



CCT: a coordinate conversion tool for hepatitis B virus

Trevor Graham Bell, Mukhlid Yousif & Anna Kramvis

To cite this article: Trevor Graham Bell, Mukhlid Yousif & Anna Kramvis (2019): CCT: a coordinate conversion tool for hepatitis B virus, Southern African Journal of Infectious Diseases, DOI: [10.1080/23120053.2018.1558641](https://doi.org/10.1080/23120053.2018.1558641)

To link to this article: <https://doi.org/10.1080/23120053.2018.1558641>



© 2019 The Author(s). Co-published by NISC Pty (Ltd) and Informa UK Limited, trading as Taylor & Francis Group



Published online: 07 Jan 2019.



Submit your article to this journal [↗](#)



Article views: 109



View Crossmark data [↗](#)

CCT: a coordinate conversion tool for hepatitis B virus[‡]

Trevor Graham Bell^{*} , Mukhlid Yousif^S and Anna Kramvis 

Hepatitis Virus Diversity Research Unit, Department of Internal Medicine, University of the Witwatersrand, Johannesburg, South Africa

^SCurrent affiliation: Centre for HIV and STI: HIV Virology Section, National Institute for Communicable Diseases, National Health Laboratory Services; Department of Pathology, University of the Witwatersrand, Johannesburg, South Africa

^{*}Corresponding author, Email: TrevorGrahamBell@gmail.com



The hepatitis B virus (HBV) genome has ~3 200 nucleotides, coding for seven proteins in four overlapping open reading frames (ORFs). Comparison of genomic coordinates between different samples and/or published literature requires manual conversion. An online tool is presented to convert nucleotide or amino acid positions between ORFs, regions and domains of the HBV genome. The user enters a position into an interactive web page, which then shows this position in all other applicable ORFs, regions or domains and plots it on a diagrammatic representation of the HBV genome. This tool assists researchers to convert coordinates, thereby facilitating comparisons between samples.

Keywords: bioinformatics, genotypes, HBV, hepatitis B virus

Introduction

Hepatitis B virus (HBV), a member of the family *Hepadnaviridae*, infects the liver of humans, primates and other animals, and may cause liver disease or liver cancer. To date, nine human HBV genotypes (strains) have been described, which are geographically distributed globally and named genotype A to I, with a putative tenth genotype 'J' proposed.¹ The HBV genome, which is circular and varies in length by genotype from 3 182 to 3 248 nucleotides, codes for seven proteins in four partially or completely overlapping open reading frames (ORFs): *Core* gene (HBcAg and HBeAg proteins), *Surface* gene (PreS1, PreS2 and S proteins), *Polymerase* and *X*. The *Polymerase* region consists of four protein subdomains (terminal protein, spacer, reverse transcriptase (rt) and ribonuclease H). The PreS1/PreS2/S ORF is completely overlapped by the *Polymerase* gene, which also partially overlaps the 3' end of *Core* and the 5' end of *X*. Mutations in one ORF can therefore lead to mutations in the alternative ORF.² Genes in ORFs, which overlap, are not in-frame with each other.

Historically, the start of the *Core* gene was considered as position 1 of the HBV genome. However, the accepted numbering system now in use considers the first 'T' of the *EcoRI* restriction site ('GAATTC') as position 1. The start of the *Core* gene is at position 1 901 in the contemporary system.

Depending on the region and context, mutations or positions of interest in the genome may be referenced as a nucleotide position or an amino acid position relative to the ORF or protein sub-domain, or as an absolute nucleotide position relative to the entire genome. However, as multiple sequence alignments of nucleotide or amino acid data are typically sub-genomic and may not include an entire region, comparing or contextualising positions is often difficult. This frequently hinders comparisons, with different numbering systems being used at various times and by different investigators, and for the different regions of the genome. Therefore, the aim of this tool was to develop an algorithm to output standardised numbering for a position of interest from a genomic or sub-

genomic fragment, across all four ORFs and seven proteins, for each HBV genotype.

