

The **TeXshade** package*

Typesetting
nucleotide and peptide alignments

Eric Beitz[†]

v1.8; 2004/08/26

Abstract

Setting alignments of nucleotides and peptides for publication or presentation purposes is usually a time consuming two-step process. First, a scientific software is used for the calculation of the alignment. This is done in a few minutes. Then, in order to highlight special sequence relationships and to label positions and regions of interest a second software with high quality output capability is needed. Manipulating sequence alignments with standard word processing or graphics programs takes its time—often several hours—and simple layout changes such as re-breaking lines, say from 50 to 40 residues per line, elongate the working time considerably.

TeXshade is an alignment shading software written in **TeX/L^ATeX** which can process multiple sequence alignments in the MSF, ALN and FASTA file format. It provides in addition to common shading algorithms special shading modes featuring functional aspects, e. g. charge or hydrophathy, and a plenitude of commands for handling shading colors, text styles, labels, legends and even allows the user to define completely new shading modes. **TeXshade** combines highest flexibility and the habitual **TeX** output quality—with reasonable time expenditure.

*Please cite: Eric Beitz (2000), **TeXshade**: shading and labeling multiple sequence alignments using **L^ATeX 2_ε**. *Bioinformatics*: **16**, 135–139.

[†]University of Tübingen, Pharmaceutical Chemistry, Morgenstelle 8, D-72076 Tübingen, Germany; send electronic mail to eric.beitz@uni-tuebingen.de; for further information, updates and on-line documentation see my homepage at <http://homepages.uni-tuebingen.de/beitz/>

Contents

1	Package Overview	4
1.1	Version History	4
1.2	L ^A T _E X basics	7
1.2.1	Typesetting documents using L ^A T _E X	7
1.2.2	Memory shortness when using T _E Xshade	8
1.3	System requirements	9
1.4	The <code>texshade</code> environment	10
1.5	Shading modes predefined in this package	11
1.5.1	Identity mode	11
1.5.2	Similarity mode	11
1.5.3	Diversity mode	12
1.5.4	Functionality modes	12
1.5.5	Secondary structures	15
1.5.6	Sequence fingerprints	18
1.6	Customization of alignment outputs	19
2	Format of alignment input files	20
2.1	The MSF file format	20
2.2	The ALN file format	22
2.3	The FASTA file format	24
3	Use of a T_EXshade parameter file	25
4	texshade user commands	26
4.1	Using predefined shading modes	26
4.2	Creating new functional shading modes	31
4.3	Appearance of the consensus line	32
4.4	Appearance of the sequence lines	34
4.4.1	Names, numbers and gaps	34
4.4.2	Hiding, killing, separating and ordering	37
4.4.3	Residues per line and further settings	38
4.4.4	Fingerprinting	39
4.5	Individual shading and labeling of sequence stretches	39
4.5.1	Manual shading of regions and blocks	40
4.5.2	Emphasizing regions and blocks	40
4.5.3	Graphical labeling of sequence features	41
4.5.4	Including secondary structure information	46
4.6	Displaying and building legends	49
4.7	Adding captions to the alignment	50

4.8	Font handling	50
4.8.1	Changing font styles	50
4.8.2	Using PostScript fonts	52
4.9	Goodies—molweight and charge	52
5	The PostScript color selection scheme	54
6	Listing of the texshade default settings	57
6.1	Standard definitions	57
6.2	Colors used in the different shading modes	57
7	Quick Reference	61
8	References	67

1 Package Overview

After `texshade.ins` is run through $\text{\TeX}$ the following files should appear in the directory:

<code>texshade.sty</code>	the style file with all $\text{\TeXshade}$ commands
<code>texshade.def</code>	an example parameter file with the standard parameter settings
<code>AQPDNA.MSF</code>	an example nucleotide alignment (MSF-format)
<code>AQPpro.MSF</code>	an example protein alignment (MSF-format)
<code>AQP2spec.ALN</code>	a further protein alignment (minimal ALN-file)
<code>AQP1.phd</code>	secondary structure information (PHD-format)
<code>AQP1.top</code>	topology data extracted from <code>AQP1.phd</code>
<code>standard.cod</code>	standard genetic code definitions
<code>ciliate.cod</code>	ciliate macronuclear genetic code

The alignment file examples as well as the topology data file are needed for $\text{\TeX}$ ing this documentation and can serve as illustrations for the MSF and ALN file format.

The following subsections give an overview of the capabilities of the $\text{\TeXshade}$ package. All commands are described in detail later on.

1.1 Version History

v1.8 2004/08/26

Corrections: Only minor bugs were fixed.

Introductions: (a) More colors have been designed, i.e. ‘light’ versions of the existing PostScript colors. (b) Three color ramps in 5% steps have been introduced: i) Blue-Red, ii) Green-Red and iii) Cold-Hot. (c) Two new feature label styles `bar` and `color` have been introduced which allow one to display number values as bar graphs or color scales along the alignment¹.

v1.7 2004/01/05

Corrections: Several bugs were fixed. In gaps the wrong character was plotted in ‘donotshade’ mode. Gaps were colored incorrectly when a single sequence was set as consensus. Another ‘donotshade’ problem was solved which led to a halt of the LaTeX run². Due to several

¹Inspired by Christoph Gille’s `STRAP`

²Thanks to Jeferson J. Arenzon and Naomi Siew

requests, the gap and match labels in **diverse** mode were switched (‘-’ in gaps; ‘.’ at matching positions) in order to follow convention. *Introduction:* **TeXshade** speaks spanish (`\spanishlanguage`). Necessary translations were contributed by Mikel Egaña Aranguren. A new feature label style **helix** has been introduced.

v1.6 2002/03/26

Corrections: The unnecessary restriction to the DVIPS driver for `color.sty` has been removed³. Any `color.sty` compatible driver option can be given with the `\usepackage{texshade}` call and is then passed to the `color` package. The ‘`\namecolor`’ and ‘`\numbercolor`’ commands do now support sequence lists.⁴

Introductions: (a) The FASTA file format is supported by **TeXshade** as alignment inputs. (b) Two commands set the space between sequence blocks either to be flexible (as so far) ‘`\flexblockspace`’ or the be fixed ‘`\fixblockspace`’. (c) One can now refer to sequences by their name in addition to the number in the input file. (d) Using ‘`\firstcolumnDSSP`’ and ‘`\secondcolumnDSSP`’ one can choose which of the first to columns should refer to the sequence numbering (the second column remains default setting)⁵.

v1.5a 2001/03/08

Corrections: ‘X’s in the alignment file caused a run-time error. Fixed.

Introductions: (a) The vertical space between feature lines can be controlled by four new commands: `\ttopspace`, `\topspace`, `\bottomspace` and `\bbottomspace`⁶. (b) It is now easily possible to add a caption to the alignment with the `\showcaption` command. (c) **TeXshade** stores the sequence lengths in the `.aux` file in order to have correct breaks of the gaps after the sequences.

v1.4&4a 2000/9/12 & 2000/10/3

Introductions: (a) The alignment legend can now be moved by the command ‘`\movelegend`’. (b) In commands with parameters that contain series of sequence numbers, e.g. `\orderseqs`, a dash can be used, e.g. `{1-3,6-4,7}` instead of `{1,2,3,6,5,4,7}`.

v1.3a&b 2000/7/28 & 2000/7/30

³As suggested by Eckhart Guthöhrlein.

⁴Thanks to Denys Bashtovyy.

⁵c and d were suggested by Christoph Gille.

⁶Suggested by Ulrike Folkers.

Introductions: (a) It is now possible to force `TeXshade` to display gap symbols before and after the actual sequence by the commands ‘`\showleadinggaps`’ and ‘`\hideleadinggaps`’ (4.4.1). (b) The sequence names input routine is now more tolerant concerning special characters.

v1.3 2000/3/3

Corrections: Line scrambling occurred when features were set in the `ttop` row without a feature in the `top` row. Fixed. The incompatible command ‘`\language`’ with the `babel` package has been replaced by ‘`\germanlanguage`’ and ‘`\englishlanguage`’⁷.

Introductions: (a) Now, translations of sequence stretches are possible. Either nucleotide or amino acid sources can be translated. This is done by the new `{translate}` option for the feature command. (b) The codons are defined by the new command ‘`\codon`’. Complete codon sets can be loaded by ‘`\geneticcode`’. (c) Further, the size and style of the nucleotide triplets of backtranslations can be set by ‘`\backtranslabel`’ and ‘`\backtranstext`’. (d) Two more feature counter styles were introduced: ‘`\Romancount`’ and ‘`\romancount`’. (e) `TeXshade` is now compatible with `TeXtopo`, a new `TeX` software for drawing and shading topology plots of membrane proteins.

v1.2a 1999/6/24 (not released)

Minor corrections: ‘`\namecolor`’ and ‘`\numbercolor`’ are now really correctly reordered. Brackets (and) are now allowed in sequence names. The option `{case}` in ‘`\funcshadingstyle`’ works now.

v1.2 1999/6/12

Corrections: (a) Functional group definitions of more than seven groups produced an error when displaying group number eight. These residues were skipped in the alignment. Fixed.

Introductions: (a) Protein secondary structure files in the DSSP, STRIDE and PHD format can be included and displayed automatically within the alignment by ‘`\includeDSSP`’ (and similar commands for STRIDE, PHDsec and PHDtopo, 4.5.4). (b) Which types of secondary structures are to be included or skipped in the alignment is chosen by ‘`\showonDSSP`’ and ‘`\hideonDSSP`’ (and respective commands for STRIDE, PHDsec and PHDtopo). (c) The appearance of the labels is defined by ‘`\appearance`’. (d) Internal counters for

⁷Thanks to Eckhart Guthöhrlein.

repeatedly occurring structure types can be activated by ‘`\numcount`’, ‘`\alphacount`’ and ‘`\Alphacount`’. All commands are described in 4.5.4.

v1.1 1999/5/26

Corrections: (a) The activation of ‘`emphregion`’ lead to an emphasized following alignment. This has been corrected. (b) ‘`\namecolor`’ and ‘`\numbercolor`’ were not reordered with the command ‘`orderseqs`’. Fixed. (c) Sequence gaps at the beginning or the end of a sequence, i. e. before the first and after the last residue where labeled with the gap symbol. Now these positions are left blank.

Introductions: (a) In order to treat the preceeding and sequence following gaps correctly, `TeXshade` needs to know the length of the sequences. Therefore, the command ‘`\seqlength`’ was introduced (4.4). (b) With ‘`\gapcolors`’ (also 4.4) the color selection for gap symbols is independent from non conserved residues. (c) The divisions of the ruler where so far fixed to 10. Now, this value is changeable by ‘`\rulersteps`’ (again 4.4). (d) ‘`\hideresidues`’ and ‘`\showresidues`’ turn off or on the residue names, i. e. one can choose between a display of shaded boxes only or with letters in the boxes (4.4.2). (e) The changes (c) through (d) were necessary for the introduction of ‘`\fingerprint`’. This command allows one to display the complete sequence in one line for an easy survey of the alignment (4.4.4).

v1.0 1999/5/12

First release.

1.2 L^AT_EX basics

1.2.1 Typesetting documents using L^AT_EX

In order to use any of the macros provided by the BioTeX-project (`TeXshade`/`TeXtopo`) efficiently a basic understanding of the TeX typesetting system and its usage is required. Several books are available on this topic, but a rather quick and easy introduction is the *Not so short introduction to L^AT_EX*. This document is available from all Comprehensive TeX Archive Network (CTAN) servers, e. g. from `ftp://ftp.dante.de/pub/tex/documentation/lshort/`, in many different languages and formats besides L^AT_EX, such as

POSTSCRIPT and on-line viewable PDF. I also put a link from the BioTeX (TeXshade/TeXtopo) homepage to the document collection (<http://homepages.uni-tuebingen.de/beitz/biotex.html>).

1.2.2 Memory shortness when using TeXshade

If you are using TeXshade to align several large sequences (about 1000 residues/sequence), LaTeX will probably stop compiling and quit with one of the following messages:

```
!\ TeX capacity exceeded, sorry [main memory size=384000]
or
```

```
!\ TeX capacity exceeded, sorry [stack size=300].
```

TeX allocates space for different kinds of internal variables. Setting alignments needs lots of memory, usually more than for typesetting plain text. Thus, the parameter settings of a standard TeX installation might not be sufficient for certain projects. This manifests in TeX error messages about insufficient memory and the setting process is interrupted. There is no reason to be concerned. The parameters can be set by hand. Unfortunately, each TeX system hides its default parameter file in a different place in the system.

In the following, an excerpt from the FAQ-list to TeXshade is added. This explains how to increase the settings in OzTeX for the Macintosh, MikTeX for Windows and TeTeX for *NIX TeX distributions. Please contribute to this list!

1. OzTeX 4.0 for the Macintosh:

Find the file ‘OzTeX:TeX:Configs:Default’. This file contains all memory settings. Look for the section ‘% TeX parameters’ and increase the values that TeX complains about during the run. You will have to restart OzTeX before the changes are active.

For older versions of OzTeX the configuration file has the same name but the path is somewhat different.

2. TeTeX for *NIX: (contributed by Joerg Daehn)

Find the file: ‘/usr/share/texmf/web2c/texmf.cnf’ or use

```
locate texmf.cnf
```

 at the command prompt to find it.

Login as super user. Backup ‘texmf.cnf’ in case you destroy something and then open the ‘texmf.cnf’ file in your favorite text editor and use its search function to locate `main_memory`. This variable is set to 384000. Change this to some higher value,

i.e. 4000000 (works fine for me!). The total amount of memory should not exceed 8000000, so check the other values in that section.

Next, you want to change the stack size. Search for `stack_size`. This will be set to 300. I changed it to 4000 and it works fine.

There might be complains by $\text{\TeX}$ about further specific parameters such as `stack_size`. You find all those in the same file.

After this you have to run ‘texconfig init’.

Logout as root.

After this all should be set for large alignments. Happy $\text{\TeX}$ ing!

The information on how to achieve this was derived from a mail in the $\text{te}\text{\TeX}$ mail archive. The original question was posted by Pascal Francq and answered by Rolf Nieprasch.

3. MiK $\text{\TeX}$ for Windows:

The MiK $\text{\TeX}$ documentation describes very detailed how the memory settings can be changed. In brief, you must locate the configuration file ‘miktex/config/miktex.ini’. In the [MiK $\text{\TeX}$] section of this file you find all the parameters you need, e.g. `mem_min`, `mem_max`, `buf_size`, `stack_size` etc.

It appears, that the standard settings of MiK $\text{\TeX}$ are bigger than that of other $\text{\TeX}$ installations, so it may not always be necessary to increase the values.

1.3 System requirements

$\text{\TeXshade}$ requires $\text{\LaTeX} 2_{\epsilon}$ and `color.sty` for shading. David Carlisle’s `color.sty` is part of the Standard $\text{\LaTeX}$ ‘Graphics Bundle’ [1]. This package can be downloaded from any $\text{\TeX}$ archive, e.g. ftp.dante.de; usually it is already included in a comprehensive $\text{\TeX}$ installation.

The `color` style allows one to use several [*options*], e.g. `dvips`, `pdftex` or `dviwin`. These provide the commands which different devices/programs need to display colored output. It is advisable to make yourself familiar with the `color.sty` manual. You should define a default driver in the file `color.cfg`. Since there is no direct call of `color.sty` by the user, the option can be stated when $\text{\TeXshade}$ is

loaded, see next subsection. If no option is stated the DVIPS driver will be loaded as was default before.

