# VaxiJen
# ViralDB - база данни с вирусни имуногени

**маг.-фарм. Николет Донева**
**докторант в Лаборатория по лекарствен дизайн и биоинформатика, Фармацевтичен факултет, Медицински университет - София**

http://www.ddg-pharmfac.net

**„Обърната ваксинология“** – нова концепция за откриване на протективни антигени, при която дизайнът на ваксините започва с биоинформатичен анализ на геномната последователност на микроорганизми без нужда от тяхното *in vitro* култивиране.

**Имуноинформатика –** раздел от биоинформатиката, фокусиран върху имунологията и ваксинологията, който включва:

- методи за съставяне на база данни;
- извличане на информация от тези данни и анализ;
- *in silico* методи и алгоритми за предсказване на имуногенност;
- методи за предсказване на В-клетъчни и Т-клетъчни епитопи

**VaxiJen** е биоинформатичен инструмент за *in silico* идентифициране на имуногени от бактериален, **вирусен**, гъбичен и туморен произход и за предсказване на протективни антигени, туморни антигени и субединични ваксини. Създаден е в Лабораторията по лекарствен дизайн и биоинформатика, Медицински университет - София

ИКТ


Проект BG05M2OP001-1.001-0003 „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.

**Цел:** *Актуализация на сървъра VaxiJen 2.0 за предсказване на имуногенност на протеини от вирусен произход*

**Основни стъпки:**

1. Компилиране на бази данни с вирусни имуногени;
2. Включване на базата данни в сървъра на VaxiJen като web апликация.
3. Компилиране на обучаваща и тестова група;
4. Прилагане на набор от методи за контролирано машинно обучение и извеждане на модели за разпознаване на имуногенни протеини с вирусен произход;
5. Последващо включване на модела/ите с най-добра предсказваща способност в сървъра VaxiJen.

**Приложение** на базата данни с вирусни антигени:

- Търсене на експериментално идентифицирани имуногенни протеини за даден патогенен вирус;
- Машинно обучение на *in silico* модели за имуногенно предсказване;
- Верифицирана база данни за задълбочен анализ на имуногенност на протеини от вирусен произход.


*Проект BG05M2OP001-1.001-0003 „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.*

# Източник на данни



Статии от Национална медицинска библиотека на САЩ **PubMed**. За търсенето на нови имуногенни протеини са използвани следните ключови думи:

(virus OR viral) AND (candidate OR candidates OR subunit) AND (protects OR protect OR protection OR protective) AND (vaccine OR vaccines) AND (human)

Филтри:

Results by year: 2010-2021;

Text availability: all;

Article types: all

Резултат:

**4101** статии за вирусни антигени

Тези статии и съответните референции са ръчно проверени само за вирусни антигени **с доказаната имуногенност при хора**.



*Проект BG05M2OP001-1.001-0003 „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.*

## Аминокиселинни последователности

- Аминокиселинните последователности на всички открити вирусни протеини/антигени са потърсени в **NCBI и UniProtKB**, с добавен филтър “Reviewed”;
- Представени са във FASTA формат;
- В случаите, в които съществуват няколко изоформи за един протеин, всичките изоформи са включени;
- Всички налични щамове/серотипове/изолати на даден вирус са включени.
- В случаите, в които няма отделен запис само за антигена, е включена цялата полипротеинова последователност.

Резултат:

Броят на **вирусни антигенни последователности**, принадлежащи към **31 вирусни вида**, с включени всички налични щамове/серотипове/изолати за даден вирус е **1788**.


*Проект BG05M2OP001-1.001-0003 „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.*

Базите данни **Uniprot-Proteomes** и/или **UniProtKB** са претърсени за вирусни протеоми, с цел създаване на негативен сет за последващото моделиране.

**Uniprot-Proteomes:**

- Използваните филтри са Complete Proteome Detector – “Standart” или “Close to Standard”.
- В случаите, в които съществуват различни протеоми за щам/серотип/изолат, тези с най-висок “protein count” са включени.

**UniProtKB:**

- Използваните филтри са “Reviewed records”.
- Всички налични данни за полипротеини/геномни полипротеини за щам/серотип/изолат са включени.

Събраните протеоми са обработени със сървъра VaxiJen 2.0 и са избрани **468** неимуногенни протеина.


*Проект BG05M2OP001-1.001-0003 „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.*

# Допълнителни източници на данни



1. Базата данни **IEDB** е претърсена за секвенции на **Т- клетъчни епитопи**.

   Филтри: Epitope Source: virus, Host: human; MHC Restriction: any; Epitope: linear peptide, disease: any; Assay: Tcell epitopes

   Резултат: 48657 позитивни и **54407** негативни записа. След премахване на повторенията от позитивния сет броят им е редуциран до **21828**.

2. Базата данни **IEDB** е претърсена за секвенции на **В- клетъчни епитопи**.

