Review article

CaspNeuroD: a knowledgebase of predicted caspase cleavage sites in human proteins related to neurodegenerative diseases

Sonu Kumar and Piotr Cieplak*

SBP Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

Present address: S. Kumar, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

*Corresponding author: Tel: +1 858-646-3100; Fax: +1 858-795-5249; E-mail: pcieplak@sbpdiscovery.org

Abstract

Background: A variety of neurodegenerative diseases (NDs) have been associated with deregulated caspase activation that leads to neuronal death. Caspases appear to be involved in the molecular pathology of NDs by directly cleaving important proteins. For instance, several proteins involved in Alzheimer's disease, including β-amyloid precursor protein (APP) and presenilins, are known to be cleaved by caspases. Therefore, cell death pathway may play a central role in many neurological diseases, and targeting the important proteins that control the cell survival and death may potentially represent a therapeutic approach for chronic neurodegenerative disorders.

Findings: We developed CaspNeuroD, a relational database of *in silico* predicted caspase cleavage sites in human proteins associated with NDs. The prediction has been done on collection of 249 human proteins reported in clinical studies of NDs using the recently published CaspDB Random Forest machine-learning model. This database could be used for identifying new caspase substrates and further our understanding of the caspase-mediated substrate cleavage in NDs.

Conclusion: Our database provides information about potential caspase cleavage sites in a verified set of human proteins involved in NDs. It provides also information about the conservation of cleavage positions in corresponding orthologs, and information about the positions of single nucleotide polymorphisms and posttranslational modifications (PTMs) that may modulate the caspase cleavage efficiency.

Database URL: caspdb.sanfordburnham.org/caspneurod.php .

Introduction

Many neurodegenerative diseases (NDs), including brain trauma, Huntington's disease (HD), Parkinson's disease, Alzheimer's disease (AD), stroke, spinal cord injury and amyotrophic lateral sclerosis (ALS)—are associated with neuronal cell death [1]. Necrosis and apoptosis are two main mechanisms of cell death [2–4]. Necrotic cell death in the central nervous system follows acute ischemia or traumatic injury to the brain or spinal cord [5, 6]. In contrast, apoptotic cell death, also known as programmed cell death, can be a feature of both acute and chronic neurologic diseases [1, 3, 7]. In chronic NDs, it is the predominant form of cell death [8, 9]. In apoptosis, a biochemical cascade activates proteases that destroy proteins, that are required for cell survival, and activates other types of proteins that mediate programmed cell death. Caspases actively contribute to the molecular pathogenesis of these diseases.

Caspases are proteolytic enzymes that perform hydrolysis of the peptide bonds in proteins to regulate their function in biological pathway(s), including the immune response, DNA replication, cell cycle progression, cell proliferation and apoptosis [10, 11]. Until now, at least 15 distinct caspases have been identified in mammals [12]. Human caspases are divided into apoptotic (Caspase-2, -3, -6, -7, -8, -9 and -10) and inflammatory (Caspase-1, -4 and -5) members. The apoptotic members have been further sub-divided into initiators (Caspase-2, -8, -9 and -10) and effectors (Caspase-3, -6 and -7) [13]. The most prominent feature of caspase-specificity is that caspases cleave their substrates almost exclusively after Asp residues. The consensus cleavage motif, determined by analysis of known cleavage sites, is DXXD-G/A/S/T/N, pointing to the overlapping specificity of this family of enzymes [14–16].

During apoptosis, caspases initiate, coordinate and accelerate cell death and dismantling by cleaving crucial structural and enzymatic proteins. There are variety of ways in which caspase activity may contribute to chronic NDs such as HD and AD. One way is to eliminate damaged neurons that are beyond repair, which suggests, that cells can no longer cope with their toxic loads and caspase pathway is therefore activated. Importantly, several NDs are characterized by the accumulation of abnormal protein deposits, such as Aβ42 in senile plaques in AD and polyglutamine-containing aggregates in HD. An additional way by which caspase activity may contribute to neurodegeneration is generating toxic fragments from key substrates. For example, caspase cleavage products of huntingtin and other truncated polyglutamine-containing proteins are known to have increased toxicity in cell culture models [17–19]. Thus, preventing the caspase cleavage

of huntingtin, atrophin-1 and the androgen receptor protects cells from an apoptotic challenge [20–22]. Similarly, caspase cleavage of APP may generate fragments with toxic potential by facilitating the amyloidogenic production of Aβ42 [23].

