GraBCas: a bioinformatics tool for score-based prediction of Caspase- and Granzyme B-cleavage sites in protein sequences

Christina Backes, Jan Kuentzer, Hans-Peter Lenhof, Nicole Comtesse and Eckart Meese*

Department of Human Genetics, Building 60, Medical School, University of Saarland, 66421 Homburg/Saar, Germany and 1Center for Bioinformatics, Building 36.1, University of Saarland, 66041 Saarbrücken, Germany

Received February 12, 2005; Revised and Accepted March 24, 2005

ABSTRACT
Caspases and granzyme B are proteases that share the primary specificity to cleave at the carboxyl terminal of aspartate residues in their substrates. Both, caspases and granzyme B are enzymes that are involved in fundamental cellular processes and play a central role in apoptotic cell death. Although various targets are described, many substrates still await identification and many cleavage sites of known substrates are not identified or experimentally verified. A more comprehensive knowledge of caspase and granzyme B substrates is essential to understand the biological roles of these enzymes in more detail. The relatively high variability in cleavage site recognition sequence often complicates the identification of cleavage sites. As of yet there is no software available that allows identification of caspase and/or granzyme with cleavage sites differing from the consensus sequence. Here, we present a bioinformatics tool ‘GraBCas’ that provides score-based prediction of potential cleavage sites for the caspases 1–9 and granzyme B including an estimation of the fragment size. We tested GraBCas on already known substrates and showed its usefulness for protein sequence analysis. GraBCas is available at http://wwwalt.med-rz.uniklinik-saarland.de/med_fak/humangenetik/software/index.html.

INTRODUCTION
Caspases are enzymes orchestrating the cellular pathways leading to apoptosis and inflammatory signals. Besides these functions they are supposed to be involved in other cellular processes, such as development, cell cycle, cell proliferation, cell migration and receptor internalization (1,2). Caspases are cysteine proteases with specificity for an aspartic acid residue at position P1 of the substrate. This primary specificity is shared by the serine protease granzyme B, which induces cytotoxic T lymphocyte-mediated target cell DNA fragmentation and apoptosis (3,4). Granzyme B-mediated cleavage also plays a role in induction of autoimmunity (5).

To date, at least 14 mammalian caspases can be grouped into three classes based on their substrate specificities. Group I consisting of caspases -1, -4, -5 (-14 and murine -11 and -12) cleaves the substrate sequence (W/L)EHD, group II (caspases -2, -3, -7) cleaves the DEXD motif and group III (caspase -6, -8, -9, -10) preferentially cleaves the (L/V)E(T/H)D sequence (6,7). Caspases of group I play an important role in the generation of inflammatory signals and in the immune regulation. Caspases -8, -9 and -10 are so-called initiator caspases mainly cleaving and activating procaspases, whereas caspases -3, -6 and -7 as effector caspases cleave numerous cellular proteins. The serine protease granzyme B prefers substrates with sequence IEXD, and is released by cytotoxic lymphocytes to kill virus-infected or tumor cells.

Although more than 280 caspase targets are described [for comprehensive review see (8)] many substrates still await identification and many cleavage sites of known substrates are not identified or experimentally verified. Likewise, the identification of granzyme B substrates is still at its infancy. Intracellular substrates of granzyme B include other caspases, mainly caspase 3 (9), ADPRT (ADP-ribosyltransferase 1, PARP) (10), BID (BH3 interacting domain death agonist) (11) and ICAD (DNA fragmentation factor) (12). Notably, the majority of autoantigens in systemic autoimmune diseases are efficiently cleaved by granzyme B (5).

A more comprehensive knowledge of caspase and granzyme B substrates is essential to understand the biological...
roles of these enzymes in more detail. The relatively high variability in cleavage site recognition sequence often complicates the identification of cleavage sites. As of yet there is no software available that allows identification of caspase and/or granzyme cleavage sites differing from the consensus sequence. The PeptidCutter program provided by the ExPasy Server (http://www.expasy.org/tools/peptidcutter) considers only the preferred peptide substrate sites. A recent tool of Lohmüller et al. (13) is restricted to caspase 3 and cathepsin B and -L substrates. Here, we present a bioinformatics tool GraBCas that provides score-based prediction of potential cleavage sites for the caspases 1–9 and granzyme B including an estimation of the fragment size. We validated our tool by scoring known substrates and demonstrated its usefulness for protein sequence analysis.

