

2. CONCEPTS AND SPECIFICATIONS

2.2.7.4.14. Greek letters

(31) In general, the corresponding letter of the Latin alphabet, prefixed by a backslash character. The complete set is:

α	A	<code>\a</code>	<code>\A</code>	alpha	ν	N	<code>\n</code>	<code>\N</code>	nu
β	B	<code>\b</code>	<code>\B</code>	beta	o	O	<code>\o</code>	<code>\O</code>	omicron
χ	X	<code>\c</code>	<code>\C</code>	chi	π	Π	<code>\p</code>	<code>\P</code>	pi
δ	Δ	<code>\d</code>	<code>\D</code>	delta	θ	Θ	<code>\q</code>	<code>\Q</code>	theta
ϵ	E	<code>\e</code>	<code>\E</code>	epsilon	ρ	R	<code>\r</code>	<code>\R</code>	rho
φ	Φ	<code>\f</code>	<code>\F</code>	phi	σ	Σ	<code>\s</code>	<code>\S</code>	sigma
γ	Γ	<code>\g</code>	<code>\G</code>	gamma	τ	T	<code>\t</code>	<code>\T</code>	tau
η	H	<code>\h</code>	<code>\H</code>	eta	υ	U	<code>\u</code>	<code>\U</code>	upsilon
ι	I	<code>\i</code>	<code>\I</code>	iota	ω	Ω	<code>\w</code>	<code>\W</code>	omega
κ	K	<code>\k</code>	<code>\K</code>	kappa	ξ	Ξ	<code>\x</code>	<code>\X</code>	xi
λ	Λ	<code>\l</code>	<code>\L</code>	lambda	ψ	Ψ	<code>\y</code>	<code>\Y</code>	psi
μ	M	<code>\m</code>	<code>\M</code>	mu	ζ	Z	<code>\z</code>	<code>\Z</code>	zeta

2.2.7.4.15. Accented letters

(32) Accents should be indicated by using the following codes before the letter to be modified (*i.e.* use `\'e` for an acute e):

<code>\'</code>	acute (é)	<code>\"</code>	umlaut (ü)
<code>\=</code>	overbar or macron (ā)	<code>\`</code>	grave (à)
<code>\~</code>	tilde (ñ)	<code>\.</code>	overdot (ô)
<code>\^</code>	circumflex (â)	<code>\;</code>	ogonek (ų)
<code>\<</code>	hacek or caron (ǎ)	<code>\,</code>	cedilla (ç)
<code>\></code>	Hungarian umlaut or double accented (ő)	<code>\(</code>	breve (ô)

These codes will always be followed by an alphabetic character.

2.2.7.4.16. Other characters

(33) Other special alphabetic characters should be indicated as follows:

<code>\%a</code>	a-ring (å)	<code>\?i</code>	dotless i (ı)	<code>\&s</code>	German 'ss' (ß)
<code>\/o</code>	o-slash (ø)	<code>\/l</code>	Polish l (ł)	<code>\/d</code>	barred d (đ)

Capital letters may also be used in these codes, so an ångström symbol (Å) may be given as `\%A`.

(34) Superscripts and subscripts should be indicated by bracketing relevant characters with circumflex or tilde characters, thus:

superscripts	<code>Csp³</code>	for	<code>Csp³</code>
subscripts	<code>U_{eq}</code>	for	<code>U_{eq}</code>

The closing symbol is essential to return to normal text.

(35) Some other codes are accepted by convention. These are:

<code>\%</code>	degree (°)	<code>\\times</code>	×
<code>--</code>	dash	<code>+-</code>	±
<code>---</code>	single bond	<code>++</code>	≡
<code>\\db</code>	double bond	<code>\\square</code>	□
<code>\\tb</code>	triple bond	<code>\\neq</code>	≠
<code>\\ddb</code>	delocalized double bond	<code>\\rangle</code>	>
<code>\\sim</code>	~	<code>\\langle</code>	<
(Note: ~ is the code for subscript)		<code>\\rightarrow</code>	→
<code>\\simeq</code>	≈	<code>\\leftarrow</code>	←
<code>\\infty</code>	∞		

Note that `\\db`, `\\tb` and `\\ddb` should always be followed by a space, *e.g.* C=C is denoted by `c\\db c`.

2.2.7.4.17. Typographic style codes

(36) The codes indicated above are designed to refer to special characters not expressible within the CIF character set, and the initial specification did not permit markup for typographic style such as italic or bold-face type. However, in some cases the ability to indicate type style is useful, and in addition to the codes above HTML-like conventions are allowed of surrounding text by `<i>` `</i>` to indicate the beginning and end of italic, and by `` `` to indicate the beginning and end of bold-face type.

(37) If it is necessary to convey more complex typographic information than is permitted by these special character codes and conventions, the entire text field should be of a richer content type allowing detailed typographic markup.

References

- Bernstein, H. J. (2002). *Some comments on parsing for computer programming languages*. <http://www.bernstein-plus-sons.com/TMM/Parsing>.
- Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P. & Berners-Lee, T. (1999). *Hypertext Transfer Protocol – HTTP/1.1*. RFC 2616. Network Working Group. <http://www.w3.org/Protocols/rfc2616/rfc2616.html>.
- Freed, N. & Borenstein, N. (1996). *Multipurpose Internet Mail Extensions (MIME) Part two: media types*. RFC 2046. Network Working Group. <http://www.ietf.org/rfc/rfc2046.txt>.
- Hall, S. R. (1991). *The STAR File: a new format for electronic data transfer and archiving*. *J. Chem. Inf. Comput. Sci.* **31**, 326–333.
- Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *The Crystallographic Information File (CIF): a new standard archive file for crystallography*. *Acta Cryst.* **A47**, 655–685.
- Hall, S. R. & Cook, A. P. F. (1995). *STAR dictionary definition language: initial specification*. *J. Chem. Inf. Comput. Sci.* **35**, 819–825.
- Hall, S. R. & Spadaccini, N. (1994). *The STAR File: detailed specifications*. *J. Chem. Inf. Comput. Sci.* **34**, 505–508.
- McMahon, B., Westbrook, J. D. & Bernstein, H. J. (2000). *Report of the COMCIFS Working Group on Dictionary Maintenance*. <http://www.iucr.org/iucr-top/cif/spec/dictionaries/maintenance.html>.
- Ulrich, E. L. *et al.* (1998). XVIIth Intl Conf. Magn. Res. Biol. Systems. Tokyo, Japan.
- Westbrook, J. D. & Hall, S. R. (1995). *A dictionary description language for macromolecular structure*. <http://ndbserver.rutgers.edu/mmcif/ddl/>.