In this study, we focus on the *in silico* prediction of caspase mediated proteolytic events in human proteins associated with NDs. We used our recently developed, accurate caspase substrate prediction algorithm [24] to understand the importance of the caspase cleavage events and their regulation in NDs. We created CaspNeuroD, a database of predicted caspase cleavage sites in human proteins involved in NDs. This database integrates information about the caspase cleavage positions; their conservation in orthologous proteins in 11 organisms and information about the single nucleotide polymorphisms (SNPs) and PTMs that may modulate caspase mediated proteolytic events.

Methods

Human neurodegerative diseases related proteins were extracted from the literature-based resource 'NeuroDNet' [25]. We collected >300 genes, which were reported in clinical studies of NDs (Figure 1A). We mapped them to verified set of human proteins from the Uniprot database [26] and retrieved 249 protein sequences for further analysis. We applied machine learning prediction method and the Random Forest (RF) classifier, as implemented in our CaspDB [24], to predict caspase cleavage sites in these proteins. The cleavage prediction method provides appropriate cleavage efficiency probability scores in the range 0–1 for every peptide bond in a protein. The score above the 0.5 threshold indicates that the peptide bond is cleaved. The CaspDB RF prediction model was constructed by combining the positional weight matrix characterizing each P5 to P3' positions and information related to predicted structural features, including secondary structure and disorder parameters. The CaspDB RF model is trained using known human caspase cleavage sequences from the curated CASBAH database [27]. The prediction model has been evaluated and discussed in our recent publication [24].

The results of the caspase cleavage predictions in proteins related to NDs are collected and presented in 'CaspNeuroD' knowledgebase. It is available to all users at caspdb.sanfordburnham.org/caspneurod.php. The information about proteins sequences in 11 organisms orthologous to human ND proteins were extracted from the OMG browser [28]. The curated dbPTM [29] and Humsavar [26, 30] databases were used to obtain information about experimentally known PTMs and SNPs in each protein,
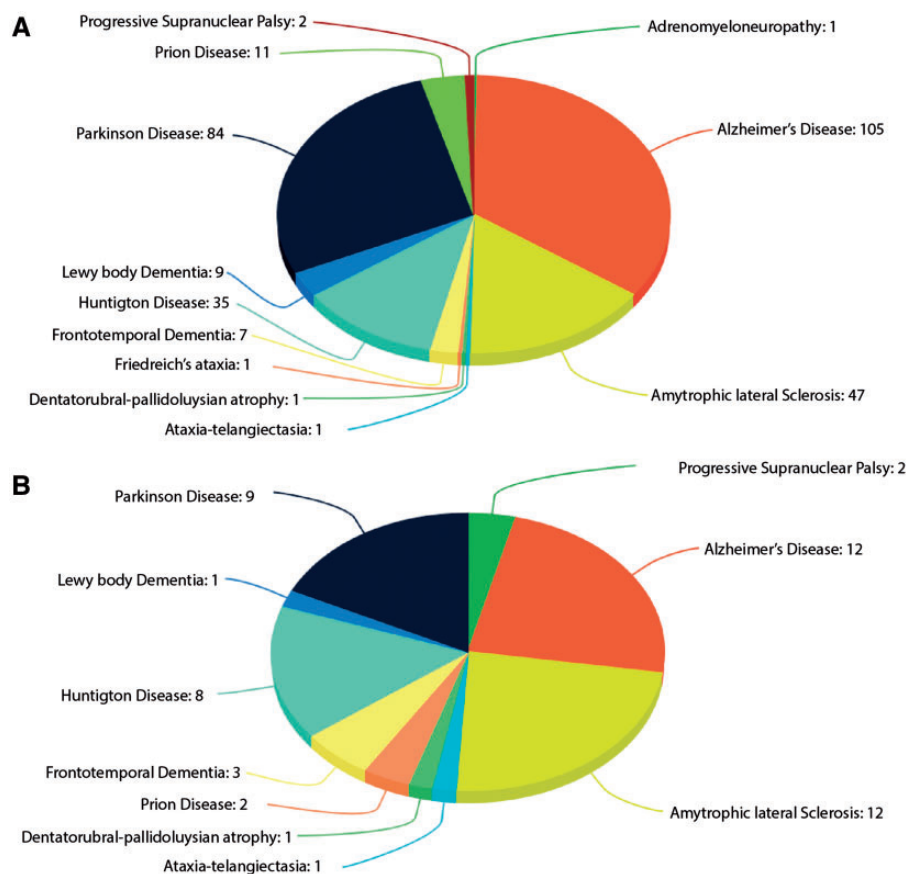
Figure 1. Proteins in ND. **(A)** The number of proteins known to be associated with ND; **(B)** The number of proteins associated with ND known to be caspase substrates.

