Helix (2022) 12 (3): 15-20

https://doi.org/10.29042/2022-12-3-15-20

Proteomic analysis and Identification of potential peptide Vaccine against pathogenic proteins of Pseudomonas

Alle Rahul

Bharath Institute of Higher Education and Reesearch, selaiyur, Chennai-600073,India,
rahulyuva21@gmail.com

Received: 14th April 2021, Accepted: 20th May 2022, Published: 30th June 2022

Abstract

Pseudomonas is a gram negative facultative anaerobic bacterium with ubiquitous occurrence. It is one of the common pathogens infecting humans and animals. It would cause acute to chronic infections and forms an opportunistic parasite for immune suppressed individuals. Some of the most common infections include malignant external otitis, pneumonia, septicemia etc. In alignment with its role in threatening of human and animal health the study aimed to annotate the complete proteome of the organism for identification of pathogenic protein cluster among them. Further based on the identified proteins the suitable peptide vaccines are to be designed which can prevent the atrocities caused by this bacterium. The 3D structure of proposed peptide vaccine can further be designed using ARGUS LAB followed by it validation. The planned work could screen and identify major pathogenic peptides from the bacterium which can be designed and validated. The use of these designed peptides as vaccines can be one of the preventive approaches for all infections caused by this opportunistic pathogen.

Key words: Facultative anaerobic, Opportunistic pathogen, Alignment, Pathogenic, 3D structure, Argus Lab, Peptide Vaccine.

Introduction

Microorganisms are ubiquitous in nature and are the major biotic components whose association with human life is indispensable. There exists no region without the presence of microbes naturally. Though all the microbes are not pathogenic in nature some of the common inhabitants can also be highly pathogenic and may cause life threatening infections and diseases. All microbial infections can be better prevented using vaccines which are protein or whole organism based preparations that can immunize an individual and prevent the infection by the pathogen upon entry. One among the common pathogenic bacteria causing human infections is *Pseudomonas aeruginosa.*

Pseudomonas species is a group of heterophilic, aerobic bacilli ranging in size from 0.5- 0.8 by 1.5-3.0 µm [7]. It is motile with its single polar flagella. Antibodies to the bacterial outer membrane proteins and lipopolysaccharides exhibit cross reactivity [3]. With focus to its nutritional requirement, it can utilize a wide variety of organic molecules as a sole carbon source. It can be easily grown on minimal salt medium with any basic carbon source [2]. It can withstand the temperatures ranging from 4 to 42^0C. It is a common soil bacterium which is capable of breaking down of polycyclic aromatic hydrocarbons. It is also found in other habitats like reserved water contaminated by animals and humans like sewage, sink etc [1]. Being an opportunistic parasite this bacteria is found especially in the hospital dumps and wastes.

Pseudomonas is known to possess several virulence factors however the exact mechanism of pathogenesis is unclear [9]. The bacterium takes control over the hosts weakened immune system and causes diseases. Blood stream infections, Pneumonia, Urinary tract infections, and surgical wound infections are some of the infections caused by Pseudomonas species.

Materials and methods

Collection of bacterial proteome and screening

Pseudomonas is known to possess a genome of the size 6.3 Mbp corresponding to approximately 2000 non redundant proteins [10]. Among all the proteins of the bacteria, it is those proteins that are foreign to human proteome which can be pathogenic and possess a potential to cause infection in the host. All the proteins of the bacterium were collected from NCBI database and were screened with the human proteome to check for their degree of similarity. The proteins that share no or less than 30% similarity to human proteins are selected as foreign and subjected for further annotation. BLAST P performs similarity search between the proteins and provides the ranking in terms of score [6]. Higher score is an indication of evolutionary similarity and non pathogenic nature. All the proteins that share a very less degree of similarity to human proteins are collected and considered for further annotation.

Antigenic site prediction and propensity calculation

All the sequences identified to be foreign can be tested for the presence of antigenic region/ peptide. PVS is one such online server used to predict the various structural and functional properties of the proteins. Tool can efficiently predict total number of antigenic regions present in protein sequence along with the details including antigenic propensity (antigenic intensity). A graphical representation along with total number of antigenic peptides and their propensities is displayed in the results. Based on peptide with highest peak in the graph most antigenic peptide can be selected. Tool can be accessed from link http://imed.med.ucm.es/PVS/. The peptides selected can be further subjected for antigenic site identification within the selected region. EMBOSS ANTIGENIC [11] can be used for the purpose. EMBOSS is a facility provided at the link https://www.bioinformatics.nl/cgi-bin/emboss/antigenic.

