Biomedical Engineering and Computational Biology Insights

Libertas Academica
FREEDOM TO RESEARCH

OPEN ACCESS
Full open access to this and thousands of other papers at http://www.la-press.com.

ORIGINAL RESEARCH

Application of a Bioinformatics-Based Approach to Identify Novel Putative in vivo BACE1 Substrates

Joseph L. Johnson, Emily Chambers and Keerthi Jayasundera

Department of Chemistry and Biochemistry, University of Minnesota Duluth, Duluth, Minnesota, USA.
Corresponding author email: jljohns3@d.umn.edu

Abstract: BACE1, a membrane-bound aspartyl protease that is implicated in Alzheimer's disease, is the first protease to cut the amyloid precursor protein resulting in the generation of amyloid-β and its aggregation to form senile plaques, a hallmark feature of the disease. Few other native BACE1 substrates have been identified despite its relatively loose substrate specificity. We report a bioinformatics approach identifying several putative BACE1 substrates. Using our algorithm, we successfully predicted the cleavage sites for 70% of known BACE1 substrates and further validated our algorithm output against substrates identified in a recent BACE1 proteomics study that also showed a 70% success rate. Having validated our approach with known substrates, we report putative cleavage recognition sequences within 962 proteins, which can be explored using in vivo methods. Approximately 900 of these proteins have not been identified or implicated as BACE1 substrates. Gene ontology cluster analysis of the putative substrates identified enrichment in proteins involved in immune system processes and in cell surface protein-protein interactions.

Keywords: bioinformatics, BACE1, protease, Alzheimer's disease, protease substrates

Biomedical Engineering and Computational Biology Insights 2013:5 1–15

doi: 10.4137/BECB.S8383

This article is available from http://www.la-press.com.

Introduction

BACE1 (memapsin 2, β-secretase, Asp 2 protease) is a Type I membrane-bound aspartyl protease. It is highly expressed in the brain and pancreas, and the bulk of the enzyme, including catalytic domain, is extracytoplasmic (extracellular or luminal), with a short C-terminal tail containing a cell trafficking domain that directs it to the trans-Golgi network and endosomes.[1] Just over ten years ago it was identified by several groups as the protease responsible for the initial cleavage of the amyloid precursor protein (APP, also a Type I membrane protein) in the brain.[2–6] Subsequent cleavage of APP within its transmembrane domain by γ-secretase, a novel aspartyl protease protein complex with multiple membrane spanning α-helices, yields short peptide fragments primarily consisting of 40 or 42 amino acids termed amyloid-β (Aβ). Aggregation of the Aβ peptides forms plaques in the brain which are one of the hallmark pathological features of Alzheimer's disease (AD). The precise mechanisms by which these Aβ peptides exert their pathogenic effects in the brain are unknown, but soluble oligomers of Aβ have been shown to be involved in the synaptic dysfunction associated with AD.[7]

Due to its association with the production of Aβ and with AD, BACE1 has gained significant attention as an attractive AD therapeutic target for at least two reasons. Firstly, since it is the first protease to cleave APP on the pathway leading to Aβ formation, inhibiting it precludes γ-secretase cleavage from leaving APP to be processed via the non-pathogenic α-secretase pathway. Secondly, BACE1 knockout mice showed a mild, albeit complex phenotype and no detectable Aβ in the brain, whereas knocking out γ-secretase was embryonic lethal.[8–13] As is the case with many other aspartyl proteases, BACE1 has a relatively open active site and fairly loose specificity. Turner et al initially reported the subsite specificity for BACE1 by measuring the second order rate constant for the peptide hydrolysis within pools of octapeptide libraries, in which seven residues were held constant while substituting one of the 19 standard amino acids (cysteine omitted) for the remaining residue.[14] This was initially done for each of the P4 to P1 and P1' to P4' residues. Subsequent studies expanded the peptide substrates tested to include changes in residues P8 to P5.[15,16] These studies of BACE1 subsite specificity provide a cleavage sequence profile that can be adapted for bioinformatic studies.

Though the precise physiological function of BACE1 remains elusive, some have suggested that it acts as a sheddase.[17] Despite its relatively loose specificity, only a handful of in vivo BACE1 substrates have been identified, primarily through top down approaches. As mentioned above, APP is a known physiological BACE1 substrate. Another extensively characterized BACE1 substrate is the growth factor Neuregulin-1 (NRG1), a Type I membrane protein expressed on the surface of axons that interacts with the ErbB family of receptor tyrosine kinases. NRG1 is involved in the stimulation of Schwann cell proliferation and ultimately myelination.[18,19] This connection between BACE1 and NRG1 is borne out in the observation of hypo myelination in BACE1$^{-/-}$ knockout mice.[20] Another set of proteins identified as BACE1 substrates are the beta-subunits of voltage gated sodium channels (VGSCβ).[21,22] Wong et al demonstrated that BACE1 knockout cell lines showed a 50% reduction in the proteolytic processing responsible for the generation of the C-terminal fragment (CTF) of β1, β2, β3, and β4 VGSC subunits, but the residual 40%–50% activity suggests that other proteases are also involved in CTF formation.[22] Although the VGSCβ4 subunit has been predicted to be a better BACE1 substrate than β2, VGSCβ2 appears to be the only subunit that acts as a substrate in the brain cortex.[21] Other documented BACE1 substrates include beta-galactoside alpha-2,6-sialytrasferase 1 (ST6Gal I),[23,24] P-selectin glycoprotein ligand 1 (PSGL-1),[25] the APP-like proteins 1 and 2 (APLP1, APLP2),[26,27] low-density lipoprotein related receptor (LRP1),[28] interleukin-1 receptor type 2 (IL-1R-2),[29] the anti-aging protein Klotho,[30] and most recently membrane-bound prostaglandin E2 synthase-2 (mPGES-2).[31] These substrates are all Type I membrane proteins with the exception of ST6Gal I, which is Type II.

Since BACE1 remains an attractive target for AD therapeutics, knowing its in vivo substrates would be valuable for predicting and/or suggesting possible side effects to be aware of during clinical trials and beyond. Successful elucidation of the native substrates of any protease often requires a multifaceted approach. Proteomics studies can yield a less biased accounting of proteins cleaved upon overexpression of a given protease, but a potential drawback of this

approach is that overexpression can alter the native properties of the protease, such as its subcellular location, so that observed "hits" are not necessarily reflective of the native physiological activity. Another potential problem is that it can be difficult to definitively prove whether an observed proteolytic event was directly or indirectly associated with the overexpressed protease. An alternate approach already mentioned is to investigate subsite specificity using synthetic peptide libraries to give a systematic view of a protease's activity and specificity, but by necessity only a small subset of the possible peptide substrates can be synthesized and tested. For example, an aspartyl protease that binds eight amino acids in its active site would require the impossible feat of synthesizing 20^8 peptides to completely define its subsite preferences. Another approach that has been successfully employed in testing whether individual proteins are substrates for a given protease involves co-expressing the protease and its potential substrate in cell culture. This approach is only feasible if there is some a priori result or hypothesis suggesting that a protein is a substrate of a given protease. Finally, animal models can confirm that a protease-substrate pair do indeed give rise to a particular phenotype, but as was seen with BACE1/NRG1, sometimes the phenotype is not noticed until after the substrate has been identified by other means, which provides suggestions on where to search.[20]