With the [dvips] option for example the output DVI-file can be converted to POSTSCRIPT using the DVIPS program and can later be viewed or printed with the public domain Ghostview program which is available for almost all computer platforms. Further, more and more standard T_EX viewers are to a certain extent POSTSCRIPT compatible, e.g. OzT_EX on the Macintosh. The option pdf_{tex} makes the conversion to a PDF file easy etc.

1.4 The texshade environment

The commands provided by the T_EX_{shade} package are enabled by the following command in the document header section:

```
\usepackage[<option>]{texshade}
```

Make sure that the file ‘texshade.sty’ is present in a directory searched by T_EX (see the installation notes in the file ‘texshade.txt’). The *<option>* given here is passed to color.sty which handles the color commands for a particular output device, see previous subsection and the color.sty manual.

The T_EX_{shade} package provides only one single new environment: texshade. This environment has one mandatory and one optional argument, both of them designating file names which must be present in a directory searched by T_EX. The required file *<alignmentfile>* contains the aligned nucleotide or peptide sequences (see section 2). This file is needed, because T_EX_{shade} does no alignment by itself, it has to take a preprocessed alignment as input. The optional file is a parameter file (section 3) with definitions for the customized calculation of the consensus, special sequence features or labels etc. In this parameter file all T_EX_{shade} commands which are allowed in the texshade environment can be used and are fully functional. Within the environment further T_EX_{shade} commands can be given to replace or complete settings from the parameter file.

Thus, setting an alignment with T_EX_{shade} is as simple as this:

```
\begin{texshade}[<parameterfile>] {<alignmentfile>}
    further TEXshade commands, if needed
\end{texshade}
```

1.5 Shading modes predefined in this package

1.5.1 Identity mode

This basic type of shading is provided by almost any alignment program. All identical residues at a position are shaded if the number of matching residues is higher than a given threshold percentage.

```
80  TLGLLLSCQISILRAVMYIIAQCVGAIVASAIL  AQP1.PRO
72  TVACLVGCHVSFLRAAFYVAAQLLGAVAGAAIL  AQP2.PRO
80  TFAMCFLAREPWIKLPITYTLAQTGAFILGAGIV  AQP3.PRO
101 TVAMVCTRKISIAKSVFYITAQCLGATIGAGIL  AQP4.PRO
73  TLALLIGNQISLLRAVFYVAAQLVGAIIAGAGIL  AQP5.PRO
```

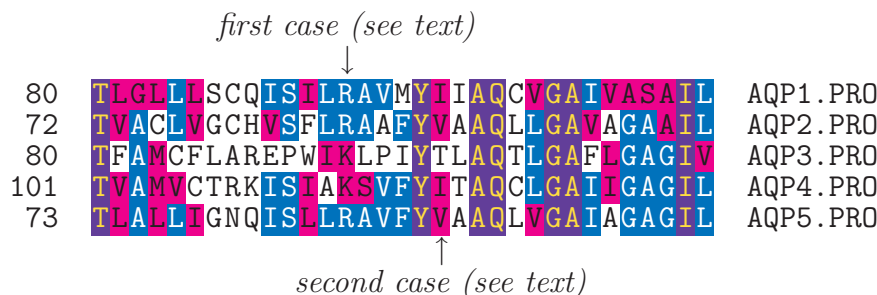
If you like, positions where all residues are identical can be shaded in a special color:

```
80  TLGLLLSCQISILRAVMYIIAQCVGAIVASAIL  AQP1.PRO
72  TVACLVGCHVSFLRAAFYVAAQLLGAVAGAAIL  AQP2.PRO
80  TFAMCFLAREPWIKLPITYTLAQTGAFILGAGIV  AQP3.PRO
101 TVAMVCTRKISIAKSVFYITAQCLGATIGAGIL  AQP4.PRO
73  TLALLIGNQISLLRAVFYVAAQLVGAIIAGAGIL  AQP5.PRO
```

1.5.2 Similarity mode

In many cases it is expedient—mostly when comparing protein sequences—to shade also residues which are not identical but similar to the consensus sequence. Consider a position where three out of five residues are basic arginines and two more residues are also basic but lysines. In similarity mode `TeXshade` shades similar residues in a different color to distinguish them from the consensus residue. Even when none of the residues alone reaches the threshold but a group of similar residues does these are shaded in the ‘similarity’ color. This case is given for instance when at a position in a five sequence alignment two aliphatic valines and two also aliphatic isoleucins are present and the threshold is set to 50%. Neither residue exceeds this percentage but as a group of similars they do.

In grayscale printouts some colors of the following alignment may appear undistinguishable. Don’t worry if you usually use grayscale—all colors/grays can be selected freely (see 5).

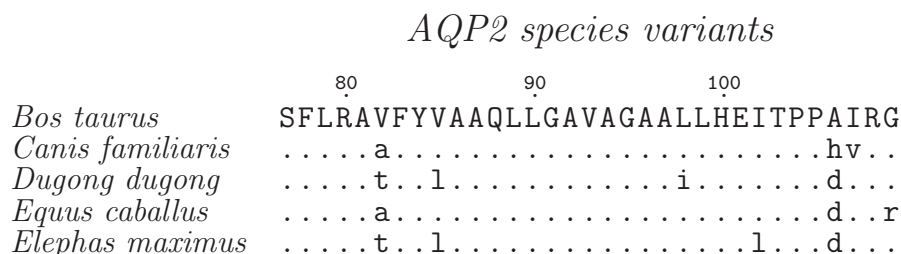


Probably you know this kind of shading from the VMS/Unix and DOS public domain program `BoxShade` by KAY HOFMANN or from the Macintosh version `MacBoxShade` by MICHAEL D. BARRON. `TeXshade` provides the same functionality—and goes truly beyond—for the `TeX` community.

1.5.3 Diversity mode

Contrary to the above described modes this shading style displays sequence differences. Thus, it is most suitable for comparing very similar sequences, e. g. species variants of a protein.

One sequence is used as consensus. Matching residues in other sequences are blanked out, mismatches are shown in lowercase.



1.5.4 Functionality modes

Displaying functional peptide similarities is one of `TeXshade`'s strong capabilities. Six functional shading modes are predefined; further user specific modes can easily be created. The examples may not look very impressive when printed in grayscale. Enjoy them on your screen or use color printouts. As mentioned before, all colors can be changed to others or to grays without restrictions (see chapter 5).

charge: residues which are charged at physiological pH (7.4) are shaded if their number at a position is higher than the threshold



138	GLGI	E	I	I	G	T	L	Q	L	V	L	C	V	L	A	T	T	D	R	.	R	R	R	D	L	G	G	S	A	P	L	AQP1.PRO
130	AVTV	E	L	F	L	T	M	Q	L	V	L	C	I	F	A	S	T	D	E	.	R	R	G	D	N	L	G	S	P	A	L	AQP2.PRO
153	GFFD	Q	F	I	G	T	A	A	L	I	V	C	V	L	A	I	V	D	P	Y	N	N	P	V	P	R	G	L	E	A	F	AQP3.PRO
159	GLLV	E	L	I	I	T	F	Q	L	V	F	T	I	F	A	S	C	D	S	.	K	R	T	D	V	T	G	S	V	A	L	AQP4.PRO
131	AMVV	E	L	I	L	T	F	Q	L	A	L	C	I	F	S	S	T	D	S	.	R	R	T	S	P	V	G	S	P	A	L	AQP5.PRO

X acidic (-)
X basic (+)

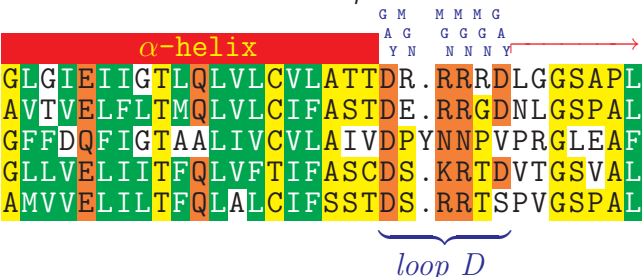
hydropathy: discrimination between acidic and basic, polar uncharged and hydrophobic nonpolar residues

138	GLGI	E	I	I	G	T	L	Q	L	V	L	C	V	L	A	T	T	D	R	.	R	R	R	D	L	G	G	S	A	P	L	AQP1.PRO
130	AVTV	E	L	F	L	T	M	Q	L	V	L	C	I	F	A	S	T	D	E	.	R	R	G	D	N	L	G	S	P	A	L	AQP2.PRO
153	GFFD	Q	F	I	G	T	A	A	L	I	V	C	V	L	A	I	V	D	P	Y	N	N	P	V	P	R	G	L	E	A	F	AQP3.PRO
159	GLLV	E	L	I	I	T	F	Q	L	V	F	T	I	F	A	S	C	D	S	.	K	R	T	D	V	T	G	S	V	A	L	AQP4.PRO
131	AMVV	E	L	I	L	T	F	Q	L	A	L	C	I	F	S	S	T	D	S	.	R	R	T	S	P	V	G	S	P	A	L	AQP5.PRO

X acidic (-)
X basic (+)
X polar uncharged
X hydrophobic nonpolar

structure: displays the potential localization within the tertiary structure of the protein

transmembrane domain 4 *trans. dom. 5*



138	GLGI	E	I	I	G	T	L	Q	L	V	L	C	V	L	A	T	T	D	R	.	R	R	R	D	L	G	G	S	A	P	L	AQP1.PRO
130	AVTV	E	L	F	L	T	M	Q	L	V	L	C	I	F	A	S	T	D	E	.	R	R	G	D	N	L	G	S	P	A	L	AQP2.PRO
153	GFFD	Q	F	I	G	T	A	A	L	I	V	C	V	L	A	I	V	D	P	Y	N	N	P	V	P	R	G	L	E	A	F	AQP3.PRO
159	GLLV	E	L	I	I	T	F	Q	L	V	F	T	I	F	A	S	C	D	S	.	K	R	T	D	V	T	G	S	V	A	L	AQP4.PRO
131	AMVV	E	L	I	L	T	F	Q	L	A	L	C	I	F	S	S	T	D	S	.	R	R	T	S	P	V	G	S	P	A	L	AQP5.PRO

loop D

X external
X ambivalent
X internal

chemical: residues are shaded due to chemical properties of their functional groups

```

138  GLGIEIIGTLQLVLCVLATTDRLRRDLGGSAPL  AQP1.PRO
130  AVTVELFLTMQLVLCIFASTDERRGDNLGSPAL  AQP2.PRO
153  GFFDQFIGTAALIVCVLAIVDPYNNPVPRGLEAF  AQP3.PRO
159  GLLVELIITFQLVFTIFASCDS.KRTDVTGSVAL  AQP4.PRO
131  AMVVELILTFLQALCIFSSTDS.RRTSPVGSVAL  AQP5.PRO
  
```

X acidic (-)
X aliphatic
X amide
X aromatic
X basic (+)
X hydroxyl
X imino
X sulfur

With `\shadeallresidues` the threshold is ignored and all residues are shaded due to their group assignment. This is *not* identical to a threshold of 0% where only the majority group would be shaded. See the difference:

```

138  GLGIEIIGTLQLVLCVLATTDRLRRDLGGSAPL  AQP1.PRO
130  AVTVELFLTMQLVLCIFASTDERRGDNLGSPAL  AQP2.PRO
153  GFFDQFIGTAALIVCVLAIVDPYNNPVPRGLEAF  AQP3.PRO
159  GLLVELIITFQLVFTIFASCDS.KRTDVTGSVAL  AQP4.PRO
131  AMVVELILTFLQALCIFSSTDS.RRTSPVGSVAL  AQP5.PRO
  
```

standard area: this shading displays the differences in the surface area of the different amino acid's sidechains

```

138  GLGIEIIGTLQLVLCVLATTDRLRRDLGGSAPL  AQP1.PRO
130  AVTVELFLTMQLVLCIFASTDERRGDNLGSPAL  AQP2.PRO
153  GFFDQFIGTAALIVCVLAIVDPYNNPVPRGLEAF  AQP3.PRO
159  GLLVELIITFQLVFTIFASCDS.KRTDVTGSVAL  AQP4.PRO
131  AMVVELILTFLQALCIFSSTDS.RRTSPVGSVAL  AQP5.PRO
  
```

X	88.1 (G); Standard sidechain area (Å ²)
X	118.2 (A); 129.8 (S)
X	146.1 (C); 146.8 (P)
X	152.5 (T); 158.7 (D); 164.5 (V); 165.5 (N)
X	181.0 (I); 186.2 (E)
X	193.1 (L); 193.2 (Q); 202.5 (H); 203.3 (M)
X	222.8 (F); 225.8 (K)
X	238.8 (Y)
X	256.0 (R); 266.2 (W)

accessible area: here, the surface area which can be accessed by solvent molecules is used as a basis for shading; low accessibility means hydrophobic (i. e. strongly buried residues), whereas highly accessible sidechains are hydrophilic (compare to **hydropathy** and **structure**)

	<i>membr.</i>	<i>loop</i>	<i>membr.</i>	
	oooooooooooooooooooo	—————	ooooooo	
138	GLGIEIIGTLQLVLCVLATTD	R.RRD	LGGSAPL	AQP1.PRO
130	AVTVELFLTMQLVLCIFASTDE	.RRGD	NLGSPAL	AQP2.PRO
153	GFFDQFIGTAALIVCVLAIVDP	YNNPV	PRGLEAF	AQP3.PRO
159	GLLVELIITFQLVFTIFASCDS	.KRTD	VTGSVAL	AQP4.PRO
131	AMVVELILTFQLALCIFSSTDS	.RRTSP	VGSPAL	AQP5.PRO

X	13.9 (C); Accessible sidechain area (Å ²)
X	23.0 (I); 23.5 (V); 25.2 (G)
X	28.7 (F); 29.0 (L); 30.5 (M); 31.5 (A)
X	41.7 (W); 44.2 (S); 46.0 (T); 46.7 (H)
X	53.7 (P)
X	59.1 (Y); 60.9 (D); 62.2 (N)
X	72.3 (E); 74.0 (Q)
X	93.8 (R)
X	110.3 (K)

1.5.5 Secondary structures

Predicted protein secondary structures in the DSSP, STRIDE or PHD file format can be included and displayed in the alignment. As an example, the following few commands show an aquaporin alignment with the PHD topology data for aquaporin type 1 (top sequence).

```
\begin{texshade}{AQPpro.MSF}
```

```
\shadingmode[allmatchspecial]{similar}
\includePHDtopo{1}{AQP1.phd}
\end{texshade}
```