VAXIJEN

A new era antigen prediction server which is purely free from alignment protocol but runs on the principle of Autocross covariance transformation of proteins [4]. The tool is based on the datasets and models of Bacteria, virus and tumor proteins for prediction of whole protein antigenecity. This prediction method is purely based on physicochemical properties and eliminates the alignment based approach to obtain high accuracy in antigenic region prediction. The submission can be either single sequence or multiple upload with the option for selecting a single organism for comparison based on users requirement. The tool is tested and validated and reveals its accuracy of antigenic prediction to a degree of 75 to 80%.

TMHMM [8] was used to predict the regions of protein that form the trans membrane domain and those falling on the external surface. CTL epitopes on the protein were predicted based on **NetCTL** [5] approach. The prediction is based on the identification of MHC binding to an extended class no of 12 MHC super types which includes the supertype A26 and B39.

Additionally the advanced version 1.2 used in the study has demonstrated a higher rate of accuracy than the previous version 1.0 imparting high degree of accuracy to the data. Principle of artificial neural network is involved in prediction of MHC Class 1 binding and proteosomal cleavage. TAP transport efficacy is predicted using weight matrix.

Model building of the predicted peptide was performed using the software Argus lab [12]. The software has a free access and is a downloadable version with high end accuracy in peptide building and chemical designing. The designed 3D structure can be cleaned and optimized to obtain a most stable structure and facility for energy calculation of molecule is provided. Cleaned and optimized structure can be downloaded and used for any further study.

Results and Discussion

The screening of the proteome resulted in the selection of 41 proteins that shared least similarity to human proteome and are suspected to be antigenic thus causing pathogenecity. All the 41 protein sequences were further analyzed using different aspects and identification of the most antigenic region within the selected proteins was performed.

Initial screening was based on the results of PVS and EMBOSS that identifies the antigenic regions/ peptides within the proteins. The resultant data was screened and filtered for the finalization of most antigenic regions showing high degree of antigenic propensity. Simultaneously the same list of proteins

were also screened using EMBOSS antigenic that not only identifies the antigenic region but also the site showing antigenicity.

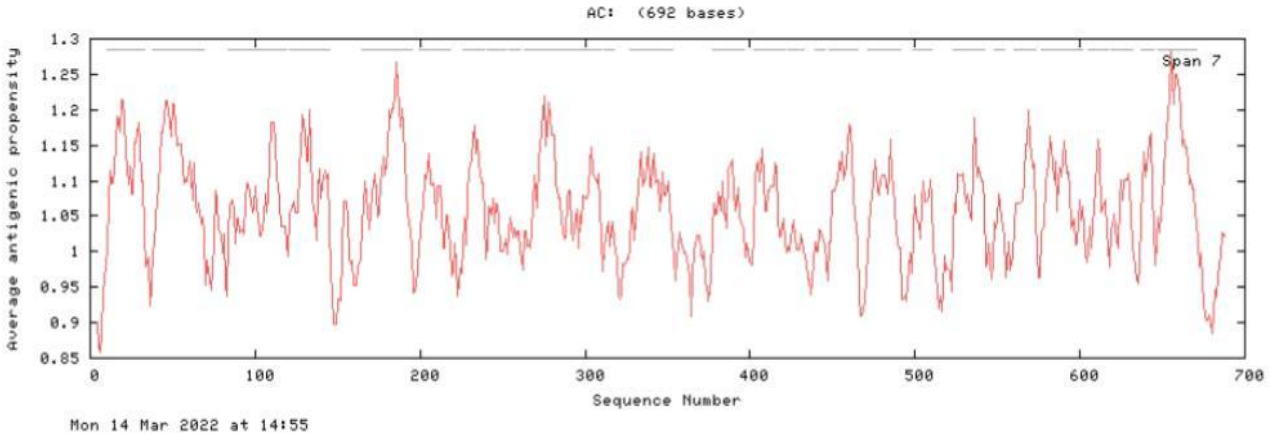
The proteins with peptides having highest antigenic propensity were found to be **Carbon starvation protein A and Two-component sensor histidine kinase** based on the initial screening by PVS. The antigenic propensity for both the proteins was 1.27. Thus they can be filtered out for further analysis. The protein sequences and the result of PVS are furnished in the Figures below.