Materials and methods

We developed a free, online tool to assist with conversion of nucleotide or amino acid positions (coordinates) between ORFs, regions and domains of the HBV genome for genotypes A to I. The tool is an interactive web page, making use of client-side JavaScript in the user's browser to perform the calculations and to plot the query position on a diagrammatic representation of the HBV genome (see [Figure 1](#)). JavaScript was used as it allows calculations to be performed directly on the web page in the user's browser. No user data are transferred to the web server to perform calculations. The tool may be accessed from any web browser on desktop or mobile platforms.

The genotype length and the coordinates of all ORFs and domains, for each genotype, are coded into the tool using values from the literature.³ The appropriate values are then populated into the table when the genotype is selected on the web page. The genotype length is required to correctly calculate coordinates spanning the *EcoRI* position (position 1). Calculations are performed according to the settings the user has specified.

Results

The HBV Co-ordinate Conversion Tool is shown in [Figure 1](#). Availability of the tool and the source code are described under 'Availability' below.

Overview

Usage of the tool involves selecting a genotype and entering a query position, either as an amino acid position (relative to the ORF, region or domain), or as an absolute or relative nucleotide position. When the 'Calculate' button is pressed, the query position is converted to nucleotide and amino acid positions in all other overlapping ORFs. The 'codon offset' value

[‡]MY and TB conceived the tool. TB wrote the code and the web-page, maintains the server, and drafted the manuscript. AK is the principal investigator. All authors contributed to, read, and approved the final manuscript.