Abbr.: *int.* – internal; *ext.* – external; *TM* – transmembrane domain

1 MAS.....EIKKKLFW AQP1.PRO
1 MW.....ELRSIAFS AQP2.PRO
1 M.....NRCG.....EMLHIRYR.....LL AQP3.PRO
1 MSDGAAARRWGKCGPPCSRESIMVAFKGVWITQAFW AQP4.PRO
1 MK.....KEVCSLAFF AQP5.PRO
! *** * *** consensus

int. A

ext. B

TM1

12 RAVVAEFLAMTLFVFISIGSALGFNYPLERNQTLV AQP1.PRO
11 RAVLAEFLATLLFVFFGLGSALQWA...SS....P AQP2.PRO
16 RQALAECLGTLILVMFCGSVAQVVLRSRGTHGGF. AQP3.PRO
36 KAVTAEFLAMLIFVLLSVGSTINWG...GSENPLP AQP4.PRO
12 KAVFAEFLATLIFVFFGLGSALKWP...SA....L AQP5.PRO
****!!*!*****!*****!!***** * consensus

int. A

ext. B

TM2

47 QDNVKKVSLAFGLSISATLAQSVGHISGAHSNPAVTL AQP1.PRO
39 PSVLQIAVAFGLGIGILVQALGHVSGAHINPAVTV AQP2.PRO
50 ...LTINLAFGFAVTLAILVAQGVSGAHLNPAVTF AQP3.PRO
68 VDMVLISLCFGLSISATMVQCFGHISGGHINPAVTV AQP4.PRO
40 PTILQISIAFGLAIGTLAQAALGPVSGGHINPAITL AQP5.PRO
** *****!!***** *!*****!!*!***** consensus

int. C

ext. D

TM3

82 GLLLSCQISILRAVMYIIAQCVGAIVASAILSGI. AQP1.PRO
74 ACLVGCHVSFLRAAFYVAAQLLGAVAGAAIILHEI. AQP2.PRO
82 AMCFLAREPWIKLPIYITLAQT LGAF LGAGIVFGLY AQP3.PRO
103 AMVCTRKISIAKSVFYITAQCLGAIIGAGILYLV. AQP4.PRO
75 ALLTIGNQISLLRAVFYVAAQLVGAIIAGAGILYWL. AQP5.PRO
**** *****!!*!*****!!*!***** consensus

116 TSSLENS LGRND LARGVNS GQ... AQP1.PRO
 108 TPVEIRGD LAVNALHN NATAGQ... AQP2.PRO
 117 YDAIWAFAGNEL LVSGPNGTAGIFATYPS GHLDMV AQP3.PRO
 137 TPPSVVGG LGVTTVHGNL TAGH... AQP4.PRO
 109 APLNARGN LAVNALNNNTTP GK... AQP5.PRO
 *** * * ***** * * ! consensus

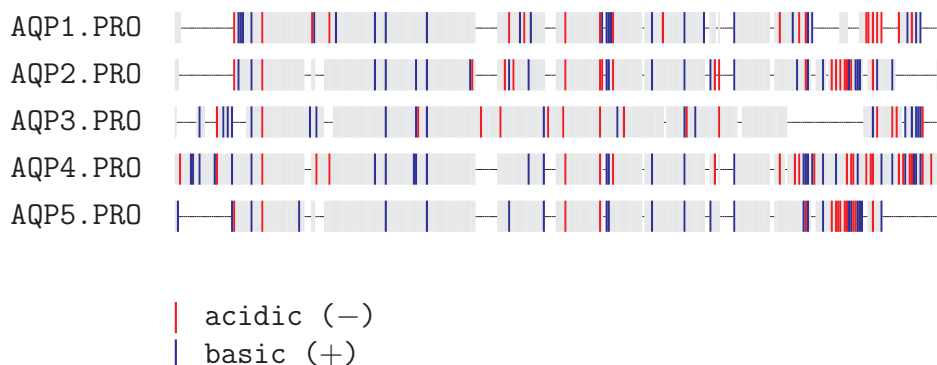
Position	AQP1	AQP2	AQP3	AQP4	AQP5	Consensus
138	G	A	N	G	A	*
130	V	T	G	L	V	*
152	F	V	F	L	V	*
159	I	F	I	T	I	*
131	L	V	L	I	L	*

TM4
 AQP1.PRO
 AQP2.PRO
 AQP3.PRO
 AQP4.PRO
 AQP5.PRO
 consensus

 $ext. F$ [illegible][illegible]

	non conserved
	similar
	conserved
	all match

The higher the similarity the darker the vertical lines. In this overview it becomes obvious that the transmembrane regions of the aquaporin isoforms are most conserved.



A fingerprint of charge distribution on different aquaporins. Sequence gaps can be left blank (first example) or drawn as lines between the sequence blocks (shown here).

1.6 Customization of alignment outputs

Extensive possibilities are given to the user to customize the final output of an alignment. Thus, all parameters defining the appearance of letters can be changed individually for sequence residues, names and numbering or the describing feature texts. Additional manual shading can be applied to any region or block of residues. Sequences are easily re-ordered, separated, hidden or blanked out without recalculation of the entire alignment; sections of the alignment can also be shown. Numbering and rulers can be displayed and set to any value. A powerful tool is the `\feature` command which allows one to label stretches of residues with bars, arrows, braces or any fill character and describing text. Legends are set automatically if desired, but user commands are also provided to build individual legends.

2 Format of alignment input files

TeXshade can handle two common alignment input formats, i.e. the MSF format (multiple sequences format) and the ALN format (alignment format). The MSF format is used by PILEUP of the Unix GCG sequence analysis package⁸. Files in the ALN format are produced by CLUSTAL which is available for free for Unix, DOS and Macintosh. Further, upon request, the FASTA format is supported since version 1.6. In addition to the mentioned software many alignment programs have export filters for the MSF, ALN or FASTA format, e.g. MACAW produces ALN files. If you are not sure whether your favorite sequence aligner produces one of the required formats compare its output to the following examples. TeXshade determines the format from the internal file structure, thus extensions like MSF, ALN or FASTA are not required. If you can choose the alignment format MSF is recommended, because this format gives information about the sequence type, i.e. peptide or nucleotide sequences, and length (for the correct setting of gaps at the sequence end).

2.1 The MSF file format

Files of this type are divided into a header section and the multiple sequence alignment. The header may contain the following components:

File Type: (optional) The first header line reads for nucleic acids alignments
!!NA_MULTIPLE_ALIGNMENT 1.0 and for amino acid sequences
!!AA_MULTIPLE_ALIGNMENT 1.0 (all uppercase).

Description: (optional) Informative text describing what is in the file.

Dividing line: (required!) Must include the following attributes:

MSF: Displays the number of bases or residues in the multiple sequence alignment.

Type: Displays the sequence type, 'P' for a peptide and 'N' for a nucleotide alignment.

Checksum: Displays an integer value that characterizes the contents of the file.

⁸For a description see <http://gene.md.huji.ac.il/Computer/GCG9doc>

.. The two periods act as a divider between the descriptive information and the following sequence information.

Name/Weight: (required!) Must include the name of each sequence included in the alignment, as well as its length, checksum and weight.

Two slashes (//): (required!) This separating line divides the name/weight information from the sequence alignment

The alignment section consists of sequence blocks divided by an empty line. Each sequence line starts out with the sequence name. An example file is shown here:

```
AQP.MSF  MSF: 87  Type: P  May 1st, 1998 Check: 2586 ..
Name: AQP1.PRO  Len: 66  Check: 1367  Weight: 1.00
Name: AQP2.PRO  Len: 58  Check: 2176  Weight: 1.00
Name: AQP3.PRO  Len: 83  Check: 1893  Weight: 1.00
Name: AQP4.PRO  Len: 63  Check: 3737  Weight: 1.00
Name: AQP5.PRO  Len: 59  Check: 3413  Weight: 1.00
//
           1                                     45
AQP1.PRO  MAS.....EIKKKLFWRAVVAEFLAM
AQP2.PRO  MW.....ELRSIAFSRAVLAEFLAT
AQP3.PRO  M.....NRCG.....EMLHIRYR.....LLRQALAECLGT
AQP4.PRO  MSDGAAARRWGKCGPPCSRESIMVAFKGVWTQAFWKAVTAEFLAM
AQP5.PRO  MK.....KEVCSLAFFKAVFAEFLAT

           45                                     87
AQP1.PRO  TLFVFISIGSALGFNYPLERNQTLVQDNVKVSLAFGLSIATL
AQP2.PRO  LLFVFFGLGSALQWA...SS...PPSVLQIAVAFGLGIGIL
AQP3.PRO  LILVMFGCGSVAQVLSRGTHGGF...LTINLAFGFVAVTLA
AQP4.PRO  LIFVLLSVGSTINWG...GSENPLPVDMLISLCFGLSIATM
AQP5.PRO  LIFVFFGLGSALKWP...SA...LPTILQISIAFGLAIGTL
```

TeXshade extracts only the information from the file it really needs. So, do not mind all the checksums listed in the file—**TeXshade** does not either. The same is true for **Weight**. Required are the string **MSF:** for the identification of the file format and **Type:** for the determination of the sequence type (both in the dividing line), further all **Name:** definitions and finally **//**. The MSF format allows one to comment out sequences. This is done by putting an exclamation point directly in front of the respective **Name**. These sequences are neither

displayed nor used for the calculation of the consensus. This works for `TeXshade`, too. To comment out sequences without changing the input file use the `TeXshade` command `\killseq{<seqref>}` (4.4.2).

```

AQP.MSF  MSF: 87  Type: P  May 1st, 1998 Check: 2586 ..
Name: AQP1.PRO  Len: 66  Check:  1367      Weight: 1.00
!Name: AQP2.PRO  Len: 58  Check:  2176      Weight: 1.00
!Name: AQP3.PRO  Len: 83  Check:  1893      Weight: 1.00
Name: AQP4.PRO  Len: 63  Check:  3737      Weight: 1.00
Name: AQP5.PRO  Len: 59  Check:  3413      Weight: 1.00
//
          1                                     45
AQP1.PRO  MAS.....EIKKKLFWRAVVAEFLAM
AQP2.PRO  MW.....ELRSIAFSRAVLAEFLAT
AQP3.PRO  M.....NRCG.....EMLHIRYR.....LLRQALAECLGT
AQP4.PRO  MSDGAAARRWGKCGPPCSRESIMVAFKGVWTQAFWKAVTAEFLAM
AQP5.PRO  MK.....KEVCSLAFFKAVFAEFLAT

          45                                     87
AQP1.PRO  TLFVFISIGSALGFNYPLERNQTLVQDNVKVSLAFGLSIATL
AQP2.PRO  LLFVFFGLGSALQWA...SS...PPSVLQIAVAFGLGIGIL
AQP3.PRO  LILVMFGCGSVAQVLSRGTHGGF...LTINLAFGFVAVTLA
AQP4.PRO  LIFVLLSVGSTINWG...GSENPLPVDMLISLCFGLSIATM
AQP5.PRO  LIFVFFGLGSALKWP...SA...LPTILQISIAFGLAIGTL

```

The sequence lengths given after `Len:` are not used by `TeXshade`. Due to the fact that most alignment programmes calculate the sequence length by summing up residues and additionally gaps which is not really correct. In order to have the sequence break right after the last residue without printing further gap symbols `TeXshade` counts the number of residues by itself. You can also use the command `\seqlength` in the `TeXshade` environment to set the values manually if you do not trust a machine.

2.2 The ALN file format

ALN files are quite similar to the above described MSF files. They simply lack a defined header section. Nevertheless, describing text is allowed before the alignment part. `TeXshade` determines the number of sequences and their names from the last sequence block—so, no further text lines are allowed after this block! Due to a lacking

declaration in the file the sequence type has to be set in the `texshade` environment by `\seqtype{<type>}` with ‘P’ for peptide and ‘N’ for nucleotide sequences; for the example below: `\seqtype{P}`. If no `\seqtype` command is used `TeXshade` assumes a nucleotide sequence.

```

                                profalign  May 1st, 1998, 16:58

of AQPpro.MSF{}

Multiple alignment parameter:

Gap Penalty (fixed):           10.00
Gap Penalty (varying):         .05
Gap separation penalty range:   8
Percent. identity for delay:    0%
List of hydrophilic residue:    GPSNDQEKRH
Protein Weight Matrix:         blossom

                                10      20      30      40
                                .      .      .      .
AQP1.PRO  MAS.....EIKKKLFWRAVVAEFLAM
AQP2.PRO  MW.....ELRSIAFSRAVLAEFLAT
AQP3.PRO  M.....NRCG.....EMLHIRYR.....LLRQALAECLGT
AQP4.PRO  MSDGAAARRWGKCGPPCSRESIMVAFKGVWTQAFWKAVTAEFLAM
AQP5.PRO  MK.....KEVCSLAFFKAVFAEFLAT
          *                               .    ** *.

AQP1.PRO  TLFVFISIGSALGFNYPLERNQTLVQDNVKVSLAFGLSIATL
AQP2.PRO  LLFVFFGLGSALQWA...SS...PPSVLQIAVAFGLGIGIL
AQP3.PRO  LILVMFGCGSVAQVLSRGTHGGF....LTINLAFGFAVTLA
AQP4.PRO  LIFVLLSVGSTINWG...GSENPLPVDMLISLCFGLSIATM
AQP5.PRO  LIFVFFGLGSALKWP...SA....LPTILQISIAFGLAIGTL
          .. *    .**                               .    ** .

```

The minimal contents of an ALN file are shown below; this is fully sufficient. Many sequence alignment programs can produce such an output. Have a look at `seqpup` by DON GILBERT if you need a comprehensive conversion program⁹.

⁹Sorry, `seqpup` is much more!

```

AQP1.PRO MAS.....EIKKKLFWRAVVAEFLAM
AQP2.PRO MW.....ELRSIAFSRAVLAEFLAT
AQP3.PRO M.....NRCG.....EMLHIRYR.....LLRQALAECLGT
AQP4.PRO MSDGAAARRWGKCGPPCSRESIMVAFKGVWTQAFWKAVTAEFLAM
AQP5.PRO MK.....KEVCSLAFFKAVFAEFLAT

AQP1.PRO TLFVFISIGSALGFNYPLERNQTLVQDNVKVSLAFGLSIATL
AQP2.PRO LLFVFFGLGSALQWA...SS...PPSVLQIAVAFGLGIGIL
AQP3.PRO LILVMFGCGSVAQVLSRGTHGGF...LTINLAFGFAVTLA
AQP4.PRO LIFVLLSVGSTINWG...GSENPLPDMVLISLCFGLSIATM
AQP5.PRO LIFVFFGLGSALKWP...SA...LPTILQISIAFGLAIGTL

```

2.3 The FASTA file format

In FASTA files each sequence is led by a single description line starting with a '>'. **TeXshade** uses the first word delimited by the leading '>' and a space as the sequence name. If no descriptive text is present **TeXshade** generates a sequence name consisting of 'seq' plus a consecutive number. The lines following the description line contain the sequence.

```

>AQP1.PRO
MAS.....EIKKKLFWRAVVAEFLAM
TLFVFISIGSALGFNYPLERNQTLVQDNVKVSLAFGLSIATL

>AQP2.PRO
MW.....ELRSIAFSRAVLAEFLAT
LLFVFFGLGSALQWA...SS...PPSVLQIAVAFGLGIGIL

>AQP3.PRO
M.....NRCG.....EMLHIRYR.....LLRQALAECLGT
LILVMFGCGSVAQVLSRGTHGGF...LTINLAFGFAVTLA

>AQP4.PRO
MSDGAAARRWGKCGPPCSRESIMVAFKGVWTQAFWKAVTAEFLAM
LIFVLLSVGSTINWG...GSENPLPDMVLISLCFGLSIATM

>AQP5.PRO
MK.....KEVCSLAFFKAVFAEFLAT
LIFVFFGLGSALKWP...SA...LPTILQISIAFGLAIGTL

```

3 Use of a T_EXshade parameter file

Using predefined parameter files for repeatedly occurring situations can save a lot of typing and makes the output throughout the publication or presentation more consistent. Further, such files are an easy way to exchange self-defined shading modes or new color schemes (i. e. for a satisfying grayscale output) with other users. If you have created a parameter file, which you think is of interest for others, please submit it to me¹⁰ as an e-mail attachment together with a short description. I will take care of those files and post them—with a reference to the author—together with the next T_EXshade distribution to make them available for all interested users.