Amino acid sequences of the two proteins:

>OHQ72594.1 carbon starvation protein A [Pseudomonas aeruginosa]
MNNNNSLLRHLAWLVVAIVGAAALGVVALRRGEAINALWIVVAAVAIYLVAYRYYSLFIASK
VMQLDPNRATPAVLNNDGLDYVPTNKHILFGHHFAAIAGAGPLVGPVLAAQMGYLPGTLWLI
AGVVLAGAVQDFMVLFISSRRNGRSLGELVREEMGQVAGTIALFGAFLIMIIILAVLALIVVKA
LADSPWGMFTVLATIPIALFMGVYMRFIRPGRIGEISIIGVFLLLGSIWLGGQVAASPEWAPHFT
FSGIQITWMLIGYGAVASVLPVWLLLAPRDYLSTFLKIGTIIGLAIGILIVMPELKMPALTQFTDG
TGPVWKGSLFPFLFITIACGAVSGFHALISSGTTPKLLNREPDARYIGYGGMLMESFVAIMAMV
AASVIEPGIYFAMNSPPAVVGADVNAVAATVSSWGFAITPEQLTQTAQDIGETTILARAGGAPT
LAVGIAHILHQVLPGENTMAFWYHFAILFEALFILTAVDAGTRAGRFMLQDLLGNFVPALKKT
ESWTANIIGTGGCVALWGWLLYQGVVDPLGGINTLWPLFGISNQMLAGIALMLATVVLIKMK
RQQYVWVTILPAAWLLICTTTAGLIKIFDSNPAVGFVALGEKYATALDAGQVLAPAKDIGQM
QHVVLNAYINAGLTVLFLLVVFSVLFYAIKVGIAAWGKSERTDKETPFEPIPDA

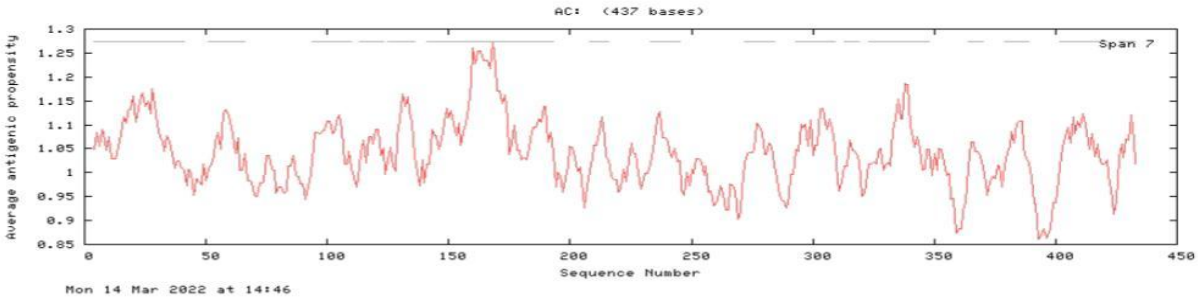
>OFM47081.1 two-component sensor histidine kinase [Pseudomonas aeruginosa]
MKLAVPRPRSLAARLALILFAGLVLAYGLSFASQFYERYQTAKHMMLDSLEQDVAISVAMLD
RLTPAEREAWLPRLERRTYRYRLDAGEPGQPLALADAPVAAHSIERALDGQYPLTLRTVADSR
PHFQVLLRLRDGSPLTIDVTPAPVPLSGWLPLVLLVQLLLLLLCTGLAVRTAIGPLTRLVKAVE
HLDPNRPAQPLAETGPREVAHAAAAFNAMQARIADYLKERMQLLAAISHDLQTPITRMKLRV
EFMDASSDRDKLWNDLEEMQHLVREGVAYARSMHGSTETSCRVDLDAFLDSLVFDYQDSGK
QVQLDGRTGAVIDTRPHALRRVLVNLVDNALKFAGAARLEVERRTDGGTRIQVLDNGPGIPA
EELDEVLKPFYRVENSRNRDTGGTGLGLAIAQQLSLALGGSLTLANRAGGGLCARIELDP

Fig 1: PVS result of Carbon Starvation protein



Inference: Antigenic propensity of approximately 1.27 was found to be highest for the above protein in the region 640-700. Thus this region of the protein can be among most pathogenic.