respectively. We also located and present the positions of cleavage sites using graphical representation of the protein domains architecture according to information retrieved from the Pfam database (31).

CaspNeuroD is currently configured on an Apache (CentOS) server hosted at the SBP Medical Discovery Institute. It has been developed based on a combination of three layers. The underlying layer is the MySQL database system that stores all the information about the putative cleavage sites in proteins related to NDs and their orthologs along with the Pfam domains, SNPs and PTMs in the back-end. The intermediate layer is an Apache-PHP application that receives the query and connects to the database to fetch data from the upper layer, which comprises populated HTML and PHP pages, to the web browser client. ClustalW (32) and Jalview (33) were implemented to show the pairwise alignment and multiple sequence alignment, respectively.

Findings

CaspNeuroD is a web-based, platform-independent, database of predicted caspase cleavage sites in 249 human proteins

related to NDs. Among them we found 51 proteins, which are already known caspase substrates and are included in two caspase substrates databases: CASBAH and Degrabase (34). This subset of proteins is involved in 10 types of NDs (Figure 1B). The list of known caspase substrates involved in various types of NDs is presented in Supplementary file 1.

Recently, Julien *et al.* (35) published the results of the quantitative MS-based analysis of substrates of Caspase -2 and -6. In this article, it was demonstrated that some 235 and 871 proteins have been detected as substrates of Caspase-2 and -6, respectively. Among them 128 and 553 substrates of Caspase-2 and -6, respectively, have not been previously reported. These new substrates have been added to Degrabase database. Caspase-6 is implicated in ND, including Huntington's and Alzheimer's diseases (36). In the data presented by Julien *et al.* there are 24 and 23 substrates in common with our CaspNeuroD for Caspase-2 and -6, respectively. Among them 16 substrates are common for both of these enzymes.

On the front page of the CaspNeuroD database the user can choose one out of 12 NDs listed from a drop down query box in order to retrieve a list of appropriate proteins. All the proteins associated with a selected ND are shown

in tabular form. This table contains Uniprot identifier, gene name, gene id (linked to NCBI gene information), chromosome location, onset, and literature reference. To retrieve information about predicted caspase cleavage sites in a protein user can click on the individual Uniprot identifier. All the results related to proteolytic events are shown in a tabular form. The result page contains the information about: (i) the caspase mediated cleavage position (P1) with score values (in the range 0–1) arranged by default in descending order of score value, predicted secondary structure (α-helix: 'H', β-sheet: 'E', loop: '_') and disorder characteristics ('.'-ordered or '*'-disorder) for each residue at every P5-P3' position, and substrate prediction class ('yes'for cleavages with probabilities scores above 0.5 or 'no' otherwise), (ii) the presence of a signal peptide, description of the domains structure in graphical and tabular form according to PFAM annotation, (iii) the list of PTMs and SNPs, including disease annotation of the latter, (iv) multiple sequence alignment with available orthologs.

We used the SignalP v.4.0 program (37) to evaluate the presence of signal peptide characterizing secreted proteins. We also used PFAM knowledgebase for domain annotation to determine the inter- or intra-domain location of cleavage sites. If a given protein is experimentally annotated as caspase substrate and is reported in one of four known databases (MEROPS, CASBAH, Degrabase and TopFIND (27, 34, 38, 39)) then appropriate links to these databases are provided.

