Sequence Alignment/Map Optional Fields Specification

The SAM/BAM Format Specification Working Group

7 Jan 2025

The master version of this document can be found at `https://github.com/samtools/hts-specs`.
This printing is version 7f4af4b from that repository, last modified on the date shown above.

This document is a companion to the *Sequence Alignment/Map Format Specification* that defines the SAM and BAM formats, and to the *CRAM Format Specification* that defines the CRAM format.[1] Alignment records in each of these formats may contain a number of optional fields, each labelled with a *tag* identifying that field's data. This document describes each of the predefined standard tags, and discusses conventions around creating new tags.

1 Standard tags

~~Predefined standard tags are listed in the following table and described in greater detail in later subsections. Optional fields are usually displayed as~~ All optional fields follow the `TAG:TYPE:VALUE` ~~; the~~ format where ~~type~~`TAG` ~~may be one~~ is a two-character string that matches `/[A-Za-z][A-Za-z0-9]/`. In an optional field, `TYPE` is a single case-sensitive letter which defines the format of ~~`A`~~`VALUE`~~(character), :~~

Type	Regexp matching ~~`B`~~`VALUE`	Description
A	`[!-~]`	Printable character
i	`[-+]?[0-9]+`	Signed integer[2]
f	`[-+]?[0-9]*\.?[0-9]+([eE][-+]?[0-9]+)?`	Single-precision floating number
Z	`[ !-~]*`	Printable string, including space
H	`([0-9A-F][0-9A-F])*`	Byte array in the Hex format[3]
B	`[cCsSiIf](,[-+]?[0-9]*\.?[0-9]+([eE][-+]?[0-9]+)?)*`	Integer or numeric array

For an integer or numeric array (type '~~`f`~~`B`~~(real number), `~~'), the first letter indicates the type of numbers in the following comma separated array. The letter can be one of '~~`H`~~`cCsSiIf`', corresponding to `int8_t` (~~hexadecimal array~~signed 8-bit integer), ~~`i`~~`uint8_t` (unsigned 8-bit integer), ~~or `Z`~~`int16_t`~~(string)~~, `uint16_t`, `int32_t`, `uint32_t` and `float`, respectively.[4] During import/export, the element type may be changed if the new type is also compatible with the array.

Predefined standard tags are listed in the following table and described in greater detail in later subsections.

Tag	Type	Description
`AM`	i	The smallest template-independent mapping quality in the template
`AS`	i	Alignment score generated by aligner
`BC`	Z	Barcode sequence identifying the sample
`BQ`	Z	Offset to base alignment quality (BAQ)

[1] See `SAMv1.pdf` and `CRAMv3.pdf` at `https://github.com/samtools/hts-specs`.

[2] The number of digits in an integer optional field is not explicitly limited in SAM. However, BAM can represent values in the range $[-2^{31}, 2^{32})$, so in practice this is the realistic range of values for SAM's 'i' as well.

[3] For example, the six-character Hex string '`1AE301`' represents the byte array `[0x1a,0xe3,0x1]`.

[4] Explicit typing eases format parsing and helps to reduce the file size when SAM is converted to BAM.