In Silico Identification of Non-cross-reactive Epitopes for Monkeypox Cell Surface-Binding Protein

Andrew Gao (andrewgao22@gmail.com)
Canyon Crest Academy
Sarah Gao
Canyon Crest Academy

Research Article

Keywords: monkeypox, epitope, pentapeptide, cross-reactivity, immunology, reverse vaccinology

Posted Date: May 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1693979/v1

Abstract

Smallpox, a disease caused by the Variola virus, was eradicated after global vaccination efforts. Recently, the monkeypox virus (MPV) has been spreading worldwide and is becoming a significant concern. The cell surface-binding protein E8L is found on MPV virion membranes and is essential for virion attachment to host cells, potentially making it an attractive vaccine target. In this study, E8L was investigated to assess its viability as a vaccine target and identify epitopes. BLAST analysis showed that the MPV Zaire-96-I-16 strain E8L amino acid sequence exhibits 38.2% percent identity with the human enzyme carbonic anhydrase III (CA3). Multiple sequence alignment and Expasy Translate were used to identify the corresponding E8L amino acid sequence of the May 2022 Massachusetts MPV case draft genome. The Massachusetts MPV E8L sequence was identical to the Zaire strain E8L sequence except for a single residue. Peptide Match software was used to identify pentapeptide subunits of E8L that are not found within the human proteome, thus addressing concerns about cross-reactivity with human proteins (such as CA3) and improving antigen specificity. Overall, 28 such pentapeptides were identified. The Immune Epitope Database was queried for experimentally-characterized epitopes/antigens containing these 28 pentapeptides, yielding 21 matching epitopes, demonstrating that some of these pentapeptides can evoke immune responses. This study identified human-foreign pentapeptides within E8L which could be explored to find viable epitopes for MPV vaccines. While existing vaccines confer protection against MPV, the risk of vaccine escape (as seen with SARS-CoV-2) suggests that additional vaccines should be investigated.

Introduction

The COVID-19 pandemic highlighted the need for rapid development of effective vaccines. Today, while research is still highly preliminary, the monkeypox virus (MPV) could potentially become a major public health concern in the United States and worldwide. MPV is a zoonotic virus belonging to the poxviruses family and Orthopoxvirus genus [1]. Smallpox and JYNNEOS vaccines confer protection against MPV but vaccine research should still continue because of vaccine escape. Also, because smallpox was eradicated in 1980, many younger people were never vaccinated against smallpox and thus lack protection. MPV is endemic to western and central Africa but cases have been identified recently in the United States, Portugal, and more countries. The World Health Organization has published an article titled “Multi-country monkeypox outbreak in non-endemic countries” describing the evolving situation [2]. The article states that as of May 21, 13:00, “92 laboratory confirmed cases” of MPV “have been reported to WHO” from twelve countries that MPV is not endemic in. So far, these cases have primarily been identified in North America and Western Europe. The WHO article also reports that the cases “confirmed by PCR” have been identified to come from the West African clade of MPV.

Grant et al used mathematical modeling to conclude that “in a population with diminishing herd immunity against orthopoxvirus species, the epidemic potential of monkeypox will continue increasing” [3]. The authors suggest that increases in MPV transmission could be attributed to the “cessation of smallpox vaccination” since the 1980s.

In order to contribute to the rapid design of a vaccine for MPV if it becomes necessary, we used bioinformatics methods to identify potential pentapeptides for vaccine epitopes. Specifically, we searched for pentapeptides that are absent in the human proteome in order to minimize cross-reactivity and side effects as well as optimize the antigen specificity.

We focused on cell surface-binding protein E8L because it is well-characterized, found on the virion membrane, and important for MPV infection [4]. Previously, one study has applied this method of pentapeptide searching to identify epitopes for SARS-Cov-2 vaccines [5]. Lucchese et al reported that “short peptides that are foreign to the host immune system have been experimentally validated not only as positive immunomodulants (i.e., adjuvants) in conjunction to vaccines, but are also evidenced as providing direct protection against lethal viral infections, at least in animal models.” Patel et al found that human-foreign pentapeptides can “enhance antigen specific immune responses and adjuvant vaccines” [6]. The small length of pentapeptides (only five amino acids) could be helpful for mRNA vaccines, for which size is an issue for translation rate and efficiency of encapsulation [7].