Fig 2: PVS plot of protein two-component sensor histidine kinase



Inference: The region with highest antigenic propensity in the above protein was found to be between 140 to 200. Which can be further analyzed

The final list of peptides selected based on PVS and EMBOSS ANTIGENIC with highest peak and antigenic propensity are SPLTIDVTPAPVPLSGWLPLVLLVQLLLLLLCTGLAVRTAIGPLTRLVKAVEH (141 to 193) of two-component sensor histidine kinase and INAGLTVLFLLVVFSVLFYAIKVGI (647-671) of carbon starvation protein A. In order to identify weather the same peptides are CTL epitopes Net CTL results are analyzed. This will predict their efficacy to act as T cell epitopic regions.

NetCTL result analysis revealed that the epitopic region of two-component sensor histidine kinase identified by PVS and EMBOSS does not have a T cell epitope as indicated by NetCTL. Thus the protein carbon starvation protein A whose antigenic sites showed to possess a T cell epitope is used for further analysis.

Emboss result for the protein carbon starvation protein A showing the peptide with highest antigenicity is shown below

Fig 3 EMBOSS Result of carbon starvation protein A

OUTPUT FILE outfile

```
########################################
# Program: antigenic
# Rundate: Fri  7 Oct 2022 08:05:49
# Commandline: antigenic
#    -auto
#    -sequence /var/lib/emboss-explorer/output/290231/.sequence
#    -minlen 6
#    -outfile outfile
#    -rformat2 motif
# Report_format: motif
# Report_file: outfile
########################################

#=======================================
#
# Sequence: OHQ72594.1     from: 1   to: 688
# HitCount: 27
#=======================================

Max_score_pos at "*"

(1) Score 1.284 length 25 at residues 644->668
                     *
 Sequence: NAGLTVLFLLVVFSVLFYAIKVGIA
           |                       |
         644                     668
 Max_score_pos: 652

(2) Score 1.266 length 31 at residues 162->192
                                 *
 Sequence: AGTIALFGAFLIMIIILAVLALIVVKALADS
           |                               |
         162                             192
tart Max score pos: 183
```

Inference: The above result shows that the region having highest antigenecity in the protein is from 644 to 668 with the sequence NAGLTVLFLLVVFSVLFYAIKVGIA and the antigenic site identified to be L at 652 position.

The above protein is further screened using NetCTL to check the T cell epitopic regions in it. The results are furnished below.