   Филтри: Epitope Source: virus, Host: human; MHC Restriction: any; Epitope: linear peptide, disease: any; Assay: Bcell epitopes

   Резултат: 123625 позитивни и **472148** негативни записи. След премахване на повторенията от позитивния сет броят им е редуциран до **81000**.

3. Базата данни **ClinicalTrials.gov** е претърсена ръчно за текущи и приключили **клинични изпитвания** с вирусни имуногени.

ИКТ


*Проект BG05M2OP001-1.001-0003 „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.*

## VaxiJen 3.0: Database of viral antigens

**VaxiJen 3.0: Database of viral antigens**

Choose a virus species:

- Ebola virus
- Chikungunya virus
- Dengue virus
- Enterovirus A71
- Epstein-Barr virus
- Equine encephalitis virus
- Hantaan virus
- Hepatitis B virus
- Hepatitis C virus
- Hepatitis E virus
- Human immunodeficiency virus
- Human cytomegalovirus
- Human herpes simplex virus
- Varicella-zoster virus
- Human metapneumovirus

Help

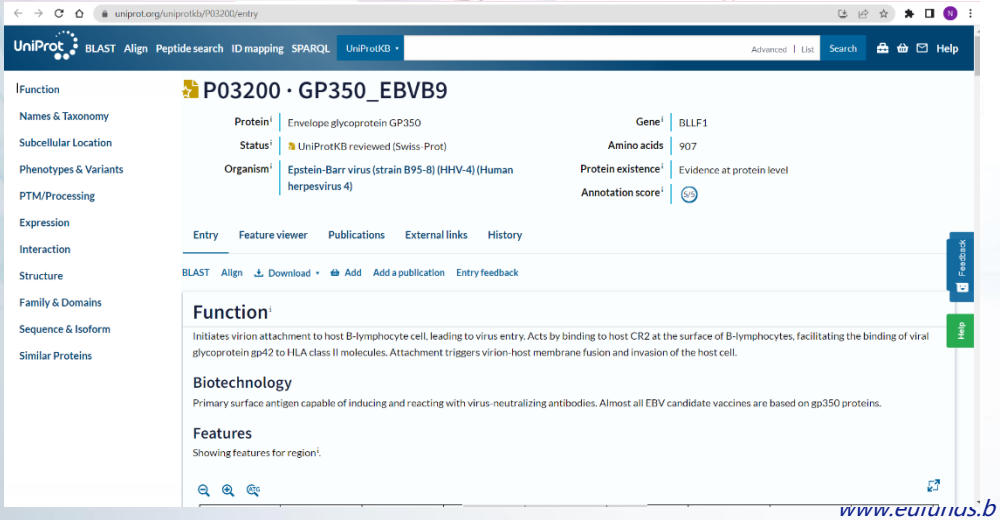
Submit



*Проект BG05M2OP001-1.001-0003 „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.*

## VaxiJen 3.0 Database of viral antigens — Epstein-Barr virus

| Virus type | Virus strain/Serotype/Isolate | Protein id | Protein_name | Antigen name | Reference | Clinical Trials Data |
|---|---|---|---|---|---|---|
| Epstein-Barr virus | strain B95-8 | P03200 | Envelope glycoprotein GP350 | Envelope glycoprotein GP350 | 17485150 18190254 19855249 25671130 25694592 26485517 26920891 29718378 32367191 7796951 | NCT00430534 NCT04645147 |
| Epstein-Barr virus | strain B95-8 | P03211 | Epstein-Barr nuclear antigen 1 | Epstein-Barr nuclear antigen 1 | 23348421 25124688 25671130 32367191 | NCT01147991 NCT01800071 |
| Epstein-Barr virus | strain AG876 | Q1HVJ2 | Latent membrane protein 2 | Latent membrane protein 2 | 23348421 25124688 25671130 32367191 | NCT01147991 NCT01800071 |