Though the identification of native protease substrates can seem unwieldy, the combined results of the experimental approaches discussed can lead to success and ultimately positively impact the design of therapeutic agents. An underutilized method in the case of BACE1 is the use of bioinformatics to leverage the wealth of information contained in proteome databases. As with other methods, the goal with bioinformatics-based methods is to distil the vast amount of data to a point that minimizes false positives and false negatives while not missing the true substrates. We report here an approach that uses published in vitro subsite specificity data to drive a bioinformatics-based search of the human proteome for BACE1 in vivo substrates. We validated our approach by comparing our results to data for known in vivo BACE1 substrates and subsequently tested the method against a recently reported whole cell proteomics study aimed at elucidating putative in vivo BACE1 substrates by monitoring for proteins cleaved upon BACE1 overexpression in HeLa and HEK cell lines.[32]

Methods

Database of protein sequences from complete human proteome

We obtained 20,300 human protein sequences from the Universal Protein Resources. (UniProt, http://www.uniprot.org) complete proteome set (July 2010 release).[33,34] The dataset contained manually annotated and reviewed protein sequences comprised of only the full length isoforms.

Transmembrane domain prediction

Human protein sequences in FASTA format were submitted to the web-based transmembrane domain prediction server TMHMM v. 2.0 that is available from the Center for Biological Sequence Analysis (http://www.cbs.dtu.dk/services/TMHMM).[35] The short output format returned the number of TM domains, the predicted residue numbers of the TM domains identified, and the topology of the TM domains. Proteins were grouped according to the number of transmembrane domains, and the subset of proteins that had a single TM domain were evaluated for their potential as BACE1 substrates as outlined below. GPI anchored proteins are potential membrane-bound substrates for BACE1 as well. During GPI-anchored protein maturation, the C-terminal domain is removed and replaced by a GPI anchor. These proteins were included as single TM domain proteins with the site of the GPI anchor being numbered as though it were the first amino acid in the TM domain.

Signal peptide sequence prediction

Most proteins containing transmembrane domains also have signal peptides that target them to the ER and the secretory pathway. These hydrophobic sequences, which are removed as part of the transport process, tend to be misidentified by certain algorithms as TM domains. To prevent these sequences from being identified as potential BACE1 cleavage sequences, we sought to identify and annotate them according to their function as distinct from other protein regions. The human protein sequences in FASTA format were submitted to the signal peptide sequence prediction server SignalP v. 3.0 (http://www.cbs.dtu.

dk/services/SignalP).[36] The server used both neural network and Hidden Markov models trained on eukaryotic signal peptide sequences. The data output from the TMHMM and SignalP prediction servers were imported into the Microsoft Excel matrix described below.

Scoring matrix and Microsoft Excel macro

The final data required for the bioinformatics analysis were experimental measurements for the cleavage of various peptide sequences by BACE1. As mentioned previously, Turner et al performed such a study shortly after BACE1 was identified, in which they synthesized octapeptide libraries based on the human APP sequence (EVNLDAEF) that randomized a single position with all of the standard amino acids except for cysteine (because of its potential for disulfide bonding), while holding the amino acids in the other 7 positions constant.[14] These eight libraries were incubated with BACE1 and the resulting peptide fragments were quantified by MALDI-TOF mass spectrometry. Based on these results, the second order rate constant for each peptide was calculated and reported as a "preference index" for each subsite. These reported preference indices for each amino acid at each subsite were converted to numerical values that were then weighted by the coefficient of variation (CV). The standard deviation of the preference indices for a given subsite was divided by the mean for those same values. The CV is a measure of the dispersion of a given set of data; therefore, subsites that show more selectivity by preferring fewer amino acids at that subsite will have a higher weighting factor. The weighting factors for the P4-P4' sites were 0.84, 1.06, 1.14, 1.77, 1.15, 0.99, 0.61, and 0.58, respectively. These factors agree with the recent observation by Li et al that the P3-P2' sites of BACE1 are most critical in determining substrate reactivity.[15] Using these values, a score for each octapeptide was calculated by multiplying the weighted preference indices for all of the subsites together and was reported as the "score" for a given octapeptide. The preference indices with a value of zero were assigned a minimal value of 0.001. This reflected the lack of activity for a given amino at a particular subsite, while preventing potential "hits" that would be missed after multiplication by zero, essentially allowing for the possibility of some error in the original mass spectrometric measurements of the second order rate constant.

We wrote a macro in Microsoft Excel Visual Basic to import and analyze the protein sequences, to calculate the score for each sequence, and to sort them according to their location in each protein sequence. For the proteins with a single TM domain, text files containing the UniProt ID, protein sequence in FASTA format, TM domain residue numbering, SignalP signal peptide prediction data, and orientation of membrane protein were imported into an Excel spreadsheet. Proteins that had an undefined orientation in the membrane were determined by manually comparing the UniProt annotations to the TMHMM prediction and were included in the database in both orientations. The macro returned the score for each octapeptide sequence, and sequences that had a score above the threshold value of 1.0×10^{-5} were retained in a matrix. This threshold value was selected relative to the score of 1.0×10^{-3} for the native APP sequence (EVKMDAEF) known to be cleaved by BACE1. We reasoned that an additional two orders of magnitude below this value was a reasonable range to reduce false negative results while minimizing the total number of sequences returned. Scores based on sum of the weighted preference indices were measured but not used because, as expected, they did not correlate well due to their inability to distinguish between sequences with acceptable preference indices at each subsite from those that had mixtures of very poor and very good preference indices. For protein sequences reaching the threshold, the results were sorted based on their position relative to the TM domain, according to which side of the membrane they were on, and whether they were Type I or Type II proteins. Octapeptide sequences less than eight residues away from the TM were rejected because one or more residues were part of the TM domain. BACE1 cleavage of proteins as far as 50 amino acids away from their TM domain have been reported and therefore the upper limit was set at 52, which allowed for some flexibility due to imprecise prediction of the exact beginning and end of TM domains. Predicted substrate sequences that fell within the TM domain itself or within a signal peptide sequence were removed and not considered further.

Gene ontology analysis

Hits returned by the algorithm were analyzed and grouped according to gene ontology (GO) terms. The UniProt IDs were submitted to the Gene Functional

Classification algorithm that is part of the DAVID Bioinformatics Resources (http://david.abcc.ncifcrf.gov/home.jsp). A total of 37 Sequences for the 962 proteins were submitted and 33 of these were not found in the database because they have unknown functions and therefore no GO terms associated with them. This list was analyzed by the Functional Annotation Tool by generating the list of terms that showed up more than would be predicted by chance in the human proteome and then grouping them into clusters of overlapping or synonymous terms. The scores are reported as *P*-values, but they are actually a relative measure.