MATERIALS AND METHODS

Design of cleavage site scoring matrices

We developed position specific scoring matrices (PSSM) for the endopeptidases granzyme B and caspase 1–9 based on experimentally determined substrate specificities (6). Thornbery et al. (6) determined the substrate specificities using positional scanning synthetic combinatorial libraries. Cleavage was fluorimetrically determined with maximum value annotated with scanning synthetic combinatorial libraries. Cleavage was experimentally determined substrate specificities (6). Thornbery et al. (6) determined the substrate specificities using positional scanning synthetic combinatorial libraries. Cleavage was fluorimetrically determined with maximum value annotated with scanning synthetic combinatorial libraries. Cleavage was experimentally determined substrate specificities (6). Thornbery et al. (6) determined the substrate specificities using positional scanning synthetic combinatorial libraries. Cleavage was fluorimetrically determined with maximum value annotated with scanning synthetic combinatorial libraries. Cleavage was experimentally determined substrate specificities (6). Thornbery et al. (6) determined the substrate specificities using positional scanning synthetic combinatorial libraries. Cleavage was fluorimetrically determined with maximum value annotated with scanning synthetic combinatorial libraries. Cleavage was experimentally determined substrate specificities (6). Thornbery et al. (6) determined the substrate specificities using positional scanning synthetic combinatorial libraries. Cleavage was fluorimetrically determined with maximum value annotated with scanning synthetic combinatorial libraries. Cleavage was experimentally determined substrate specificities (6). Thornbery et al. (6) determined the substrate specificities using positional scanning synthetic combinatorial libraries. Cleavage was fluorimetrically determined with maximum value annotated with scanning synthetic combinatorial libraries.

Table 1. Scoring matrices for granzyme B and caspases 1–9

| Position Pi | AA of consensus recognition motif | C | S | T | P | A | G | N | D | E | Q | H | R | K | M | I | L | V | F | Y | W |
| Granzyme B  |                                  | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1000 | 52 | 500 | 12 | 1 | 1 |
| Caspase 1   |                                  | 4 | W | 1 | 48 | 48 | 16 | 16 | 80 | 96 | 16 | 128 | 1 | 1 | 96 | 288 | 80 | 496 | 576 | 1000 |
| Caspase 2   |                                  | 4 | D | 1 | 1 | 50 | 10 | 1 | 1 | 10 | 1000 | 200 | 1 | 1 | 1 | 1 | 1 | 1 | 180 | 400 | 80 | 40 | 40 | 1 |
| Caspase 3   |                                  | 4 | D | 1 | 40 | 50 | 1 | 10 | 20 | 1000 | 40 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 20 | 10 | 1 | 1 |
| Caspase 4   |                                  | 4 | W | 1 | 306 | 357 | 1 | 153 | 153 | 255 | 1000 | 408 | 17 | 17 | 1 | 153 | 153 | 306 | 272 | 255 | 187 | 153 |
| Caspase 5   |                                  | 4 | V | 1 | 14 | 378 | 406 | 182 | 1 | 1 | 1 | 0 | 14 | 196 | 42 | 0 | 1 | 714 | 224 | 1000 | 182 | 154 | 84 |
| Caspase 6   |                                  | 4 | W | 1 | 80 | 208 | 144 | 96 | 48 | 80 | 288 | 256 | 96 | 48 | 1 | 1 | 304 | 848 | 224 | 384 | 352 | 1000 |
| Caspase 7   |                                  | 4 | V | 1 | 144 | 880 | 64 | 96 | 16 | 64 | 224 | 256 | 48 | 48 | 1 | 1 | 656 | 304 | 1000 | 80 | 48 | 48 | 16 |
| Caspase 8   |                                  | 4 | V | 1 | 48 | 48 | 16 | 80 | 16 | 16 | 176 | 1000 | 144 | 48 | 1 | 1 | 16 | 16 | 48 | 16 | 16 | 16 | 16 |
| Caspase 9   |                                  | 4 | D | 1 | 54 | 576 | 18 | 72 | 18 | 108 | 18 | 18 | 36 | 1000 | 54 | 36 | 1 | 648 | 486 | 918 | 288 | 216 | 558 |
| Caspase 10  |                                  | 4 | V | 1 | 117 | 78 | 1 | 39 | 26 | 26 | 1000 | 104 | 26 | 39 | 1 | 1 | 13 | 13 | 13 | 13 | 13 | 13 |
| Caspase 11  |                                  | 4 | E | 1 | 221 | 357 | 1 | 323 | 51 | 153 | 255 | 1000 | 425 | 187 | 17 | 1 | 306 | 857 | 198 | 224 | 153 | 84 |
| Caspase 12  |                                  | 4 | V | 1 | 16 | 448 | 448 | 128 | 16 | 48 | 1 | 1 | 16 | 208 | 80 | 16 | 1 | 704 | 176 | 1000 | 160 | 128 | 48 |
| Caspase 13  |                                  | 4 | L | 1 | 208 | 304 | 480 | 448 | 96 | 304 | 704 | 448 | 96 | 144 | 1 | 1 | 576 | 1000 | 720 | 224 | 256 | 144 |
| Caspase 14  |                                  | 4 | T | 1 | 45 | 75 | 0 | 45 | 15 | 15 | 180 | 1000 | 150 | 45 | 1 | 1 | 45 | 15 | 105 | 45 | 45 | 15 |
| Caspase 15  |                                  | 4 | L | 1 | 180 | 1000 | 216 | 324 | 18 | 126 | 72 | 198 | 108 | 306 | 72 | 72 | 1 | 720 | 108 | 792 | 180 | 198 | 306 |
| Caspase 16  |                                  | 4 | L | 1 | 198 | 216 | 594 | 576 | 144 | 108 | 414 | 468 | 180 | 126 | 36 | 18 | 1 | 576 | 1000 | 684 | 252 | 216 | 144 |
| Caspase 17  |                                  | 4 | E | 1 | 85 | 136 | 51 | 85 | 17 | 17 | 272 | 1000 | 187 | 119 | 1 | 1 | 102 | 119 | 204 | 102 | 85 | 51 |
| Caspase 18  |                                  | 4 | H | 1 | 85 | 136 | 102 | 85 | 17 | 17 | 272 | 1000 | 187 | 119 | 1 | 1 | 102 | 119 | 204 | 102 | 85 | 51 |

