maximum for β-sheet was estimated at 1.348, a value higher than the maximum of α-helix frequency (1.188). There were no extensive areas just for one type of structure, but only an alternate disposition on sequence length. Based on the criterion that states the β-type structures are characteristic to antigenic sites, we identified the locations for the following antigenic determinants: 10-50, 75-90, 95-115 (2 picks), 180-200, 205-225, 285-300, 305-320, 325-350, 375-385, 400-450, 465-490 and 530-560. The β-turn structures had a maximum frequency of 1.225 and we identified the following antigenic sites: 40-60, 65-85, 100-130, 170-185, 190-205, 225-250, 275-285, 295-315, 315-335, 340-365, 380-390, 415-430, 440-475 and 500-520. It was noticed a slight alternance of β-type structures, with the dominance of β-turn ones. However, the N and C-terminal ends are of α-helix type.

It is known that to interact with antigens, antibodies must have access to the specific antigenic sites. Proteins are three-dimensional structures and have areas which are not accessible, for example, the central area of globular proteins to which antibodies, being large molecules, cannot reach. The most accessible areas are those areas of the protein molecule that are exposed on the surface structure, and being in contact with the aqueous environment, are usually hydrophilic. Also, the flexibility of peptidic segments plays an important role for antigen-antibody interactions, because during the complex formation it is necessary a structural alteration of antigenic determinants in order to obtain the most energetically favourable conformation and as well the best match of the areas coming into contact. Therefore, we analyzed the accessibility and flexibility of amino acid residues that might be involved in the antigenic sites. The virtual profiles of amino acids accessibility and flexibility are shown in Figure 3.

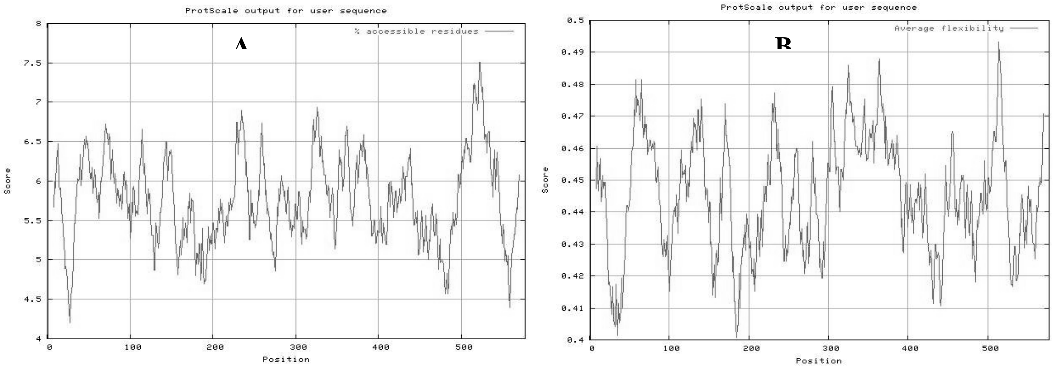

Fig. 3. The accessibility and flexibility of peptidic segments for hemagglutinin-neuraminidase, Newcastle disease virus, LaSota strain (ExPAsy – ProtScale). (A – accessibility, B – flexibility)

We noticed a high accessibility of the HN glycoprotein chain, the maximum index value being 7.507. It can be observed the presence of a large number of accessible areas characterized by high flexibility, but it does not necessarily suggest the existence of an antigenic site. The fragment 40-160 has a middle accessibility (compared with the highest index) and is characterized by the presence of two broad areas of high flexibility. The area with the greatest flexibility, both in terms of occupied surface and indexes values, is located at position 300-400, where amino acid residues have an acceptable accessibility; this area is reach in β-type structures, so that there is a great possibility that antibodies against NDV to recognize and bind to this sequence. The most accessible residues are found next to C-terminal end, position 490-540. The first 20-25 amino acids of this sequence have an increased flexibility, the remaining sequence being quite rigid. However, due to the great accessibility, this polypeptidic area may be the subject of an antigen-antibody interaction.

Nevertheless, we did not notice any major antigenic determinant, but only the presence of a large number of sequences with a length of 10-15 amino acids, which meet the physical-chemical and structural features of antigenic sites. To more accurately locate them, we made a comparative analysis of these sequences, which is depicted in Table 1.