## VaxiJen 3.0 Database of viral antigens:

| Antigen name | Antigen sequence | Known B cell epitopes | Known T cell epitopes |
|---|---|---|---|
| Envelope glycoprotein GP350 | MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTADVNVTINFDVGGKKHQ<br>LDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGELALTMRSKKLPINVTTGEE<br>QQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYIKWDNCNSTNITAVVRAQGLD<br>VTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEISQVLPGDNKFNITCSGYESHVP<br>SGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLGNNSILYVFYSGNGPKASGGDYCIQ<br>SNIVFSDEIPASQDMPTNTTDITYVGDNATYSVPMVTSEDANSPNVTVTAFWAWPNNTET<br>DFKCKWTLTSGTPSGCENISGAFASNRTFDITVSGLGTAPKTLIITRTATNATTTTHKVI<br>FSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPASTGPTVSTADVTSPTPAG<br>TTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPTPAVTTPTPNATSPTPA<br>VTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSP<br>TSAVTTPTPNATGPTVGETSPQANATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSS<br>STSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSP<br>APRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKAN<br>STTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVT<br>TAQATVPVPPTSQPRFSNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPY<br>DDAETYV | 176845 244105 65466 33650 10888 49539<br>52792 76275 696127 52367 694870 694223<br>694417 694273 694462 696613 695079 35210<br>695402 696616 75533 693953 695263 694572<br>695400 696407 696777 694465 694428 696957<br>693750 695580 695249 696745 695534 694189<br>695987 694214 694219 695560 695008 694608<br>696723 696652 694790 693996 696206 696165<br>695435 697051 695979 693856 695617 696553<br>693984 695607 694076 695141 696701 695848<br>694296 694516 694398 696896 694777 695128<br>695011 693952 696736 694860 694486 695671<br>696734 696600 696871 695390 696697 695931<br>695884 695725 695224 696670 696503 696703<br>695367 696320 696592 695557 696716 695600<br>696529 695531 695795 696376 694577 696918<br>696595 694843 696722 696310 696739 695790<br>697023 695363 693961 696861 695398 695275<br>695438 693768 695569 696394 694898 | 16044<br>69731<br>65746<br>149795 |

IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

Help | More IEDB

Home | Specialized Searches | Analysis Resource

**EPITOPE SUMMARY**

STT is a linear peptidic epitope (epitope ID 176845) studied as part of Capsid protein VP1 from Norwalk virus (Norwalk calicivirus)., tested in 12 B cell assays

**COMPILED DATA**

B Cell Assay(s) 12

| Assay Type | Positive / All |
|---|---|
| qualitative binding | 6/6 |
| neutralization | 4/4 |
| hemagglutination inhibition | 2/2 |

Provide Feedback | Help Request | Solutions Center | Tool Licensing Information

Last Updated: November 20, 2022

Supported by a contract from the National Institute of Allergy and Infectious Diseases, a component of the National Institutes of Health in the Department of Health and Human Services.

IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

Help | More IEDB

Home | Specialized Searches | Analysis Resource

**EPITOPE SUMMARY**

FGQLTPHTKAV is a linear peptidic epitope (epitope ID 16044) studied as part of Envelope glycoprotein GP350 from Human herpesvirus 4 (Epstein Barr virus)., tested in 3 T cell assays

**COMPILED DATA**

T Cell Assay(s) 3

| Assay Type | Positive / All |
|---|---|
| proliferation | 3/3 |

**EXTERNAL RESOURCES**

| Resource | Link |
|---|---|
| IEDB-AR: MHC-I Processing | Predict MHC class I processing |
| IEDB-AR: MHC-I | Predict MHC class I binding affinity |
| IEDB-AR: B cell scales | Predict B cell epitopes |

Provide Feedback | Help Request | Solutions Center | Tool Licensing Information

Last Updated: November 20, 2022

Supported by a contract from the National Institute of Allergy and Infectious Diseases, a component of the National Institutes of Health in the Department of Health and Human Services.

ИКТ

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

PubMed.gov

Search

Advanced User Guide

Save Email Send to Display options

Clinical Trial > Vaccine. 2007 Jun 11;25(24):4697-705. doi: 10.1016/j.vaccine.2007.04.008.
Epub 2007 Apr 18.

## Phase I/II studies to evaluate safety and immunogenicity of a recombinant gp350 Epstein-Barr virus vaccine in healthy adults

Michel Moutschen [1], Philippe Léonard, Etienne M Sokal, Françoise Smets, Michèle Haumont, Pasqualina Mazzu, Alex Bollen, Francoise Denamur, Pascal Peeters, Gary Dubin, Martine Denis

Affiliations + expand

PMID: 17485150 DOI: 10.1016/j.vaccine.2007.04.008

### Abstract

Two double-blind randomised controlled studies (phase I and I/II) were performed to assess for the

FULL TEXT LINKS

ELSEVIER FULL TEXT ARTICLE

ACTIONS

Cite

Collections

SHARE

PAGE NAVIGATION

NIH U.S. National Library of Medicine
ClinicalTrials.gov

Find Studies About Studies Submit Studies Resources About Site PRS Login

Home > Search Results > Study Record Detail

Save this study

**Evaluation of Safety, Immune-Response and Efficacy of GSK Biologicals' EBV (Epstein Barr Virus) Vaccine (268664).**

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT00430534

Recruitment Status : Completed
First Posted : February 2, 2007
Last Update Posted : February 2, 2007

View this study on Beta.ClinicalTrials.gov

**Sponsor:**
Henogen

**Information provided by:**
Henogen

Study Details Tabular View No Results Posted Disclaimer How to Read a Study Record

# *Благодаря за вниманието!*



*Проект BG05M2OP001-1.001-0003 „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.*