Amino acid preference distribution for each position Pi was extracted from Thornberry et al. (6) giving the most common amino acid a value of 1000.
(100^3) of the product of the consensus recognition motif for normalization and multiplied by 100, yielding a total score between 0 and 100.

\[
\text{Score}(A4A3A2D) = 100 \times \frac{\text{Score}_{P4}(A4) \times \text{Score}_{P3}(A3) \times \text{Score}_{P2}(A2)}{100^3}.
\]

**Using additional filter options for granzyme B and caspase 3**

To improve the power of the prediction we analyzed the amino acid distribution of known granzyme B and caspase 3 cleavage sites at positions P6–P2' taken from the literature (see also Supplementary Material 1 and 2).

For granzyme B we found a preference for V (15\times) and I (11\times) at position P4, for E (9\times) at position P3 and for P (11\times) at position P2 in accordance with the results of Thornberry et al. (6). We detected S at position P1' and G at position P2', respectively in 9 out of 30 cleavage sites. The result list of the PSSM-based cleavage sites can optionally be filtered with two 'stringency' filters that take the occurrence of amino acids at position P2' into account. We installed a 'low stringency' filter that excludes hits with the amino acids C, Q, I, M, V all of which are medium sized or large amino acids. A second 'high stringency' filter selects hits with a G at position P2'.

The analysis of the 59 cleavage sites of caspase 3 substrate confirmed the preferences for D at P4 (31\times), E at P3 (17\times) and V at P2 (16\times). For P1' we found an abundance of G (18\times) and S (17\times) and in lower amount A (5\times) and N (4\times). As for the granzyme B prediction, two additional 'stringency' filters for the prediction of caspase 3 cleavage sites are available. The 'high stringency' filter screens the predicted hits for occurrences of G, S, A or N at position P1', and the 'low stringency' filter screens for absence of R, E, H, K, Q, I, L, M, F, W and Y at this position.

**GraBCas software tool**

The GraBCas program was written in Java™ and is available as an application or as an applet. Both are available at http://wwwalt.med-rz.uniklinik-saarland.de/med_fak/humangenetik/software/index.html. If your browser does not support Java™ you need to install the Java Runtime Environment (JRE) 1.4.x, which can be downloaded at http://java.sun.com.

The graphical user interface is easy to use. There are several register cards for each endopeptidase and one register card presenting the input form, where the amino acid sequence can be pasted and a cutoff for the PSSM scores can be chosen.

After pressing the OK-button in the input form, the program calculates the scores of potential cleavage sites for all endopeptidases and presents them in the corresponding register card sorted with the highest scoring sites on top. The user can open an additional window for viewing the positions of the predicted cleavage sites within the amino acid sequence. The window also shows the fragment length and size in kDa (0.11 kDa per amino acid) of the predicted fragments.

As described above, for caspase 3 and granzyme B additional filter options are available in their register cards. The two filter types for these enzymes, a 'high-stringency' and a 'low-stringency' filter, are based on the extended substrate specificity. For granzyme B the amino acids at position P2' were taken into account in addition to the positions P4–P1. For caspase 3, amino acids at position P1' are evaluated.