Results

Generation of the single TM domain subset of the complete human proteome

Submission of the complete human proteome set of protein sequences to the TMHMM prediction server yielded 2364 proteins (~11.5%) with 1 TM domain. Approximately 77% had 0 TM domains, while there were about 2% each of proteins containing 2, 6, or 7 TM domains. The remaining 6% was scattered among proteins with 3–5 or 8–23 TM domains. These data were evaluated to determine how well the TMHMM prediction server performed relative to the annotations contained in the UniProt database by taking the UniProt IDs from the 0 TM domain subset and searching for the term "transmembrane", which returned 171 proteins or 0.8% of the TM containing protein sequences that were missed. These proteins were added to the 1 TM domain dataset using the annotations from UniProt. The 2 TM and 3 TM subsets were then analyzed for instances where TMHMM overpredicted the number of TM domains. There were 220 proteins (1.1%) in the 2 TM subset, which according to UniProt annotations, only had 1 TM domain. For the large majority of these proteins, one of the TM domains was predicted by SignalP and annotated by UniProt as being a signal peptide sequence. This was not surprising when considering that signal peptide sequences tend to be rather hydrophobic. Only 7 proteins predicted by TMHMM to have 3 TM domains had 1 TM according to UniProt. Overall, the TMHMM algorithm categorized approximately 13% of the 1 TM proteins differently than UniProt. Roughly half of these were identified as 2 TM proteins which were actually 1 TM proteins with signal sequences. The remaining 6% discrepancy likely represents minor differences in how the 1 TM proteins are identified with each method having its own minor sources of error.

Ninety-seven 1 TM proteins had an ambiguous orientation in the membrane according to UniProt. These protein sequences were analyzed as both Type I and Type II proteins. Amazingly, none of these proteins returned peptide sequences that exceeded the threshold limits when analyzed by the macro as Type II proteins. GPI anchored proteins were the last to be included in the single TM subset. Although these proteins do not have a transmembrane α-helix, they are associated with the membrane through a GPI anchor attached to the C-terminus of the protein. This GPI anchor is added with concomitant removal of a C-terminal protein domain. As mentioned in the Methods, the distance from the TM domain was counted from the residue attached to the GPI anchor.

Summary of results

There were over 11,000,000 amino acids in the 20,300 proteins from the complete human proteome and more than 10,860,000 octapeptide sequences to analyze for their predicted ability to serve as BACE1 substrates. The initial stage of screening, done to identify proteins with a single TM domain, reduced the number of proteins to analyze down to 3085 protein sequences, 97 of which were duplicated because of their ambiguous orientation in the membrane. A total of 39,864 octapeptide sequences (of the approximately 1,600,000 possible) had scores exceeding the threshold of 1.0×10^{-5}. Of these 10.8% were within the TM domain, 12.2% fell within the signal peptide sequence, 20.0% were cytoplasmic, and 56.9% were extracytoplasmic (extracellular or luminal). Of the 56.9% of sequences that were extracytoplasmic, 7.7% (4.4% of the total) met both threshold requirements, having a score $> 1.0 \times 10^{-5}$ and being within 8–52 residues of the TM domain. This equated to 1748 octapeptide sequences of the roughly 1,600,000 possible (~0.11%) contained within 962 different proteins—a significant reduction in number of sequences to consider.

Hits among known BACE1 substrates

Once the data collection and sorting were completed, the results were surveyed to evaluate how well the algorithm had successfully predicted the known BACE1 substrates as hits. As shown in Table 1,

Table 1. Predicted BACE1 cut sites for known substrates.

UniProt ID	Protein	Topology	Predicted cleavage recognition		
			Site	Sequence	Score
P05067	APP	Type I	13	LVFFAEDV	8.44E-03
			33	EVKMDAEF	1.02E-03
			41	NIKTEEIS	6.04E-05
Q06481	APLP2	Type I	9	REDFSLSS	1.20E-03
			30	MIFNAERV	7.23E-05
			44	DENMVIDE	3.55E-03
P27930	IL-1R-2	Type I	16	TLSFQTLR	1.02E-03
Q02297	NRG1	Type I	11	QEKAEELY	6.14E-05
P56975	NRG3	Type I	11	FMESEEVY	2.07E-05
			13	IEFMESEE	2.52E-04
			14	GIEFMESE	4.50E-02
Q8IWT1	VGSCβ4	Type I	15	TIFLQVVD	3.58E-01
Q9NY72	VGSCβ3	Type I	44	EFEFEAHR	1.09E-05
Q07699	VGSCβ1	Type I	21	EHNTSVVK	1.03E-04
			28	LLFFENYE	1.09E-05
			29	RLLFFENY	1.73E-05
Q14242	PSGL-1	Type I	21	ASNLSVNY	8.30E-05
O60939	VGSCβ2	Type I	None		
Q9H7Z7	mPGES-2	Type I	None		
P51693	APLP1	Type I	None		
Q07954	LRP1	Type I	None		
P15907	ST6Gal I	Type II	None		

the macro correctly identified 9 Type I substrates out the 13 known in vivo substrates. Each of these had a score over the threshold and at least one predicted cut site in the extracytoplasmic juxtamembrane domain. A cleavage recognition site of 13 for a Type I membrane protein, for example, means that the 13th amino acid from the transmembrane domain is the P4 residue and that the octapeptide sequence would span the range 13–6 with the protein cleavage occurring between residues 10 and 9. APP and APLP2 were each identified with three potential cut sites, while the closely related APLP1 was not identified as having any predicted cut sites. The BACE1 cleavage sequences for APP at sites 13 and 33 were LVFFAEDV and EVKMDAEF, respectively. These are recognition sequences that have been described previously,[8] the second corresponding to the canonical site for the generation of Aβ. The sequence for the mutant Swedish APP protein was not included in the standard proteome database. Three of the four beta subunits of the voltage gated sodium channels (β1, β3, and β4) were successfully identified; VGSCβ2, however, was not. NRG1, IL-1R-2, and PSGL-1 did have predicted recognition sequences while mPGES-2 and the Type II protein ST6Gal I did not. The octapeptide recognition sequences for all of the hits can be found in Table S1.

Validation of the algorithm for BACE1 substrates identified by proteomics

Hemming et al recently reported a quantitative proteomics study utilizing two human epithelial cell lines overexpressing BACE1.[32] This study reported 68 putative substrates, many of which had not been identified previously. This provided an excellent opportunity to evaluate the validity of the substrate prediction algorithm beyond the more well-characterized BACE1 substrates with a larger dataset. The macro successfully predicted 70% of the BACE1 protein substrates reported. One of these, Glypican-3, was a GPI anchored protein and the remainder were Type I membrane proteins. No Type II membrane proteins were positively identified, but this is not surprising given that only a very small percentage of BACE1 substrates have been identified to date using quantitative proteomics or other methods. For the remaining 30% of proteomics-based substrates that

were not identified, two were GPI anchored proteins, one was a Type II membrane protein, and the rest were Type I membrane proteins. As was the case with the known BACE1 substrates, the predicted cleavage recognition sequences did not a show a clear consensus in their scores or in their distance from the TM domain. Others have reported this observation as well and suggested that at least some of this variability could be attributed to the fact that both enzyme and substrate are membrane-bound and so the energetics and properties of recognition, binding, and cleavage would be different from those of non-membrane associated enzymes and substrates.[15] It is not likely that all of the BACE1 substrates identified by quantitative proteomics will prove to be native substrates, a point that was made by the authors themselves.[32] For example, although BACE1 is listed as a substrate for itself, further work showed that there was not a direct correlation and that proteolysis of BACE1 was catalyzed by a different protease.