Fig 4: Results of NetCTL showing T cell epitopic regions

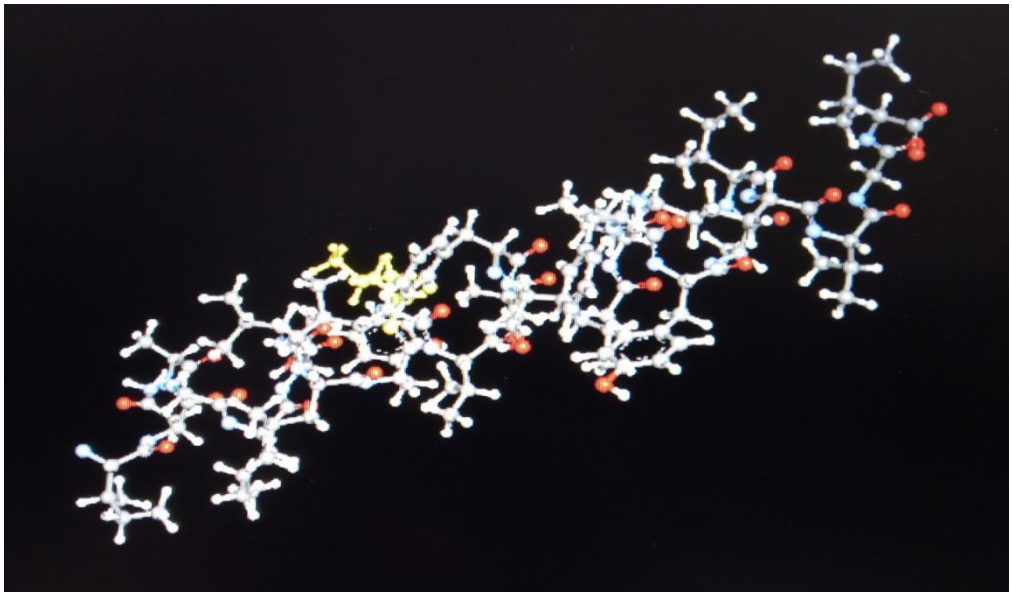
```
638 ID Sequence pep VLNAYINAG aff   0.0656 aff_rescale   0.2785 cle 0.0638 tap  -1.3610 COMB   0.2200
639 ID Sequence pep LNAYINAGL aff   0.0569 aff_rescale   0.2415 cle 0.1148 tap   0.8700 COMB   0.3022
640 ID Sequence pep NAYINAGLT aff   0.0643 aff_rescale   0.2732 cle 0.0245 tap  -0.3810 COMB   0.2578
641 ID Sequence pep AYINAGLTV aff   0.0643 aff_rescale   0.2728 cle 0.7122 tap   0.9330 COMB   0.4263
642 ID Sequence pep YINAGLTVL aff   0.0772 aff_rescale   0.3278 cle 0.8217 tap   1.1230 COMB   0.5072
643 ID Sequence pep INAGLTVLF aff   0.0620 aff_rescale   0.2631 cle 0.3675 tap   2.3700 COMB   0.4368
644 ID Sequence pep NAGLTVLFL aff   0.0781 aff_rescale   0.3316 cle 0.1346 tap   0.8900 COMB   0.3963
645 ID Sequence pep AGLTVLFLL aff   0.0662 aff_rescale   0.2809 cle 0.3166 tap   0.9910 COMB   0.3779
646 ID Sequence pep GLTVLFLLV aff   0.0870 aff_rescale   0.3694 cle 0.7804 tap   0.2730 COMB   0.5001
647 ID Sequence pep LTVLFLLVV aff   0.1166 aff_rescale   0.4953 cle 0.5887 tap   0.3450 COMB   0.6008
648 ID Sequence pep TVLFLLVVF aff   0.0564 aff_rescale   0.2395 cle 0.2160 tap   2.7180 COMB   0.4078
649 ID Sequence pep VLFLLVVFS aff   0.0594 aff_rescale   0.2521 cle 0.0263 tap  -2.0420 COMB   0.1540
650 ID Sequence pep LFLLVVFSV aff   0.0581 aff_rescale   0.2465 cle 0.7296 tap   0.5510 COMB   0.3835
651 ID Sequence pep FLLVVFSVL aff   0.0683 aff_rescale   0.2898 cle 0.9446 tap   0.9450 COMB   0.4787
652 ID Sequence pep LLVVFSVLF aff   0.0833 aff_rescale   0.3537 cle 0.8596 tap   2.5310 COMB   0.6092
653 ID Sequence pep LVVFSVLFY aff   0.2887 aff_rescale   1.2258 cle 0.7920 tap   3.1930 COMB   1.5042 <-E
654 ID Sequence pep VVFSVLFYA aff   0.0635 aff_rescale   0.2697 cle 0.1158 tap  -0.1750 COMB   0.2783
655 ID Sequence pep VFSVLFYAI aff   0.0552 aff_rescale   0.2343 cle 0.2644 tap   0.8070 COMB   0.3143
656 ID Sequence pep FSVLFYAIK aff   0.0883 aff_rescale   0.3749 cle 0.1237 tap   0.5350 COMB   0.4202
657 ID Sequence pep SVLFYAIKV aff   0.0707 aff_rescale   0.3002 cle 0.9347 tap   0.4190 COMB   0.4614
658 ID Sequence pep VLFYAIKVG aff   0.0513 aff_rescale   0.2179 cle 0.0424 tap  -1.1480 COMB   0.1669
659 ID Sequence pep LFYAIKVGI aff   0.0527 aff_rescale   0.2239 cle 0.7675 tap   0.9400 COMB   0.3860
660 ID Sequence pep FYAIKVGIA aff   0.0548 aff_rescale   0.2327 cle 0.0795 tap  -0.3010 COMB   0.2295
```

Inference: The above result of NetCTL shows that the pattern LVVFSVLFY with the score 0.288 is identified to be T cell epitope and this same region was also recognized as the one with highest antigenecity according to PVS . Thus the peptide is identified to be a better target for vaccine development.