**Sensitivity–specificity plots**

For determining the specificity and sensitivity of the GraBCas predictions we used the known cleavage sites of granzyme B (4–6, 9–12) summarized in Table 2 and the known non-substrates of granzyme B (5) presented in Table 4. Due to the lack of information on known non-substrates for caspase 3 the sensitivity–specificity plot could only be calculated for granzyme B (Figure 1).

The x-axis of the plots represents the cutoff values (with respect to the PSSM scores), while the y-axis represents the percentage of the specificity or sensitivity of the predictions made by GraBCas, respectively. The specificity is computed as follows:

\[
\text{Number of true negatives} + \text{Number of true negatives}.
\]

The true negatives are the known non-substrates where the maximal PSSM score of all tetrapeptides ending with a D is smaller than the chosen cutoff value. A specificity of 1 means that all known non-substrates were below the cutoff, i.e. all known non-substrates were correctly classified as negatives.

The sensitivity is defined as:

\[
\text{Number of true positives} + \text{Number of false negatives}.
\]

where true positives are the known cleavage sites with a score larger than the chosen cutoff value. A sensitivity of 1 means that all cleavage sites of our test set (Table 2) have a score higher than the chosen cutoff and that they have been correctly classified as positives.

**RESULTS AND DISCUSSION**

We analyzed the cleavage sites of known substrates of granzyme B and caspase 3 to compare the experimentally identified peptide specificity with the cleavage site predicted by the program GraBCas.

In total, we collected 29 substrates with 30 cleavage sites for granzyme B (Table 2) and 47 substrates with 59 cleavage sites for caspase 3 (Table 3) and computed the GraBCas scores of the cleavage sites. For granzyme B we collected additionally 17 sequences which are non-substrates of this endopeptidase (Table 4), computed the scores of all putative cleavage sites in these sequences and extracted the best hit by GraBCas for each of these non-substrates.

The sensitivity–specificity plot for granzyme B is shown in Figure 1. When using a cutoff value of 1.2 in the GraBCas program, we obtain a sensitivity of ~80% and a specificity of ~82%. The cutoff value can be adjusted if a higher specificity or sensitivity is needed for the cleavage site prediction.

A closer look at the sensitivity–specificity plot shows that the best score (28.8 for IEED in glycogen phosphorylase) of the alleged non-substrates is extremely high. The top value of the best hit IEED is due to the fact that this tetrapeptide has three identical positions with the granzyme B consensus
recognition motif IEPD. Furthermore, the amino acid E on P2 has a middle-sized value and the tetrapeptides LEED, IEAD and IETD are known substrates of granzyme B. We assume that glycogen phosphorylase is probably a substrate of granzyme B. This warrants further experimental analysis.

We also studied the occurrences of amino acids at position P1 and P2 of the known cleavage sites of granzyme B and caspase 3. Additional filtering options have been added to GraBCas that are based on these statistics. For granzyme B, we detected G at position P2 detected by the high stringency filter. Numbers in brackets indicate cleavage site position in the amino acid sequence.