Methods

Uniprot Sequence Retrieval

The search query “monkeypox virus cell surface binding protein” was submitted to the online Uniprot database. The results were filtered to show reviewed entries only. Entry Q8V4Y0, cell surface-binding protein from gene E8L, was selected [4]. The amino acid FASTA sequence was retrieved. The protein was isolated from strain Zaire-96-I-16 of the monkeypox virus.

BLAST and Multiple Sequence Alignment

The Zaire strain’s E8L amino acid sequence was submitted to NCBI BLAST with default settings to search for highly similar proteins [8]. BLAST results were sorted by BLAST score. Percent identities for the selected BLAST hits ranged from 47.4–95.4%. The top twelve entries (including the Zaire strain’s sequence) were submitted to Clustal Omega Multiple Sequence Alignment software to generate an alignment for the sequences [9]. Perfectly conserved contiguous regions were identified, the longest being DDYGSNHL.

Draft Genome Sequence Translation And Analysis

The complete draft genome of the monkeypox virus was sequenced and uploaded to the NCBI GenBank database on May 21, 2022, by Gigante et al [10]. It was obtained from a confirmed monkeypox case in Massachusetts, USA. There are 197,124 base pairs and the genome was sequenced using the Oxford Nanopore device. The ID is ON563414.

Link to the genome: https://www.ncbi.nlm.nih.gov/nuccore/ON563414 [10]

The FASTA sequence of the genome was obtained and submitted to the Expasy Translate tool to convert the DNA sequence to the protein sequence [11]. Expasy Translate automatically identified and mapped all open reading frames. Due to it being extremely new, the FASTA sequence has not yet been annotated.

In order to find the region of DNA encoding the cell surface-binding protein E8L, the protein sequence was searched for DDYGSNHL, the oligopeptide found to be identical across all ten cell surface-binding proteins from the multiple sequence alignment.

Antigenicity Prediction

The VaxiJen 2.0 software was used to predict the antigenicity of the Massachusetts E8L based on the FASTA sequence [12]. VaxiJen has over 80% accuracy at predicting antigenicity. The default threshold for antigenicity is set at 0.4. VaxiJen predictions are based on auto-cross covariance transformation of input sequences into amino acid vectors.

Cross-Reactivity Search Against Human Proteome

Peptide Match is a software from the Protein Information Resource that searches proteome databases for matches with provided protein or peptide sequences [13]. First, Python code was written in Google Colab to split the cell surface-binding protein E8L sequence into overlapping k-mers of length five, also known as pentapeptides. The list of k-mers was submitted to Peptide Match with the settings of (1) restricted to homo sapiens proteome (2) UniProt/SwissProt only (3) include isoforms (4) not only UniRef100 representative sequences only, and (5) treat leucine and isoleucine as different.

IEDB Search

Pentapeptides not found in the human proteome were recorded. The Immune Epitope Database (IEDB) was searched for antigens and epitopes containing these pentapeptides [14]. The linear peptide search function was used with the substring setting. The assays to include were T cell, B cell, and MHC Ligand. Any host or disease or MHC class was considered.

Results And Discussion

The first objective was to identify the amino acid sequence of ON563414's E8L protein which could be different from other monkeypox strains. We aimed to identify a highly conserved amino acid motif across monkeypox E8L proteins that could be used to locate ON563414 E8L. We started with an E8L amino acid sequence (Q8V4Y0) from the strain Zaire-96-I-16 which is well-annotated and documented on the Uniprot database.