To investigate the conservation of a substrate's cleavage sites in other organisms, orthologous proteins from 11 organisms (Supplementary file 2) were retrieved. A special 'Compare' button is available for evaluating pair-wise comparison of cleavage sites between a given substrate and its orthologous proteins. A standard ClustalW pair-wise alignment aids in analysis of the conservation of cleavage sites. To display the multiple sequence alignment of a substrate and all its orthologous proteins, a 'Start Jalview' button is provided. The output page includes a list of SNPs and PTMs, with appropriate annotations, because both types of protein modification may influence the outcome of caspase-mediated proteolysis.

In summary, we provide a user-friendly knowledgebase for retrieving information about potential caspase cleavage sites for all verified human proteins associated with NDs. This database provides additional information that would be helpful in generating new hypotheses and in verification of new experimental findings concerning caspase-mediated cleavages of putative substrates. Information about cleavages in orthologous proteins is useful in assessing conservation of the cleavage positions across species, and thus assessing the confidence of the prediction. Overall, our database will complement ongoing experimental efforts in identifying role of new caspase substrates and further our understanding of the biochemistry of caspase-mediated substrate cleavages in NDs.

As more information about caspases and their substrates related to ND becomes available we will update our database.

Supplementary data

Supplementary data are available at *Database* Online.

Funding

References

1. Yuan,J., and Yankner,B.A. (2000) Apoptosis in the nervous system. *Nature*, 407, 802–809.
2. Kanduc,D., Mittelman,A., Serpico,R. *et al*. (2002) Cell death: apoptosis versus necrosis (review). *Int. J. Oncol.*, 21, 165–170.
3. Martin,L.J. (2001) Neuronal cell death in nervous system development, disease, and injury (Review). *Int. J. Mol. Med.*, 7, 455–478.
4. Wyllie,A.H., Kerr,J.F., and Currie,A.R. (1980) Cell death: the significance of apoptosis. *Int. Rev. Cytol.*, 68, 251–306.
5. Emery,E., Aldana,P., Bunge,M.B. *et al*. (1998) Apoptosis after traumatic human spinal cord injury. *J. Neurosurg.*, 89, 911–920.
6. Linnik,M.D., Zobrist,R.H., and Hatfield,M.D. (1993) Evidence supporting a role for programmed cell death in focal cerebral ischemia in rats. *Stroke*, 24, 2002–2008. discussion 2008–2009.
7. Martin,J.B. (1999) Molecular basis of the neurodegenerative disorders. *N. Engl. J. Med.*, 340, 1970–1980.
8. Smale,G., Nichols,N.R., Brady,D.R. *et al*. (1995) Evidence for apoptotic cell death in Alzheimer's disease. *Exp. Neurol.*, 133, 225–230.
9. Thomas,L.B., Gates,D.J., Richfield,E.K. *et al*. (1995) DNA end labeling (TUNEL) in Huntington's disease and other neuropathological conditions. *Exp. Neurol.*, 133, 265–272.
10. Dix,M.M., Simon,G.M., and Cravatt,B.F. (2008) Global mapping of the topography and magnitude of proteolytic events in apoptosis. *Cell*, 134, 679–691.
11. Los,M., Stroh,C., Janicke,R.U. *et al*. (2001) Caspases: more than just killers?. *Trends Immunol.*, 22, 31–34.
12. Chowdhury,I., Tharakan,B., and Bhat,G.K. (2008) Caspases - an update. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.*, 151, 10–27.
13. Pop,C., and Salvesen,G.S. (2009) Human caspases: activation, specificity, and regulation. *J. Biol. Chem.*, 284, 21777–21781.
14. McStay,G.P., Salvesen,G.S., and Green,D.R. (2008) Overlapping cleavage motif selectivity of caspases: implications for analysis of apoptotic pathways. *Cell Death Differ.*, 15, 322–331.

15. Stennicke,H.R., Renatus,M., Meldal,M., and Salvesen,G.S. (2000) Internally quenched fluorescent peptide substrates disclose the subsite preferences of human caspases 1, 3, 6, 7 and 8. *Biochem. J.*, 350(Pt 2), 563–568.

16. Thornberry,N.A., Rano,T.A., Peterson,E.P. *et al.* (1997) A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. *J. Biol. Chem.*, 272, 17907–17911.