No special file format is required for parameter files. T_EXshade simply calls the file using the `\input` command right after resetting all parameters to default. An example parameter file is present containing the standard parameters of T_EXshade called `texshade.def`. This file can be changed freely and can be used as a template for the creation of personal parameter files.

Five steps are executed by T_EXshade when processing the `texshade` environment:

```
\begin{texshade}[\langle parameterfile \rangle]{\langle alignmentfile \rangle}
```

1. Analysis of the `\langle alignmentfile \rangle`; determination of the number of sequences and sequence names
2. Setting parameters to default
3. Setting parameters to the definitions of the `\langle parameterfile \rangle`, if existent
4. Execution of further T_EXshade commands within the environment, if existent

```
\end{texshade}
```

5. Loading and setting the alignment on a line by line basis

¹⁰`eric.beitz@uni-tuebingen.de`

4 texshade user commands

The `TeXshade` package must be loaded by the `\usepackage` command in the document header section.

```
\usepackage[<option>]{texshade}
```

Then, the `texshade` environment is ready to use as described in 1.4. See also section 3 for a description of the optional parameter file. All other commands provided by `TeXshade` (except `\molweight`, `\charge` [4.9] and `\shadebox` [4.6]) must be used within the `texshade` environment.

4.1 Using predefined shading modes

If no `\shadingmode` command is given in the `texshade` environment the default shading mode (*identical*, see 1.5.1) is active. For the selection of one of the other predefined shading modes the following command is provided.

```
\shadingmode[<option>]{<mode>}
```

You can choose from four shading modes and declare one option which depends on the selected mode.

1. `\shadingmode[<allmatchspecial>]{identical}`

There is not much to explain here (see 1.5.1). Use the option `allmatchspecial` to shade positions with a special color where all residues are identical. `\allmatchspecial` can also be used as a command. As both, option or command `allmatchspecial` is only active in the *identical* and *similar* shading modes.

One can choose from five predefined shading color schemes with the command `\shadingcolors{<scheme>}`. The sets are named ‘blues’ (used in the example, 1.5.1), ‘reds’, ‘greens’, ‘grays’ and ‘black’. Default is `\shadingcolors{blues}`. Further, the colors for the non matching, the conserved and all matching residues can be set individually plus the letter case (lower or upper) or any character can be chosen:

```
\nomatchresidues{<res.col.>}{<shad.col.>}{<case>}{<style>}  
\conservedresidues{<res.col.>}{<shad.col.>}{<case>}{<style>}
```

`\allmatchresidues{⟨res.col.⟩}{⟨shad.col.⟩}{⟨case⟩}{⟨style⟩}`

For how to handle colors for the foreground `⟨res.col.⟩` and the background `⟨shad.col.⟩` see section 5. The third parameter `⟨case⟩` tells `TeXshade` to print the corresponding residue as a lowercase or an uppercase letter or even to print any other character. Finally, the `⟨style⟩` parameter tells `TeXshade` which shape to use for the letters. Use one of the following styles for `⟨style⟩`.

<code>⟨style⟩</code>	<i>effect</i>
<code>bf</code>	bold face series
<code>md</code>	normal series
<code>up</code>	upright shape (normal shape)
<code>it</code>	italics shape
<code>sl</code>	slanted shape
<code>rm</code>	modern roman family
<code>sf</code>	sans serif family
<code>tt</code>	typewriter family

In order to change only some of the parameters it is sufficient to declare these and use empty braces for the others. Examples:

`\conservedresidues{White}{Blue}{upper}{bf}`: the conserved residues are printed as bold face white uppercase letters on blue.

`\nomatchresidues{}{}{{\bullet}}`: instead of the non matching residues a ‘•’ is printed. The colors and style are not changed. Note the double curly braces which make `TeXshade` interpret this complex symbol description as one single character.

2. `\shadingmode[⟨allmatchspecial⟩]{similar}`

See 1.5.2 for an example output and an explanation of the shading. In addition to the described commands for changing shading colors this shading mode provides the command `\similarresidues`. Use it in analogy to the commands above.

How does `TeXshade` know which residues are considered to be similar? These definitions are set by two command couples, i. e.

`\pepsims`, `\pepgroups` for peptides and `\DNAsims`, `\DNAGroups` for nucleotides. With `\pepsims` and `\DNAsims` residues are defined which are similar to the consensus residue. Examples:

`\pepsims{S}{TA}` If a serine is the consensus residue then all threonines and alanines at this position are shaded in the color for similars. This definition does *not* imply that threonine and alanine are similar to each other! This becomes obvious when you inspect the next definition:

`\pepsims{T}{S}` Serine but not alanine is declared to be similar to threonine.

What happens if there is no consensus residue? How does `TeXshade` decide if a group of similars is greater than the threshold? Therefore groups are pre-defined:

`\pepgroups{FYW,ILVM,RK,DE,GA,ST,NQ}` This command allows one to set up to nine groups of similars, separated by commas. Each residue can belong to only one group. If one residue is assigned to several groups only the last assignment is carried out.

`\DNAGroups{GAR,CTY}` This command is used in analogy to the amino acid groups. Here, two ambiguity codes (‘R’ for purine base and ‘Y’ for pyrimidine base) are assigned in addition.

Residues which do not appear in any of the four commands are considered not to belong to a group. The default settings for similars are listed below:

```
\pepgroups{FYW,ILVM,RK,DE,GA,ST,NQ}
```

```
\pepsims{F}{YW} % Y and W are similar to F
\pepsims{Y}{WF} % W and F are similar to Y
\pepsims{W}{YF} % Y and F are similar to W
```

```
\pepsims{I}{LVM} % L, V and M are similar to I
\pepsims{L}{VMI} % V, M and I are similar to L
\pepsims{V}{MIL} % M, I and L are similar to V
```

```
\pepsims{R}{KH} % K and H are similar to R
\pepsims{K}{HR} % H and R are similar to K
\pepsims{H}{RK} % R and K are similar to H
```

```

\pepsims{A}{GS} % G and S are similar to A
\pepsims{G}{A}  % A (but not S) is similar to G

\pepsims{S}{TA} % T and A are similar to S
\pepsims{T}{S}  % S (but not A) is similar to T

\pepsims{D}{EN} % E and N (but not Q) are similar to D
\pepsims{E}{DQ} % D and Q (but not N) are similar to E
\pepsims{N}{QD} % Q and D (but not E) are similar to N
\pepsims{Q}{NE} % N and E (but not D) are similar to Q

\DNAGroups{GAR,CTY}

\DNAsims{A}{GR} % G and R are similar to A
\DNAsims{G}{AR} % A and R are similar to G
\DNAsims{R}{AG} % A and G are similar to R

\DNAsims{C}{TY} % T and Y are similar to C
\DNAsims{T}{CY} % C and Y are similar to T
\DNAsims{Y}{CT} % C and T are similar to Y

```

3. `\shadingmode[⟨seqref⟩]{diverse}`

1.5.3 depicts an example alignment. Choose the number or the name of the sequence `⟨seqref⟩` which will be treated as the consensus and to which the other sequences are compared. If no `⟨seqref⟩` is declared the first sequence is set as consensus (`⟨seqref⟩ = 1`).

Standard definitions for `diverse` mode are:

```

\nomatchresidues{Black}{White}{lower}{up}
\similarresidues{Black}{White}{lower}{up}
\conservedresidues{Black}{White}{{.}}{up}
\allmatchresidues{Black}{White}{{.}}{up}
\gapchar{-}

```

After calling `\shadingmode{diverse}` these commands can be used to redefine the `diverse` mode settings (mind the double curly braces around the dot-symbol!).

4. `\shadingmode[⟨type⟩]{functional}` There are six different functional shading modes available for peptide sequences; nucleotide sequences can not be shaded due to functional aspects. Four of `TeXshade`'s functional modes correspond to the four 'alphabets' employed by KARLIN and GHANDOUR for peptide alignments [2]. Additional 'alphabets' to the standard 20-letter array of amino acids can highlight peptide similarities which were otherwise not visible. For the 'alphabet' definitions see below:

- `⟨type⟩ = charge` Acidic (D, E) and basic (H, K, R).
- `⟨type⟩ = hydropathy` Acidic and basic (as above), polar uncharged (C, G, N, Q, S, T, Y) and hydrophobic nonpolar (A, F, I, L, M, P, V, W), see also KYTE and DOOLITTLE [3].
- `⟨type⟩ = structure` External (D, E, H, K, N, Q, R), internal (F, I, L, M, V) and ambivalent (A, C, G, P, S, T, W, Y).
- `⟨type⟩ = chemical` Acidic (D, E), aliphatic (A, G, I, L, V), amide (N, Q), aromatic (F, W, Y), basic (H, K, R), hydroxyl (S, T), imino (P) and sulfur (C, M).

The two modes described below highlight sidechain sizes and hydrophobicity, respectively, according to ROSE *et al.* [4,5]. Standard area stands for the surface area of the residue in Å², i.e. it is a measure for the size of a residue's sidechain. The accessible area value (also in Å²) gives information about the size of the surface area which is accessible by solvent molecules within the folded protein. A very small area means that the residue is strongly buried and is thus very hydrophobic. Hydrophilic residues in turn possess large accessible areas due to their preferred location at the protein surface. Therefore, this kind of shading provides another method, in addition to `hydropathy` and `structure`, for the visualization of structural protein properties.

- `⟨type⟩ = standard area` for the area values see legend of the alignment in 1.5.4
- `⟨type⟩ = accessible area` for values see 1.5.4

If no $\langle type \rangle$ or an unknown $\langle type \rangle$ is designated as option all functional groups and shading colors are cleared. This is also achieved by the command `\clearfuncgroups`. With all groups cleared one can start to build new shading modes from scratch. How to do this is explained in the next section.

In order to exchange the colors but to keep the group definitions and descriptions the command `\funcshadingstyle` can be employed. Usage:

```
\funcshadingstyle{\langle residue \rangle}{\langle res.col. \rangle}{\langle shad.col. \rangle}
{\langle case \rangle}{\langle style \rangle}
```

$\langle residue \rangle$ is one representative of the whole amino acid group. The colors which are declared by the next four parameters are used for all residues in this group. $\langle case \rangle$ and $\langle style \rangle$ are as described for example in `\nomatchresidues`.

4.2 Creating new functional shading modes

The grouping of amino acids due to other properties can make sense as suggested by KARLIN and GHANDOUR [2], e. g. physical properties (molecular weight, shape), kinetic properties (reaction velocity, Michaelis-Menton constant), or structure (α -helices, β -sheets, turns). New amino acid groups are defined with the `\funcgroup` command. This command needs six parameters:

```
\funcgroup{\langle descr \rangle}{\langle residues \rangle}{\langle res.col. \rangle}{\langle shad.col. \rangle}
{\langle case \rangle}{\langle style \rangle}
```

$\langle descr \rangle$ contains descriptive text which is displayed in the legend. The second parameter $\langle residues \rangle$ holds the amino acids to be grouped. The colors for the foreground and background are set with the following two parameters, the case and style is declared by the last parameters. The example below defines a functional group named ‘acidic (–)’ containing the amino acids aspartic and glutamic acid with white letters on a red background:

```
\funcgroup{acidic ($-$)}{DE}{White}{Red}{upper}{up}
```

For the usage of colors see section 5. Up to nine individual groups can be defined. New groups are simply added to the already existing groups, i. e. if an extension of the group definitions of an

existing shading mode is desired there is no need to clear these groups und re-define them again. Just add the new groups with the `\funcgroup` command. To create completely new modes use the command `\shadingmode{functional}` without an option *before* setting the new groups. The new definitions are active only in the functional shading mode—so be sure to have it switched on before setting the new groups. Remember, `\shadingmode{functional}` without an optional parameter clears all groups defined before, see above. The following example shows the definitions needed to produce an output which is identical to the functional mode ‘charge’:

```
\begin{texshade}{\langle alignmentfile \rangle}
  \shadingmode{functional}
  \funcgroup{acidic ($-$)}{DE}{White}{Red}{upper}{up}
  \funcgroup{basic ($+$)}{HKR}{White}{Blue}{upper}{up}
\end{texshade}
```

4.3 Appearance of the consensus line

An important parameter for the calculation of the consensus is the threshold percentage. Default setting is 50%, i.e. to become the consensus residue more than half of the residues at this position must be identical or similar, depending on the shading mode. Any percentage between 0 and 100 is allowed and can be set with `\threshold{\langle percentage \rangle}`, e.g. `\threshold{50}`.

Another possibility is to set one sequence of the alignment as consensus and compare the other sequences to this one. Therefore, the command `\constosingleseq{\langle seqref \rangle}` is provided. The `\langle seqref \rangle` selects the sequence to be used as consensus (numbering according to the appearance in the alignment file; top sequence is number 1, or use the sequence name). Nevertheless, the threshold percentage is also taken into account, i.e. with a threshold of 50% half of the sequences must be identical or similar compared to the specified consensus sequence in order to be shaded. With `\constoallseqs` the consensus is calculated considering all sequences (the case described in the paragraph above).

Consensus lines are displayed either on the top or at the bottom of the alignment by calling `\showconsensus{\langle position \rangle}` with `\langle position \rangle` **top** or **bottom**. To hide the consensus use `\hideconsensus`.

The consensus line is named ‘consensus’ in english texts, ‘consenso’ in spanish or ‘Konsensus’ if the `german.sty` is used. With `\nameconsensus{⟨name⟩}` any name can be set.

You can tell `TeXshade` which symbols or letters to use in the consensus line for different matching qualities by

```
\defconsensus{⟨symbol1⟩}{⟨symbol2⟩}{⟨symbol3⟩}.
```

The following parameters are allowed for symobols 1–3:

1. $\langle symbol1 \rangle$ = no match symbol (if below threshold)
 - any character or letter
 - `{ }` (empty braces) for blank space
2. $\langle symbol2 \rangle$ = conserved symbol (if threshold is exceeded)
 - `upper` (prints the consensus residue in uppercase)
 - `lower` (prints the consensus residue in lowercase)
 - any character or letter
 - `{ }` (empty braces) for blank space
3. $\langle symbol3 \rangle$ = all match symbol (if all residues match and `\allmatchspecial` is active)
 - see $\langle symbol2 \rangle$

Example: `\defconsensus{{}}{*}{upper}` does not show non matching residues in the consensus line, marks conserved residues with ‘*’, and displays the uppercase letter of the consensus residue at positions where all residues match.

Finally, the colors of the above defined symbols are adjustable by the command:

```
\consensuscolors{⟨res.col.1⟩}{⟨shad.col.1⟩}
                {⟨res.col.2⟩}{⟨shad.col.2⟩}
                {⟨res.col.3⟩}{⟨shad.col.3⟩}
```

The color definitions are in the same order as in the `\defconsensus` command:

1. $\langle res.col.1 \rangle$ = no match residue color (if below threshold)
 $\langle shad.col.1 \rangle$ = no match background color
2. $\langle res.col.2 \rangle$ = conserved residue color (if threshold is exceeded)
 $\langle shad.col.2 \rangle$ = conserved background color
3. $\langle res.col.3 \rangle$ = all match residue color (if all residues match and `\allmatchspecial` is active)
 $\langle shad.col.3 \rangle$ = all match background color

For colors which are not to be changed empty braces can be used.