>sp|Q8V4Y0|CAHH_MONPZ Cell surface-binding protein OS=Monkeypox virus (strain Zaire-96-I-16) OX=619591 GN=E8L PE=2 SV=1

MPQQLSPINIETKKAISDTRLKTLDIHYNESKPTTIQNTGKLVRINFKGGYISGGFLPNEYVLSTIHI

YWGKEDDYGSNHLIDVYKYSGEINLVHWNKKKYSSYEEAKKHDDGIIIIAIFLQVSDHKNVYFQ

KIVNQLDSIRSANMSAPFDSVFYLDNLLPSTLDYFTYLGTTINHSADAAWIIFPTPINIHSDQLSKF

RTLLSSSNHEGKPHYITENYRNPYKLNDDTQVYYSGEIIRAATTSPVRENYFMKWLSDLREACFS

YYQKYIEGNKTFAIIAIVFVFILTAILFLMSQRYSREKQN

Table 1: Uniprot entry sequence for Q8V4Y0 cell surface-binding protein [4]

There were 250 BLAST hits for Q8V4Y0, 19 reviewed and 231 unreviewed. The worst hit had 38.4% percent identity while the best (excluding Q8V4Y0 itself) had 95.4% (rabbitpox virus, Utrecht strain).

Most of the top hits originated from the Vaccinia virus, but rabbitpox, taterapox, camelpox, and other poxviruses scored highly as well (Fig. 1). We noticed many BLAST hits were labeled as “carbonic anhydrase homolog”. Additionally, the Homo sapiens (human) enzyme carbonic anhydrase 3 and 13 (CA3, CA13) were hits, with percent identity of 38.2% and 37.7% respectively, suggesting that CA3 and CA13 are homologous with E8L. This homology has significant implications for vaccine and drug development. Researchers should be careful to avoid off-target effects and cross-reactivity with an important human enzyme.

Sequence alignment of CA3 and E8L showed significant homology, with 38.2% identity, a score of 388, 56.6% positives, and a low E-value of 6.1e-42 (Fig. 2).

The Clustal Omega multiple sequence alignment of the top twelve cell surface-binding protein sequences identified several perfectly conserved sequences, the longest being DDYGSNHL (Fig. 3). The amino acids around DDYGSNHL are also very similar among viruses. It appears that the cell surface-binding protein is well-conserved in poxviruses and is perhaps slower to mutate.

Given the perfect conservation of DDYGSNHL within twelve poxviruses, it was hypothesized that it would also be identical in the Massachusetts virus. So, in order to locate the position and sequence of E8L in the Massachusetts monkeypox virus genome, a text string search was performed within Expasy Translate for DDYGSNHL. The Massachusetts genome had been translated with Expasy to amino acids. DDYGSNHL was successfully located within an open reading frame.

The open reading frame (Fig. 4), highlighted in red, beginning with MPQQ and ending with EKQN) was compared to the Zaire strain cell surface-binding protein sequence. The sequences were perfectly identical except for a single residue. The nineteenth amino acid in the Zaire strain is threonine but it is alanine in the Massachusetts sequence.

> VIRT-11693:3'5' Frame 1, start_pos=31791

MPQQLSPINIETKKAISDARLKTLDIHYNESKPTTIQNTGKLVRINFKGGYISGGFLPNEYVLSTI

HIYWGKEDDYGSNHLIDVYKYSGEINLVHWNKKKYSSYEEAKKHDDGIIIIAIFLQVSDHKNV

YFQKIVNQLDSIRSANMSAPFDSVFYLDNLLPSTLDYFTYLGTTINHSADAAWIIFPTPINIHSD

QLSKFRTLLSSSNHEGKPHYITENYRNPYKLNDDTQVYYSGEIIRAATTSPVRENYFMKWLSD

LREACFSYYQKYIEGNKTFAIIAIVFVFILTAILFLMSQRYSREKQN

Table 2: E8L amino acid sequence for Massachusetts virus

VaxiJen 2.0 predicted Massachusetts E8L to be antigenic with a score of 0.5316, well over the standard threshold of 0.4. Next, the protein sequence was divided into all possible pentapeptides, or 5-mers, using Python. Three hundred pentapeptides were obtained.