Table 2. Analysis of cleavage sites of known granzyme B substrates with GraBCas

<table>
<thead>
<tr>
<th>Granzyme B substrate</th>
<th>Acc_number</th>
<th>Known cleavage site</th>
<th>Score by GraBCas</th>
<th>P6–P2 of cleavage site</th>
</tr>
</thead>
<tbody>
<tr>
<td>AARS: alanyl-tRNA synthetase</td>
<td>NP_001596</td>
<td>VADP (632)</td>
<td>7.65</td>
<td>SLVAPDRL</td>
</tr>
<tr>
<td>ADPRT: ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase)</td>
<td>NP_001609</td>
<td>VDPD (536)</td>
<td>9.9</td>
<td>AAVDPDGE</td>
</tr>
<tr>
<td>BID: BH3 interacting domain death agonist</td>
<td>NP_001187</td>
<td>IED (75)</td>
<td>57.6</td>
<td>GRIEADSE</td>
</tr>
<tr>
<td>CASP3: caspase 3, apoptosis-related cysteine protease</td>
<td>NP_004337</td>
<td>IED (175)</td>
<td>54.4</td>
<td>CGIIETeD</td>
</tr>
<tr>
<td>CASP7: caspase 7, apoptosis-related cysteine protease</td>
<td>NP_001218</td>
<td>IQAD (198)</td>
<td>4.6656</td>
<td>DGQADeGe</td>
</tr>
<tr>
<td>CENPB: centromere protein B, 80 kDa</td>
<td>NP_001801</td>
<td>VDSD (457)</td>
<td>7.4448</td>
<td>GVVDSSDe</td>
</tr>
<tr>
<td>CHD4: chromodomain helicase DNA binding protein 4</td>
<td>NP_001264</td>
<td>VDPD (1312)</td>
<td>9.9</td>
<td>EVVDPPDYN</td>
</tr>
<tr>
<td>DFFA: DNA fragmentation factor, 45 kDa, alpha polypeptide</td>
<td>NP_004392</td>
<td>DETD (117)</td>
<td>0.0544</td>
<td>MEVTCDG</td>
</tr>
<tr>
<td>DFFA: DNA fragmentation factor, 45 kDa, alpha polypeptide</td>
<td>NP_004392</td>
<td>VTD (6)</td>
<td>0.0432</td>
<td>DVDETDG</td>
</tr>
<tr>
<td>FBG: fibrilin</td>
<td>NP_001427</td>
<td>VGPD (184)</td>
<td>23.85</td>
<td>DIVGDGL</td>
</tr>
<tr>
<td>FLNA: filamin A, alpha (actin binding protein 280)</td>
<td>NP_001447</td>
<td>VADPDRL</td>
<td>11.4048</td>
<td>TIKQDKY</td>
</tr>
<tr>
<td>G22P1: thyroid autoantigen 70 kDa (Ku antigen)</td>
<td>NP_001460</td>
<td>IssD (79)</td>
<td>22.3344</td>
<td>KISIeD</td>
</tr>
<tr>
<td>GRIA3: glutamate receptor, ionotropic, AMPA 3</td>
<td>NP_000819</td>
<td>ISND (416)</td>
<td>18.5328</td>
<td>QIISND</td>
</tr>
<tr>
<td>HARS: histidyl-tRNA synthetase</td>
<td>NP_001002</td>
<td>LGPD (48)</td>
<td>2.4804</td>
<td>AQLGPPRS</td>
</tr>
<tr>
<td>IARS: isoleucine-tRNA synthetase</td>
<td>NP_001522</td>
<td>VTD (983)</td>
<td>2.7</td>
<td>LDVTPDQS</td>
</tr>
<tr>
<td>L4 100K [Human adenovirus C]</td>
<td>AAQ19301</td>
<td>IEQD (481)</td>
<td>57.6</td>
<td>VIIEQD</td>
</tr>
<tr>
<td>MKI67: antigen identified by monoclonal antibody Ki-67</td>
<td>NP_004208</td>
<td>VCDE (1481)</td>
<td>0.2727</td>
<td>TPVCEeD</td>
</tr>
<tr>
<td>NUMA1: nuclear mitotic apparatus protein 1</td>
<td>NP_000167</td>
<td>VATD (1705)</td>
<td>4.1616</td>
<td>FQVADTAL</td>
</tr>
<tr>
<td>PMS1: PMS1 postmeiotic segregation increased 1</td>
<td>NP_000525</td>
<td>ISAD (496)</td>
<td>17.1072</td>
<td>SSISADe</td>
</tr>
<tr>
<td>PMS2: PMS2 postmeiotic segregation increased 2</td>
<td>NP_000526</td>
<td>VEK (493)</td>
<td>0.05</td>
<td>ARVEK</td>
</tr>
<tr>
<td>PMSCL2: polymyositis/scleroderma autoantigen 2, 100 kDa</td>
<td>NP_002676</td>
<td>VEG (252)</td>
<td>28.8</td>
<td>QQVEQDMF</td>
</tr>
<tr>
<td>POLR1A: polymerase (RNA I) I polypeptide A, 194 kDa</td>
<td>NP_056240</td>
<td>ICPD (448)</td>
<td>0.1</td>
<td>SVICPDMY</td>
</tr>
<tr>
<td>POLR2A: polymerase (RNA II) I (DNA directed) polypeptide A, 220 kDa</td>
<td>NP_000925</td>
<td>ITPD (370)</td>
<td>5.4</td>
<td>TVITPPD</td>
</tr>
<tr>
<td>PRKDC: protein kinase, DNA-activated, catalytic polypeptide</td>
<td>NP_008835</td>
<td>VGPD (2698)</td>
<td>23.85</td>
<td>KISVGPPD</td>
</tr>
<tr>
<td>SNRF70: small nuclear ribonucleoprotein 70 kDa polypeptide (RNP antigen)</td>
<td>NP_003083</td>
<td>LGND (409)</td>
<td>1.574796</td>
<td>ECLOND</td>
</tr>
<tr>
<td>SS: signal recognition particle 72 kDa</td>
<td>NP_008787</td>
<td>VPD (573)</td>
<td>2.7</td>
<td>PKYTPP</td>
</tr>
<tr>
<td>SSB: Sjogren syndrome antigen B (autoantigen La)</td>
<td>NP_003133</td>
<td>LEED (220)</td>
<td>1.4976</td>
<td>QKLEED</td>
</tr>
<tr>
<td>TOP1: topoisomerase (DNA I)</td>
<td>NP_003277</td>
<td>IEAD (15)</td>
<td>57.6</td>
<td>SIQIEAD</td>
</tr>
<tr>
<td>UBE4B: ubiquitination factor E4B (UDF2 homolog, yeast)</td>
<td>NP_006039</td>
<td>VDPD (123)</td>
<td>3.0096</td>
<td>SQVDVAG</td>
</tr>
<tr>
<td>UBT1: upstream binding transcription factor, RNA polymerase I</td>
<td>NP_055048</td>
<td>VPVD (220)</td>
<td>0.05</td>
<td>LKVRPDA</td>
</tr>
</tbody>
</table>