Example:

```
\consensuscolors{}{}{Blue}{White}{Red}{Green}
```

Non matching symbol colors are not changed, conserved residues are displayed blue on white and where all residues match red symbols on green ground are displayed in the consensus line.

4.4 Appearance of the sequence lines

4.4.1 Names, numbers and gaps

Many parameters that influence the appearance of the actual sequence lines can be changed for customization. Thus, the sequence names and numbering can be shown and placed either left or right by

```
\shownames{\langle position \rangle}
```

```
\shownumbering{\langle position \rangle}
```

with $\langle position \rangle$ set to `left` or `right`. Both, names and numbering can be displayed on the same side.

`TeXshade` uses the sequence names from the alignment input file. This can cause some problems during the `TeX`-run when special characters are present in those names! `TeXshade` does not accept the following characters in sequence names: `\ { } @` spaces and the tilde. Those have to be replaced in the input file. The characters `#` and `%` can only be used with a leading backslash, e.g. `\#`. This must also be changed in the input file. All other special characters should be displayed properly.

Sequence names that are accepted by `TeXshade` can further be changed in the `texshade` environment:

```
\nameseq{\langle seqref \rangle}{\langle name \rangle}
```

`<seqref>` selects the sequence whose name is to be changed. The basis for the `<seqref>` is the appearance in the alignment input file with the top sequence = 1, or the old name. The colors are set by `\namescolor{<color>}` and `\numberingcolor{<color>}`, respectively.

In order to change the colors only of some sequence names or numbers the commands `\namecolor{<seq1>, ... ,<seq n>}{<color>}` and `\numbercolor{<seq1>, ... ,<seq n>}{<color>}` are provided.

In order to hide all names or the numbering use the command `\hidenames` or `\hidenumbering`. If only the names or numbers of some sequences should be hidden apply

`\hidename{<seq1>, ... ,<seq n>}` or
`\hidenumber{<seq1>, ... ,<seq n>}`, respectively.

In some situations, e.g. when only sections of sequences are displayed, one may not want to have the residue numbering start out with number 1. The command `\startnumber{<seqref>}{<first residue number>}` allows one to set the starting number of any sequence to any value incl. negative values but except '0' which is not used in sequence numbering (the transition from negative to positive values is like this: ... -2, -1, 1, 2 ...).

TeXshade needs to know the correct length of the sequences to be able to break them right after the last residue. If MSF files are used as an input the length is already given but the calculation is usually wrong because the gaps are also counted. Thus, **TeXshade** counts the number of residues during each run by itself and stores the values in the `.aux` file. That means that it needs two runs to get the numbers right. Again, this is only important if the gap symbol after the sequence end should be suppressed, see below (`\hideleadinggaps`). If you know the correct length of the sequences you can use the command

`\seqlength{<seqref>}{<length>}`

in order to set the values by hand and have the gaps break properly already in the first **TeX** run.

Example: `\seqlength{1}{346}` means that sequence no. 1 is 346 residues long.

TeXshade can display a section of the complete alignment without the need to edit the alignment input file or even to re-calculate the entire alignment. This allows one to use one single alignment of the full length proteins or open reading frames for multiple visualizations of different sections in a document as done in this manual. Thus, the file `AQPpro.MSF` contains the full-length multiple protein alignment of

five aquaporins but only sections are displayed as examples in 1.5.1 through 1.5.4. The definition of a section is done by

```
\setends{<seqref>}{<startnumber>..<stopnumber>}
```

Again, $\langle seqref \rangle$ is the sequence number based on the appearance in the alignment file, or the name; further, in order to use the consensus as a measure for the sequence section the string ‘consensus’ as $\langle seqref \rangle$ is accepted. The specified sequence is truncated at positions $\langle startnumber \rangle$ and $\langle stopnumber \rangle$. All other sequences are cut accordingly. If the number of the first residue in the sequence is set to a new value with the `\startnumber` command (s. a.) this is taken into account. Some examples:

a) `\setends{1}{20..100}`

b) `\startnumber{1}{15} \setends{1}{35..115}`

Both commands select the same section from the alignment but the numbering for sequence 1 starts at position 20 in the first example and at position 35 in the latter.

c) `\setends{consensus}{20..100}`

This may describe a completely different section of the multiple sequence alignment.

Another possibility to label sequence positions is to switch on a ruler on the top or at the bottom of the sequence block using `\showruler{<position>}{<seqref>}`. The residue ruler of one sequence $\langle seqref \rangle$ or the consensus (declare ‘consensus’ as $\langle seqref \rangle$) can be displayed at $\langle position \rangle$ **top** or **bottom**. The ruler is hidden with `\hideruler`. The steps between two numbers are set by `\rulersteps{<number>}`. In order to change the color which is used for the ruler write `\rulercolor{<color>}`.

Further, the symbol which is displayed in sequence gaps is freely selectable with `\gapchar{<symbol>}`. $\langle symbol \rangle$ can be any character or symbol. If math symbols are to be used math mode must be activated by $\$$ characters, i.e. `\gapchar{${\triangle}$}`. Note the double curly braces in the last command. Everytime a ‘complex’ character is used, i.e. a character definition consisting of more than one letter, it must be braced in order to be interpreted as one character. One exception is `\gapchar{rule}`; with this parameter lines are drawn in the sequence gaps with a certain thickness defined by `\gaprule{<thickness>}`, e.g. `\gaprule{1.5pt}`. The colors of the gaps and gap symbols are set by `\gapcolors{<symbol color>}{<background color>}`.

There are some discussions whether or not to display gap symbols before and after the actual sequence. Since v1.3a one can control the appearance of those gap symbols by the commands `\showleadinggaps` and `\hideleadinggaps`. By default, leading gaps are indicated by symbols despite my personal thinking that it could suggest that there are some not displayed residues upstream resp. downstream of the gap.

4.4.2 Hiding, killing, separating and ordering

If one or more sequences from the alignment input file should be used for the calculation of the consensus but it is desired not to display these sequences in the final output use the command `\hideseq{<seq1>,<seq2>,...,<seq n>}`. For consecutive sequence numbers a dash can be used, e.g. `\hideseq{1-3}` instead of `\hideseq{1,2,3}`. Decending series are also permitted, e.g. `\hideseq{3-1}`. This command allows one for example to hide the sequence which has been defined as the consensus sequence with `\constosingleseq`. In order to completely exclude sequences the command `\killseq{<seq1>,<seq2>,...,<seq n>}` is provided. Again, for number series the dash can be used (s.a.). The designated sequences are neither displayed nor considered for the calculation of the consensus. This is another possibility to comment out sequences in addition to the use of an exclamation point in front of the **Name**: definition in an MSF-file (see figure on page 22).

The command `\donotshade{<seq1>,<seq2>,...,<seq n>}` makes one or more sequences (remember the dash, s.a.) appear unshaded in black letters on white background. This does not influence any other sequences or the consensus calculation.

If a very graphical output of the sequences is desired, the residue symbols or letters can be blanked out by `\hideresidues`. Now, only the shaded boxes are printed. In combination with `\gapchar{rule}` one obtains alignments in a style à la Mondrian. The residues reappear with `\showresidues`.

If an alignment contains members of several subgroups of a protein or a gene family it may be rather helpful to visualize the group divisions by a separation line. Therefore, the command `\separationline{<seqref>}` is applicable. This command inserts vertical space after the sequence which is referred to by `<seqref>`. How much space is inserted is defined by one of the following commands: `\smallsep`, `\medsep` (default) or `\bigsep`. These lengths corre-

spond to the known `\small-`, `\med-` and `\bigskip` commands. With `\vsepspace{⟨length⟩}` any length with any T_EX unit can be assigned, e.g. `\vsepspace{2mm}`.

The sequence order given by the alignment input file is easily reorganized by `\orderseqs{⟨seq1⟩,⟨seq2⟩,...,⟨seq n⟩}` without the need for editing the alignment input file (which would be a big copy’n’paste job). Make sure that all sequences are assigned in this command. If there are more sequences present than numbers or names in the command an error message will occur. Here also, the dash can be used for sequence number series. Example: `\orderseqs{1-3,6-4,7}` is equivalent to `\orderseqs{1,2,3,6,5,4,7}`. Reordering of sequences only changes the output; all commands using the parameter `⟨seqref⟩` are not influenced, because `⟨seqref⟩` always corresponds to the appearance in the alignment file. Thus, to completely reverse the order of a five sequence alignment simply type `\orderseqs{5-1}`.

4.4.3 Residues per line and further settings

By default T_EX`shade` puts the highest possible by five divisible number of residues in one line depending on the `\textwidth`. With `\residuesperline{⟨number⟩}` a new value can be set. If this value exceeds the highest possible number of residues per line it is ignored; lower values are accepted of course. But also in the latter case the number of residues printed per line is rounded such to be divisible by five. To force T_EX`shade` to set lines with exactly the desired number of residues use the asterisk-extended command `\residuesperline*{⟨number⟩}`. Expect multiple *overflow hbox* errors after this command, because in this mode T_EX`shade` does not check the length of the lines any more.

T_EX`shade` calculates the dimensions of a shaded box from the width and height of the uppercase letter ‘M’ and the depth of the lowercase ‘g’. Depending on the font used for the sequence residues the box dimensions might not be fully satisfactory. With `\charstretch{⟨factor⟩}` and `\linestretch{⟨factor⟩}` the width and height/depth, respectively, of the boxes can be multiplied individually by a `⟨factor⟩` to stretch (> 1) or shrink (< 1) the dimensions.

The reserved space for the sequence numbering is set by the command `\numberingwidth{⟨n digits⟩}`. Here, the default setting is four-digit numbering, i.e. –999 through 9999. If this range is to be changed assign the desired number as parameter `⟨n digits⟩`, e.g. `\numberingwidth{111111}` reserves space for 6 digit numbering.

The vertical space between the sequence blocks can be controlled by the commands `\smallblockskip`, `\medblockskip` (default setting), `\bigblockskip` or `\noblockskip`. Further, the command `\vblockspace{<length>}` allows one to set a defined space length using any T_EX unit, e.g. `\vblockspace{0.4in}`.

Two more commands set the space between the sequence blocks to be flexible (`\flexblockspace`) (default) or fixed (`\fixblockspace`). Flexible means, that only the vertical white space between the blocks is kept to the settings by e.g. `\medblockskip`. This results in flexible space between the actual blocks depending on the presence of feature lines. When switching to fixed space the distance of the blocks is kept constant by using more white space between blocks without feature lines. Thus, a difference between flexible and fixed space will only be noticeable when features are used.

The position of the output can be aligned left, right or centered on the page by `\alignment{<position>}` with the *<position>* parameter `left`, `center` or `right`.

4.4.4 Fingerprinting

An easy way to gain an overview on complete alignments is provided by displaying a so called alignment ‘fingerprint’. In this style the whole sequence can be shown in one line. Due to the lacking space the residue names are hidden and the shaded boxes are reduced to thin vertical colored lines. The command `\fingerprint{<res. per line>}` takes one argument stating the desired number of residues per line, e.g. `\fingerprint{1000}`. All T_EX`shade` commands are compatible with `\fingerprint`, i.e. all shading modes are applicable for displaying overviews on similarity or every functional aspect. Also, all kinds of labeling—as described in the following—work with this command.

4.5 Individual shading and labeling of sequence stretches

Computer calculated alignment shading is informative—but even more information can be visualized by manual labeling of positions and regions of interest with different colors, text styles or graphical marks and descriptive text. All this is provided by easy to handle T_EX`shade` commands.

4.5.1 Manual shading of regions and blocks

Besides the shading calculated by `TeXshade` any region can be shaded manually with a color specified by the user. This is very useful to highlight secondary protein modification sites such as phosphorylation or glycosylation sites or longer motifs for example protein/protein interaction sites. This is done by the use of the following command:

```
\shaderegion{<seqref>}{<start1>..<stop1>,<start2>..<stop2>,  
...,<start n>..<stop n>}{<res.col.>}{<shad.col.>}
```

Example: in order to shade residue number 13 and the region 20–30 of sequence number 1 in red letters on green ground type the following command:

```
\shaderegion{1}{13..13,20..30}{Red}{Green}
```

If the consensus is to be shaded use `consensus` as `<seqref>`.

In analogy to `\shaderegion` which is restricted to one single sequence `\shadeblock` shades the corresponding region in all other sequences as well except the consensus. If also the consensus is to be shaded define the region using `consensus` as `<seqref>`.

```
\shadeblock{<seqref>}{<start1>..<stop1>,<start2>..<stop2>,  
...,<start n>..<stop n>}{<res.col.>}{<shad.col.>}
```

4.5.2 Emphasizing regions and blocks

If it is preferred to keep the calculated shading colors but distinct regions or blocks are yet to be emphasized use the following commands to change the font style of such regions:

```
\emphregion{<seqref>}{<start1>..<stop1>,<start2>..<stop2>,  
...,<start n>..<stop n>}
```

and

```
\emphblock{<seqref>}{<start1>..<stop1>,<start2>..<stop2>,  
...,<start n>..<stop n>}
```

Which style `TeXshade` uses for emphasizing regions is defined by `\emphdefault{<style>}`. Default setting is the *italics* font shape (set by `\emphdefault{it}`). In order to change this setting choose one of the styles `bf`, `md`, `up`, `it`, `sl`, `rm`, `sf`, `tt`.

Example: `\emphdefault{bf}`

4.5.3 Graphical labeling of sequence features

The `\feature` command is designed to fulfill most needs for the graphical labeling of sequence stretches and the setting of descriptive text. It needs five parameters:

```
\feature{<position>}{<secref>}{<start1>..stop1,  
          <start2>..stop2},...,<start n>..stop n}{<labelstyle>}{<text>}
```

In the following paragraphs all possible parameter settings of this rather complex but mighty command are discussed in detail. The parameter `<position>` tells `TeXshade` where to display the feature label, i.e. on the top of the alignment (`top`), or at the bottom (`bottom`). Further, there can be a feature line on the top of the top feature line (`ttop`) or below the bottom feature line (`bbottom`). Thus, up to four features overlapping in four different lines may be displayed. Depending on the content of the feature lines the gaps between them might be not satisfactory. Therefore, four separate commands can be employed to change the space between `ttop` and `top` (`\ttopspace{<length>}`), between `top` and the alignment (`\topspace{<length>}`), between the alignment and `bottom` (`\bottomspace{<length>}`) and between `bottom` and `bbottom` (`\bbottomspace{<length>}`). Use positive values to further separate the lines, e.g. `\ttopspace{3mm}` or negative values to reduce the space, e.g. `\bottomspace{-0.1in}`.

The argument `<secref>` and the third parameter containing the definitions of the specified regions are identical to the ones described before in several commands, e.g. `\ruler` (4.4.1) or `\shaderegion` (4.5.1). New is the fourth parameter for the definition of the label style. There are many possibilities like braces, helices, boxes, arrows, bars, any fill character, bar graphs, color scales or even translations of the specified regions.