Novel BACE1 substrates predicted by bioinformatics

As mentioned earlier, our study returned 1748 potential octapeptide recognition sequences in 962 different protein sequences (Table S1). Table 3 gives the results for those sequences which had a score greater than 0.01 and were not listed previously. The only sequence with a score greater than 1 came from the T cell immunoreceptor with Ig and ITIM domains protein. The next seven peptide sequences with scores between 1 and 0.1 come from proteins involved in immune response, calcium-dependent exocytosis, disulfide formation, cytokine signaling, and trafficking. As an example from peptides scoring between 0.1 and 0.01, a conserved sequence (PLDLAVFW) in the family of nine UDP-glucuronosyltransferase 1 proteins is predicted to be a strong BACE1 substrate. Many of the top scoring sequences are composed of negatively charged and hydrophobic amino acids, consistent with the preference table values. As is the case for the known BACE1 substrates, there were a variety of predicted cleavage recognition sites ranging from 8–50 in Table 3 and 8–52 in the Table S1.

Gene ontology (GO) analysis

GO analysis of the complete set of 962 proteins identified by the prediction algorithm as BACE1 substrates was performed using DAVID bioinformatics resources from the NIAID at NIH to search and then cluster GO terms to identify the enrichment of biological themes within a list of genes or proteins.[37] As expected based on the predicted BACE1 substrates dataset, the terms "membrane protein" and "transmembrane" were associated with almost all of the proteins. The other common clusters that were returned are shown in Table 4 with their enrichment score and representative terms that were included in a given cluster. The enrichment score for a group is based on the combination of the EASE scores (a modified Fisher Exact *P*-Value scores) from the members of the group, with a higher score indicating a greater enrichment. Processes involved in cell-surface protein-protein or small molecule interactions, such as immunoglobulins, integrins, leucine-rich repeat proteins, and receptors, were the most highly enriched terms in the list of predicted BACE1 substrates.

Discussion

Identification of in vivo substrates for proteases is a difficult task, especially those that have relatively loose subsite specificity and/or a large active site that accommodates a longer peptide chain. Both of these conditions apply to BACE1.[14,38] In addition to these challenges, sub-cellular localization also determines whether proteins with the potential to be substrates are actually proteolyzed in vivo. Because of its promising potential as a therapeutic target for Alzheimer's disease, BACE1 has been studied extensively to elucidate its subsite specificity as well as its ability to cleave proteins in cell-based proteomics assays. Very recently, Turner et al extended their analysis of the subsite specificity of BACE1 from eight subsites (P4-P4') to twelve (P8-P4').[14,15] Both studies utilized synthetic peptide libraries in which one position of the peptide was randomized with each of the standard amino acids (except cysteine) while holding the other positions constant. These libraries were then incubated with BACE1 and analyzed by mass spectrometry to determine a relative second order rate constant normalized to the Swedish APP sequence (EVNLDAEF). Inherent in this approach was the assumption that neighboring peptide residues did not significantly interact with one another. The fact that they and we have used these preference indices to successfully identify a significant number of known

Table 2. Predicted BACE1 cut sites for substrates identified by Hemming et al[32] proteomics study.

UniProt ID	Protein	Topology	Predicted cleavage recognition		
			Site	Sequence	Score
P05067	APP	Type I	13	LVFFAEDV	8.44E-03
			33	EVKMDAEF	1.02E-03
			41	NIKTEEIS	6.04E-05
Q06481	APLP2	Type I	9	REDFSLSS	1.20E-03
			30	MIFNAERV	7.23E-05
			44	DENMVIDE	3.55E-03
P40189	Interleukin-6 receptor beta chain	Type I	17	GPEFTFTT	9.00E-05
			35	DTLYMVRM	2.17E-03
P08581	Hepatocyte growth factor receptor	Type I	29	NSELNIEW	1.22E-05
O75976	Carboxypeptidase D	Type I	22	DAASSVVI	4.17E-05
P29317	Ephrin type A receptor 2	Type I	15	VHEFQTLS	2.32E-03
			28	QALTQEGQ	1.43E-04
P54764	Ephrin type A receptor 4	Type I	44	NPLTSYVF	6.06E-05
Q15375	Ephrin type A receptor 7	Type I	16	GKMFEATA	5.55E-03
			25	DVATLEEA	2.89E-05
			40	RAFTAAGY	2.89E-05
P54760	Receptor protein tyrosine kinase variant EPHB4V1	Type I	16	QTQLDESE	6.70E-04
			41	GASYLVQV	1.20E-05
Q92823	Neuronal cell adhesion molecule 1	Type I	14	GPAMASRQ	2.46E-05
P32004	Neuronal cell adhesion molecule L1	Type I	24	RHQMAVKT	5.75E-05
			38	DTDYEIHL	2.83E-04
			40	QPDTDYEI	2.96E-04
Q9NPR2	Semaphorin-4B	Type I	39	GVADQTDE	7.20E-05
Q9C0C4	Semaphorin-4C	Type I	25	EGYLVAVV	1.17E-05
Q9H2E6	Semaphorin-6A	Type I	31	DPLGAVSS	2.07E-05
Q96JA1	Leucine-rich repeats and immunoglobulin-like domains protein 1	Type I	51	TPDNQLLV	5.72E-05
O94898	Leucine-rich repeats and immunoglobulin-like domains protein 2	Type I	28	HIYLNVIS	1.28E-04
Q6UXM1	Leucine-rich repeats and immunoglobulin-like domains protein 3	Type I	51	IVDSDVSD	7.11E-05
Q9Y6N7	Roundabout homolog 1	Type I	9	QISDVVKQ	2.36E-05
			15	QVSLAQQI	4.11E-04
			47	EVAASTGA	1.99E-05
Q9HCK4	Roundabout homolog 2	Type I	47	EVAASTSA	1.75E-05
Q7Z5N4	Sidekick-1	Type I	17	NPSTAVSA	3.82E-05
Q58EX2	Sidekick-2	Type I	38	GVSYDFRV	3.74E-04
			52	EVSSYTFS	3.77E-05
P15151	Poliovirus receptor	Type I	23	QAELTVQV	5.00E-04
Q92673	Sortilin-related receptor	Type I	14	GADASATQ	2.07E-05
			22	LLYDELGS	1.02E-05
			23	ILLYDELG	1.89E-04
			46	GHNYTFTV	8.20E-05
Q96JP9	Protocadherin 21 (cadherin-related family member 1)	Type I	15	MAAFLIQT	6.23E-05
			17	SPMAAFLI	1.45E-05
			26	ITDAETLS	2.20E-05
			39	SPSFSTTA	5.71E-05
Q9Y5H2	Protocadherin gamma A11	Type I	11	LANSETSD	3.08E-05
			20	LADLGSLE	3.89E-05
			22	EVLADLGS	9.31E-05
			40	PPLSATVT	1.54E-05
Q9Y5G8	Protocadherin gamma A5	Type I	8	PEDLDLTL	1.03E-02
			22	DILADLGS	7.29E-05
Q9Y5G5	Protocadherin gamma A8	Type I	9	DPNDSSLT	6.06E-05
			22	EVLTELGS	1.67E-03
			40	PPLSATVT	1.54E-05
Q9UN70	Protocadherin Gamma C3	Type I	40	EPSLSTTA	3.88E-03

(*Continued*)

Table 2. (*Continued*)