17. Hackam,A.S., Singaraja,R., Wellington,C.L. *et al.* (1998) The influence of huntingtin protein size on nuclear localization and cellular toxicity. *J. Cell Biol.*, 141, 1097–1105.

18. Ikeda,H., Yamaguchi,M., Sugai,S. *et al.* (1996) Expanded polyglutamine in the Machado-Joseph disease protein induces cell death in vitro and in vivo. *Nat. Genet.*, 13, 196–202.

19. Martindale,D., Hackam,A., Wieczorek,A. *et al.* (1998) Length of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates. *Nat. Genet.*, 18, 150–154.

20. Bano,D., Zanetti,F., Mende,Y., and Nicotera,P. (2011) Neurodegenerative processes in Huntington's disease. *Cell Death Dis.*, 2, e228.

21. Ellerby,L.M., Andrusiak,R.L., Wellington,C.L. *et al.* (1999) Cleavage of atrophin-1 at caspase site aspartic acid 109 modulates cytotoxicity. *J. Biol. Chem.*, 274, 8730–8736.

22. Ellerby,L.M., Hackam,A.S., Propp,S.S. *et al.* (1999) Kennedy's disease: caspase cleavage of the androgen receptor is a crucial event in cytotoxicity. *J. Neurochem.*, 72, 185–195.

23. Gervais,F.G., Xu,D., Robertson,G.S. *et al.* (1999) Involvement of caspases in proteolytic cleavage of Alzheimer's amyloid-beta precursor protein and amyloidogenic A beta peptide formation. *Cell*, 97, 395–406.

24. Kumar,S., van Raam,B.J., Salvesen,G.S., and Cieplak,P. (2014) Caspase cleavage sites in the human proteome: CaspDB, a database of predicted substrates. *PloS One*, 9, e110539.

25. Vasaikar,S.V., Padhi,A.K., Jayaram,B., and Gomes,J. (2013) NeuroDNet - an open source platform for constructing and analyzing neurodegenerative disease networks. *BMC Neurosci.*, 14, 3.

26. Magrane,M. and Consortium, U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database*, 2011, bar009.

27. Luthi,A.U., and Martin,S.J. (2007) The CASBAH: a searchable database of caspase substrates. *Cell Death Differ.*, 14, 641–650.

28. Altenhoff,A.M., Gil,M., Gonnet,G.H., and Dessimoz,C. (2013) Inferring hierarchical orthologous groups from orthologous gene pairs. *PloS One*, 8, e53786.

29. Lu,C.T., Huang,K.Y., Su,M.G. *et al.* (2013) DbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. *Nucleic Acids Res.*, 41, D295–D305.

30. Yip,Y.L., Scheib,H., Diemand,A.V. *et al.* (2004) The Swiss-Prot variant page and the ModSNP database: a resource for sequence and structure information on human protein variants. *Hum. Mutat.*, 23, 464–470.

31. Finn,R.D., Bateman,A., Clements,J. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.*, 42, D222–D230.

32. Larkin,M.A., Blackshields,G., Brown,N.P. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947–2948.

33. Waterhouse,A.M., Procter,J.B., Martin,D.M. *et al.* (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189–1191.

34. Crawford,E.D., Seaman,J.E., Agard,N. *et al.* (2013) The DegraBase: a database of proteolysis in healthy and apoptotic human cells. *Mol. Cell. Proteomics*, 12, 813–824.

35. Julien,O., Zhuang,M., Wiita,A.P. *et al.* (2016) Quantitative MS-based enzymology of caspases reveals distinct protein substrate specificities, hierarchies, and cellular roles. *Proc. Natl. Acad. Sci. U S A*, 113, E2001–E2010.

36. Graham,R.K., Ehrnhoefer,D.E., and Hayden,M.R. (2011) Caspase-6 and neurodegeneration. *Trends Neurosci.*, 34, 646–656.

37. Petersen,T.N., Brunak,S., von Heijne,G., and Nielsen,H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8, 785–786.

38. Lange,P.F., Huesgen,P.F., and Overall,C.M. (2012) TopFIND 2.0–linking protein termini with proteolytic processing and modifications altering protein function. *Nucleic Acids Res.*, 40, D351–D361.

39. Rawlings,N.D., Waller,M., Barrett,A.J., and Bateman,A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.*, 42, D503–D509.