Braces:

In order to display an over- or underbrace as a label use the parameter `{brace}`. Depending on the `<position>` (`ttop`, `top`, `bottom` or `bbottom`) the respective brace is displayed. The standard color of braces is black. It can be changed by an optional parameter directly after the definition of the symbol, e.g. `{brace[Red]}`.

Protein α -Helices:

The parameter `{helix}` will plot a symbolized α -helix as a label. The standard color of the helix spiral is black. It can be changed by an

optional parameter directly after the definition of the symbol, e.g. `{helix[Red]}`.

Filling a stretch with a symbol:

A region can be filled with any character for labeling purposes using the parameter `{fill:<symbol>}`. The `<symbol>` is freely selectable; the usage is like in `\gapchar` (4.4.1). Do not use spaces before or after the expression `<symbol>`; this will shift the symbols to the respective direction. The standard color of the fill symbol is black. It can be changed by an optional parameter directly after the definition of the symbol, e.g. `{fill:$\bullet$[Red]}`.

The `\feature` command does not like special characters in text mode, e.g. `\dag`. One has to use the math version of those symbols between `$`-signs. The following quite common text symbols have also a math equivalent¹¹:

<i>symbol</i>	<i>command</i>	<i>description</i>
†	<code>\$_daggers\$</code>	dagger
‡	<code>\$_ddaggers\$</code>	double dagger
¶	<code>\$_mathparagraph\$</code>	paragraph mark
§	<code>\$_mathsection\$</code>	section mark
\$	<code>\$_mathdollar\$</code>	dollar
{	<code>\$_lbraces\$</code>	left brace
}	<code>\$_rbraces\$</code>	right brace

Boxes:

Boxed text is printed using the parameter `{box:<text>}`. By default black letters in a white framed box are displayed. In order to change these colors optional parameters can be included in the argument: `{box[<boxcolor>]:<text>[<textcolor>]}`.

Example: `{box[Blue]:$_alpha$~helix[Yellow]}`

Horizontal bars and arrows:

For displaying bars and arrows a simple selection scheme consisting of three consecutive characters is used as the `<labelstyle>` parameter. Each bar or arrow is defined by its left end, the middle part, and the right end. The following table gives some examples for the construction of arrows and bars.

¹¹Thanks to Darrell Conklin for giving notice to me about this problem

middle left end ↓ right end	
---	plain bar
===	double bar
-->	right arrow
'->	right arrow with up hook
<-	left <i>maps to</i> arrow
<=>	double arrow, two heads
,-,	plain bar with down hooks
=	double bar with vertical ends

All combinations of the left-end-characters ($-=<' , |$), the middle-characters ($-$), and the right-end-characters ($-=>' , |$) are allowed and produce the desired arrow or bar. The color is changed as described above. For plain bars the thickness can be set by the separate command `\featurerule{<length>}` with any T_EX measure as *<length>*, e.g. `\featurerule{3pt}`.

Sequence translations:

With the option `{translate}`, sequence stretches can be translated from nucleotide to peptide sequences as well as backtranslations from peptide to nucleotide sequences are possible. Default setting for the translations is the standard genetic code. Of course, the codons can be re-defined by the user. The command `\codon{<amino acid>}{<triplet1, ..., triplet n>}` has been implemented for this issue. The usage is simple. Replace *<amino acid>* by the single letter code of the amino acid to be defined and add a list of triplets for this residue. Example definition for the amino acid *alanine*:

```
\codon{A}{GCA,GCG,GCC,GCT,GCU,GCN}
```

Note the last triplet in the list. It contains an ambiguity code N which stands for *any* nucleotide. This triplet has been added at the last position because the last triplet is used for the generation of the backtranslated nucleotide sequence from a peptide. Two files are included in the T_EXshade distribution as examples (`standard.cod`, `ciliate.cod`). If you want to define a new genetic code store your commands in a file like the examples. Such files with the suffix `.cod` can be loaded in the T_EXshade environment by `\geneticcode{<filename>}`, e.g. `\geneticcode{ciliate}`. Do not designate the suffix `.cod` in *<filename>*. Please note, when inspecting the example files, that only the exchanges compared to the standard code need to be defined in a new genetic code file.

When DNA sequences are translated to protein the resulting amino acids are aligned to the second nucleotide of each triplet. It is more difficult to produce a satisfactory display of backtranslated nucleotide sequences due to the lack of space. You need thrice as much space than the original peptide sequence, because single letter amino acid code is translated to a triplet code. Therefore, the user can choose from five display styles for backtranslations depending on personal preferences:

`\backtranslabel[ $\langle size \rangle$ ]{ $\langle style \rangle$ }`, with

```
{ $\langle style \rangle$ } = {horizontal}
              = {alternating}
              = {zigzag}
              = {oblique}
              = {vertical}
```

$\langle size \rangle$ can be any $\text{\TeX}$ size from `tiny` up to `Huge`, but `tiny` is recommended (and default setting). Translations can be colored as all other labels, see above.

Bar graphs and color scales:

Sequence related numeral data, such as hydropathy or solvent accessibility data etc., can be shown in a feature line as bar graphs or color scales. The data are either provided in a separate file (highly recommended) or may be entered by hand in the `\feature` command. The format of a data file is simple: every value must appear in a separate line. Comments are allowed, because `\text{\TeXshade}` ignores all lines starting with a letter (avoid a ‘-’ as the first character of a comment line as this is interpreted as a negative number). Make sure that there are as many values as residues in your sequence. `\text{\TeXshade}` will read this file and determine the minimal and maximal values. These data are then normalized for plotting. Due to $\text{\TeX}$ ’s limited calculation capabilities no values above 10 737 are allowed and the difference between minimum and maximum must not exceed this very number. Values below 0.001 may be susceptible to major rounding errors. Thus, try to provide your data already normalized to moderate scales, e.g. 0.0–1.0 or -100–100. Data, which is directly entered in the `\feature` command must be normalized to integer values with a maximal difference of 100 between the highest and lowest value, e.g. -50–50 or 0–100.

The range to be plotted can be set by hand as an optional parameter in the `\feature` command. This can be necessary when the data file

contains values between e.g. -0.44 and 0.87 . Without help `TeXshade` will assume -0.44 as minimum and 0.87 as maximum. But if the actual range to be plotted should be $-1.0-1.0$ this needs to be set manually, see examples below. Be aware of the fact, that if you define a scale by hand, which is more narrow than the values of the input, this will stretch the bars accordingly. It is NOT recommended to use this method for stretching bars vertically. Instead another command has been introduced. The plotted bars can be stretched by a factor if the appearance is not as desired: `\bargraphstretch{<factor>}`. Here, the factor is multiplied with the bar length, e.g. `\bargraphstretch{2}` will double the bar height, `\bargraphstretch{0.5}` will make them half as high.

The default color of bar graphs is gray and can be changed by an optional parameter at the end of the `label` definition. Default for color scales is a 5% gray scale from very light gray to black (`Gray`). Three more colorful scales have been implemented, i.e. `BlueRed`, `GreenRed` and `ColdHot`, the latter being particularly useful for ranges from negative to positive values.

The general format of this feature label definition for bar graphs is:

```
{bar[<min>,<max>]:<data/file>[<color>]}
```

and for color scales:

```
{color[<min>,<max>]:<data/file>[<scale>]}
```

Some examples:

```
{bar:10,20,30,40,50[Red]}
{bar[-20,40]:-10,0,10,20,30[Red]}
{bar:scale.txt}
{bar[-10,10]:scale.txt[Red]}
{color[-10,10]:scale.txt[ColdHot]}
{color[-0.1,0.1]:otherscale.txt[ColdHot]}
```

No graphical label, only text:

If no graphical label is wanted the fourth parameter of `\feature` can be empty braces.

Finally, the fifth parameter of the `\feature` command contains the descriptive text for the labeled region. Type whatever you want incl. symbols and math chars. The text field can also contain sequence translations. In this case just set `<text> = {translate}`. There is a

command for setting the size and style of backtranslated sequences in the feature $\langle text \rangle$ which corresponds to the one described above:

```
\backtranstext[ $\langle size \rangle$ ]{ $\langle style \rangle$ }
```

Again, the color can be set by an optional parameter appended to the text. For how to change the font size of text or symbols in the feature style line (**featurestyles**) or the in descriptive text line (**features**) see section 4.8.1, page 50.

Examples for the appearance of features are given in the overview section (1), see:

similarity mode (1.5.2): fill-character; here, only one position is labeled. It is also possible to label a longer stretch, then, the character is printed several times to fill the specified region.

```
\feature{top}{1}{93..93}{fill:$\downarrow$}{first...}
\feature{bottom}{1}{98..98}{fill:$\uparrow$}{second...}
```

diversity mode (1.5.3): no graphical label, text only

```
\feature{top}{1}{77..109}{-}{AQP2 species variants}
```

functional mode (1.5.4): bar graph, color scale, box, arrow, translation, brace, helix

```
\feature{top}{3}{153..165}
    {bar[-50,50]:-50,-45,-40,...,40,45,50}{-}
\feature{top}{3}{167..186}
    {color:5,10,15,...,90,95,100[ColdHot]}{-}
\feature{top}{1}{138..157}
    {box[Red]:$\alpha$~helix[Yellow]}
    {transmembrane domain 4}
\feature{top}{1}{164..170}{-,->[Red]}{trans. dom. 5}
\feature{top}{1}{158..163}{translate[Blue]}{-}
\backtranslabel{oblique}
\feature{bottom}{1}{158..163}
    {brace[Blue]}{loop D[Blue]}
\feature{top}{1}{138..157,164..170}{helix}{membr.}
```

4.5.4 Including secondary structure information

The DSSP [7], STRIDE [8] and PHD [9] algorithms produce very reliable secondary protein structure predictions. PHD files contain both, secondary structure information and topology data. This information can be displayed in an alignment by one of the commands:

<code>\includeDSSP</code>	sec. structure calculated by DSSP
<code>\includeSTRIDE</code>	sec. structure calculated by STRIDE
<code>\includePHDsec</code>	sec. structure calculated by PHD
<code>\includePHDtopo</code>	topology data calculated by PHD

The syntax is `\includeDSSP{<seqref>}{<filename>}`, with **seqref** indicating the number or name of the sequence for which the secondary structure data is calculated and **filename** designating the corresponding structure file to be included.

Several types of secondary structures are predicted by these programs; in order to designate them in **TeXshade** use the names from the right column:

secondary structure	designation
<i>DSSP and STRIDE</i>	
4-helix (α -helix)	alpha
isolated β -bridge	bridge
extended strand (β -strand)	beta
3-helix (3_{10} -helix)	3-10
5-helix (π -helix)	pi
H-bonded turn	turn
<i>PHDsec</i>	
helix	alpha
sheet	beta
<i>PHDtopo</i>	
internal region	internal
external region	external
transmembrane domain	TM

By default all three types of helices and the strands are displayed whereas turns and bridges are skipped. If it is desired to show them as well, call for example `\shownonDSSP{bridge,turn}`. In analogy to this example all structure features can be activated in DSSP, STRIDE, PHDsec and PHDtopo. In order to hide certain structure types use for example `\hideonDSSP{3-10,pi}`.

The DSSP format has two columns of sequence numberings. The first column is consecutive, whereas the second column contains the actual sequence numbering. This can be different from the first column when sequence parts are missing in the DSSP file. One can choose

which column will be read by `TeXshade` by ‘`\firstcolumnDSSP`’ and ‘`\secondcolumnDSSP`’. The second column is still default.

One word on PHD predictions: when starting the PHD software do not restrict the prediction to secondary structure or topology alone. This leads to changes in the PHD output file which are not correctly interpretable by `TeXshade` due to ambiguities. There is no way around it—thus, run the full prediction.

Now, some information about how `TeXshade` extracts and displays secondary structure features. In short, it is a two step process. First, `TeXshade` analyzes the secondary structure file and extracts all necessary data. This data is converted into a format which is readable and processable by `TeXshade` using the `feature` command (see 4.5.3). This command allows one to label sequence stretches graphically. For a detailed explanation see the indicated reference. A list of feature commands is saved in a file with the ending ‘`.sec`’ for DSSP, STRIDE and PHDsec or ‘`.top`’ for PHDtopo. Then, in a second step, this file is loaded again and executed. When `TeXshade` encounters this file a second time, i.e. in a second `TeX` run, it uses the already existing file for the output. The great advantage of this method is its flexibility. Due to the simple reason that the feature file can be edited in the meantime. Thus, the user has the ability to change the computer generated file according to his personal needs. On the other hand, one can force `TeXshade` to write a new file every time by the optional argument `[make new]` in the include command, e.g. `\includePHDsec[make new]{1}{AQP.phd}`.

Finally, the appearance of the feature labels can be assigned by the command

`\appearance{<filetype>}{<type>}{<position>}{<labelstyle>}{<text>}`.

Here, `<filetype>` stands for one of the following secondary structure file types: DSSP, STRIDE, PHDsec or PHDtopo and `<type>` designates the secondary structure type as shown in the right column of the table above. The other arguments `<position>`, `<labelstyle>` and `<text>` are almost as described in 4.5.3. One further possibility is to include internal counters for each secondary structure type. Just add one of the following commands to the text in the feature description.

<i>counter</i>	<i>display</i>
<code>\numcount</code>	1, 2, 3 ...
<code>\alphacount</code>	a, b, c ...
<code>\Alphacount</code>	A, B, C ...
<code>\romancount</code>	i, ii, iii ...
<code>\Romancount</code>	I, II, III ...

Examples:

```
\appearance{DSSP}{alpha}{ttop}
      {-->}{$\alpha$-helix~\Alphacount}
\appearance{PHDtopo}{TM}{bottom}
      {box[Blue]:TM\numcount[Yellow]}{}
```

4.6 Displaying and building legends

For each predefined shading mode `TeXshade` can print an appropriate legend to explain the used shading colors. The commands `\showlegend` and `\hidelegend` display or clear the legend at the end of the alignment. The legend is displayed by default beneath the first residue of the last alignment line. The location can be changed by `\movelegend{<x-offset>}{<y-offset>}`. Both parameters require a TeX length, e.g. `\movelegend{5cm}{-2cm}` moves the legend 5 cm to the right and 2 cm up.

The language for the descriptions is english by default; if the `\german.sty` package is active legend texts are in german. So far, german, spanish and english are implemented. With the commands `\germanlanguage`, `\spanishlanguage` and `\englishlanguage` switching between the languages is made possible. For the addition of other languages contact me. Finally, the color of the describing legend texts can be set with the command `\legendcolor{<color>}`.

User defined legends are easily built with the following command `\shadebox{<color>}`. Use this command outside the `TeXshade` environment, e.g. in the text or in the caption. As `<color>` any color can be designated (see section 5) or one of the following parameters:

- `nomatch` = the color used for nonmatching residues
- `similar` = the color used for similar residues
- `conserved` = the color used for conserved residues

- `allmatch` = the color used for the case that all residues match (if `\allmatchspecial` is active)

The command simply prints a shaded box in the specified color then a describing text can be appended. Examples:

```
\shadebox{nomatch}---nonmatching residues
\shadebox{similar}: similar residues
\shadebox{conserved}~conserved residues
\shadebox{Yellow}\quad PKA phosphorylation sites
```

4.7 Adding captions to the alignment

Since `TeXshade` v1.5 captions can be added to the alignment. So far, captions were difficult to use when the alignment was bigger than one page and therefore did not fit into a figure environment. The `TeXshade` captions behave exactly as normal figure captions. They adopt their style, use the figure counter number and appear in the list of figures as any other figure.