The bold printed amino acids in the extended cleavage site indicate hits with a G residue at position P2 detected by the high stringency filter. Numbers in brackets indicate cleavage site position in the amino acid sequence.

Figure 1. Sensitivity–specificity plot for granzyme B. x-axis: scores by the GraBCas program; y-axis: percentage of specificity or sensitivity.
Table 3. Analysis of cleavage sites of known caspase 3 substrates with GraBCas

<table>
<thead>
<tr>
<th>Caspase 3 substrate</th>
<th>Acc_number</th>
<th>Known cleavage site</th>
<th>Score by GraBCas</th>
<th>P6-P2' of cleavage site</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD1: adducing 1 (alpha)</td>
<td>NP_001110</td>
<td>DSVD (633)</td>
<td>0.357</td>
<td>TGDSDDAG</td>
</tr>
<tr>
<td>APAF1: apoptotic protease activating factor</td>
<td>NP_001151</td>
<td>SVD (271)</td>
<td>0.462672</td>
<td>DKVSVTSDV</td>
</tr>
<tr>
<td>ARHGIDB: Rho GDP dissociation inhibitor (GDI) beta</td>
<td>NP_001166</td>
<td>DELD (19)</td>
<td>22.4</td>
<td>DDDELSK</td>
</tr>
<tr>
<td>ATP2B4: ATPase, Ca++ transporting, plasma membrane 4</td>
<td>NP_001675</td>
<td>DEID (1080)</td>
<td>71.4</td>
<td>GLDEIDHA</td>
</tr>
<tr>
<td>BAD: BCL2-antagonist of cell death</td>
<td>NP_004313</td>
<td>EQU (14)</td>
<td>0.001632</td>
<td>FSEQESSS</td>
</tr>
<tr>
<td>BAX: BCL2-associated X protein</td>
<td>NP_004315</td>
<td>FQID (33)</td>
<td>0.002142</td>
<td>QFQIDQRA</td>
</tr>
<tr>
<td>BCL2: B-cell CLL/Jurkat tumor protein</td>
<td>NP_000624</td>
<td>DAGD (34)</td>
<td>0.0357</td>
<td>EDWAGDVG</td>
</tr>
<tr>
<td>BCL2L1: BCL2-like 1</td>
<td>NP_001182</td>
<td>HLAD (61)</td>
<td>0.027846</td>
<td>SWHLADSP</td>
</tr>
<tr>
<td>BCL2L1: BCL2-like 2</td>
<td>NP_001182</td>
<td>SSLL (76)</td>
<td>0.274176</td>
<td>HSSLLADR</td>
</tr>
<tr>
<td>BIRC2: baculoviral IAP repeat-containing 2</td>
<td>NP_001157</td>
<td>ENAD (372)</td>
<td>0.111384</td>
<td>GEENADFP</td>
</tr>
<tr>
<td>BLM: Bloom syndrome</td>
<td>NP_000048</td>
<td>TEVD (415)</td>
<td>5</td>
<td>LTLTVEFDN</td>
</tr>
<tr>
<td>BRCA1: breast cancer 1, early onset</td>
<td>NP_009225</td>
<td>DLLD (1154)</td>
<td>3.4272</td>
<td>PDDLLDDG</td>
</tr>
<tr>
<td>CAMK4: calcium/calmodulin-dependent protein kinase IV</td>
<td>NP_001735</td>
<td>YWID (31)</td>
<td>0.009496</td>
<td>FDWIDVGS</td>
</tr>
<tr>
<td>CASP8: cysteine-aspartate-specific protease 8</td>
<td>NP_001813</td>
<td>LTVEVD (2146)</td>
<td>1,428</td>
<td>1VEWDGAG</td>
</tr>
<tr>
<td>CTD: capping terminal domain</td>
<td>NP_001813</td>
<td>VISEVD (442)</td>
<td>4</td>
<td>VISEVGN</td>
</tr>
<tr>
<td>CDC6: CDC6 cell division cycle 6 homolog (S.cerevisiae)</td>
<td>NP_001245</td>
<td>LVRD (99)</td>