UniProt ID	Protein	Topology	Predicted cleavage recognition		
			Site	Sequence	Score
Q86VZ4	Low-density lipoprotein receptor-related protein 11	Type I	23	EESYIFES	3.20E-05
O75096	Low-density lipoprotein receptor-related protein 4	Type I	37	RTSLEEVE	9.63E-03
			47	TTLYSSTT	1.08E-05
P31431	Syndecan-4	Type I	43	PKKLEENE	1.67E-05
MULTIPLE	HLA class I histocompatibility antigen (Combined)	Type I	9	EPSSQSTV	3.00E-05
Q13332	Receptor-type tyrosine protein phosphatase S	Type I	8	IVDGEEGL	2.82E-05
Q13740	CD166 antigen	Type I	19	DEADEISD	1.29E-04
Q12907	Vesicular integral-membrane protein VIP36	Type I	52	MKLFQLMV	1.20E-03
Q5VU97	Cache domain containing 1	Type I	19	DDMGAIGD	2.22E-05
Q9BYH1	Seizure 6-like protein 2	Type I	12	EAAAETSL	1.25E-05
			19	EHALEVAE	5.97E-02
			51	ELMGEVTI	3.82E-03
Q92859	Neogenin	Type I	45	MPNDQASG	1.60E-05
Q6UVK1	Chondroitin sulfate proteoglycan 4	Type I	9	LSFLEANM	3.03E-04
			12	GGFLSFLE	9.84E-05
Q24JP5	Transmembrane protein 132A	Type I	8	VTELELGM	4.24E-04
Q13145	BMP and activin membrane-bound inhibitor homolog	Type I	14	QELTSSKE	1.42E-04
Q14126	Desmoglein 2	Type I	10	QHDSYVGL	9.29E-05
			46	EIQFLISD	2.81E-03
Q9NZV1	Cysteine-rich motor neuron 1 protein	Type I	45	EVDLEVPL	1.12E-03
Q92896	Golgi apparatus protein 1	Type I	13	DLAMQVMT	4.21E-03
			15	FSDLAMQV	1.88E-04
Q9NR96	Toll-like receptor 9	Type I	47	DFLLEVQA	1.55E-03
			48	MDFLLEVQ	8.73E-05
			49	FMDFLLEV	1.41E-04
			51	AAFMDFLL	3.58E-04
O75509	Tumor necrosis factor receptor superfamily member 21	Type I	37	LPSMEATG	3.14E-04
P51654	Glypican-3	GPI	31	AYDLDVDD	2.48E-05
			33	ELAYDLDV	1.30E-03
			35	LAELAYDL	3.35E-04
P51693	APLP1	Type I	None		
Q99523	Sortilin	Type I	None		
Q5ZPR3	CD276 antigen	Type I	None		
P19021	Peptidyl-glycine alpha-amidating monooxygenase	Type I	None		
Q6UX71	Plexin domain-containing protein 2	Type I	None		
P35613	Basigin	Type I	None		
O95185	Netrin receptor UNC5C	Type I	None		
Q8TB96	T-cell immunomodulatory protein	Type I	None		
O14672	Disintegrin and metalloproteinase domain-containing protein 10	Type I	None		
O43291	Kunitz-type protease inhibitor 2	Type I	None		
O43493	Trans-golgi network integral membrane protein 2	Type I	None		
Q12860	Contactin-1	GPI	None		
Q8NFY4	Semaphorin-6D	Type I	None		
O00592	Podocalyxin-like protein 1	Type I	None		
P56817	Beta-secretase 1	Type I	None		
Q2VWP7	Protogenin	Type I	None		
P78504	Jagged-1	Type I	None		
P11717	Cation-independent mannose-6-phosphate receptor	Type I	None		
Q86YC3	Leucine-rich repeat-containing protein 33	Type I	None		
P52803	Ephrin-A5	GPI	None		
O00461	Golgi phosphoprotein 4	Type II	None		

Table 3. Predicted cut sites and scores for novel putative BACE1 substrates from the human proteome.

UniProt ID	Protein	Predicted cleavage recognition		
		Site	Sequence	Score
Q495A1	T cell immunoreceptor with Ig and ITIM domains	21	RIFLEVLE	1.03E+00
O95470	Sphingosine-1-phosphate lyase 1	28	EPYLEILE	8.08E-01
P12314	High affinity immunoglobulin gamma Fc receptor I	14	ELELQVLG	5.50E-01
Q9BZM6	NKG2D ligand 1	22	EEFLMYWE	4.48E-01
Q9NP60	X-linked interleukin-1 receptor accessory protein-like 2	46	EVELALIF	2.07E-01
Q13445	Transmembrane emp24 domain-containing protein 1	49	EEMLDVKM	1.58E-01
Q5T7P8	Synaptotagmin-6	46	QEALAVLA	1.16E-01
Q6ZRP7	Sulfhydryl oxidase 2	8	GVDFSSLD	1.09E-01
A0PJX4	Protein shisa-3 homolog	50	PEDFDTLD	9.03E-02
Q96A26	Protein FAM162A	17	TVSLEMLD	7.63E-02
	UDP-glucuronosyltransferase 1 family (combined)	37	PLDLAVFW	7.42E-02
P60509	HERV-R(b)_3p24.3 provirus ancestral env polyprotein	40	NISLALED	7.41E-02
Q4ADV7	Protein RIC1 homolog	35	DENFSTLS	6.68E-02
Q3SXP7	Uncharacterized protein KIAA1644	26	ETEFQAVM	6.15E-02
O95140	Mitofusin-2	32	QEEFMVSM	6.07E-02
Q96FB5	UPF0431 protein C1orf66	16	PLNLAALQ	6.01E-02
O75578	Integrin alpha-10	15	ESLLEVVQ	5.55E-02
Q15363	Transmembrane emp24 domain-containing protein 2	21	QEYMEVRE	4.86E-02
Q5DX21	Immunoglobulin superfamily member 11	19	LLDLQVIS	4.74E-02
O43699	Sialic acid-binding Ig-like lectin 6	17	QISLSLFV	4.58E-02
O95971	CD160 antigen	35	GHFFSILF	4.32E-02
O60499	Syntaxin-10	37	GIMLDAFA	4.31E-02
Q6ZNB6	NF-X1-type zinc finger protein NFXL1	35	QAELEAFE	3.98E-02
O95866	Protein G6b	48	ELLLSAGD	3.68E-02
Q86UW2	Organic solute transporter subunit beta	16	QELLEEML	3.62E-02
P26006	Integrin alpha-3	15	DIDSELVE	3.44E-02
Q9Y639	Neuroplastin	36	IVNLQITE	3.32E-02
Q6UWI2	Prostate androgen-regulated mucin-like protein 1	25	LIDMETTT	3.01E-02
A2A2Y4	FERM domain-containing protein 3	45	FEDLEADE	3.00E-02
Q6P7N7	Transmembrane protein 81	21	EVNLDSYS	2.88E-02
A6NFR6	Putative uncharacterized protein C5orf60	24	AVDMDILF	2.81E-02
Q8N386	Leucine-rich repeat-containing protein 25	20	QHNLSAFL	2.76E-02
Q9HBW1	Leucine-rich repeat-containing protein 4	12	QTSLDEVM	2.68E-02
Q9Y5Y7	Lymphatic vessel endothelial hyaluronic acid receptor 1	32	EVFMETST	2.65E-02
P0C6S8	Leucine-rich repeat neuronal protein 2	40	DTYFATLT	2.56E-02
Q6NUS6	Tectonic-3	43	EVSLTTLV	2.56E-02
Q8IYS5	Osteoclast-associated immunoglobulin-like receptor	48	EFFLEEVT	2.47E-02
Q9H5V8	CUB domain-containing protein 1	16	DLLFSVTL	2.34E-02
Q15399	Toll-like receptor 1	41	QVSSEVLE	2.29E-02
Q9Y2C9	Toll-like receptor 6	41	QVSSEVLE	2.29E-02
Q13651	Interleukin-10 receptor subunit alpha	47	HENFSLLT	2.28E-02
Q9Y5I0	Protocadherin alpha-13	34	TVLLSLVE	2.09E-02
Q68DV7	RING finger protein 43	28	EKLMEFVY	2.08E-02
Q6UX41	Butyrophilin-like protein 8	47	EISLTVQE	1.86E-02
Q15262	Receptor-type tyrosine-protein phosphatase kappa	45	NIYFQAMS	1.85E-02
Q5TH69	Brefeldin A-inhibited guanine nucleotide-exchange protein 3	14	DLLFELLR	1.76E-02
Q9Y5F3	Protocadherin beta-1	21	EPYLQFQD	1.63E-02
P29376	Leukocyte tyrosine kinase receptor	34	QAELQLAE	1.60E-02
Q86XX4	Extracellular matrix protein FRAS1	17	NLEMQELA	1.56E-02
P60507	HERV-F(c)1_Xq21.33 provirus ancestral Env polyprotein	34	ETSLLTLD	1.40E-02
Q5SWX8	Protein odr-4 homolog	47	IEDLEIAE	1.37E-02

