Nucleic Acid Sequence Composition of the Oxford – AstraZeneca Vaccine ChAdOx1 nCoV-19 (AZD1222, Vaxzevria)

Marco T. Radukic  
Bielefeld University

Dinh To Le  
Bielefeld University

Kristian M. Müller  
kristian@syntbio.net  
Bielefeld University

Research Article

Keywords: COVID-19, SARS-CoV-2, Adenovirus, Gene Therapy

DOI: https://doi.org/10.21203/rs.3.rs-799338/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

The vaccine ChAdOx1 nCoV-19 has been widely used, but its purity has been disputed because of rare side effects. We used Nanopore sequencing to assess the sequence and genetic purity of a vaccine dose. As we were lacking a reference sequence for the antigen cassette, we provide the obtained annotated sequence of the full vector to aid further studies on this topic. Our sample adhered to the published data, was highly pure (>99.97%), and no copy of the E1 gene as a predictor on replication-competent escape mutants was found.

Introduction

Vaccination with ChAdOx1 nCoV-19 against COVID-19 is safe and effective with millions of people around the world already immunized. Very rarely however, a serious medical condition referred to as vaccine-induced immune thrombotic thrombocytopenia (VITT) is observed after the first inoculation (1–3), and even more rarely, for the Ad26.COV2.S vector vaccine by Johnson & Johnson (4). Research into the causes is ongoing. Preliminary findings hint that residual host cell protein in the vaccine solution, and the physical vaccine structure may contribute (5, 6). Secretion of soluble spike protein from incorrectly spliced spike transcripts was also recently proposed as an adverse factor in VITT development (7). Although the general design of ChAdOx1 nCoV-19 is described (8, 9), prediction of splicing, or development of other sequence-based assays like PCR distribution studies (10), or studies on rare integration events (11), requires exact sequence information, which we provide here by nanopore sequencing and curated assembly.

The sequence of the underlying ChAdOx1 vector has been disclosed in patent EP3321367A2 / US20150044766A1 (patent Seq. ID 38 and 40). Likely, publication of specific ChAdOx1 nCoV-19 patent applications will further disclose sequences of the expressed transgene (e.g. UK patent application no. GB2003670.3 as mentioned in a preprint conflict of interest statement by Fischer et al. (12)). However, to our knowledge the latter documents have not yet been released to the public.

Previously, we described that direct nanopore sequencing also reveals information on sequence and quantity of DNA impurities in viral vectors (13). Thus sequencing of the final product adds to the recently released transcriptomic (14) and proteomic (5) characterizations of the vaccine.

Methods

ChAdOx1 nCoV-19 vaccine solution was retrieved from expired remainders in a vaccine vial (Vaxzevria lot 210163). Adenovirus vector was precipitated from the bulk at 40% ammonium sulfate saturation (242.3 g/l bulk), 2 h incubation at room temperature and centrifuging 15 min, 1614 rcf (RT) as previously described (15). The pellet was then resuspended in 1/20 the original volume with Tris-buffered saline (50 mM Tris, 150 mM NaCl, 2 mM MgCl₂, pH 8.0) containing 0.02 U/µl of proteinase K (NEB) and incubated for 1 h, 50 °C and 10 min, 95 °C. Vector DNA was purified from this solution by a PCR cleanup kit.
(Macherey-Nagel) according to the manufacturer’s instructions, and eluate DNA concentration and purity were determined photometrically on a NanoDrop 2000c (Thermo).

200 ng of extracted DNA was used for nanopore sequencing on a MinION Mk1B device with a Flongle FLG-001 flow cell (R9.4.1 pore chemistry) and SQK-RBK004 library prep kit (Oxford Nanopore) according to the manufacturer’s instructions. Raw reads were basecalled with Guppy 5.0.7 with the super accuracy preset (Oxford Nanopore) and reads with a Q score >= 10 (by Guppy) and a length > 500 nt (by NanoFilt 2.8.0, 100 nt head crop) were analyzed further. The 500 nt cut-off exceeds a 200 nt cut-off suggested for contamination monitoring (16) and was chosen for technical reasons.