<td>0.0055692</td>
<td>RRLVFEDQ</td>
</tr>
<tr>
<td>CDC9: CDC9 cell division cycle protein</td>
<td>NP_000235</td>
<td>DEHF (212)</td>
<td>18.15</td>
<td>EEFEDVGS</td>
</tr>
<tr>
<td>CENP-E: centromere protein E</td>
<td>NP_001849</td>
<td>AVVD (499)</td>
<td>0.306</td>
<td>GNADVGDG</td>
</tr>
<tr>
<td>GLN: gelsolin</td>
<td>NP_000168</td>
<td>DQTD (403)</td>
<td>15.4224</td>
<td>QNADVGDG</td>
</tr>
<tr>
<td>GCLC: glutamate-cysteine ligase, catalytic subunit</td>
<td>NP_001489</td>
<td>AVVD (499)</td>
<td>0.306</td>
<td>GNADVGDG</td>
</tr>
<tr>
<td>EIF2S1: eukaryotic translation initiation factor 2, subunit 1 alpha, 35 kDa</td>
<td>NP_001124</td>
<td>SEVD (442)</td>
<td>4</td>
<td>VISEVGN</td>
</tr>
<tr>
<td>EIF2S1: eukaryotic translation initiation factor 2, subunit 1, alpha isoform, 35 kDa</td>
<td>NP_004085</td>
<td>AEVD (301)</td>
<td>1</td>
<td>ENAEDVGS</td>
</tr>
<tr>
<td>EIF2S1: eukaryotic translation initiation factor 2, subunit 1, alpha isoform, 35 kDa</td>
<td>NP_004085</td>
<td>DGDD (304)</td>
<td>0.0085</td>
<td>EVSDDDDA</td>
</tr>
<tr>
<td>FNTA: farnesyltransferase, CAAX box, alpha</td>
<td>NP_002018</td>
<td>VSD (59)</td>
<td>0.137088</td>
<td>GYVSFEDP</td>
</tr>
<tr>
<td>GCLC: glutamate-cysteine ligase, catalytic subunit</td>
<td>NP_001489</td>
<td>AVVD (499)</td>
<td>0.306</td>
<td>GNADVGDG</td>
</tr>
<tr>
<td>IFN: interferon</td>
<td>NP_001151</td>
<td>YVVD (391)</td>
<td>0.012436</td>
<td>GDYVVDSP</td>
</tr>
<tr>
<td>ISC: intermediate symptomatic cleaves</td>
<td>NP_001151</td>
<td>YVVD (391)</td>
<td>0.012436</td>
<td>GDYVVDSP</td>
</tr>
<tr>
<td>JUN: Jun oncogene</td>
<td>NP_001151</td>
<td>YVVD (391)</td>
<td>0.012436</td>
<td>GDYVVDSP</td>
</tr>
<tr>
<td>KRT18: keratin 18</td>
<td>NP_001245</td>
<td>SEVD (442)</td>
<td>4</td>
<td>VISEVGN</td>
</tr>
<tr>
<td>LRRK2: leucine rich repeat kinase 2</td>
<td>NP_001245</td>
<td>SEVD (442)</td>
<td>4</td>
<td>VISEVGN</td>
</tr>
<tr>
<td>LRRK2: leucine rich repeat kinase 2</td>
<td>NP_001245</td>
<td>SEVD (442)</td>
<td>4</td>
<td>VISEVGN</td>
</tr>
<tr>
<td>MAP2: microtubule-associated protein 2</td>
<td>NP_005901</td>
<td>DMVD (421)</td>
<td>0.1</td>
<td>SIMDVMSP</td>
</tr>
<tr>
<td>MDM2: Mdm2, transformed 3T3 cell double minute 2, subunit 1, alpha, 53 kDa</td>
<td>NP_002383</td>
<td>DVPD (361)</td>
<td>12,436</td>
<td>GFVPFDCK</td>
</tr>
<tr>
<td>NFkBIA: nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha</td>
<td>NP_006390</td>
<td>DRHD (32)</td>
<td>0.3332</td>
<td>LDRHBDHG</td>
</tr>
<tr>
<td>NFkBIA: nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha</td>
<td>NP_006390</td>
<td>DRHD (32)</td>
<td>0.3332</td>
<td>LDRHBDHG</td>
</tr>
</tbody>
</table>