PeptideMatch software searched for these three hundred pentapeptides in the UniProt and SwissProt databases' human proteomes. For 272 pentapeptides, matches were found. There were 28 unique pentapeptides for which no match in the human proteome was found. The most promiscuous pentapeptide was AATTS for which 63 matching proteins were identified. Some of the pentapeptides were overlapping and could be combined to form a longer oligopeptide. For instance, IHIYW, HIYWG, and IYWGK are part of a longer oligopeptide IHIYWGK that is also not found in any human proteome.

Table 3
28 unmatched pentapeptides not found in any human proteins

Unmatched Pentapeptides			
PINIE	IYWGK	VYFQK	NYRNP
DIHYN	VYKYS	NMSAP	RENYF
IHYNE	INLVH	TLDYF	NYFMK
GGYIS	KKHDD	DYFTY	YFMKW
PNEYV	HDDGI	WIIFP	FMKWL
IHIYW	SDHKN	PTPIN	MKWLS
HIYWG	HKNVY	PINIH	YIEGN

Table 4
4 longer oligopeptides formed by linking consecutive unmatched pentapeptides

Longer Unmatched Oligopeptides	Length
IHIYWGK	7
DIHYNE	6
SDHKNVY	7
NYFMKWLS	8

A search of the IEDB found 21 epitopes containing at least one of the 28 pentapeptides. They came from ectromelia mousepox, vaccinia virus, Clostridium perfringens, and other pathogens. These epitopes are experimentally demonstrated to have immunogenic potential. Interestingly, two human epitopes appeared in the results. Supposedly, all human proteins should have been filtered out by the previous steps. It is possible that the human epitopes were not recorded in the UniProt and SwissProt databases.

Table 5

IEDB access IDs, epitopes sequence, and organism for the epitopes containing at least one of the 28 unique pentapeptides that are foreign to humans. The pentapeptides within the epitopes are bolded and highlighted for visibility.

IEDB ID	Pentapeptide (or longer)	Epitope	Organism
874264	IHYNE	IHYNESKPTTIQNTG	Ectromelia virus (Ectromelia mousepox virus)
1400881	GGYIS	KGGYISLNYL	Mus musculus (mouse)
62863	PNEYV	TAGPNEYVYYKVYATYRKYQ	Clostridium perfringens
102955	PNEYV	YISGGFLPNEYVLSSLHIYW	Vaccinia virus (vaccinia virus VV)
102972	HIYWGK	YVLSSLHIYWGKEDDYGSNH	Vaccinia virus (vaccinia virus VV)
145659	HIYWGK	YVLSSLHIYWGKE	Vaccinia virus (vaccinia virus VV)
102535	INLVH	INLVHWNKKKYSSYEEAKKH	Vaccinia virus (vaccinia virus VV)
80628	KKHDD	CDLFKKHDDAIVRLR	Argentinian mammarenavirus (Junin arenavirus)
85490	KKHDD	LLNLLCDLFKKHDDA	Argentinian mammarenavirus (Junin arenavirus)
102627	SDHKNVY	LQVSDHKNVYFQKIVNQLDS	Vaccinia virus (vaccinia virus VV)
112306	SDHKN	SDHKNYL	Homo sapiens (human)
112421	SDHKN	YSSDHKN	Homo sapiens (human)
915942	VYFQK	IRVYFQKL	Mus musculus (mouse)
102627	VYFQK	LQVSDHKNVYFQKIVNQLDS	Vaccinia virus (vaccinia virus VV)
102371	TLDYF	DSVFYLDNLLPSTLDYFTYL	Vaccinia virus (vaccinia virus VV)
102371	DYFTY	DSVFYLDNLLPSTLDYFTYL	Vaccinia virus (vaccinia virus VV)
49730	PTPIN	PTPINNEKDI	Plasmodium knowlesi
49731	PTPIN	PTPINNEKDII	Plasmodium knowlesi
80445	RENYF	ARENYFMRW	Vaccinia virus (vaccinia virus VV)
102323	RENYF	ATTSPARENYFMRWLSDLRE	Vaccinia virus (vaccinia virus VV)
146581	RENYF	TSPARENYF	Vaccinia virus (vaccinia virus VV)