De-novo assembly of the genome was performed on 395 reads > 25 000 nt using Flye 2.8.3 (nano-raw preset) (17) and the assembly was polished with all reads using the medaka_consensus wrapper of Medaka 1.4.3 with model r941_min_sup_g507 and the -g option (Oxford Nanopore). Reads were re-mapped to the polished assembly or HuAd5 (GenBank AC_000008) E1 gene with minimap2 2.17 (18) with the map-ont preset and mapped to the human genome built hg38 with blastn 2.9.0+ (19) (E value cut-off of 10^-5). Adenoviral features were annotated to the polished assembly from ChAd Y25 genome (GenBank NC_017825) and HuAd5 genome with Geneious R9 software. Web based protein blast (blast.ncbi.nlm.nih.gov) was used to confirm the spike protein sequence. Expression cassette features were annotated with the PlasMapper application of Geneious R9 and by transferring web based BLASTn alignments.

**Results And Discussion**

DNA extraction from the vaccine solution yielded pure vector DNA at 170 ng/µl (OD260/280 of 1.89). This represents a theoretical DNA recovery rate of 32% when compared to the EMA product information that states 5 × 10^{10} viral particles per 0.5 ml dose (20) and our initial guess of 36 kb genome size.

30 hours of sequencing on a new flow cell (without multiplexing) yielded 44 219 basecalled reads (268 Mb) that passed the initial filtering. These reads had an average length of 6 078 bp. Assembly resulted in a single contig of 35 501 bp. Alignment of the assembly to chimpanzee adenovirus Y25 isolate (GenBank NC_017825) and human adenovirus 5 (GenBank AC_000008) confirmed the deletion of the E1 and E3 genes and the substitution of the Y25 E4 ORF 4, 6, 7 and 34K CDS regions with their human Ad5 counterparts, as described for the creation of the original ChAdOx1 vector system (8). These alignments were also used to transfer annotations to the final vector sequence where applicable, which is presented in the GenBank file format as **Supplementary Information S1** to this manuscript. Notably, compared to the Y25 isolate, 15 bp are missing from each terminus in our assembly, which is probably a limitation of the library preparation method used here. The sequence disclosed in the vector patent lists 5’ further 16 additional bases compared to Y25.

In the initial assembly, nine sites of nucleotide deletions were observed throughout the ChAdOx1 backbone (G8391, GG9587-8, G10102, C10369, C15647, C16997, C19751, C19974, G31290 compared to
supplied sequence). All were G or C, located in homopolymeric regions, and within coding sequences. They very likely originated from the biased error pattern of the utilized nanopore flow cell and are therefore probably not present in the vaccine. For clarity we decided to manually correct these deletions in the final annotated sequence as provided (insertions are marked in the GenBank file). No further deviations from the reference DNA sequences (and in case of spike, protein sequence) were found.

As expected, the expression cassette for the SARS-CoV-2 surface glycoprotein (also S or spike, GenBank QHD43416.1 was used as a reference) is inserted at the E1 site and consists of a wild-type CMV promoter and enhancer (nt 175 298 – 174 206 of GenBank MN920393.1), two repeats of TetO, a CMV intronic sequence (nt 174 211 – 173 227 of GenBank MN920393.1), HindIII and KpnI recognition sites, a tPA leader peptide sequence (GenBank E04506, a Kozak sequence, e.g. present in the wildtype-tPA, is notably absent), the codon-modified S protein coding sequence (amino acid 2 – 1273) and the polyadenylation signal of bovine growth hormone as described (9). The 3′-UTR of the spike gene harbors additional 78 bp that are not thoroughly described in recent literature. The sequence appears to be a multi-cloning site, including a bacteriophage SP6 promoter in reverse orientation to the transgene. This UTR sequence is identical (by web-based BLASTn) to several patented sequences for adenovirus and DNA vaccines by the University of Oxford and may be a remnant of the original direct DNA vaccine vector pTH, from which the expression cassette design for ChAdOx1 seems to originate (21).

We also note that the coding sequence for the spike protein as sequenced has only 97.8% pairwise identity (90 mismatches) with a DNA sequence obtained for the same amino acid sequence with the GeneArt sequence optimizer (ThermoFisher, analysis performed on August 2, 2021, homo sapiens setting), which was previously assumed to be identical (7). However, the algorithm of the GeneArt sequence optimizer might be subject to continual changes.

Of 44 219 reads for ChAdOx1 nCoV-19 that passed initial filtering, 44 205 reads (99.97%) mapped to the here provided assembly (average of 99.7% alignment block length per read length and average of 71% matched residues), giving the impression of a very homogenous payload. Two of the reads that mapped to the assembly also mapped to the human genome built hG38, but these hits were false positives due to sequencing artifacts of low complexity within the reads (as identified by manual inspection). Of the 14 unmapped reads, 13 gave a hit with the nucleotide database (by web-based megablast.). However, we note that occasional spurious reads due to our sample handling under non-cleanroom conditions cannot be ruled out.