The bold printed amino acids in the extended cleavage site indicate hits detected by the high stringency filter. Numbers in brackets indicate cleavage site position in the amino acid sequence.
Table 4. Analysis of cleavage sites of known non-substrates of granzyme B with GraBCas

<table>
<thead>
<tr>
<th>Granzyme B non-substrate</th>
<th>Acc_number</th>
<th>Best hit</th>
<th>Score by GraBCas</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIM21: 52 kD Ro/SSA autoantigen</td>
<td>NP_003132</td>
<td>LDPD(294)</td>
<td>1,0296</td>
</tr>
<tr>
<td>SSA2: 60 kD Ro/SSA autoantigen</td>
<td>NP_004591</td>
<td>VTTD(427)</td>
<td>1,4688</td>
</tr>
<tr>
<td>XRCC5: ATP-dependent DNA helicase II Ku80</td>
<td>NP_006964</td>
<td>FGTD(62)</td>
<td>0,313856</td>
</tr>
<tr>
<td>VCL: vinculin isoform VCL</td>
<td>NP_003364</td>
<td>LQSD(98)</td>
<td>0,316742</td>
</tr>
<tr>
<td>VCL: vinculin isoform meta-VCL</td>
<td>NP_054706</td>
<td>LQSD(98)</td>
<td>0,316742</td>
</tr>
<tr>
<td>TUBB2: tubulin, beta 2</td>
<td>NP_001060</td>
<td>VISD(26)</td>
<td>0,0376</td>
</tr>
<tr>
<td>CRP: C-reactive protein, pentraxin-related</td>
<td>NP_000286</td>
<td>LAED(26)</td>
<td>0,229132</td>
</tr>
<tr>
<td>SERPINA1: serine (or cysteine) proteinase inhibitor, clade A</td>
<td>NP_00002235</td>
<td>LAED(26)</td>
<td>0,229132</td>
</tr>
<tr>
<td>GSTA1: glutathione S-transferase A1</td>
<td>NP_665683</td>
<td>VEID(61)</td>
<td>0,8</td>
</tr>
<tr>
<td>PYGB: brain glycogen phosphorylase</td>
<td>NP_002853</td>
<td>IEED(129)</td>
<td>28,8</td>
</tr>
<tr>
<td>TF: transferring</td>
<td>NP_001054</td>
<td>VTLD(82)</td>
<td>0,2592</td>
</tr>
<tr>
<td>LTF: lactotransferrin</td>
<td>NP_002334</td>
<td>VTLD(79)</td>
<td>0,2592</td>
</tr>
<tr>
<td>LYZ: lysozyme precursor</td>
<td>NP_000230</td>
<td>RSTD(71)</td>
<td>0,016158</td>
</tr>
<tr>
<td>ORM1: orosomucoid 1 precursor</td>
<td>NP_000598</td>
<td>LAFD(133)</td>
<td>0,114566</td>
</tr>
<tr>
<td>F2: coagulation factor II precursor (Thrombin B-chain)</td>
<td>NP_000497</td>
<td>LDDE(306)</td>
<td>0,296528</td>
</tr>
</tbody>
</table>

Numbers in brackets indicate cleavage site position in the amino acid sequence.

In accordance with the results of Stennicke et al. (15). Absent amino acids included the charged residues R, E, H, K and the large residues Q, I, L, M, F, W and Y.

With GraBCas we provide a position specific scoring scheme for the prediction of cleavage sites for granzyme B and caspases 1–9. GraBCas offers an easy to use, concise user interface in register card format. The design of GraBCas specifically acknowledged the high variability of cleavage site recognition sequences. We validated our tool by scoring known substrates and demonstrated its usefulness for protein sequence analysis. GraBCas may contribute to a more comprehensive knowledge of caspase and granzyme B substrates and a better understanding of the biological roles of these enzymes.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

This study was supported by a grant of the Center of Bioinformatics/Saarbrücken supported by the Deutsche Forschungsgemeinschaft and by a grant from the Deutsche Krebshilfe (10-1966-Me4). Funding to pay the Open Access publication charges for this article was provided by the University of Saarland.

Conflict of interest statement. None declared.

REFERENCES