Conclusion

Monkeypox virus is endemic in West and Central Africa. However, many cases have been reported recently in non-endemic countries such as the United States of America and Portugal. It could be a cause for concern if it has mutated to become more infectious. Much more research is needed. In the present study, the E8L sequence was extracted from the May 2022 Massachusetts complete draft genome and was found to be almost identical to the Zaire strain E8L. E8L appeared to be strongly conserved among poxviruses and it was identified by VaxiJen to be antigenic. 28 pentapeptides from E8L that are not found in the human proteome were identified. These pentapeptides could be used to find epitopes for peptide vaccines for MPV that exhibit low cross-reactivity with human proteins and have high antigen specificity. A concern we identified was E8L's homology with carbonic anhydrase [15]. However, because the 28 pentapeptides are not found in the human proteome, this is less of a concern. Cell surface-binding protein E8L should be further investigated as a vaccine target, partly due to its integral role in MPV binding to host cells for invasion and also the degree of conservation we identified through BLAST and multiple sequence alignment. In the related poxvirus Vaccinia, the cell surface-binding protein has been identified as antigenic with "predicted antigenic sites located near amino acids 108–110 (CA domain) and 298–299 (transmembrane domain)" [15]. The protein binds to chondroitin sulfate, a surface glycosaminoglycan found on host cells. E8L's homology with eukaryotic carbonic anhydrases could point to monkeypox viruses' ability to acquire genes from host organisms through horizontal gene transfer. Indeed, analysis has revealed several independent horizontal gene transfer events of host genes into poxviruses [16]. While poxviruses, as DNA viruses, mutate relatively slowly, they should not be underestimated because of their propensity for horizontal gene transfer. Fixsen et al found that poxviruses can "capture host genes by LINE-1 retrotransposition" [17]. Future steps include obtaining a 3D structure model of E8L, assessing the accessibility of the pentapeptide, and evaluating the allergenicity and immunogenic potential of epitopes. Overall, the monkeypox virus should be carefully monitored and precautions should be taken. It could be wise to invest in vaccine and drug development.

Declarations

Conflicts of Interest

The authors have no conflicts of interest to report.

References

1. Arita, I., et al. "Human Monkeypox: A Newly Emerged Orthopoxvirus Zoonosis in the Tropical Rain Forests of Africa." The American Journal of Tropical Medicine and Hygiene, vol. 34, no. 4, 1 July 1985, pp. 781–789, 10.4269/ajtmh.1985.34.781. Accessed 19 Mar. 2022.
2. World Health Organization (21 May 2022). Disease Outbreak News; Multi-country monkeypox outbreak in non-endemic countries. Available at: https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON385