The usage is slightly different from normal captions but intuitive:

```
\showcaption[<position>]{<text>}
```

The optional *<position>* tells `TeXshade` to put the caption on **top** or at the **bottom** of the alignment. If nothing is stated here the caption will appear at the bottom. The parameter *<text>* just holds the caption text as in the normal `\caption`. The command can be used at any position within the `texshade` environment. A simple example would be:

```
\showcaption{A beautiful \TeXshade{} alignment.}
```

4.8 Font handling

4.8.1 Changing font styles

The font styles for the numbering, the sequence names, the sequence residues, the descriptive feature texts and the legends can be changed by several commands.

```
\setfamily{<text>}{<family>}
\setseries{<text>}{<series>}
\setshape{<text>}{<shape>}
\setsize{<text>}{<size>}
```

The first parameter selects the text whose style is to be changed. Possible first parameters are `numbering`, `names`, `residues`, `features`, `featurestyles` and `legend`.

The style is set by the second parameter:

command	<i><2. parameter></i>	
<code>\setfamily</code>	<code>rm</code>	modern roman font family
	<code>sf</code>	sans serif font family
	<code>tt</code>	typewriter font family
<code>\setseries</code>	<code>bf</code>	bold face series
	<code>md</code>	normal series
<code>\setshape</code>	<code>it</code>	italics shape
	<code>sl</code>	slanted shape
	<code>sc</code>	small capitals shape
	<code>up</code>	upright shape
<code>\setsize</code>	<code>tiny</code>	the known T _E X sizes
	<code>scriptsize</code>	
	<code>footnotesize</code>	
	<code>small</code>	
	<code>normalsize</code>	
	<code>large</code>	
	<code>Large</code>	
	<code>LARGE</code>	
	<code>huge</code>	
	<code>Huge</code>	

Example: `\setfamily{features}{it} \setseries{features}{bf}`

With the command

`\setfont{<text>}{<family>}{<series>}{<shape>}{<size>}`

all four font attributes of one *<text>* can be changed simultaneously. The order of the parameters is as indicated.

Example: `\setfont{features}{rm}{it}{bf}{normalsize}`

Further, short commands are provided to change single font attributes quickly. The following commands set attributes of feature texts.

```

\featuresrm   \featurestiny
\featuressf   \featuresscriptsize
\featurestt   \featuresfootnotesize
\featuresbf   \featuresmall

```

<code>\featuresmd</code>	<code>\featuresnormalsize</code>
<code>\featuresit</code>	<code>\featureslarge</code>
<code>\featuressl</code>	<code>\featuresLarge</code>
<code>\featuressc</code>	<code>\featuresLARGE</code>
<code>\featuresup</code>	<code>\featureshuge</code>
	<code>\featuresHuge</code>

Corresponding sets are provided for the numbering (`\numberingrm` etc.), featurestyles (`featurestylesrm` etc.), names (`\namesrm` etc.), residues (`\residuesrm` etc.) and legend texts (`legendrm` etc.).

4.8.2 Using PostScript fonts

As already mentioned `TeXshade` makes intensive use of POSTSCRIPT for shading. Now, that POSTSCRIPT output is active anyway, including POSTSCRIPT fonts is very easy. Just declare in the document header

```
\usepackage{<PS-font>}
```

The typewriter font of `TeX` is always a topic of discussions. By including the package `\usepackage{courier}` `TeX`'s typewriter font is replaced by the widely accepted `COURIER`. Have a look into the directory `..texinputs:latex:psnfss`; there, some styles are located which exchange the common `TeX` fonts by POSTSCRIPT fonts, e.g. `avant.sty`, `bookman.sty`, `chancery.sty`, `courier.sty`, `helvet.sty` or `utopia.sty`. Depending on the style used the `\rmdefault`-, `\sfdefault`-, and `\ttdefault` fonts are substituted partly or completely. Thus, `courier.sty` for instance exchanges only the typewriter font, whereas `bookman.sty` sets `BOOKMAN` as `\rmdefault`, `AVANTGARDE` as `\sfdefault` and `COURIER` as `\ttdefault`. For further information see TOMAS ROKICKI's `dvips` manual [10].

4.9 Goodies—molweight and charge

During the process of sequence setting `TeXshade` sums up the molecular weight and charge of the aligned proteins. This data can be accessed by the following two commands.

```
\molweight{<seqref>}{<Da/kDa>}
\charge{<seqref>}{<i/o/N/C>}
```

The first parameter `<seqref>` selects the sequence. The second parameter in the `\molweight` command allows one to switch the units

between Dalton (Da) and kilo-Dalton (kDa). The `\charge` command needs the second parameter for the correct consideration of the charged protein termini. Thus, 'i' refers to internal sequences, 'o' to the overall charge, 'N' to N-terminal sequence parts, and 'C' to the C-terminal end of a protein.

Example: Charge: `\charge{1}{o}`; Weight: `\molweight{1}{Da}`

5 The PostScript color selection scheme

POSTSCRIPT provides 64 standard colors. All these colors are pre-defined in the `color.sty`. Each color has a pictorial name such as `Bittersweet` and a distinct composition, e.g. 0% cyan + 75% magenta + 100% yellow + 24% black—the so-called CMYK scheme. `TeXshade` enhances this color scheme by gray scales in 5% steps. The following colors and grays can be used in `TeXshade` by simply declaring the name of the color in the respective command, e.g. `\consensuscolors`:

<i>name</i>	<i>CMYK</i>	<i>name</i>	<i>CMYK</i>
GreenYellow	0.15,0,0.69,0	Yellow	0,0,1,0
Goldenrod	0,0.10,0.84,0	Dandelion	0,0.29,0.84,0
Apricot	0,0.32,0.52,0	Peach	0,0.50,0.70,0
Melon	0,0.46,0.50,0	YellowOrange	0,0.42,1,0
Orange	0,0.61,0.87,0	BurntOrange	0,0.51,1,0
Bittersweet	0,0.75,1,0.24	RedOrange	0,0.77,0.87,0
Mahagony	0,0.85,0.87,0.35	Maroon	0,0.87,0.68,0.32
BrickRed	0,0.89,0.94,0.28	Red	0,1,1,0
OrangeRed	0,1,0.50,0	RubineRed	0,1,0.13,0
WildStrawberry	0,0.96,0.39,0	Salmon	0,0.53,0.38,0
CarnationPink	0,0.63,0,0	Magenta	0,1,0,0
VioletRed	0,0.81,0,0	Rhodamine	0,0.82,0,0
Mulberry	0.34,0.90,0,0.02	RedViolet	0.07,0.90,0,0.34
Fuchsia	0.47,0.91,0,0.08	Lavender	0,0.48,0,0
Thistle	0.12,0.59,0,0	Orchid	0.32,0.64,0,0
DarkOrchid	0.40,0.80,0.20,0	Purple	0.45,0.86,0,0
Plum	0.50,1,0,0	Violet	0.79,0.88,0,0
RoyalPurple	0.75,0.90,0,0	BlueViolet	0.86,0.91,0,0.04
Periwinkle	0.57,0.55,0,0	CadetBlue	0.62,0.57,0.23,0
CornflowerBlue	0.65,0.13,0,0	MidnightBlue	0.98,0.13,0,0.43
NavyBlue	0.94,0.54,0,0	RoyalBlue	1,0.50,0,0
Blue	1,1,0,0	Cerulean	0.94,0.11,0,0
Cyan	1,0,0,0	ProcessBlue	0.96,0,0,0
SkyBlue	0.62,0,0.12,0	Turquoise	0.85,0,0.20,0
TealBlue	0.86,0,0.34,0.02	Aquamarine	0.82,0,0.30,0
BlueGreen	0.85,0,0.33,0	Emerald	1,0,0.50,0
JungleGreen	0.99,0,0.52,0	SeaGreen	0.69,0,0.50,0
Green	1,0,1,0	ForestGreen	0.91,0,0.88,0.12
PineGreen	0.92,0,0.59,0.25	LimeGreen	0.50,0,1,0
YellowGreen	0.44,0,0.74,0	SpringGreen	0.26,0,0.76,0
OliveGreen	0.64,0,0.95,0.40	RawSienna	0,0.72,1,0.45
Sepia	0,0.83,1,0.70	Brown	0,0.81,1,0.60
Tan	0.14,0.42,0.56,0		
White (Gray0)	0,0,0,0	Black (Gray100)	0,0,0,1
Gray5	0,0,0,0.05	Gray10	0,0,0,0.10
Gray15	0,0,0,0.15	Gray20	0,0,0,0.20

Gray25	0,0,0,0.25	Gray30	0,0,0,0.30
LightGray	0,0,0,0.33	Gray35	0,0,0,0.35
Gray40	0,0,0,0.40	Gray45	0,0,0,0.45
Gray50	0,0,0,0.50	Gray	0,0,0,0.50
Gray55	0,0,0,0.55	Gray60	0,0,0,0.60
Gray65	0,0,0,0.65	DarkGray	0,0,0,0.66
Gray70	0,0,0,0.70	Gray75	0,0,0,0.75
Gray80	0,0,0,0.80	Gray85	0,0,0,0.85
Gray90	0,0,0,0.90	Gray95	0,0,0,0.95
LightGreenYellow	0.08,0,0.35,0	LightYellow	0,0,0.50,0
LightGoldenrod	0,0.05,0.42,0	LightDandelion	0,0.15,0.42,0
LightApricot	0,0.16,0.26,0	LightPeach	0,0.25,0.35,0
LightMelon	0,0.23,0.25,0	LightYellowOrange	0,0.21,0.50,0
LightOrange	0,0.31,0.44,0	LightBurntOrange	0,0.26,0.50,0
LightBittersweet	0,0.38,0.50,0.12	LightRedOrange	0,0.39,0.44,0
LightMahogany	0,0.43,0.44,0.18	LightMaroon	0,0.44,0.34,0.16
LightBrickRed	0,0.45,0.47,0.14	LightRed	0,0.50,0.50,0
LightOrangeRed	0,0.50,0.25,0	LightRubineRed	0,0.50,0.07,0
LightWildStrawberry	0,0.48,0.20,0	LightSalmon	0,0.27,0.19,0
LightCarnationPink	0,0.32,0,0	LightMagenta	0,0.50,0,0
LightVioletRed	0,0.40,0,0	LightRhodamine	0,0.41,0,0
LightMulberry	0.17,0.45,0,0.01	LightRedViolet	0.04,0.45,0,0.17
LightFuchsia	0.24,0.46,0,0.04	LightLavender	0,0.24,0,0
LightThistle	0.06,0.30,0,0	LightOrchid	0.16,0.32,0,0
LightDarkOrchid	0.20,0.40,0.10,0	LightPurple	0.23,0.43,0,0
LightPlum	0.25,0.50,0,0	LightViolet	0.40,0.44,0,0
LightRoyalPurple	0.38,0.45,0,0	LightBlueViolet	0.43,0.46,0,0.02
LightPeriwinkle	0.29,0.28,0,0	LightCadetBlue	0.31,0.29,0.12,0
LightCornflowerBlue	0.33,0.07,0,0	LightMidnightBlue	0.49,0.07,0,0.22
LightNavyBlue	0.47,0.27,0,0	LightRoyalBlue	0.50,0.25,0,0
LightBlue	0.50,0.50,0,0	LightCerulean	0.47,0.06,0,0
LightCyan	0.50,0,0,0	LightProcessBlue	0.48,0,0,0
LightSkyBlue	0.31,0,0.06,0	LightTurquoise	0.43,0,0.10,0
LightTealBlue	0.43,0,0.17,0.01	LightAquamarine	0.41,0,0.15,0
LightBlueGreen	0.43,0,0.17,0	LightEmerald	0.50,0,0.25,0
LightJungleGreen	0.50,0,0.26,0	LightSeaGreen	0.35,0,0.25,0
LightGreen	0.50,0,0.50,0	LightForestGreen	0.46,0,0.44,0.06
LightPineGreen	0.46,0,0.30,0.13	LightLimeGreen	0.25,0,0.50,0
LightYellowGreen	0.22,0,0.37,0	LightSpringGreen	0.13,0,0.38,0
LightOliveGreen	0.32,0,0.48,0.20	LightRawSienna	0,0.36,0.50,0.23
LightSepia	0,0.44,0.50,0.35	LightBrown	0,0.41,0.50,0.30
LightTan	0.07,0.21,0.28,0		
<i>name</i> <i>RGB</i>		<i>name</i> <i>RGB</i>	
BlueRed5	0.15,0.17,0.55	BlueRed10	0.20,0.23,0.57
BlueRed15	0.24,0.29,0.60	BlueRed20	0.33,0.35,0.64
BlueRed25	0.43,0.43,0.68	BlueRed30	0.52,0.52,0.73
BlueRed35	0.60,0.60,0.78	BlueRed40	0.70,0.70,0.84

BlueRed45	0.80,0.80,0.85	BlueRed50	0.86,0.82,0.82
BlueRed55	0.87,0.73,0.73	BlueRed60	0.89,0.64,0.64
BlueRed65	0.90,0.55,0.55	BlueRed70	0.91,0.47,0.46
BlueRed75	0.91,0.39,0.37	BlueRed80	0.90,0.33,0.28
BlueRed85	0.89,0.25,0.20	BlueRed90	0.88,0.23,0.14
BlueRed95	0.87,0.21,0.09	BlueRed100	0.87,0.16,0.04
GreenRed5	0,1,0	GreenRed10	0.05,0.95,0
GreenRed15	0.10,0.90,0	GreenRed20	0.15,0.85,0
GreenRed25	0.20,0.80,0	GreenRed30	0.25,0.75,0
GreenRed35	0.30,0.70,0	GreenRed40	0.35,0.65,0
GreenRed45	0.40,0.60,0	GreenRed50	0.45,0.55,0
GreenRed55	0.50,0.50,0	GreenRed60	0.55,0.45,0
GreenRed65	0.60,0.40,0	GreenRed70	0.65,0.35,0
GreenRed75	0.70,0.30,0	GreenRed80	0.75,0.25,0
GreenRed85	0.80,0.20,0	GreenRed90	0.85,0.15,0
GreenRed95	0.90,0.10,0	GreenRed100	0.95,0.05,0
ColdHot5	0,0.08,1	ColdHot10	0,0.29,1
ColdHot15	0,0.49,1	ColdHot20	0,0.70,1
ColdHot25	0,0.90,1	ColdHot30	0,1,0.87
ColdHot35	0,1,0.68	ColdHot40	0,1,0.46
ColdHot45	0,1,0.25	ColdHot50	0,1,0.04
ColdHot55	0.16,1,0	ColdHot60	0.35,1,0
ColdHot65	0.56,1,0	ColdHot70	0.79,1,0
ColdHot75	0.98,1,0	ColdHot80	1,0.82,0
ColdHot85	1,0.60,0	ColdHot90	1,0.40,0
ColdHot95	1,0.20,0	ColdHot100	0.91,0,0

Type the color names with the upper case letters exactly as described above. For the definition of new colors use one of the `color.sty` commands:

```
\definecolor{<name>}{cmyk}{<C,M,Y,K>}
```

```
\definecolor{<name>}{rgb}{<R,G,B>}
```

The `<name>` can be chosen freely, the values for the color composition must be in the range 0–1, i.e. 0–100% of the respective component (‘C’ – cyan, ‘M’ – magenta, ‘Y’ – yellow, ‘K’ – black; or ‘R’ – red, ‘G’ – green, ‘Blue’ – blue) separated by commas.