(*Continued*)

Table 3. (*Continued*)

UniProt ID	Protein	Predicted cleavage recognition		
		Site	Sequence	Score
Q9H4D0	Calsyntenin-2	49	EFNLEVSI	1.35E-02
Q9P246	Stromal interaction molecule 2	44	EPSFMISQ	1.27E-02
A6BM72	Multiple epidermal growth factor-like domains protein 11	25	QAALMMEE	1.22E-02
Q9UQV4	Lysosome-associated membrane glycoprotein 3	23	DVQLQAFD	1.17E-02
Q6IEE7	Transmembrane protein 132E	8	LTDLEIGM	1.13E-02
Q96KV6	Butyrophilin subfamily 2 member A3	50	DSLFMVTT	1.11E-02
Q96MU8	Kremen protein 1	48	QANLSVSA	1.08E-02
P13598	Intercellular adhesion molecule 2	15	PKMLEIYE	1.06E-02
Q13421	Mesothelin	31	QDDLDTLG	1.05E-02
Q01638	Interleukin-1 receptor-like 1	34	EEDLLLQY	1.04E-02

BACE1 substrates and, even more importantly, to make specific predictions about the location of cut sites using computational methods, supports the validity and utility of their data.

Because attempting to identify the in vivo substrates for a protease with loose substrate specificity is difficult, a combination of approaches such as proteomics, bioinformatics, and in vitro biochemical measurements can and indeed have driven the ultimate identification of native substrates. The cleavage of APP at the β-site was known for several years before the discovery that the novel membrane bound aspartyl protease BACE1 was responsible for the observed β-secretase activity. Although several BACE1 substrates have been identified through careful observation, phenotypes arising from BACE1 activity can be subtle or nonexistent because some actual BACE1 substrates can be proteolyzed by other proteases such as BACE2 or α-secretase. Alternate strategies are needed to focus and inform in vivo studies. The complete kinetic assessment of BACE1 subsite specificity employing synthetic peptide libraries provides the powerful opportunity to extend their application to protein sequences as well. These data have demonstrated the promise of this approach, but it is apparent that further refinement is required. For example, despite the success of the algorithm in predicting the most likely cleavage sites for APLP2, it did not identify any for APLP1, which is known to be cleaved by BACE1.[32] Li et al made predictions for the BACE1 cleavage sites in two other known substrates, mPGES-2 and ST6Gal I, but how these cleavages happen at the proposed sites is not clear.[15] mPGES-2, a Type I membrane protein with a short extracytoplasmic domain and large cytoplasmic domain, is known to be cut between amino acids 87 and 88 to release it from the membrane, but this cleavage site is on the cytoplasmic side of the lipid bilayer.[39] Though a BACE1 cleavage site was predicted, it does not match the known site and how BACE1 can cleave at this intracellular peptide sequence is unclear. For the Type II membrane protein ST6Gal I, the original peptide sequence identified as a BACE1 substrate is actually from rat.[24] Surprisingly, this cleavage recognition sequence is not even conserved between rat and human, and according to our algorithm the changes to the human sequence would make it a worse substrate. The predicted cleavage site is 11 residues away from the transmembrane domain. Because the orientation of the peptide sequence is reversed for a Type II protein, it is unclear how BACE1 could cut so close to

Table 4. Gene ontology cluster analysis of putative BACE1 substrates from the bioinformatics analysis.

Enrichment score	Annotation cluster terms
85.1	Immunoglobulin domain (230)
72.8	Receptor (302), signal transducer (314)
61.5	Cell adhesion (209), cadherin (73), cation binding (186)
35.6	Fibronectin type III (76)
24.2	Immune response (108), immune system process (145), response to stimulus (232)
13.8	Integrin mediated signaling (27), regulation of actin cytoskeleton (31)
13.3	Cytokine binding (35), cytokine-cytokine receptor interactions (48), growth factor binding (32)
11.7	Leucine-rich repeat (51)

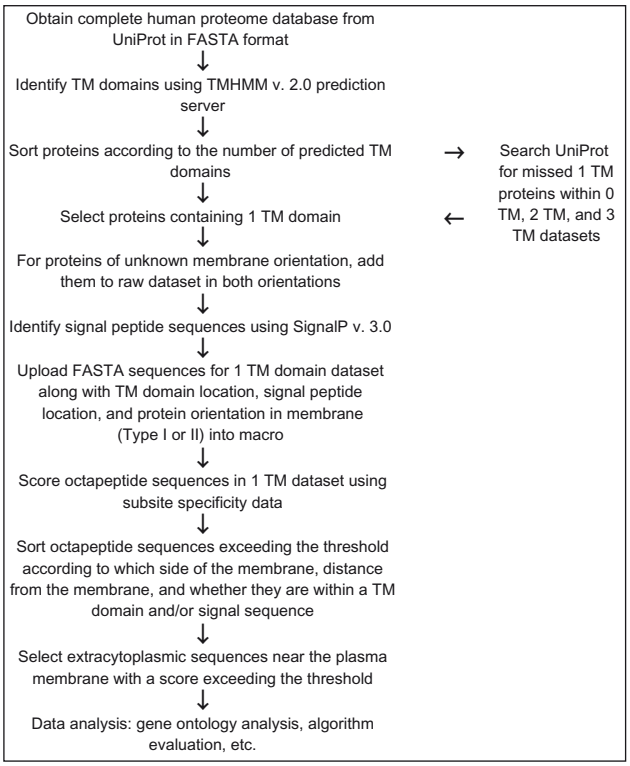

Figure 1. Schematic view of bioinformatics workflow.

the membrane and have the peptide sit in its active site in the proposed orientation.