3. Grant, Rebecca, et al. “Modelling Human-To-Human Transmission of Monkeypox.” Bulletin of the World Health Organization, vol. 98, no. 9, 8 July 2020, pp. 638–640, 10.2471/blt.19.242347. Accessed 21 Jan. 2021.
4. “E8L - Cell Surface-Binding Protein - Monkeypox Virus (Strain Zaire-96-I-16) (MPX) - E8L Gene & Protein.” Www.uniprot.org, www.uniprot.org/uniprot/Q8V4Y0. Accessed 24 May 2022.
5. Lucchese, Guglielmo. “Epitopes for a 2019-NCoV Vaccine.” Cellular & Molecular Immunology, vol. 17, no. 5, 24 Feb. 2020, pp. 539–540, 10.1038/s41423-020-0377-z. Accessed 15 June 2020.
6. Patel, Ami, et al. “Pentamers Not Found in the Universal Proteome Can Enhance Antigen Specific Immune Responses and Adjuvant Vaccines.” PLoS ONE, vol. 7, no. 8, 24 Aug. 2012, p. e43802, 10.1371/journal.pone.0043802. Accessed 23 Feb. 2021.
7. Lucchese, Guglielmo, et al. “An Epitope Platform for Safe and Effective HTLV-1-Immunization: Potential Applications for MRNA and Peptide-Based Vaccines.” Viruses, vol. 13, no. 8, 1 Aug. 2021, p. 1461, www.mdpi.com/1999-4915/13/8/1461/htm, 10.3390/v13081461. Accessed 24 May 2022.
8. Altschul, Stephen F., et al. “Basic Local Alignment Search Tool.” Journal of Molecular Biology, vol. 215, no. 3, Oct. 1990, pp. 403–410, 10.1016/s0022-2836(05)80360-2.
9. Madeira, Fábio, et al. “Search and Sequence Analysis Tools Services from EMBL-EBI in 2022.” Nucleic Acids Research, 12 Apr. 2022, 10.1093/nar/gkac240.
10. Gigante, C.M.. et al. “Monkeypox Virus Isolate MPXV_USA_2022_MA001, Complete Genome.” NCBI Nucleotide, 21 May 2022, www.ncbi.nlm.nih.gov/nuccore/ON563414. Accessed 23 May 2022.
11. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31:3784-3788(2003).
12. Irini A Doytchinova and Darren R Flower. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007 8:4. http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html
13. Chuming Chen, Zhiwen Li, Hongzhan Huang, Baris E. Suzek, Cathy H. Wu, UniProt Consortium, A fast Peptide Match service for UniProt Knowledgebase, Bioinformatics, Volume 29, Issue 21, 1 November 2013, Pages 2808–2809, https://doi.org/10.1093/bioinformatics/btt484
14. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 2019 Jan 8;47(D1):D339-D343. doi: 10.1093/nar/gky1006. PMID: 30357391
15. Ohradanova, Anna, et al. “Reconstitution of Carbonic Anhydrase Activity of the Cell-Surface-Binding Protein of Vaccinia Virus.” Biochemical Journal, vol. 407, no. 1, 12 Sept. 2007, pp. 61–67, www.ncbi.nlm.nih.gov/pmc/articles/PMC2267410/, 10.1042/bj20070816. Accessed 24 May 2022.
16. Bratke, Kirsten A, and Aoife McLysaght. “Identification of Multiple Independent Horizontal Gene Transfers into Poxviruses Using a Comparative Genomics Approach.” BMC Evolutionary Biology, vol. 8, no. 1, 2008, p. 67, 10.1186/1471-2148-8-67. Accessed 24 Mar. 2022.
17. Fixsen, Sarah M., et al. Poxviruses Capture Host Genes by LINE-1 Retrotransposition. 27 Oct. 2020, 10.1101/2020.10.26.356402. Accessed 24 May 2022.

Figures

Figure 1

Top BLAST results for Q8V4Y0. Entry IDs, protein names, and percent identity are listed. A higher percent identity means that more of the amino acids match.

Figure 2

Sequence alignment of Homo sapiens CA3 and Zaire strain monkeypox E8L shows significant homology. The plus sign (+) indicates that the two amino acids are similar but not the same. An empty space indicates that the two amino acids are not similar. A dash (-) indicates that there is a gap in the alignment.

Figure 3

Portion of Clustal Omega Multiple Sequence Alignment showing perfect conservation of amino acid sequence “DDYGSNHL” (highlighted in orange) across twelve poxviruses. Asterisks indicate perfectly conserved residues. Colons indicate moderately strong conservation between amino acids with similar properties. Periods indicate weak conservation between amino acids with slightly similar properties.

Figure 4

The DDYSNHL “motif” within an open reading frame in the translated amino acid sequence of the Massachusetts monkeypox virus genome. The interface is Expasy Translate.