Table 1. Antigenic determinants prediction of hemagglutinin-neuraminidase glycoprotein (Newcastle disease virus) depending on physical-chemical and structural properties

Algorithm	Location of predicted antigenic determinants
Hydropathy (hydrophobicity)	25-50, 80-115, 125-150, 210-225, 240-255, 290-315, 375-390, 390-400, 405-415, 535-560
Hydrophilicity	1-25, 50-90, 125-150, 225-240, 250-270, 280-290, 325-375, 500-510, 560-577
β-sheet	10-50, 75-90, 95-115, 180-200, 205-225, 285-300, 305-320, 325-350, 375-385, 400-450, 465-490, 530-560
β-turn	40-60, 65-85, 100-130, 170-185, 190-205, 225-250, 275-285, 295-315, 315-335, 340-365, 380-390, 415-430, 440-475, 500-520
Accessibility	1-15, 30-75, 75-90, 105-120, 130-155, 215-240, 250-260, 275-290, 310-345, 350-390, 410-445, 490-540, 560-577
Flexibility	1-25, 45-90, 105-150, 160-175, 215-235, 250-260, 280-290, 300-390, 450-480, 500-515, 550-577

Table 2. Hemagglutinin-neuraminidase glycoprotein (Newcastle disease virus, LaSota strain), antigenic determinants. Features

Antigenic determinant	Features
10 – 25	Highly hydrophilic sequence with α-helix and β-turn alternating structure; accessible and flexible
25 – 50	Highly hydrophobic sequence, predominantly β-sheet, partly accessible; not flexible
50 – 90	Highly hydrophilic sequence with β-sheet and β-turn structures; accessible and very flexible
100 – 115	Highly hydrophobic sequence with β-sheet and β-turn structures; high accessibility and flexibility
125 – 150	Mid hydrophobicity and hydrofilicity, α-helix (predominantly) and β-turn structure; mid accessibility, high flexibility
170 – 185	Hydrophobic – hydrophilic alternating sequence, with β-sheet and β-turn structures; very flexible, not accessible
215 – 225	Hydrophobic sequence, with β-sheet structure, accessible; high flexibility
230 – 240	Hydrophilic sequence with β-turn structure; accessible and flexible
250 – 260	Hydrophilic sequence with mainly α-helix structure, accessible; mid flexibility
275 – 290	Hydrophilic sequence, β-sheet and β-turn structures; high flexibility, not accessible
295 – 310	Hydrophobic sequence, β-sheet and β-turn structures; high flexibility, not accessible
315 – 325	Hydrophobic sequence, predominantly β-sheet and β-turn structures; high accessibility and flexibility
330 – 345	Hydrophilic sequence, β-sheet structure; high accessibility and flexibility
350 – 365	Hydrophilic sequence, β-turn structure; high accessibility and flexibility
370 – 390	Mainly hydrophobic sequence with β-sheet and β-turn structures, accessible; low flexibility
410 – 430	Hydrophilic sequence, β-sheet and β-turn structures partly accessible; not flexible
440 – 450	Low hydrophilic sequence with β-sheet structure, low accessibility; not flexible
455 – 475	Low hydrophilic sequence, β-sheet and β-turn structures; low flexibility, not accessible
500 – 510	Hydrophilic sequence, α-helix and β-turn structures with high accessibility and flexibility
530 – 540	Hydrophobic sequence with β-sheet structure; very accessible and flexible
540 – 560	Hydrophilic sequence, β-sheet structure with low flexibility; not accessible
560 – 577	Hydrophilic sequence, α-helix (predominantly) and β-turn structures accessible and flexible at β-turn level

One can remark that these antigenic determinants contain both hydrophobic and hydrophilic amino acids, and predominant secondary structures are those of β-type. Some of these sequences (170-185, 275-290, 295-310, 455-475, 540-560), even if they have antigenic features, they are not accessible, suggesting that following the process of polypeptidic chain folding, they are located into hidden areas inside molecule, forming the protein "core". The fact that there are rigid sequences does not exclude the possibility of their recognition by specific antibodies, as long as they are located on the surface molecule.

In addition, we developed a virtual profile of the antigenicity for glycoprotein HN-NDV by using the *Antigenic Prediction* and *CTLPred* programs. In the following table there are shown the results obtained by running these programs, compared with the antigenic determinants predicted above (by sequential analysis of their properties).