In addition to the in vitro studies characterizing BACE1 activity with short peptide substrates, proteomic methods have also been used to guide the search for in vivo substrates.[32] One strength of this approach is that it does not bias the choice of peptide sequences to test for BACE1 activity, which was a necessary simplification when utilizing synthetic peptide libraries. Another advantage is that BACE1 is in its membrane-bound form and presumably exposed primarily to substrates that are membrane-bound as well. However, one drawback to this approach includes needing to limit the analysis to a few cell lines, some of which may not typically express BACE1. In addition, the in vivo data generated can only be for those proteins expressed in the particular cell line(s) chosen. This may be one explanation for the lack of identification of some of the known BACE1 substrates such as VGSCβ subunits, IL-1R-2, PSGL-1, LRP1, and NRG1, leading to false negative results. Another potential source of incorrect identification of BACE1 substrates that could yield false positive results arises from the overexpression of BACE1. Lee et al showed that BACE1 overexpression shifted the subcellular localization of APP cleavage to earlier points in the secretory pathway.[40] Since this happens for APP upon BACE1 overexpression, caution should be used when interpreting the results for other substrates identified by proteomics. Because the purpose of proteomic and in vitro studies is to narrow the list of potential proteins to investigate further for their in vivo activity, the studies' drawbacks do not present insurmountable problems as both the proteomics and in vitro approaches successfully identified known BACE1 substrates.

Combining bioinformatics with existing proteomics and in vitro data should give a more robust prediction of BACE1 in vivo substrates. This report adds to the BACE1 in vivo substrate discussion by utilizing a bioinformatics approach to both successfully predict the BACE1 cleavage sites for a large number of known substrates and to identify potential novel BACE1 substrates by extending the analysis to the entire human proteome. We first compared our results to the known BACE1 in vivo substrates. Nine of the thirteen substrates in Table 1 were positively identified using our algorithm. The predicted recognition cleavage sites and the cut sites for these nine proteins match the published data. Four of the proteins had three sites that met our criteria. In the case of APP, multiple BACE1 cleavage sites are known to be present.[8] Our method did not return positive identifications for mPGES-2, ST6Gal I, VGSCβ2, and APLP1. Our proposed explanation for not identifying mPGES-2 and ST6Gal I as potential substrates has been described above. For VGSCβ2, the score for the cleavage site reported by Li et al was 3.3×10^{-6}, just below our threshold. This result may necessitate changing the threshold, but we are currently investigating other methods that will reduce rather than increase the number of hits returned while capturing all of the known substrates. From both the proteomics and the in vitro studies, one would predict that APLP2 is a better substrate than APLP1. This is also the case with our bioinformatics data, which is not surprising since the preference indices from Turner et al were used in our scoring matrix as well. APLP1 was identified via proteomics, but it is not apparent why our method did not identify it as a substrate. One explanation could be due to the large number of cysteine residues in the juxtamembrane region for APLP1. Since cysteine was left out of the octapeptide

substrate libraries, scoring cysteine-rich sequences is not possible with our algorithm.

The proteomics data of Hemming et al were used to validate the efficacy of our method.[32] Approximately 70% of their reported substrates were correctly identified, and importantly, we report the predicted recognition sites for those cleavages. Because of the way the algorithm is currently written, no Type II protein would be identified as a substrate. Though the data for rat ST6Gal I is convincing, exactly how BACE1 recognizes and cleaves this sequence that is in the opposite orientation is not clear. Additionally, the human ST6Gal I sequence is not conserved in the rat sequence where the proteolysis by BACE1 was described. The fact that BACE1 substrates such as BACE1 itself were identified by proteomics, which upon further analysis were shown to be associated with a protease other than BACE1, highlights the need for complimentary information about substrates predicted via proteomics, whether from further biochemical or bioinformatics studies. With the solid foundation provided by this study, further refinement of our substrate prediction algorithm is underway to address the lack of identification of the remaining 30% of proteomics and known BACE1 substrates. Some of the substrates identified by proteomics may or may not turn out to be actual in vivo BACE1 substrates and definitive conclusions about the relative value of the bioinformatics or proteomics methods must be determined in further studies. Each method has value and unique strengths and weaknesses in guiding the search for native BACE1 substrates.

As is the case with the known substrates identified by in vivo and proteomics methods, the distance from the membrane for the cut recognition sites span the entire range between 8 and 52. Between Table 3 and the summary of the GO analysis in Table 4, the annotation clusters yielded a significant number of proteins in relatively few categories: A large number (230 of 962) contained immunoglobulin domains or were involved in immune response or immune system processes; just over 300 had functions related to receptors and signal transduction; proteins involved in protein-protein interactions including cell adhesion proteins accounted for 209 proteins, including some further subcategorized as cadherins, cation binding proteins, integrin proteins, and leucine-rich repeat proteins; and finally cytokines and their receptors are involved in processes such as growth factor binding. Efforts to refine the algorithm to improve its accuracy are underway, and though experiments to evaluate these putative BACE1 substrates in vivo are planned, they are beyond the scope of the present study.

Acknowledgements

This work was supported by the Swenson Family Foundation, the Swenson College of Science and Engineering, and the University of Minnesota Duluth.

Author Contributions

Conceived and designed the experiments: JLJ. Analysed the data: JLJ, EC, KJ. Wrote the first draft of the manuscript: JLJ. Contributed to the writing of the manuscript: JLJ, EC, KJ. Agree with manuscript results and conclusions: JLJ, EC, KJ. Jointly developed the structure and arguments for the paper: JLJ, EC, KJ. Made critical revisions and approved final version: JLJ. All authors reviewed and approved of the final manuscript.

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Willem M, Lammich S, Haass C. Function, regulation and therapeutic properties of [beta]-secretase (BACE1). *Semin Cell Dev Biol*. Apr 2009; 20(2):175–82. Epub Jan 20, 2009.
2. Haniu M, Denis P, Young Y, et al. Characterization of Alzheimer's β-secretase protein BACE. A pepsin family member with unusual properties. *J Biol Chem*. Jul 14, 2000;275(28):21099–106.