Examples:

```
\definecolor{Salmon}{cmyk}{0,0.53,0.38,0}
```

```
\definecolor{ColdHot15}{rgb}{0,0.49,1}
```

6 Listing of the texshade default settings

6.1 Standard definitions

The file `texshade.def` mirrors all commands which are carried out at the beginning of the `texshade` environment. Short comments are also included, thus, it is referred to this file for further information.

6.2 Colors used in the different shading modes

Color scheme *blues*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
Black	Magenta	similar
White	RoyalBlue	identical
Goldenrod	RoyalPurple	all match

Color scheme *greens*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
Black	GreenYellow	similar
White	PineGreen	identical
YellowOrange	OliveGreen	all match

Color scheme *reds*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
Black	YellowOrange	similar
White	BrickRed	identical
YellowGreen	Mahagony	all match

Color scheme *grays*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
Black	LightGray	similar
White	DarkGray	identical
White	Black	all match

Color scheme *black*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
Black	White	similar
White	Black	identical
White	Black	all match

Functional mode *charge*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
White	Red	acidic
White	Blue	basic

Functional mode *hydropathy*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
White	Red	acidic
White	Blue	basic
Black	Yellow	polar uncharged
White	Green	hydrophobic nonpolar

Functional mode *chemical*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
White	Red	acidic
White	Black	aliphatic
White	Green	amide
White	Brown	aromatic
White	Blue	basic
Black	Magenta	hydroxyl
Black	Orange	imino
Black	Yellow	sulfur

Functional mode *structure*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
Black	Orange	external
Black	Yellow	ambivalent
White	Green	internal

Functional mode *standard area*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
Black	BrickRed	G
Black	Orange	A, S
Black	Yellow	C, P
Black	YellowGreen	T, D, V, N
White	PineGreen	I, E
Black	SkyBlue	L, Q, H, M
White	RoyalPurple	F, K
White	RedViolet	Y
White	Black	R, W

Functional mode *accessible area*:

<i>res.color</i>	<i>shad.color</i>	<i>residues</i>
Black	White	no match
Black	BrickRed	C
Black	Orange	I, V, G
Black	Yellow	F, L, M, A
Black	YellowGreen	W, S, T, H
White	PineGreen	P
Black	SkyBlue	Y, D, N
White	RoyalPurple	E, Q
White	RedViolet	R
White	Black	K

7 Quick Reference

The **TeXshade** logo

`\TeXshade`

The **TeXshade** environment (10 ff.)

```
\begin{texshade}[\langle parameterfile \rangle] {\langle alignmentfile \rangle}
    further TeXshade commands, if needed
\end{texshade}
```

Predefined shading modes

`\seqtype{\langle type \rangle}` (P – peptide, N – nucleotide) [23]

`\shadingmode[\langle option \rangle]{\langle mode \rangle}` [26]

<i>\langle mode \rangle</i>	<i>\langle option \rangle</i>	
identical	allmatchspecial	
similar	allmatchspecial	
diverse	<i>\langle seqref \rangle</i>	
functional	<i>\langle type \rangle</i>	charge hydropathy structure chemical standard area accessible area

`\shadeallresidues` [14]

Shading colors (26 ff.)

```
\shadingcolors{\langle scheme \rangle} (blues, reds, greens, grays, black)
\nomatchresidues{\langle res.col. \rangle}{\langle shad.col. \rangle}{\langle case \rangle}{\langle style \rangle}
\similarresidues{\langle res.col. \rangle}{\langle shad.col. \rangle}{\langle case \rangle}{\langle style \rangle}
\conservedresidues{\langle res.col. \rangle}{\langle shad.col. \rangle}{\langle case \rangle}{\langle style \rangle}
\allmatchresidues{\langle res.col. \rangle}{\langle shad.col. \rangle}{\langle case \rangle}{\langle style \rangle}
\funcshadingstyle{\langle residue \rangle}{\langle res.col. \rangle}{\langle shad.color \rangle}
    {\langle case \rangle}{\langle style \rangle} [31]
```

Residue grouping

<code>\pepsims{⟨residue⟩}{⟨similar⟩}</code>	[27]
<code>\pepgroups{⟨group1⟩,⟨group2⟩, ... , ⟨groupn⟩}</code>	[27]
<code>\DNAsims{⟨residue⟩}{⟨similar⟩}</code>	[27]
<code>\DNAGroups{⟨group1⟩,⟨group2⟩, ... , ⟨groupn⟩}</code>	[27]

Definition of new functional shading modes

<code>\clearfuncgroups</code>	[31]
<code>\funcgroup{⟨descr⟩}{⟨residues⟩}{⟨res.col.⟩}{⟨shad.col.⟩}</code> <code>{⟨case⟩}{⟨style⟩}</code>	[31]

Appearance of the consensus line

<code>\threshold{⟨percentage⟩}</code>	[32]
<code>\constosingleseq{⟨seqref⟩}</code>	[32]
<code>\showconsensus{⟨position⟩}</code>	[32]
<code>\hideconsensus</code>	[32]
<code>\nameconsensus{⟨name⟩}</code>	[32]
<code>\defconsensus{⟨symbol1⟩}{⟨symbol2⟩}{⟨symbol3⟩}</code>	[33]
<code>\consensuscolors{⟨res.col.1⟩}{⟨shad.col.1⟩}</code> <code>{⟨res.col.2⟩}{⟨shad.col.2⟩}</code> <code>{⟨res.col.3⟩}{⟨shad.col.3⟩}</code>	[33]

Appearance of the sequence lines

<code>\shownames{⟨position⟩}</code>	[34]
<code>\shownumbering{⟨position⟩}</code>	[34]
<code>\nameseq{⟨seqref⟩}{⟨name⟩}</code>	[34]
<code>\namescolor{⟨color⟩}</code>	[35]
<code>\namecolor{⟨seq1⟩, ... ,⟨seq n⟩}{⟨color⟩}</code>	[35]
<code>\hidenames</code>	[35]
<code>\hidename{⟨seq1⟩, ... ,⟨seq n⟩}</code>	[35]
<code>\numberingcolor{⟨color⟩}</code>	[35]
<code>\numbercolor{⟨seq1⟩, ... ,⟨seq n⟩}{⟨color⟩}</code>	[35]
<code>\hidenumbering</code>	[35]
<code>\hidenumber{⟨seq1⟩, ... ,⟨seq n⟩}</code>	[35]
<code>\hideresidues</code>	[37]
<code>\showresidues</code>	[37]
<code>\startnumber{⟨seqref⟩}{⟨first residue number⟩}</code>	[35]

<code>\seqlength{⟨seqref⟩}{⟨length⟩}</code>	[35]
<code>\setends{⟨seqref⟩}{⟨startnumber⟩}..⟨stopnumber⟩}</code>	[35]
<code>\showruler{⟨position⟩}{⟨seqref⟩}</code>	[36]
<code>\rulersteps{⟨number⟩}</code>	[36]
<code>\rulercolor{⟨color⟩}</code>	[36]
<code>\hideruler</code>	[36]
<code>\gapchar{⟨symbol⟩}</code> (incl. rule)	[36]
<code>\gapcolors{⟨symbol color⟩}{⟨background color⟩}</code>	[36]
<code>\showleadinggaps</code>	[37]
<code>\hideleadinggaps</code>	[37]
<code>\fingerprint{⟨res. per line⟩}</code>	[39]

Hiding, killing, separating and ordering

<code>\hideseq{⟨seq1⟩,⟨seq2⟩,...,⟨seq n⟩}</code>	[37]
<code>\killseq{⟨seq1⟩,⟨seq2⟩,...,⟨seq n⟩}</code>	[37]
<code>\donotshade{⟨seq1⟩,⟨seq2⟩,...,⟨seq n⟩}</code>	[37]
<code>\separationline{⟨seqref⟩}</code>	[37]
<code>\smallsep</code>	[37]
<code>\medsep</code>	[37]
<code>\bigsep</code>	[37]
<code>\vsepspace{⟨length⟩}</code>	[37]
<code>\orderseqs{⟨seq1⟩,⟨seq2⟩,...,⟨seq n⟩}</code>	[38]

Residues per line and further settings

<code>\residuesperline{⟨number⟩}</code>	[38]
<code>\residuesperline*{⟨number⟩}</code>	[38]
<code>\charstretch{⟨factor⟩}</code>	[38]
<code>\linestretch{⟨factor⟩}</code>	[38]
<code>\numberingwidth{⟨n digits⟩}</code>	[38]
<code>\smallblockskip</code>	[39]
<code>\medblockskip</code>	[39]
<code>\bigblockskip</code>	[39]
<code>\noblockskip</code>	[39]
<code>\vblockspace{⟨length⟩}</code>	[39]
<code>\flexblockspace</code>	[39]
<code>\fixblockspace</code>	[39]
<code>\alignment{⟨position⟩}</code>	[39]

Individual shading and labeling of sequence stretches

```

\shaderegion{<seqref>}{<start1>..<>stop1>,<start2>..<>stop2>,
...,<start n>..<>stop n>}{<res.col.>}{<shad.col.>} [40]
\shadeblock{<seqref>}{<start1>..<>stop1>,<start2>..<>stop2>,
...,<start n>..<>stop n>}{<res.col.>}{<shad.col.>} [40]
\emphregion{<seqref>}{<start1>..<>stop1>,<start2>..<>stop2>,
...,<start n>..<>stop n>} [40]
\emphblock{<seqref>}{<start1>..<>stop1>,<start2>..<>stop2>,
...,<start n>..<>stop n>} [40]
\emphdefault{<style>} [40]
\feature{<position>}{<seqref>}{<start1>..<>stop1>,
<start2>..<>stop2>,...,<start n>..<>stop n>}{<labelstyle>}{<text>} [41]

{<labelstyle>} = {brace[<color>]}
                = {fill:<symbol>[<textcolor>]}
                = {helix[<helixcolor>]}
                = {box[<boxcolor>]:<text>[<textcolor>]}
                = arrows and bars (==<' ,|)(==)(==>' ,|)
                = {translate[<color>]}
                = {bar[<min>,<max>]:<data/file>[<color>]}
                = {color[<min>,<max>]:<data/file>[<scale>]}

\ttopspace{<length>} [41]
\topspace{<length>} [41]
\bottomspace{<length>} [41]
\bbottomspace{<length>} [41]
\featurerule{<length>} [43]
\bargraphstretch{<factor>} [45]
\codon{<amino acid>}{<triplet1>,...,<triplet n>} [43]
\geneticcode{<filename>} [43]
\backtranslabel[<size>]{<style>} [44]
\backtranstext[<size>]{<style>} [46]

```

```

{<style>} = {horizontal}
           = {alternating}
           = {zigzag}
           = {oblique}
           = {vertical}

```

Including secondary structure information

<code>\includeDSSP[make new]{\seqref}{\filename}</code>	[46]
<code>\includeSTRIDE[make new]{\seqref}{\filename}</code>	[46]
<code>\includePHDsec[make new]{\seqref}{\filename}</code>	[46]
<code>\includePHDtopo[make new]{\seqref}{\filename}</code>	[46]
<code>\showonDSSP{\structures}</code>	[47]
<code>\showonSTRIDE{\structures}</code>	[47]
<code>\showonPHDsec{\structures}</code>	[47]
<code>\showonPHDtopo{\structures}</code>	[47]
<code>\hideonDSSP{\structures}</code>	[47]
<code>\hideonSTRIDE{\structures}</code>	[47]
<code>\hideonPHDsec{\structures}</code>	[47]
<code>\hideonPHDtopo{\structures}</code>	[47]
<code>\appearance{\type}{\position}{\labelstyle}{\text}</code>	[48]
<code>\numcount</code>	[48]
<code>\alphacount</code>	[48]
<code>\Alphacount</code>	[48]
<code>\romancount</code>	[48]
<code>\Romancount</code>	[48]
<code>\firstcolumnDSSP</code>	[48]
<code>\secondcolumnDSSP</code>	[48]

Displaying and building legends

<code>\showlegend</code>	[49]
<code>\hidelegend</code>	[49]
<code>\movelegend{\x-offset}{\y-offset}</code>	[49]
<code>\germanlanguage, \spanishlanguage, \englishlanguage</code>	[49]
<code>\legendcolor{\color}</code>	[49]
<code>\shadebox{\color}</code>	[49]

Adding captions to the alignment

<code>\showcaption[\position]{\text}</code>	[50]
---	------

Font handling

<code>\setfamily{\text}{\family}</code>	[50]
<code>\setseries{\text}{\series}</code>	[50]
<code>\setshape{\text}{\shape}</code>	[50]
<code>\setsize{\text}{\size}</code>	[50]
<code>\setfont{\text}{\family}{\series}{\shape}{\size}</code>	[51]

<code>\featuresrm</code>	<code>\featurestiny</code>	[51]
<code>\featuresssf</code>	<code>\featuresscriptsize</code>	
<code>\featurestt</code>	<code>\featuresfootnotesize</code>	
<code>\featuresbf</code>	<code>\featuressmall</code>	
<code>\featuresmd</code>	<code>\featuresnormalsize</code>	
<code>\featuresit</code>	<code>\featureslarge</code>	
<code>\featuressl</code>	<code>\featuresLarge</code>	
<code>\featuressc</code>	<code>\featuresLARGE</code>	
<code>\featuresup</code>	<code>\featureshuge</code>	
	<code>\featuresHuge</code>	

Corresponding sets are provided for the numbering (`\numberingrm` etc.), featurestyles (`featurestylesrm` etc.), names (`\namesrm` etc.), residues (`\residuesrm` etc.) and legend texts (`legendrm` etc.).

Goodies—molweight and charge

<code>\molweight{\seqref}\{Da/kDa\}</code>	[52]
<code>\charge{\seqref}\{i/o/N/C\}</code>	[52]

8 References

- [1] CARLISLE, D. The Standard L^AT_EX ‘Graphics Bundle’, `color.sty`.
- [2] KARLIN, S.; GHANDOUR, G. (1985) Multiple-alphabet amino acid sequence comparisons of the immunoglobulin κ -chain constant domain. *Proc. Natl. Acad. Sci. USA*: **82**, 8597–8601.
- [3] KYTE, J.; DOOLITTLE, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*: **157**, 105–132.
- [4] ROSE, G. D.; GESELOWITZ, A. R.; LESSER, G. J.; LEE, R. H.; ZEHFUS, M. H. (1985) Hydrophobicity of amino acid residues in globular proteins. *Science*: **229**, 835–838.
- [5] LESSER, G. J.; ROSE, G. D. (1990) Hydrophobicity of amino acid subgroups in proteins. *Proteins: structure, function and genetics*: **8**, 6–13.
- [6] FRÖHLICH, K.-U. (1994) Sequence similarity presenter: a tool for the graphic display of similarities of long sequences for use in presentations. *Comput. Applic. Biosci.*: **10**, 179–183.
- [7] KABSCH, W.; SANDER, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*: **22**, 2577–2637.
- [8] FRISHMAN, D.; ARGOS, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins: structure, function and genetics*: **23**, 566–579.
- [9] ROST, B.; SANDER, C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins: structure, function and genetics*: **19**, 55–72.
- [10] ROKICKI, T. DVIPS: A T_EX driver.