Lastly, no read mapped to the human adenovirus 5 E1 gene, which would be necessary to form a replication-competent escape mutant. This finding corroborates previously published transcriptomics studies on this safety-aspect of the ChAdOx1 vector system (14).

**Conclusion**

Sequence-level assessment of payload homogeneity for viral gene transfer vectors is a relatively new area of research. Our sample proved to be rather unspectacular in this regard. Regulatory bodies may
accept small amounts of residual host cell DNA in vector preparations (up to 10 ng per dose for EMA (22)) and the vaccine sample undercuts this target in our analysis. Also, no other significant DNA contaminations were found. On the other hand, the exact sequence information for ChAdOx1 nCoV-19 is important for studies on several aspects of vaccine safety but will probably become easily accessible only through patent applications, which is a lengthy process. Given the pace and scale of the SARS-CoV-2 pandemic, high throughput sequencing, which is cheap and available (e.g., material for this experiment cost less than 100 €), can fill the gap when justified by public interest.

Due to the safety profile, work with the ChAdOx1 vector is classified as biosafety level 1 in many countries (23, 24) and the sequence may serve as a blueprint for what can be regarded as safe.

Declarations

Ethics statement. The vial taken for sequencing was first used for vaccination, but the remainder expired due to low demand. It was kindly provided to us by a medical doctor.

Data availability. Read data can be made available upon written request to the corresponding author.

Conflict of Interest. The authors declare no conflict of interest.

Acknowledgement. MTR was funded by grant No 0400293 from the federal state of North Rhine-Westphalia and the European Fund for Regional Development (EFRE.NRW).

References

1. Schultz,N.H., Sørvoll,I.H., Michelsen,A.E., Munthe,L.A., Lund-Johansen,F., Ahlen,M.T., Wiedmann,M., Aamodt,A.-H., Skattør,T.H., Tjønnfjord,G.E., et al. (2021) Thrombosis and Thrombocytopenia after ChAdOx1 nCoV-19 Vaccination. N. Engl. J. Med., 384, 2124–2130.

2. Greinacher,A., Thiele,T., Warkentin,T.E., Weiss,K., Kyrle,P.A. and Eichinger,S. (2021) Thrombotic Thrombocytopenia after ChAdOx1 nCov-19 Vaccination. N. Engl. J. Med., 384, 2092–2101.

3. Pharmacovigilance Risk Assessment Committee (PRAC) (2021) Signal assessment report on embolic and thrombotic events (SMQ) with COVID-19 Vaccine (ChAdOx1-S [recombinant]) – COVID-19 Vaccine AstraZeneca (Other viral vaccines). Eur. Med. Agency.

4. Sadoff,J., Davis,K. and Douoguih,M. (2021) Thrombotic Thrombocytopenia after Ad26.COV2.S Vaccination — Response from the Manufacturer. N. Engl. J. Med., 384, 1965–1966.

5. Michalik,S., Siegerist,F., Franzke,K., Schindler,M., Seifert,U. and Beer,M. (2021) Comparative analysis of ChAdOx1 nCoV-19 and Ad26.COV2.S SARS-CoV-2 vector vaccines. researchsquare, 10.21203/rs.3.rs-736157/v1.
6. Baker, A.T., Boyd, R.J., Sarkar, D., Vant, J., Crespo, A.T., Truong, C.D., Bates, E., Wilson, E., Chan, C.K., Lipka-Lloyd, M., et al. (2021) The Structure of ChAdOx1/AZD-1222 Reveals Interactions with CAR and PF4 with Implications for Vaccine-induced Immune Thrombotic Thrombocytopenia. bioRxiv, 10.1101/2021.05.19.444882.

7. Kowarz, E., Krutzke, L., Reis, J., Bracharz, S., Kochanek, S. and Marschalek, R. (2021) Vaccine-Induced Covid-19 Mimicry Syndrome: Splice reactions within the SARS-CoV-2 Spike open reading frame result in Spike protein variants that may cause thromboembolic events in patients immunized with vector-based vaccines. researchsquare, 10.21203/rs.3.rs-558954/v1.