Table 3. The comparison of antigenic determinant of hemagglutinin-neuraminidase (Newcastle disease virus) predicted by different methods

Prediction method (antigenic determinant location)		
Sequential prediction	**Antigenic Prediction**	**CTLPred**
10 – 25	4 – 10	2 – 11
25 – 50	22 – 47	15 – 49
50 – 90	52 – 61	144 – 153
100 – 115	76 – 97	157 – 166
125 – 150	107 – 114	212 – 221
170 – 185	124 – 130	225 – 237
215 – 225	137 – 165	269 – 285
230 – 240	182 – 212	346 – 354
250 – 260	217 – 229	387 – 398
275 – 290	235 – 253	408 – 427
295 – 310	262 – 269	430 – 439
315 – 325	282 – 290	461 – 478
330 – 345	312 – 318	481 – 493
350 – 365	330 – 337	497 – 506
370 – 390	369 – 394	540 – 549
410 – 430	403 – 429	550 - 560
440 – 450	436 – 444	
455 – 475	451 – 480	
500 – 510	501 – 507	
530 – 540	513 – 519	
540 – 560	525 – 546	
560 – 577	554 - 565	

Analyzing the presented sequences, we can say that there is a good correlation between the results obtained y the three prediction methods with a difference of ±5-8 amino acids. From the five sequences listed as not accessible by sequential prediction, only the site 170-185 was not indicated as antigenic, the others being recognized by *Antigenic Prediction* software as well. The *CTLPred* software identified a smaller number of antigenic determinants, most of them being located in the second half of the molecule, to the C-terminal end, where amino acids accessibility is higher.

Once an antigen determinant has been predicted, its presence should be verified by chemical synthesis of its sequence and testing its activity by appropriate immunochemical methods as ELISA or Western blot.

Based on these predictions, there can be obtained synthetic peptides that offer the advantage of a high purity state, which decreases the risk of nonspecific binding within protein – protein interactions.

These predicted peptides may serve to obtain antibodies with high affinity binding useful in diagnostic techniques, or to obtain very specific antibodies as monoclonal antibodies which can be used both as diagnosis reagents and as prophylactic products coupled with various drug molecules.

Moreover, the antigenic determinants predicted for viruses may be useful for synthetic vaccines manufacture, thus eliminating the use of whole viral particles which also can produce side effects, especially in case of highly pathogenic viral strains.

Conclusions

In order to identify the antigenic determinants of major glycoprotein hemagglutinin-neuraminidase of Newcastle disease virus there was made a virtual prediction study, based on physical-chemical and structural features of amino acids sequence. There were predicted 22 antigenic sites for glycoprotein HN - NDV. The predicted antigenic determinants generally have an alternating hydrophobic - hydrophilic sequence, with β-sheet and β-turn secondary structures, with variable flexibility, being exposed to the surface of viral particle.

Such predictive studies can be particularly useful for increasing the sensitivity and specificity of diagnostic techniques. Thus, this predicted antigenic determinants may be used in order to obtain synthetic peptides with antigenic function, useful in diagnosis, research and vaccines production.

References

1. A.S. KOLASKAR, P.C. TONGAONKAR, *FEBS Lett*., 276, 172-174 (1990)
2. D.T. JONES, W.R. TAYLOR, J.M. THORNTON, *Biochemistry*, 33, 3038-3049 (1994)
3. H. KAUR, G.P. RAGHAVA, *Bioinformatics,* 18, 1508-1514 (2002)
4. V.KRCHNAK, O.MACH, A.MALY, *Anal. Biochem*., 165, 200-207 (1987)
5. V.MOREA, A.TRAMONTANO, M.RUSTICI, C.CHOTHIA, A.M.LESK, *Biophys Chem*., 68, 9-16 (1997)
6. ANA SAGRERA, C. COBALEDA, J. M. GONZALEZ DE BUITRAGO, A. GARCIA-SASTRE, E.VILLAR, *Glycoconjugate Journal*, 18, 283–289 (2001)
7. H.J. TSAI, K.H. CHANG, C.H. TSENG, KAREN M. FROST, RUTH J. MANVELL, D. J. ALEXANDER, *Veterinary Microbiology,* 104(1-2), 19-30 (2004)
8. NCBI, BLAST, http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi
9. T.P. HOPP, K.R. WOODS, *Proc. Natl. Acad. Sci. USA,* 78, 3824-3828 (1981)
10. J. KYTE, R.F. DOOLITTLE, *J. Mol. Biol*., 157, 105-132 (1982)
11. S. MITAKU, T. HIROKAWA, *Prot. Eng*., 12, 953-957 (1999)
12. ExPaSy Molecular Tools, www.expasy.org
13. P.Y. CHOU, G.D. FASMAN, *Adv. Enzymol. Relat. Areas Mol. Biol.,* 47, 145-148 (1978)
14. P.Y. CHOU, G.D. FASMAN, *Biophys J*., 26, 367-373 (1979)
15. J. JANIN, *Nature*, 277, 491-492 (1979)