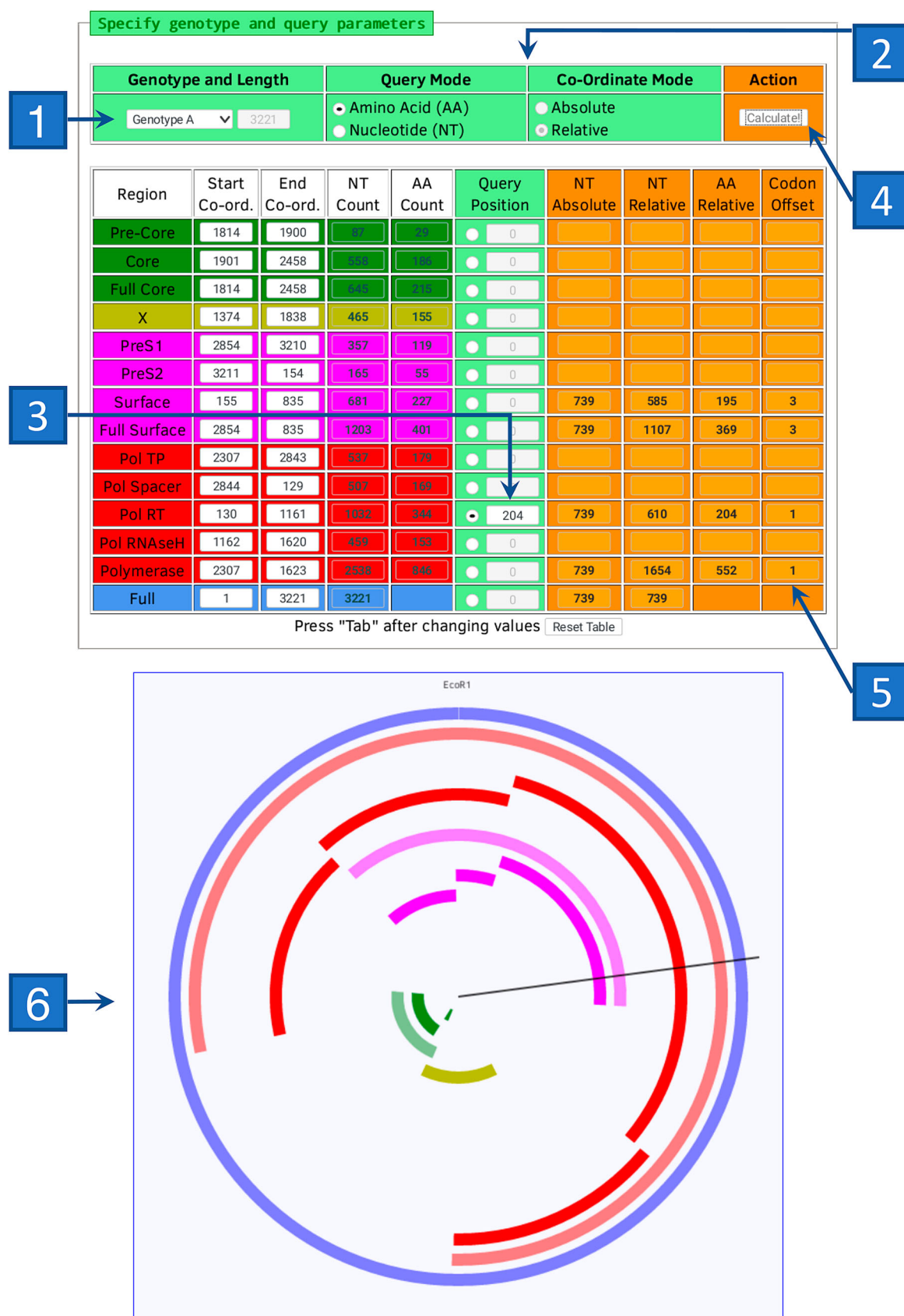


Figure 1: The coordinate conversion tool web page. Briefly, a genotype (flag 1) and modes (flag 2) are selected, the query position is input (flag 3) and the 'Calculate' button (flag 4) is pressed. Coordinate results (flag 5) and diagrammatic position (flag 6) are then shown. The 'query mode' indicates whether the query position entered represents an amino acid position or a nucleotide position. The 'co-ordinate mode' indicates whether the query position entered should be considered relative to position 1 of the genome (*EcoRI*) or relative to the relevant ORF. Amino acid positions are always considered relative to the respective ORF. The four orange columns on the right of the table show the query position as an absolute and relative nucleotide position, and a relative amino acid position. The position of the nucleotide within the codon (position 1, 2 or 3 of the codon) is shown in the 'Codon Offset' column. When an amino acid position is entered, this will be a 1 for the specified ORF, as the first nucleotide position of the amino acid is considered. ORFs and regions in the diagram are colour-coded to match the colours used in the input table above. See text for detailed description of the tool.

indicates the query position within a codon (1, 2 or 3) for the ORF, region or domain of the original query, and for all overlapping ORFs.

Detailed usage

The tool is illustrated in Figure 1 with flags to demonstrate typical usage, as described below.

- **Flag 1:** The HBV genotype (A to I) is selected. The length and coordinates in the table below will change according to the genotype selected.³
- **Flag 2:** The 'Query Mode' and 'Co-Ordinate Mode' are selected. These modes indicate how the query value (Flag 3) should be interpreted. Amino acid queries are always relative to the ORF, region or domain. Nucleotide queries may be relative to the ORF, region or domain, or absolute from *EcoRI* as position 1.
- **Flag 3:** The query amino acid or nucleotide position of interest is entered into the box next to the relevant ORF or region after clicking on the relevant radio button to activate the edit box. In this example, amino acid number 204 in the 'rt' domain of the Polymerase gene has been entered.
- **Flag 4:** The 'Calculate' button is pressed to convert the query to all other overlapping ORFs, regions and domains. An error message is displayed below this button when there are errors in any of the input parameters.
- **Flag 5:** The position of interest (amino acid number 204 in 'Pol RT' in this example) is converted to all other overlapping ORFs, regions and domains, with values shown in the orange areas of the table. 'Codon offset' indicates the codon position (either 1, 2 or 3) in the relevant ORF, region or domain. For amino acid queries, this will always be 1. For nucleotide queries, this will be the position of the nucleotide within the codon of the relevant amino acid.
- **Flag 6:** The query position is shown by the black line on the diagrammatic representation of the HBV genome. ORFs and regions are colour-coded to match the colours used in the input table. This output assists in visualising the position of interest (and other ORFs and regions) in the context of the complete HBV genome.