8. Dicks, M.D.J., Spencer, A.J., Edwards, N.J., Wadell, G., Bojang, K., Gilbert, S.C., Hill, A.V.S. and Cottingham, M.G. (2012) A novel chimpanzee adenovirus vector with low human seroprevalence: improved systems for vector derivation and comparative immunogenicity. PLoS One, 7, e40385.

9. van Doremalen, N., Lambe, T., Spencer, A., Belij-Rammerstorfer, S., Purushotham, J.N., Port, J.R., Avanzato, V.A., Bushmaker, T., Flaxman, A., Ulaszewska, M., et al. (2020) ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. Nature, 586, 578–582.

10. Badbaran, A., Mailer, R., Dahlke, C., Woens, J., Fathi, A., Mellinghoff, S.C., Renné, T., Addo, M.M., Riecken, K. and Fehse, B. (2021) Digital PCR to quantify ChAdOx1 nCoV-19 copies in blood and tissues. bioRxiv, 10.1101/2021.05.28.446155.

11. Stephen, S.L., Montini, E., Sivanandam, V.G., Al-Dhalimy, M., Kestler, H.A., Finegold, M., Grompe, M. and Kochanek, S. (2010) Chromosomal integration of adenoviral vector DNA in vivo. J. Virol., 84, 9987–94.

12. Fischer, R.J., van Doremalen, N., Adney, D.R., Yinda, C.K., Port, J.R., Holbrook, M.G., Schulz, J.E., Williamson, B.N., Thomas, T., Barbian, K., et al. (2021) ChAdOx1 nCoV-19 (AZD1222) protects hamsters against SARS-CoV-2 B.1.351 and B.1.1.7 disease. bioRxiv Prepr. Serv. Biol., 10.1101/2021.03.11.435000.

13. Radukic, M.T., Brandt, D., Haak, M., Müller, K.M. and Kalinowski, J. (2020) Nanopore sequencing of native adeno-associated virus (AAV) single-stranded DNA using a transposase-based rapid protocol. NAR genomics Bioinforma., 2, iqaa074.

14. Almuqrin, A., Davidson, A.D., Williamson, M.K., Lewis, P.A., Heesom, K.J., Morris, S., Gilbert, S.C. and Matthews, D.A. (2021) SARS-CoV-2 vaccine ChAdOx1 nCoV-19 infection of human cell lines reveals low levels of viral backbone gene transcription alongside very high levels of SARS-CoV-2 S glycoprotein gene transcription. Genome Med., 13, 43.

15. Schagen, F.H.E., Rademaker, H.J., Rabelink, M.J.W.E., van Ormondt, H., Fallaux, F.J., van der Eb, A.J. and Hoeben, R.C. (2000) Ammonium sulphate precipitation of recombinant adenovirus from culture medium: an easy method to increase the total virus yield. Gene Ther., 7, 1570–1574.
16. U.S. Food and Drug Administration (2010) Guidance for Industry; Cell Characterization and Qualification of Cell Substrates and Other Biological Materials Used in the Production of Viral Vaccines for Infectious Disease Indications.

17. Kolmogorov,M., Yuan,J., Lin,Y. and Pevzner,P.A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.*, 37, 540–546.

18. Li,H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 3094–3100.

19. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, 215, 403–10.

20. European Medicines Agency (2021) COVID-19 Vaccine AstraZeneca - Product Information as approved by the CHMP on 29 January 2021, pending endorsement by the European Commission.

21. Hanke,T., Schneider,J., Gilbert,S.C., Hill,A. V and McMichael,A. (1998) DNA multi-CTL epitope vaccines for HIV and Plasmodium falciparum: immunogenicity in mice. *Vaccine*, 16, 426–35.

22. European Directorate for the Quality of Medicines & Healthcare EDQM (2020) EUROPEAN PHARMACOPOEIA: Free access to supportive pharmacopoeial texts in the field of vaccines for human use during the coronavirus disease (COVID-19) pandemic.

23. Zentrale Kommission für die Biologische Sicherheit (2020) Empfehlung der ZKBS zur Risikobewertung des rekombinanten Adenovirus ChAdOx1 nCoV-19 gemäß § 5 Absatz 1 GenTSV, Az. 45242.0181.

24. AstraZeneca (2020) AZD1222 Safety Data Sheet, Australia, Ver 1.1, Rev. 02.12.2020, SDS No. 25943.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- ChAdOx1nCoV19assemblygenbank.txt