3. Hussain I, Powell D, Howlett DR, et al. Identification of a novel aspartic protease (Asp 2) as β-secretase. *Mol Cell Neurosci*. Dec 1999;14(6): 419–27.
4. Lin X, Koelsch G, Wu S, et al. Human aspartic protease memapsin 2 cleaves the β-secretase site of β-amyloid precursor protein. *Proc Natl Acad Sci U S A*. Feb 15, 2000;97(4):1456–60.
5. Sinha S, Anderson JP, Barbour R, et al. Purification and cloning of amyloid precursor protein beta-secretase from human brain. *Nature*. Dec 2, 1999; 402(6761):537–40.
6. Vassar R, Bennett BD, Babu-Khan S, et al. Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science*. Oct 22, 1999;286(5440):735–41.
7. Lauren J, Gimbel DA, Nygaard HB, et al. Cellular prion protein mediates impairment of synaptic plasticity by amyloid-[bgr] oligomers. *Nature*. Feb 26, 2009;457(7233):1128–32.
8. Cai H, Wang Y, McCarthy D, et al. BACE1 is the major beta-secretase for generation of Abeta peptides by neurons. *Nat Neurosci*. Mar 2001;4(3): 233–4.
9. Chiocco MJ, Kulnane LS, Younkin L, et al. Altered amyloid-β metabolism and deposition in genomic-based β-secretase transgenic mice. *J Biol Chem*. Dec 10, 2004;279(50):52535–42. Epub Sep 27, 2004.
10. Dominguez D, Tournoy J, Hartmann D, et al. Phenotypic and biochemical analyses of BACE1- and BACE2-deficient mice. *J Biol Chem*. Sep 2, 2005; 280(35):30797–806. Epub Jun 29, 2005.
11. Luo Y, Bolon B, Damore MA, et al. BACE1 (β-secretase) knockout mice do not acquire compensatory gene expression changes or develop neural lesions over time. *Neurobiol Dis*. Oct 2003;14(1):81–8.
12. Luo Y, Bolon B, Kahn S, et al. Mice deficient in BACE1, the Alzheimer's beta-secretase, have normal phenotype and abolished beta-amyloid generation. *Nat Neurosci*. 2001;4(3):231–2.
13. Roberds SL, Anderson J, Basi G, et al. BACE knockout mice are healthy despite lacking the primary beta-secretase activity in brain: implications for Alzheimer's disease therapeutics. *Hum Mol Genet*. Jun 1, 2001;10(12): 1317–24.
14. Turner RT 3rd, Koelsch G, Hong L, et al. Subsite specificity of memapsin 2 (beta-secretase): implications for inhibitor design. *Biochemistry*. Aug 28, 2001;40(34):10001–6.
15. Li X, Bo H, Zhang XC, et al. Predicting memapsin 2 (β-secretase) hydrolytic activity. *Prot Sci*. Nov 2010;19(11):2175–85.
16. Turner RT 3rd, Hong L, Koelsch G, et al. Structural locations and functional roles of new subsites S(5), S(6), and S(7) in memapsin 2 (beta-secretase). *Biochemistry*. Jan 11, 2005;44(1):105–12.
17. Lichtenthaler SF, Steiner H. Sheddases and intramembrane-cleaving proteases: RIPpers of the membrane. *EMBO Rep*. Jun 2007;8(6):537–41. Epub May 11, 2007.
18. Garratt AN, Britsch S, Birchmeier C. Neuregulin, a factor with many functions in the life of a Schwann cell. *Bioessays*. Nov 2000;22(11): 987–96.
19. Lemke G. Neuregulin-1 and myelination. *Sci STKE*. 2006;2006(325):pe11.
20. Willem M, Garratt AN, Novak B, et al. Control of peripheral nerve myelination by the {beta}-secretase BACE1. *Science*. Oct 27, 2006; 314(5799):664–6. Epub Sep 21, 2006.
21. Kim DY, Carey BW, Wang H, et al. BACE1 regulates voltage-gated sodium channels and neuronal activity. *Nat Cell Biol*. Jul 2007;9(7):755–64. Epub Jun 18, 2007.
22. Wong HK, Sakurai T, Oyama F, et al. β subunits of voltage-gated sodium channels are novel substrates of β-site amyloid precursor protein-cleaving enzyme (BACE1) and γ-secretase. *J Biol Chem*. Jun 17, 2005;280(24):23009–17. Epub Apr 11, 2005.
23. Kitazume S, Tachida Y, Oka R, et al. Characterization of alpha 2,6-sialyltransferase cleavage by Alzheimer's beta -secretase (BACE1). *J Biol Chem*. Apr 25, 2003;278(17):14865–71. Epub Dec 7, 2002.
24. Kitazume S, Tachida Y, Oka R, et al. Alzheimer's beta-secretase, beta-site amyloid precursor protein-cleaving enzyme, is responsible for cleavage secretion of a Golgi-resident sialyltransferase. *Proc Natl Acad Sci U S A*. Nov 20, 2001;98(24):13554–9. Epub Nov 6, 2001.
25. Lichtenthaler SF, Dominguez DI, Westmeyer GG, et al. The cell adhesion protein P-selectin glycoprotein ligand-1 is a substrate for the aspartyl protease BACE1. *J Biol Chem*. Dec 5, 2003;278(49):48713–9. Epub Sep 24, 2003.
26. Li Q, Sudhof TC. Cleavage of amyloid-β precursor protein and amyloid-β precursor-like protein by BACE 1. *J Biol Chem*. Mar 12, 2004; 279(11):10542–50. Epub Dec 29, 2003.
27. Pastorino L, Ikin AF, Lamprianou S, et al. BACE (β-secretase) modulates the processing of APLP2 in vivo. *Mol Cell Neurosci*. 2004;25(4): 642–9.
28. von Arnim CA, Kinoshita A, Peltan ID, et al. The low density lipoprotein receptor-related protein (LRP) is a novel beta-secretase (BACE1) substrate. *J Biol Chem*. May 6, 2005;280(18):17777–85. Epub Mar 4, 2005.
29. Kuhn PH, Marjaux E, Imhof A, et al. Regulated intramembrane proteolysis of the interleukin-1 receptor II by {alpha}-, beta-, and {gamma}-secretase. *J Biol Chem*. Apr 20, 2007;282(16):11982–95. Epub Feb 16, 2007.
30. Bloch L, Sineshchekova O, Reichenbach D, et al. Klotho is a substrate for α-, β- and γ-secretase. *FEBS Lett*. Oct 6, 2009;583(19):3221–4. Epub Sep 6, 2009.
31. Kihara T, Shimmyo Y, Akaike A, et al. Abeta-induced BACE-1 cleaves N-terminal sequence of mPGES-2. *Biochem Biophys Res Commun*. Mar 19, 2010;393(4):728–33. Epub Feb 18, 2010.
32. Hemming ML, Elias JE, Gygi SP, et al. Identification of β-secretase (BACE1) substrates using quantitative proteomics. *PLoS ONE*. Dec 29, 2009;4(12):e8477.
33. Consortium TU. The universal protein resource (UniProt) in 2010. *Nucl Acids Res*. Jan 2010;38(Database issue):D142–8. Epub Oct 20, 2009.
34. Jain E, Bairoch A, Duvaud S, et al. Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*. May 8, 2009;10:136.
35. Krogh A, Larsson B, von Heijne G, et al. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol*. 2001;305(3):567–80.
36. Emanuelsson O, Brunak S, von Heijne G, et al. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protocols*. 2007;2(4): 953–71.
37. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protocols*. 2009;4(1):44–57.
38. Hong L, Koelsch G, Lin X, et al. Structure of the protease domain of memapsin 2 (beta -secretase) complexed with inhibitor. *Science*. Oct 6, 2000;290(5489):150–3.
39. Murakami M, Nakashima K, Kamei D, et al. Cellular prostaglandin E2 production by membrane-bound prostaglandin E synthase-2 via both cyclooxygenases-1 and -2. *J Biol Chem*. Sep 26, 2003;278(39):37937–47. Epub Jun 30, 2003.
40. Lee EB, Zhang B, Liu K, et al. BACE overexpression alters the subcellular processing of APP and inhibits Aβ deposition in vivo. *J Cel Biol*. Jan 17, 2005;168(2):291–302. Epub Jan 10, 2005.

Supplementary Data

Table S1.xls