Based on the above all analysis the final peptide selected for peptide vaccine designing was INAGLTVLFLLVVFSVLFYAIKVGI corresponding to 647- 671, of the protein carbon starvation protein A [Pseudomonas aeruginosa].

The next step in the study is to construct the 3D structure of this selected peptide in argus lab.

Fig 5: 3D structure of peptide vaccine INAGLTVLFLLVVFSVLFYAIKVGI in Argus lab:



Inference: The final peptide structure in argus lab.

Conclusion

Pseudomonas aeruginosa being one of the most common pathogenic ubiquitously distributed bacteria, forms the target organism for the study. Proteome of the organism was screened for the identification of foreign proteins to human proteome. 41 Protein sequences were selected to be foreign and were subjected for antigenic region prediction. The proteins that show highest antigenicity were identified to be 2 in number and were Carbon starvation Protein A and two component sensor histidine kinase. The proteins were further screened for the identification of T cell epitopic regions and transmembrane regions. All the results analysis revealed that the peptide INAGLTVLFLLVVFSVLFYAIKVGI of carbon starvation protein A was most antigenic, whose 3D structure was developed in Argus software. This peptide can be synthesized and tested on animals for its efficacy in acting as a potential vaccine against the infections caused by the organism *Pseudomonas aeruginosa.*

References

1) Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep 1;25(17):3389-402. doi: 10.1093/nar/25.17.3389. PMID: 9254694; PMCID: PMC146917.

2) Diggle SP, Whiteley M. Microbe Profile: Pseudomonas aeruginosa: opportunistic pathogen and lab rat. Microbiology (Reading). 2020 Jan;166(1):30-33. doi: 10.1099/mic.0.000860. Erratum in: Microbiology (Reading). 2021 Aug;167(8): PMID: 31597590; PMCID: PMC7273324.

3) Huszczynski SM, Lam JS, Khursigara CM. The Role of Pseudomonas aeruginosa Lipopolysaccharide in Bacterial Pathogenesis and Physiology. Pathogens. 2019 Dec 19;9(1):6. doi: 10.3390/pathogens9010006. PMID: 31861540; PMCID: PMC7168646.

4) Iglewski BH. Pseudomonas. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 27. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8326/

5) Irini A Doytchinova and Darren R Flower. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine. 2007 25:856-866

6) Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001 Jan 19;305(3):567-80. doi: 10.1006/jmbi.2000.4315. PMID: 11152613.

7) Kung VL, Ozer EA, Hauser AR. The accessory genome of Pseudomonas aeruginosa. Microbiol Mol Biol Rev. 2010 Dec;74(4):621-41. doi: 10.1128/MMBR.00027-10. PMID: 21119020; PMCID: PMC3008168.

8) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. BMC Bioinformatics. Oct 31;8:424. 2007

9) Mena KD, Gerba CP. Risk assessment of Pseudomonas aeruginosa in water. Rev Environ Contam Toxicol. 2009;201:71-115. doi: 10.1007/978-1-4419-0032-6_3. PMID: 19484589.

10) Rice P., Bleasby A and Ison J. The EMBOSS Users Guide. Cambridge University Press

11) Sadikot RT, Blackwell TS, Christman JW, Prince AS. Pathogen-host interactions in Pseudomonas aeruginosa pneumonia. Am J Respir Crit Care Med. 2005 Jun 1;171(11):1209-23. doi: 10.1164/rccm.200408-1044SO. Epub 2005 Feb 1. PMID: 15695491; PMCID: PMC2718459.

12) Thompson, M.A. (2004) Molecular Docking Using ArgusLab, an Efficient Shape-Based Search Algorithm and the a Score Scoring Function. ACS Meeting, Philadelphia.