Additional usage

The tool may be used with genotypes that do not contain the typical number of nucleotides because of the presence of deletions and/or insertions ('indels'). This can be achieved by editing the length and region values in the table directly. If the genotype length is changed a warning dialogue is presented and the background colour of the length box is changed to indicate this. No sanity checks on values are performed by the tool when length and/or coordinate values are changed in this way, as this usage is intended to accommodate unusual or atypical genomes. Length and coordinate values in the table can be reset by selecting another genotype from the drop-down selector or by pressing the 'Reset Table' button below the table.

This tool requires that either the genotype from which the sample originates is known, or that the positions of all ORFs and domains of a sample of unknown genotype are known.

The tool cannot be used to genotype samples and does not make predictions regarding the position of ORFs.

Discussion and conclusion

This online tool will assist researchers to quickly and accurately convert nucleotide or amino acid coordinates of the genome between different ORFs, regions and domains for genotypes A to I of HBV. This will facilitate the analysis of mutations in the overlapping ORFs, which is important, especially in the era of bioinformatics and biomarker discovery. Often, mutations are numbered differently by various researchers and as a result are not recognised as the same mutation, leading to their exclusion from various analyses. For example, Lamivudine-related resistance mutations at amino acid rtL180M (previously amino acid 528, 526, 515, or 525) and rtM204 V/I (previously 552, 550, 539, or 549),⁴ are still being referred to using the old nomenclature. This new tool allows nucleotide and amino acid mutations to be referred to in standardised nomenclature for each genotype, thereby assisting comparisons. As it is web-based and available free of charge, it can be used from any platform or region of the world where Internet access is available.

Availability

The tool is available online at <http://hvdr.bioinf.wits.ac.za/cct> or <http://bit.ly/hbvcct> (case sensitive). The source code is available under the GNU General Public License, version 2, from *GitHub* at <https://github.com/DrTrevorBell/CCT>. Researchers working on other suitable genomes are encouraged to fork the project and adapt it accordingly. This paper should be cited when any research that makes use of the tool is published.

Future work

The following enhancements to the tool are planned:

- HBV coordinates from non-human primates and animals, sourced from the literature, will be included in the tool as additional 'genotypes' in the drop-down selector.
- The consensus nucleotide and amino acid for the query position will be included on the output page, for reference. This sequence data will be drawn from our curated multiple sequence alignments of each HBV genotype.⁵
- To assist with the comparison of positions between genotypes, a 'Universal Numbering System', which is under development, will be included.

Acknowledgements – TB would like to thank Dr Dieter Glebe for useful suggestions and feedback.

Disclosure statement – No potential conflict of interest was reported by the authors.

Funding – TB received funding from the National Research Foundation, South Africa (NRF) (Grant GUN #86215), the Deutsche Forschungsgemeinschaft (DFG) and the University Research Committee (URC). MY received funding from the DFG. AK received funding from the NRF (Grants GUN #65530 and GUN #93516) and the DFG.

ORCID

Trevor Graham Bell  <http://orcid.org/0000-0003-2630-7513>
 Anna Kramvis  <http://orcid.org/0000-0001-6006-3765>

References

1. Kramvis A. Genotypes and genetic variability of hepatitis B virus. *Intervirology*. 2014;57:141–50.
2. Mizokami M, Orito E, Ohba K-i, Ikeo K, Lau JYN, Gojobori T. Constrained evolution with respect to gene overlap of hepatitis B virus. *J. Mol. Evol.* 1997;44:S83–90.
3. Kramvis A, Kew M, François G. Hepatitis B virus genotypes. *Vaccine*. 2005;23:2409–23.
4. Stuyver LJ, Locarnini SA, Lok A, et al. Nomenclature for antiviral-resistant human hepatitis B virus mutations in the polymerase region. *Hepatology*. 2001;33:751–57.
5. Bell TG, Yousif M, Kramvis A. Bioinformatic curation and alignment of genotyped hepatitis B virus (HBV) sequence data from the GenBank public database. *Springerplus*. 2016;5(1):1896.

Received: 24-08-2018 Accepted: 10-12-2018